Advanced Algorithms for Multi-Antenna and Multi-Carrier Communication Systems

Dem Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

M.Sc. **Yue Rong**
geboren am 23. September 1976 in Jiangsu, VR China

Referent:	Prof. Dr. Alex B. Gershman
Korreferent:	Prof. Dr.-Ing. Martin Haardt
Tag der Einreichung:	4. Juli 2005
Tag der mündlichen Prüfung:	2. November 2005

D17
Darmstädter Dissertationen

To my wife Ye and my parents.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Alex B. Gershman for his constant support and guidance during my Ph.D. study. His encouragement, questions, comments, and suggestions have greatly influenced the contents of this thesis. I have benefited tremendously from his knowledge, vision, insights and rigor in scientific research.

I also thank Prof. Martin Haardt from Ilmenau University of Technology for his great interest in my work, and for taking the time to be the second reviewer of my dissertation.

I thank my colleague Dr. Sergiy A. Vorobyov for fruitful cooperation which led to several interesting results obtained during my Ph.D. studies. His sound mathematical background made discussions with him always beneficial. Many thanks go to my colleague and friend Dr. Yisheng Xue. His enthusiasm in research encouraged me greatly. I will miss the discussions and many constructive suggestions from him. I thank Dr. Shahram Shahbazpanahi for cooperation and advice.

I would express my gratitude to Prof. Andreas Czylwik, Dr. Thomas Kaiser, and secretary Petra Hoetger from the University of Duisburg-Essen. I am indebted for much help from them during my three years of graduate study and work at the Smart Antenna Research Team (SmART) at the Department of Communication Systems. I always feel fortunate that I have worked in such a group whose members gave me so much inspiration. I want to express my gratitude to all of them.

Prof. Nicholas D. Sidiropoulos's papers introduced me to the world of multi-way analysis. His expertise in parallel factor analysis gave me great help.

Among all, I am grateful to my parents and my wife Ye. Their patience and tremendous unconditional *love* have been the strongest support to me. Without them, this thesis would never have come to existence.

Darmstadt, Germany — Yue Rong

July 2005

Zusammenfassung

In den letzten Jahren hat sich die Nachfrage in der drahtlosen Kommunikation nach zuverlässigen Multimedia- und Daten- Übertragungsverfahren mit hohen Transferraten extrem vergrößt. Der Einsatz von Vielfachantennen- und Mehrträger- Kommunikation stellt theoretisch attraktive und technisch praktische Lösungen zur Verfügung, um den Anforderungen gerecht zu werden. Diese Arbeit hat das Ziel, Algorithmen für Vielfachantennen- und Mehrträger- Kommunikation zu beschreiben und entwerfen.

Im ersten Teil der Dissertation werden für räumliche Signaturschätzung bandbreiteneffiziente blinde Algorithmen beschrieben, die auf zeitvarianter Nutzer-Sendeleistung und paralleler Faktor-Analysis (PARAFAC) basieren. Verglichen mit früheren Verfahren, erfordern die Algorithmen keine Kenntnis des Ausbreitungskanals und/oder der Struktur des Sensor-Arrays. Des weiteren sind die Algorithmen in allgemeineren Szenarien als bereits bekannte Methoden einsetzbar.

Im zweiten Teil, basierend auf der robusten Anpassung der PARAFAC-Modelle, werden blinde Mehrbenutzer-Detektionsalgorithmen für Direktsequenz-Codedivisions-Vielfachzugriff- (DS-CDMA) Systeme eingesetzt. Verglichen mit früheren Verfahren weisen die Algorithmen eine verbesserte Bandbreiteneffizienz und Robustheit gegenüber gepulstem Umgebungsrauschen auf.

Der dritte Teil der Dissertation beschäftigt sich mit linearen Mehrbenutzerempfängern für die gemeinsame Raum-Zeit-Dekodierung und Interferenzunterdrückung in multiple-input multiple-output (MIMO) Systemen, die orthogonale Raum-Zeit-Blockcodes (OSTBCs) und fehlerhafte channel state information (CSI) verwenden. Mit unterschiedlichen CSI-Fehlanpassungs-Modellierungen werden robuste lineare Empfänger basierend auf worst-case Optimierung oder stochastic programming beschrieben. Dadurch wird die Robustheit der Kommunikationssysteme gegenüber CSI-Fehlanpassungen beträchtlich erhöht.

Der vierte Teil der Dissertation konzentriert sich auf die Verringerung der Kanal-Fading-Effekte für orthogonale Frequenzaufteilungs-Vielfachzugriff- (OFDM) Systeme. Insbesondere wird ein neuer linearer block precoder entwickelt, der auf dem Kriterium der Maximierung der mean cutoff rate basiert. Verglichen mit vorherigen Precodertechniken bietet der Precoder eine größere Effizienz und erfordert weniger Kenntnis über den Übertragungskanal am Sender. Darüber hinaus werden adaptive Verfahren für OFDM-basierende Kommunikationssysteme untersucht, bei denen der Sender nur ein Bit CSI pro Unterträger zur Verfügung hat, das er durch eine Rückkopplung mit niedriger Datenrate erhalten hat. Optimale Parametereinstellungen werden für diese adaptiven Algorithmen abgeleitet. Unvollkommenheiten der Rückkopplung werden betrachtet und ihre Auswirkung auf die Leistung der adaptiven Techniken wird untersucht.

Abstract

In recent years, the demands for reliable high rate multimedia and data transmission in wireless communications have increased tremendously. Multi-antenna and multi-carrier communications provide both theoretically attractive and technically practical solutions to satisfy these requirements. This thesis aims at designing and studying advanced algorithms for multi-antenna and multi-carrier communications.

In the first part of the thesis, we propose bandwidth-efficient blind spatial signature estimation algorithms based on time-varying user power loading and parallel factor (PARAFAC) analysis. Compared with the earlier approaches, our algorithms do not require any knowledge of the propagation channel and/or sensor array manifold and are applicable to more general class of scenarios.

In the second part of the thesis, blind multiuser separation-detection algorithms for direct-sequence code-division multiple access (DS-CDMA) systems based on the robust fitting of PARAFAC models are proposed. These algorithms provide an improved bandwidth efficiency and robustness against impulsive ambient noise as compared with the earlier approaches.

The third part of the thesis is devoted to linear multiuser receivers for joint space-time decoding and interference rejection in multiple-input multiple-output (MIMO) systems that use orthogonal space-time block codes (OSTBCs) and erroneous channel state information (CSI). Using different approaches to model the CSI mismatch, robust linear receivers based on worst-case performance optimization and stochastic programming are proposed, respectively. The proposed receivers greatly enhance the robustness of the communication systems against CSI mismatches.

The fourth part of the thesis focuses on the issue of channel fading mitigation for the orthogonal frequency division multiplexing (OFDM) communication systems. In particular, a new linear block precoding technique based on the maximization of the mean cutoff rate is developed. Compared with the earlier precoding techniques, our precoder provides better performance and requires less channel knowledge at the transmitter. Also, adaptive approaches for OFDM-based communication systems are studied in the case when the transmitter has only one bit of CSI per subcarrier obtained through a low-rate feedback. Optimal parameters for these adaptive algorithms are derived. Imperfections of the feedback channel are considered and their impact on the performance of the adaptive techniques is investigated.

Table of Contents

List of Tables

List of Figures

Acronyms

AGC	automatic gain control
AMS	adaptive modulation selection
APA	adaptive power allocation
ASCS	adaptive subcarrier selection
AWGN	additive white Gaussian noise
BER	bit error rate
bps	bits per second
BPSK	binary phase shift keying
BS	base station
CC	channel coding
CDMA	code-division multiple access
CP	cyclic prefix
CRB	Cramér-Rao bound
CSI	channel state information
DFT	discrete Fourier transform
DL	diagonal loading
DLMV	diagonal loading minimum variance
DOA	direction-of-arrival
DS	direct-sequence

ESPRIT	estimation of signal parameters via rotational invariance techniques
ETSI	European Telecommunications Standards Institute
FDD	frequency-division duplex
FDMA	frequency-division multiple access
FFT	fast Fourier transform
FH	frequency-hopping
FIM	Fisher information matrix
GAM	generalized array manifold
GSM	global system for mobile communications
HDDS	hard-decision decoding scheme
HIPERLAN/2	high performance radio local area network type 2
HOS	higher-order statistics
ICA	independent component analysis
IEEE	Institute of Electrical and Electronics Engineers
IFFT	inverse fast Fourier transform
i.i.d.	independent identically distributed
INR	interference-to-noise ratio
ISI	intersymbol interference
LAE	least absolute error
LAN	local area network
LL	log-likelihood
LP	linear programming
LS	least square
MAI	multiple-access interference
MAN	metropolitan area network
MAP	maximum a posteriori

MBER	minimum bit error rate
MF	matched filter
MIMO	multiple-input multiple-output
ML	maximum-likelihood
MMSE	minimum mean-square error
MOE	minimum output energy
MSE	mean-square error
MUSIC	multiple signal classification
MV	minimum variance
MVDR	minimum variance distortionless response
NLP	nonlinear programming
OFDM	orthogonal frequency-division multiplexing
OSTBC	orthogonal space-time block code
PARAFAC	parallel factor
PCF	power change factor
PCU	per channel use
PDF	probability density function
PEP	pairwise error probability
PSK	phase shift keying
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadriphase shift keying
RMSE	root-mean-square error
SDDS	soft-decision decoding scheme
SDMA	space-division multiple access
SER	symbol error rate

SINR	signal-to-interference-plus-noise ratio
SISO	soft-input soft-output
SNR	signal-to-noise ratio
SOC	second-order cone
SOCP	second-order cone programming
STBC	space-time block code
SVD	singular value decomposition
TALAE	trilinear alternating least absolute error
TALS	trilinear alternating least square
TDD	time-division duplex
TDMA	time-division multiple access
ULA	uniform linear array
UMTS	universal mobile telecommunications system
WLAN	wireless LAN
WMF	weighted median filtering
ZF	zero-forcing

Chapter 1

Introduction

This thesis aims at designing and studying advanced algorithms for multi-antenna and multi-carrier communication systems. In this introductory chapter, we briefly present a necessary background on multi-antenna and multi-carrier communication systems and overview the contributions of this thesis.

1.1 Background on Multi-Antenna and Multi-Carrier Communication Systems

Multi-antenna systems are used in a wide range of applications including communications, radar, sonar, seismology, biomedicine, astronomy, medical imaging, and other fields. The first application of multi-antenna techniques dates back to the second world-war (the Bartlett beamformer) [1]-[3]. Recently, in wireless communication field, multi-antenna systems attract significant interest both of academic researchers and practitioners, because multi-antenna communications are able to provide a significant increase in system capacity and coverage without additional consumption of the available radio spectrum [4]-[12], [24]-[27]. For example, multiple antennas at base stations (BSs) have already been used in wireless communication systems such as the global system for mobile communications (GSM) [13] since early 1990's to improve the signal reception at BSs [14]. The use of multiple antennas both at the transmitters and receivers is considered as one of the approaches

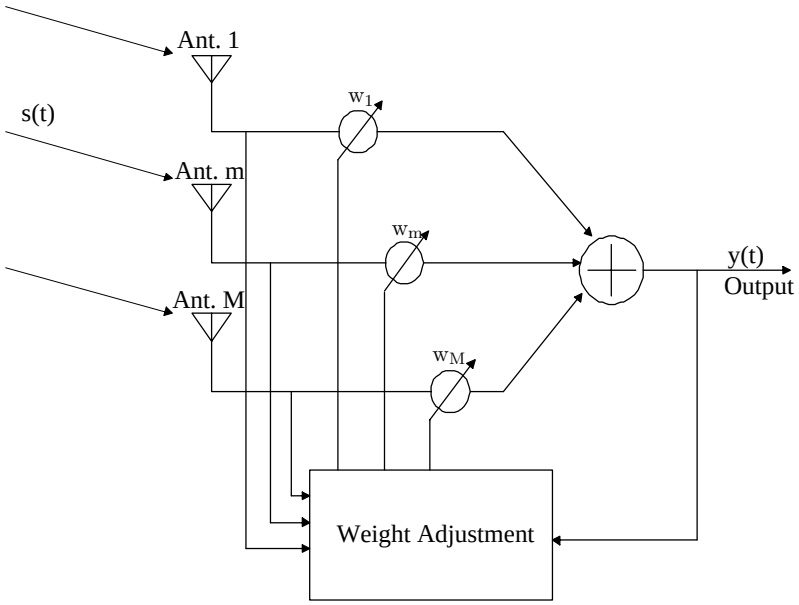

Figure 1.1: Basic diagram of narrowband adaptive multi-sensor antenna.

to improve the system throughput and spectral efficiency for 3G and further generations of wireless communications [15].

Multi-antenna communication systems take advantage of spatial filtering and spatial diversity. Spatial filtering is often called beamforming, which aims at enhancing the signal of interest and suppressing interferences based on the differences in their spatial signatures (spatial locations) [1]-[3]. Therefore, multi-antenna techniques allow different users to share the same time-frequency resources [10]. Figure 1.1 illustrates the diagram of an M-antenna narrowband receiver which performs spatial filtering of the incoming signal $s(t)$. Here, $[w_1, \cdots, w_M]^T$ is the weight vector of the spatial filter.

Besides performing spatial filtering, multiple antennas also provide spatial diversity [5]-[10]. An inherent problem in wireless communications is channel fading, which arises due to constructive or destructive effects of signals travelling through multiple paths from the transmitter to the receiver. Diversity techniques are frequently applied in wireless communications to combat fading [5]. The essence of diversity techniques is to collect signals which fade independently. Different types of diversities such as time diversity and

frequency diversity can be exploited to combat channel fading [5]. With multi-antenna receiver and/or transmitter, the signals received at each antenna experience independent fading. Therefore, spatial diversity can be applied to mitigate fading.

Multi-antenna techniques can be combined with multicarrier techniques such as orthogonal frequency-division multiplexing (OFDM) for high data rate transmission. OFDM converts a frequency-selective fading channel into parallel flat fading sub-channels through the inverse fast Fourier transform (IFFT) at the transmitter and the fast Fourier transform (FFT) at the receiver [18]-[20]. Thus, the equalizer design at the receiver is greatly simplified. Such an advantage brought by OFDM techniques is particularly important for multi-antenna communication systems, due to the fact that the number of equalizers increases proportionally to the number of receive antennas.

Another advantage of OFDM is that the subcarriers have the minimum frequency separation required to maintain orthogonality of their corresponding time domain waveforms, yet the signal spectra corresponding to different subcarriers overlap in frequency [18]. Therefore, the available bandwidth is used very efficiently. Due to the merits mentioned above, OFDM has been adopted in many standards, for example, IEEE802.11a LAN, IEEE802.16a (WiMAX), and ETSI HIPERLAN/2 [21]-[23].

1.2 Thesis Overview and Contributions

In this thesis, advanced algorithms for multi-antenna and multi-carrier communications are presented and studied. In Chapter 2, we address the blind spatial signature estimation problem and develop bandwidth-efficient algorithms based on time-varying user power loading and parallel factor (PARAFAC) analysis. Chapter 3 studies blind direct-sequence code-division multiple access (DS-CDMA) multiuser separation-detection in impulsive ambient noise and proposes algorithms based on the robust fitting of PARAFAC models. Robust linear receiver design problem for multiple-access space-time block coded multiple-input multiple-output (MIMO) system is investigated in Chapter 4. Using different approaches to model the channel state information (CSI) mismatch, we develop robust linear receivers

based on worst-case performance optimization and stochastic programming, respectively. In Chapter 5, we propose a cutoff rate based linear block precoding technique for OFDM communication systems. The proposed technique only requires the knowledge of the average relative channel multipath powers and delays. Chapter 6 studies the performance of adaptive OFDM systems with one-bit-per-subcarrier channel state feedback. Chapter 7 summarizes the thesis and gives the outlook to some future work.

Chapter 2: Blind Spatial Signature Estimation

In multiple-access communication systems, signals from different users can be separated at the receive antenna array based on the knowledge of their spatial signatures [28]-[31]. However, user spatial signatures are typically unknown at the receiver and, therefore, have to be estimated. We develop a new approach to blind spatial signature estimation using PARAFAC analysis [43]-[46]. Compared with the existing methods for blind spatial signature estimation, for example [28], [31]-[34], our approach is bandwidth-efficient and does not require any restrictive assumptions on the array geometry and the propagation environment.

Chapter 2 is based on the journal publication:

- Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Signal Processing*, vol. 53, pp. 1697-1710, May 2005.

and two conference publications:

- Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation using time-varying user power loading and parallel factor analysis," in *Proc. 58th IEEE Vehicular Technology Conference (VTC)*, Orlando, USA, Oct. 4-9, 2003, vol. 1, pp. 79-83.

- Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Deterministic Cramér-Rao bound for symmetric PARAFAC model with application to blind spatial signature estimation," in *Proc. 3rd IEEE Int. Symposium on Signal Processing and*

Information Technology (ISSPIT), Darmstadt, Germany, Dec. 14-17, 2003, pp. 411-414.

Chapter 3: Blind Multiuser Detection in Impulsive Noise

In many multiuser detection techniques, the channel noise is assumed to be Gaussian [16]. However, the Gaussian noise assumption has often been proven inadequate due to the significant impulsive nature of the channel noise [88], [89]. This implies that a robust multiuser detection algorithm should take into account the impulsive feature of the channel noise. In this chapter, we propose a new blind DS-CDMA multiuser separation-detection technique which makes use of the receive antenna array. We link this multiuser detection problem to the PARAFAC model. Iterative algorithms for robust PARAFAC model fitting under impulsive noise are proposed. Compared with conventional DS-CDMA multiuser detection algorithms [44], our approaches achieve significant performance improvement in impulsive noise with only a moderate performance degradation under Gaussian noise.

The material in Chapter 3 is based on the journal publication:

- S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Processing*, vol. 53, pp. 2678-2689, Aug. 2005.

and two conference publications:

- S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models based on linear programming," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Quebec, Canada, May 17-21, 2004, vol. 2, pp. 113-116.

- S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust fitting of multilinear models with application to blind multiuser receivers: Iterative weighted median filtering approach," in *Proc. 5th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Lisbon, Portugal, July 11-14, 2004, pp. 478-482.

Chapter 4: Robust Linear Receivers for MIMO Systems

If both the transmitter and the receiver have multiple antennas, then a MIMO system arises naturally. In this case, space-time block codes (STBCs) can be used as a powerful approach to exploit spatial diversity and combat fading [5]-[10], [95]-[97]. Recently, several linear receivers have been proposed for joint space-time decoding and interference rejection [99]-[103]. These receiver schemes provide good trade-offs between the complexity and performance. However, a common shortcoming of these linear receivers is that they use the assumption that the exact CSI is available at the receiver side. However, in practice the exact CSI is difficult to obtain, that is, there is always a mismatch between the exact and presumed CSI. In this chapter, new robust linear receivers are proposed for joint space-time decoding and interference rejection in the case of erroneous CSI. Using different approaches to model the CSI mismatch, we design robust linear receivers based on worst-case performance optimization and stochastic programming, respectively. The proposed receivers are shown to provide a substantially improved robustness against CSI errors as compared with the existing linear multiple-access MIMO receivers.

Chapter 4 is based on the journal publication:

- Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multiaccess MIMO systems with imperfect channel state information," *IEEE Trans. Signal Processing*, vol. 53, pp. 3081-3090, Aug. 2005.

and four conference publications:

- Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multiple-access MIMO wireless systems," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Quebec, Canada, May 17-21, 2004, vol. 2, pp. 9-12.

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "A robust linear receiver for multi-access space-time block coded MIMO systems based on probability constrained optimization," in *Proc. 59th IEEE Vehicular Technology Conference (VTC)*, Milan, Italy, May 17-19, 2004, vol. 1, pp. 118-122.

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "A robust linear receiver for uplink multi-user MIMO systems based on probability-constrained optimization and second-order cone programming," in *Proc. 3rd IEEE Signal Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, July 18-21, 2004, pp. 153-157.

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receiver design for multi-access space-time block coded MIMO systems using stochastic optimization," in *Proc. 13th IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, July 17-20, 2005.

Chapter 5: Linear Block Precoding for OFDM Systems

OFDM is a promising communication scheme which facilitates the equalizer design at the receiver. However, a well known disadvantage of this scheme is that, at each subcarrier, the channel may be subject to a deep fading. To mitigate the channel fading, we propose a new linear precoding technique based on the mean cutoff rate maximization criterion. Compared with other precoding techniques, for example [126]-[129], which need the full CSI at the transmitter, the proposed technique only requires the knowledge of the average relative channel multipath powers and delays. The combination of the proposed precoding scheme with error-correcting coding techniques is studied. Simulation results show an improved performance of our precoding approach.

The material in Chapter 5 is based on the journal publication:

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Linear block precoding for OFDM systems based on maximization of mean cutoff rate," *IEEE Trans. Signal Processing*, accepted.

and two conference publications:

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Linear OFDM precoder design for multiuser wireless communications using cutoff rate optimization," in *Proc. 12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, Sep. 6-10, 2004, pp. 2071-2074.

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Combining error-correction coding and cutoff rate maximization based precoding", in *Proc. Int. ITG/IEEE Workshop on Smart Antennas (WSA)*, Duisburg, Germany, April 4-5, 2005.

Chapter 6: Adaptive OFDM with Channel State Feedback

Adaptive techniques can be applied to mitigate channel fading in OFDM communication systems if CSI is available at the transmitter. We study the performance of an OFDM-based communication system whose transmitter has only one bit of CSI per subcarrier obtained through a low-rate feedback. Three adaptive approaches are considered to exploit such a CSI feedback: adaptive subcarrier selection (ASCS), adaptive power allocation (APA), and adaptive modulation selection (AMS). It is shown that one bit CSI feedback can greatly enhance the system performance. Imperfections of the feedback channel are considered and their impact on the performance of the APA and AMS techniques is studied. We show that exploiting the knowledge that the feedback channel is imperfect, the performance of these adaptive techniques can be substantially improved.

Chapter 6 is based on the journal submission:

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Adaptive OFDM techniques with one-bit-per-subcarrier channel state feedback," *IEEE Trans. Communications*, vol. 53, Dec. 2005.

and two conference publications:

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "The impact of imperfect one bit per subcarrier channel state information feedback on adaptive OFDM wireless communication systems," in *Proc. 60th IEEE Vehicular Technology Conference (VTC)*, Los Angeles, CA, USA, Sep. 26-29, 2004, vol. 1, pp. 626-630.

- Y. Rong, S. A. Vorobyov, and A. B. Gershman, "On average one bit per subcarrier channel state information feedback in OFDM wireless communication systems" in *Proc. IEEE Global Telecommunication Conference (GLOBECOM)*, Dallas, Texas, USA, Nov. 29-Dec. 3, 2004, vol. 6, pp. 4011-4015.

1.3 Notations

We use the following common notations. Lower case letters are used to denote scalars. Bold face lower case letters denote vectors. Bold face upper case letters denote matrices. For matrices, $(\cdot)^T, (\cdot)^*, (\cdot)^H, (\cdot)^{-1}$, and $(\cdot)^\dagger$ denote transpose, conjugate, Hermitian transpose, inverse, and pseudo inverse operations, respectively. rank$\{\cdot\}$ and tr$\{\cdot\}$ denote the rank and trace, respectively. For a matrix $\boldsymbol{A}$, $k_{\boldsymbol{A}}$ stands for the Kruskal rank. $\overline{(\cdot)}$ denotes a three-dimensional array, E$\{\cdot\}$ represents the statistical expectation, $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix, $\|\cdot\|_1$ denotes the ℓ_1 norm of a real-valued vector, and $\mathcal{D}_p\{\cdot\}$ is the operator which makes a diagonal matrix by selecting the pth row of a matrix and putting it on the main diagonal while putting zeros elsewhere. For complex numbers and matrices, Re$\{\cdot\}$ and Im$\{\cdot\}$ denote the real and imaginary parts, respectively. $\otimes, \odot$, and $\circ$ represent the Kronecker, Khatri-Rao, and Schur-Hadamard matrix products, respectively, Pr$\{\cdot\}$ denotes the probability operator, and $\mathcal{CN}(\cdot,\cdot)$ stands for complex Gaussian distribution. An N-dimensional identity matrix is denoted as $\boldsymbol{I}_N$, $\boldsymbol{0}_{M\times N}$ is denoted as an $M \times N$ matrix with all zero entries, and $\succeq$ stands for the point-wise ordering.

Chapter 2

Blind Spatial Signature Estimation

In this chapter, the problem of blind spatial signature estimation is addressed. After a review of the traditional (non-blind) and blind approaches to spatial signature estimation in Section 2.1, the data model is introduced in Section 2.2. In order to make the model identifiable, a time-varying user power loading method in the uplink mode is proposed in Section 2.3 that enables to use the PARAFAC analysis to blindly estimate the spatial signature. Then identifiability issues are studied in detail in Section 2.4. Two PARAFAC spatial signature estimators are presented in Section 2.5. The first technique is based on the trilinear alternating least squares (TALS) regression procedure, while the second one makes use of the joint approximate diagonalization algorithm. These techniques do not require any knowledge of the propagation channel and/or sensor array manifold and are applicable to a more general class of scenarios than earlier approaches to blind spatial signature estimation. In Section 2.6 we derive the modified Cramér-Rao bound (CRB) which serves as a benchmark of the problem at hand. Simulation results are presented in Section 2.7. Section 2.8 briefly summarizes the chapter. Detailed derivations of the modified CRB are listed in Section 2.A.

2.1 Existing Spatial Signature Estimation Techniques

In a multiple-access communication system, signals from different users can be separated at the receive antenna array based on the knowledge of their spatial signatures [28]-[31]. In

particular, known spatial signatures can be used for beamforming to separate each user-of-interest from the other (interfering) users. However, user spatial signatures are usually unknown at the receiver and, therefore, have to be estimated.

Traditional (non-blind) approaches to spatial signature estimation make use of training sequences which are periodically transmitted by each user and are known at the receiver [29]. However, the use of training sequences reduces the information transmission rate, and strict coordination of the training epochs of several users in a multiuser setting requires tight synchronization. As a result, blind spatial signature estimation techniques have attracted a significant attention in the literature [31]-[39].

There are several blind approaches to spatial signature estimation. The most common one is based on the parametric modelling of spatial signatures using direction-of-arrival (DOA) parameters [28], [31], [32]. For example, in [28] the coherently distributed source model is used to parameterize the spatial signature. Unfortunately, the source angular spread should be small for the first-order Taylor series expansion used in [28] to be valid. This is a limitation for mobile communications applications in urban environments with low base station antenna mast heights, where angular spreads up to 25° are typically encountered [40], [41]. Furthermore, the approach of [28] requires precise array calibration.

Two other DOA-based blind spatial signature estimation methods are developed in [31] and [32]. In these papers, the source spatial signature is modelled as a plane wave distorted by unknown direction-independent gains and phases. The latter assumption can be quite restrictive in wireless communications where spatial signatures may have an arbitrary form and, therefore, such gains and phases should be modelled as DOA-dependent quantities. As a result, the techniques of [31] and [32] are applicable to a particular class of scenarios only.

Another popular approach to blind spatial signature estimation makes use of the cyclostationary nature of communication signals [33], [34]. This approach does not make use of any DOA-based model of spatial signatures but it is applicable only to users which all have different cyclic frequencies. The latter condition implies that the users must have different carrier frequencies (which is not the case for space-division multiple access – SDMA) and/or

baud rates [34]. This can limit practical applications of the methods of [33] and [34].

Another well-known approach to this problem employs higher-order statistics (cumulants) to estimate spatial signatures in a blind way [35]-[39]. Cumulant-based methods are only applicable to non-Gaussian signals. Moreover, all such algorithms are restricted by the requirement of a large number of snapshots. This requirement is caused by a slow convergence of sample estimates of higher-order cumulants.

The aforementioned restrictions of cumulant-based methods have been a strong motivation for further attempts to develop blind spatial signature estimators which are based on second-order statistics only and which do not require any DOA-related or cyclostationarity assumptions. In [38], such a method was proposed using joint approximate diagonalization of a set of spatial auto- and cross-covariance matrices. This method requires an existence of a long-time coherence of the source signals to obtain enough cross-covariance matrices at multiple lags for the joint diagonalization process and to guarantee identifiability. In practical wireless communication systems, the signal time coherence is severely limited, i.e., the correlation time of the received signals typically does not largely exceed the sampling interval. For example, communication signals sampled at the symbol rate are uncorrelated[1] and, hence, higher-lag correlations are all zero. In such cases, multiple covariance matrices are unavailable and the method of [38] is not applicable. Furthermore, [38] offers limited identifiability – for example, it requires that the matrix of spatial signatures be full column rank and, therefore, the number of sources should be less or equal to the number of antennas.

In this chapter, we develop a new bandwidth-efficient approach to blind spatial signature estimation using PARAFAC analysis [43]-[46]. Our approach does not require any restrictive assumptions on the array geometry and the propagation environment. Time-varying user power loading is exploited to obtain multiple spatial zero-lag covariance matrices required for the PARAFAC model.

[1]Channel-coded signals, which include redundancy for error correction, are in fact interleaved before transmission, with the goal of making the transmitted signal approximately uncorrelated.

Blind PARAFAC multi-sensor reception and steering vector estimation have been considered earlier in [44] and [46]. However, the approach of [44] is applicable to DS-CDMA systems only, as spreading sequence is explicitly used as the third dimension of the data array; while [46] requires multiple shifted but otherwise identical subarrays, and a DOA parameterization. Below, we show that the proposed user power loading enables us to give up the CDMA and multiple-invariance/DOA parameterization assumptions, and extend the blind approach to any type of SDMA system employing multiple antennas at the receiver.

Blind source separation of non-stationary sources using multiple covariance matrices has also been considered in [47] but, again, under limited identifiability conditions stemming from the usual ESPRIT-like solution. Our identifiability results are considerably more general as they do not rely on this limited viewpoint.

2.2 Data Model

Let an array of M sensors receive the narrowband signals from P users. We assume that the observation interval is shorter than the coherence time of the channel (i.e., the scenario is time-invariant) and the time dispersion introduced by the multipath propagation is small in comparison with the reciprocal of the bandwidth of the emitted signals [28]. Under such assumptions, the $M \times 1$ snapshot vector of antenna array outputs can be written as [28]

$$\boldsymbol{y}(n) = \boldsymbol{A}\boldsymbol{s}(n) + \boldsymbol{v}(n) \tag{2.1}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_P] \in \mathbb{C}^{M \times P}$ is the matrix of the user spatial signatures, $\boldsymbol{a}_p = [a_{1,p}, \ldots, a_{M,p}]^T \in \mathbb{C}^{M \times 1}$ is the spatial signature of the pth user, $\boldsymbol{s}(n) = [s_1(n), \ldots, s_P(n)]^T \in \mathbb{C}^{P \times 1}$ is the vector of the equivalent baseband user waveforms, $\boldsymbol{v}(n) = [v_1(n), \ldots, v_M(n)]^T \in \mathbb{C}^{M \times 1}$ is the vector of additive spatially and temporally white Gaussian noise. Note that, in contrast to direction finding problems, the matrix $\boldsymbol{A}$ is unstructured. Assuming that there is a block of J snapshots available, the model (2.1) can be written as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{V} \tag{2.2}$$

where $\boldsymbol{Y} = [\boldsymbol{y}(1), \ldots, \boldsymbol{y}(J)] \in \mathbb{C}^{M \times J}$ is the array data matrix, $\boldsymbol{S} = [\boldsymbol{s}(1), \ldots, \boldsymbol{s}(J)] \in \mathbb{C}^{P \times J}$ is the user waveform matrix, and $\boldsymbol{V} = [\boldsymbol{v}(1), \ldots, \boldsymbol{v}(J)] \in \mathbb{C}^{M \times J}$ is the sensor noise matrix. A quasi-static channel is assumed throughout the chapter. This assumption means that the spatial signatures are block time-invariant (i.e., the elements of $\boldsymbol{A}$ remain constant over a block of J snapshots).

Assuming that the user signals are uncorrelated with each other and sensor noise, the array covariance matrix of the received signals can be written as

$$\boldsymbol{R} \triangleq \mathrm{E}\{\boldsymbol{y}(n)\boldsymbol{y}^H(n)\} = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{A}^H + \sigma_v^2 \boldsymbol{I}_M \tag{2.3}$$

where $\boldsymbol{Q} \triangleq \mathrm{E}\{\boldsymbol{s}(n)\boldsymbol{s}^H(n)\}$ is the diagonal covariance matrix of the signal waveforms, σ_v^2 is the sensor noise variance.

The problem studied here is the estimation of the matrix $\boldsymbol{A}$ from noisy array observations $\boldsymbol{Y}$.

2.3 Time-Varying User Power Loading Method

Before proceeding, we need to clarify that by *identifiability* we mean the uniqueness (up to inherently unresolvable source permutation and scale ambiguities) of all user spatial signatures given the exact covariance data. Identifiability in this sense is impossible to achieve with only one known covariance matrix (2.3) because the matrix $\boldsymbol{A}$ can be estimated from $\boldsymbol{R}$ only up to an arbitrary unknown unitary matrix [45]. The approach we will use to provide a unique user spatial signature estimation is based on an artificial user power loading and PARAFAC model analysis. Therefore, next we explain how this model is related to our problem.

Let us divide uniformly the whole data block of J snapshots into K sub-blocks, so that each sub-block contains $N_{\mathrm{s}} = \lfloor \frac{J}{K} \rfloor$ snapshots, where $\lfloor x \rfloor$ denotes the largest integer less than x. We fix the transmit power of each user within each sub-block while changing it artificially[2]

[2]Note that the effect of time-varying user powers has been exploited in [47] where an ESPRIT-type algorithm has been proposed for blind source separation of nonstationary sources. Similar ideas have been used in [38] and [48]. However, the authors of [38], [47] and [48] assume that the source powers vary because of signal nonstationarity rather than artificial power loading.

between different sub-blocks. It should be stressed that the proposed artificial time-varying user power loading does not require precise synchronization among the users, but the users should roughly know the boundaries of epochs over which the powers are kept constant (this can be achieved, for example, using the standard power control feedback channel). Therefore, a certain level of user coordination is required from the transmitter side[3]. We stress that the proposed user power loading can be easily implemented by overlaying a small power variation on top of the usual power control, without any other modifications to existing hardware, or communication system/network parameters. Also, as it will be seen in the sequel, the user powers need not vary much to enable blind identification. In particular, power variations that will be used are in the order of 30%. Such power variations will not significantly affect the bit error rate (BER) which is seriously affected only when order-of-magnitude power variations are encountered.

If power control is fast enough (in the sense that there are several power changes per channel coherence dwell), we can exploit it as a sort of user power loading. However, power control is usually much slower than the channel coherence time, because its purpose is to combat long-term shadowing. For this reason, in practice it may not be possible to rely on the power control variations, and we need to induce a faster (but much smaller in magnitude) power variation on top of power control. This extra power variation need not "follow the channel", i.e., it can be pseudo-random and, hence, the channel need not be measured any faster than required for regular power control.

Using the proposed power loading, the received snapshots within any kth sub-block correspond to the following covariance matrix

$$\boldsymbol{R}(k) = \boldsymbol{A}\boldsymbol{Q}(k)\boldsymbol{A}^H + \sigma_v^2 \boldsymbol{I}_M \tag{2.4}$$

where $\boldsymbol{Q}(k)$ is the diagonal covariance matrix of the user waveforms in the kth sub-block. Using all K sub-blocks, we will have K different covariance matrices $\{\boldsymbol{R}(1), \ldots, \boldsymbol{R}(K)\}$. Note that these matrices differ from each other only because the signal waveform covariance matrices $\boldsymbol{Q}(k)$ differ from one sub-block to another.

[3]As it will be seen from our simulations, the methods proposed in this chapter will work well even in the case when there is no user coordination (i.e., in the unsynchronized user case).

In practice, the noise power can be estimated and then subtracted from the covariance matrix (2.4). Let us stack the K matrices $\boldsymbol{R}(k) - \sigma_v^2 \boldsymbol{I}_M$, $k = 1, \ldots, K$ together to form a three-way array $\overline{\boldsymbol{R}}$, which is natural to call the *covariance array*. The (i, l, k)th element of such an array can be written as

$$r_{i,l,k} \triangleq \left[\overline{\boldsymbol{R}}\right]_{i,l,k} = \sum_{p=1}^{P} a_{i,p} \nu_p(k) a_{l,p}^* \tag{2.5}$$

where $\nu_p(k) \triangleq [\boldsymbol{Q}(k)]_{p,p}$ is the power of the pth user in the kth sub-block. Defining the matrix $\boldsymbol{P} \in \mathbb{R}^{K \times P}$ as

$$\boldsymbol{P} \triangleq \begin{bmatrix} \nu_1(1) & \ldots & \nu_P(1) \\ \vdots & \ddots & \vdots \\ \nu_1(K) & \ldots & \nu_P(K) \end{bmatrix}$$

we can write the following relationship between $\boldsymbol{Q}(k)$ and $\boldsymbol{P}$

$$\boldsymbol{Q}(k) = \mathcal{D}_k\{\boldsymbol{P}\} \tag{2.6}$$

for all $k = 1, \ldots, K$. In (2.6), $\mathcal{D}_k\{\cdot\}$ is the operator which makes a diagonal matrix by selecting the kth row and putting it on the main diagonal while putting zeros elsewhere. The structure of the three-dimensional covariance array $\overline{\boldsymbol{R}}$ is shown in Figure 2.1.

Equation (2.5) implies that $r_{i,l,k}$ is a sum of rank-1 triple products. If P is sufficiently small[4], equation (2.5) represents a low-rank decomposition of $\overline{\boldsymbol{R}}$. Therefore, the problem of spatial signature estimation can be reformulated as the problem of low-rank decomposition of the three-way covariance array $\overline{\boldsymbol{R}}$.

2.4 Identifiability of PARAFAC Model

In this section, we give a brief review on PARAFAC model and study the identifiability of the PARAFAC model-based spatial signature estimation. Towards this end, we discuss conditions under which the trilinear decomposition of $\overline{\boldsymbol{R}}$ is unique. Identifiability conditions on the number of sub-blocks and the number of array sensors are derived.

[4]Exact conditions for P are given in the next section.

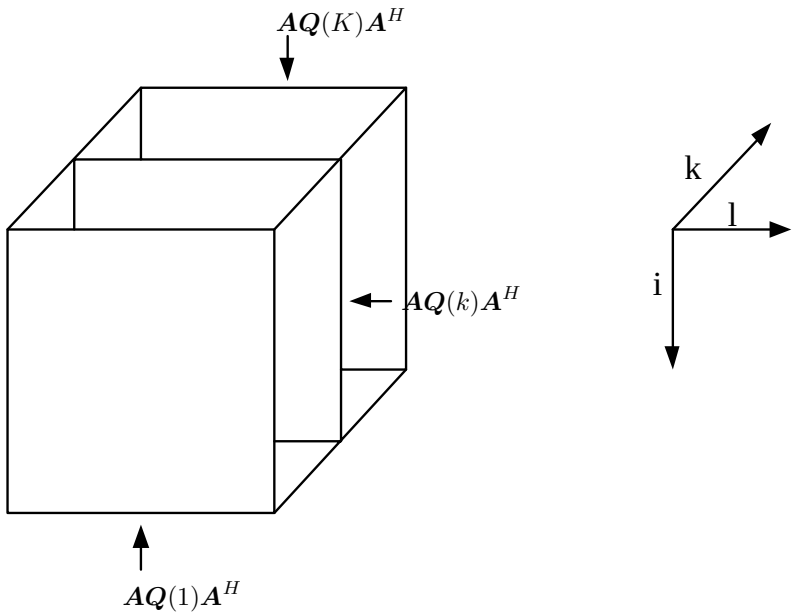

Figure 2.1: Three-dimensional covariance array.

The PARAFAC model [42], [43] is a useful data analysis tool that has recently found applications in communications and array signal processing, e.g., [44], [46]. Generalizing the concept of low-rank decomposition to higher-way arrays or *tensors*, PARAFAC is instrumental in the analysis of data arrays indexed by three or more independent variables, just like singular value decomposition (SVD) is instrumental in ordinary matrix (two-way array) analysis. Unlike SVD, PARAFAC does not impose orthogonality constraints; the reason is that low-rank decomposition of higher-order tensorial data is essentially unique under certain relatively mild conditions [43], [44], in contrast to low-rank matrix decomposition.

Because of its direct link to low-rank decomposition, PARAFAC analysis has found applications in numerous and diverse disciplines, e.g., cf. [44], [46] and references therein. Related work on joint diagonalization, symmetric, super-symmetric, and rank-one tensorial decomposition, has also appeared in the signal processing literature, mostly in the context of higher-order statistics (HOS) and independent component analysis (ICA)-based blind source separation [75]-[78].

In order to study the identifiability of PARAFAC model, let us first define the *Kruskal*

rank of a matrix [43].

Definition 2.1. *The Kruskal rank (or k-rank) of a matrix* $\boldsymbol{C}$ *is* $k_{\boldsymbol{C}}$ *if and only if every* $k_{\boldsymbol{C}}$ *columns of* $\boldsymbol{C}$ *are linearly independent, and either* $\boldsymbol{C}$ *has* $k_{\boldsymbol{C}}$ *columns or* $\boldsymbol{C}$ *contains a set of* $k_{\boldsymbol{C}}+1$ *linearly dependent columns. Note that k-rank is always less than or equal to the conventional matrix rank. It can be easily checked that if* $\boldsymbol{C}$ *is full column rank, then it is also full k-rank.*

Using (2.6) and assuming that the noise term is subtracted from the matrix $\boldsymbol{R}(k)$, we can rewrite (2.4) as

$$\boldsymbol{R}(k) = \boldsymbol{A}\mathcal{D}_k(\boldsymbol{P})\boldsymbol{A}^H \tag{2.7}$$

for all $k = 1, \ldots, K$. Let us introduce the matrix

$$\boldsymbol{R}_a \triangleq \begin{bmatrix} \boldsymbol{A}\mathcal{D}_1(\boldsymbol{P})\boldsymbol{A}^H \\ \boldsymbol{A}\mathcal{D}_2(\boldsymbol{P})\boldsymbol{A}^H \\ \vdots \\ \boldsymbol{A}\mathcal{D}_K(\boldsymbol{P})\boldsymbol{A}^H \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}\mathcal{D}_1(\boldsymbol{P}) \\ \boldsymbol{A}\mathcal{D}_2(\boldsymbol{P}) \\ \vdots \\ \boldsymbol{A}\mathcal{D}_K(\boldsymbol{P}) \end{bmatrix} \boldsymbol{A}^H = (\boldsymbol{P} \odot \boldsymbol{A})\boldsymbol{A}^H \tag{2.8}$$

where $\odot$ is the Khatri-Rao (column-wise Kronecker) matrix product [46].

To establish identifiability, we have to obtain under which conditions the decomposition (2.8) of the matrix $\boldsymbol{R}_a$ via matrices $\boldsymbol{P}$ and $\boldsymbol{A}$ is unique (up to the scaling and permutation ambiguities). In [43], the uniqueness of trilinear decomposition for the case of real-valued arrays has been established. These results have been later extended to the complex-valued matrix case [44]. In the context of our present application, which involves a conjugate-symmetric PARAFAC model, the results of [43] and [44] specialize to the following theorem. See also [51] for a discussion of the corresponding real-symmetric model.

Theorem 2.1. *Consider the set of matrices (2.7). If for* $P > 1$

$$k_{\boldsymbol{A}} + k_{\boldsymbol{P}} + k_{\boldsymbol{A}^*} = 2k_{\boldsymbol{A}} + k_{\boldsymbol{P}} \geq 2P + 2 \tag{2.9}$$

then $\boldsymbol{A}$ *and* $\boldsymbol{P}$ *are unique up to inherently unresolvable permutation and scaling of columns, i.e., if there exists any other pair* $\{\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{P}}\}$ *which satisfies (2.9), then this pair is related to the pair* $\{\boldsymbol{A}, \boldsymbol{P}\}$ *via*

$$\widetilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{\Omega}_1, \quad \widetilde{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{\Pi}\boldsymbol{\Omega}_2$$

where $\boldsymbol{\Pi}$ *is a permutation matrix, and* $\boldsymbol{\Omega}_1$ *and* $\boldsymbol{\Omega}_2$ *are diagonal scaling matrices satisfying*

$$\boldsymbol{\Omega}_1\boldsymbol{\Omega}_1^*\boldsymbol{\Omega}_2 = \boldsymbol{I}_P\,.$$

For $P = 1$, $\boldsymbol{A}$ *and* $\boldsymbol{P}$ *are always unique, irrespective of (2.9).*

PROOF: See [43], [44] and [51]. □

Note that the scaling ambiguity can be easily avoided by taking one of the array sensors as a reference and normalizing user spatial signatures with respect to it. The permutation ambiguity is unremovable but it is usually immaterial because typically the ordering of the estimated spatial signatures is unimportant.

It is worth noting that condition (2.9) is sufficient for identifiability, and is necessary only if $P = 2$ or $P = 3$, but is not necessary if $P \geq 4$ [50]. Furthermore, for $P > 1$ the condition $k_{\boldsymbol{P}} \geq 2$ becomes necessary [49]. In terms of the number of sub-blocks the latter condition requires that

$$K \geq 2\,.$$

The practical conclusion is that in the multiuser case, not less than two covariance matrices must be collected to uniquely identify $\boldsymbol{A}$ which means that the users have to change their powers at least once during the transmission. Similarly, it is necessary that $M > 1$.

The following result gives sufficient conditions for the number of sensors to guarantee almost sure-identifiability[5].

Theorem 2.2. *Suppose that*

[5]The definition of almost-sure identifiability in the context discussed is given in [52].

- *The elements of* $\boldsymbol{A}$ *are drawn from distribution* $P_{\mathcal{L}}(\mathbb{C}^{MP})$*, which is assumed continuous with respect to the Lebesgue measure in* $\mathbb{C}^{MP}$*;*

- *The elements of* $\boldsymbol{P}$ *are drawn from distribution* $P_{\mathcal{L}}(\mathbb{R}^{KP})$*, which is assumed continuous with respect to the Lebesgue measure in* $\mathbb{R}^{KP}$*.*

Then

- *For* $1 < P \leq K$*, the value of*

$$M \geq \frac{P+2}{2} \tag{2.10}$$

is sufficient for almost-sure identifiability.

- *For* $P > K$ *and* $K \geq 2$*, the value of*

$$M \geq \frac{2P+2-K}{2} \tag{2.11}$$

is sufficient for almost-sure identifiability.

PROOF: The assumptions of Theorem 2.2 mean that the following equalities hold almost surely [52]

$$k_{\boldsymbol{A}} = \text{rank}\{\boldsymbol{A}\} = \min(M, P) \tag{2.12}$$

$$k_{\boldsymbol{P}} = \text{rank}\{\boldsymbol{P}\} = \min(K, P)\,. \tag{2.13}$$

Substituting (2.12) and (2.13) into (2.9), we have

$$2\min(M, P) + \min(K, P) \geq 2P + 2\,. \tag{2.14}$$

The following cases should be considered:

1. $M \geq P$. In this case, $k_{\boldsymbol{A}} = P$. Furthermore, as $K \geq 2$, we have that $k_{\boldsymbol{P}} \geq 2$. Therefore, condition (2.14) is always satisfied.

2. $M < P$; $P \leq K$. In this case, $k_{\boldsymbol{A}} = M$, $k_{\boldsymbol{P}} = P$ and condition (2.14) becomes

$$2M + P \geq 2P + 2\,.$$

This inequality is equivalent to (2.10).

3. $M < P$; $P > K$. In this case, $k_{\boldsymbol{A}} = M$, $k_{\boldsymbol{P}} = K$ and (2.14) can be written as

$$2M + K \geq 2P + 2\,.$$

This inequality is equivalent to (2.11). □

2.5 Estimators

We will now develop two techniques for blind spatial signature estimation based on the PARAFAC model of Section 2.3.

In practice, the exact covariance matrices $\boldsymbol{R}(k)$ are unavailable but can be estimated from the array snapshots $\boldsymbol{y}(n)$, $n = 1, \ldots, J$. The sample covariance matrices are given by

$$\hat{\boldsymbol{R}}(k) = \frac{1}{N_{\mathrm{s}}} \sum_{n=(k-1)N_{\mathrm{s}}+1}^{kN_{\mathrm{s}}} \boldsymbol{y}(n)\boldsymbol{y}^H(n), \qquad k = 1, \ldots, K\,.$$

These matrices can be used to form a sample three-way covariance array denoted as $\hat{\overline{\boldsymbol{R}}}$.

If $M > P$ then the noise power σ_v^2 can be estimated as the average of the smallest $M - P$ eigenvalues of the matrix

$$\tilde{\boldsymbol{R}} = \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{R}}(k) = \frac{1}{J} \sum_{n=1}^{J} \boldsymbol{y}(n)\boldsymbol{y}^H(n)$$

and the estimated noise component $\hat{\sigma}_v^2 \boldsymbol{I}_M$ can be subtracted from sub-blocks of the sample covariance array $\hat{\overline{\boldsymbol{R}}}$. In case $M \leq P$, noise power can be estimated on system start-up, before any transmission begins.

To formulate our techniques, we will need "slices" of the matrices $\overline{\boldsymbol{R}}$ and $\hat{\overline{\boldsymbol{R}}}$ along different dimensions [44]. Towards this end, let us define the "slice" matrices

$$\begin{aligned}\boldsymbol{R}_a^{(k)} &\triangleq [r_{:,:,k}]\\ \boldsymbol{R}_b^{(l)} &\triangleq [r_{:,l,:}]\\ \boldsymbol{R}_c^{(i)} &\triangleq [r_{i,:,:}]\end{aligned}$$

where $i,l = 1,\ldots,M$; $k = 1,\ldots,K$; and $r_{i,l,k} \triangleq \left[\overline{\boldsymbol{R}}\right]_{i,l,k}$. Similarly,

$$\begin{aligned}\hat{\boldsymbol{R}}_a^{(k)} &\triangleq [\hat{r}_{:,:,k}]\\ \hat{\boldsymbol{R}}_b^{(l)} &\triangleq [\hat{r}_{:,l,:}]\\ \hat{\boldsymbol{R}}_c^{(i)} &\triangleq [\hat{r}_{i,:,:}]\end{aligned}$$

where $i,l = 1,\ldots,M$; $k = 1,\ldots,K$; and $\hat{r}_{i,l,k} \triangleq \left[\hat{\overline{\boldsymbol{R}}}\right]_{i,l,k}$.

For the sake of convenience, let us introduce $\boldsymbol{K} \triangleq \boldsymbol{A}^H$ and rewrite (2.8) as

$$\boldsymbol{R}_a \triangleq \begin{bmatrix}\boldsymbol{R}_a^{(1)}\\ \boldsymbol{R}_a^{(2)}\\ \vdots\\ \boldsymbol{R}_a^{(K)}\end{bmatrix} = (\boldsymbol{P}\odot\boldsymbol{A})\boldsymbol{K}\,.$$

In the same way, let us define the matrices

$$\begin{aligned}\boldsymbol{R}_b &\triangleq \begin{bmatrix}\boldsymbol{R}_b^{(1)}\\ \boldsymbol{R}_b^{(2)}\\ \vdots\\ \boldsymbol{R}_b^{(M)}\end{bmatrix} = (\boldsymbol{K}^T\odot\boldsymbol{P})\boldsymbol{A}^T\\ \boldsymbol{R}_c &\triangleq \begin{bmatrix}\boldsymbol{R}_c^{(1)}\\ \boldsymbol{R}_c^{(2)}\\ \vdots\\ \boldsymbol{R}_c^{(M)}\end{bmatrix} = (\boldsymbol{A}\odot\boldsymbol{K}^T)\boldsymbol{P}^T\end{aligned}$$

and their sample estimates

$$\hat{\boldsymbol{R}}_a \triangleq \begin{bmatrix} \hat{\boldsymbol{R}}_a^{(1)} \\ \hat{\boldsymbol{R}}_a^{(2)} \\ \vdots \\ \hat{\boldsymbol{R}}_a^{(K)} \end{bmatrix}, \quad \hat{\boldsymbol{R}}_b \triangleq \begin{bmatrix} \hat{\boldsymbol{R}}_b^{(1)} \\ \hat{\boldsymbol{R}}_b^{(2)} \\ \vdots \\ \hat{\boldsymbol{R}}_b^{(M)} \end{bmatrix}, \quad \hat{\boldsymbol{R}}_c \triangleq \begin{bmatrix} \hat{\boldsymbol{R}}_c^{(1)} \\ \hat{\boldsymbol{R}}_c^{(2)} \\ \vdots \\ \hat{\boldsymbol{R}}_c^{(M)} \end{bmatrix}.$$

Note that for the sake of algorithm simplicity, we will not exploit the fact that our PARAFAC model is symmetric. That is, the algorithm that follows treats $\boldsymbol{A}$ and $\boldsymbol{K}$ as independent variables; symmetry will only be exploited in the calculation of the final estimate of $\boldsymbol{A}$.

2.5.1 Trilinear Alternating Least Square Estimator

The basic idea behind the TALS procedure for PARAFAC fitting is to update each time a subset of parameters using least square (LS) regression while keeping the previously obtained estimates of the rest of parameters fixed. This alternating projections-type procedure is iterated for all subsets of parameters until convergence is achieved [42], [44], [46], [53].

In application to our problem, the PARAFAC TALS procedure can be formulated as follows.

- **Step 1:** Initialize $\boldsymbol{P}$ and $\boldsymbol{A}$.
- **Step 2:** Find the estimate of $\boldsymbol{K}$ by solving the following LS problem

  $$\hat{\boldsymbol{K}} = \arg\min_{\boldsymbol{K}} \left\| \hat{\boldsymbol{R}}_a - (\boldsymbol{P} \odot \boldsymbol{A})\boldsymbol{K} \right\|^2$$

  whose analytic solution is given by

  $$\hat{\boldsymbol{K}} = (\boldsymbol{P} \odot \boldsymbol{A})^{\dagger} \hat{\boldsymbol{R}}_a$$

  Set $\boldsymbol{K} = \hat{\boldsymbol{K}}$.
- **Step 3:** Find the estimate of $\boldsymbol{A}$ by solving the following LS problem

  $$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \left\| \hat{\boldsymbol{R}}_b - (\boldsymbol{K}^T \odot \boldsymbol{P})\boldsymbol{A}^T \right\|^2$$

whose analytic solution is given by

$$\hat{\boldsymbol{A}} = \hat{\boldsymbol{R}}_b^T \left((\boldsymbol{K}^T \odot \boldsymbol{P})^\dagger \right)^T$$

Set $\boldsymbol{A} = \hat{\boldsymbol{A}}$.

- **Step 4:** Find the estimate of $\boldsymbol{P}$ by solving the following LS problem

$$\hat{\boldsymbol{P}} = \arg\min_{\boldsymbol{P}} \left\| \hat{\boldsymbol{R}}_c - (\boldsymbol{A} \odot \boldsymbol{K}^T)\boldsymbol{P}^T \right\|^2$$

whose analytic solution is given by

$$\hat{\boldsymbol{P}} = \hat{\boldsymbol{R}}_c^T \left((\boldsymbol{A} \odot \boldsymbol{K}^T)^\dagger \right)^T$$

Set $\boldsymbol{P} = \hat{\boldsymbol{P}}$.

- **Step 5:** Repeat steps 2-4 until convergence is achieved and then compute the final estimate of $\boldsymbol{A}$ as $\hat{\boldsymbol{A}} = (\boldsymbol{A} + \boldsymbol{K}^H)/2$.

The complexity of the TALS algorithm is $\mathcal{O}(P^3 + M^2PK)$ per iteration. It is worth noting that, when P is small relative to M and K, only a few iterations of this algorithm are usually required to achieve convergence [46].

2.5.2 Joint Diagonalization-Based Estimator

Using the idea of [38], we can obtain the estimate of $\boldsymbol{A}$ by means of a joint diagonalizer of the matrices $\boldsymbol{R}(k)$, $k = 1, \ldots, K$.

The estimator can be formulated as the following sequence of steps:

- **Step 1:** Calculate the eigendecomposition of $\tilde{\boldsymbol{R}}$ and find the estimate $\hat{\sigma}_v^2$ of the noise power as the average of the $M - P$ smallest eigenvalues of this matrix.

- **Step 2:** Compute the whitening matrix as

$$\boldsymbol{W} = [(\lambda_1 - \hat{\sigma}_v^2)^{-1/2}\boldsymbol{g}_1, \ldots, (\lambda_P - \hat{\sigma}_v^2)^{-1/2}\boldsymbol{g}_P]^H$$

where $\{\lambda_p\}_{p=1}^P$ are the largest (signal-subspace) eigenvalues of $\tilde{\boldsymbol{R}}$ and $\{\boldsymbol{g}_p\}_{p=1}^P$ are the corresponding eigenvectors.

- **Step 3:** Compute the prewhitened sample covariance matrices as

$$\hat{\boldsymbol{R}}_w(k) = \boldsymbol{W}\hat{\boldsymbol{R}}(k)\boldsymbol{W}^H, \qquad k = 1, \ldots, K\,.$$

- **Step 4:** Obtain a unitary matrix $\boldsymbol{U}$ as a joint diagonalizer of the set of matrices $\{\hat{\boldsymbol{R}}_w(k)\}_{k=1}^K$.

- **Step 5:** Estimate the matrix $\boldsymbol{A}$ as

$$\hat{\boldsymbol{A}} = \boldsymbol{W}^\dagger \boldsymbol{U}\,.$$

Several efficient joint diagonalization algorithms can be used in step 4, see [54] and [55]. For example, the complexity of the AC-DC algorithm of [55] is $\mathcal{O}(M^2PK + M^3)$ per iteration.

It should be pointed out that the joint diagonalization-based estimator requires stronger conditions in terms of the number of sensors as compared with the TALS estimator. Indeed, $M \geq P$ is required for the joint diagonalization algorithms [38], [55], whereas this constraint is not needed for TALS.

Both the TALS and joint diagonalization algorithms can be initialized randomly [46], [55]. Alternatively, if power control is fast enough (in the sense that there are several power changes per channel coherence dwell), we can use the fact that the power changes are known at the receiver to initialize the matrix $\boldsymbol{P}$ in TALS. However, as mentioned in Section 2.3, power control algorithms are usually much slower than the channel coherence time because their purpose is to combat long-term shadowing. For this reason, such an initialization of $\boldsymbol{P}$ may not be possible.

2.6 Modified Cramér-Rao Bound

In this section, we present a modified deterministic CRB on estimating the user spatial signatures[6]. The model (2.1) for the nth sample of the kth sub-block can be rewritten as

$$\boldsymbol{y}(k,n) = \boldsymbol{A}\boldsymbol{Q}^{1/2}(k)\tilde{\boldsymbol{s}}(n) + \boldsymbol{v}(n), \qquad n = (k-1)N_{\mathrm{s}} + 1, \ldots, kN_{\mathrm{s}} \tag{2.15}$$

[6]The deterministic CRB is a relevant bound in cases when the signal waveforms are unknown deterministic or random with unknown statistics, see, e.g., [56] and [57].

where

$$\tilde{\boldsymbol{s}}(n) \triangleq [\tilde{s}_1(n), \ldots, \tilde{s}_P(n)]^T = \boldsymbol{Q}^{-1/2}(k)\boldsymbol{s}(n)$$

is the vector of normalized signal waveforms and the normalization is done so that all waveforms have unit powers.

Hence, the observations in the kth sub-block satisfy the following model

$$\boldsymbol{y}(k,n) \sim \mathcal{CN}(\boldsymbol{\mu}(k,n), \sigma_v^2 \boldsymbol{I}_M) \tag{2.16}$$

where

$$\boldsymbol{\mu}(k,n) = \boldsymbol{A}\boldsymbol{Q}^{1/2}(k)\tilde{\boldsymbol{s}}(n)\,, \qquad n = (k-1)N_{\mathrm{s}} + 1, \ldots, kN_{\mathrm{s}}\,. \tag{2.17}$$

The unknown parameters of the model (2.15) are all entries of $\boldsymbol{A}$, diagonal elements of $\boldsymbol{Q}(k)$ $(k = 1, \ldots, K)$ and the noise power σ_v^2. Note that, to make the model (2.15) identifiable, we assume that the signal waveforms are known. Therefore, we study a modified (optimistic) CRB. Note, however, that, as follows from our simulation results in the next section, such an optimistic CRB still remains relevant for the problem considered, because the performances of our estimators are rather close to it.

Also, note that the parameter σ_v^2 is decoupled with other parameters in the Fisher information matrix (FIM) [57]. Therefore, without any loss of generality, σ_v^2 can be excluded from the vector of unknown parameters.

A delicate point regarding the CRB for model (2.15) is the inherent permutation and scaling ambiguities. To get around the problem of scaling ambiguity, we assume that each spatial signature vector is normalized so that its first element is equal to one (after such a normalization the first row of $\boldsymbol{A}$ becomes $[1, \ldots, 1]$). To avoid the permutation ambiguity, we assume that the first row of $\boldsymbol{P}$ is known and consists of distinct elements. Then, the vector of the parameters of interest can be written as

$$\boldsymbol{\alpha} = [\boldsymbol{\alpha}_2^T, \ldots, \boldsymbol{\alpha}_M^T]^T \in \mathbb{R}^{2(M-1)P \times 1} \tag{2.18}$$

where

$$\boldsymbol{\alpha}_m \triangleq [\mathrm{Re}\{\tilde{\boldsymbol{a}}_m\}^T, \mathrm{Im}\{\tilde{\boldsymbol{a}}_m\}^T]^T\,, \qquad \tilde{\boldsymbol{a}}_m \triangleq [a_{m,1}, \ldots, a_{m,P}]^T\,.$$

The vector of nuisance parameters can be expressed as

$$\boldsymbol{\zeta} = [\tilde{\boldsymbol{p}}(2), \ldots, \tilde{\boldsymbol{p}}(K)]^T \in \mathbb{R}^{(K-1)P\times 1} \tag{2.19}$$

where $\tilde{\boldsymbol{p}}(k)$ is the kth row of the matrix $\boldsymbol{P}$.

Using (2.18) and (2.19), the vector of unknown parameters can be written as

$$\boldsymbol{\theta} = [\boldsymbol{\alpha}^T,\ \boldsymbol{\zeta}^T]^T \in \mathbb{R}^{(2(M-1)P+(K-1)P)\times 1}\,.$$

Theorem 2.3. *The* $(2(M-1)P+(K-1)P)\times(2(M-1)P+(K-1)P)$ *FIM is given by*

$$\text{FIM} = \left[\begin{array}{ccc|ccc} \boldsymbol{J}_{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2} & & 0 & & & \\ & \ddots & & & \boldsymbol{J}_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(2)} \ \cdots \ \boldsymbol{J}_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(K)} & \\ 0 & & \boldsymbol{J}_{\boldsymbol{\alpha}_M,\boldsymbol{\alpha}_M} & & & \\ \hline & \boldsymbol{J}^T_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(2)} & & \boldsymbol{J}_{\tilde{\boldsymbol{p}}(2),\tilde{\boldsymbol{p}}(2)} & & 0 \\ & \vdots & & & \ddots & \\ & \boldsymbol{J}^T_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(K)} & & 0 & & \boldsymbol{J}_{\tilde{\boldsymbol{p}}(K),\tilde{\boldsymbol{p}}(K)} \end{array}\right] \tag{2.20}$$

where

$$\boldsymbol{J}_{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2} = \cdots = \boldsymbol{J}_{\boldsymbol{\alpha}_M,\boldsymbol{\alpha}_M} = \frac{2}{\sigma_v^2}\begin{bmatrix} \mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} & -\mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \\ \mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} & \mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \end{bmatrix} \tag{2.21}$$

$$\boldsymbol{J}_{\tilde{\boldsymbol{p}}(k),\tilde{\boldsymbol{p}}(k)} = \frac{2}{\sigma_v^2}\mathrm{Re}\{(\boldsymbol{G}(k))^H\boldsymbol{G}(k)\} \tag{2.22}$$

$$\boldsymbol{J}_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(k)} = \frac{2}{\sigma_v^2}(\boldsymbol{I}_{M-1}\otimes\tilde{\boldsymbol{F}}(k))\tilde{\boldsymbol{H}}(k) \tag{2.23}$$

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{f}_1(1) & \ldots & \boldsymbol{f}_P(1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{f}_1(K) & \ldots & \boldsymbol{f}_P(K) \end{bmatrix} \in \mathbb{C}^{KN_\mathrm{s}\times P} \tag{2.24}$$

$$\boldsymbol{G}(k) = \begin{bmatrix} \boldsymbol{h}_{1,1}(k) & \ldots & \boldsymbol{h}_{1,P}(k) \\ \vdots & \ddots & \vdots \\ \boldsymbol{h}_{M,1}(k) & \ldots & \boldsymbol{h}_{M,P}(k) \end{bmatrix} \in \mathbb{C}^{MN_\mathrm{s}\times P} \tag{2.25}$$

$$\tilde{\boldsymbol{F}}(k) = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}^H(k)\} & -\mathrm{Im}\{\boldsymbol{F}^H(k)\} \\ \mathrm{Im}\{\boldsymbol{F}^H(k)\} & \mathrm{Re}\{\boldsymbol{F}^H(k)\} \end{bmatrix} \tag{2.26}$$

$$\boldsymbol{F}(k) = [\boldsymbol{f}_1(k), \ldots, \boldsymbol{f}_P(k)] \in \mathbb{C}^{N_\mathrm{s} \times P} \tag{2.27}$$

$$\tilde{\boldsymbol{H}}(k) = [\tilde{\boldsymbol{H}}_2^T(k), \ldots, \tilde{\boldsymbol{H}}_M^T(k)]^T \tag{2.28}$$

$$\tilde{\boldsymbol{H}}_m(k) = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{H}_m(k)\} \\ \mathrm{Im}\{\boldsymbol{H}_m(k)\} \end{bmatrix} \tag{2.29}$$

$$\boldsymbol{H}_m(k) = [\boldsymbol{h}_{m,1}(k), \ldots, \boldsymbol{h}_{m,P}(k)] \in \mathbb{C}^{N_\mathrm{s} \times P} \tag{2.30}$$

$$\boldsymbol{f}_p(k) = \left[\sqrt{\nu_p(k)}\tilde{s}_p((k-1)N_\mathrm{s}+1), \ldots, \sqrt{\nu_p(k)}\tilde{s}_p(kN_\mathrm{s})\right]^T \in \mathbb{C}^{N_\mathrm{s} \times 1} \tag{2.31}$$

$$\boldsymbol{h}_{m,p}(k) = \left[\frac{a_{m,p}\tilde{s}_p((k-1)N_\mathrm{s}+1)}{2\sqrt{\nu_p(k)}}, \ldots, \frac{a_{m,p}\tilde{s}_p(kN_\mathrm{s})}{2\sqrt{\nu_p(k)}}\right]^T \in \mathbb{C}^{N_\mathrm{s} \times 1} \tag{2.32}$$

and $\otimes$ denotes the Kronecker product.

The $2(M-1)P \times 2(M-1)P$ spatial signature-related block of the CRB matrix is given in closed form as

$$\mathrm{CRB}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \Bigg[\boldsymbol{J}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} - \frac{2}{\sigma_v^2}\sum_{k=2}^{K}(\boldsymbol{I}_{M-1} \otimes \tilde{\boldsymbol{F}}(k))\tilde{\boldsymbol{H}}(k)\left[\mathrm{Re}\{\boldsymbol{G}^H(k)\boldsymbol{G}(k)\}\right]^{-1} \times \tilde{\boldsymbol{H}}^H(k)(\boldsymbol{I}_{M-1} \otimes \tilde{\boldsymbol{F}}(k))^H\Bigg]^{-1} \tag{2.33}$$

where the upper-left block of (2.20) can be expressed as

$$\boldsymbol{J}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{J}_{\boldsymbol{\alpha}_M,\boldsymbol{\alpha}_M} \end{bmatrix} = \frac{2}{\sigma_v^2}\boldsymbol{I}_{M-1} \otimes \begin{bmatrix} \mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} & -\mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \\ \mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} & \mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \end{bmatrix}. \tag{2.34}$$

PROOF: See appendix to this chapter (Section 2.A). □

The obtained CRB expressions will be compared with the performance of the TALS and joint diagonalization-based estimators in the next section.

2.7 Numerical Examples

In this section, the performance of the developed blind spatial signature estimators is compared with that of the ESPRIT-like estimator of [31], the generalized array manifold (GAM) MUSIC estimator of [28], and the derived modified CRB.

Although the proposed blind estimators are applicable to general array geometries, the ESPRIT-like estimator is based on the uniform linear array (ULA) assumption. Therefore, to compare the estimators in a proper way, we assume a ULA of M omnidirectional sensors spaced half a wavelength apart, and $P = 2$ binary phase shift keying (BPSK) user signals impinging on the array from the angles θ_1 and θ_2 relative to the broadside, where in each simulation run θ_1 and θ_2 are randomly uniformly drawn from the whole field-of-view $[-90^\circ, 90^\circ]$. Throughout the simulations, the users are assumed to be synchronized (except Figures 2.5 and 2.6 where the case of unsynchronized users is considered), $K = 10$ sub-blocks are used in our techniques (except Figure 2.11 where K is varied), and the user powers are changed between different sub-blocks uniformly with a constant power change factor (PCF) of 1.2 (except Figure 2.10 where the PCF is varied). Note that $\boldsymbol{P} = \mathrm{SNR}(\boldsymbol{E}_I + \mathrm{PCF} \cdot \boldsymbol{D}_R)$ where SNR is the average user signal-to-noise ratio (SNR) in a single sensor, $\boldsymbol{E}_I$ is the matrix whose elements are all equal to one, $\boldsymbol{D}_R$ is a random matrix whose elements are uniformly and independently drawn from the interval $[-0.5, 0.5]$, and it is assumed that $\sigma_v^2 = 1$.

To implement the PARAFAC TALS and joint diagonalization-based estimators, we use the COMFAC algorithm of [53] and AC-DC algorithm of [55], respectively. Throughout the simulations, both our algorithms are initialized randomly. The stopping criterion of the TALS algorithm is the relative error in fitting the matrices $\hat{\boldsymbol{R}}_a$, $\hat{\boldsymbol{R}}_b$ and $\hat{\boldsymbol{R}}_c$. The stopping criterion of the joint diagonalization algorithm is the relative joint diagonalization error. The algorithms are stopped if such errors become small. Typically, both algorithms converged in less than 30 iterations.

In most figures, the estimator performances are compared in terms of the root-mean-square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{LPM}\sum_{l=1}^{L}\left\|\hat{\boldsymbol{A}}(l)-\boldsymbol{A}\right\|^2} \tag{2.35}$$

where $L = 500$ is the number of independent simulation runs and $\hat{\boldsymbol{A}}(l)$ is the estimate of $\boldsymbol{A}$ obtained from the lth run. Note that permutation and scaling of columns is fixed by means of a least-squares ordering and normalization of the columns of $\hat{\boldsymbol{A}}(l)$. A greedy least-squares algorithm [44] is used to match the (normalized) columns of $\hat{\boldsymbol{A}}$ to those of $\boldsymbol{A}$. We first form a $P \times P$ distance matrix whose (p,q)th element contains the Euclidean distance between the pth column of $\boldsymbol{A}$ and the qth column of $\hat{\boldsymbol{A}}$. The smallest element of this distance matrix determines the first match, and the respective row and column of this matrix are deleted. The process is then repeated with the reduced-size distance matrix.

The CRB is averaged over simulation runs as well.

To verify that the RMSE is a proper performance measure in applications to communications problems, one of our figures also illustrates the performance in terms of the BER when the estimated steering vectors are used together with a typical detection strategy to estimate the transmitted bits.

Example 1: Unknown Sensor Gains and Phases

Following [31], we assume in our first example that the array gains and phases are unknown, i.e., the received data are modelled as (2.2) with

$$\boldsymbol{A} = \boldsymbol{\Gamma}\boldsymbol{A}_0$$

where $\boldsymbol{A}_0$ is the matrix of nominal (plane-wavefront) user spatial signatures and $\boldsymbol{\Gamma}$ is the diagonal matrix containing the array unknown gains and phases, i.e., $\boldsymbol{\Gamma} = \text{diag}\{g_1 e^{j\phi_1}, \ldots, g_M e^{j\phi_M}\}$. The unknown gains $g_1, \ldots, g_M$ are independently drawn in each simulation run from the uniform random generator with the mean equal to $\sqrt{3}$ and standard deviation equal to one, while the unknown phases $\phi_1, \ldots, \phi_M$ are independently and uniformly drawn from the interval $[0, 2\pi)$.

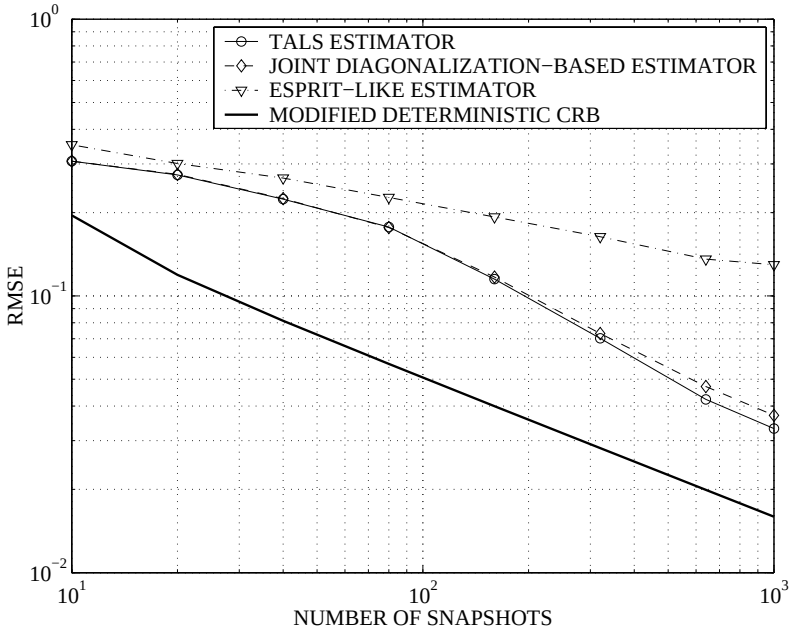

Figure 2.2: RMSEs versus J for $M = 10$ and SNR = 10 dB. First example; synchronized users.

Figure 2.2 displays the RMSEs of our estimators and the ESPRIT-like estimator of [31] along with the CRB versus J for $M = 10$ and SNR = 10 dB. Figure 2.3 shows the performances of the same estimators and the CRB versus the SNR for $M = 10$ and $J = 1000$.

Figure 2.4 illustrates the performance in terms of the BER when the estimated spatial signatures are used to detect the transmitted bits via the zero-forcing (ZF) detector $\text{sign}(\hat{\boldsymbol{A}}^{\dagger}\boldsymbol{y}(n))$. In order to avoid errors in computing the pseudoinverse of the matrix $\boldsymbol{A}$, the runs in which $\boldsymbol{A}^H\boldsymbol{A}$ was ill-conditioned have been dropped. The resulting BERs are displayed versus the SNR for $M = 10$ and $J = 1000$. Additionally, the results of the so-called *clairvoyant* ZF detector $\text{sign}(\boldsymbol{A}^{\dagger}\boldsymbol{y}(n))$ are displayed in this figure. Note that the latter detector corresponds to the ideal case when the source spatial signatures are exactly known and, therefore, it does not correspond to any practical situation. However, its performance is included in Figure 2.4 for the sake of comparison as a benchmark.

To demonstrate that the proposed techniques are insensitive to user synchronization, Figures 2.5 and 2.6 show the RMSEs of the same methods and in the same scenarios as in

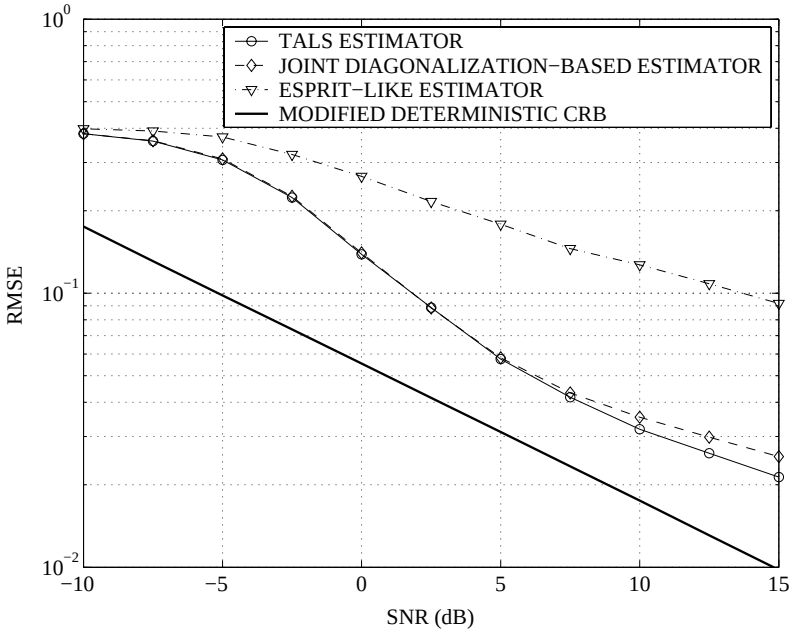

Figure 2.3: RMSEs versus the SNR for $M = 10$ and $J = 1000$. First example; synchronized users.

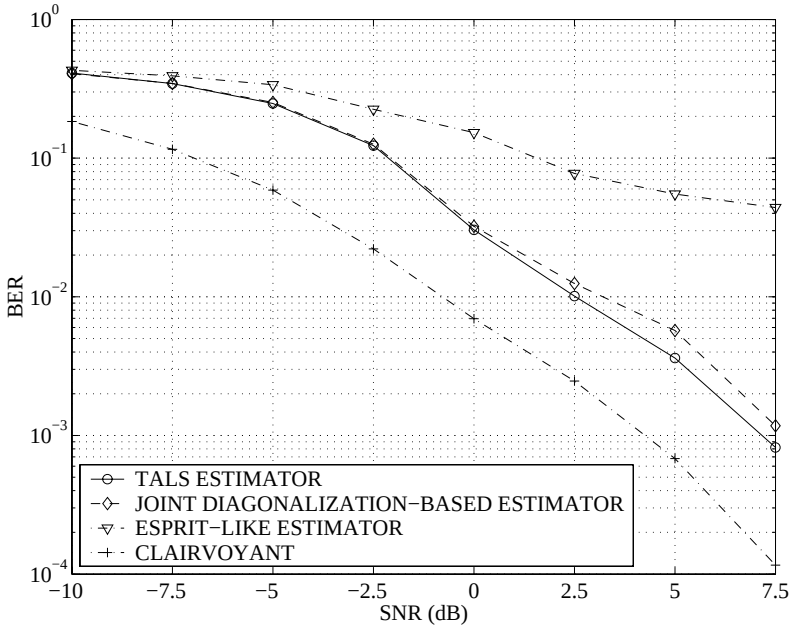

Figure 2.4: BERs versus the SNR for $M = 10$ and $J = 1000$. First example; synchronized users.

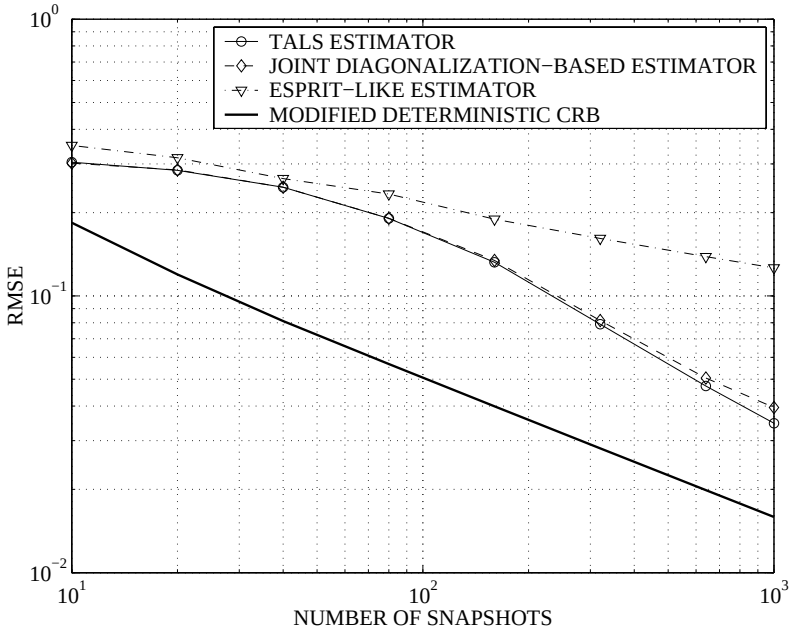

Figure 2.5: RMSEs versus J for $M = 10$ and SNR $= 10$ dB. First example; unsynchronized users.

Figures 2.2 and 2.3, respectively, but for the case of unsynchronized users[7].

To evaluate the performance with a smaller number of sensors, Figure 2.7 compares the RMSEs of the estimators tested versus J for $M = 4$ and SNR $= 10$ dB. Figure 2.8 displays the performances of these estimators versus the SNR for $M = 4$ and $J = 1000$.

To illustrate how the performance depends on the number of sensors, the RMSEs of the estimators tested are plotted in Figure 2.9 versus M. Figures 2.10 and 2.11 compare the performances of the proposed PARAFAC estimators versus the PCF and the number of sub-blocks K, respectively. In these figures, $J = 1000$ and SNR $= 10$ dB.

Example 2: Unknown Coherent Local Scattering

In our second example, we address the scenario where the spatial signature of each nominal (plane-wavefront) user is distorted by local scattering effects [40], [41]. In this example, the pth user spatial signature is formed by five signal paths of the same amplitude including the

[7]That is, the user powers vary without any synchronization between the users.

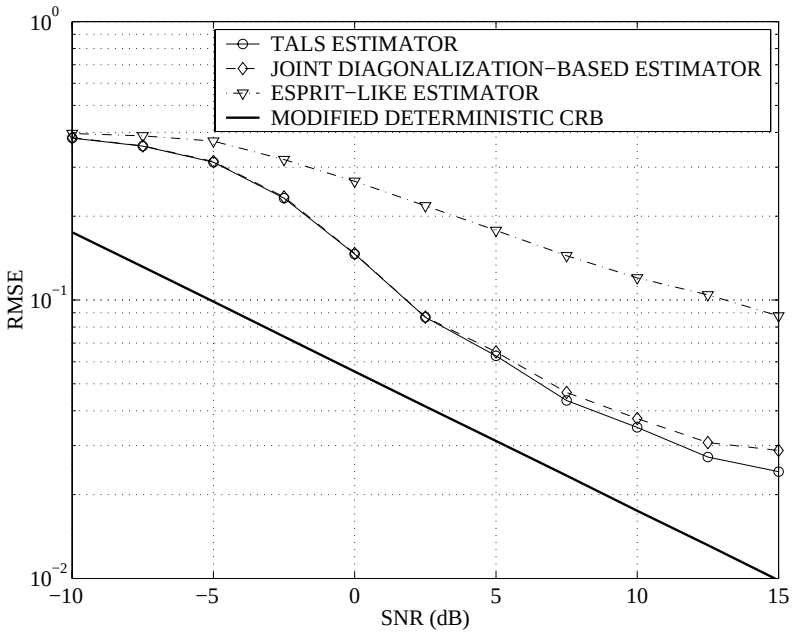

Figure 2.6: RMSEs versus the SNR for $M = 10$ and $J = 1000$. First example; unsynchronized users.

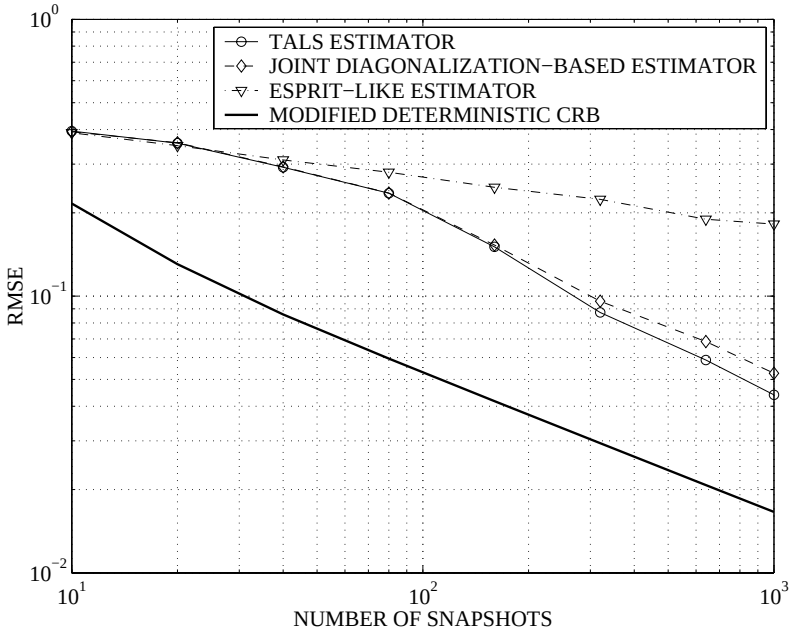

Figure 2.7: RMSEs versus J for $M = 4$ and SNR = 10 dB. First example; synchronized users.

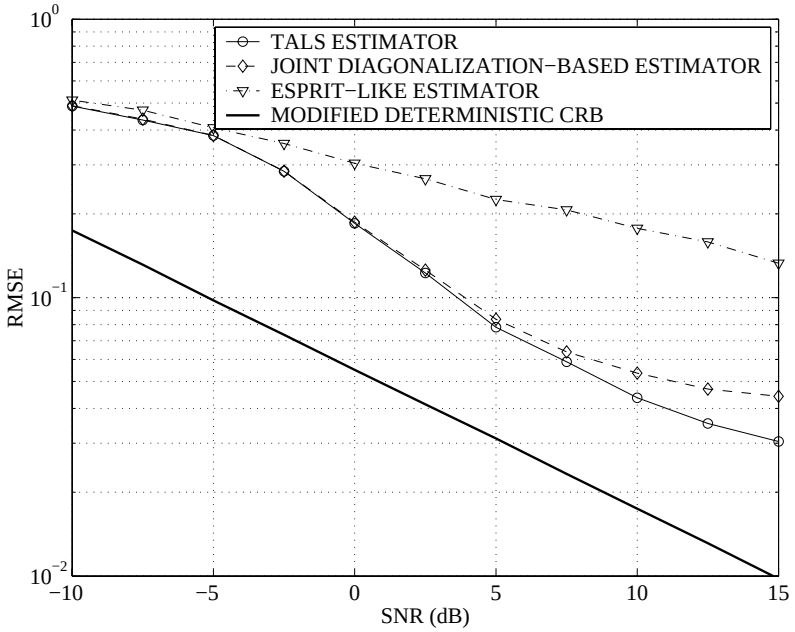

Figure 2.8: RMSEs versus the SNR for $M = 4$ and $J = 1000$. First example; synchronized users.

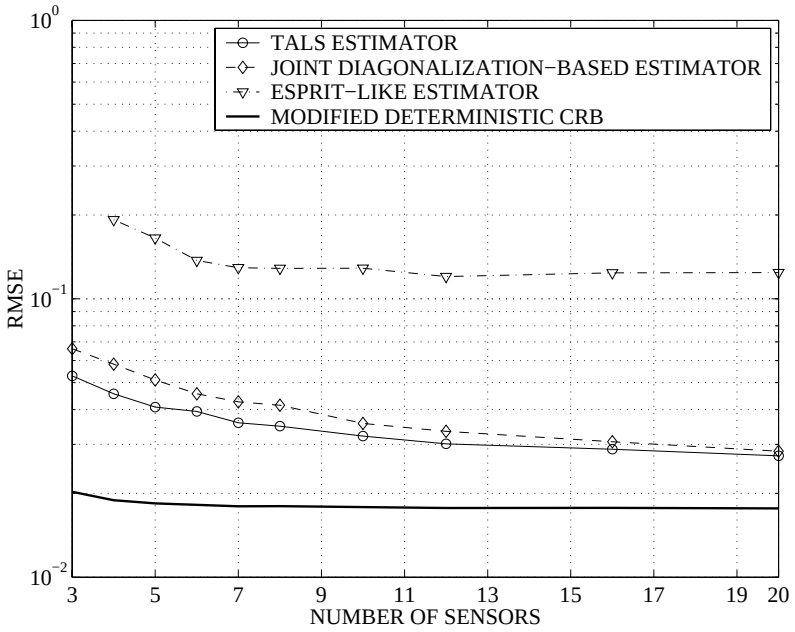

Figure 2.9: RMSEs versus M for SNR = 10 dB and $J = 1000$. First example; synchronized users.

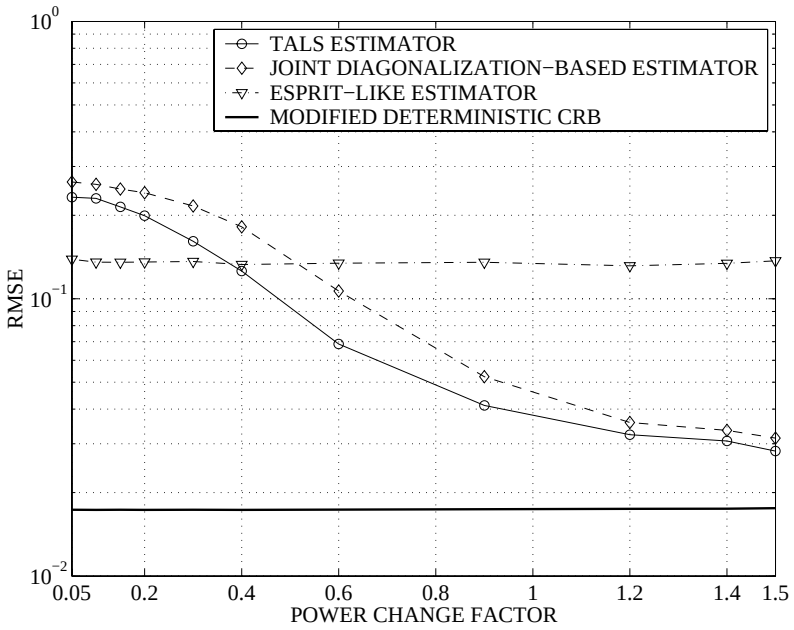

Figure 2.10: RMSEs versus the PCF for SNR = 10 dB and $J = 1000$. First example; synchronized users.

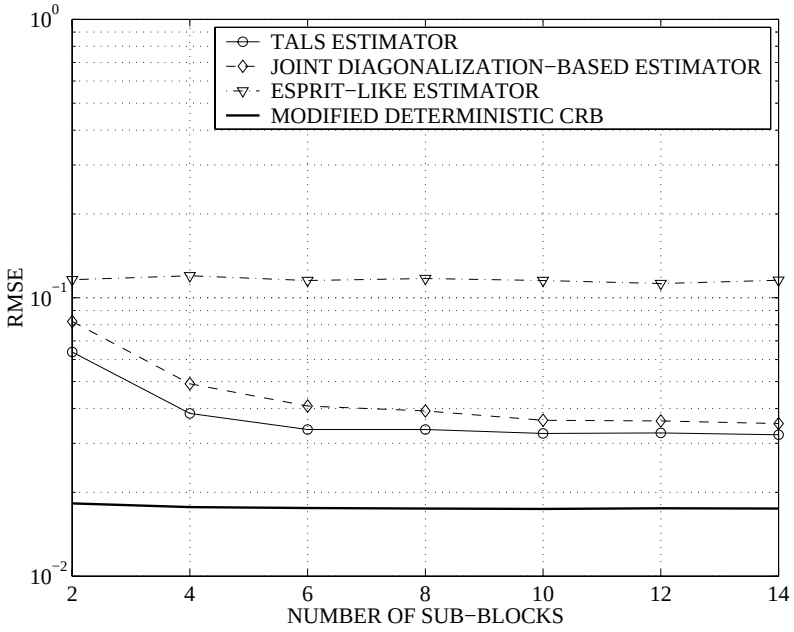

Figure 2.11: RMSEs versus K for SNR = 10 dB and $J = 1000$. First example; synchronized users.

single direct path and four coherently scattered paths. Each of these paths is characterized by its own angle and phase. The angle of the direct path is equal to the nominal user DOA while the angles of scattered paths are independently drawn in each simulation run from a uniform random generator with the mean equal to the nominal user DOA and the standard deviations equal to 8° and 10° for the first and second users, respectively. The path phases for each user are uniformly and independently drawn in each simulation run from the interval $[0, 2\pi)$.

Note that in the second example it is improper to compare the proposed techniques with the ESPRIT-like estimator of [31] because the latter estimator is not a relevant technique for the scenario considered. Therefore, in this example we compare our techniques to the GAM-MUSIC estimator of [28].

Figure 2.12 displays the performance of the spatial signature estimators tested versus the number of snapshots J for $M = 10$ and SNR = 10 dB. Note that the SNR is defined here by taking into account all signal paths. The performance of the same methods versus the SNR for $M = 10$ and $J = 1000$ is displayed in Figure 2.13.

Discussion

Our simulation results clearly demonstrate that the proposed blind PARAFAC spatial signature estimators substantially outperform the ESPRIT-like estimator and the GAM-MUSIC estimator. These improvements are especially pronounced at high values of SNR, number of snapshots, and number of sensors.

Comparing Figures 2.2 and 2.3 with Figures 2.5 and 2.6, respectively, we observe that the requirement of user synchronization is not critical to the performance of both the TALS and joint diagonalization-based algorithms. As a matter of fact, the performances of these techniques do not differ much in the cases of synchronized and unsynchronized users. This means that our techniques can easily accommodate intercell interference provided that out-of-cell users also play up and down their powers, because the fact that out-of-cell users will not be synchronized is not critical performance-wise.

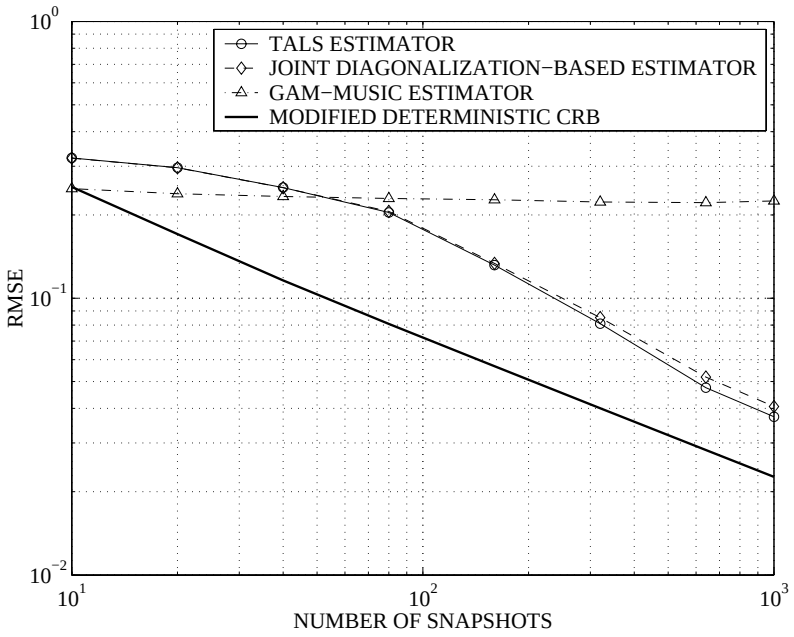

Figure 2.12: RMSEs versus J for $M = 10$ and SNR $= 10$ dB. Second example; synchronized users.

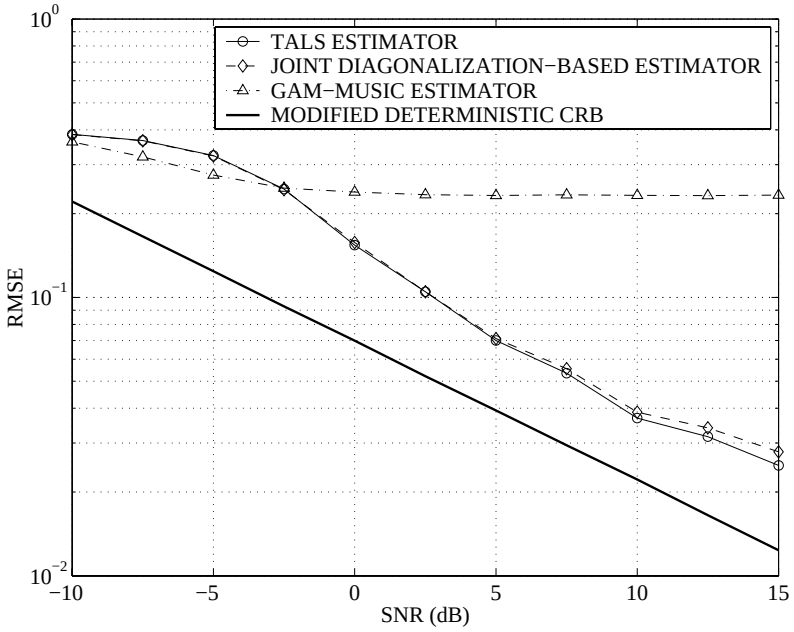

Figure 2.13: RMSEs versus the SNR for $M = 10$ and $J = 1000$. Second example; synchronized users.

From Figure 2.10, it is clear that the performance of the proposed techniques can be improved by increasing the PCF. This figure clarifies that the performance improvements of our estimators over the ESPRIT-like estimator are achieved by means of using the power loading proposed. From Figure 2.10, it follows that even moderate values of PCF ($1.2 \ldots 1.4$) are sufficient to guarantee that the performances of the proposed PARAFAC estimators are comparable with the CRB and are substantially better than that of the ESPRIT-like estimator.

From Figure 2.11, we can observe that the performance of the proposed PARAFAC estimators is also improved when increasing the number of sub-blocks while keeping the total block length fixed. However, this is only true for small numbers of K; for $K \geq 8$ curves saturate. Note that this figure makes it clear that even a moderate number of sub-blocks ($K = 4 \ldots 6$) is sufficient to guarantee that the performance is comparable with the CRB and is better than that of the ESPRIT-like estimator. We stress that the effects of the PCF and K cannot be seen from the CRB in Figures 2.10 and 2.11 because the time-averaged user powers and the total number of snapshots do not change in these figures.

Figures 2.12 and 2.13 show that both the TALS and joint-diagonalization based estimators substantially outperform the GAM-MUSIC estimator if the values J and SNR are sufficiently high. Interestingly, the performance of GAM-MUSIC does not improve much when increasing J or SNR. This observation can be explained by the fact that the GAM-MUSIC estimator is biased. Note that from Figure 2.12 it follows that GAM-MUSIC may perform better than the proposed PARAFAC estimators in the case when J is small because power loading approach does not work properly if there are only a few snapshots per sub-block (in this case, the covariance matrix estimates for each sub-block become very poor).

Interestingly, as it follows from Figure 2.4, the proposed PARAFAC-based techniques combined with ZF detector have the same BER slope as the clairvoyant ZF detector, while the performance losses with respect to the latter detector do not exceed 3 dB at high SNRs.

There are several reasons why the proposed techniques perform better than the ESPRIT-like algorithm. First of all, even in the case when the array is fully calibrated, the performance of ESPRIT is poorer than MUSIC and/or maximum-likelihood (ML) because ESPRIT does not take advantage of the full array manifold, but only of the array shift-invariance property. Second, our algorithm takes advantage of the user power loading while the ESPRIT-like algorithm does not.

As far as the comparison GAM-MUSIC method is concerned, better performances of the proposed techniques can be explained by the above-mentioned fact that GAM-MUSIC uses the first-order Taylor series approximation which is only adequate for asymptotically small angular spreads. As a result, the GAM-MUSIC estimator is biased. Also, similarly to the ESPRIT-like algorithm, GAM-MUSIC does not take any advantage of the user power loading.

Although the performance of the proposed estimators can be made comparable to the CRB with proper choice of PCF and system parameters, it does not come close to attaining the CRB. This can be attributed to the fact that the TALS estimator does not exploit the symmetry of the model ($\boldsymbol{K} = \boldsymbol{A}^H$), whereas joint diagonalization relies on an approximate pre-whitening step. Both methods rely on finite-sample covariance and noise-power estimates. This explains the observation that the CRB cannot be attained.

2.8 Chapter Summary

In this chapter, the problem of blind user spatial signature estimation using the PARAFAC analysis model has been addressed. A time-varying user power loading in the uplink mode has been proposed to make the model identifiable and to enable the application of the PARAFAC analysis model. Identifiability issues and the relevant CRB have been studied and two blind spatial signature estimation algorithms have been presented. The first technique is based on the PARAFAC fitting TALS regression while the second one makes use of joint matrix diagonalization. These techniques have been shown to provide better performance than the popular ESPRIT-type blind estimator of [31] and GAM-MUSIC estimator

of [28]. Moreover, they are also applicable to a more general class of scenarios.

2.A Derivation of the Modified Cramér-Rao Bound

The (j,l)th element of the FIM is given by [57]

$$\mathrm{FIM}_{j,l} = \frac{2}{\sigma_v^2}\sum_{k=1}^{K}\sum_{n=(k-1)N_\mathrm{s}+1}^{kN_\mathrm{s}} \mathrm{Re}\left\{\frac{\partial \boldsymbol{\mu}^H(k,n)}{\partial\theta_j}\frac{\partial\boldsymbol{\mu}(k,n)}{\partial\theta_l}\right\}. \tag{2.36}$$

Using (2.17) along with (2.36), we have

$$\frac{\partial\boldsymbol{\mu}(k,n)}{\partial \mathrm{Re}\{a_{m,p}\}} = \sqrt{\nu_p(k)}\tilde{s}_p(n)\tilde{\boldsymbol{e}}_m \tag{2.37}$$

$$\frac{\partial\boldsymbol{\mu}(k,n)}{\partial \mathrm{Im}\{a_{m,p}\}} = j\sqrt{\nu_p(k)}\tilde{s}_p(n)\tilde{\boldsymbol{e}}_m \tag{2.38}$$

$$\frac{\partial\boldsymbol{\mu}(k,n)}{\partial \nu_p(k)} = \left[\frac{a_{1,p}\tilde{s}_p(n)}{2\sqrt{\nu_p(k)}},\ldots,\frac{a_{M,p}\tilde{s}_p(n)}{2\sqrt{\nu_p(k)}}\right]^T \tag{2.39}$$

where $\tilde{\boldsymbol{e}}_m$ is the vector containing one in the mth position and zeros elsewhere.

Using (2.37) and (2.38) along with (2.36) we obtain that

$$\boldsymbol{J}_{\mathrm{Re}\{a_{m,p}\},\mathrm{Re}\{a_{m,q}\}} = \boldsymbol{J}_{\mathrm{Im}\{a_{m,p}\},\mathrm{Im}\{a_{m,q}\}} \tag{2.40}$$

$$= \frac{2}{\sigma_v^2}\sum_{k=1}^{K}\sum_{n=(k-1)N_\mathrm{s}+1}^{kN_\mathrm{s}} \mathrm{Re}\left\{\sqrt{\nu_p(k)\nu_q(k)}\tilde{s}_p^*(n)\tilde{s}_q(n)\right\}$$

$$= \frac{2}{\sigma_v^2}\mathrm{Re}\{\boldsymbol{\xi}_p^H\boldsymbol{\xi}_q\} \tag{2.41}$$

where

$$\boldsymbol{\xi}_p \triangleq [\boldsymbol{f}_p^T(1),\ldots,\boldsymbol{f}_p^T(K)]^T \in \mathbb{C}^{KN_\mathrm{s}\times 1}.$$

Similarly,

$$\boldsymbol{J}_{\mathrm{Im}\{a_{m,p}\},\mathrm{Re}\{a_{m,q}\}} = -\boldsymbol{J}_{\mathrm{Re}\{a_{m,p}\},\mathrm{Im}\{a_{m,q}\}} = \frac{2}{\sigma_v^2}\mathrm{Im}\{\boldsymbol{\xi}_p^H\boldsymbol{\xi}_q\}. \tag{2.42}$$

Therefore,

$$\begin{aligned}\boldsymbol{J}_{\mathrm{Re}\{\boldsymbol{\alpha}_m\},\mathrm{Re}\{\boldsymbol{\alpha}_m\}} = \boldsymbol{J}_{\mathrm{Im}\{\boldsymbol{\alpha}_m\},\mathrm{Im}\{\boldsymbol{\alpha}_m\}} &= \frac{2}{\sigma_v^2}\begin{bmatrix} \mathrm{Re}\{\boldsymbol{\xi}_1^H\boldsymbol{\xi}_1\} & \dots & \mathrm{Re}\{\boldsymbol{\xi}_1^H\boldsymbol{\xi}_P\} \\ \vdots & \ddots & \vdots \\ \mathrm{Re}\{\boldsymbol{\xi}_P^H\boldsymbol{\xi}_1\} & \dots & \mathrm{Re}\{\boldsymbol{\xi}_P^H\boldsymbol{\xi}_P\} \end{bmatrix} \\ &= \frac{2}{\sigma_v^2}\mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \end{aligned} \tag{2.43}$$

and

$$\begin{aligned}\boldsymbol{J}_{\mathrm{Im}\{\boldsymbol{\alpha}_m\},\mathrm{Re}\{\boldsymbol{\alpha}_m\}} = -\boldsymbol{J}_{\mathrm{Re}\{\boldsymbol{\alpha}_m\},\mathrm{Im}\{\boldsymbol{\alpha}_m\}} &= \frac{2}{\sigma^2}\begin{bmatrix} \mathrm{Im}\{\boldsymbol{\xi}_1^H\boldsymbol{\xi}_1\} & \dots & \mathrm{Im}\{\boldsymbol{\xi}_1^H\boldsymbol{\xi}_P\} \\ \vdots & \ddots & \vdots \\ \mathrm{Im}\{\boldsymbol{\xi}_P^H\boldsymbol{\xi}_1\} & \dots & \mathrm{Im}\{\boldsymbol{\xi}_P^H\boldsymbol{\xi}_P\} \end{bmatrix} \\ &= \frac{2}{\sigma_v^2}\mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\}\,. \end{aligned} \tag{2.44}$$

Using (2.43) and (2.44), we obtain (2.21). Note that the right-hand side of (2.21) does not depend on the index m. Hence,

$$\boldsymbol{J}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{J}_{\boldsymbol{\alpha}_M,\boldsymbol{\alpha}_M} \end{bmatrix} = \frac{2}{\sigma_v^2}\boldsymbol{I}_{M-1} \otimes \begin{bmatrix} \mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} & -\mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \\ \mathrm{Im}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} & \mathrm{Re}\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\} \end{bmatrix}. \tag{2.45}$$

Next, using (2.39) along with (2.36) we can write for $k = 2, \dots, K$ and $p, q = 1, \dots, P$

$$\begin{aligned}\left[\boldsymbol{J}_{\tilde{\boldsymbol{p}}(k),\tilde{\boldsymbol{p}}(k)}\right]_{p,q} &= \frac{2}{\sigma_v^2}\sum_{n=(k-1)N_\mathrm{s}+1}^{kN_\mathrm{s}}\sum_{m=1}^{M}\mathrm{Re}\left\{\frac{(a_{m,p}\tilde{s}_p(n))^*}{2\sqrt{\nu_p(k)}}\frac{a_{m,q}\tilde{s}_q(n)}{2\sqrt{\nu_q(k)}}\right\} \\ &= \frac{2}{\sigma_v^2}\mathrm{Re}\{\boldsymbol{c}_p^H(k)\boldsymbol{c}_q(k)\} \end{aligned} \tag{2.46}$$

where

$$\boldsymbol{c}_p(k) \triangleq [\boldsymbol{h}_{1,p}^T(k), \dots, \boldsymbol{h}_{M,p}^T(k)]^T \in \mathbb{C}^{MN_\mathrm{s}\times 1}\,.$$

Stacking all P^2 elements given by (2.46) in one matrix we have

$$\begin{aligned}\boldsymbol{J}_{\tilde{\boldsymbol{p}}(k),\tilde{\boldsymbol{p}}(k)} &= \frac{2}{\sigma_v^2}\begin{bmatrix} \mathrm{Re}\{\boldsymbol{c}_1^H(k)\boldsymbol{c}_1(k)\} & \dots & \mathrm{Re}\{\boldsymbol{c}_1^H(k)\boldsymbol{c}_P(k)\} \\ \vdots & \ddots & \vdots \\ \mathrm{Re}\{\boldsymbol{c}_P^H(k)\boldsymbol{c}_1(p)\} & \dots & \mathrm{Re}\{\boldsymbol{c}_P^H(k)\boldsymbol{c}_P(k)\} \end{bmatrix} \\ &= \frac{2}{\sigma_v^2}\mathrm{Re}\{\boldsymbol{G}^H(k)\boldsymbol{G}(k)\}, \qquad k = 2, \dots, K\,. \end{aligned} \tag{2.47}$$

Finally, using (2.37)-(2.39) along with (2.36) we can write for $k = 2, \dots K$, $m = 2, \dots, M$ and $p, q = 1, \dots, P$

$$\begin{aligned}
\left[\boldsymbol{J}_{\mathrm{Re}\{\boldsymbol{a}_m\},\tilde{\boldsymbol{p}}(k)}\right]_{p,q} &= \frac{2}{\sigma_v^2} \sum_{n=(k-1)N_\mathrm{s}+1}^{kN_\mathrm{s}} \mathrm{Re}\left\{ \frac{1}{2} \frac{\sqrt{\nu_p(k)}}{\sqrt{\nu_q(k)}} \tilde{s}_p^*(n) a_{m,q} \tilde{s}_q(n) \right\} \\
&= \frac{2}{\sigma_v^2} \mathrm{Re}\{\boldsymbol{f}_p^H(k) \boldsymbol{h}_{m,q}(k)\} \qquad (2.48) \\
\left[\boldsymbol{J}_{\mathrm{Im}\{\boldsymbol{a}_m\},\tilde{\boldsymbol{p}}(k)}\right]_{p,q} &= \frac{2}{\sigma_v^2} \sum_{n=(k-1)N_\mathrm{s}+1}^{kN_\mathrm{s}} \mathrm{Re}\left\{ -j \frac{1}{2} \frac{\sqrt{\nu_p(k)}}{\sqrt{\nu_q(k)}} \tilde{s}_p^*(n) a_{m,q} \tilde{s}_q(n) \right\} \\
&= \frac{2}{\sigma_v^2} \mathrm{Im}\{\boldsymbol{f}_p^H(k) \boldsymbol{h}_{m,q}(k)\} \,. \qquad (2.49)
\end{aligned}$$

Collecting all $(M-1)P^2$ elements given by (2.48) and $(M-1)P^2$ elements given by (2.49) in one matrix, we obtain

$$\boldsymbol{J}_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(k)} = \frac{2}{\sigma_v^2} \begin{bmatrix} \begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}^H(k)\boldsymbol{H}_2(k)\} \\ \mathrm{Im}\{\boldsymbol{F}^H(k)\boldsymbol{H}_2(k)\} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}^H(k)\boldsymbol{H}_M(k)\} \\ \mathrm{Im}\{\boldsymbol{F}^H(k)\boldsymbol{H}_M(k)\} \end{bmatrix} \end{bmatrix}, \qquad k = 2, \dots, K \,. \qquad (2.50)$$

Observing that

$$\begin{aligned}
\begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}^H(k)\boldsymbol{H}_m(k)\} \\ \mathrm{Im}\{\boldsymbol{F}^H(k)\boldsymbol{H}_m(k)\} \end{bmatrix} &= \begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}^H(k)\} & -\mathrm{Im}\{\boldsymbol{F}^H(k)\} \\ \mathrm{Im}\{\boldsymbol{F}^H(k)\} & \mathrm{Re}\{\boldsymbol{F}^H(k)\} \end{bmatrix} \begin{bmatrix} \mathrm{Re}\{\boldsymbol{H}_m^H(k)\} \\ \mathrm{Im}\{\boldsymbol{H}_m^H(k)\} \end{bmatrix} \\
&= \tilde{\boldsymbol{F}}(k) \tilde{\boldsymbol{H}}_m(k)
\end{aligned}$$

we can further simplify (2.50) to

$$\boldsymbol{J}_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(k)} = \frac{2}{\sigma_v^2} \left(\boldsymbol{I}_{M-1} \otimes \tilde{\boldsymbol{F}}(k) \right) \tilde{\boldsymbol{H}}(k) \,. \qquad (2.51)$$

Also, note that

$$\boldsymbol{J}_{\boldsymbol{\alpha},\tilde{\boldsymbol{p}}(k)}^T = \boldsymbol{J}_{\tilde{\boldsymbol{p}}(k),\boldsymbol{\alpha}} \,. \qquad (2.52)$$

Using (2.45), (2.47), (2.51) and (2.52) we obtain the expressions (2.20)-(2.32).

Computing the CRB for $\boldsymbol{\theta}$ requires the inverse of the $(2(M-1)P+(K-1)P)\times(2(M-1)P+(K-1)P)$ matrix (2.20). Our objective is to obtain the CRB associated with the vector parameter $\boldsymbol{\alpha}$ only, avoiding the inverse of the full FIM matrix. Exploiting the fact that the lower-right sub-block

$$\begin{bmatrix} \boldsymbol{J}_{\boldsymbol{p}(2),\boldsymbol{p}(2)} & & \\ & \ddots & \\ & & \boldsymbol{J}_{\boldsymbol{p}(K),\boldsymbol{p}(K)} \end{bmatrix}$$

of (2.20) is a block-diagonal matrix and using the partitioned matrix inversion lemma (see [57], p. 572), after some algebra we obtain (2.33) and (2.34), and the proof is complete.

Chapter 3

Blind Multiuser Detection in Impulsive Noise

In this chapter, we consider the problem of blind DS-CDMA multiuser detection in impulsive ambient noise. We link this problem to robust PARAFAC model fitting in impulsive noise. In Section 3.1, we give a brief overview of the known techniques for multiuser detection in impulsive noise. The PARAFAC model for multiuser detection is introduced in Section 3.2. Several common impulsive noise models are presented in Section 3.3. In Section 3.4, we develop two iterative algorithms for robust PARAFAC model fitting based on least absolute error (LAE) criterion. The first one is based on efficient interior point methods for linear programming (LP), employed in an alternating fashion. The second one is based on weighted median filtering (WMF) iterations. These two algorithms provide tradeoffs between the performance and complexity. The CRB of the problem at hand is presented in Section 3.5. Simulation results are given in Section 3.6. Section 3.7 briefly summarizes the chapter. The proof of property (3.12) and the detailed derivation of CRBs are given in appendices (Sections 3.A and 3.B, respectively).

3.1 Overview of Known Techniques

In conventional channel sharing approaches such as time-division multiple access (TDMA) and frequency-division multiple access (FDMA), no more than one user is allowed to occupy

a certain given time-frequency slot. These approaches have the advantage that there is no interference among different users. However, from the system capacity point of view, TDMA and FDMA approaches work in a low capacity region. CDMA scheme, which increases the achievable capacity region [17], has attracted much research interest in last decades [58].

Among two common CDMA schemes: frequency-hopping CDMA (FH-CDMA) and DS-CDMA, the latter one is of particular interest. One of the reasons is that FH spread spectrum systems experience occasional strong bursty errors, while DS spread spectrum radio systems experience continuous but lower-level random errors [58].

In a typical DS-CDMA system, each user is assigned a distinct spreading code. The user spreading codes are employed at the receiver to simultaneously detect the signals transmitted by all users. This approach is known as multiuser detection [16]. The spreading codes are either known a priori or can be estimated at the receiver.

Considerable amount of literature has addressed multiuser detection problem in last decades [16], [59]-[61]. Most of these works assume that the additive noise is Gaussian, primarily due to the simplicity and elegancy of the mathematics involved in handling Gaussian noise. Moreover, Gaussian noise assumption naturally stems from the central limit theorem. However, in many physical channels such as power lines [62] and indoor and outdoor radio channels [63]-[65], the channel noise may have non-negligible impulsive components. These impulsive components arise due to physical or man-made impulsive interference such as automobile ignition, neon light or some electronic emissions [65]. The impulsiveness nature of the channel noise in wireless channel was shown in several channel measurement campaigns [63], [64]. Several recent models for impulsive noise in wireless communications can be found in [66], [67].

It is well known that impulsive noise can be quite detrimental to the conventional systems based on the Gaussian noise assumption. For example, the error rate results presented in [69] indicate that the presence of impulsive noise can cause significant performance degradation of DS-CDMA communications over that predicted from a Gaussian noise model. This implies that in order to achieve reliable communication quality, the impulsive feature of the channel noise should be taken into consideration.

Recently, some works have addressed the problem of multiuser detection in impulsive noise. The most common practice is to apply some zero-memory nonlinearities for robust decorrelation, such as the hard limiter, the hole puncher, and the clipper [71], [88], [89]. In [89], a nonlinear preprocessor is designed based on the M-estimation method for robust regression proposed in [81]. In [88], the threshold of a nonlinear filter is obtained adaptively by joint optimization of BER and/or mean-square error (MSE). In [71], it is observed that the system performance can be improved by postponing the nonlinear processing after the decorrelation operation.

The nonlinear filtering techniques mentioned above can greatly improve the robustness of the multiuser detector against impulsive noise. However, in order to design such nonlinear filters, a priori knowledge of the system parameters such as spreading codes, channel gain and SNR is required at the receiver. Even for some "semi-blind" implementations of robust multiuser detectors such as proposed in [89], the knowledge of signature waveforms is still required. In wireless communications, however, it is a non-trivial task to obtain this knowledge (see Chapter 2). Therefore, *fully blind* multiuser receivers which are robust against impulsive noise are of great interest. By fully blind we mean that the detector does not have any knowledge of the channel gains, the spreading codes, and the SNR.

In this chapter, we propose a fully blind robust multiuser detection algorithm. We link the robust multiuser detection problem to the robust PARAFAC model fitting problem. Compared with conventional fully blind DS-CDMA multiuser detection algorithms which assume Gaussian ambient noise [44], our approach achieves significant performance improvement in impulsive noise with only a moderate performance degradation with respect to Gaussian noise. Compared with the robust multiuser detectors based on zero-memory nonlinearities, our multiuser detectors also show better performance.

3.2 PARAFAC Model for Multiuser Detection

Let us consider a multiple-access wireless communication system with P users using DS-CDMA as the multiple-access scheme. The antenna array at the receiver has M sensors.

Each information bearing symbol from the pth user is spread by a distinct code sequence $d_p(i),\ i = 1, \cdots, I$, where I is the spreading factor (spreading gain). Then the signal chips of the spread symbols are pulse-shaped and transmitted through multipath wireless channels. We assume that the maximal channel delay time is much less than a symbol duration. This is a quite common assumption in CDMA systems [72] which corresponds to the flat fading channel where the signal chips from the previous symbol only leak to the immediately following symbol but have no effects on other following symbols. In this scenario, the intersymbol interference (ISI) can be eliminated by either discarding the prefix or guard chips inserting techniques [44], [72]. We also assume a block fading channel, which is static during J symbols.

The waveform observed at the ith chip interval of the jth[1]symbol at the mth receive antenna senor consists of the received signal from all P users superimposed with additive noise. We assume that the signals arriving at the receiver are synchronized to within a fraction of chip interval. The discrete time baseband signal at the receive antenna sensor can be written as

$$\tilde{x}_{m,i,j} = x_{m,i,j} + v_{m,i,j} \tag{3.1}$$

where $x_{m,i,j}$ is the signal component, while $v_{m,i,j}$ is the i.i.d. additive zero mean noise, which may have significant impulsive component. The signal component $x_{m,i,j}$ can be written as

$$x_{m,i,j} = \sum_{p=1}^{P} a_{m,p} b_{i,p} c_{j,p} \tag{3.2}$$

for all $m = 1, \cdots, M,\ i = 1, \cdots, I,\ j = 1, \cdots, J$. Here $a_{m,p}$ is the channel fading/gain between user p and antenna element m, $b_{i,p}$ is the equivalent signature of the pth user, and $c_{j,p}$ is the jth information bearing symbol of user p.

Let us introduce matrices and arrays which are useful for the development of our algorithms. Let $\overline{\boldsymbol{X}}$ denote an $M \times I \times J$ *three-dimensional array* whose (m, i, j)th element is $x_{m,i,j}$ in (3.2), $\boldsymbol{A}$ – an $M \times P$ matrix whose (m, p)th element is $a_{m,p}$ in (3.2), $\boldsymbol{B}$ – an $I \times P$ matrix whose (i, p)th element is $b_{i,p}$ in (3.2), and $\boldsymbol{C}$ – a $J \times P$ matrix whose (j, p)th

[1]With a little abuse of notations, in Chapter 3, j denotes the index of received signals.

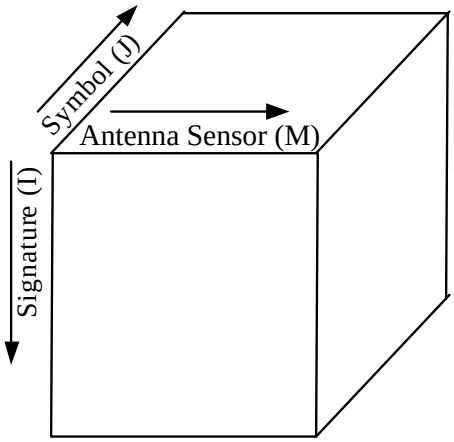

Figure 3.1: Received data array.

element is $c_{j,p}$ in (3.2). Figure 3.1 shows the structure of the three-dimensional array $\overline{\boldsymbol{X}}$. Similar to $\overline{\boldsymbol{X}}$, we also introduce two other $M \times I \times J$ three-dimensional arrays $\widetilde{\overline{\boldsymbol{X}}}$ and $\overline{\boldsymbol{V}}$ with elements given by (3.1). Then we can recast (3.1) into array form as

$$\widetilde{\overline{\boldsymbol{X}}} = \overline{\boldsymbol{X}} + \overline{\boldsymbol{V}}\,. \tag{3.3}$$

Note that matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ have their physical meanings in that $\boldsymbol{A}$ is the matrix of fading channel gains between all P users and M sensors of the antenna at the receiver, $\boldsymbol{B}$ is the equivalent signature of P users, and $\boldsymbol{C}$ is the matrix containing all the symbols transmitted by P users during J symbol duration.

The fully blind multiuser detection problem can be formulated in the following way. Given the noisy signal $\widetilde{\overline{\boldsymbol{X}}}$ at the receive antenna array, we want to estimate matrix $\boldsymbol{C}$, without a priori knowledge about $\boldsymbol{A}$ and $\boldsymbol{B}$.

Equation (3.2) expresses $x_{m,i,j}$ as a sum of P rank-1 triple products. This reminds us the link between blind multiuser detection and trilinear decomposition or *PARAFAC analysis*.

The PARAFAC fitting problem is to estimate $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ given the noisy data $\widetilde{\overline{\boldsymbol{X}}}$. From the data model in (3.2) and (3.3), we can find that the blind DS-CDMA multiuser detection problem has the same mathematical formulation as the PARAFAC fitting problem. Therefore, this interesting link opens an avenue for the development of DS-CDMA multiuser detection algorithms in impulsive noise by solving the PARAFAC fitting in impulsive noise.

It is important to point out that the authors of [44] also linked the blind DS-CDMA multiuser detection with PARAFAC analysis. However, in that paper the additive noise was assumed to be Gaussian.

PARAFAC fitting problem for Gaussian noise have been solved by using alternating least squares regression procedure (e.g., cf. [44], [53]). LS regression is optimal (in the ML sense) when the additive ambient noise is i.i.d. Gaussian. However, when the additive noise has strong impulsive components, the performance of LS regression may degrade significantly.

In this chapter, we develop two new iterative procedures for PARAFAC fitting in impulsive noise. One procedure is based on LP; the other makes use of WMF (e.g., [83]). Their relative merits are investigated via simulations and compared with the pertinent CRB, which are also derived herein.

3.3 Modelling of Impulsive Noise

Before developing our algorithms, let us briefly review some models of impulsive noise. Examples of models with impulsive noise include Laplacian distribution model [92], the Class A model [66], the α-stable model [68], and the ϵ-contaminated model [89]. Laplacian distribution is more heavy tailed than Gaussian distribution and is therefore suited to model impulsive noise. The Class A impulsive noise model is built by the infinite summation of weighted Gaussian random variables with increasing variances. The α-stable model describes a class of stable distribution with α ($0 < \alpha \leq 2$) parameterizing the heaviness of the tail of the distribution. Among all α-stable distributions, Gaussian distribution and Cauchy distribution are special cases with $\alpha = 2$ and $\alpha = 1$, respectively. The ϵ-contaminated model represents the ambient noise which is the mixture of nominal Gaussian noise with ϵ-contaminated outliers. Among the ϵ-contaminated noise model, the two-term Gaussian mixture model or the mixed-Gaussian model is widely used [89]. The probability density function (PDF) of this noise model can be written as

$$f_{\text{mixed}-\text{Gaussian}} = (1-\epsilon)\mathcal{CN}(0,\nu^2) + \epsilon\,\mathcal{CN}(0,\kappa\nu^2) \tag{3.4}$$

with $0 \leq \epsilon \leq 1$, and $\kappa \geq 1$. Here, $\mathcal{CN}(0, \nu^2)$ represents nominal Gaussian background noise and $\mathcal{CN}(0, \kappa\nu^2)$ models the impulsive component.

For α-stable distribution, two important issues should be noticed. First, no closed-form expression for the PDF exists for α other than $\alpha = 1$ and $\alpha = 2$. For other α, we can only write the characteristic function in closed-form as

$$\phi(\omega) = \exp\{-\gamma|\omega|^\alpha\}$$

where γ is a positive constant related to the scale of the distribution (also known as *dispersion* in the case of the Cauchy distribution). Second, except for the Gaussian distribution, only the moments of order less than α exist for α-stable distribution. This means that the power of an α-stable noise can not be calculated in the conventional way as the second-order moment of the noise. In [91], the so-called *geometric* power and *geometric* SNR are used to define the power and SNR, respectively. The geometric power of symmetric α-stable noise is defined as [91, pp. 38]:

$$S_0 = \frac{(C_{\mathrm{g}}\gamma)^{1/\alpha}}{C_{\mathrm{g}}}$$

where $C_{\mathrm{g}} = e^{C_{\mathrm{e}}} \approx 1.78$ is the exponential of the Euler constant $C_{\mathrm{e}} \approx 0.5772$. In particular, the geometric power of complex Gaussian noise can be calculated as

$$S_{0,\mathrm{Gaussian}} = \exp\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\ln\sqrt{\mathrm{Re}\{v_{m,i,j}\}^2 + \mathrm{Im}\{v_{m,i,j}\}^2}}{\pi\sigma_v^2} \times \exp\left\{-\frac{\mathrm{Re}\{v_{m,i,j}\}^2 + \mathrm{Im}\{v_{m,i,j}\}^2}{\sigma_v^2}\right\} d\,\mathrm{Re}\{v_{m,i,j}\}\, d\,\mathrm{Im}\{v_{m,i,j}\}\right\} = \frac{\sigma_v}{\sqrt{C_{\mathrm{g}}}} \quad (3.5)$$

For the complex Cauchy case the geometric power can be written as

$$S_{0,\mathrm{Cauchy}} = \exp\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \ln\sqrt{\mathrm{Re}\{v_{m,i,j}\}^2 + \mathrm{Im}\{v_{m,i,j}\}^2} \times \frac{\gamma}{2\pi(\mathrm{Re}\{v_{m,i,j}\}^2 + \mathrm{Im}\{v_{m,i,j}\}^2 + \gamma^2)^{3/2}}\, d\,\mathrm{Re}\{v_{m,i,j}\}\, d\,\mathrm{Im}\{v_{m,i,j}\}\right\} = 2\gamma\,. \quad (3.6)$$

The geometric SNR is defined as [91, pp. 68]

$$\mathrm{SNR} = \frac{1}{C_{\mathrm{g}}}\left(\frac{A}{S_0}\right)^2 \quad (3.7)$$

where A is the magnitude of the noise-free signal. Substituting (3.5) into (3.7) we find that for the Gaussian case the geometric SNR is equivalent to the standard SNR. For the complex Cauchy case the geometric SNR can be obtained by substituting (3.6) into (3.7)

$$\mathrm{SNR}_{\mathrm{Cauchy}} = \frac{A^2}{4C_{\mathrm{g}}\gamma^2}\,. \tag{3.8}$$

In Section 3.6, we will study the performance of different multiuser detection algorithms in Laplacian, Cauchy and ϵ-contaminated noise, respectively.

3.4 Trilinear Alternating Least Absolute Error (TALAE) Regression

In this section, we develop algorithms which solve the problem of blind DS-CDMA multiuser detection in impulsive noise. These algorithms are based on the robust PARAFAC model fitting in impulsive noise and make use of the TALAE regression.

We start by introducing some useful notations. Let $\boldsymbol{A}_m = \mathcal{D}_m(\boldsymbol{A})$. Then, by "slicing" the three-dimensional array $\overline{\boldsymbol{X}}$ in a series of "slabs" (two-dimensional arrays) we obtain

$$\boldsymbol{X}_m = \boldsymbol{B}\boldsymbol{A}_m\boldsymbol{C}^T, \qquad m = 1, \ldots, M\,.$$

Here such a slicing is made perpendicular to the mth dimension, i.e., $\boldsymbol{X}_m := [x_{m,\cdot,\cdot}]$ is the $I \times J$ two-dimensional slice of $\overline{\boldsymbol{X}}$ corresponding to the given index m. Two other types of slicing of $\overline{\boldsymbol{X}}$ are useful in understanding the algorithms which will be developed in this section. They are given by

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{C}\boldsymbol{B}_i\boldsymbol{A}^T, \qquad i = 1, \ldots, I \\ \boldsymbol{Z}_j &= \boldsymbol{A}\boldsymbol{C}_j\boldsymbol{B}^T, \qquad j = 1, \ldots, J \end{aligned}$$

where $\boldsymbol{B}_i = \mathcal{D}_i(\boldsymbol{B})$, $\boldsymbol{C}_j = \mathcal{D}_j(\boldsymbol{C})$, while the $J \times M$ matrix $\boldsymbol{Y}_i$ and $M \times I$ matrix $\boldsymbol{Z}_j$ are defined as $\boldsymbol{Y}_i := [x_{\cdot,i,\cdot}]$ and $\boldsymbol{Z}_j := [x_{\cdot,\cdot,j}]$, respectively.

As mentioned in Section 3.2, the performance of the LS regression degrades dramatically when the noise is impulsive. The LAE criterion is often used as a robust alternative to LS

[81]. LAE regression is optimal (in the ML sense) when the additive noise is i.i.d. Laplacian (e.g., see [82]). An easy way to see this is to consider mean estimation under LS and LAE criteria. These correspond to arithmetic mean and median operators, respectively. The median operator rejects impulses regardless of strength[2]; whereas the arithmetic mean is skewed by even one outlying sample.

Interestingly, as we will see, regression under the LAE criterion often performs well even when measurement errors are not Laplacian, but rather drawn from the class of α-stable or ϵ-contaminated distributions. It is therefore of interest to develop PARAFAC regression procedures that optimize the LAE fitting criterion.

Let us introduce the tall matrix

$$\boldsymbol{X} = \left[\begin{array}{c} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_M \end{array}\right]_{IM\times J} = \left[\begin{array}{c} \boldsymbol{B}\boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{B}\boldsymbol{A}_M \end{array}\right] \boldsymbol{C}^T = (\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{C}^T .$$

Similarly, we introduce the matrix of noisy data

$$\widetilde{\boldsymbol{X}} = \left[\begin{array}{c} \widetilde{\boldsymbol{X}}_1 \\ \vdots \\ \widetilde{\boldsymbol{X}}_M \end{array}\right] = \left[\begin{array}{c} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_M \end{array}\right] + \left[\begin{array}{c} \boldsymbol{V}_1 \\ \vdots \\ \boldsymbol{V}_M \end{array}\right] .$$

Then the conditional ML estimation problem for the matrix $\boldsymbol{C}$ given matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and assuming i.i.d. Gaussian measurement noise is the LS fitting problem

$$\min_{\boldsymbol{C}} \left\| \widetilde{\boldsymbol{X}} - (\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{C}^T \right\|^2 . \tag{3.9}$$

If the measurement noise is i.i.d. Laplacian (with i.i.d. Laplacian-distributed real and imaginary parts in the complex case), then ML estimation is equivalent to LAE regression. Some manipulations are necessary in order to express the absolute error criterion in the

[2]Up to roughly $J/2$ impulses can be rejected, where J is the sample size.

form of a convenient vector ℓ_1 norm. Towards this end, introduce the operator $\mathcal{F}(\cdot)$

$$\boldsymbol{s} = \mathcal{F}(\boldsymbol{S}) \triangleq \begin{bmatrix} \breve{\boldsymbol{S}}_{\cdot,1} \\ \vdots \\ \breve{\boldsymbol{S}}_{\cdot,L} \end{bmatrix} \tag{3.10}$$

$$\breve{\boldsymbol{S}}_{\cdot,l} \triangleq \begin{bmatrix} \mathrm{Re}\{\boldsymbol{S}_{\cdot,l}\} \\ \mathrm{Im}\{\boldsymbol{S}_{\cdot,l}\} \end{bmatrix} \tag{3.11}$$

where $\boldsymbol{S}$ is a complex-valued $M \times L$ matrix, and $\boldsymbol{S}_{\cdot,l}$ denotes its lth column. The following property holds (see Section 3.A for derivation):

$$\mathcal{F}\{\boldsymbol{DF}\} = (\boldsymbol{I}_L \otimes \mathcal{G}\{\boldsymbol{D}\})\mathcal{F}\{\boldsymbol{F}\} \tag{3.12}$$

where $\boldsymbol{D}$ and $\boldsymbol{F}$ are any $M \times N$ and $N \times L$ complex-valued matrices, respectively, and $\mathcal{G}\{\boldsymbol{D}\}$ stands for the following operator

$$\mathcal{G}\{\boldsymbol{D}\} \triangleq \begin{bmatrix} \mathrm{Re}\{\boldsymbol{D}\} & -\mathrm{Im}\{\boldsymbol{D}\} \\ \mathrm{Im}\{\boldsymbol{D}\} & \mathrm{Re}\{\boldsymbol{D}\} \end{bmatrix}.$$

Using property (3.12), we find that the absolute error model fitting criterion can be written as

$$\left\| \tilde{\boldsymbol{x}} - (\boldsymbol{I}_J \otimes \mathcal{G}\{\boldsymbol{A} \odot \boldsymbol{B}\})\boldsymbol{c} \right\|_1 \tag{3.13}$$

i.e., through the ℓ_1 norm of a real-valued vector. Here, $\tilde{\boldsymbol{x}} = \mathcal{F}(\widetilde{\boldsymbol{X}})$ and $\boldsymbol{c} = \mathcal{F}(\boldsymbol{C}^T)$.

Using the other two ways of slicing the array $\overline{\boldsymbol{X}}$, we introduce the matrices

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_I \end{bmatrix}_{JI \times M} = (\boldsymbol{B} \odot \boldsymbol{C})\boldsymbol{A}^T$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_J \end{bmatrix}_{MJ \times I} = (\boldsymbol{C} \odot \boldsymbol{A})\boldsymbol{B}^T$$

and correspondingly

$$\widetilde{\boldsymbol{Y}} = \begin{bmatrix} \widetilde{\boldsymbol{Y}}_1 \\ \vdots \\ \widetilde{\boldsymbol{Y}}_I \end{bmatrix}, \qquad \widetilde{\boldsymbol{Z}} = \begin{bmatrix} \widetilde{\boldsymbol{Z}}_1 \\ \vdots \\ \widetilde{\boldsymbol{Z}}_J \end{bmatrix}$$

where $\widetilde{\boldsymbol{Y}}_i$, $i = 1, \ldots, I$, and $\widetilde{\boldsymbol{Z}}_j$, $j = 1, \ldots, J$ are the noisy slabs of $\widetilde{\widetilde{\boldsymbol{X}}}$ along corresponding dimensions.

Now we have all notations necessary to explain the new fitting algorithms.

3.4.1 TALAE Regression Based on Linear Programming

The idea behind this algorithm is similar to that of TALS regression for Gaussian noise[3] [46], [53] and is as follows: each time, update a subset of parameters using the LAE criterion, conditioned on previously obtained estimates of the remaining parameters; proceed to update another subset of parameters; repeat until convergence.

In more detail, we first initialize matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ randomly or by single-invariance ESPRIT when applicable [44], [46]. Then, given the matrix $\widetilde{\boldsymbol{X}}$, and these initial estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ (which we denote hereafter as $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{B}}$), our purpose is to find the estimate of the matrix $\boldsymbol{C}$ which minimizes the norm (3.13). Specifically, we have to find the estimate of $\boldsymbol{C}$ by solving the following optimization problem

$$\widehat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \left\| \tilde{\boldsymbol{x}} - (\boldsymbol{I}_J \otimes \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\})\boldsymbol{c} \right\|_1, \qquad \widehat{\boldsymbol{C}} = (\mathcal{F}^{-1}\{\widehat{\boldsymbol{c}}\})^T \tag{3.14}$$

for given $\tilde{\boldsymbol{x}}$, $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{B}}$. In (3.14), $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse operator to $\mathcal{F}\{\cdot\}$ of (3.10). Introducing the vector $\boldsymbol{e} = [1, 1, \ldots, 1]^T$ and the vector of slack variables $\boldsymbol{q}_1$ (both of commensurate dimensions), we can equivalently write the problem (3.14) as

$$\begin{aligned} \min_{\boldsymbol{c},\ \boldsymbol{q}_1} \boldsymbol{e}^T\boldsymbol{q}_1 \quad \text{subject to} \quad & \tilde{\boldsymbol{x}} - (\boldsymbol{I}_J \otimes \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\})\boldsymbol{c} \preceq \boldsymbol{q}_1 \\ & \tilde{\boldsymbol{x}} - (\boldsymbol{I}_J \otimes \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\})\boldsymbol{c} \succeq -\boldsymbol{q}_1 \end{aligned} \tag{3.15}$$

where $\succeq$ denotes the usual *point-wise* ordering. The optimization problem in (3.15) is an LP problem that can be very efficiently solved using interior-point methods [84], [85].

[3] However, the norm of type (3.13) is now used instead of the Frobenius norm of (3.9).

Using the second way of slicing the three-dimensional array (i.e., working with the data $\tilde{\boldsymbol{y}} = \mathcal{F}(\widetilde{\boldsymbol{Y}})$) and exploiting the property (3.12) we obtain that the estimate of $\boldsymbol{A}$ can be found by solving the following optimization problem

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \left\| \tilde{\boldsymbol{y}} - (\boldsymbol{I}_M \otimes \mathcal{G}\{\widehat{\boldsymbol{B}} \odot \widehat{\boldsymbol{C}}\})\boldsymbol{a} \right\|_1, \qquad \widehat{\boldsymbol{A}} = (\mathcal{F}^{-1}\{\widehat{\boldsymbol{a}}\})^T \tag{3.16}$$

with given $\tilde{\boldsymbol{y}}$ and previously estimated $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{C}}$. This problem can be rewritten as the following LP problem:

$$\begin{aligned} \min_{\boldsymbol{a},\ \boldsymbol{q}_2} \boldsymbol{e}^T\boldsymbol{q}_2 \quad \text{subject to} \quad & \tilde{\boldsymbol{y}} - (\boldsymbol{I}_M \otimes \mathcal{G}\{\widehat{\boldsymbol{B}} \odot \widehat{\boldsymbol{C}}\})\boldsymbol{a} \preceq \boldsymbol{q}_2 \\ & \tilde{\boldsymbol{y}} - (\boldsymbol{I}_M \otimes \mathcal{G}\{\widehat{\boldsymbol{B}} \odot \widehat{\boldsymbol{C}}\})\boldsymbol{a} \succeq -\boldsymbol{q}_2 \end{aligned}$$

where $\boldsymbol{q}_2$ is the vector of slack variables of commensurate dimension.

Finally, using the third way of slicing the three-dimensional array and applying the property (3.12) we obtain that the estimate of $\boldsymbol{B}$ can be found by solving the following optimization problem:

$$\widehat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} \left\| \tilde{\boldsymbol{z}} - (\boldsymbol{I}_I \otimes \mathcal{G}\{\widehat{\boldsymbol{C}} \odot \widehat{\boldsymbol{A}}\})\boldsymbol{b} \right\|_1, \qquad \widehat{\boldsymbol{B}} = (\mathcal{F}^{-1}\{\widehat{\boldsymbol{b}}\})^T \tag{3.17}$$

with given $\tilde{\boldsymbol{z}}$ and previously estimated $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{C}}$. This problem is equivalent to the following LP problem:

$$\begin{aligned} \min_{\boldsymbol{b},\ \boldsymbol{q}_3} \boldsymbol{e}^T\boldsymbol{q}_3 \quad \text{subject to} \quad & \tilde{\boldsymbol{z}} - (\boldsymbol{I}_I \otimes \mathcal{G}\{\widehat{\boldsymbol{C}} \odot \widehat{\boldsymbol{A}}\})\boldsymbol{b} \preceq \boldsymbol{q}_3 \\ & \tilde{\boldsymbol{z}} - (\boldsymbol{I}_I \otimes \mathcal{G}\{\widehat{\boldsymbol{C}} \odot \widehat{\boldsymbol{A}}\})\boldsymbol{b} \succeq -\boldsymbol{q}_3 \end{aligned}$$

where $\boldsymbol{q}_3$ is the vector of slack variables of commensurate dimension.

Fitting proceeds by updating one matrix at a time, conditioned on interim estimates of the other two, in a round-robin fashion. Note that the conditional update of any given matrix may either improve or maintain but cannot worsen the current fit. Monotone convergence of the fit (but not necessarily to the global minimum) follows directly from this observation. The per-iteration complexity of TALAE-LP is equal to the cost of solving LP problems [85]. This is, however, of higher order of complexity as computing a matrix pseudo-inverse in the TALS method [53] and can be estimated as $\mathcal{O}(M^3I^3 + I^3J^3 + M^3J^3)$

while the complexity of TALS is $\mathcal{O}(PMIJ)$. Overall complexity depends on the number of iterations, which varies depending on problem-specific parameters and the given batch of data.

3.4.2 TALAE Regression Based on Weighted Median Filtering

LP yields the optimal solution for each of the conditional optimization problems in (3.14), (3.16) and (3.17). In the following, we show how one can *iteratively* solve (3.14) (and likewise (3.16) and (3.17)), using simple WMF. Unlike the LP-based solution, the iterative solution derived below is not necessarily an optimal solution of (3.14). However, the WMF iteration is simpler, monotonically convergent, and it does not appear to affect the performance of the overall model-fitting loop.

Fixing all parameters in (3.14) except for $c_{(j-1)2P+p}$ ($j \in \{1,\dots,J\}$, $p \in \{1,\dots,2P\}$), we can simplify this problem as

$$\min_{c_{(j-1)2P+p}} \left\| \breve{\boldsymbol{X}}_{\cdot,j} - \sum_{f=1,f\neq p}^{2P} \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\}_{\cdot,f} c_{(j-1)2P+f} - \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\}_{\cdot,p} c_{(j-1)2P+p} \right\|_1 \tag{3.18}$$

where $\breve{\boldsymbol{X}}_{\cdot,j} = [\mathrm{Re}\{\widetilde{\boldsymbol{X}}_{\cdot,j}\}^T,\ \mathrm{Im}\{\widetilde{\boldsymbol{X}}_{\cdot,j}\}^T]^T$, and $\mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\}_{\cdot,p}$ stands for the pth column of $\mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\}$.

Defining

$$\boldsymbol{h}_j = \breve{\boldsymbol{X}}_{\cdot,j} - \sum_{f=1,f\neq p}^{2P} \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\}_{\cdot,f}\, c_{(j-1)2P+f}$$

and

$$\boldsymbol{m}_p = \mathcal{G}\{\widehat{\boldsymbol{A}} \odot \widehat{\boldsymbol{B}}\}_{\cdot,p}$$

the problem in (3.18) becomes

$$\min_{c_{(j-1)2P+p}} \left\| \boldsymbol{h}_j - \boldsymbol{m}_p c_{(j-1)2P+p} \right\|_1 . \tag{3.19}$$

The minimization problem in (3.19) can be further written as

$$\min_{c_{(j-1)2P+p}} \sum_{l=1}^{2MI} \left| h_j(l) - m_p(l) c_{(j-1)2P+p} \right| \tag{3.20}$$

where $h_j(l)$ and $m_p(l)$ are the lth elements of the vectors $\boldsymbol{h}_j$ and $\boldsymbol{m}_p$, respectively. Equivalently, (3.20) can be expressed as

$$\min_{c_{(j-1)2P+p}} \sum_{l=1}^{2MI} \left|m_p(l)\right| \left| \frac{h_j(l)}{m_p(l)} - c_{(j-1)2P+p} \right| \tag{3.21}$$

provided that none of the elements $m_p(l)$ is zero. Note that if one of these elements is zero, then the corresponding summand in (3.21) can be dropped because it becomes a constant in this case (independent of $c_{(j-1)2P+p}$). Note that the optimization problem (3.21) is solved by WMF (e.g., see [83]) where $\{h_j(l)/m_p(l)\}_{l=1}^{2MI}$, $\{|m_p(l)|\}_{l=1}^{2MI}$, and $c_{(j-1)2P+p}$ are the filter inputs, weights, and output value, respectively. The WMF operation boils down to sorting, and can thus be efficiently implemented at a complexity cost of $2MI\log_2(2MI)$.

Iterating the WMF over real and imaginary parts of all elements of the matrix $\boldsymbol{C}$, e.g., in a circular fashion, and likewise for the elements of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ involved in the decomposition, one obtains a LAE trilinear regression algorithm that is monotonically convergent in terms of the LAE cost function. The per-iteration complexity of the TALAE-WMF algorithm is then estimated as $\mathcal{O}(PMIJ\log_2(MIJ))$, which can be much lower than the corresponding complexity of TALAE-LP.

Note that the proposed robust PARAFAC algorithms (as any alternating-optimization-based technique) can use any additional side-information by keeping the corresponding columns of the respective matrices fixed during iterations or initializing them with preliminary (possibly coarse) estimates. The advantage of PARAFAC fitting versus other approaches in this case is that it uses all the model structure and aims for the ML solution.

3.5 Cramér-Rao Bounds

In order to study and compare the performance of different multiuser detection algorithms, in this section we present the results of the CRBs for the problem at hand. Detailed derivations are listed in Section 3.B.

The CRB for Gaussian noise has been derived in [86]. Corresponding CRBs for impulsive noise are of interest as benchmarks in our present context. Since the closed-form expression

for the density function of mixed-Gaussian noise is not available and the symmetric α-stable distribution does not have an analytic expression for its density function. We will derive the CRBs for the fully blind DS-CDMA multiuser detection problem for Laplacian and the special case of Cauchy noise. First, the Cauchy distribution has a closed-form expression for its density function. Second, estimators that perform well under the Cauchy distribution are robust in different impulsive noise environments, i.e., the performance of such estimators does not change significantly when other symmetric α-stable distributions are used [79].

One delicate point regarding the CRB for the trilinear decomposition model is the inherent permutation and scale ambiguity. To derive a meaningful CRB, we assume that the first rows of $\boldsymbol{A}$ and $\boldsymbol{B}$ are normalized to $[1,\ldots,1]_{1\times P}$, which resolves the scale ambiguity [86]. Further we assume that the first row of $\boldsymbol{C}$ is known and consists of distinct elements, which resolves the permutation ambiguity [86]. Then, we can write the $1\times(M+I+J-3)P$ row-vector of unknown complex parameters as

$$\boldsymbol{\theta} = [\boldsymbol{a}_2,\ldots,\boldsymbol{a}_M,\boldsymbol{b}_2,\ldots,\boldsymbol{b}_I,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_J]$$

where $\boldsymbol{a}_m, m=2,\cdots,M$ is the mth row of matrix $\boldsymbol{A}$, $\boldsymbol{b}_i, i=2,\cdots,I$ is the ith row of matrix $\boldsymbol{B}$, and $\boldsymbol{c}_j, j=2,\cdots,J$ is the jth row of matrix $\boldsymbol{C}$.

Theorem 3.1. *The $(M+I+J-3)P\times(M+I+J-3)P$ FIM for the estimation of $\boldsymbol{\theta}$ in Gaussian, Laplacian and Cauchy noise is given by*

$$\text{FIM} = \begin{bmatrix} \boldsymbol{F}_1 & \boldsymbol{F}_2 \\ \boldsymbol{F}_2^H & \boldsymbol{F}_{\boldsymbol{c},\boldsymbol{c}} \end{bmatrix} \tag{3.22}$$

where

$$\boldsymbol{F}_1 = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{a}} & \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b}} \\ \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b}}^H & \boldsymbol{F}_{\boldsymbol{b},\boldsymbol{b}} \end{bmatrix}, \qquad \boldsymbol{F}_2 = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{c}} \\ \boldsymbol{F}_{\boldsymbol{b},\boldsymbol{c}} \end{bmatrix} \tag{3.23}$$

$$\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{a}} = \beta \left[\boldsymbol{I}_{M-1} \otimes ((\boldsymbol{B} \odot \boldsymbol{C})^H (\boldsymbol{B} \odot \boldsymbol{C})) \right] \tag{3.24}$$

$$\boldsymbol{F}_{\boldsymbol{b},\boldsymbol{b}} = \beta \left[\boldsymbol{I}_{I-1} \otimes ((\boldsymbol{C} \odot \boldsymbol{A})^H (\boldsymbol{C} \odot \boldsymbol{A})) \right] \tag{3.25}$$

$$\boldsymbol{F}_{\boldsymbol{c},\boldsymbol{c}} = \beta \left[\boldsymbol{I}_{J-1} \otimes ((\boldsymbol{A} \odot \boldsymbol{B})^H (\boldsymbol{A} \odot \boldsymbol{B})) \right] \tag{3.26}$$

$$\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b}} = \beta \left[\boldsymbol{I}_{M-1} \otimes (\boldsymbol{B} \odot \boldsymbol{C})^H \right] \boldsymbol{\Psi}_{1,2} \left[\boldsymbol{I}_{I-1} \otimes (\boldsymbol{C} \odot \boldsymbol{A}) \right] \tag{3.27}$$

$$\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{c}} = \beta \left[\boldsymbol{I}_{M-1} \otimes (\boldsymbol{B} \odot \boldsymbol{C})^H \right] \boldsymbol{\Psi}_{1,3} \left[\boldsymbol{I}_{J-1} \otimes (\boldsymbol{A} \odot \boldsymbol{B}) \right] \tag{3.28}$$

$$\boldsymbol{F}_{\boldsymbol{b},\boldsymbol{c}} = \beta \left[\boldsymbol{I}_{I-1} \otimes (\boldsymbol{C} \odot \boldsymbol{A})^H \right] \boldsymbol{\Psi}_{2,3} \left[\boldsymbol{I}_{J-1} \otimes (\boldsymbol{A} \odot \boldsymbol{B}) \right] \tag{3.29}$$

$$\beta = \begin{cases} \frac{1}{\sigma_v^2}\,, & \textit{for Gaussian noise} \\ \frac{2}{\sigma_v^2}\,, & \textit{for Laplacian noise} \\ \frac{3}{10\gamma^2}, & \textit{for complex Cauchy noise} \\ \frac{1}{2\gamma^2}\,, & \textit{for real Cauchy noise} \end{cases} \tag{3.30}$$

σ_v is the standard deviation of the Gaussian or Laplacian distribution, $\gamma > 0$ is the dispersion of the Cauchy distribution and

$$\boldsymbol{\Psi}_{1,2} = \begin{bmatrix} \boldsymbol{\Psi}_{1,2}(2,2) & \dots & \boldsymbol{\Psi}_{1,2}(2,I) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{1,2}(M,2) & \dots & \boldsymbol{\Psi}_{1,2}(M,I) \end{bmatrix} \tag{3.31}$$

$$\boldsymbol{\Psi}_{1,3} = \begin{bmatrix} \boldsymbol{\Psi}_{1,3}(2,2) & \dots & \boldsymbol{\Psi}_{1,3}(2,J) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{1,3}(M,2) & \dots & \boldsymbol{\Psi}_{1,3}(M,J) \end{bmatrix} \tag{3.32}$$

$$\boldsymbol{\Psi}_{2,3} = \begin{bmatrix} \boldsymbol{\Psi}_{2,3}(2,2) & \dots & \boldsymbol{\Psi}_{2,3}(2,J) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{2,3}(I,2) & \dots & \boldsymbol{\Psi}_{2,3}(I,J) \end{bmatrix}. \tag{3.33}$$

Here

$$\boldsymbol{\Psi}_{1,2}(m,i) = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & \ddots & & & & & \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ & & & \ddots & & & & & \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ & & & \ddots & & & & & \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow (i-1)J+1 \\ \leftarrow (i-1)J+2 \\ \vdots \\ \leftarrow (i-1)J+J \\ \\ \\ \end{matrix} \tag{3.34}$$

$$\begin{matrix} & & \uparrow & & \uparrow & & \uparrow & & \\ & & m & & M+m & & (J-1)M+m & & \end{matrix}$$

$$\boldsymbol{\Psi}_{1,3}(m,j) = \begin{bmatrix} 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & & \ddots & & & & & \\ 0 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & & \ddots & & & & & \\ 0 & \dots & 0 & 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & & \ddots & & & & & \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 0 \end{bmatrix} \begin{matrix} \\ \leftarrow \quad j \\ \vdots \\ \leftarrow \quad j+(i-1)J \\ \vdots \\ \leftarrow \quad j+(I-1)J \end{matrix} \tag{3.35}$$

$$\begin{matrix} & & \uparrow & \uparrow & & & \uparrow & & \\ & (m-1)I+1 & & (m-1)I+2 & & & (m-1)I+I & & \end{matrix}$$

$$\boldsymbol{\Psi}_{2,3}(i,j) = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & \ddots & & & & & \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ & & & \ddots & & & & & \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ & & & \ddots & & & & & \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow (j-1)M+1 \\ \leftarrow (j-1)M+2 \\ \vdots \\ \leftarrow (j-1)M+M \\ \\ \\ \end{matrix} \tag{3.36}$$

$$\begin{matrix} & \uparrow & & \uparrow & & \uparrow & \\ & i & I+i & & (M-1)I+i & & \end{matrix}$$

PROOF: See Section 3.B. □

Finally, the CRB matrix of the unknown elements of $\boldsymbol{C}$ is given by

$$\mathrm{CRB}_{\boldsymbol{c},\boldsymbol{c}} = (\boldsymbol{F}_{\boldsymbol{c},\boldsymbol{c}} - \boldsymbol{F}_2^H \boldsymbol{F}_1^{-1} \boldsymbol{F}_2)^{-1}. \tag{3.37}$$

3.6 Numerical Examples

In this section, we compare the performance of the proposed blind robust multiuser detection algorithms with that of the conventional blind multiuser detection algorithm [44], the robust multiuser detection algorithms using nonlinear filtering technique [88], and against the pertinent CRB.

The data $\overline{\boldsymbol{X}}$ are contaminated by channel noise. Four models of the channel noise are used: Gaussian noise, Laplacian noise, Cauchy noise, and mixed-Gaussian noise.

For TALS fitting we use the COMFAC algorithm [53], which is a fast implementation of TALS. The MOSEK convex optimization MATLAB toolbox [90] is used to solve LP problems associated with our TALAE-LP algorithm. The interior-point based solver of the MOSEK toolbox is applied. Scale and permutation ambiguities are inherent to this blind multiuser detection problem [44]; the scale ambiguity manifests itself as a complex constant

that multiplies each individual row of $\boldsymbol{C}$. For constant-modulus transmissions, this ambiguity can be removed via automatic gain control (AGC) and differential encoding/decoding. We assume differentially-encoded user signals throughout the simulations. For the purpose of performance evaluation only, the permutation ambiguity is resolved using a greedy LS $(\boldsymbol{C}, \widehat{\boldsymbol{C}})$ column-matching algorithm.

We present Monte Carlo simulations that are designed to assess the RMSE performance of the aforementioned algorithms. The parameters used in the simulations are as follows: L = number of Monte Carlo trials = 100; $I = 8$; $J = 20$; and $\alpha = 1$, which yields Cauchy noise. For Cauchy noise, geometric SNR (3.8) is used. While for the other three noises we use the standard SNR.

Throughout the simulations we assume that the noise power is normalized to be equal to 1. User signals are redrawn from an i.i.d. Bernoulli distribution and differentially encoded for each Monte Carlo trial. BPSK modulation is used for all user signals. The gains of the channel matrix $\boldsymbol{A}$ and the elements of the equivalent spreading code matrix $\boldsymbol{B}$ are generated as i.i.d. Gaussian unit variance random variables and are fixed in each Monte Carlo trial, and re-drawn from one trial to another.

Even though dimensions and ranks are such that algebraic (ESPRIT-like) initialization is possible for all three algorithms in our simulation setup, we choose to initialize all three competing algorithms randomly for each batch of data. The reason is that we wish to assess the global convergence characteristics of the three iterations.

The RMSE for each simulated point and for each method tested is calculated according to the following expression

$$\text{RMSE} = \sqrt{\frac{1}{L(J-1)P} \sum_{l=1}^{L} \left\| \widehat{\boldsymbol{C}}(l) - \boldsymbol{C} \right\|^2}$$

while the (averaged) CRB is calculated as

$$\text{CRB} = \sqrt{\sum_{l=1}^{L} \frac{\text{tr}\{\text{CRB}_{\boldsymbol{c},\boldsymbol{c}}(l)\}}{L(J-1)P}}\,.$$

In the first example, we assume that $M = 8$ sensors and $P = 2$ users. Figures 3.2-3.4 display the performance of the aforementioned algorithms in terms of RMSE versus the

SNR[4] for the case of Gaussian, Laplacian and Cauchy noise, respectively, and compare the performance with the corresponding CRBs. Figure 3.5 shows the performance of the proposed TALAE methods as compared with the TALS procedure with clipper nonlinearity [88]. The functional form of the clipper nonlinearity can be written as

$$\varphi(x) = \begin{cases} x & , \quad |x| < \tau \\ \text{sgn}(x) \cdot \tau, & \text{otherwise} \end{cases}$$

where τ is the threshold of the clipper, and $\text{sgn}(\cdot)$ denotes the sign function. In order to study the effect of τ, we set different values of τ: $0.25\gamma, 0.5\gamma, \gamma, 10\gamma, 30\gamma$, where γ is the dispersion of the Cauchy noise. Figures 3.6 and 3.7 demonstrate the performance of the proposed algorithms compared with the conventional TALS procedure and the TALS procedure with clipper nonlinearity in mixed-Gaussian ambient noise. The values of τ are set to be $\nu, 3\nu, 10\nu$, where ν is the standard deviation of the nominal background noise (3.4). In Figures 3.6 and 3.7, we change the shape of the distribution by varying ϵ with fixed total noise variance $\sigma_v^2 = (1-\epsilon)\nu^2 + \epsilon\kappa\nu^2$. κ is set to $\kappa = 100$ in both figures, while ϵ is set to be $\epsilon = 0.1$ and $\epsilon = 0.01$ in Figures 3.6 and 3.7, respectively.

In the second example, SNR = 15 dB is fixed while the numbers of sensors and users are varied. Figures 3.8 and 3.9 show the RMSEs of the methods tested versus the number of antenna elements for the case of Cauchy noise, and for $P = 2$ and $P = 4$, respectively.

Figures 3.2 and 3.3 demonstrate that in the case of Gaussian noise, the TALS method performs slightly better than the proposed robust algorithms, while in the case of Laplacian noise, the proposed robust algorithms have slightly better performance as compared with the TALS method. In the case of Cauchy noise (Figures 3.4, 3.5, 3.8, and 3.9), the TALS method breaks down, while the performance of the proposed robust algorithms is not affected and is close to the CRB (despite the fact that our techniques are designed for Laplacian noise). In mixed-Gaussian noise (Figures 3.6 and 3.7) the performance of all algorithms depends on the parameters κ and ϵ in (3.4). For the channel noise with less frequent impulsive components ($\epsilon = 0.01$), the conventional TALS algorithm performs quite well in high SNR

[4]Geometric SNR in Cauchy case and standard SNR in other cases.

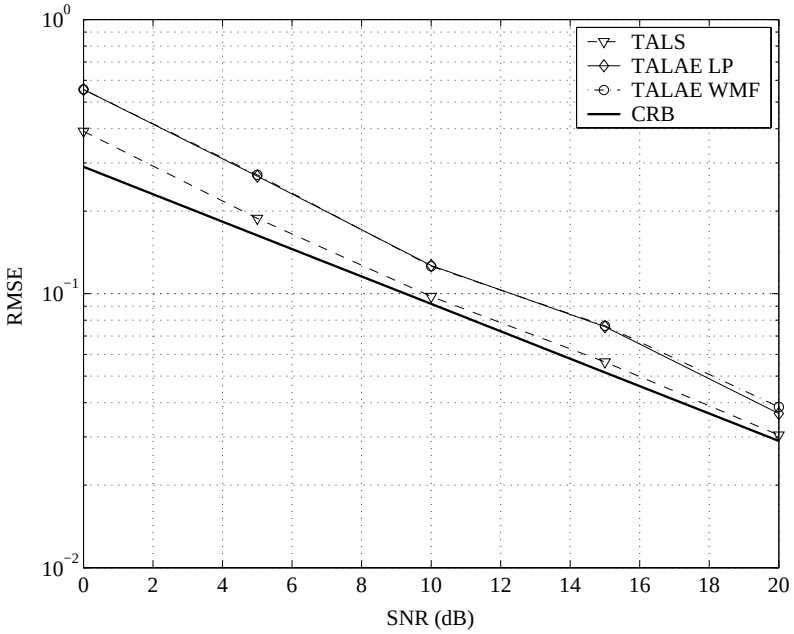

Figure 3.2: RMSEs versus SNR. First example; Gaussian channel noise.

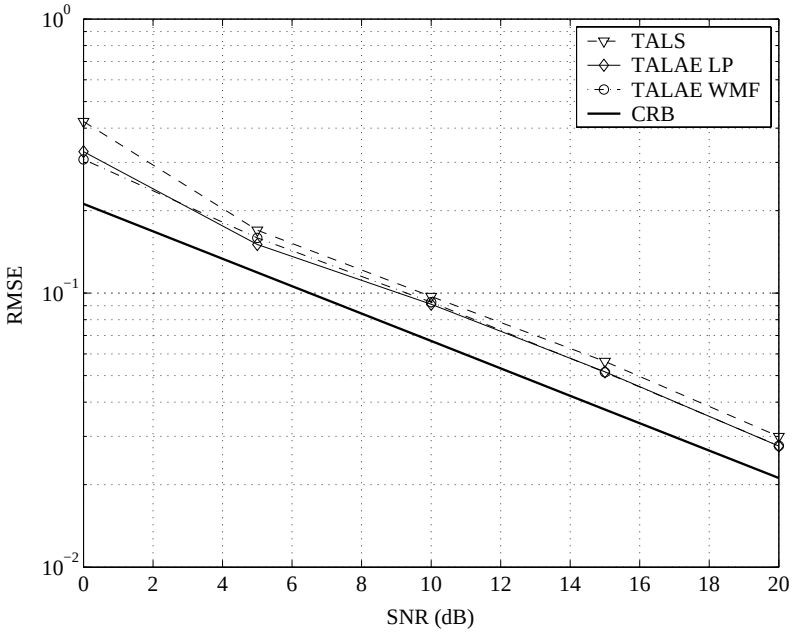

Figure 3.3: RMSEs versus SNR. First example; Laplacian channel noise.

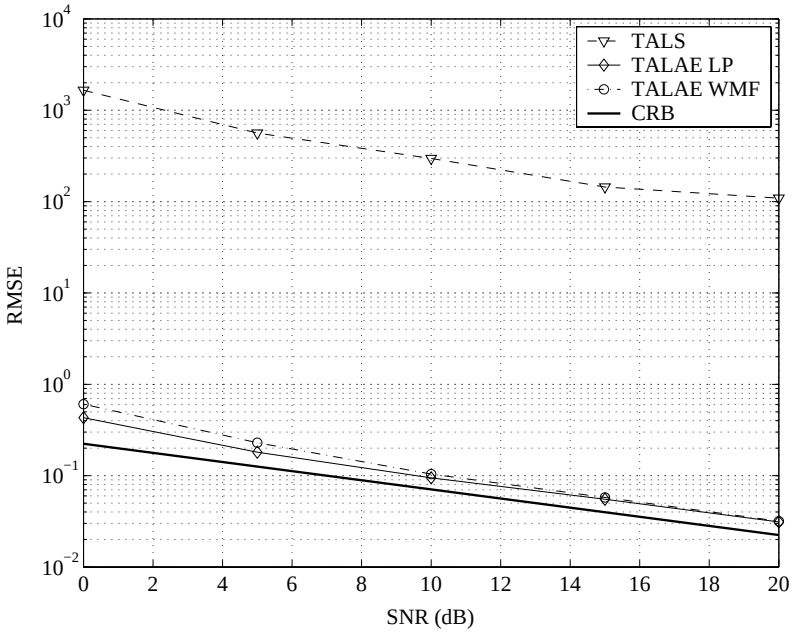

Figure 3.4: RMSEs versus SNR. First example; Cauchy channel noise.

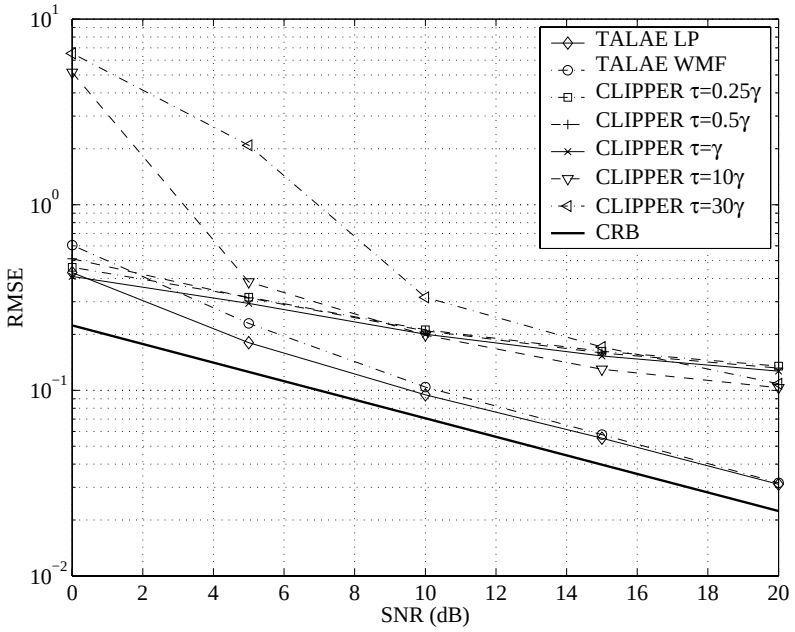

Figure 3.5: RMSEs versus SNR. First example; Cauchy channel noise.

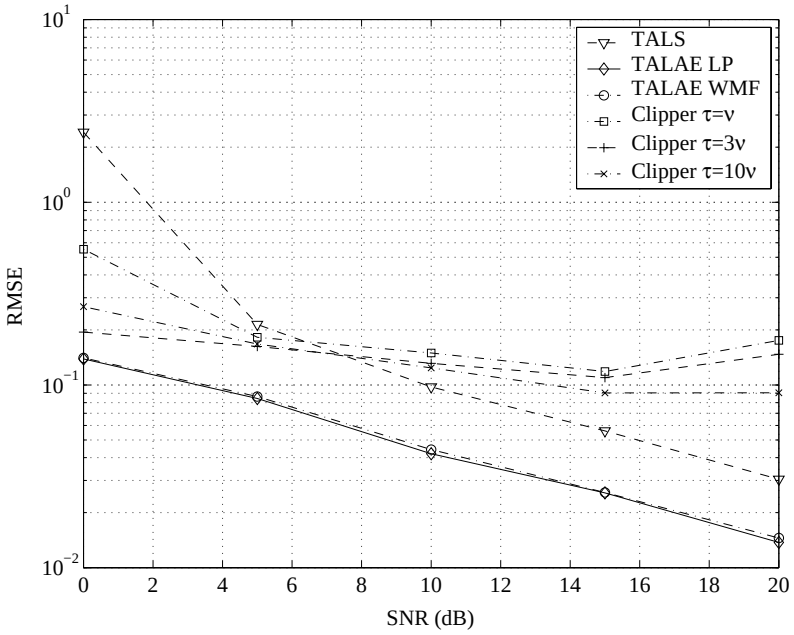

Figure 3.6: RMSEs versus SNR. First example; mixed-Gaussian channel noise, $\kappa = 100$, $\epsilon = 0.1$.

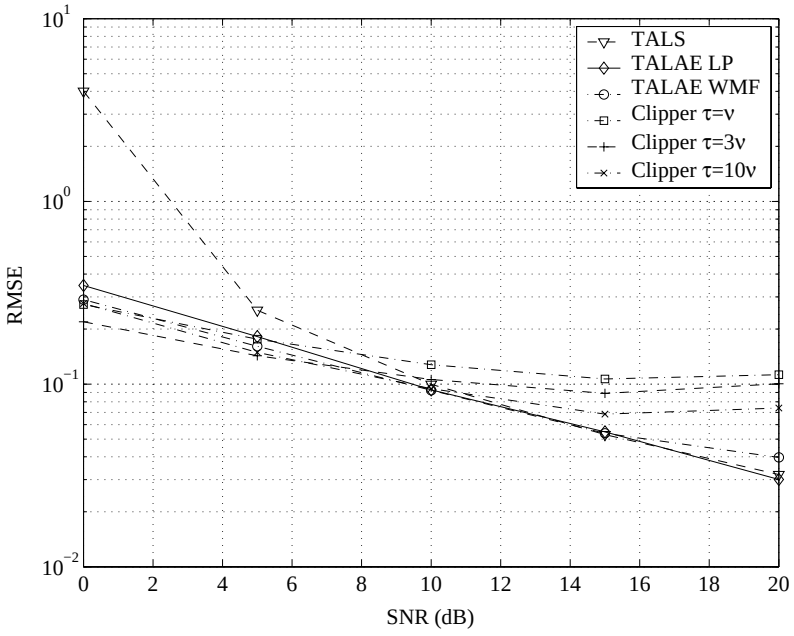

Figure 3.7: RMSEs versus SNR. First example; mixed-Gaussian channel noise, $\kappa = 100$, $\epsilon = 0.01$.

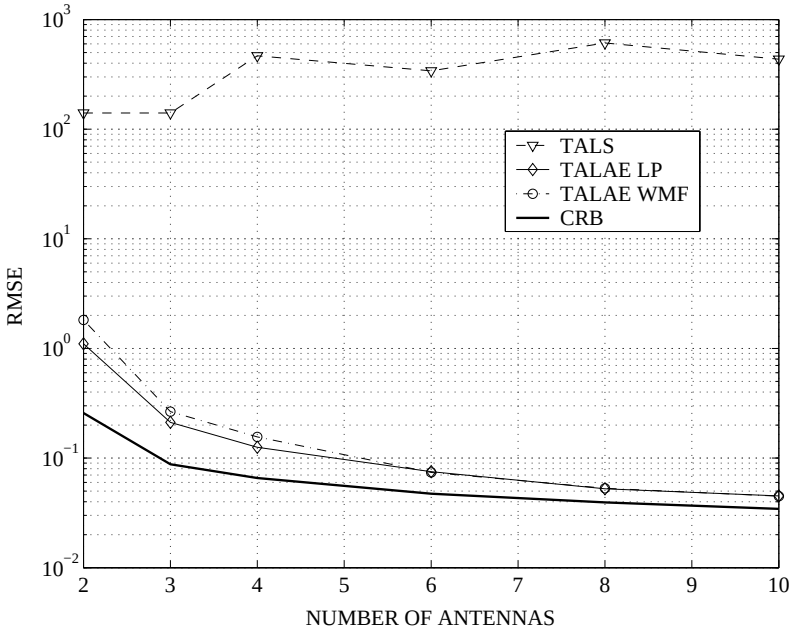

Figure 3.8: RMSEs versus M for $P = 2$. Second example; Cauchy channel noise.

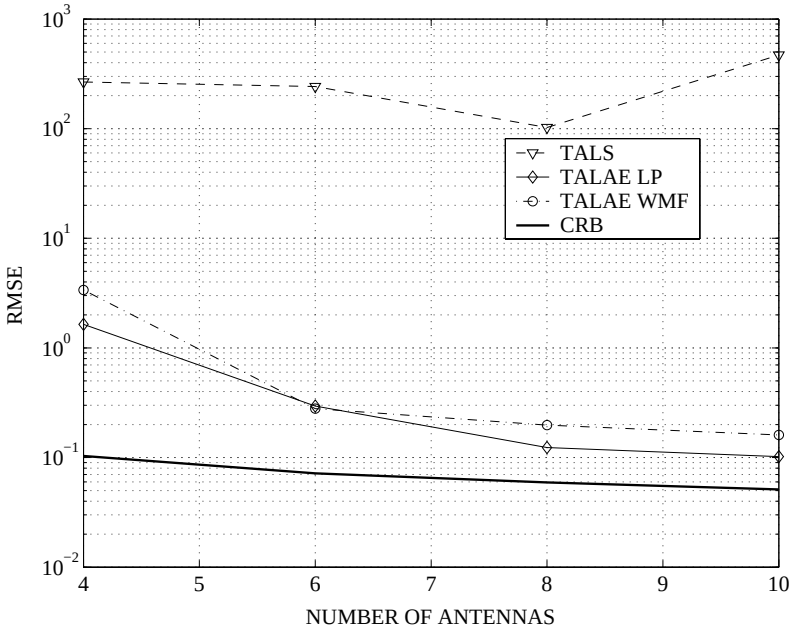

Figure 3.9: RMSEs versus M for $P = 4$. Second example; Cauchy channel noise.

region. While its performance degrades dramatically when the background noise has more frequent impulsive components ($\epsilon = 0.1$).

The performance degradation of TALAE algorithms relative to TALS in the Gaussian case can be considered as a moderate price for greatly improved robustness against heavy-tailed Cauchy noise. The TALS procedure with clipper nonlinearity shows good performance for the case of Cauchy noise as well (Figure 3.5). However, we can see that the performance of such method depends on the choice of the threshold of the clipper τ. Thus, the selection of τ is critical for good performance, especially in low SNR region. For the case of mixed-Gaussian noise, the TALS procedure with clipper performs quite well in low SNR region. However, in high SNR region, it shows some error-floors. Moreover, the TALS approach with clipper nonlinearity is *ad hoc* and not optimal in the ML sense, while our approach does not use any *ad hoc* parameters, shows better performance, and moreover, in the case of TALAE-WMF method has the same computational complexity.

Comparing the two robust regressions (TALAE-LP versus TALAE-WMF), we see that they behave very similarly performance-wise in all cases considered in our simulations. This was not necessarily expected, because TALAE-LP jointly updates many parameters, and is therefore capable of making "superdiagonal" optimization steps which are not possible with TALAE-WMF. The latter updates one parameter at a time, and thus it may be more easily trapped in ridges which do not allow fit improvements by means of updating only a single parameter. Nevertheless, this possibility does not appear to affect performance in our simulations. Further note that the two robust regressions appear robust against random initialization. Intuitively, this fact can be attributed to the fact that we work with a relatively over-determined model.

We have seen that each complete update cycle of TALAE-WMF (in which all parameters are visited for update once, in any order) is computationally much simpler than the corresponding cycle of TALAE-LP. To get a real sense of computational complexity, the number of iterations required for convergence of both methods is also needed. This number varied between 10 and 20 depending on simulated noise model and initialization of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Throughout the simulations, it was observed that TALAE-WMF requires

2-5 more iterations than TALAE-LP for Gaussian or Laplacian noise, and about the same number of iterations for Cauchy noise. This is indeed a positive result taking into account that TALAE-WMF can be implemented with simple sorting hardware, whereas TALAE-LP and TALS require a sophisticated computing capability.

3.7 Chapter Summary

In this chapter, the problem of robust blind DS-CDMA multiuser detection has been addressed. By exploiting the link with PARAFAC analysis, the problem boils down to robust fitting of PARAFAC model. Two algorithms based on LAE criterion have been proposed. The algorithms rely on alternating optimization, using LP or WMF. Our results show that under heavy-tailed noise, both algorithms outperform conventional blind DS-CDMA multiuser detection algorithm based on alternating LS PARAFAC fitting procedure, and, even though they are matched to the Laplacian distribution, they still perform well under Cauchy and mixed-Gaussian noise. Moreover, their performance degrades only moderately under Gaussian noise. Between these two algorithms, the WMF iteration is particular appealing from a simplicity viewpoint.

3.A Proof of Property (3.12)

The lth column of $\boldsymbol{DF}$ satisfies

$$[\breve{\boldsymbol{DF}}]_{\cdot,l} = \begin{bmatrix} \mathrm{Re}\{[\boldsymbol{DF}]_{\cdot,l}\} \\ \mathrm{Im}\{[\boldsymbol{DF}]_{\cdot,l}\} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{D}\} & -\mathrm{Im}\{\boldsymbol{D}\} \\ \mathrm{Im}\{\boldsymbol{D}\} & \mathrm{Re}\{\boldsymbol{D}\} \end{bmatrix} \begin{bmatrix} \mathrm{Re}\{\boldsymbol{F}_{\cdot,l}\} \\ \mathrm{Im}\{\boldsymbol{F}_{\cdot,l}\} \end{bmatrix} = \mathcal{G}\{\boldsymbol{D}\}\breve{\boldsymbol{F}}_{\cdot,l}$$

and therefore

$$\mathcal{F}\{\boldsymbol{DF}\} = \begin{bmatrix} [\breve{\boldsymbol{DF}}]_{\cdot,1} \\ \vdots \\ [\breve{\boldsymbol{DF}}]_{\cdot,L} \end{bmatrix} = \begin{bmatrix} \mathcal{G}\{\boldsymbol{D}\} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \mathcal{G}\{\boldsymbol{D}\} \end{bmatrix} \begin{bmatrix} \breve{\boldsymbol{F}}_{\cdot,1} \\ \vdots \\ \breve{\boldsymbol{F}}_{\cdot,L} \end{bmatrix} = (\boldsymbol{I}_L \otimes \mathcal{G}\{\boldsymbol{D}\})\mathcal{F}\{\boldsymbol{F}\}\,.$$

3.B Derivation of the Cramér-Rao Bounds

The proof for the case of Gaussian noise is given in [86]. Here we provide proofs for the Laplacian and Cauchy cases only.

Laplacian case:

The likelihood function in this case can be written as [92]

$$f_{\text{Laplacian}}(\widetilde{\overline{\boldsymbol{X}}}|\boldsymbol{\theta}) = \prod_{m=1}^{M}\prod_{i=1}^{I}\prod_{j=1}^{J}\frac{1}{\sigma_v^2}\exp\left\{-\frac{2}{\sigma_v}\left(\left|\text{Re}\Big\{\tilde{x}_{m,i,j}-\sum_{p=1}^{P}a_{m,p}b_{i,p}c_{j,p}\Big\}\right| + \left|\text{Im}\Big\{\tilde{x}_{m,i,j}-\sum_{p=1}^{P}a_{m,p}b_{i,p}c_{j,p}\Big\}\right|\right)\right\}.$$

The corresponding log-likelihood (LL) function is given by

$$L_{\text{Laplacian}}(\boldsymbol{\theta}) = MIJ\ln\frac{1}{\sigma_v^2} - \frac{2}{\sigma_v}\sum_{m=1}^{M}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\left|\text{Re}\Big\{\tilde{x}_{m,i,j}-\sum_{p=1}^{P}a_{m,p}b_{i,p}c_{j,p}\Big\}\right| + \left|\text{Im}\Big\{\tilde{x}_{m,i,j}-\sum_{p=1}^{P}a_{m,p}b_{i,p}c_{j,p}\Big\}\right|\right).$$

The complex FIM for LL $L_{\text{Laplacian}}(\boldsymbol{\theta})$ can be expressed as

$$\text{FIM} = \text{E}\left\{\left(\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^{H}\left(\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right\}$$

Taking partial derivatives of $L_{\text{Laplacian}}(\boldsymbol{\theta})$ with respect to the unknown parameters, we obtain

$$\begin{aligned}
\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial a_{m,p}} &= \frac{1}{2}\left[\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial\text{Re}\{a_{m,p}\}} - \sqrt{-1}\,\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial\text{Im}\{a_{m,p}\}}\right] \\
&= \frac{1}{\sigma_v}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\text{Re}\{v_{m,i,j}\}b_{i,p}c_{j,p}}{\left|\text{Re}\{v_{m,i,j}\}\right|} - \sqrt{-1}\,\frac{\text{Im}\{v_{m,i,j}\}b_{i,p}c_{j,p}}{\left|\text{Im}\{v_{m,i,j}\}\right|} \\
\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial b_{i,p}} &= \frac{1}{\sigma_v}\sum_{m=1}^{M}\sum_{j=1}^{J}\frac{\text{Re}\{v_{m,i,j}\}a_{m,p}c_{j,p}}{\left|\text{Re}\{v_{m,i,j}\}\right|} - \sqrt{-1}\,\frac{\text{Im}\{v_{m,i,j}\}a_{m,p}c_{j,p}}{\left|\text{Im}\{v_{m,i,j}\}\right|} \\
\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial c_{j,p}} &= \frac{1}{\sigma_v}\sum_{m=1}^{M}\sum_{i=1}^{I}\frac{\text{Re}\{v_{m,i,j}\}a_{m,p}b_{i,p}}{\left|\text{Re}\{v_{m,i,j}\}\right|} - \sqrt{-1}\,\frac{\text{Im}\{v_{m,i,j}\}a_{m,p}b_{i,p}}{\left|\text{Im}\{v_{m,i,j}\}\right|}.
\end{aligned}$$

Next let us derive expressions for the elements of FIM, starting from the element

$$
\begin{aligned}
&\mathrm{E}\left\{\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial a^*_{m_1,p_1}}\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial a_{m_2,p_2}}\right\}\\
&= \mathrm{E}\left\{\frac{1}{\sigma_v^2}\sum_{i_1=1}^{I}\sum_{j_1=1}^{J}\sum_{i_2=1}^{I}\sum_{j_2=1}^{J}\left(\frac{\mathrm{Re}\{v_{m_1,i_1,j_1}\}b^*_{i_1,p_1}c^*_{j_1,p_1}}{\left|\mathrm{Re}\{v_{m_1,i_1,j_1}\}\right|}+\sqrt{-1}\,\frac{\mathrm{Im}\{v_{m_1,i_1,j_1}\}b^*_{i_1,p_1}c^*_{j_1,p_1}}{\left|\mathrm{Im}\{v_{m_1,i_1,j_1}\}\right|}\right)\right.\\
&\quad\left.\times\left(\frac{\mathrm{Re}\{v_{m_2,i_2,j_2}\}b_{i_2,p_2}c_{j_2,p_2}}{\left|\mathrm{Re}\{v_{m_2,i_2,j_2}\}\right|}-\sqrt{-1}\,\frac{\mathrm{Im}\{v_{m_2,i_2,j_2}\}b_{i_2,p_2}c_{j_2,p_2}}{\left|\mathrm{Im}\{v_{m_2,i_2,j_2}\}\right|}\right)\right\}\\
&= \frac{2}{\sigma_v^2}\sum_{i=1}^{I}\sum_{j=1}^{J}b^*_{i,p_1}c^*_{j,p_1}b_{i,p_2}c_{j,p_2}\delta_{m_1,m_2}\\
&= \frac{2}{\sigma_v^2}\,\tilde{\boldsymbol{e}}_{p_1}^T(\boldsymbol{B}\odot\boldsymbol{C})^H(\boldsymbol{B}\odot\boldsymbol{C})\tilde{\boldsymbol{e}}_{p_2}\delta_{m_1,m_2} \qquad (3.38)
\end{aligned}
$$

where $\tilde{\boldsymbol{e}}_p$ is the pth unit coordinate vector, and $\delta_{m,l}$ stands for the Kronecker delta. In (3.38) we use the fact that the real and imaginary parts of $v_{m,i,j}$ are statistically independent.

Thus, the matrix containing all the elements (3.38) ($m_1, m_2 = 2,\ldots,M$ and $p_1, p_2 = 1,\ldots,P$) $\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{a}}$ can be expressed in the form of (3.24) with the coefficient $\beta = 2/\sigma_v^2$. In the same way, we compute the matrices $\boldsymbol{F}_{\boldsymbol{b},\boldsymbol{b}}$ and $\boldsymbol{F}_{\boldsymbol{c},\boldsymbol{c}}$ in the form of (3.25) and (3.26), respectively.

Similarly, we can find the element

$$
\begin{aligned}
\mathrm{E}\left\{\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial a^*_{m,p_1}}\frac{\partial L_{\text{Laplacian}}(\boldsymbol{\theta})}{\partial b_{i,p_2}}\right\} &= \frac{2}{\sigma_v^2}\sum_{j=1}^{J}b^*_{i,p_1}c^*_{j,p_1}a_{m,p_2}c_{j,p_2}\\
&= \frac{2}{\sigma_v^2}\,\tilde{\boldsymbol{e}}_{p_1}^T(\boldsymbol{B}\odot\boldsymbol{C})^H\boldsymbol{\Psi}_{1,2}(m,i)(\boldsymbol{C}\odot\boldsymbol{A})\tilde{\boldsymbol{e}}_{p_2}\,. \qquad (3.39)
\end{aligned}
$$

Then the matrix containing all the elements (3.39) ($m = 2,\ldots,M$, $i = 2,\ldots,I$ and $p_1, p_2 = 1,\ldots,P$) $\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b}}$ can be written in the form of (3.27) with the coefficient $\beta = 2/\sigma_v^2$. In the same way, we find the matrices $\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{c}}$ and $\boldsymbol{F}_{\boldsymbol{b},\boldsymbol{c}}$ given by (3.28) and (3.29), respectively. Applying the matrix inversion lemma to the FIM matrix we obtain (3.37).

Cauchy case:

We have seen that the Laplacian CRB for the PARAFAC model only differs by a multiplicative constant from the corresponding Gaussian CRB. In [93], it is shown that this is true for general signal models observed in i.i.d. additive noise, provided that the noise PDF

possesses everywhere continuous first and second derivatives. This is not the case for the Laplacian PDF, due to the discontinuity at the origin; however it is true for the Cauchy. In fact, in [93] and [94] it is shown that the noise PDF-dependent multiplicative constant that appears in the FIM can be computed as

$$\beta = \int_{-\infty}^{+\infty} \frac{|f^{'}(v)|^2}{f(v)}\, dv, \tag{3.40}$$

where $f(\cdot)$ is the noise PDF, and $f^{'}(\cdot)$ is its first derivative. Hence we can proceed in this fashion, calculating the integral above for the Cauchy PDF.

The expression for complex isotropic Cauchy density function in our case is given by [79]

$$f_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\}, \text{Im}\{v_{m,i,j}\}\big) = \frac{\gamma}{2\pi\big(\text{Re}\{v_{m,i,j}\}^2 + \text{Im}\{v_{m,i,j}\}^2 + \gamma^2\big)^{3/2}}\,. \tag{3.41}$$

The first derivative of the PDF (3.41) can be easily calculated as

$$\begin{aligned}
&f^{'}_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\}, \text{Im}\{v_{m,i,j}\}\big) \\
&= \frac{1}{2}\left[\frac{\partial f_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\}, \text{Im}\{v_{m,i,j}\}\big)}{\partial \text{Re}\{v_{m,i,j}\}} - \sqrt{-1}\,\frac{\partial f_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\}, \text{Im}\{v_{m,i,j}\}\big)}{\partial \text{Im}\{v_{m,i,j}\}}\right] \\
&= -\frac{3\gamma}{4}\left[\frac{2\text{Re}\{v_{m,i,j}\}}{2\pi\big(\text{Re}\{v_{m,i,j}\}^2 + \text{Im}\{v_{m,i,j}\}^2 + \gamma^2\big)^{5/2}}\right. \\
&\quad \left. -\sqrt{-1}\,\frac{2\text{Im}\{v_{m,i,j}\}}{2\pi\big(\text{Re}\{v_{m,i,j}\}^2 + \text{Im}\{v_{m,i,j}\}^2 + \gamma^2\big)^{5/2}}\right] \\
&= \frac{3\gamma}{4\pi}\frac{-\text{Re}\{v_{m,i,j}\} + \sqrt{-1}\text{Im}\{v_{m,i,j}\}}{\big(\text{Re}\{v_{m,i,j}\}^2 + \text{Im}\{v_{m,i,j}\}^2 + \gamma^2\big)^{5/2}}
\end{aligned}$$

and, correspondingly,

$$\frac{\big|f^{'}_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\}, \text{Im}\{v_{m,i,j}\}\big)\big|^2}{f_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\}, \text{Im}\{v_{m,i,j}\}\big)} = \frac{9\gamma}{8\pi}\frac{\text{Re}\{v_{m,i,j}\}^2 + \text{Im}\{v_{m,i,j}\}^2}{\big(\text{Re}\{v_{m,i,j}\}^2 + \text{Im}\{v_{m,i,j}\}^2 + \gamma^2\big)^{7/2}}\,. \tag{3.42}$$

Substituting (3.42) into (3.40) and calculating the integral we finally find the coefficient β

for the Cauchy noise case

$$\begin{aligned}
\beta &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\left|f'_{\text{Cauchy}}\big(\text{Re}\{v_{m,i,j}\},\text{Im}\{v_{m,i,j}\}\big)\right|^2}{f_{\text{Cauchy}}(\text{Re}\{v_{m,i,j}\},\text{Im}\{v_{m,i,j}\})}\, d\,\text{Re}\{v_{m,i,j}\}\, d\,\text{Im}\{v_{m,i,j}\} \\
&= \frac{9\gamma}{4}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\text{Re}\{v_{m,i,j}\}^2+\text{Im}\{v_{m,i,j}\}^2}{2\pi\big(\text{Re}\{v_{m,i,j}\}^2+\text{Im}\{v_{m,i,j}\}^2+\gamma^2\big)^{7/2}}\, d\,\text{Re}\{v_{m,i,j}\}\, d\,\text{Im}\{v_{m,i,j}\} \\
&= \frac{9\gamma}{4}\int_0^{+\infty}\int_0^{2\pi} \frac{r^2}{2\pi(\gamma^2+r^2)^{7/2}}\, r\, dr\, d\vartheta \\
&= \frac{9\gamma}{4}\int_0^{+\infty} \frac{r^3}{(\gamma^2+r^2)^{7/2}}\, d\, r = \frac{9\gamma}{4}\frac{2}{15\gamma^3} = \frac{3}{10\gamma^2}
\end{aligned}$$

where $r=\sqrt{\text{Re}\{v_{m,i,j}\}^2+\text{Im}\{v_{m,i,j}\}^2}$ and $\vartheta=\arctan\frac{\text{Im}\{v_{m,i,j}\}}{\text{Re}\{v_{m,i,j}\}}$ are the polar coordinates.

In the proof above we considered complex noise. However, it worth noting that in the particular case of real Cauchy noise the coefficient β is equal to $1/2\gamma^2$. Indeed, the real Cauchy PDF is

$$f_{\text{Cauchy}}(v_{m,i,j}) = \frac{\gamma}{\pi(v_{m,i,j}^2+\gamma^2)} \tag{3.43}$$

and its first derivative is

$$f'_{\text{Cauchy}}(v_{m,i,j}) = -\gamma\frac{2v_{m,i,j}}{\pi(v_{m,i,j}^2+\gamma^2)^2}\,. \tag{3.44}$$

Substituting (3.43) and (3.44) into (3.40) we obtain

$$\beta = \int_{-\infty}^{+\infty}\frac{(f'_{\text{Cauchy}}(v_{m,i,j}))^2}{f_{\text{Cauchy}}(v_{m,i,j})}\, d\, v_{m,i,j} = \frac{4\gamma}{\pi}\int_{-\infty}^{+\infty}\frac{v_{m,i,j}^2}{(v_{m,i,j}^2+\gamma^2)^3}\, d\, v_{m,i,j} = \frac{8\gamma}{\pi}\frac{\pi}{16\gamma^3} = \frac{1}{2\gamma^2}$$

and the proof is complete.

Chapter 4

Robust Linear Receivers for MIMO Systems

If both the transmitter and the receiver have multiple antennas, then a MIMO wireless communication system arises naturally. In this chapter, we focus on multiuser MIMO systems. New robust linear receiver techniques are developed for joint space-time decoding and interference rejection in multiple-access MIMO systems that use orthogonal space-time block codes (OSTBCs) and erroneous CSI.

In Section 4.1, we give a brief overview of recently developed linear receivers for multiple-access space-time block coded MIMO systems. Some background knowledge and system models on point-to-point and multiple-access space-time block coded MIMO systems are given in Section 4.2. An overview of the minimum variance (MV) linear multiuser receivers for MIMO systems are also given in Section 4.2. Using different approaches to model the CSI mismatch, in Sections 4.3 and 4.4, we propose robust linear receivers based on worst-case performance optimization and stochastic programming, respectively. These receivers are shown to provide different tradeoffs in terms of robustness, flexibility, performance, and computational complexity. Simulation results are presented in Section 4.5. Section 4.6 briefly summarizes the chapter. Proofs of Lemma 4.1, Lemma 4.2, and Theorem 4.1 are listed in Sections 4.A-4.C, respectively.

4.1 Introduction

Space-time coding has recently emerged as a powerful approach to exploit spatial diversity and combat fading in MIMO wireless communication systems [6]-[10], [95]-[97]. OSTBCs [96], [97] represent an attractive class of space-time coding techniques because they enjoy full diversity and low decoding complexity. In the point-to-point MIMO communication case, the optimal ML detector for this class of codes consists of a simple linear receiver which maximizes the output SNR and the symbol-by-symbol detector. For each symbol, this ML detector can be interpreted as a matched filter (MF) receiver [98].

In the multiple-access MIMO case, the ML receiver has much more complicated structure and prohibitively high complexity as compared with the ML receiver for the point-to-point MIMO case. Therefore, in multiple-access scenarios suboptimal but simple linear receivers can be a good choice [99]-[103].

Several linear receiver techniques have been recently developed for space-time coded multiple-access MIMO systems [99]-[103]. For example, a MV linear receiver has been developed in [99] for DS-CDMA systems which use multiple antennas and space-time block coding. However, the scheme proposed in [99] is restricted by transmitters that consist of two antennas only. The latter restriction is dictated by the Alamouti's OSTBC scheme that is adopted in [99].

Another linear technique has been proposed in [100] where a decorrelator receiver has been developed for a DS-CDMA based communication system. This receiver also uses the Alamouti's code and is limited by the assumption that the transmitter consists of two antennas and that not more than two antennas are used at the receiver. Another restriction of the receiver of [100] is that it is applicable only to the BPSK signal case.

One more linear receiver technique for the multiple-access MIMO case has been proposed in [101]. Similar to [99] and [100], the approach of [101] is restricted to the case of Alamouti's code. Another restriction of this approach is that it cannot suppress more than one interferer.

A more general class of MV linear receivers has been recently proposed in [103]. In

contrast to [99]-[101], the techniques of [103] are applicable to the general case of arbitrary OSTBCs and multiple interferers.

A common shortcoming of the techniques [99]-[103] is that they use the assumption that the exact CSI is available at the receiver. In practice, this condition can be violated because of channel estimation errors that are caused by limited/outdated training as well as the effects of multiple-access interference (MAI) and noise.

In this chapter, we develop two classes of robust linear receiver techniques for joint space-time decoding and interference rejection by using different CSI mismatch models. First, we apply a deterministic model for the CSI mismatch and propose robust receivers under the framework of worst-case performance optimization. Simulation results show significant performance improvement of the robust linear receivers as compared with the conventional (non-robust) receivers.

Second, taking into account that the worst-case approach may be overly pessimistic, we propose a class of robust receivers using a stochastic model for the CSI mismatch. We formulate the robust receiver design problems through probability-constrained stochastic optimization problems [119], [120]. We prove the convexity of these problems and convert them into nonlinear programming (NLP) and more simple second-order cone programming (SOCP) problems.

4.2 Background

In this section, we review the models for point-to-point and multiple-access space-time block coded MIMO wireless systems. The latter model is used to formulate the multiple-access MIMO linear receiver design problem.

4.2.1 Space-Time Block Coded MIMO Model

The relationship between the input and the output of a single-access (point-to-point) MIMO system with N transmit and M receive antennas and flat block-fading channel can be

expressed as [97]

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{H} + \boldsymbol{V} \tag{4.1}$$

where

$$\begin{aligned} \boldsymbol{Y} &\triangleq [\boldsymbol{y}^T(1) \ \cdots \ \boldsymbol{y}^T(T)]^T \\ \boldsymbol{X} &\triangleq [\boldsymbol{x}^T(1) \ \cdots \ \boldsymbol{x}^T(T)]^T \\ \boldsymbol{V} &\triangleq [\boldsymbol{v}^T(1) \ \cdots \ \boldsymbol{v}^T(T)]^T \end{aligned}$$

are the matrices of the received signals, transmitted signals, and noise, respectively, $\boldsymbol{H}$ is the $N \times M$ complex channel matrix, and T is the block length. Here,

$$\begin{aligned} \boldsymbol{y}(t) &= [y_1(t) \ \cdots \ y_M(t)] \\ \boldsymbol{x}(t) &= [x_1(t) \ \cdots \ x_N(t)] \\ \boldsymbol{v}(t) &= [v_1(t) \ \cdots \ v_M(t)] \end{aligned}$$

are the complex row vectors of the received signal, transmitted signal, and noise, respectively.

Let us denote complex information-bearing symbols prior to space-time encoding as $s_1, s_2, \ldots, s_K$ and assume that these symbols belong to (possibly different) constellations $\mathcal{U}_k$, $k = 1, 2, \ldots, K$. Let

$$\boldsymbol{s} \triangleq [s_1 \ \cdots \ s_K]^T.$$

Note that $\boldsymbol{s} \in \mathcal{S}$ where $\mathcal{S} = \{\boldsymbol{s}^{(1)} \cdots \boldsymbol{s}^{(Q)}\}$ is the set of all possible symbol vectors and Q is the cardinality of this set. The $T \times N$ matrix $\boldsymbol{X}(\boldsymbol{s})$ is called an OSTBC if [97]

- all elements of $\boldsymbol{X}(\boldsymbol{s})$ are linear functions of the K complex variables $s_1, s_2, \ldots, s_K$ and their complex conjugates;
- for any arbitrary $\boldsymbol{s}$, it satisfies

$$\boldsymbol{X}^H(\boldsymbol{s})\boldsymbol{X}(\boldsymbol{s}) = \|\boldsymbol{s}\|^2 \boldsymbol{I}_N \, .$$

It can be readily verified that the matrix $\boldsymbol{X}(\boldsymbol{s})$ can be written as [103], [114], [115]

$$\boldsymbol{X}(\boldsymbol{s}) = \sum_{k=1}^{K} (\boldsymbol{C}_k \mathrm{Re}\{s_k\} + \boldsymbol{D}_k \mathrm{Im}\{s_k\}) \tag{4.2}$$

where $\boldsymbol{C}_k \triangleq \boldsymbol{X}(\tilde{\boldsymbol{e}}_k)$, $\boldsymbol{D}_k \triangleq \boldsymbol{X}(j\tilde{\boldsymbol{e}}_k)$, and $\tilde{\boldsymbol{e}}_k$ is the $K \times 1$ vector having one in the kth position and zeros elsewhere. Using (4.2), one can rewrite (4.1) as [103], [114], [115]

$$\underline{\boldsymbol{Y}} = \boldsymbol{A}(\boldsymbol{H})\underline{\boldsymbol{s}} + \underline{\boldsymbol{V}} \tag{4.3}$$

where the "underline" operator for any matrix $\boldsymbol{P}$ is defined as

$$\underline{\boldsymbol{P}} \triangleq \begin{bmatrix} \mathrm{vec}(\mathrm{Re}\{\boldsymbol{P}\}) \\ \mathrm{vec}(\mathrm{Im}\{\boldsymbol{P}\}) \end{bmatrix} \tag{4.4}$$

and $\mathrm{vec}(\cdot)$ is the vectorization operator stacking all columns of a matrix on top of each other. Here, the $2MT \times 2K$ real matrix $\boldsymbol{A}(\boldsymbol{H})$ is defined as [103], [115]

$$\boldsymbol{A}(\boldsymbol{H}) \triangleq [\underline{\boldsymbol{C}_1\boldsymbol{H}} \cdots \underline{\boldsymbol{C}_K\boldsymbol{H}} \ \underline{\boldsymbol{D}_1\boldsymbol{H}} \cdots \underline{\boldsymbol{D}_K\boldsymbol{H}}]\,. \tag{4.5}$$

The matrix $\boldsymbol{A}$ captures both the effects of the space-time code and the channel. An important property of this matrix is that its columns have the same norms and are orthogonal to each other:

$$\boldsymbol{A}^T(\boldsymbol{H})\boldsymbol{A}(\boldsymbol{H}) = \|\boldsymbol{H}\|^2 \boldsymbol{I}_{2K}\,. \tag{4.6}$$

In the presence of the exact CSI at the receiver, the optimal (ML) space-time decoder uses channel knowledge to find the closest point to the received signal in the noise-free observation space $\mathcal{Y} = \{\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \ldots, \boldsymbol{Y}^{(Q)}\}$, i.e., it obtains [97]

$$l_{\mathrm{opt}} = \underset{q \in \{1,\ldots,Q\}}{\arg\min} \|\boldsymbol{Y} - \boldsymbol{Y}^{(q)}\| \tag{4.7}$$

and then uses this index to decode the transmitted bits. Here $\boldsymbol{Y}^{(q)}$ is the noise-free received signal matrix that corresponds to the vector of information-bearing symbols $\boldsymbol{s}^{(q)}$.

The ML receiver can also be viewed as a matched filter whose output SNR is maximized [98]. It can be shown [103], [115] that (4.7) is equivalent to the MF linear receiver, which computes the following estimate of $\underline{\boldsymbol{s}}$:

$$\underline{\hat{\boldsymbol{s}}} = \frac{1}{\|\boldsymbol{H}\|^2} \boldsymbol{A}^T(\boldsymbol{H})\underline{\boldsymbol{Y}} \tag{4.8}$$

and builds the estimate of the vector $\boldsymbol{s}$ as

$$\hat{\boldsymbol{s}} = [\boldsymbol{I}_K \; j\boldsymbol{I}_K]\, \underline{\hat{\boldsymbol{s}}}\,. \tag{4.9}$$

The kth element of $\hat{\boldsymbol{s}}$ is then compared with all points in $\mathcal{U}_k$. The closest point is accepted as an estimate of kth entry of $\boldsymbol{s}$. This procedure is repeated for all $k = 1, 2, \ldots, K$, that is, the decoding is done symbol-by-symbol.

According to (4.6), the matrix $\frac{1}{\|\boldsymbol{H}\|^2}\boldsymbol{A}^T(\boldsymbol{H})$ in (4.8) is the pseudoinverse of $\boldsymbol{A}(\boldsymbol{H})$. Therefore, (4.8) can be alternatively viewed as a decorrelator receiver.

Let us now consider an uplink multiple-access MIMO communication system shown in Figure 4.1. The transmitters (users) are assumed to have the same number of transmitting antennas and to encode the information-bearing symbols using the same OSTBC[1]. The received signal is given by [103]

$$\boldsymbol{Y} = \sum_{p=1}^{P} \boldsymbol{X}_p \boldsymbol{H}_p + \boldsymbol{V} \tag{4.10}$$

where $\boldsymbol{X}_p$ is the matrix of transmitted signals of the pth transmitter, $\boldsymbol{H}_p$ is the channel matrix between the pth transmitter and the receiver, and P is the number of transmitters.

Applying the "underline" operator of (4.4) to (4.10), we have [103]

$$\underline{\boldsymbol{Y}} = \sum_{p=1}^{P} \boldsymbol{A}(\boldsymbol{H}_p)\, \underline{\boldsymbol{s}_p} + \underline{\boldsymbol{V}} \tag{4.11}$$

where $\boldsymbol{s}_p$ is a $K \times 1$ vector of information-bearing symbols of the pth transmitter and, according to (4.5), the matrix $\boldsymbol{A}(\boldsymbol{H}_p)$ can be expressed as

$$\boldsymbol{A}(\boldsymbol{H}_p) = [\underline{\boldsymbol{C}_1\boldsymbol{H}_p} \cdots \underline{\boldsymbol{C}_K\boldsymbol{H}_p}\; \underline{\boldsymbol{D}_1\boldsymbol{H}_p} \cdots \underline{\boldsymbol{D}_K\boldsymbol{H}_p}] \triangleq [\boldsymbol{a}_1(\boldsymbol{H}_p) \cdots \boldsymbol{a}_{2K}(\boldsymbol{H}_p)]\,. \tag{4.12}$$

In equation (4.12), $\boldsymbol{a}_k(\boldsymbol{H}_p)$ represents the space-time signature that corresponds to the kth real-valued symbol transmitted by the pth user (i.e., the space-time signature that corresponds to the kth component of the vector $\underline{\boldsymbol{s}_p}$). Both the effects of the space-time code and the channel of the pth user are captured in the matrix (4.12).

[1]These assumptions are only needed for notational simplicity and can be relaxed, see [103].

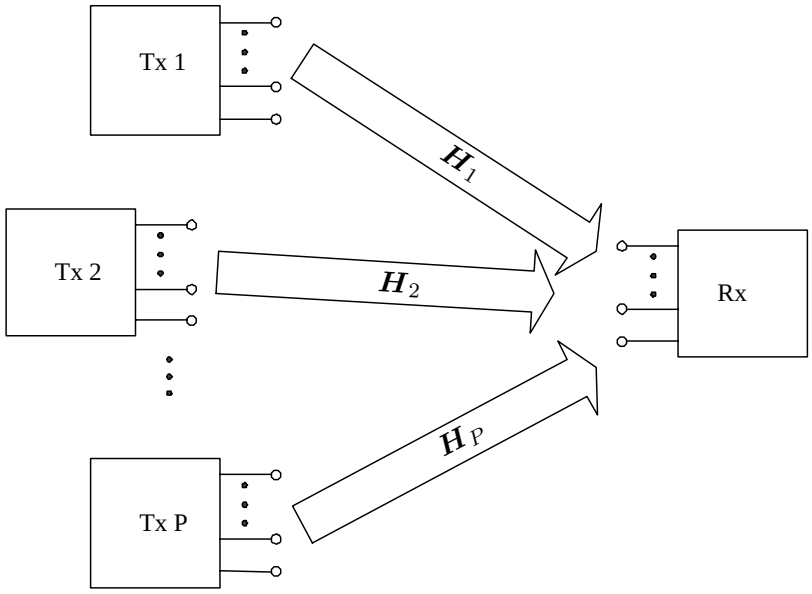

Figure 4.1: Multiple-access MIMO system.

4.2.2 Conventional Multiple-Access MIMO Linear Receivers

In the multiple-access MIMO case, the MF receiver of (4.7) becomes highly non-optimal because it ignores the effect of MAI treating it as a noise. In this case, the receiver performance is determined by the signal-to-interference-plus-noise ratio (SINR) rather than the SNR and some cancellation of MAI is required.

Using the model (4.11) and assuming without any loss of generality that the first transmitter is the transmitter-of-interest, we can express the output vector of a linear receiver as [103]

$$\underline{\hat{\boldsymbol{s}}_1} = \boldsymbol{W}^T \underline{\boldsymbol{Y}} \tag{4.13}$$

where

$$\boldsymbol{W} = [\boldsymbol{w}_1\, \boldsymbol{w}_2\, \cdots\, \boldsymbol{w}_{2K}]$$

is the $2MT \times 2K$ real matrix of the receiver coefficients and $\underline{\hat{\boldsymbol{s}}_1}$ is the estimate of the vector $\underline{\boldsymbol{s}_1}$ at the receiver output. The vector $\boldsymbol{w}_k$ can be interpreted as the weight vector for the kth entry of $\underline{\boldsymbol{s}_1}$.

Given the matrix $\boldsymbol{W}$, the estimate of the vector of information-bearing symbols of the

transmitter-of-interest can be computed as

$$\hat{\boldsymbol{s}}_1 = [\boldsymbol{I}_K \; j\boldsymbol{I}_K]\, \underline{\hat{\boldsymbol{s}}_1}\,.$$

Using such a linear estimate, the kth information-bearing symbol can be detected as a point in $\mathcal{U}_k$ which is the nearest neighbor to the kth entry of $\hat{\boldsymbol{s}}_1$.

Using the framework of (4.13), we can interpret the MF receiver in (4.8) as a linear receiver with the following coefficient matrix:

$$\boldsymbol{W}_{\mathrm{MF}} = \frac{1}{\|\boldsymbol{H}_1\|^2}\, \boldsymbol{A}(\boldsymbol{H}_1)\,. \tag{4.14}$$

The similarity of the vectorized multiple-access MIMO model (4.11) and models used in array processing gives an opportunity to design the matrix $\boldsymbol{W}$ using the MV principle [1]-[3]. Using the MV approach, in [103] it has been proposed to estimate each entry of $\underline{\boldsymbol{s}_1}$ by minimizing the receiver output power while preserving a unity gain for this particular entry of $\underline{\boldsymbol{s}_1}$, that is [103]

$$\min_{\boldsymbol{w}_k} \boldsymbol{w}_k^T \hat{\boldsymbol{R}} \boldsymbol{w}_k \qquad \text{subject to} \qquad \boldsymbol{a}_k^T(\boldsymbol{H}_1)\boldsymbol{w}_k = 1 \quad \text{for all} \quad k = 1, \ldots, 2K \tag{4.15}$$

where

$$\hat{\boldsymbol{R}} = \frac{1}{J} \sum_{i=1}^{J} \underline{\boldsymbol{Y}_i}\, \underline{\boldsymbol{Y}_i}^T$$

is the sample estimate of the $2MT \times 2MT$ full rank covariance matrix

$$\boldsymbol{R} \triangleq \mathrm{E}\{\underline{\boldsymbol{Y}}\; \underline{\boldsymbol{Y}}^T\}$$

of the vectorized data, $\boldsymbol{Y}_i$ is the ith received data block.

The solution to (4.15) is given by [103]

$$\boldsymbol{W}_{\mathrm{MV}} = [\boldsymbol{w}_{\mathrm{MV},1} \; \cdots \; \boldsymbol{w}_{\mathrm{MV},2K}] \tag{4.16}$$

where

$$\boldsymbol{w}_{\mathrm{MV},k} = \frac{1}{\boldsymbol{a}_k^T(\boldsymbol{H}_1)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_k(\boldsymbol{H}_1)}\, \hat{\boldsymbol{R}}^{-1}\boldsymbol{a}_k(\boldsymbol{H}_1)\,, \quad k = 1, \ldots, 2K\,. \tag{4.17}$$

The form of the obtained MV receiver (4.16)-(4.17) is similar to that of the minimum variance distortionless response (MVDR) receiver used in beamforming [1]-[3] and minimum output energy (MOE) receiver used in multiuser detection [16]. Although the receiver (4.16) is able to reject MAI, it does not cancel *self-interference* [99] which, for each $\boldsymbol{w}_{\mathrm{MV},k}$, is caused by other entries of $\underline{\boldsymbol{s}_1}$ than the kth one. Note that the complete cancellation of self-interference is a very desirable feature because, otherwise, the symbol-by-symbol detector becomes non-optimal [103].

To incorporate the self-interference cancellation feature into (4.15), it was proposed in [103] to use additional zero-forcing constraints $\boldsymbol{a}_l^T(\boldsymbol{H}_1)\boldsymbol{w}_k = 0$ for all $l \neq k$. These constraints guarantee that self-interference is completely rejected.

With such additional constraints, the problem (4.15) can be reformulated as [103]

$$\min_{\boldsymbol{W}} \mathrm{tr}\{\boldsymbol{W}^T\hat{\boldsymbol{R}}\boldsymbol{W}\} \quad \text{subject to} \quad \boldsymbol{A}^T(\boldsymbol{H}_1)\boldsymbol{W} = \boldsymbol{I}_{2K}\,. \tag{4.18}$$

Using the Lagrange multiplier method, the solution to (4.18) can be written in the following form [103]

$$\tilde{\boldsymbol{W}}_{\mathrm{MV}} = \hat{\boldsymbol{R}}^{-1}\boldsymbol{A}(\boldsymbol{H}_1)\left[\boldsymbol{A}^T(\boldsymbol{H}_1)\hat{\boldsymbol{R}}^{-1}\boldsymbol{A}(\boldsymbol{H}_1)\right]^{-1}. \tag{4.19}$$

Note that the linear receivers (4.17) and (4.19) can be used not only in the case of multiple-access MIMO systems with OSTBCs, but also in a more general case where linear (not necessarily orthogonal) STBCs are used.

To improve the performance in the case of imperfect CSI and sample size, it was proposed in [103] to apply fixed diagonal loading (DL) to (4.17) and (4.19). Then, the DL-based modification of the MV receiver (4.16) can be written as

$$\boldsymbol{W}_{\mathrm{DLMV}} = [\boldsymbol{w}_{\mathrm{DLMV},1} \ \cdots \ \boldsymbol{w}_{\mathrm{DLMV},2K}] \tag{4.20}$$

where

$$\boldsymbol{w}_{\mathrm{DLMV},k} = \frac{1}{\boldsymbol{a}_k^T(\boldsymbol{H}_1)(\hat{\boldsymbol{R}} + \varrho\boldsymbol{I}_{2MT})^{-1}\boldsymbol{a}_k(\boldsymbol{H}_1)}\,(\hat{\boldsymbol{R}} + \varrho\boldsymbol{I}_{2MT})^{-1}\boldsymbol{a}_k(\boldsymbol{H}_1)\,,\ k=1,\ldots,2K \tag{4.21}$$

and ϱ is the fixed DL factor.

Similarly, the DL-based modification of the MV receiver (4.19) takes the form

$$\tilde{\boldsymbol{W}}_{\mathrm{DLMV}} = (\hat{\boldsymbol{R}} + \varrho \boldsymbol{I}_{2MT})^{-1} \boldsymbol{A}(\boldsymbol{H}_1) \left[\boldsymbol{A}^T(\boldsymbol{H}_1)(\hat{\boldsymbol{R}} + \varrho \boldsymbol{I}_{2MT})^{-1} \boldsymbol{A}(\boldsymbol{H}_1) \right]^{-1}. \tag{4.22}$$

Simulation results in [103] have demonstrated that the receiver (4.22) usually outperforms (4.20). Unfortunately, it is not clear from [103] what is the proper choice of ϱ in (4.21) and (4.22) and how it depends on the norm of the CSI errors. Furthermore, it is well known that the optimal choice of the DL factor is scenario-dependent [104], [111]. Therefore, the robustness of the fixed DL receivers (4.20) and (4.22) may be insufficient.

4.3 Robust Linear Receivers Based on Worst-Case Performance Optimization

In this section, we use a deterministic model for the CSI mismatch and propose a class of robust linear receivers under the framework of worst-case performance optimization [104], [107], [111]. In particular, we assume that the actual channel matrix resides in an uncertainty set, which can be viewed as a sphere in a matrix space. The center of this sphere is the presumed channel matrix, while its radius reflects the uncertainty level of the CSI mismatch. Based on this model, we propose robust receivers which can be seen as a generalization of the techniques of (4.15) and (4.18).

Let us assume that the exact channel matrix $\boldsymbol{H}_1$ is not available at the receiver. The only available quantity is its estimate $\hat{\boldsymbol{H}}_1$ which represents a distorted (mismatched) copy of $\boldsymbol{H}_1$. Let us introduce the error matrix

$$\boldsymbol{\Delta}_1 \triangleq \boldsymbol{H}_1 - \hat{\boldsymbol{H}}_1 \tag{4.23}$$

between the actual channel matrix $\boldsymbol{H}_1$ and its presumed (estimated) value $\hat{\boldsymbol{H}}_1$ and let the Frobenius norm of this error matrix be upper bounded by a known constant ε, that is

$$\|\boldsymbol{\Delta}_1\| \leq \varepsilon\,. \tag{4.24}$$

Let us define the mismatched space-time signatures $\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)$ $(k = 1, \ldots, 2K)$ of the desired user through the matrix

$$\begin{aligned} \boldsymbol{A}(\hat{\boldsymbol{H}}_1) &= [\underline{\boldsymbol{C}_1\hat{\boldsymbol{H}}_1} \cdots \underline{\boldsymbol{C}_K\hat{\boldsymbol{H}}_1}\ \underline{\boldsymbol{D}_1\hat{\boldsymbol{H}}_1} \cdots \underline{\boldsymbol{D}_K\hat{\boldsymbol{H}}_1}] \\ &\triangleq [\boldsymbol{a}_1(\hat{\boldsymbol{H}}_1) \cdots \boldsymbol{a}_{2K}(\hat{\boldsymbol{H}}_1)]\,. \end{aligned} \tag{4.25}$$

The following Lemma will be needed to derive our robust MV receivers.

Lemma 4.1. *For any OSTBC,*

$$\|\boldsymbol{\Delta}_1\| = \|\boldsymbol{e}_{k,1}\| \quad \textit{for all} \quad k = 1, \ldots, 2K \tag{4.26}$$

where

$$\boldsymbol{e}_{k,1} \triangleq \boldsymbol{a}_k(\boldsymbol{H}_1) - \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)\,. \tag{4.27}$$

PROOF: see Section 4.A. □

The sought robust modification of (4.15) should minimize the output power subject to the constraint that the distortionless response is maintained for the set of mismatched real-valued space-time signature vectors $\mathcal{A}(\varepsilon) = \{\boldsymbol{c}_k \mid \boldsymbol{c}_k = \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1 + \boldsymbol{\Delta}),\ \|\boldsymbol{\Delta}\| \leq \varepsilon\}$. This formulation corresponds to the spherical uncertainty set case [104]. Then, the robust modification of (4.15) can be written as the following optimization problem

$$\min_{\boldsymbol{w}_k} \boldsymbol{w}_k^T \hat{\boldsymbol{R}} \boldsymbol{w}_k \quad \text{subject to} \quad \min_{\|\boldsymbol{\Delta}\| \leq \varepsilon} \boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1 + \boldsymbol{\Delta}) \geq 1 \quad \text{for all} \quad k = 1, \ldots, 2K\,. \tag{4.28}$$

The main modification of (4.28) with respect to (4.15) is that, for each k, instead of requiring fixed distortionless response towards the single mismatched space-time signature $\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)$, in (4.28) such distortionless response is maintained by means of inequality constraints for a continuum of all space-time signatures given by the set $\mathcal{A}(\varepsilon)$. If (4.24) is satisfied, then from (4.28) it follows that the distortionless response is also maintained for the actual space-time signature $\boldsymbol{a}_k(\boldsymbol{H}_1) = \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1 + \boldsymbol{\Delta}_1)$. The constraints in (4.28) guarantee that the distortionless response will be maintained in the *worst case*, i.e., for the particular vector $\boldsymbol{c}_k \in \mathcal{A}\{\varepsilon\}$ which corresponds to the smallest value of $\boldsymbol{w}_k^T \boldsymbol{c}_k$.

Using Lemma 4.1, this problem can be transformed to

$$\min_{\boldsymbol{w}_k} \boldsymbol{w}_k^T \hat{\boldsymbol{R}} \boldsymbol{w}_k \quad \text{subject to} \qquad \min_{\|\boldsymbol{e}_k\| \le \varepsilon} \boldsymbol{w}_k^T \left(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_k \right) \ge 1 \qquad \text{for all} \quad k = 1, \ldots, 2K \,. \tag{4.29}$$

The problem mathematically similar to (4.29) arises in adaptive beamforming [104], [107] and multiuser detection [111]. Using the results of [104], it can be straightforwardly shown that in the case of a moderate mismatch parameter ε (i.e., in the case when $\boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) > \varepsilon \|\boldsymbol{w}_k\|$), this problem is equivalent to

$$\min_{\boldsymbol{w}_k} \boldsymbol{w}_k^T \hat{\boldsymbol{R}} \boldsymbol{w}_k \qquad \text{subject to} \qquad \boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) - \varepsilon \|\boldsymbol{w}_k\| = 1 \quad \text{for all} \quad k = 1, \ldots, 2K \,. \tag{4.30}$$

Applying the Lagrange multiplier method to (4.30), for each $k = 1, \ldots, 2K$ we obtain that the solution to (4.30) is given by the equation [111]

$$2\hat{\boldsymbol{R}} \boldsymbol{w}_k + \mu\varepsilon \frac{\boldsymbol{w}_k}{\|\boldsymbol{w}_k\|} = \mu \, \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) \tag{4.31}$$

where μ is the unknown Lagrange multiplier. To get around the problem of computing μ, let us assume that constant modulus symbol constellations are used. Hence, the vector $\boldsymbol{w}_k$ can be rescaled by an arbitrary constant without affecting the receiver performance [111]. Using this fact and rescaling $\boldsymbol{w}_k$ as $\boldsymbol{w}_k := \boldsymbol{w}_k/\mu$, we can rewrite (4.31) as

$$\boldsymbol{w}_k = \left(2\hat{\boldsymbol{R}} + \frac{\varepsilon}{\|\boldsymbol{w}_k\|} \boldsymbol{I}_{2MT} \right)^{-1} \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) \,. \tag{4.32}$$

Note that the term $\varepsilon/\|\boldsymbol{w}_k\|$ can be interpreted as an *adaptive* DL factor which is optimally matched to the given level ε of the channel uncertainty. To solve (4.32), we can apply a technique similar to that developed in [111]. From (4.32) we obtain that the optimal value of $\|\boldsymbol{w}_k\|$ can be found as the root of the following nonlinear equation [111]

$$\sum_{i=1}^{2MT} \left(\frac{[\tilde{\boldsymbol{a}}_k(\hat{\boldsymbol{H}}_1)]_i}{2\lambda_i \|\boldsymbol{w}_k\| + \varepsilon} \right)^2 = 1 \tag{4.33}$$

where

$$\hat{\boldsymbol{R}} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^T \tag{4.34}$$

is the eigenvalue decomposition of $\hat{\boldsymbol{R}}$; $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1\ \lambda_2\ \cdots\ \lambda_{2MT}\}$ is the diagonal matrix of the eigenvalues of $\hat{\boldsymbol{R}}$; $\tilde{\boldsymbol{a}}_k(\hat{\boldsymbol{H}}_1) = \boldsymbol{U}^T\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)$; and $[\cdot]_i$ denotes the ith element of a vector.

Standard methods such as Newton-Raphson technique can be applied to solve equation (4.33), see [111] for more details. Once this equation is solved, the obtained value of $\|\boldsymbol{w}_k\|$ can be inserted into the right-hand side of (4.32) to compute the optimal vector $\boldsymbol{w}_k$. Repeating this procedure for all $k = 1, \cdots, 2K$, we obtain the optimal weight matrix $\boldsymbol{W}$ which is the solution of (4.28).

Next, let us develop a robust modification of the receiver (4.19). To obtain such a modification, we should add worst-case zero-forcing constraints for self-interference. Following this idea and taking into account that in this case it is impossible to reject self-interference completely, we add to (4.29) additional constraints to limit the contribution of self-interference to the receiver output power. Then, for each k, our problem takes the following form

$$\min_{\boldsymbol{w}_k} \boldsymbol{w}_k^T\hat{\boldsymbol{R}}\boldsymbol{w}_k \quad \text{subject to} \quad \begin{array}{l} \min\limits_{\|\boldsymbol{e}_k\|\le\varepsilon} \boldsymbol{w}_k^T\left(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_k\right) \ge 1 \\ \max\limits_{\|\boldsymbol{E}_k\|\le\eta} \|(\boldsymbol{B}_k^T + \boldsymbol{E}_k^T)\boldsymbol{w}_k\| \le \delta_k \end{array} \tag{4.35}$$

where the $2MT \times (2K-1)$ matrices $\boldsymbol{B}_k$ and $\boldsymbol{E}_k$ are defined as

$$\begin{aligned} \boldsymbol{B}_k &\triangleq [\boldsymbol{a}_1(\hat{\boldsymbol{H}}_1)\ \cdots\ \boldsymbol{a}_{k-1}(\hat{\boldsymbol{H}}_1)\ \boldsymbol{a}_{k+1}(\hat{\boldsymbol{H}}_1)\ \cdots\ \boldsymbol{a}_{2K}(\hat{\boldsymbol{H}}_1)] \\ \boldsymbol{E}_k &\triangleq [\boldsymbol{e}_1\ \cdots\ \boldsymbol{e}_{k-1}\ \boldsymbol{e}_{k+1}\ \cdots\ \boldsymbol{e}_{2K}] \end{aligned}$$

respectively, δ_k is the value which limits the contribution of self-interference in the uncertainty region $\|\boldsymbol{E}_k\| \le \eta$, and η is the upper bound for $\|\boldsymbol{E}_k\|$.

Lemma 4.2. *For any OSTBC,*

$$\eta = \varepsilon\sqrt{2K-1}\,. \tag{4.36}$$

PROOF: see Section 4.B. □

Using triangle and Cauchy-Schwartz inequalities along with $\|\boldsymbol{E}_k\| \le \eta$ and Lemma 4.2,

we have

$$\begin{aligned}\|(\boldsymbol{B}_k^T + \boldsymbol{E}_k^T)\boldsymbol{w}_k\| &\leq \|\boldsymbol{B}_k^T\boldsymbol{w}_k\| + \|\boldsymbol{E}_k^T\boldsymbol{w}_k\| \\ &\leq \|\boldsymbol{B}_k^T\boldsymbol{w}_k\| + \|\boldsymbol{E}_k\|\|\mathbf{w}_k\| \\ &\leq \|\boldsymbol{B}_k^T\boldsymbol{w}_k\| + \eta\|\boldsymbol{w}_k\| \\ &= \|\boldsymbol{B}_k^T\boldsymbol{w}_k\| + \varepsilon\sqrt{2K-1}\,\|\boldsymbol{w}_k\|\,. \end{aligned} \tag{4.37}$$

It can be readily verified that all the inequalities in (4.37) become equalities if

$$\boldsymbol{E}_k = \frac{\eta\boldsymbol{w}_k\boldsymbol{w}_k^T\boldsymbol{B}_k}{\|\boldsymbol{w}_k\|\|\boldsymbol{B}_k^T\boldsymbol{w}_k\|}\,.$$

Using the latter observation and (4.37), we have that

$$\max_{\|\boldsymbol{E}_k\|\leq\eta} \|(\boldsymbol{B}_k^T + \boldsymbol{E}_k^T)\boldsymbol{w}_k\| = \|\boldsymbol{B}_k^T\boldsymbol{w}_k\| + \varepsilon\sqrt{2K-1}\,\|\boldsymbol{w}_k\|\,. \tag{4.38}$$

Note that, to zero-force self-interference in the uncertainty region as much as possible, the parameter δ_k in (4.35) should be chosen as small as possible (subject to the constraint that this problem remains feasible). The problem of potential infeasibility and, correspondingly, the problem of choice of δ_k can be avoided by treating δ_k as a variable to be minimized. Following this idea, let us add δ_k to the objective function in (4.35). Also, let us use (4.38) to simplify the second constraint in (4.35). Then, we obtain the following problem:

$$\min_{\boldsymbol{w}_k,\delta_k} \sqrt{\boldsymbol{w}_k^T\hat{\boldsymbol{R}}\boldsymbol{w}_k} + \delta_k \quad \text{subject to} \quad \begin{aligned} &\boldsymbol{w}_k^T\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) - \varepsilon\|\boldsymbol{w}_k\| \geq 1 \\ &\|\boldsymbol{B}_k^T\boldsymbol{w}_k\| + \varepsilon\sqrt{2K-1}\,\|\boldsymbol{w}_k\| \leq \delta_k\,. \end{aligned} \tag{4.39}$$

Now, let us convert this problem to the convex SOCP form [116]. The canonical SOCP problem has the following formulation [116]

$$\begin{aligned} \min_{\boldsymbol{z}} \quad & \boldsymbol{f}^T\boldsymbol{z} \\ \text{subject to} \quad & \|\boldsymbol{Q}_i\boldsymbol{z} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^T\boldsymbol{z} + d_i, \quad i = 1,\ldots,I \end{aligned} \tag{4.40}$$

where $\boldsymbol{z} \in \mathbb{R}^n$ are the optimization variables, and $\boldsymbol{f} \in \mathbb{R}^n$, $\boldsymbol{Q}_i \in \mathbb{R}^{(n_i-1)\times n}$, $\boldsymbol{b}_i \in \mathbb{R}^{n_i-1}$, $\boldsymbol{c}_i \in \mathbb{R}^n$, and $d_i \in \mathbb{R}$ are the problem parameters. Here, $\mathbb{R}$ is the set of all real numbers. The constraint

$$\|\boldsymbol{Q}_i\boldsymbol{z} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^T\boldsymbol{z} + d_i \tag{4.41}$$

is known as the second-order cone (SOC) of dimension n_i.

To convert the optimization problem (4.39) to the canonical SOCP problem, we now use the Cholesky decomposition of $\hat{\boldsymbol{R}}$:

$$\hat{\boldsymbol{R}} = \boldsymbol{L}^T \boldsymbol{L} \tag{4.42}$$

where $\boldsymbol{L}$ is an upper-triangular matrix. Using (4.42), we obtain that

$$\sqrt{\boldsymbol{w}_k^T \hat{\boldsymbol{R}} \boldsymbol{w}_k} = \sqrt{\boldsymbol{w}_k^T \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{w}_k} = \|\boldsymbol{L}\boldsymbol{w}_k\| \,. \tag{4.43}$$

Making use of (4.43) and introducing a new auxiliary variable τ_1 which satisfies the inequality $\tau_1 \geq \|\boldsymbol{L}\boldsymbol{w}_k\|$, the optimization problem (4.39) can be equivalently rewritten as

$$\begin{aligned} \min_{\tau_1, \delta_k, \boldsymbol{w}_k} \tau_1 + \delta_k \quad \text{subject to} \quad & \|\boldsymbol{L}\boldsymbol{w}_k\| \leq \tau_1 \\ & \varepsilon \|\boldsymbol{w}_k\| \leq \boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) - 1 \\ & \|\boldsymbol{B}_k^T \boldsymbol{w}_k\| + \varepsilon \sqrt{2K-1}\, \|\boldsymbol{w}_k\| \leq \delta_k \,. \end{aligned} \tag{4.44}$$

The first and the second constraints in (4.44) are already written in the SOCP form. Let us now convert the third constraint of (4.44) into SOCP constraints. Introducing auxiliary variables τ_2 and τ_3, this constraint can be written as

$$\begin{aligned} \|\boldsymbol{B}_k^T \boldsymbol{w}_k\| &\leq \tau_2 \\ \varepsilon \sqrt{2K-1}\, \|\boldsymbol{w}_k\| &\leq \tau_3 \\ \tau_2 + \tau_3 &\leq \delta_k \,. \end{aligned} \tag{4.45}$$

Replacing the third constraint in (4.44) with (4.45), we finally obtain the following equivalent form of the problem (4.35):

$$\begin{aligned} \min_{\boldsymbol{\tau}, \delta_k, \boldsymbol{w}_k} \tau_1 + \delta_k \quad \text{subject to} \quad & \|\boldsymbol{L}\boldsymbol{w}_k\| \leq \tau_1 \\ & \varepsilon \|\boldsymbol{w}_k\| \leq \boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) - 1 \\ & \|\boldsymbol{B}_k^T \boldsymbol{w}_k\| \leq \tau_2 \\ & \varepsilon \sqrt{2K-1}\, \|\boldsymbol{w}_k\| \leq \tau_3 \\ & \tau_2 + \tau_3 \leq \delta_k \end{aligned} \tag{4.46}$$

where $\boldsymbol{\tau} = [\tau_1, \tau_2, \tau_3]^T$. Comparing (4.46) with the canonical form of a SOCP problem given in (4.40), one can easily see that if $\boldsymbol{z} = [\boldsymbol{w}_k^T, \boldsymbol{\tau}^T, \delta_k]^T$, then we have five SOCs with $\boldsymbol{Q}_1 = [\boldsymbol{L} \;\; \boldsymbol{0}_{2MT\times 4}]$, $\boldsymbol{b}_1 = \boldsymbol{0}_{2MT\times 1}$, $\boldsymbol{c}_1 = [\boldsymbol{0}_{1\times 2MT} \;\; 1 \;\; \boldsymbol{0}_{1\times 3}]^T$, $d_1 = 0$; $\boldsymbol{Q}_2 = [\varepsilon \boldsymbol{I}_{2MT} \;\; \boldsymbol{0}_{2MT\times 4}]$, $\boldsymbol{b}_2 = \boldsymbol{0}_{2MT\times 1}$, $\boldsymbol{c}_2 = [\boldsymbol{a}_k^T(\hat{\boldsymbol{H}}_1) \;\; \boldsymbol{0}_{1\times 4}]^T$, $d_2 = -1$; $\boldsymbol{Q}_3 = [\boldsymbol{B}_k^T \;\; \boldsymbol{0}_{(2K-1)\times 4}]$, $\boldsymbol{b}_3 = \boldsymbol{0}_{(2K-1)\times 1}$, $\boldsymbol{c}_3 = [\boldsymbol{0}_{1\times(2MT+1)} \;\; 1 \;\; \boldsymbol{0}_{1\times 2}]^T$, $d_3 = 0$; $\boldsymbol{Q}_4 = [\varepsilon\sqrt{2K-1}\,\boldsymbol{I}_{2MT} \;\; \boldsymbol{0}_{2MT\times 4}]$, $\boldsymbol{b}_4 = \boldsymbol{0}_{2MT\times 1}$, $\boldsymbol{c}_4 = [\boldsymbol{0}_{1\times(2MT+2)} \;\; 1 \;\; 0]^T$, $d_4 = 0$; $\boldsymbol{Q}_5 = \boldsymbol{0}_{1\times(2MT+4)}$, $\boldsymbol{b}_5 = 0$, $\boldsymbol{c}_5 = [\boldsymbol{0}_{1\times(2MT)} \;\; 0 \;\; -1 \;\; -1 \;\; 1]^T$, $d_5 = 0$.

The problem (4.46) represents a convex SOCP problem which can be straightforwardly and efficiently solved using interior point algorithms [116], [117].

Note that (4.46) can be solved for each value of k ($k = 1, \ldots, 2K$) independently. Also, in contrast to our first robust receiver (4.32), the receiver (4.46) is not restricted by constant modulus symbol constellations.

It is also worth noting that the proposed receivers (4.32) and (4.46) do not need any knowledge of the channel matrices of interfering users.

The main computational cost of our first receiver (4.32) is determined by the matrix inversion and eigendecomposition operations in (4.32) and (4.34), respectively. Therefore, the complexity of this receiver is $\mathcal{O}(M^3T^3)$. The complexity of our second receiver (4.46) is mainly determined by the complexity of the corresponding interior point algorithm used to solve the SOCP problem (4.46) and is equal to $\mathcal{O}(M^3T^3)$ per iteration [116]. Typically, less than ten to fifteen iterations are required to converge (a commonly accepted fact in the optimization community [104] which is also gained by our extensive simulations.)

Summarizing, our second receiver may have slightly higher computational complexity than the first one and also requires a specific built-in convex optimization software. This moderate increase in the implementational complexity of (4.46) is compensated by its more general application to non-constant modulus signal constellations and, as shown in Section 4.5, by remarkable performance improvements over (4.32).

4.4 Robust Linear Receivers Based on Stochastic Programming

Worst-case designs may be overly conservative because the probability of occurrence of the worst-case mismatch may be very low. In this section, we propose a less conservative approach to robust linear MIMO receiver design based on stochastic programming [119], [120]. It guarantees the robustness against CSI errors with a certain selected probability. Our approach is based on a stochastic model of the CSI mismatch. The justification of this model follows from the fact that in MIMO communication systems, orthogonal sequences are optimal for training if nothing is known *a priori* about the channel. Then, it can be proven that for orthogonal training sequences and additive white Gaussian noise (AWGN), the CSI errors are Gaussian [122], [123, eq. (12)]. In order to show explicitly the functional relationship between $\boldsymbol{e}_k$ and $\boldsymbol{\Delta}_1$, we denote hereafter $\boldsymbol{e}_k(\boldsymbol{\Delta}_1)$, $k = 1, \cdots, 2K$ as the mismatch vectors of the space-time signature.

Let us obtain the receiver coefficient vector $\boldsymbol{w}_k$ for the kth entry of $\underline{\boldsymbol{s}_1}$ as the solution of the following probability-constrained optimization problem

$$\min_{\boldsymbol{w}_k, \boldsymbol{\delta}} \quad \boldsymbol{w}_k^T \hat{\boldsymbol{R}} \boldsymbol{w}_k + \|\boldsymbol{\delta}\|^2 \tag{4.47}$$

$$\text{subject to} \quad \Pr\{\boldsymbol{w}_k^T(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_k(\boldsymbol{\Delta}_1)) \geq 1\} \geq \zeta, \tag{4.48}$$

$$\Pr\{\sigma_1 |\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_l(\boldsymbol{\Delta}_1))| \leq \delta_l\} \geq \zeta, \tag{4.49}$$

$$l = 1, \ldots, 2K, \;\; l \neq k$$

where $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_{k-1}, \delta_{k+1}, \cdots, \delta_{2K}]^T$ is the $(2K-1) \times 1$ vector whose values limit the contribution of self-interference, σ_1 is the standard deviation of the waveform of the desired user, ζ is a certain probability value which should be selected according to the quality of service (QoS) requirements, and $\Pr\{\cdot\}$ denotes the probability operator whose form is assumed to be known. It is important to note that the probability bound ζ in the problem (4.47)-(4.49) can be selected from the interval $(0, 1)$ and it determines the amount of channel mismatch that is allowed at the receiver.

In the formulation (4.47)-(4.49), the receiver output power is minimized, while the

distortionless response for the kth entry of $\underline{\boldsymbol{s}_1}$ is kept with a certain probability. The term $\|\boldsymbol{\delta}\|^2$ in the objective function (4.47), together with the constraints in (4.49) guarantee that the self-interference is suppressed with a certain selected probability. The motivation for this is that in MIMO communications, the guaranteed self-interference suppression is very important for the symbol-by-symbol detector (see the discussion in Sections 4.2 and 4.3). Problem (4.47)-(4.49) is called in the optimization literature the *chance-constrained* or *probability-constrained stochastic programming* problem [119], [120]. The constraints (4.48) and (4.49) can be also interpreted as *non-outage probability constraints*.

Note that in the objective function (4.47), $\hat{\boldsymbol{R}}$ is proportional to the power of MAI and self-interference. Therefore, to achieve a good performance in the whole SNR range, $\|\boldsymbol{\delta}\|^2$ should also be proportional to the self-interference power (power of the user-of-interest). This is taken into account by the term σ_1 in the constraints in (4.49), because in order to satisfy the lth constraint in (4.49), δ_l is proportional to σ_1 for a given ζ.

Theorem 4.1. *If the elements of $\boldsymbol{\Delta}_1$ are uncorrelated and each element has circular complex Gaussian distribution: $[\boldsymbol{\Delta}_1]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2)$, then the constraints in (4.49) are convex if $\zeta \in [0.5, 1)$.*

PROOF: See Section 4.C. □

Theorem 4.2. *If the elements of $\boldsymbol{\Delta}_1$ are uncorrelated and each element has circular complex Gaussian distribution: $[\boldsymbol{\Delta}_1]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2)$, then the optimization problem (4.47)-(4.49) is convex if $\zeta \in (0.5, 1)$.*

PROOF: We first observe that the objective function (4.47) is a summation of two convex quadratic functions. Thus, it is convex.

Now let us prove that the constraint (4.48) is also convex under the assumptions of the theorem. First, from Section 4.A, we find that $\boldsymbol{e}_k(\boldsymbol{\Delta}_1)$ is a linear combination of the real and imaginary parts of the elements of the channel mismatch matrix $\boldsymbol{\Delta}_1$. If the elements of $\boldsymbol{\Delta}_1$ are uncorrelated and have circular complex Gaussian distribution

$$[\boldsymbol{\Delta}_1]_{n,m} \sim \mathcal{CN}(0, \sigma_h^2), \quad n = 1, \ldots N, \quad m = 1, \ldots, M$$

then, we find that $\boldsymbol{e}_k(\boldsymbol{\Delta}_1)$ has multivariate real Gaussian distribution [124]. Its mean vector and covariance matrix can be calculated, respectively, as

$$\mathrm{E}\{\boldsymbol{e}_k(\boldsymbol{\Delta}_1)\} = \boldsymbol{\Psi}_k \mathrm{E}\{\underline{\boldsymbol{\Delta}_1}\} = \mathbf{0}_{2MT\times 1}$$

$$\begin{aligned}\mathrm{E}\{\boldsymbol{e}_k(\boldsymbol{\Delta}_1)\boldsymbol{e}_k^T(\boldsymbol{\Delta}_1)\} &= \boldsymbol{\Psi}_k \mathrm{E}\{\underline{\boldsymbol{\Delta}_1}\,\underline{\boldsymbol{\Delta}_1}^T\}\boldsymbol{\Psi}_k^T \\ &= \frac{\sigma_h^2}{2}\boldsymbol{\Psi}_k\boldsymbol{\Psi}_k^T \\ &= \frac{\sigma_h^2}{2}\big(\boldsymbol{I}_{2M}\otimes \boldsymbol{F}_k\boldsymbol{F}_k^T\big)\end{aligned}$$

where

$$\boldsymbol{F}_k = \begin{cases} \boldsymbol{C}_k, & k = 1,\ldots,K \\ \mathrm{Im}\{\boldsymbol{D}_{k-K}\}, & k = K+1,\ldots,2K. \end{cases}$$

Since only $\boldsymbol{e}_k(\boldsymbol{\Delta}_1)$ is a random variable in the product $\boldsymbol{w}_k^T(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_k(\boldsymbol{\Delta}_1))$, and both $\boldsymbol{w}_k$ and $\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)$ are deterministic values, the random variable $\boldsymbol{w}_k^T(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_k(\boldsymbol{\Delta}_1))$ has real Gaussian distribution. Its mean and covariance can be computed, respectively, as

$$\begin{aligned}\mathrm{E}\{\boldsymbol{w}_k^T(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_k(\boldsymbol{\Delta}_1))\} &= \boldsymbol{w}_k^T\mathrm{E}\{\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_k(\boldsymbol{\Delta}_1)\} \\ &= \boldsymbol{w}_k^T\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) \qquad (4.50)\end{aligned}$$

$$\begin{aligned}\mathrm{E}\{\boldsymbol{w}_k^T\boldsymbol{e}_k(\boldsymbol{\Delta}_1)\boldsymbol{e}_k^T(\boldsymbol{\Delta}_1)\boldsymbol{w}_k\} &= \boldsymbol{w}_k^T\mathrm{E}\{\boldsymbol{e}_k(\boldsymbol{\Delta}_1)\boldsymbol{e}_k^T(\boldsymbol{\Delta}_1)\}\boldsymbol{w}_k \\ &= \frac{\sigma_h^2}{2}\boldsymbol{w}_k^T\big(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_k\boldsymbol{F}_k^T\big)\boldsymbol{w}_k \\ &= \frac{\sigma_h^2}{2}\big\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_k^T)\boldsymbol{w}_k\big\|^2. \qquad (4.51)\end{aligned}$$

Using the standard error function for Gaussian distribution

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \qquad (4.52)$$

the left hand side of the constraint (4.48) can be written as

$$\Pr\{\boldsymbol{w}_k^T(\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_k(\boldsymbol{\Delta}_1))\geq 1\} = \frac{1}{2}-\frac{1}{2}\,\mathrm{erf}\left(\frac{1-\boldsymbol{w}_k^T\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)}{\sigma_h\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_k^T)\boldsymbol{w}_k\|}\right). \qquad (4.53)$$

Substituting (4.53) into (4.48), we obtain the following constraint

$$\frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{1 - \boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)}{\sigma_h \|(\boldsymbol{I}_{2M} \otimes \boldsymbol{F}_k^T)\boldsymbol{w}_k\|}\right) \geq \zeta$$

which can be equivalently written as

$$\mathrm{erf}\left(\frac{\boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) - 1}{\sigma_h \|(\boldsymbol{I}_{2M} \otimes \boldsymbol{F}_k^T)\boldsymbol{w}_k\|}\right) \geq 2\zeta - 1. \tag{4.54}$$

If $\zeta > 0.5$, we have $2\zeta - 1 > 0$. In this case, (4.54) can be written as

$$\sigma_h \|(\boldsymbol{I}_{2M} \otimes \boldsymbol{F}_k^T)\boldsymbol{w}_k\| \leq \frac{\boldsymbol{w}_k^T \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1) - 1}{\mathrm{erf}^{-1}(2\zeta - 1)} \tag{4.55}$$

where $\mathrm{erf}^{-1}(\cdot)$ denotes the inverse error function. The constraint (4.55) is a SOC constraint (see (4.41)), and thus it is convex.

Summarizing, both the objective function and the constraints are convex if $\zeta \in (0.5, 1)$. Therefore, the problem (4.47)-(4.49) is convex. □

Although the problem (4.47)-(4.49) is convex, it is nonlinear and does not have closed-form solution. However, the problem can be solved numerically by the NLP approach [121] in the following way.

Let us introduce a new vector

$$\boldsymbol{\varpi} \triangleq [(\boldsymbol{L}\boldsymbol{w}_k)^T, \boldsymbol{\delta}^T]^T.$$

where $\boldsymbol{L}^T\boldsymbol{L}$ is the Cholesky factorization of $\hat{\boldsymbol{R}}$ (4.42). Apparently, minimizing $\|\boldsymbol{\varpi}\|$ is equivalent to minimizing the objective function of (4.47). Hence, introducing a new non-negative scalar variable χ and a new constraint

$$\|\boldsymbol{\varpi}\| \leq \chi \tag{4.56}$$

we can write a new objective function as follows

$$\min_{\boldsymbol{w}_k, \boldsymbol{\delta}, \chi} \chi. \tag{4.57}$$

The deterministic equivalent form of the constraint (4.48) is given by (4.55). The probability constraints (4.49) also can be converted into their deterministic equivalents. Using

(4.50)-(4.52), the left hand side of (4.49) can be written as

$$\begin{aligned} & \Pr\{\sigma_1|\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_l(\boldsymbol{\Delta}_1))|\le\delta_l\} \\ = \; & \Pr\{\sigma_1\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_l(\boldsymbol{\Delta}_1))\le\delta_l\}-\Pr\{\sigma_1\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_l(\boldsymbol{\Delta}_1))\le-\delta_l\} \\ = \; & \frac{1}{2}\,\mathrm{erf}\left(\frac{\delta_l-\sigma_1\boldsymbol{w}_k^T\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)}{\sigma_h\sigma_1\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l^T)\boldsymbol{w}_k\|}\right)-\frac{1}{2}\,\mathrm{erf}\left(\frac{-\delta_l-\sigma_1\boldsymbol{w}_k^T\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)}{\sigma_h\sigma_1\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l^T)\boldsymbol{w}_k\|}\right) \\ & \quad l=1,\ldots,2K,\; l\ne k. \end{aligned} \tag{4.58}$$

Combining (4.55)-(4.58) together, we can rewrite the stochastic programming problem (4.47)-(4.49) as the following deterministic NLP problem

$$\begin{aligned} \min_{\boldsymbol{w}_k,\boldsymbol{\delta},\chi} \quad & \chi \\ \text{s.t.} \quad & \|\boldsymbol{\varpi}\|\le\chi \\ & \sigma_h\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_k^T)\boldsymbol{w}_k\|\le\frac{\boldsymbol{w}_k^T\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)-1}{\mathrm{erf}^{-1}(2\zeta-1)} \\ & \mathrm{erf}\left(\frac{\delta_l-\sigma_1\boldsymbol{w}_k^T\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)}{\sigma_h\sigma_1\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l^T)\boldsymbol{w}_k\|}\right)-\mathrm{erf}\left(\frac{-\delta_l-\sigma_1\boldsymbol{w}_k^T\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)}{\sigma_h\sigma_1\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l^T)\boldsymbol{w}_k\|}\right)\ge2\zeta, \\ & l=1,\ldots,2K,\quad l\ne k. \end{aligned} \tag{4.59}$$

The problem (4.59) is equivalent to the problem (4.47)-(4.49). However, the complexity order of solving the former problem is $\mathcal{O}(M^{4.5}T^{4.5}K^{4.5})$ [162], which is quite high for practical wireless communication systems. Below we show that the problem (4.47)-(4.49) can be approximated to a SOCP problem.

The deterministic equivalent form of the constraint (4.48) is given by (4.55), which is a SOC constraint. However, the constraints (4.49) can not be directly converted to SOC constraints. To enable such conversion, we approximate the constraints (4.49) using the Chebyshev inequality that states that for any random variable ξ and any positive real number α,

$$\Pr\{|\xi|\ge\alpha\}\le\frac{\mathrm{E}\{\xi^2\}}{\alpha^2}. \tag{4.60}$$

Since the constraints (4.49) share the same structure, we further discuss only the lth constraint. Under the assumption that $\boldsymbol{e}_l(\boldsymbol{\Delta}_1)$ has Gaussian distribution and using (4.50) and

(4.51), we have

$$\mathrm{E}\Big\{\big|\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_l(\boldsymbol{\Delta}_1))\big|^2\Big\} = \boldsymbol{w}_k^T\bigg(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)\boldsymbol{a}_l^T(\hat{\boldsymbol{H}}_1)+\frac{\sigma_h^2}{2}(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l\boldsymbol{F}_l^T)\bigg)\boldsymbol{w}_k. \quad (4.61)$$

Applying (4.60) and (4.61), the left hand side of the lth constraint in (4.49) can be lower bounded by

$$\begin{aligned} & \Pr\{\sigma_1|\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_l(\boldsymbol{\Delta}_1))|\le\delta_l\} \\ = \;& 1-\Pr\{\sigma_1|\boldsymbol{w}_k^T(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)+\boldsymbol{e}_l(\boldsymbol{\Delta}_1))|\ge\delta_l\} \\ \ge \;& 1-\frac{\sigma_1^2}{\delta_l^2}\boldsymbol{w}_k^T\Big(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)\boldsymbol{a}_l^T(\hat{\boldsymbol{H}}_1)+\frac{\sigma_h^2}{2}(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l\boldsymbol{F}_l^T)\Big)\boldsymbol{w}_k\,. \end{aligned} \quad (4.62)$$

Using (4.62), we obtain the following approximated constraint

$$\frac{\sigma_1^2}{1-\zeta}\boldsymbol{w}_k^T\Big(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)\boldsymbol{a}_l^T(\hat{\boldsymbol{H}}_1)+\frac{\sigma_h^2}{2}(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l\boldsymbol{F}_l^T)\Big)\boldsymbol{w}_k\le\delta_l^2. \quad (4.63)$$

Furthermore, the approximated constraints in (4.49) can be summed up and written as

$$\boldsymbol{w}_k{}^T\boldsymbol{\Sigma}_k\boldsymbol{w}_k\le\|\boldsymbol{\delta}\|^2 \quad (4.64)$$

where

$$\boldsymbol{\Sigma}_k\triangleq\frac{\sigma_1^2}{1-\zeta}\sum_{l=1,l\ne k}^{2K}\Big[\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1)\boldsymbol{a}_l^T(\hat{\boldsymbol{H}}_1)+\frac{\sigma_h^2}{2}\big(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_l\boldsymbol{F}_l^T\big)\Big].$$

From (4.64), it can be seen that the modified constraints of (4.49) can be eliminated by including them in the objective function of (4.47). Using this trick, the objective function of (4.47) can be rewritten as

$$\boldsymbol{w}_k^T(\hat{\boldsymbol{R}}+\boldsymbol{\Sigma}_k)\boldsymbol{w}_k.$$

Let $\boldsymbol{Z}_k^T\boldsymbol{Z}_k$ be the Cholesky factorization of $\hat{\boldsymbol{R}}+\boldsymbol{\Sigma}_k$ and let us introduce a new variable τ such that $\|\boldsymbol{Z}_k\boldsymbol{w}_k\|\le\tau$. Then the optimization problem (4.47)-(4.49) can be modified to the following SOCP problem

$$\begin{aligned} \min_{\boldsymbol{w}_k,\tau}\quad & \tau \\ \text{subject to}\quad & \|\boldsymbol{Z}_k\boldsymbol{w}_k\|\le\tau \\ & \sigma_h\|(\boldsymbol{I}_{2M}\otimes\boldsymbol{F}_k^T)\boldsymbol{w}_k\|\le\frac{\boldsymbol{w}_k^T\boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)-1}{\mathrm{erf}^{-1}(2\zeta-1)}. \end{aligned} \quad (4.65)$$

The modified problem (4.65) can be easily solved using standard and highly efficient interior-point methods, for example, using the SeDuMi package [117]. Using the primal-dual potential reduction method, the problem (4.65) can be solved with the complexity order of $\mathcal{O}(M^3T^3)$ [85]. Moreover, the amount of computation per iteration in interior-point methods is $\mathcal{O}(n^2\sum_i n_i)$, where n is the dimension of the optimization variables, and n_i is the dimension of the ith SOC constraint [116]. Comparing the problem (4.46) with the problem (4.65), the latter one requires less computation, since it has two SOC constraints less than the former one.

4.5 Numerical Examples

In this section we study the performance of the proposed robust receivers through numerical simulations.

Throughout the simulations, we assume a single receiver of $M = 8$ antennas. The number of transmitters is varied in different simulation examples. The interfering transmitter uses the same OSTBC as the transmitter of interest. The interference-to-noise ratio (INR) is equal to 20 dB and the quadriphase shift keying (QPSK) modulation scheme is used. All plots are averaged over 1000 independent simulation runs. In each simulation run, the elements of the true channel matrices $\boldsymbol{H}_p$ (for $p = 1, \cdots, P$) are independently drawn from a complex Gaussian random generator with zero mean and unit variance.

The proposed robust receivers (4.32), (4.46), and (4.65) are compared with the MF receiver (4.14) and the DLMV receiver (4.22). The numbers in the figure legends refer to the equation number of the corresponding receiver in the text. Note that the imperfect CSI case is assumed, i.e., all these receivers use the presumed (erroneous) channel matrix $\hat{\boldsymbol{H}}_1$ rather than the actual channel matrix $\boldsymbol{H}_1$. In each simulation run, each element of the presumed channel matrix $\hat{\boldsymbol{H}}_1$ is generated by drawing a complex Gaussian random variable with zero mean and the variance $\sigma_h^2 = 0.1$ and adding this variable to a corresponding element of the matrix $\boldsymbol{H}_1$. Moreover, the performance of the so-called *informed* MV receiver is tested and included in all plots. The latter receiver corresponds to the ideal case when

(4.19) is used with the exactly known $\boldsymbol{H}_1$. Obviously, this receiver does not correspond to any practical situation and is included in our simulations for the sake of comparison only (as a benchmark).

The SeDuMi convex optimization MATLAB toolbox [117] has been used to solve the SOCP problems (4.46) and (4.65). SeDuMi toolbox applies interior-point method which is computationally efficient. We have observed that the interior-point method converges typically in less than 15 iterations.

The DL factor of $\varrho = 10\sigma_v^2$ is used in the DLMV receiver where σ_v^2 is the noise variance. Note that this is a popular *ad hoc* choice of ϱ [103], [104].

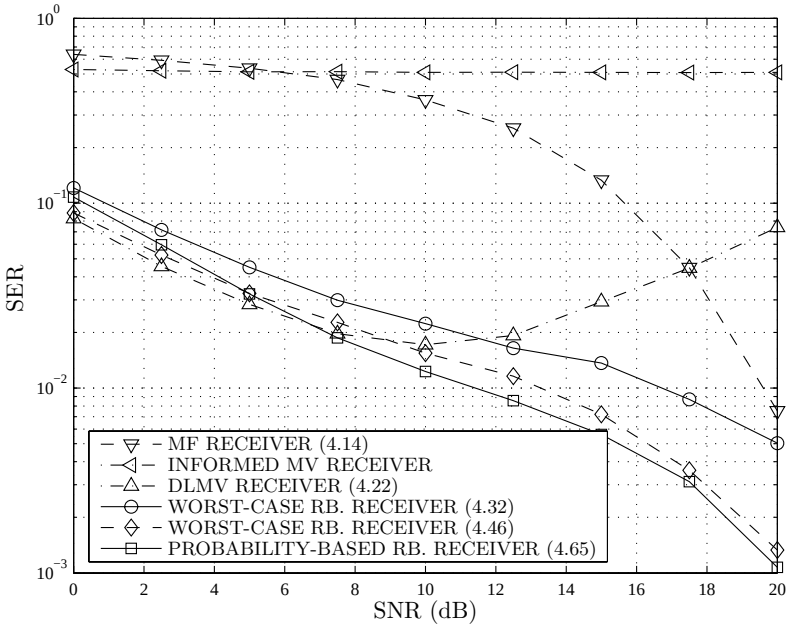

Figure 4.2: SER versus SNR; first example.

In the first example, $P = 2$ transmitters with $N = 2$ antennas per transmitter are assumed and the full-rate Alamouti's OSTBC ($T = 2$, $K = 2$) is used [96]. In this example, the parameter $\varepsilon = 6\sigma_h$ is used in the worst-case robust receivers (4.32) and (4.46), and $\zeta = 0.95$ is taken for the probability-based robust receiver (4.65). It should be noted that this value of ε is nearly optimal for this example. In Figure 4.2, SERs of all the receivers

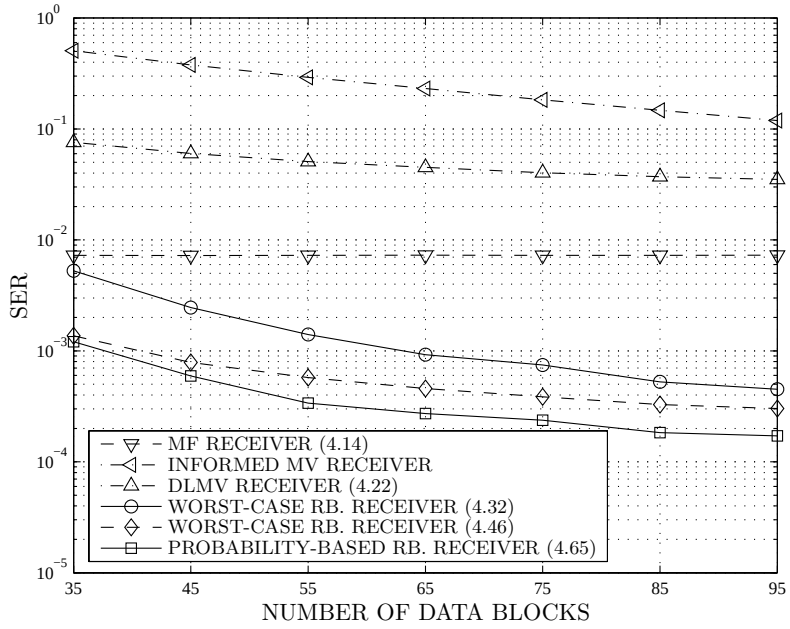

Figure 4.3: SER versus number of data blocks; first example.

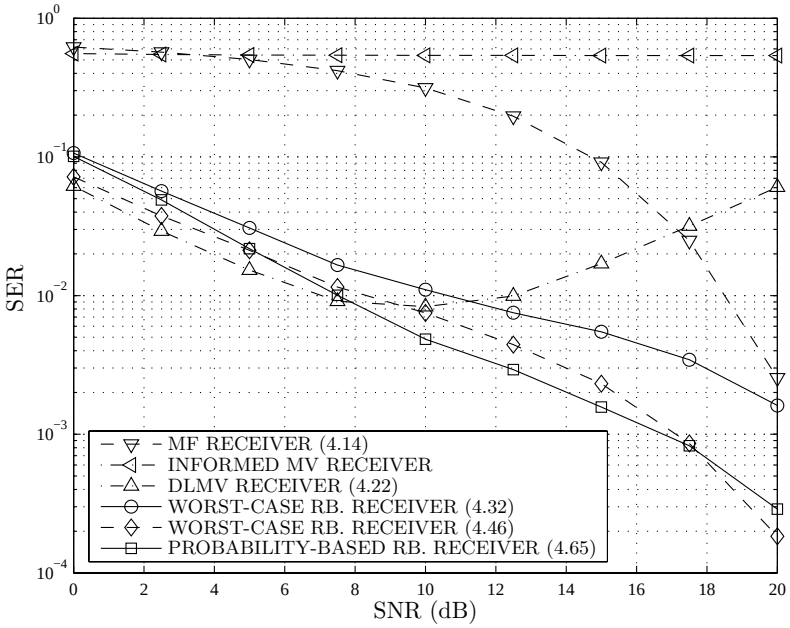

Figure 4.4: SER versus SNR; second example.

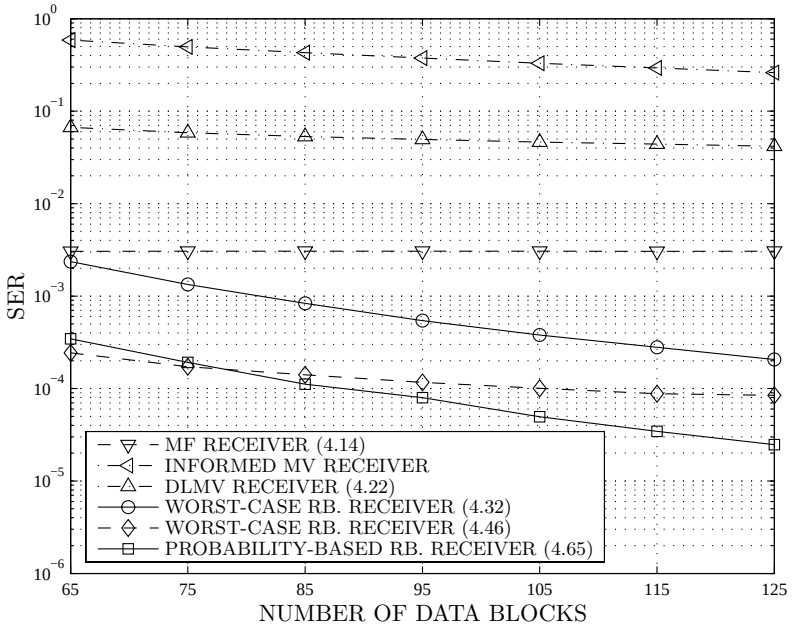

Figure 4.5: SER versus number of data blocks; second example.

tested are displayed versus the SNR for $J = 35$. Figure 4.3 shows the SERs of the same receivers versus the number of data blocks for SNR = 20 dB.

In the second example, we assume $P = 2$ transmitters. Each transmitter has $N = 3$ antennas and the 3/4-rate ($K = 3$; $T = 4$) orthogonal design STBC from [97] is used. The parameter $\varepsilon = 7\sigma_h$ is taken for the worst-case receivers (which is nearly optimal for this example), and $\zeta = 0.95$ is used for the probability-based receiver. Figure 4.4 shows the receiver SERs versus the SNR for $J = 70$, while Figure 4.5 displays the receiver SERs versus the number of data blocks for SNR = 20 dB.

In the third example, we assume $P = 4$ transmitters. Each transmitter has $N = 3$ antennas and the half-rate ($K = 4$; $T = 8$) orthogonal design STBC from [97] is used. The parameters $\varepsilon = 6\sigma_h$ and $\zeta = 0.99$ are taken in this example[2]. Figure 4.6 displays the SERs versus SNR of all the receivers tested for $J = 130$, while Figure 4.7 shows the receiver SERs

[2]We observed during simulations that the value of ζ should be set larger, with the increasing number of interferers. It can be explained that for the scenario with large number of users, the guaranteed distortionless response to the symbol of interest becomes very important.

versus the number of data blocks for SNR = 20 dB.

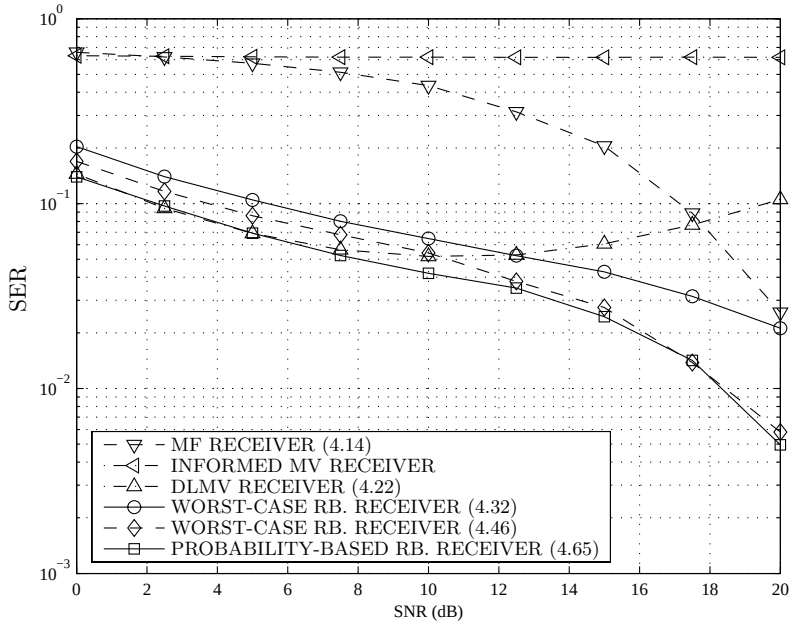

Figure 4.6: SER versus SNR; third example.

From Figures 4.2-4.7, it follows that in all examples, the proposed robust receivers (4.32), (4.46), and (4.65) provide better performance tradeoffs over the whole tested SNR range as compared with the other receivers (including the informed MV receiver). As expected, the receivers (4.46) and (4.65) substantially outperform (4.32) because the former techniques take advantage of an additional nulling of self-interference, see [103] and the discussion in Section 4.2. Furthermore, as it can be expected from [103], these performance improvements of (4.46) and (4.65) relative to (4.32) are especially pronounced at high SNRs. Comparing the worst-case receivers (4.32) and (4.46) with the probability-based receiver (4.65), we find that in most cases, the latter one shows better performance. The reason is that the probability-based robust designs are more flexible and less conservative than the worst-case approaches.

It can be observed from Figures 4.2, 4.4, and 4.6 that very substantial performance improvements over the DLMV receiver at high SNR values are achieved in our robust

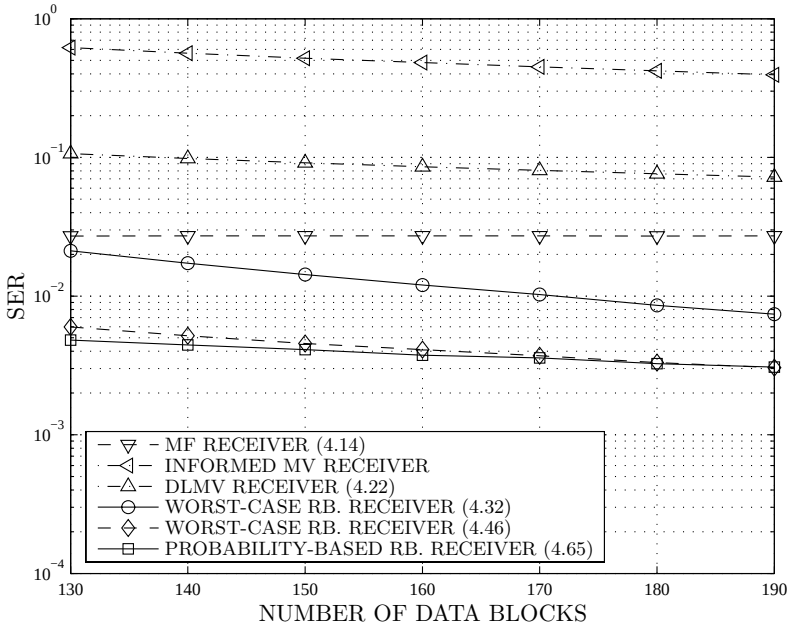

Figure 4.7: SER versus number of data blocks; third example.

receivers at the price of slightly worse performance at low SNR values (where both the DLMV and our techniques perform quite well). Such performance degradation of the DLMV receiver at high SNR values can be explained by the fact that this receiver uses the fixed DL factor. The poor performance of the informed MV receiver is due to its insufficient robustness against finite sample effects.

4.6 Chapter Summary

In this chapter, two classes of new linear receiver techniques for joint space-time decoding and interference rejection in multiple-access MIMO systems have been developed. These techniques use orthogonal space-time block codes and have an improved robustness against CSI errors. The first class of techniques uses the deterministic CSI mismatch model and the worst-case performance optimization approach is applied. The second class of receivers applies the stochastic CSI model and provides robustness against CSI errors with a certain

selected probability. Both classes of the proposed receivers have been shown to provide a substantially improved robustness against CSI mismatches as compared with the existing non-robust multiple-access MIMO receiver algorithms. Comparing the two classes of the proposed receivers, the latter one is more flexible and less conservative.

4.A Proof of Lemma 4.1

Let

$$\boldsymbol{G}_k \triangleq \begin{cases} \boldsymbol{C}_k, & k = 1, \ldots, K \\ \boldsymbol{D}_{k-K}, & k = K+1, \ldots, 2K\,. \end{cases} \tag{4.66}$$

Using the definition of the underline operator (4.4) along with well-known properties of the Kronecker product and vec$(\cdot)$ operator [118], we have

$$\begin{aligned} \boldsymbol{a}_k(\boldsymbol{H}_1) &= \begin{bmatrix} \mathrm{vec}(\mathrm{Re}\{\boldsymbol{G}_k\boldsymbol{H}_1\}) \\ \mathrm{vec}(\mathrm{Im}\{\boldsymbol{G}_k\boldsymbol{H}_1\}) \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Re}\{(\boldsymbol{I}_M \otimes \boldsymbol{G}_k)\mathrm{vec}(\boldsymbol{H}_1)\} \\ \mathrm{Im}\{(\boldsymbol{I}_M \otimes \boldsymbol{G}_k)\mathrm{vec}(\boldsymbol{H}_1)\} \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Re}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} & -\mathrm{Im}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} \\ \mathrm{Im}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} & \mathrm{Re}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\mathrm{Re}\{\boldsymbol{H}_1\}) \\ \mathrm{vec}(\mathrm{Im}\{\boldsymbol{H}_1\}) \end{bmatrix} \\ &= \boldsymbol{\Psi}_k \boldsymbol{h}_1 \end{aligned} \tag{4.67}$$

where

$$\boldsymbol{\Psi}_k \triangleq \begin{bmatrix} \mathrm{Re}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} & -\mathrm{Im}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} \\ \mathrm{Im}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} & \mathrm{Re}\{\boldsymbol{I}_M \otimes \boldsymbol{G}_k\} \end{bmatrix} \tag{4.68}$$

is a $2MT \times 2MN$ real-valued matrix, and $\boldsymbol{h}_1 \triangleq \underline{\boldsymbol{H}_1}$. Using the linearity of the underline operator (4.4), the vector $\boldsymbol{e}_{k,1} \triangleq \boldsymbol{a}_k(\boldsymbol{H}_1) - \boldsymbol{a}_k(\hat{\boldsymbol{H}}_1)$ can be written as

$$\boldsymbol{e}_{k,1} = \underline{\boldsymbol{G}_k\boldsymbol{H}_1} - \underline{\boldsymbol{G}_k\hat{\boldsymbol{H}}_1} = \underline{\boldsymbol{G}_k\boldsymbol{\Delta}_1} = \boldsymbol{\Psi}_k\boldsymbol{d}_1 \tag{4.69}$$

where

$$\boldsymbol{d}_1 \triangleq \underline{\boldsymbol{\Delta}_1}\,. \tag{4.70}$$

Since for any OSTBC $\boldsymbol{G}_k^H \boldsymbol{G}_k = \boldsymbol{I}_N, \forall k = 1, \ldots, 2K$ [97], one can easily prove from (4.68) that for any OSTBC

$$\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k = \boldsymbol{I}_{2MN} \tag{4.71}$$

Using (4.69) and (4.71), the Euclidean norm of $\boldsymbol{e}_{k,1}$ can be expressed as

$$\|\boldsymbol{e}_{k,1}\| = \sqrt{\boldsymbol{d}_1^T \boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k \boldsymbol{d}_1} = \|\boldsymbol{d}_1\| = \|\boldsymbol{\Delta}_1\| . \tag{4.72}$$

Lemma 4.1 is proven.

4.B Proof of Lemma 4.2

Using the definition of the Frobenius norm along with Lemma 4.1 (i.e., using the constraints $\|\boldsymbol{e}_k\|^2 \leq \varepsilon$; $k = 1, \ldots, 2K$), we obtain

$$\begin{aligned} \|\boldsymbol{E}_k\| &= \sqrt{\mathrm{tr}\{\boldsymbol{E}_k^T \boldsymbol{E}_k\}} \\ &= \sqrt{\sum_{i=1; i \neq k}^{2K} \boldsymbol{e}_i^T \boldsymbol{e}_i} \\ &\leq \varepsilon \sqrt{2K - 1} \end{aligned} \tag{4.73}$$

where the inequality in the third row of (4.73) becomes equality if and only if $\|\boldsymbol{e}_k\|^2 = \varepsilon$; $k = 1, \ldots, 2K$. Hence, from (4.73) we conclude that

$$\eta = \max \|\boldsymbol{E}_k\| = \varepsilon \sqrt{2K - 1} \tag{4.74}$$

and Lemma 4.2 is proven.

4.C Proof of Theorem 4.1

Let us first introduce the following lemma.

Lemma 4.3. *Let vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ have a joint real Gaussian distribution with a covariance matrix $\boldsymbol{B}$, so that*

$$\mathrm{E}\{(\boldsymbol{v}_i - \mathrm{E}\{\boldsymbol{v}_i\})(\boldsymbol{v}_l - \mathrm{E}\{\boldsymbol{v}_l\})^T\} = r_{il} \boldsymbol{B}, \qquad \forall i, l, \quad i, l = 1, \ldots, n \tag{4.75}$$

where r_{il} are some constants. Then the set

$$\mathcal{K}(\zeta) = \left\{ \boldsymbol{x} \,\middle|\, \Pr\{\boldsymbol{v}_1^T\boldsymbol{x} \geq \kappa_1 \ \wedge \ \ldots \ \wedge \ \boldsymbol{v}_n^T\boldsymbol{x} \geq \kappa_n\} \geq \zeta \right\} \tag{4.76}$$

is convex for $\zeta \geq 0.5$. Here $\wedge$ denotes the set intersection operation, $0 < \zeta \leq 1$, and κ_i are arbitrary real constants.

PROOF: See [120, p. 312]. □

The constraints (4.49) share the same structure. Thus, it is enough to show that at least one of them is convex. Let us rewrite the lth constraint in (4.49) in the following equivalent form

$$\Pr\left\{ (\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_l(\boldsymbol{\Delta}_1))^T\boldsymbol{w}_k + \delta_l \geq 0 \ \wedge \ -(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_l(\boldsymbol{\Delta}_1))^T\boldsymbol{w}_k + \delta_l \geq 0 \right\} \geq \zeta. \tag{4.77}$$

The constraint (4.77) is called *joint chance constraint* in the stochastic programming literature [119], [120].

To be consistent with the notations used in Lemma 4.3, let us denote

$$\begin{aligned} \boldsymbol{v}_1 &\triangleq [(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_l(\boldsymbol{\Delta}_1))^T, 1]^T \\ \boldsymbol{v}_2 &\triangleq [-(\boldsymbol{a}_l(\hat{\boldsymbol{H}}_1) + \boldsymbol{e}_l(\boldsymbol{\Delta}_1))^T, 1]^T \\ \boldsymbol{x} &\triangleq [\boldsymbol{w}_k^T, \delta_l]^T \\ \kappa_1 &\triangleq 0 \qquad \kappa_2 \triangleq 0\,. \end{aligned} \tag{4.78}$$

Then, (4.77) can be equivalently written as

$$\Pr\{\boldsymbol{v}_1^T\boldsymbol{x} \geq \kappa_1 \ \wedge \ \boldsymbol{v}_2^T\boldsymbol{x} \geq \kappa_2\} \geq \zeta. \tag{4.79}$$

As the vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ have joint Gaussian distribution, such that

$$\mathrm{E}\{(\boldsymbol{v}_1 - \mathrm{E}\{\boldsymbol{v}_1\})(\boldsymbol{v}_1 - \mathrm{E}\{\boldsymbol{v}_1\})^T\} = \mathrm{E}\{(\boldsymbol{v}_2 - \mathrm{E}\{\boldsymbol{v}_2\})(\boldsymbol{v}_2 - \mathrm{E}\{\boldsymbol{v}_2\})^T\} = \boldsymbol{B} \tag{4.80}$$

where

$$\boldsymbol{B} = \begin{bmatrix} \mathrm{E}\{\boldsymbol{e}_l(\boldsymbol{\Delta}_1)\boldsymbol{e}_l^T(\boldsymbol{\Delta}_1)\} & \boldsymbol{0}_{2MT\times 1} \\ \boldsymbol{0}_{1\times 2MT} & 0 \end{bmatrix} = \begin{bmatrix} \frac{\sigma_h^2}{2}(\boldsymbol{I}_{2M} \otimes \boldsymbol{F}_l\boldsymbol{F}_l^T) & \boldsymbol{0}_{2MT\times 1} \\ \boldsymbol{0}_{1\times 2MT} & 0 \end{bmatrix} \tag{4.81}$$

and

$$\mathrm{E}\{(\boldsymbol{v}_1 - \mathrm{E}\{\boldsymbol{v}_1\})(\boldsymbol{v}_2 - \mathrm{E}\{\boldsymbol{v}_2\})^T\} = -\boldsymbol{B}. \tag{4.82}$$

Lemma 4.3 can be applied. Thus, the convexity of the constraints (4.49) follows for $\zeta \in [0.5, 1)$. This completes the proof.

Chapter 5

Linear Block Precoding for OFDM Systems

In this chapter, a new linear precoding technique based on the mean cutoff rate maximization criterion is proposed. In Section 5.1, we give a shot introduction to OFDM communications and briefly review the existing precoding techniques. Section 5.2 is devoted to the design of cutoff rate maximization-based linear block precoder. Section 5.3 addresses the issue of joint channel-coded and linearly-precoded OFDM communication systems. Simulation results are presented in Section 5.4. Section 5.5 briefly summarizes the chapter.

5.1 OFDM Systems with Linear Block Precoding

OFDM is a promising multiuser communication scheme which enables to mitigate MAI by means of providing each user with a non-intersecting fraction of subcarriers [18]. Due to the IFFT at the transmitter and the FFT at the receiver, the frequency selective fading channel is converted by OFDM into parallel flat fading channels [19]. This greatly facilitates the equalizer design at the receiver.

However, a well known disadvantage of the OFDM scheme is that, at each subcarrier, the channel may be subject to a deep fading. This makes a reliable detection of the information-bearing symbols at this particular subcarrier very difficult and, as a result, the overall performance of the system may degrade substantially. Thus, the transceiver optimization

is required.

A general transceiver optimization framework is discussed in [125]. In application to OFDM systems, a popular recent approach to improve the performance of OFDM systems in fading environments is to use linear block precoding at the transmitter [126]. For example, the minimum mean-square error (MMSE) and the minimum bit error rate (MBER) precoders for ZF equalization have been proposed in [126] and [127], respectively, and the MBER precoder for MMSE equalization has been studied in [128]. Another efficient precoding technique based on the channel capacity maximization has been proposed in [129].

Unfortunately, the application of precoders [126]-[129] may be limited by the fact that they require the full channel knowledge at the transmitter. To avoid this drawback, another linear precoder has been designed in [130] based on maximization of the diversity and coding gains. In contrast to the precoders of [126]-[129], the technique of [130] requires only the knowledge of the order of the multipath channel at the transmitter.

Another MBER based technique that does not require any channel information has been proposed in [131]. However, the class of MBER-optimal channel independent precoders developed in [131] is limited by the case when the MMSE equalization and QPSK modulation are used. Moreover, the performance of MBER precoder with MMSE equalization can be significantly improved by combining it with a water-filling procedure [128]. However, in the latter case the full channel knowledge at the transmitter is required.

In this chapter, a new linear precoder is proposed that maximizes the channel mean cutoff rate and requires the knowledge of the average relative channel multipath powers and delays at the transmitter. Our simulations show that the proposed precoder substantially outperforms the approach of [130] and several other linear precoding techniques in terms of BER.

5.2 Proposed Linear Block Precoder

For the sake of simplicity and following [126]-[129], let us consider the single-user block transmission system with N_c subcarriers. The extension to the multiuser case can be done

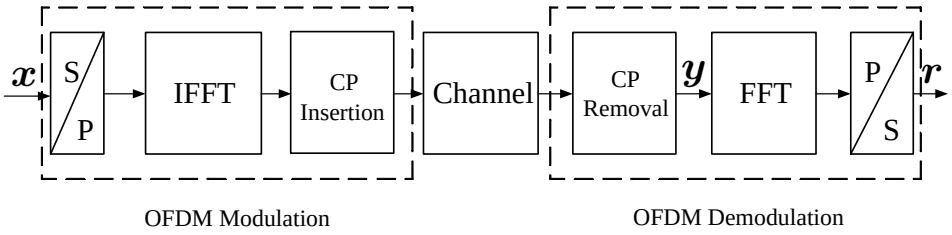

Figure 5.1: System block diagram of conventional OFDM communication systems.

straightforwardly by allocating a different group of subcarriers to each user [130]. The frequency selective wireless channel between the transmitter and the user is characterized by the path gains h_l $(l = 1, \cdots, L_P)$ and the delays τ_l $(l = 1, \cdots, L_P)$, where L_P is the total number of paths. We assume that the coefficients h_l $(l = 1, \cdots, L_P)$ are independent (but not necessarily identically distributed) zero-mean complex Gaussian random variables with the variances σ_l^2 $(l = 1, \ldots, L_P)$.

The block diagram of a conventional OFDM communication system is shown in Figure 5.1. It works in the following way. The tth block of transmitted symbols $\boldsymbol{x}(t) = [x(tN_c), \cdots, x(tN_c + N_c - 1)]^T$ is first serial-to-parallel converted. Then the converted signal is IFFT-modulated and the cyclic prefix (CP) is inserted to form one OFDM symbol. It is assumed that the length of the CP is longer than the maximum path delay. Finally the symbol is pulse-shaped and transmitted through the channel. The channel is assumed to be constant during the OFDM symbol transmission time. Hereafter, for notational simplicity the block index dependence of $\boldsymbol{x}$ is omitted.

After removing the CP, the received $N_c \times 1$ signal vector $\boldsymbol{y}$ at the receiver can be written as [19]

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{F}_t^H \boldsymbol{x} + \boldsymbol{n} \tag{5.1}$$

where $\boldsymbol{F}_t$ is the $N_c \times N_c$ normalized FFT matrix with its (i, l)th entry given by

$$[\boldsymbol{F}_t]_{i,l} = \frac{1}{\sqrt{N_c}} \exp\left(-\frac{j2\pi(i-1)(l-1)}{N_c}\right)$$

$\boldsymbol{H}$ is the $N_c \times N_c$ circulant channel matrix between the transmitter and the receiver with its (k, l)th entry given by $h_{(k-l+1) \bmod N_c}$, and $\boldsymbol{n}$ is the $N_c \times 1$ vector of the receiver AWGN

with the variance of $\sigma_v^2 \boldsymbol{I}_{N_c}$. After the FFT operation and parallel-to-serial conversion, the $N_c \times 1$ output symbol vector $\boldsymbol{r}$ can be written as

$$\boldsymbol{r} = \boldsymbol{F}_t \boldsymbol{y}. \tag{5.2}$$

Substituting (5.1) into (5.2) and using the fact that $\boldsymbol{F}_t \boldsymbol{H} \boldsymbol{F}_t^H = \boldsymbol{D}_c$ [19], where

$$\boldsymbol{D}_c = \text{diag}\{d_1, d_2, \cdots, d_{N_c}\}$$

is the diagonal matrix of the subcarrier channel gains, the received symbol block can be written as [19]

$$\boldsymbol{r} = \boldsymbol{D}_c \boldsymbol{x} + \boldsymbol{v} \tag{5.3}$$

where $\boldsymbol{v} = \boldsymbol{F}_t \boldsymbol{n}$ with $\text{E}\{\boldsymbol{v}\boldsymbol{v}^H\} = \sigma_v^2 \boldsymbol{I}_{N_c}$. The channel gain d_n $(n = 1, \cdots, N_c)$ of the nth subcarrier is given by [132][1]

$$d_n = \frac{1}{\sqrt{N_c}} \sum_{l=1}^{L_P} h_l \exp\left(-\frac{j2\pi n \tau_l}{N_c T_s}\right) \tag{5.4}$$

where T_s is the sampling interval.

If linear block precoding is used at the transmitter, then $\boldsymbol{x} = \sqrt{E_s}\boldsymbol{T}\boldsymbol{s}$, where E_s is the transmitted symbol power, $\boldsymbol{s} = [s_1, \ldots, s_W]^T$ is the $W \times 1$ vector of the information-bearing symbols, and $\boldsymbol{T}$ is the $N_c \times W$ precoding matrix [126]. Below, we assume that $W = N_c$ because in this case, the data rate is not sacrificed [19]. Then, equation (5.3) can be written as

$$\boldsymbol{r} = \sqrt{E_s}\boldsymbol{D}_c\boldsymbol{T}\boldsymbol{s} + \boldsymbol{v}. \tag{5.5}$$

The channel cutoff rate R_0 is a lower bound on the Shannon channel capacity, beyond which the sequential decoding becomes impractical [133], [134]. It also specifies an upper bound on the error rate of the optimal ML decoder and has been frequently used as a practical coding limit because it can be calculated in a simpler way than the channel capacity

[1]Note that the frequency domain channel response is often expressed through the FFT, but in this case the true multipath channel taps need to be converted to equivalent taps which have delays that are integer multiples of the sampling period. Therefore, following [132] we use the discrete Fourier transform (DFT) instead of FFT to calculate the channel impulse response in (5.4).

[134]. Therefore, the cutoff rate appears to be a proper criterion for the design of linear block precoders. Note that it has been previously used as a performance metric for OFDM systems [135], [136], and as a design criterion for transmitter optimization in MIMO channels [137].

We assume that a discrete constellation is used at the transmitter, the full channel knowledge is available at the receiver, and the ML technique is used to detect the symbols $\boldsymbol{s}$ from the received data $\boldsymbol{r}$. The conditional PDF of the received data can be written as

$$f(\boldsymbol{r}|\boldsymbol{s}^{(i)},\boldsymbol{T},\boldsymbol{D}_c) = \frac{1}{(\pi\sigma_v^2)^{N_c}} \exp\left(-\frac{\|\boldsymbol{r}-\sqrt{E_s}\boldsymbol{D}_c\boldsymbol{T}\boldsymbol{s}^{(i)}\|^2}{\sigma_v^2}\right) \tag{5.6}$$

where $\boldsymbol{s}^{(i)}$ is the ith member of the transmit vector constellation. To simplify the notation further, let us denote $f(\boldsymbol{r}|\boldsymbol{s}^{(i)},\boldsymbol{T},\boldsymbol{D}_c)$ as $f(i)$. The mean cutoff rate can be calculated as [133, p. 361]

$$\begin{aligned} R_0 &= -\log_2 \mathrm{E}_{\boldsymbol{D}_c}\left\{\int_{\boldsymbol{r}}\left[\frac{1}{M_c^{N_c}}\sum_{i=1}^{M_c^{N_c}}\sqrt{f(i)}\right]^2 d\boldsymbol{r}\right\} \\ &= -\log_2\left[\frac{1}{M_c^{N_c}} + \frac{1}{M_c^{2N_c}}\sum_{i=1}^{M_c^{N_c}}\sum_{\substack{l=1\\l\neq i}}^{M_c^{N_c}} \mathrm{E}_{\boldsymbol{D}_c}\left\{\int_{\boldsymbol{r}}\sqrt{f(i)f(l)}\,d\boldsymbol{r}\right\}\right] \end{aligned} \tag{5.7}$$

where M_c is the constellation size. Substituting (5.6) into (5.7) we obtain the following expression for the mean cutoff rate

$$R_0 = -\log_2\left[\frac{1}{M_c^{N_c}} + \frac{1}{M_c^{2N_c}}\sum_{i=1}^{M_c^{N_c}}\sum_{\substack{l=1\\l\neq i}}^{M_c^{N_c}} \mathrm{E}_{\boldsymbol{D}_c}\left\{\exp\left(-\frac{E_s\|\boldsymbol{D}_c\boldsymbol{T}(\boldsymbol{s}^{(i)}-\boldsymbol{s}^{(l)})\|^2}{4\sigma_v^2}\right)\right\}\right]. \tag{5.8}$$

Using the results of [139], the expectation of exponential quadratic form in (5.8) can be written as

$$\mathrm{E}_{\boldsymbol{D}_c}\left\{\exp\left(-\frac{E_s\|\boldsymbol{D}_c\boldsymbol{T}(\boldsymbol{s}^{(i)}-\boldsymbol{s}^{(l)})\|^2}{4\sigma_v^2}\right)\right\} = \prod_{k=1}^{\mathrm{rank}\{\boldsymbol{E}_{i,l}\}}\left(1+\frac{E_s}{4\sigma_v^2}\lambda_k\right)^{-1} \tag{5.9}$$

where

$$\boldsymbol{E}_{i,l} \triangleq \mathrm{E}_{\boldsymbol{D}_c}\{\boldsymbol{D}_c\boldsymbol{T}\boldsymbol{e}_{i,l}\boldsymbol{e}_{i,l}^H\boldsymbol{T}^H\boldsymbol{D}_c^H\}, \qquad \boldsymbol{e}_{i,l} \triangleq \boldsymbol{s}^{(i)}-\boldsymbol{s}^{(l)} \tag{5.10}$$

λ_k is the kth non-zero eigenvalue of the matrix $\boldsymbol{E}_{i,l}$. Substituting (5.9) in (5.8), after straightforward manipulations we obtain

$$R_0 = -\log_2 \left[\frac{1}{M_c^{N_c}} + \frac{1}{M_c^{2N_c}} \sum_{i=1}^{M_c^{N_c}} \sum_{\substack{l=1 \\ l \neq i}}^{M_c^{N_c}} \prod_{k=1}^{\text{rank}\{\boldsymbol{E}_{i,l}\}} \left(1 + \frac{E_s}{4\sigma_v^2} \lambda_k \right)^{-1} \right]. \tag{5.11}$$

It is worth noting that the expression (5.11) for the mean cutoff rate is directly related to the expression for the Chernoff bound on the pairwise error probability (PEP). In particular, the second term under the logarithm in (5.11) can be seen as an average of the Chernoff bounds on PEP for all distinct pairs of symbols. In other words, the maximization of mean cutoff rate is equivalent to the minimization of averaged PEP. This observation provides further motivation to choose the mean cutoff rate as a criterion for precoder design.

It is well known that the precoder based on PEP minimization also provides maximum diversity gain [138]. Therefore, the maximization of the mean cutoff rate will achieve the maximum diversity gain under the following condition [138]

$$\left| [\boldsymbol{T}]_n (\boldsymbol{s}^{(i)} - \boldsymbol{s}^{(l)}) \right| \neq 0, \qquad n = 1, \ldots, N_c, \quad i, l = 1, \ldots, M_c^{N_c}$$

where $[\boldsymbol{T}]_n$ is the nth row of $\boldsymbol{T}$.

To compute the matrix $\boldsymbol{E}_{i,l}$ explicitly, let us introduce the vector

$$\boldsymbol{d} \triangleq [[\boldsymbol{D}_c]_{1,1}, \cdots, [\boldsymbol{D}_c]_{N_c,N_c}]^T.$$

Then, $\boldsymbol{E}_{i,l}$ can be rewritten as

$$\boldsymbol{E}_{i,l} = \boldsymbol{R}_{\boldsymbol{d}} \circ (\boldsymbol{T} \boldsymbol{e}_{i,l} \boldsymbol{e}_{i,l}^H \boldsymbol{T}^H) \tag{5.12}$$

where $\circ$ stands for the Schur-Hadamard matrix product and $\boldsymbol{R}_{\boldsymbol{d}} \triangleq \mathrm{E}_{\boldsymbol{d}}\{\boldsymbol{d}\boldsymbol{d}^H\}$. The (n,k)th entry of $\boldsymbol{R}_{\boldsymbol{d}}$ is given by

$$[\boldsymbol{R}_{\boldsymbol{d}}]_{n,k} = \frac{1}{N_c} \sum_{l=1}^{L_P} \sigma_l^2 \exp\left(-\frac{j2\pi(n-k)\tau_l}{N_c T_s} \right). \tag{5.13}$$

Our task now is to design the precoding matrix $\boldsymbol{T}$ which maximizes R_0 in (5.11) subject to the power constraint $\|\boldsymbol{T}\| = \sqrt{N_c}$. This problem does not have any analytical solution,

but it can be solved by using either algebraic number-theoretic techniques or by computer search over compact parameterizations of unitary matrices [130], [138]. Here, we obtain $\boldsymbol{T}$ through computer search over the unitary[2] parameterization expressed via Givens rotation matrices. For the details of this technique, see, for example, [138]. Provided that each user occupies a moderate number of subcarriers (not more than 3 subcarriers per user), and since the precoding matrices can be designed for each user independently, we can conclude that the total number of real optimization parameters for the particular user is $N_c(N_c-1) \leq 6$. If the number of optimization parameters is small, full search is computationally feasible, and, thus, the design of our precoder becomes practically feasible as well.

It can be seen from (5.11)-(5.13) that only the knowledge of the average relative channel powers and delays is required at the transmitter for the design of our linear precoder. Although the channel state variations can be very fast due to small-scale fading, the power and multipath delay variations are typically much slower [18]. Therefore, a low-rate feedback can be used to convey this information to the transmitter.

5.3 Joint Channel-Coded and Linearly-Precoded OFDM Systems

To mitigate fading and noise effects, practical wireless systems often employ some form of outer channel coding (CC). In fact, channel coding and linear precoding techniques can be combined to combat channel fading [143].

The block diagram of a joint channel-coded and linearly-precoded OFDM communication system is shown in Figure 5.2. It operates in the following way. A sequence of information bits is first encoded by error-correction CC. The coded bits then pass through a random interleaver Π_1. The output of Π_1 are mapped to constellation symbols. After constellation mapping, successive blocks of N_c symbols are linearly precoded by the matrix $\boldsymbol{T}$. The precoded symbols are interleaved by a second interleaver Π_2, which is needed

[2]Note that unitary precoders have the advantage that they do not alter the Euclidian distance between the entries of any block of information-bearing symbols [130].

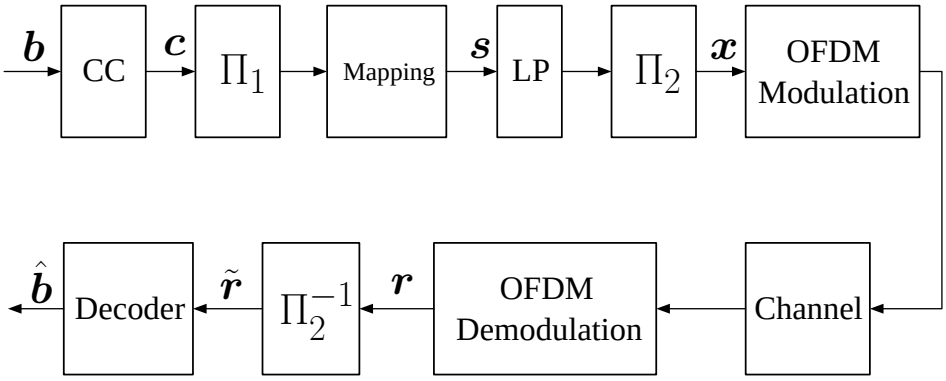

Figure 5.2: System block diagram of joint channel-coded and linearly-precoded OFDM communication systems.

to decorrelate the subcarrier channel gains. The output of Π_2 passes through the OFDM modulator. The resulting output is transmitted through the channel.

From Figure 5.2 we can observe that the channel encoder, the interleaver Π_1, and the linear precoder together represent a serial concatenated encoder [140]. In such an encoder, the linear precoder and the channel encoder can be seen as the inner encoder and the outer encoder, respectively. Therefore, the standard decoding schemes for serial concatenated codes can be used to decode the joint channel-coded and linearly-precoded symbols.

5.3.1 Hard-Decision Decoding Scheme

The block diagram of the first scheme, which is referred to as the hard-decision decoding scheme (HDDS), is shown in Figure 5.3. HDDS consists of the ML symbol detector and Viterbi decoder. In this scheme, the ML symbol detector is used to detect the symbols encoded by the linear precoder, while the Viterbi decoder is applied to decode the convolutional CCs. Note that such decoding scheme is suboptimal because hard decision is used at the output of the ML symbol detector. However, the computational complexity of this decoder is relatively low, provided that each user occupies only a moderate number of subcarriers.

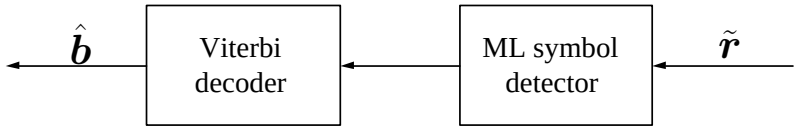

Figure 5.3: Block diagram of concatenated ML symbol detector and Viterbi decoder.

5.3.2 Soft-Decision Decoding Scheme

The second decoding scheme we refer to as the soft-decision decoding scheme (SDDS). The block diagram of SDDS is shown in Figure 5.4. It is based on the iterative (turbo) decoding technique. The main components of SDDS are two maximum a posteriori (MAP) soft-input soft-output (SISO) modules. Each SISO module is a four-port device which receives bits of soft information and outputs the updated soft information calculated by the MAP algorithm [140]. The updated soft information is exchanged between two SISO modules in an iterative way. Normally the soft information is represented through the LL ratio. Compared with the hard information, the soft information not only contains the result of a decision, but also reflects the reliability of this decision [140].

The merit of the iterative (turbo) decoding scheme is that during each iteration, the so-called extrinsic information (which is the soft information passed from one SISO module to another) increases the reliability of the decision. Therefore, after a finite number of iterations, the reliability of the decision will be high enough, and the iteration process will be terminated. Then, the final decision can be made by passing the value of the likelihood ratio of each bit through a threshold detector. Detailed discussion about the turbo principle is beyond the scope of this chapter, and we refer readers to [140] and [143] for more information.

Note that SDDS can greatly improve the performance of the joint channel-coded and linearly-precoded OFDM system. However, compared with HDDS, SDDS has higher computational complexity, because the computations involved in the MAP algorithm are at least 4 times of that involved in the Viterbi algorithm [143]. Moreover, SDDS requires more memory as compared with HDDS. Therefore, HDDS and SDDS offer different tradeoffs in terms of performance, system requirement and hardware/software complexity.

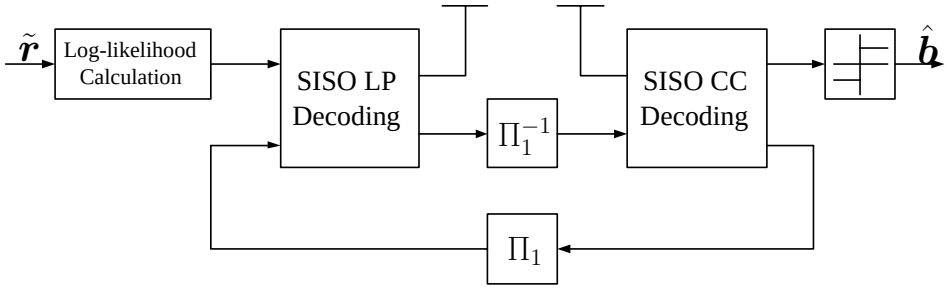

Figure 5.4: Block diagram of iterative (turbo) decoder.

5.4 Numerical Examples

In this section, we investigate the performance of the proposed precoder in multipath indoor and outdoor channels. As an example of a Rayleigh fading outdoor channel, we choose the ETSI "Vehicular A" channel environment which has been defined for the evaluation of UMTS radio interface proposals [141]. The multipath time delays and the variances of the multipath gains of the "Vehicular A" channel are shown in Table 5.1. Correspondingly,

Table 5.1: Characteristics of the ETSI "Vehicular A" channel environment.

Tap	Time Delays (T_s)	Average Power (dB)
1	0	0
2	1.55	-1
3	3.55	-9
4	5.45	-10
5	8.65	-15
6	12.55	-20

as an example of a multipath Rayleigh fading indoor channel we choose the HIPERLAN/2 "Model A" channel, which represents a typical office environment [142]. The multipath time delays and the variances of the multipath gains of the HIPERLAN/2 "Model A" channel are shown in Table 5.2. The Doppler frequencies for these two channels are set to be equal to 100 Hz and 50 Hz, respectively.

Throughout the simulations, a multiuser block transmission system with 64 subcarriers is assumed. All subcarriers are allocated among the users and interleaved [130] such that

Table 5.2: Characteristics of the HIPERLAN/2 "Model A" channel environment.

Tap	Time Delays (T_s)	Average Power (dB)
1	0	0
2	0.2	-0.9
3	0.4	-1.7
4	0.6	-2.6
5	0.8	-3.5
6	1	-4.3
7	1.2	-5.2
8	1.4	-6.1
9	1.6	-6.9
10	1.8	-7.8
11	2.2	-4.7
12	2.8	-7.3
13	3.4	-9.9
14	4	-12.5
15	4.8	-13.7
16	5.8	-18.0
17	6.8	-22.4
18	7.8	-26.7

the subcarriers assigned to the same user are as less correlated to each other as possible. Each user is provided with $N_c = 3$ subcarriers and the BPSK modulation is used.

In our first example, different precoding techniques are compared with each other when no CC is used. This enables us to study the net effect of precoding on the performance of OFDM systems. Six different techniques are compared: the approach where no precoding is used, the MMSE-ZF precoder of [126], the MBER-ZF precoder of [127], the MBER-MMSE precoder of [128], the Vandermonde precoder of [130], and the proposed precoder. The precoders of [126]-[128] are assumed to utilize the full channel knowledge at the transmitter, while our precoder uses only the average relative channel powers and delays. It is also important to stress that without any precoding, $\boldsymbol{T} = \boldsymbol{I}_{N_c}$, and the detection of each information-bearing symbol is decoupled from the detection of any other symbols. Thus, provided that a constant-modulus constellation is used, the ML and MMSE symbol detectors are equivalent in this case.

Figure 5.5 displays the channel mean cutoff rate of the aforementioned precoding schemes versus the SNR for the ETSI "Vehicular A" channel environment. It can be seen from this figure that, as expected, the proposed linear precoder has the highest mean cutoff rate among all the techniques tested.

Figure 5.6 compares the BER performances of the same techniques with different symbol detectors for the ETSI "Vehicular A" channel environment. In this figure, the performances of the proposed, MBER-MMSE, and Vandermonde precoders are displayed both in the cases when the ML and MMSE symbol detectors are used. Additionally, the BER performances of the MMSE-ZF and MBER-ZF precoders are displayed along with the BER of the standard approach where no precoding is used. All results are averaged over 1000 simulation runs.

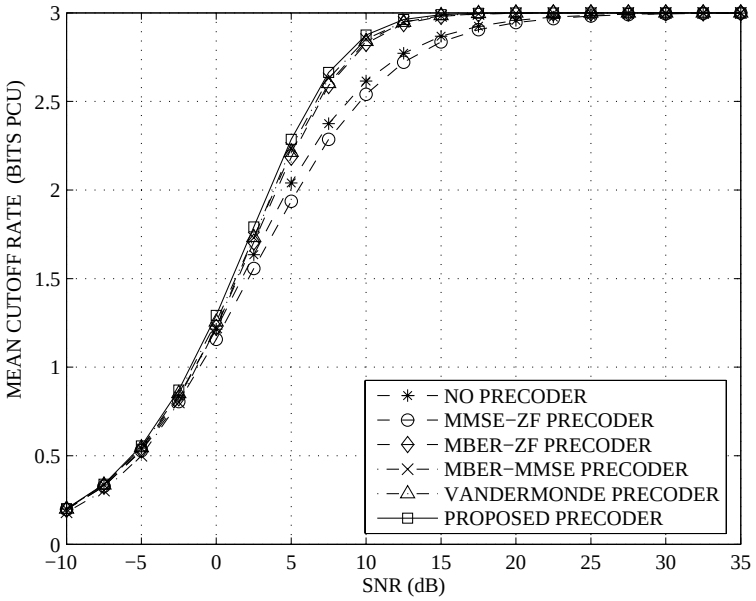

Figure 5.5: Cutoff rate versus SNR. First example with the ETSI "Vehicular A" channel environment.

Figure 5.7 shows the channel mean cutoff rate of all aforementioned precoders versus SNR for the HIPERLAN/2 "Model A" channel environment. As in the previous case, in this indoor channel scenario the proposed linear precoder has the highest R_0.

Finally, the BER performances of all tested precoders are compared in Figure 5.8 in the

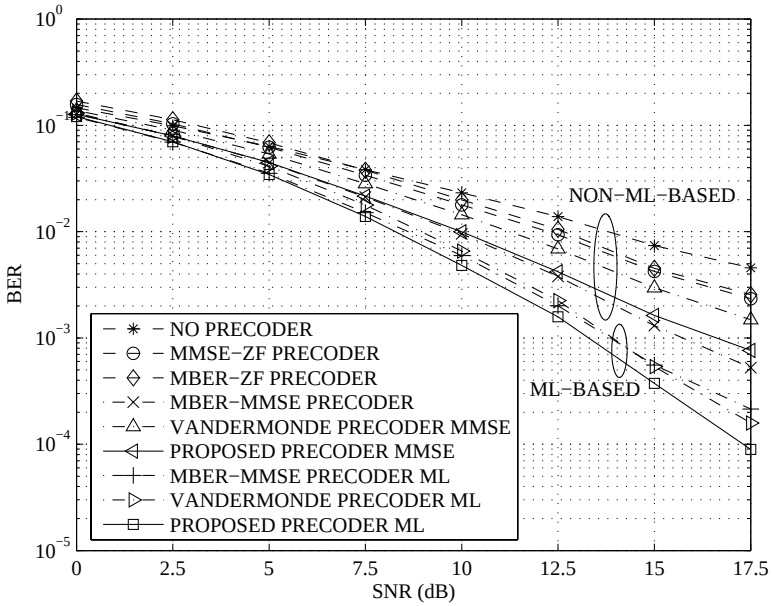

Figure 5.6: BER versus SNR. First example with the ETSI "Vehicular A" channel environment.

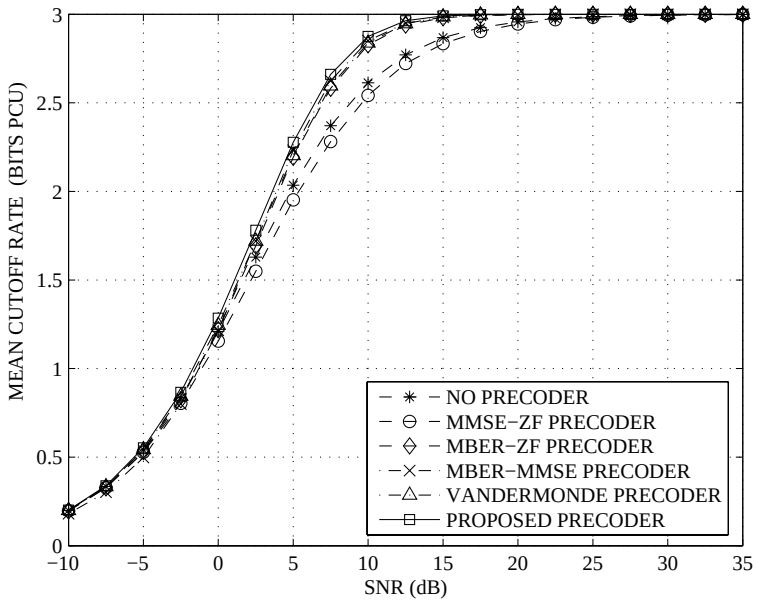

Figure 5.7: Cutoff rate versus SNR. First example with the HIPERLAN/2 "Model A" channel environment.

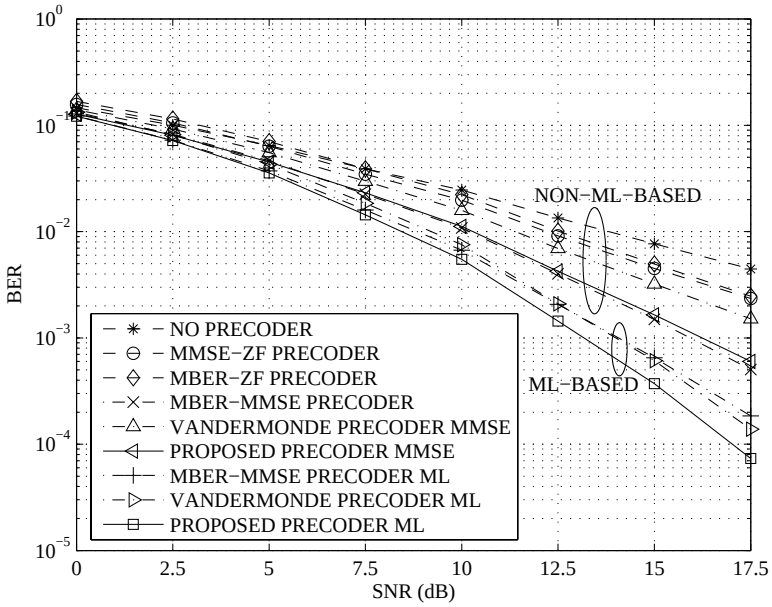

Figure 5.8: BER versus SNR. First example with the HIPERLAN/2 "Model A" channel environment.

HIPERLAN/2 "Model A" channel case.

We can observe that our linear precoder substantially outperforms all other techniques tested in terms of BER for both the ETSI "Vehicular A" and HIPERLAN/2 "Model A" channel environments. Interestingly, this conclusion is true when the ML-based as well as non-ML (MMSE) receivers are used, with the only exception for the MBER-MMSE precoder. In particular, the performance of the latter precoder is comparable to the performance of our precoder used with the MMSE receiver. This fact demonstrates that although the mean cutoff rate based precoder has been proposed for the ML-based symbol detector, it also provides a good performance when applied with the simpler MMSE symbol detector.

In our second example, we compare the performance of OFDM systems with combined coding-precoding, with CC only, with precoding only, and without CC and precoding. The proposed precoder and the Vandermonde precoder are used in this simulation example. The Vandermonde precoder is chosen for comparison to the proposed precoder as one that shows the best performance among known precoders tested in the previous example.

First, we study the system performance when HDDS is used at the receiver. For this simulation, the rate 1/2 convolutional code in the HIPERLAN/2 standard [142], [143] with generator $(133, 171)$ is used. The size of the interleaver Π_1 (see Figure 5.2) is 256 OFDM symbols. Figures 5.9 and 5.10 show the BER performance of different OFDM configurations for the ETSI "Vehicular A" and the HIPERLAN/2 "Model A" channel environments, respectively. From these two figures, it can be seen that OFDM system with combined coding-precoding shows the best performance among the techniques tested. Moreover, the scheme where the cutoff rate maximization based precoder is used outperforms the scheme based on the Vandermonde precoder of [130].

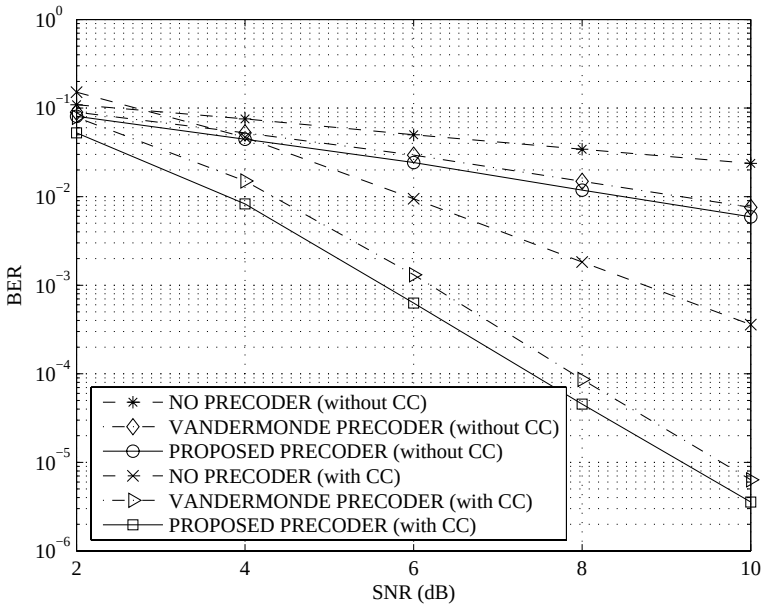

Figure 5.9: BER versus SNR. Second example with the ETSI "Vehicular A" channel environment; HDDS.

Second, we investigate the system performance when SDDS is applied at the receiver. For this simulation the rate 1/2 systematic convolutional code with generator $(5, 7)$ is used. Figure 5.11 shows the system performance for the ETSI "Vehicular A" channel environment. Two iterations are carried out before the final decision. From Figure 5.11, we observe that SDDS substantially enhances the overall system performance. However, this performance

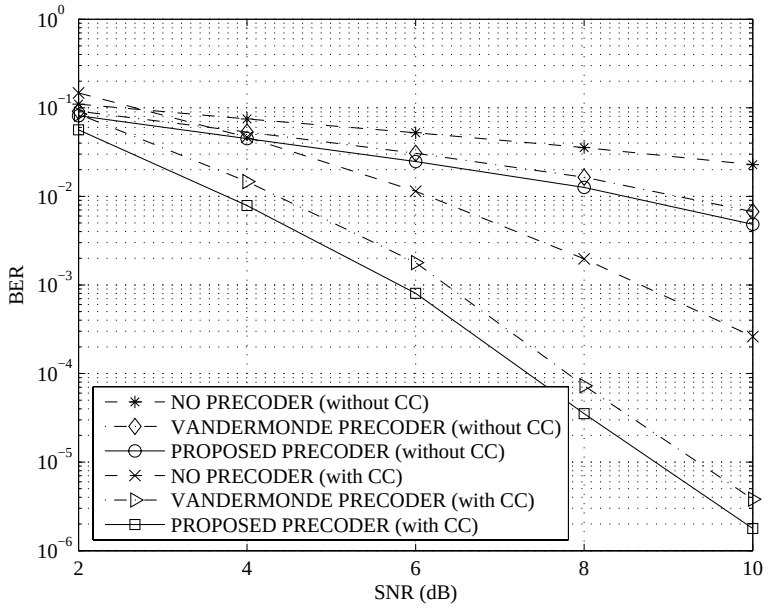

Figure 5.10: BER versus SNR. Second example with the HIPERLAN/2 "Model A" channel environment; HDDS.

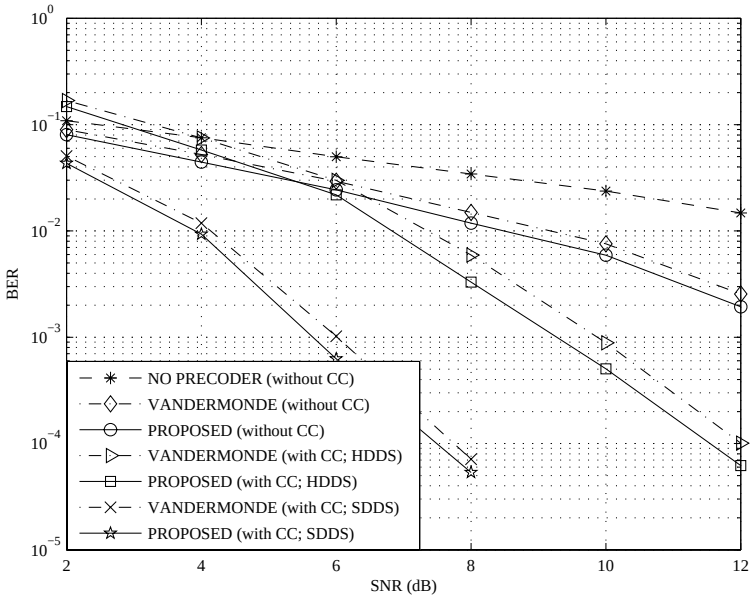

Figure 5.11: BER versus SNR. Second example with the ETSI "Vehicular A" channel environment; SDDS.

improvement is at the price of higher decoding complexity and larger memory consumption. We can also see from Figure 5.11 that the scheme where the cutoff rate maximization based precoder is used outperforms the scheme based on the Vandermonde precoder of [130].

5.5 Chapter Summary

In this chapter, a new linear precoder for block OFDM transmissions has been proposed. Our precoder is based on the maximization of the channel mean cutoff rate and requires only the knowledge of the average relative channel multipath powers and delays at the transmitter. Simulation results show substantial performance improvements achieved by the proposed precoding technique relative to the existing linear block precoders which typically require the full channel knowledge at the transmitter. The proposed linear precoding technique can be combined with channel coding technique to further enhance the performance of OFDM communication systems. It can be readily generalized to the MIMO case.

Chapter 6

Adaptive OFDM with Channel State Feedback

In this chapter, we study the performance of OFDM-based communication systems whose transmitter has only one bit of CSI per subcarrier obtained through a low-rate feedback. Section 6.1 gives a short introduction to adaptive OFDM communication systems with one-bit-per-subcarrier channel state feedback and formulates the system model. In Section 6.2, we consider the perfect feedback channel case and present our analysis of the ASCS, APA, and AMS schemes applied to OFDM systems with one-bit-per-subcarrier feedback. Section 6.3 is devoted to the analysis and optimization of the APA and AMS schemes in the case when the feedback channel suffers from errors or delays. Section 6.4 presents simulation results where the performances of the ASCS, APA, and AMS schemes are compared under the conditions of perfect and imperfect feedback channels. Section 6.5 briefly summarizes the chapter.

6.1 System Models and Background

One of the disadvantages of OFDM communication systems is that some subcarriers may suffer from deep fading. As we have already seen from Chapter 5, error-correction channel coding and linear precoding techniques can be applied to mitigate fading. Furthermore, if some CSI is available at the transmitter, adaptive modulation and resource allocation

techniques can be applied to allocate bits and transmitted powers to the subcarriers [20], [144]-[145]. However, in wireless communications it can be difficult to obtain such CSI. For example, in the downlink mode of cellular communications, if the time-division duplex (TDD) mode is used, the downlink transmit CSI can be obtained by estimating the uplink channel and using the uplink-downlink reciprocity. However, in practical situations fast channel variability and user mobility may prohibit to use the aforementioned reciprocity property. Moreover, this property does not hold if the frequency-division duplex (FDD) mode is used. In the latter case, some feedback has to be exploited to transmit the downlink CSI from the receiver to the transmitter. As the bandwidth consumed by the feedback channel is proportional to the feedback rate, it is interesting to study the performance of systems which enable only a low-rate CSI feedback. For example, the use of one-bit channel state feedback in Alamouti-type systems has been studied in [146], while the asymptotic lower bound on the minimum feedback rates for multicarrier transmission has been derived in [147].

In this chapter, the performance of OFDM communication systems with one-bit-per-subcarrier CSI feedback is studied. The uncoded transmission case is considered and the raw BER is used as the criterion to evaluate the system performance. Assuming that the feedback channel is perfect, three adaptive approaches including ASCS, APA, and AMS are used to exploit the CSI feedback and compared via simulations. For the latter two techniques, a closed-form expression for the BER is also derived and, based on it, the parameters of these techniques are optimized.

In practical situations, the feedback channel may be erroneous and may suffer from a feedback delay. Therefore, the feedback CSI may be unreliable. Motivated by these facts, the impact of an imperfect CSI feedback on the performance of the APA and AMS techniques is also studied, and it is investigated how to exploit the knowledge of the fact that the feedback is imperfect to optimize the parameters of these techniques.

Let us consider the point-to-point OFDM communication system (5.3), as that described in Section 5.2, and introduce an $N_c \times N_c$ diagonal matrix $\boldsymbol{P}_c$ of the transmitted powers of

the symbols corresponding to different subcarriers. We have

$$\boldsymbol{x} = \boldsymbol{P}_c^{1/2}\boldsymbol{s}\,. \tag{6.1}$$

Substituting (6.1) into (5.3), we obtain

$$\boldsymbol{r} = \boldsymbol{D}_c\boldsymbol{P}_c^{1/2}\boldsymbol{s} + \boldsymbol{v}.$$

From (5.4), it is obvious that d_n is a zero-mean complex Gaussian random variable with the variance of $\frac{1}{N_c}\sum_{l=1}^{L_P}\sigma_l^2$. Without any loss of generality, we normalize the variance of the channel gain at each subcarrier so that $\frac{1}{N_c}\sum_{l=1}^{L_P}\sigma_l^2 = 1$. It can be seen that $d_1, \ldots, d_{N_c}$ all are identically distributed. The absolute value of each $d_n, n = 1, \cdots, N_c$ is Rayleigh-distributed with the PDF

$$p(\alpha) = 2\alpha\exp(-\alpha^2). \tag{6.2}$$

We assume that the transmitter transmits at the constant data rate of n_r bits per second (bps) and that the transmitter has perfect knowledge of the SNR, while the receiver has perfect downlink CSI knowledge. The CSI is transmitted back to the transmitter through a low-rate feedback channel. More specifically, we consider the case when a total number of N_c bits containing the CSI for all subcarriers (i.e., one bit per subcarrier) is transmitted to the transmitter in one feedback cycle.

6.2 Perfect One-Bit-Per-Subcarrier CSI Feedback

In this section, we assume that the feedback channel is perfect (i.e., there are no feedback errors and/or delays) and study several efficient ways to make use of N_c feedback bits (one bit per subcarrier) available. Clearly, it is impossible to provide a sufficiently accurate CSI feedback to the transmitter with only N_c bits. To illustrate this fact, we note that in wireless communications, the order of the multipath channel can be about $L_P = 10$ [141], and the typical choice of the number of information-bearing subcarriers for wireless local area networks (WLANs) is $N_c = 52$ [18]. Assuming that 16 bits are used to represent a real-valued number, 320 bits are required to feedback the full CSI and, therefore, more

than six bits of feedback per subcarrier (or, equivalently, more than $6N_c$ bits in total) are required in this case. Thus, the question how to make use of only one feedback bit per subcarrier in an efficient way is of a great practical interest.

6.2.1 Adaptive Subcarrier Selection

The idea of the considered subcarrier selection strategy is that subcarriers which are affected by a deep fading should be excluded and only subcarriers with high channel gains should be used.[1]

The feedback in the system with ASCS can be organized in the following way. The receiver sorts the channel gains in all N_c subcarriers and picks R subcarriers with the highest channel gains. If some particular subcarrier has been selected, "1" is conveyed back to the transmitter to indicate that this particular subcarrier should be used; otherwise "0" is transmitted to indicate that this subcarrier should be dropped. The transmitter equally distributes the available power among the selected subcarriers. In order to keep constant data rate for different numbers of selected subcarriers, different types of signal modulation may be used.

To determine the optimal number of subcarriers, a theoretical analysis of the error probability is required. However, such an analysis appears to be a very difficult task because it involves order statistics of correlated random variables (channel gains of different subcarriers). Thus, we limit our study of the ASCS strategy by simulations presented in Section 6.4.

6.2.2 Adaptive Power Allocation

As an alternative to the ASCS strategy, the one-bit-per-subcarrier CSI feedback can be used to adaptively allocate transmitted powers according to the channel gain at each subcarrier under the constraint that the average transmitted power per subcarrier is fixed. In the practical (sufficiently high) SNR range, it is known that more power should be allocated to

[1] Such a strategy is, however, somewhat different from the techniques used in [148] and [149] where other approaches to subcarrier selection are addressed.

faded subcarriers than to non-faded ones to minimize the BER [20]. However, as we will see below, at low SNRs the situation may be reversed, that is, the BER is minimized when more power is allocated to non-faded subcarriers with high channel gains.

With one-bit-per-subcarrier CSI feedback, APA can be implemented in the following way. If the channel gain of some subcarrier is below a certain threshold u, the feedback bit "0" is conveyed to the transmitter and, in this case, the transmitter allocates the transmitted power ψ_1 to this particular subcarrier. Otherwise, the feedback bit "1" is delivered to the transmitter and it allocates the transmitted power ψ_2 to this subcarrier. We refer to this technique as conventional APA.

In what follows, we present a theoretical study of the average BER of the conventional APA strategy and further optimize this power allocation scheme.

The exact SER in the case of M-PSK modulation can be calculated as [150]

$$P_s(\text{MPSK}) = \frac{1}{\pi}\int_0^{\frac{M_c-1}{M_c}\pi}\int_0^{\infty}\exp\left(-\frac{g_{\text{PSK}}\alpha^2 E_s}{\sin^2\phi\ \sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi \tag{6.3}$$

where E_s is the transmitted signal power, M_c denotes the constellation size, and $g_{\text{PSK}} = \sin^2(\pi/M_c)$. Hereafter, we assume that the Gray mapping is always used to map bits to symbols. In this case, the BER can be approximated as [5]

$$P_b \approx \frac{1}{\log_2 M_c}P_s. \tag{6.4}$$

Substituting $M_c = 4$ into (6.3) and using (6.4), we obtain that in the QPSK modulation case,

$$P_b(\text{QPSK}) = \frac{1}{2\pi}\int_0^{\frac{3}{4}\pi}\int_0^{\infty}\exp\left(-\frac{\alpha^2 E_s}{2\sin^2\phi\ \sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi.$$

In this case, the BER of the APA technique with one-bit-per-subcarrier feedback can be calculated as

$$\begin{aligned}P_b^{\text{APA}}(\text{QPSK}, u, \psi_1, \psi_2) \;=\; &\frac{1}{2\pi}\left[\int_0^{\frac{3}{4}\pi}\int_0^{u}\exp\left(-\frac{\alpha^2\psi_1 E_s}{2\sin^2\phi\ \sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi\right.\\ &\left.+\;\int_0^{\frac{3}{4}\pi}\int_u^{\infty}\exp\left(-\frac{\alpha^2\psi_2 E_s}{2\sin^2\phi\ \sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi\right]\end{aligned} \tag{6.5}$$

Table 6.1: Optimal parameters of conventional APA.

SNR (dB)	0	5	10	15	20	25
u	0.4724	1.1774	0.7147	0.4724	0.3246	0.2265
ψ_1	0.2	1.1	1.6	2.6	4.6	10.5
ψ_2	1.2	0.7	0.6	0.6	0.6	0.5

where ψ_1 denotes the normalized transmitted power when the value of the channel gain lies in the interval $[0, u)$ and ψ_2 denotes the normalized transmitted power when the value of the channel gain lies in the interval $[u, \infty)$.

Let us now obtain the optimal threshold u and optimal power allocations ψ_1 and ψ_2 which minimize (6.5) subject to both the average and peak transmit power constraints. Such optimal values of u, ψ_1, and ψ_2 can be found as a solution to the following constrained optimization problem

$$\begin{aligned} \min_{u,\psi_1,\psi_2} \quad & P_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2) \\ \text{subject to} \quad & \int_0^u \psi_1 p(\alpha)\, d\alpha + \int_u^\infty \psi_2 p(\alpha)\, d\alpha = 1 \\ & 0 < \psi_1 < \psi_M, \quad 0 < \psi_2 < \psi_M, \quad u > 0 \end{aligned} \tag{6.6}$$

where ψ_M denotes the normalized maximum transmitted power which is determined by the transmission hardware peak power. The first constraint in (6.6) limits the normalized average transmitted power, while the second and third constraints in (6.6) limit the normalized peak transmitted powers. Substituting (6.2) into (6.6), we see that the objective function is a highly nonlinear function of u, ψ_1, and ψ_2. To solve the problem (6.6), the method proposed in [151] can be used. The idea of this method is to quantize the parameters u, ψ_1, and ψ_2 and obtain a suboptimal solution using standard dynamic programming techniques.

Table 6.1 shows the optimal parameters for the conventional APA scheme which uses one bit feedback for each subcarrier. The optimal values of parameters u, ψ_1, and ψ_2 are obtained here by solving the optimization problem (6.6).

Let us now consider the effect of correlation of the channel gains between subcarriers. From (5.13) it follows that the channel gains in adjacent subcarriers are highly correlated.

This fact can be exploited in the following way. The CSI feedback can be provided for every other subcarrier (i.e., for the subcarriers with the indices $2, 4, 6, \ldots$) rather than for each subcarrier. Then the CSI feedback is only required for $N_c/2$ subcarriers. In this case, we can use 2 bits of feedback per subcarrier and still have N_c bits of feedback in total. If such an approach is adopted, then four normalized transmitted power levels ψ_i $(i = 1, 2, 3, 4)$ and, correspondingly, three thresholds u_l $(l = 1, 2, 3)$ can be used in the APA scheme. Hereafter, we refer to this technique as modified APA. The BER for this scheme can be computed as

$$P_b^{\mathrm{APA}}(\mathrm{QPSK}, \boldsymbol{u}, \boldsymbol{\psi}) = \frac{1}{2\pi} \sum_{i=1}^{4} \int_0^{\frac{3}{4}\pi} \int_{\Omega_i} \exp\left(-\frac{\alpha^2 \psi_i E_s}{2 \sin^2\phi \, \sigma_v^2}\right) p(\alpha) \, d\alpha \, d\phi$$

where $\boldsymbol{u} = [u_1, u_2, u_3]^T$ is the vector of the thresholds, $\boldsymbol{\psi} = [\psi_1, \psi_2, \psi_3, \psi_4]^T$ is the vector of the normalized transmitted powers, and $\Omega_i = [u_{i-1}, u_i)$ $(i = 1, \cdots, 4)$ are the channel gain intervals with $u_0 = 0$ and $u_4 = \infty$. Then, the optimal values of the vector parameters $\boldsymbol{u}$ and $\boldsymbol{\psi}$ can be found by solving the following constrained optimization problem:

$$\begin{aligned} \min_{\boldsymbol{u}, \boldsymbol{\psi}} \quad & P_b^{\mathrm{APA}}(\mathrm{QPSK}, \boldsymbol{u}, \boldsymbol{\psi}) \\ \text{subject to} \quad & \sum_{i=1}^{4} \int_{\Omega_i} \psi_i p(\alpha) \, d\alpha = 1 \\ & 0 < \psi_i < \psi_M, \quad i = 1, 2, 3, 4 \\ & 0 < u_l < \infty, \quad l = 1, 2, 3. \end{aligned} \tag{6.7}$$

Table 6.2 shows the optimal parameters $\boldsymbol{u}$ and $\boldsymbol{\psi}$ for the modified APA scheme when the correlation between adjacent subcarriers is exploited and two bits of feedback for every other subcarrier is provided. The optimal values of parameters are obtained by solving problem (6.7).

Another important question is whether it is beneficial to reduce the total number of subcarriers but to increase the constellation dimension. For example, if the number of subcarriers is reduced twice (to $N_c/2$), then the same amount of information at the same rate can be transmitted by using the constellation whose dimension is four times higher than in the case of N_c subcarriers. For example, if the QPSK modulation has been used in

Table 6.2: Optimal parameters of modified APA.

SNR (dB)	0	5	10	15	20	25
u_1	0.3654	0.2792	0.5049	0.2265	0.1591	0.1591
u_2	0.6269	0.8936	0.7732	0.4031	0.3246	0.3246
u_3	1.5174	1.3774	1.1362	0.7147	0.5972	0.7147
ψ_1	0.1	0.3	2.0	5.8	16.3	28.3
ψ_2	0.7	1.3	1.1	2.2	2.7	1.5
ψ_3	1.3	0.9	0.7	1.0	0.9	0.4
ψ_4	1.0	0.6	0.4	0.4	0.3	0.1

the case of N_c subcarriers, then 16-QAM modulation should be used in the case of $N_c/2$ subcarriers to maintain the same data transmission rate.

The BER for M-QAM modulation can be computed as [150]

$$\begin{aligned} P_b(\text{MQAM}) \;=\; & \frac{1}{\log_2 M_c}\left[\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M_c}}\right)\int_0^{\frac{\pi}{2}}\int_0^{\infty}\exp\left(-\frac{g_{\text{QAM}}\alpha^2 E_s}{\sin^2\phi\;\sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi\right. \\ & \left. -\;\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M_c}}\right)^2\int_0^{\frac{\pi}{4}}\int_0^{\infty}\exp\left(-\frac{g_{\text{QAM}}\alpha^2 E_s}{\sin^2\phi\;\sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi\right] \end{aligned} \tag{6.8}$$

where $g_{\text{QAM}} = 3/(2(M_c-1))$. Substituting $M_c = 16$ into (6.8), the BER of the OFDM scheme with APA that uses N_c bits of feedback, $N_c/2$ subcarriers, and 16-QAM modulation can be written as

$$\begin{aligned} P_b^{\text{APA}}(16\text{QAM},\boldsymbol{u},\boldsymbol{\psi}) \;=\; & \frac{1}{4}\left[\frac{3}{\pi}\sum_{i=1}^{4}\int_0^{\frac{\pi}{2}}\int_{\Omega_i}\exp\left(-\frac{0.1\alpha^2\psi_i E_s}{\sin^2\phi\;\sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi\right. \\ & \left. -\;\frac{9}{4\pi}\sum_{i=1}^{4}\int_0^{\frac{\pi}{4}}\int_{\Omega_i}\exp\left(-\frac{0.1\alpha^2\psi_i E_s}{\sin^2\phi\;\sigma_v^2}\right)p(\alpha)\,d\alpha\,d\phi\right]. \end{aligned}$$

Similar to (6.7), the parameters $\boldsymbol{u}$ and $\boldsymbol{\psi}$ of this scheme can be optimized by solving the following constrained optimization problem

$$\begin{aligned} \min_{\boldsymbol{u},\boldsymbol{\psi}} \quad & P_b^{\text{APA}}(16\text{QAM},\boldsymbol{u},\boldsymbol{\psi}) \\ \text{subject to} \quad & \sum_{i=1}^{4}\int_{\Omega_i}\psi_i p(\alpha)\,d\alpha = 1 \\ & 0<\psi_i<\psi_M, \quad i=1,2,3,4 \\ & 0<u_l<\infty, \quad l=1,2,3. \end{aligned} \tag{6.9}$$

Table 6.3: Optimal parameters of APA with reduced number of subcarriers.

SNR (dB)	0	5	10	15	20	25
u_1	0.4031	0.2265	0.2792	0.4386	0.2265	0.1591
u_2	0.8027	0.5364	0.8326	0.6563	0.4386	0.3246
u_3	0.9904	1.5174	1.2213	0.9904	0.8027	0.6856
ψ_1	0.1	0.1	0.3	2.3	6.8	18.7
ψ_2	0.4	0.3	1.4	1.3	2.1	2.5
ψ_3	1.2	1.3	0.9	0.8	0.8	0.8
ψ_4	1.8	0.9	0.6	0.4	0.3	0.2

Table 6.3 shows the optimal parameters for the APA scheme with a reduced number of subcarriers. These parameters are obtained by solving the optimization problem (6.9).

6.2.3 Adaptive Modulation Selection

The AMS scheme is based on the following idea. When a certain subcarrier is corrupted by fading, a constellation with smaller dimension and higher transmitted power can be assigned to this particular subcarrier, while constellations of larger dimensions and less transmitted power can be assigned to the subcarriers whose channel gain is high. Similar to the case of ASCS, a low-rate one-bit-per-subcarrier feedback can be used to divide the subcarriers into two groups that use different constellations and transmitted powers.

For example, to achieve the data rate of 2 bps per subcarrier, we can use the BPSK modulation at faded subcarriers and the 8-PSK modulation at non-faded subcarriers. In this case, the data rate can be expressed as $(\log_2 2) \int_0^u p(\alpha)\, d\alpha + (\log_2 8) \int_u^\infty p(\alpha)\, d\alpha$ where the threshold u of the channel gain is used to divide subcarriers into "faded" and "non-faded" groups. Taking into account that the data rate of 2 bps per subcarrier is chosen, the value of u can be found by solving the following data rate constraint equation

$$\int_0^u p(\alpha)\, d\alpha + 3 \int_u^\infty p(\alpha)\, d\alpha = 2. \tag{6.10}$$

Using (6.2), we obtain from (6.10) that $u = \sqrt{\ln 2}$. Then, the BER for this particular AMS

Table 6.4: Optimal parameters of AMS.

SNR (dB)	0	5	10	15	20	25
ψ_1	1.2925	1.0945	1.0554	1.2629	1.5799	1.8049
ψ_2	0.7075	0.9055	0.9446	0.7371	0.4201	0.1951

scheme can be written as

$$P_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_1, \psi_2) = \frac{1}{\pi}\left[\int_0^{\frac{\pi}{2}}\int_0^{\sqrt{\ln 2}} \exp\left(-\frac{\alpha^2\psi_1 E_s}{\sin^2\phi\ \sigma_v^2}\right) p(\alpha)\, d\alpha\, d\phi + \frac{1}{3}\int_0^{\frac{7\pi}{8}}\int_{\sqrt{\ln 2}}^{\infty} \exp\left(-\frac{\sin^2(\pi/8)\alpha^2\psi_2 E_s}{\sin^2\phi\ \sigma_v^2}\right) p(\alpha)\, d\alpha\, d\phi\right]. \quad (6.11)$$

The following constrained optimization problem should be solved to obtain the optimum power allocation in this case

$$\begin{aligned} \min_{\psi_1,\psi_2} \quad & P_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_1, \psi_2) \\ \text{subject to} \quad & \psi_1 + \psi_2 = 2 \\ & 0 < \psi_1, \psi_2 < 2. \end{aligned} \quad (6.12)$$

Table 6.4 lists the values of optimal power allocation for BPSK and 8-PSK constellations obtained by solving the optimization problem (6.12).

6.3 Imperfect One-Bit-Per-Subcarrier CSI Feedback

In the previous section, we assumed that the feedback channel is perfect. However, in real-world applications, this channel may be erroneous and/or may suffer from a feedback delay. In this section, we extend the results of the previous section to the imperfect feedback channel case by considering two types of imperfections: feedback errors and delays.

6.3.1 Erroneous Feedback Channel

We model the erroneous feedback channel as a binary symmetric channel with the error probability ζ. Note that the performance gain obtained from one bit CSI feedback decreases

with increasing ζ, and if ζ is high enough, then such erroneous CSI feedback can sometimes even worsen the system performance. Therefore, it is important to study the impact of erroneous feedback to the performance of the APA and AMS techniques.

Adaptive Power Allocation

Here, we consider only the conventional APA scheme, because obtaining optimal parameters for modified APA and APA with reduced number of subcarriers in the case of imperfect feedback seems to be mathematically intractable.[2] Taking the feedback error into account, the BER can be calculated as

$$Q_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2; \zeta) = (1-\zeta) P_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2) + \zeta\, P_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_2, \psi_1). \tag{6.13}$$

Substituting (6.5) into (6.13), we obtain the BER of the APA scheme. In the case of erroneous CSI feedback, the power constraint should also be modified as follows

$$\zeta \left(\int_0^u \psi_1 f(\alpha)\, d\alpha + \int_u^\infty \psi_2 f(\alpha)\, d\alpha \right) + (1-\zeta) \left(\int_0^u \psi_2 f(\alpha)\, d\alpha + \int_u^\infty \psi_1 f(\alpha)\, d\alpha \right) = 1\,. \tag{6.14}$$

Therefore, the optimal parameters from Table 6.1 can not be used in the case of erroneous CSI feedback, because (6.14) is not satisfied.

However, if the error probability ζ is known at the transmitter, we can find the optimal values of u, ψ_1, and ψ_2 that optimize the performance of the APA scheme under erroneous feedback. Similar to (6.6), the optimal values of u, ψ_1, and ψ_2 can be found as a solution of the following constrained optimization problem

$$\begin{aligned} \min_{u,\psi_1,\psi_2} \quad & Q_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2; \zeta) \\ \text{subject to} \quad & \zeta \left(\int_0^u \psi_1 f(\alpha)\, d\alpha + \int_u^\infty \psi_2 f(\alpha)\, d\alpha \right) + (1-\zeta) \left(\int_0^u \psi_2 f(\alpha)\, d\alpha + \int_u^\infty \psi_1 f(\alpha)\, d\alpha \right) = 1 \\ & 0 < \psi_1 < \psi_M; \quad 0 < \psi_2 < \psi_M; \quad u > 0. \end{aligned} \tag{6.15}$$

Note that both the objective function and the first constraint of the problem (6.15) differ from that of the problem (6.6), respectively. Table 6.5 summarizes optimal parameters

[2] Of course, brute force search can be performed here to find optimal parameters. However, it is unlikely to be used in practice.

Table 6.5: Optimal parameters of APA with erroneous feedback channel.

SNR (dB)		0	5	10	15	20	25
$\zeta = 0.15$	u	0.61	1.07	0.73	0.49	0.32	0.21
	ψ_1	0.71	1.13	1.41	1.84	2.36	2.82
	ψ_2	1.17	0.78	0.68	0.64	0.62	0.60
$\zeta = 0.4$	u	0.65	1.04	0.72	0.49	0.32	0.20
	ψ_1	0.94	1.05	1.12	1.18	1.21	1.23
	ψ_2	1.05	0.94	0.89	0.86	0.85	0.84

for the APA scheme with erroneous feedback channel when the probability of error in the feedback channel is equal to 0.15 and 0.4, respectively. It can be seen that when the feedback channel becomes less reliable, it is judicious to equally distribute the power between "faded" and "non-faded" channel realizations.

Adaptive Modulation Selection

For the AMS scheme with erroneous CSI feedback the BER can be calculated as

$$Q_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_1, \psi_2; \zeta) = (1-\zeta) P_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_1, \psi_2) + \zeta\, P_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_2, \psi_1). \tag{6.16}$$

Substituting (6.11) into (6.16), we can obtain the BER for the AMS scheme. In particular, we will find the critical value of the error probability ζ above which one-bit-per-subcarrier feedback can only worsen the system performance.[3] Specifically, the feedback remains meaningful only if the following condition is satisfied

$$Q_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_1, \psi_2; \zeta) \leq P_b(\text{QPSK}). \tag{6.17}$$

If (6.17) holds as equality, we obtain the critical value of ζ. These values for different SNRs are listed in Table 6.6 in the case when the optimal parameters from Table 6.4 are used. We can see from this table that the critical error probability of the feedback channel depends on the SNR conditions of the communication channel.

[3]For the APA scheme, the critical ζ is meaningless, because the power constraint may not be satisfied.

Table 6.6: Critical probability of feedback error of AMS.

SNR (dB)	0	5	10	15	20	25
ζ	0	0	0.1975	0.2079	0.1277	0.0603

Table 6.7: Optimal parameters of AMS with erroneous feedback channel.

SNR (dB)	0	5	10	15	20	25
ψ_1	1.3110	1.0596	0.9812	1.0628	1.0812	1.0653
ψ_2	0.6890	0.9404	1.0188	0.9372	0.9188	0.9347

However, if the error probability ζ is known at the transmitter, we can find the optimal values of ψ_1 and ψ_2 that optimize the performance of the AMS scheme under erroneous feedback channel. These optimal values can be found as a solution to the following optimization problem

$$\begin{aligned} \min_{\psi_1,\psi_2} \quad & Q_b^{\text{AMS}}(\text{BPSK}, 8\text{PSK}, \psi_1, \psi_2; \zeta) \\ \text{subject to} \quad & \psi_1 + \psi_2 = 2 \\ & 0 < \psi_1, \psi_2 < 2. \end{aligned} \tag{6.18}$$

Table 6.7 shows the optimal values of ψ_1 and ψ_2 for the AMS scheme with erroneous feedback channel when $\zeta = 0.15$.

6.3.2 Delayed Feedback Channel

The second source of imperfections in the feedback channel is the delay between the actual CSI and the CSI received at the transmitter. Therefore, it is also important to study the impact of outdated CSI on the APA and AMS approaches.

Let α_0 and α_τ be the channel gains at the time slots 0 and τ, respectively. It has been shown in [152] and [153, p. 142] that the joint PDF of α_0^2 and α_τ^2 has the following form

$$f_{\alpha_0^2,\alpha_\tau^2}(x, y; \rho) = \frac{1}{1-\rho} \exp\left(-\frac{x+y}{1-\rho}\right) I_0\left(\frac{2\sqrt{\rho xy}}{1-\rho}\right) \tag{6.19}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of the order 0, and $\rho = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)\mathrm{var}(y)}}$ is the correlation coefficient which characterizes the feedback delay.

Adaptive Power Allocation

Using (6.19), we obtain that in the case of delayed one bit CSI feedback and QPSK modulation, the BER for the APA scheme can be written as

$$R_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2; \rho) = \frac{1}{2\pi}\left[\int_0^{u^2}\int_0^{\infty}\int_0^{\frac{3\pi}{4}} \exp\Big(-\frac{xE_s\psi_1}{2\sin^2\phi\ \sigma_v^2}\Big) f_{\alpha_0^2,\alpha_\tau^2}(x,y;\rho)\, d\phi\, dx\, dy \right.$$
$$\left. + \int_{u^2}^{\infty}\int_0^{\infty}\int_0^{\frac{3\pi}{4}} \exp\Big(-\frac{xE_s\psi_2}{2\sin^2\phi\ \sigma_v^2}\Big) f_{\alpha_0^2,\alpha_\tau^2}(x,y;\rho)\, d\phi\, dx\, dy\right]. \tag{6.20}$$

Using the following property of the first-order Marcum Q-function [153, p. 75]

$$\int_0^{\infty} x \exp\left(-\frac{x^2+z^2}{2}\right) I_0(zx)\, dx = 1$$

we can simplify the integral in (6.20) as

$$R_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2; \rho) = \frac{1}{2\pi}\int_0^{\frac{3\pi}{4}} \left[\frac{1}{A_1+1}(1-\exp(-B_1u^2)) + \frac{1}{A_2+1}\exp(-B_2u^2)\right] d\phi \tag{6.21}$$

where

$$A_i = \frac{\psi_i E_s}{2\sin^2\phi\ \sigma_v^2}, \qquad B_i = \frac{\psi_i E_s + 2\sin^2\phi\ \sigma_v^2}{\psi_i E_s(1-\rho) + 2\sin^2\phi\ \sigma_v^2}, \qquad i = 1, 2.$$

It can be seen that if there is no feedback delay (i.e., $\rho = 1$), equation (6.21) yields the same BER result as in (6.5). It is also worth noting that when increasing the delay τ, the coefficient ρ decreases, but the BER increases. Therefore, we can find the critical value of the correlation coefficient ρ under which the CSI feedback becomes meaningless. This can be done by solving the following equation

$$R_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2; \rho) = P_b(\mathrm{QPSK}).$$

The critical values for the coefficient ρ at different SNRs and for optimal parameters from Table 6.1 are listed in Table 6.8. As we can see from this table, for some values of

Table 6.8: Critical ρ of conventional APA.

SNR (dB)	0	5	10	15	20	25
ρ	0.8121	0.4518	0.6489	0.7556	0.8510	0.9404

Table 6.9: Optimal parameters of conventional APA with feedback delay.

SNR (dB)	0	5	10	15	20	25
u	0.5364	1.1774	0.8326	0.8326	0.8326	0.8326
ψ_1	0.7	1.1	1.3	1.5	1.6	1.6
ψ_2	1.1	0.7	0.7	0.5	0.4	0.4

SNR the critical value of the coefficient ρ can be quite large and, thus, only very short feedback delays can be tolerated by the communication system. Moreover, we can also see from Table 6.8 that the critical value of ρ depends on SNR in a nonlinear way.

If ρ is known at the transmitter, we can find the values of u, ψ_1, and ψ_2 that optimize the performance of the APA scheme under delayed CSI feedback. These values can be found by solving the following constrained optimization problem

$$\begin{aligned} \min_{u,\psi_1,\psi_2} \quad & R_b^{\mathrm{APA}}(\mathrm{QPSK}, u, \psi_1, \psi_2; \rho) \\ \text{subject to} \quad & \int_0^u \psi_1 f(y)\, dy + \int_u^\infty \psi_2 f(y)\, dy = 1 \\ & 0 < \psi_1 < \psi_M; \quad 0 < \psi_2 < \psi_M; \quad u > 0. \end{aligned}$$

Table 6.9 summarizes these optimal parameters for the conventional APA scheme when $\rho = 0.8$.

Adaptive Modulation Selection

Let us study the performance of the AMS scheme in the case of outdated CSI feedback. Using equations (6.11) and (6.19), we obtain the BER of the AMS scheme in the following

Table 6.10: Critical ρ of AMS.

SNR (dB)	0	5	10	15	20	25
ρ	1	1	0.7515	0.6725	0.7455	0.8172

form:

$$R_b^{\mathrm{AMS}}(\mathrm{BPSK},8\mathrm{PSK},\psi_1,\psi_2;\rho)=\frac{1}{\pi}\left[\int_0^{\ln 2}\int_0^{\infty}\int_0^{\frac{\pi}{2}}\exp\Big(-\frac{xE_s\psi_1}{\sin^2\phi\ \sigma_v^2}\Big)f_{\alpha_0^2,\alpha_\tau^2}(x,y;\rho)\,d\phi\,dx\,dy\right.$$
$$\left.+\frac{1}{3}\int_{\ln 2}^{\infty}\int_0^{\infty}\int_0^{\frac{7\pi}{8}}\exp\Big(-\frac{\sin^2(\pi/8)xE_s\psi_2}{\sin^2\phi\ \sigma_v^2}\Big)f_{\alpha_0^2,\alpha_\tau^2}(x,y;\rho)\,d\phi\,dx\,dy\right].$$

Then, the critical value of ρ can be found by solving the following equation

$$R_b^{\mathrm{AMS}}(\mathrm{BPSK},8\mathrm{PSK},\psi_1,\psi_2;\rho)=P_b(\mathrm{QPSK}).$$

Table 6.10 summarizes the critical values of the coefficient ρ calculated for different SNRs and for optimal parameters from Table 6.4. Comparing Table 6.10 with Table 6.8, we observe that at moderate/high SNRs, the AMS scheme is more robust to CSI feedback delays than the APA approach. Moreover, the performance of the AMS scheme can be improved if the coefficient ρ is known at the transmitter. In such a case, we can find the optimal values of ψ_1 and ψ_2 by solving the following constrained optimization problem

$$\begin{aligned}\min_{\psi_1,\psi_2}\quad & R_b^{\mathrm{AMS}}(\mathrm{BPSK},8\mathrm{PSK},\psi_1,\psi_2;\rho)\\ \text{subject to}\quad & \psi_1+\psi_2=2\\ & 0<\psi_1,\psi_2<2.\end{aligned}$$

Table 6.11 shows the optimal values of these parameters for the AMS scheme with delayed CSI feedback when $\rho=0.8$.

6.4 Numerical Examples

The channel model used in our simulations is based on the ETSI "Vehicular A" channel environment [141], see Table 5.1. In all examples we assume that the transmitter transmits at the fixed data rate of $n_r=128$ bps and the available number of subcarriers is $N_c=64$.

Table 6.11: Optimal parameters of AMS with feedback delay.

SNR (dB)	0	5	10	15	20	25
ψ_1	1.3032	1.0568	0.9953	1.0811	1.2244	1.3412
ψ_2	0.6968	0.9432	1.0047	0.9189	0.7756	0.6588

6.4.1 Perfect CSI Feedback

Adaptive Subcarrier Selection

In the first example, three different system configurations are compared: where no subcarrier selection is used, where 32 "best" subcarriers are selected, and where 16 "best" subcarriers are selected. To keep the same data rate for each system configuration, we use the QPSK modulation for no subcarrier selection, 16-QAM modulation for the selection of 32 subcarriers, and 256-QAM modulation for the selection of 16 subcarriers.

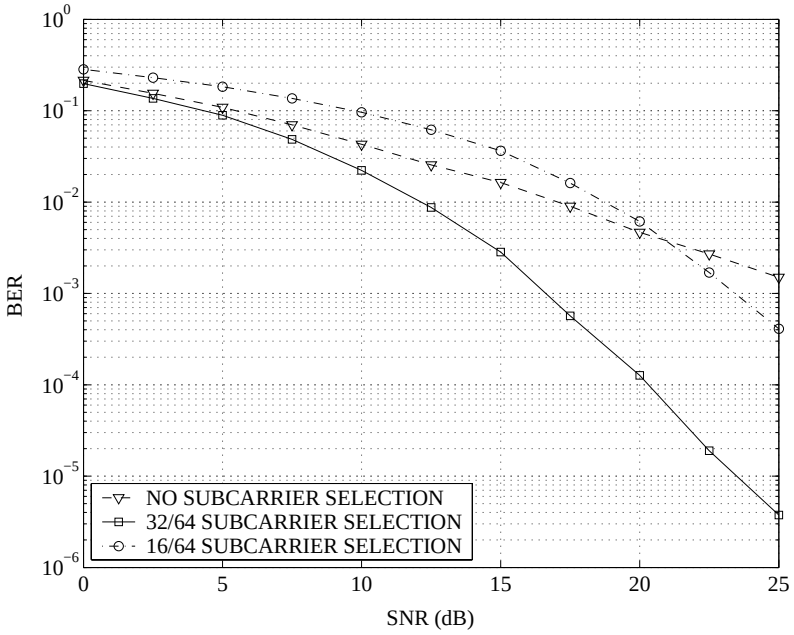

Figure 6.1: BER versus SNR; perfect CSI feedback. ASCS with different numbers of selected subcarriers.

Figure 6.1 shows the performance of all three system configurations in terms of BER

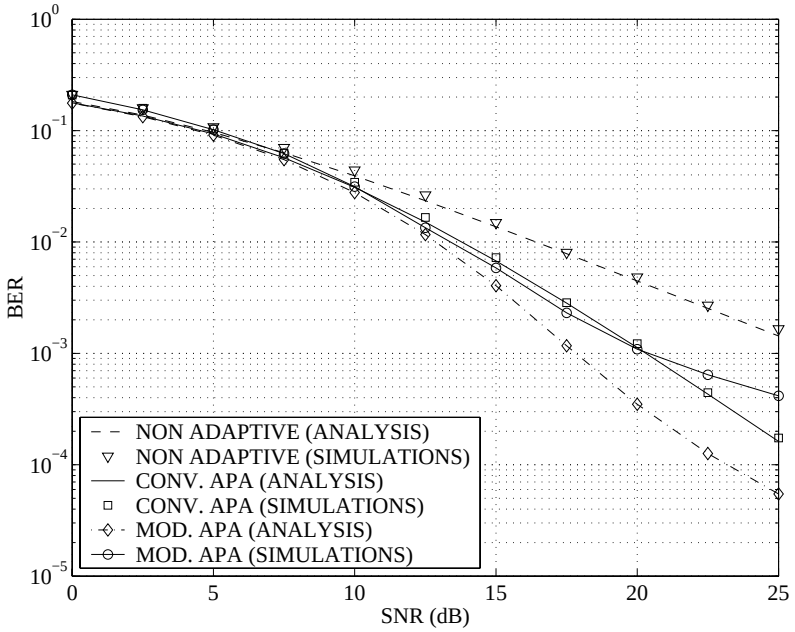

Figure 6.2: BER versus SNR; perfect CSI feedback. Conventional and modified APA.

versus SNR. The tradeoff between the number of subcarriers and the modulation used can be seen from this figure. In particular, the adaptive selection of 32 subcarriers has the best performance among the system configurations tested. However, the adaptive selection of 16 subcarriers has much worse performance than that of 32 subcarriers and at low and moderate SNRs can even perform worse than the configuration without subcarrier selection.

We stress here that, since no theoretical analysis of BER is possible for the ASCS scheme, this may limit its practical application. Moreover, in multiuser OFDM communication scenarios only a small number of subcarriers may be assigned to each user, and this scheme may not be applicable.

Adaptive Power Allocation

Hereafter, in each figure we display theoretical BER curves that correspond to the derived analytical expressions, and the numerical BER curves obtained via simulations.

BERs of the conventional and modified APA schemes are shown versus SNR in Figure 6.2. The QPSK modulation and the optimal parameters from Tables 6.1 and 6.2 are

used, respectively, for these two schemes. The theoretical BER for the modified APA scheme with the optimal parameters from Table 6.2 assumes that there is a full correlation between each pair of adjacent subcarriers.

From Figure 6.2, it can be seen that both tested APA schemes outperform the non-adaptive OFDM scheme. There is only a slight performance difference between conventional and modified APA.

Comparing the results of Figure 6.2 with that of Figure 6.1, we can see that the APA approach is less efficient than the ASCS scheme with 32 selected subcarriers. This is especially true at high SNRs. However, the APA scheme allows an easier optimization as compared with the ASCS approach.

Figure 6.3 shows BER of the APA scheme with reduced number of subcarriers versus SNR. The optimal parameters from Table 6.3 are used to obtain this figure. It can be seen that this scheme performs better than the non-adaptive OFDM scheme at moderate and high SNRs. However, it has higher BER than the conventional APA scheme and the ASCS scheme with 32 selected subcarriers in the SNR interval of $[0; 20]$ dB. In other words, the APA scheme with reduced number of subcarriers does not bring any performance improvements as compared with the conventional APA approach. Note that, due to the fact that (6.4) is an approximation, the theoretical and numerical curves do not coincide at low SNRs in the case when large constellation dimensions are used.

Adaptive Modulation Selection

Figure 6.4 displays BER of the AMS scheme with the optimal parameters taken from Table 6.4 versus SNR. In this figure, the BPSK and 8-PSK modulations are used at "faded" and "non-faded" subcarriers, respectively. It can be seen from Figure 6.4 that the AMS scheme outperforms the non-adaptive approach. However, the AMS scheme has higher BER than the conventional APA approach. Moreover, comparing Figure 6.4 with Figures 6.2 and 6.3, we can notice that the AMS scheme has higher BER then the modified APA approach, but outperforms the APA approach with reduced number of subcarriers at low and moderate

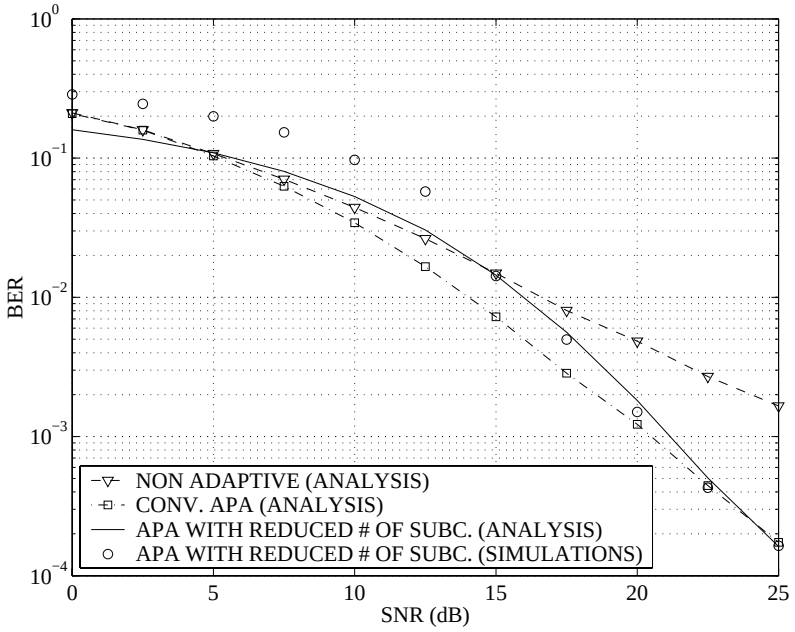

Figure 6.3: BER versus SNR; perfect CSI feedback. Conventional APA and APA with reduced number of subcarriers.

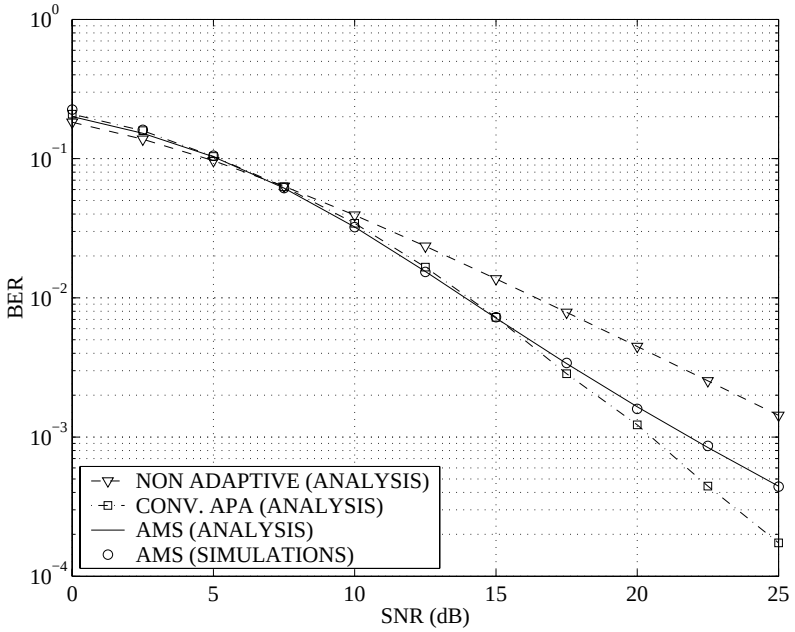

Figure 6.4: BER versus SNR; perfect CSI feedback. Conventional APA and AMS.

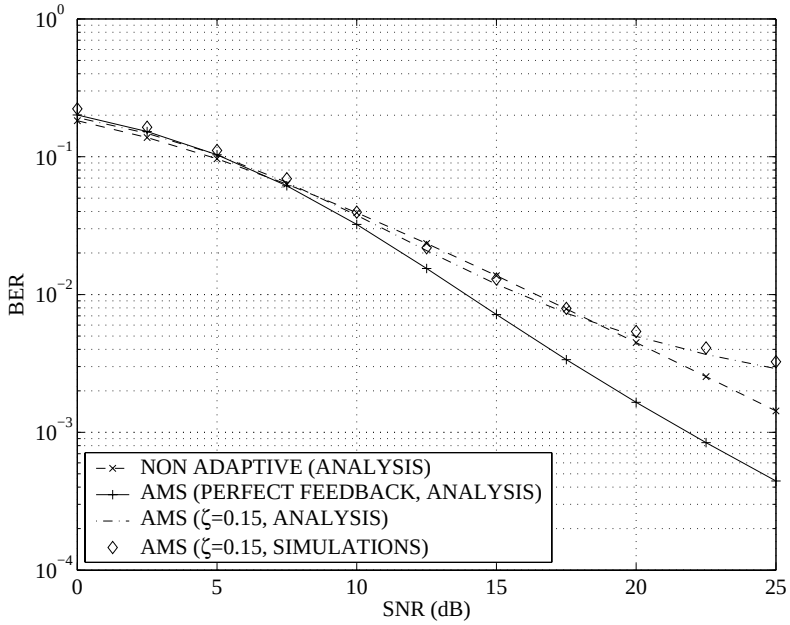

Figure 6.5: BER versus SNR; erroneous CSI feedback. AMS.

SNRs.

6.4.2 Imperfect CSI Feedback

Erroneous Feedback Channel

Figure 6.5 displays BER of the AMS scheme with erroneous feedback channel versus SNR. In this figure, $\zeta = 0.15$ is taken. The optimal parameters from Table 6.4 are selected, and the BPSK and 8-PSK modulations are used at "faded" and "non-faded" subcarriers, respectively.

We can see from this figure that the performance of the AMS approach degrades severely in the case of erroneous feedback compared with the perfect feedback case. For example, at the BER of $3 \cdot 10^{-3}$, the performance degradation of the AMS approach with erroneous feedback amounts to 7.5 dB compared with the performance of the AMS approach with perfect feedback. Moreover, at high SNRs the AMS scheme performs worse than the non-adaptive scheme.

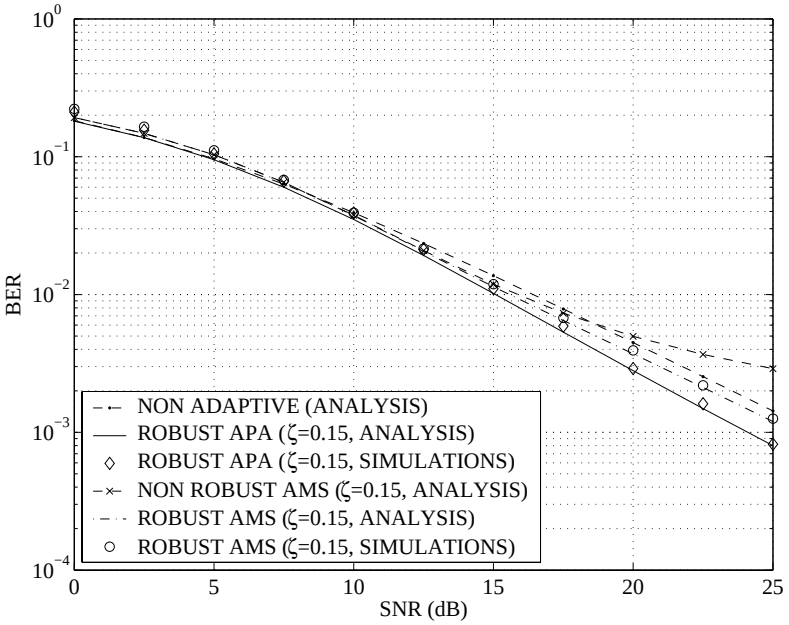

Figure 6.6: BER versus SNR; erroneous CSI feedback. Conventional APA and AMS (non-robust and robust).

It is worth noting that these simulation results also agree with the results shown in Table 6.6. In particular, from Table 6.6 it can be seen that the critical value of the error probability ζ at low SNRs of 0 and 5 dB is equal to zero, which means that the AMS scheme cannot tolerate any errors in the feedback channel. Indeed, from Figure 6.5 we can observe that the theoretical BER of the AMS scheme is larger than that of the non-adaptive scheme at the SNRs of 0 and 5 dB. Moreover, as we can see from Table 6.6, the critical values of ζ at the SNRs of 20 and 25 dB are lower than the value of ζ used in simulations and, indeed, the performance of the AMS scheme is worse than the performance of non-adaptive scheme in the SNR interval of $[20; 25]$ dB (see Figure 6.5).

The BERs of the conventional APA and AMS schemes versus SNR are shown in Figure 6.6 for $\zeta = 0.15$. The QPSK modulation is used in this figure for the APA technique. The optimal parameters from Table 6.5 are used for the APA technique which we refer to as "robust APA". This terminology reflects the fact that ζ is known and is used in the APA technique to obtain the optimal parameters u, ψ_1, and ψ_2 from solving (6.15). For

the AMS scheme, the BPSK and 8-PSK modulations are used for "faded" and "non-faded" subcarriers, respectively, and the optimal parameters from Table 6.7 are selected. The terms "robust AMS" and "non robust AMS" correspond to the cases of known and unknown ζ, respectively. In the first case, the parameters ψ_1 and ψ_2 are optimized by solving (6.18) (which uses the knowledge of ζ), while in the second case these parameters are obtained from solving (6.12) which does not use this knowledge.

As can be seen from Figure 6.6, the performance of the AMS scheme can be significantly improved if the error probability ζ is known at the transmitter. It can also be seen that the robust APA scheme outperforms the non-adaptive one.

Delayed Feedback Channel

Jakes' fading model is used to simulate the delayed feedback channel [5]. The maximal Doppler frequency of 67 Hz is used, which corresponds to the vehicular speed of 36 km/h at the carrier frequency of 2 GHz. We take $\rho = 0.8$, which corresponds to the feedback delay of 37 symbol durations in the IS-136 standard [154].

The BER versus SNR curves for the APA and AMS schemes using the optimal parameters of Tables 6.1 and 6.4, respectively, are shown in Figure 6.7. The QPSK modulation is used for the conventional APA approach, while the BPSK and 8-PSK modulations are used at "faded" and "non-faded" subcarriers, respectively, for the AMS scheme.

From Figure 6.7, we can conclude that at the moderate and high SNRs, the APA approach is more sensitive to the delay in the feedback channel as compared with the AMS approach. At the BER of $2 \cdot 10^{-3}$, the performance degradation of the APA approach due to the feedback channel delay amounts to 7 dB, while the corresponding degradation of the AMS scheme is only 3 dB. Moreover, the APA scheme shows in this case worse performance than the non-adaptive OFDM technique.

It is also worth noting that the results of Figure 6.7 agree with the results of Tables 6.8 and 6.10. Indeed, the theoretical BER for the AMS technique can be seen to be higher than that for the non-adaptive scheme at SNRs of 0 and 5 dB. For these SNRs, as can be

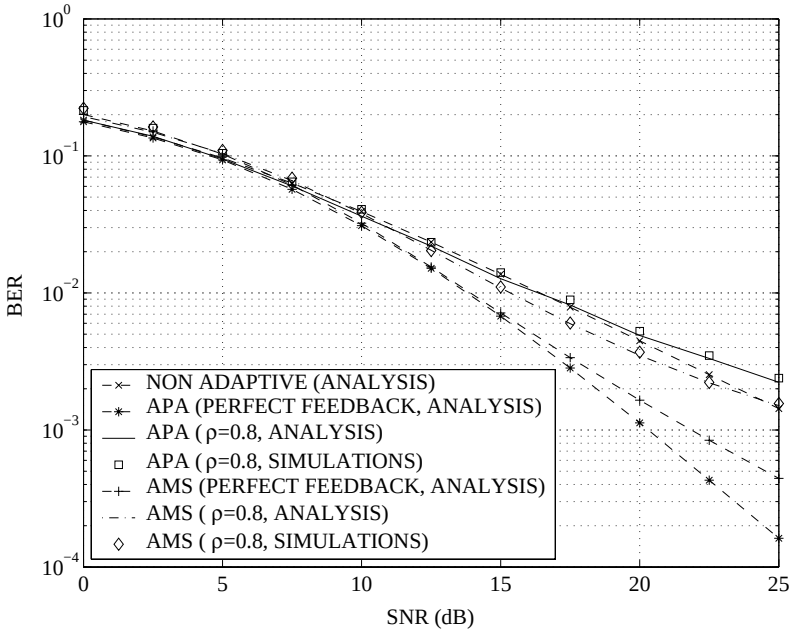

Figure 6.7: BER versus SNR; delayed CSI feedback. Conventional APA and AMS.

seen from Table 6.10, the critical value of the correlation coefficient ρ is equal to one, which means that the AMS scheme cannot tolerate any delays in the feedback channel. Moreover, the APA and AMS approaches perform worse than the non-adaptive scheme at the SNRs for which the critical value of ρ given in Tables 6.8 and 6.10 is higher than the value of 0.8 that is used in Figure 6.7.

Figure 6.8 displays the BERs of the conventional APA and AMS schemes with delayed feedback channel versus SNR. In this figure, the coefficient ρ is assumed to be known at the transmitter. Similar to the case of erroneous feedback channel, we refer to such APA and AMS approaches as robust ones. Vice versa, we call the APA and AMS schemes non-robust if the optimal parameters are found under the assumption that ρ is unknown. The optimal parameters for the robust APA and AMS techniques are shown in Tables 6.9 and 6.11, respectively.

It can be clearly seen from Figure 6.8 that the performance of both APA and AMS approaches can be substantially improved if the parameter ρ is known at the transmitter. This improvement is more pronounced for the APA scheme.

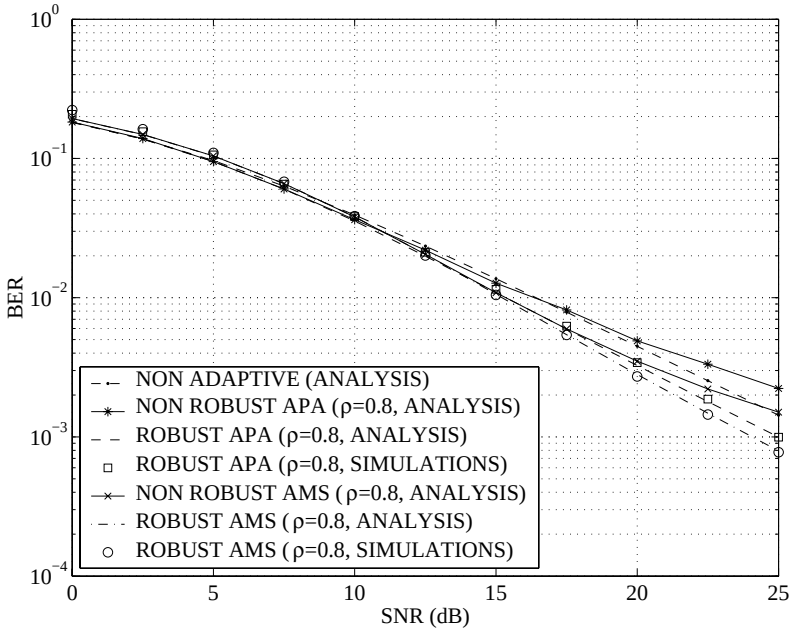

Figure 6.8: BER versus SNR; delayed CSI feedback. Conventional APA and AMS (non-robust and robust).

6.5 Chapter Summary

In this chapter, the performance of OFDM communication systems with one-bit-per-subcarrier CSI feedback has been studied. Three adaptive techniques including ASCS, APA, and AMS schemes have been used to exploit such CSI feedback. We found that even one-bit-per-subcarrier CSI feedback can greatly improve the overall system performance if the feedback channel is perfect. Among the three approaches tested, the ASCS approach has the best performance. However, the performance of OFDM systems with one-bit-per-subcarrier feedback can be even worse than the performance of the OFDM system without any feedback if the feedback channel is imperfect. It has been demonstrated that the performance of both the APA and AMS approaches can be substantially improved by exploiting the knowledge of how imperfect the feedback channel is.

Chapter 7

Concluding Remarks and Future Work

In wireless communications, multi-antenna and multi-carrier techniques attract great research interest due to their advantages in enhancing system capacity and improving communication quality. In this thesis, we have proposed and studied several advanced algorithms for multi-antenna and multi-carrier communications.

Concluding Remarks:

In Chapter 2, we have addressed the spatial signature estimation problem. Spatial signature estimation is of great interest in wireless communications, since the knowledge of spatial signatures can be used for beamforming to separate the user-of-interest from other interfering users. We have proposed bandwidth-efficient approaches to estimate the user spatial signature. The proposed approaches employ the time-varying user power loading, which makes the model identifiable. Based on the PARAFAC analysis model, two algorithms have been developed. The first one is based on the TALS regression in PARAFAC fitting, and the second one utilizes joint matrix diagonalization. Compared with other blind spatial signature estimation approaches, our estimators provide better performance and are applicable to a more general class of scenarios.

Chapter 3 has investigated the problem of robust blind DS-CDMA multiuser detection in impulsive ambient noise. By exploiting the multiple antennas at the receiver, this problem can be linked to robust PARAFAC fitting in impulsive noise. The conventional PARAFAC

fitting procedure applies LS fitting, and its performance degrades dramatically if the additive noise has strong impulsive components. Motivated by this fact, we have proposed two robust fitting procedures for PARAFAC model, which utilize the LAE criterion. These two procedures work in an iterative fashion and make use of LP and WMF, respectively. In comparison to the blind DS-CDMA multiuser detection algorithm with conventional TALS PARAFAC fitting, the proposed approaches show distinct performance improvement in impulsive noise, while only small performance degradation in Gaussian noise. Compared with the TALS PARAFAC fitting procedure which employs *ad hoc* nonlinear preprocessing techniques, the proposed procedures also show better performance. Between two proposed algorithms, the WMF iteration is particular appealing from a simplicity viewpoint.

In Chapter 4, we have studied multiple-access MIMO wireless communications when transmitters (users) also have multiple antennas and use OSTBCs. The linear receiver design problem for joint space-time decoding and interference rejection has been addressed. In particular, we have considered the practical scenario when the receiver has only presumed CSI, which is the approximated version of the actual one. Using different approaches to model the CSI mismatches, two classes of robust linear receivers have been proposed. The first class of receivers is based on the worst-case performance optimization when the deterministic CSI mismatch model is used, while the second one employs the stochastic programming technique, based on the stochastic CSI mismatch model. Compared with the existing linear receivers for multiple-access MIMO communication, both classes of linear receivers show an improved robustness against CSI errors.

In Chapter 5, we have investigated the problem of linear precoder design for OFDM systems. We have proposed a cutoff rate maximization-based linear block precoder, which only requires the knowledge of the average relative channel multipath powers and delays at the transmitter. Substantial performance improvements are achieved by our precoding technique relative to the existing linear block precoders. It has been shown that the proposed linear precoder can be combined with channel coding technique to further enhance the performance of OFDM systems.

Finally, in Chapter 6, we have analyzed the performance of an adaptive OFDM system

whose transmitter has only one-bit-per-subcarrier channel state feedback. Three adaptive techniques: ASCS, APA, and AMS have been proposed to exploit this feedback. It has been shown that this one-bit feedback can greatly improve the system performance. Moreover, the impact of imperfections of the feedback channel on the system performance, and the ways to improve the performance by exploiting the knowledge of how imperfect the feedback channel is have also been studied in Chapter 6.

Future Work:

In Chapter 3, for the TALS PARAFAC fitting procedure with nonlinear processing, the nonlinear filtering is performed *before* the TALS fitting. Thus, the signal component may also be distorted. It is therefore interesting to study the performance of the TALS procedure with nonlinear operations performed *after* the TALS fitting, which will only process the impulsive noise component. However, since the DS-CDMA multiuser detector proposed in Chapter 3 is fully blind and the fitting procedures work in an iterative fashion, nonlinear operations after LS fitting may bring some problems such as the convergence of the iterative procedure. This issue is interesting and deserves further study.

The robust linear receivers proposed in Chapter 4 were developed under the assumption of quasi-static fading channel. This means that the MIMO channel is fixed for one block duration and independently changed in the next block. However, in practical scenarios, the channel fading is continuously varying. Therefore, this observation leaves room for further study of linear receivers robust against CSI errors in continuously time-varying channels.

For the linear block precoding technique developed in Chapter 5, when the number of subcarriers assigned to one user is large, the computational complexity of the proposed precoder design is high. How to reduce the computational burden of the optimization problem can be the subject of a deeper study.

In Chapter 6, the optimal parameters for the APA and AMS schemes are derived in the case of uncoded transmissions. It would be interesting to apply these adaptive methods and derive the optimal parameters for the channel-coded OFDM systems.

Bibliography

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4-24, April. 1988.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, vol. 13, pp. 67-94, July 1996.

[3] H. L. Van Trees, *Optimum Array Processing*. New York: John Wiley & Sons, 2002.

[4] M. Haardt and A. Alexiou, "Smart antennas and related technologies," in Technologies for the Wireless Future - Wireless World Research Forum (WWRF) (R. Tafazolli, ed.), pp. 414-428, Chichester, UK: John Wiley & Sons, 2005, Chapter 7, Section 3.

[5] J. Proakis, *Digital Communications*, 4th Edition, New York: McGraw-Hill, 2001.

[6] A. J. Paulraj, R. U. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications.* Cambridge, UK: Cambridge University Press, 2003.

[7] A. B. Gershman and N. D. Sidiropoulos, *Space-Time Processing for MIMO Communications.* Chichester, UK: John Wiley & Sons, 2005.

[8] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications.* Cambridge, UK: Cambridge University Press, 2003.

[9] A. Alexiou and M. Haardt, "Smart antenna technologies for future wireless systems: Trends and challenges," IEEE Communications Magazine, vol. 42, pp. 90-97, Sep. 2004.

[10] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: The value of spatial diversity in wireless networks," *Proc. IEEE*, vol. 92, pp. 219-270, Feb. 2004.

[11] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommunications*, vol. 10, pp. 585-595, Nov. 1999.

[12] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311-335, 1998.

[13] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*. Telecom Publishing, 1992.

[14] E. G. Larsson, "The ubiquitous signal processing," Ph.D. dissertation, Uppsala University, 2002.

[15] *Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 4)*, 3GPP Standard, Document TR 25.848 V4.0.0, March 2001.

[16] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[18] R. D. J. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Boston: Artech House, 2000.

[19] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Magazine*, vol. 17, pp. 29-48, May 2000.

[20] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: A convenient framework for time-frequency processing in wireless communications," *Proc. IEEE*, vol. 88, pp. 611-640, May 2000.

[21] *Wireless LAN Medium Access Control and Physical Layer Specifications*, IEEE Standard 802.11a, 2000.

[22] *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Standard 802.16a, 2003.

[23] *Broadband Radio Access Networks; HIPERLAN Type 2; Physical Layer*, ETSI Standard, Document TS 101 475 V1.1.1, April 2000.

[24] B. Ottersten, "Array processing for wireless communications," in *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, Greece, July 1996. pp. 466-473.

[25] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, pp. 49-83, Nov. 1997.

[26] J. H. Winters, "Smart antennas for wireless systems, " *IEEE Personal Communications Magazine*, vol. 5, pp. 23-27, Feb. 1998.

[27] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Communications*, vol. 42, pp. 1740-1751, Feb.-April 1994.

[28] D. Asztèly, B. Ottersten, and A. L. Swindlehurst, "Generalised array manifold model for wireless communication channel with local scattering," *IEE Proc. Radar, Sonar Navigation*, vol. 145, pp. 51-57, Feb. 1998.

[29] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, pp. 449-462, Feb. 1998.

[30] S. S. Jeng, H. P. Lin, G. Xu, and W. J. Vogel, "Measurements of spatial signature of an antenna array," in *Proc. PIMRC*, Sep. 1995, Toronto, Canada, vol. 2, pp. 669-672.

[31] D. Astèly, A. L. Swindlehurst, and B. Ottersten, "Spatial signature estimation for uniform linear arrays with unknown receiver gains and phases," *IEEE Trans. Signal Processing*, vol. 47, pp. 2128-2138, Aug. 1999.

[32] A. J. Weiss and B. Friedlander, "'Almost blind' steering vector estimation using second-order moments," *IEEE Trans. Signal Processing*, vol. 44, pp. 1024-1027, April 1996.

[33] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE*, vol. 78, pp. 753-767, April 1990.

[34] Q. Wu and K. M. Wong, "Blind adaptive beamforming for cyclostationary signals," *IEEE Trans. Signal Processing*, vol. 44, pp. 2757-2767, Nov. 1996.

[35] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.*, vol. F-140, pp. 362-370, Dec. 1993.

[36] M. C. Dogan and J. M. Mendel, "Cumulant-based blind optimum beamforming," *IEEE Trans. Aerospace and Electronic Systems*, vol. 30, pp. 722-741, July 1994.

[37] E. Gonen and J. M. Mendel, "Applications of cumulants to array processing, Part III: Blind beamforming for coherent signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 2252-2264, Sep. 1997.

[38] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434-444, Feb. 1997.

[39] N. Yuen and B. Friedlander, "Performance analysis of blind signal copy using fourth order cumulants," *J. Adaptive Control and Signal Processing,* vol. 10, No. 2/3, pp. 239-266, 1996.

[40] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Trans. Vehicular Technology*, vol. 49, pp. 437-447, March 2000.

[41] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "Spatial channel characteristics in outdoor environments and their impact on BS antenna system performance," in *Proc. VTC*, Ottawa, Canada, May 1998, vol. 2, pp. 719-723.

[42] R. A. Harshman, "Foundation of the PARAFAC procedure: Model and conditions for an "explanatory" multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1-84, Dec. 1970.

[43] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Applicat.*, vol. 16, pp. 95-138, 1977.

[44] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 810-823, March 2000.

[45] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometrics*, vol. 14, pp. 229-239, 2000.

[46] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Processing*, vol. 48, pp. 2377-2388, Aug. 2000.

[47] M. K. Tsatsanis and C. Kweon, "Blind source separation of non-stationary sources using second-order statistics," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998, vol. 2, pp. 1574 -1578.

[48] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources," *IEEE Trans. Signal Processing*, vol. 49 pp. 1837-1848, Sep. 2001.

[49] R. L. Harshman, "Determination and proof of minimum uniqueness conditions for PARAFAC1," *UCLA Working Papers in Phonetics*, vol. 22, pp. 111-117, 1972.

[50] J. M. F. ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOMP/PARAFAC," *Psychometrika*, vol. 67, No. 3, Sep. 2002.

[51] J. M. F. ten Berge, N. D. Sidiropoulos, and R. Rocci, "Typical rank and INDSCAL dimensionality for symmetric three-way arrays of order I×2×2 or I×3×3," *Linear Algebra Applicat.*, to appear.

[52] T. Jiang, N. D. Sidiropoulos, and J. M. F. ten Berge, "Almost-sure identifiability of multi-dimensional harmonic retrieval," *IEEE Trans. Signal Processing*, vol. 49, pp. 1849-1859, Sep. 2001.

[53] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, Aussois, France, Jan. 1999. *(http://www.ece.umn.edu/users/nikos/public_ html/3SPICE/code.html)*

[54] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Analysis and Applicat.*, vol. 17, pp. 161-164, Jan. 1996.

[55] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1545-1553, July 2002.

[56] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783-1795, Oct. 1990.

[57] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ: Prentice-Hall, 1993.

[58] V. K. Garg, *IS-95 CDMA and cdma2000 Cellular/PCS Systems Implementation.* Upper Saddle River, NJ: Prentice-Hall, 2000.

[59] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Communications*, vol. 38, pp. 509-519, April 1990.

[60] U. Madhow, M. L. Honig, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Information Theory*, vol. 41, pp. 944-960, July 1995.

[61] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Communications*, vol. 42, pp. 3178-3188, Dec. 1994.

[62] M. Götz, M. Rapp, and K. Dostert, "Power line channel characteristics and their effects on communication system design," *IEEE Communications Magazine*, vol. 42, pp. 78-86, April 2004.

[63] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 991-1001, Sep. 1993.

[64] T. K. Blankenship, D. M. Kriztman, and T. S. Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," in *Proc. VTC*, Phoenix, AZ, May 1997, vol. 3, pp. 1942-1946.

[65] D. Parson, *The Mobile Radio Propagation Channel*. New York: Wiley, 1996.

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

[67] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatters," *IEEE Trans. Signal Processing*, vol. 46, pp. 1601-1611, June 1998.

[68] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman and Hall, 1994.

[69] B. Aazhang and H. V. Poor, "Performance of DS/SSMA communications in impulsive channels – part I: Linear correlation receivers," *IEEE Trans. Communications*, vol. 35, pp. 1179-1187, Nov. 1987.

[70] B. Aazhang and H. V. Poor, "Performance of DS/SSMA communications in impulsive channels – part II: Hard-limiting correlation receivers," *IEEE Trans. Communications*, vol. 36, pp. 88-96, Jan. 1988.

[71] A. M. Zoubir and A. T. Lane-Glover, "Multiuser detection in impulsive noise," in *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, Aug. 2001, pp. 102-105.

[72] H. Liu and G. Xu, "A subspace method for signature waveform estimation in synchronous CDMA systems," *IEEE Trans. Communnications*, vol. 44, pp. 1346-1354, Oct. 1996.

[73] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," *Research Methods for Multimode Data Analysis*, H. G. Law, C. W. Jr. Snyder, J. Hattie, and R. P. McDonald, Eds., Praeger, New York, 1984, pp. 122-215.

[74] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283-319, 1970.

[75] P. Comon, "Independent component analysis: A new concept?" *Signal Processing*, vol. 36, pp. 287-314, April 1994.

[76] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, ..., R_N)$ approximation and applications of higher-order tensors", *SIAM J. Matrix Analysis and Applicat.*, vol. 21, pp. 1324-1342, 2000.

[77] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis and (simultaneous) third-order tensor diagonalization", *IEEE Trans. Signal Processing*, vol. 49, pp. 2262-2271, Oct. 2001.

[78] E. Kofidis and P. A. Regalia, "On the best rank-1 approximation of higher-order supersymmetric tensors," *SIAM J. Matrix Analysis and Applicat.*, vol. 23, pp. 863-884, 2002.

[79] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Processing*, vol. 43, pp. 2700-2713, Nov. 1995.

[80] P. Tsakalides, "Array signal processing with alpha-stable distributions," Ph.D. dissertation, University of Southern California, Dec. 1995.

[81] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[82] N. D. Sidiropoulos and R. Bro, "Mathematiacl programming algorithms for regression-based filtering in $\mathbb{R}^N$," *IEEE Trans. Signal Processing*, vol. 47, pp. 771-782, March 1999.

[83] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filtering under structural constraints," *IEEE Trans. Signal Processing*, vol. 43, pp. 591-603, March 1995.

[84] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, no. 4, pp. 373-395, 1984.

[85] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

[86] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimantional arrays," *IEEE Trans. Signal Processing*, vol. 49, pp. 2074-2086, Sep. 2001.

[87] X. Wang and H. V. Poor, "Space-time multiuser detection in multipath CDMA channels," *IEEE Trans. Signal Processing*, vol. 47, pp. 2356-2374, Sep. 1999.

[88] S. N. Batalama, M. J. Medley, and I. N. Psaromiligkos, "Adaptive robust spread-spectrum receivers," *IEEE Trans. Communications*, vol. 47, pp. 905-917, June 1999.

[89] X. Wang and H. V. Poor, "Robust multiuser detection in non-Gaussian channels," *IEEE Trans. Signal Processing*, vol. 47, pp. 289-305, Feb. 1999.

[90] *The MOSEK optimization tools version 2.0 (Build 19). User's manual and reference,* EKA Consulting ApS, Denmark, 2001. *(http://www.mosek.com)*

[91] J. G. Gonzalez, "Robust techniques for wireless communications in non-Gaussian environments," Ph.D. dissertation, University of Delaware, 1997.

[92] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Upper Saddle River, NJ: Prentice-Hall, 1998.

[93] A. Swami, "Cramér-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Processing*, vol. 53, pp. 231-244, Sep. 1996.

[94] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Processing*, vol. 82, pp. 1829-1846, Dec. 2002.

[95] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rates wireless communications: Performance criterion and code construction," *IEEE Trans. Information Theory*, vol. 44, pp. 744-765, March 1998.

[96] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

[97] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Information Theory*, vol. 45, pp. 1456-1467, July 1999.

[98] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Trans. Information Theory*, vol. 47, pp. 1650-1656, May 2001.

[99] H. Li, X. Lu, and G. B. Giannakis, "Capon multiuser receiver for CDMA systems with space-time coding," *IEEE Trans. Signal Processing*, vol. 50, pp. 1193-1204, May 2002.

[100] D. Reynolds, X. Wang, and H. V. Poor, "Blind adaptive space-time multiuser detection with multiple transmitter and receiver antennas," *IEEE Trans. Signal Processing*, vol. 50, pp. 1261-1276, June 2002.

[101] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Applications of space-time block codes and interference suppression for high capacity and high data rate wireless systems," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998, vol. 2, pp. 1803-1810.

[102] A. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, pp. 76-92, May 2000.

[103] S. Shahbazpanahi, M. Beheshti, A. B. Gershman, M. Gharavi-Alkhansari, and K. M. Wong, "Minimum variance linear receivers for multi-access MIMO wireless systems with space-time block coding," *IEEE Trans. Signal Processing*, vol. 52, pp. 3306-3313, Dec. 2004.

[104] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, pp. 313-324, Feb. 2003.

[105] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Processing*, vol. 51, pp. 2257-2269, Sep. 2003.

[106] S. Vorobyov, A. B. Gershman, Z.-Q. Luo, and N. Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," *IEEE Signal Processing Letters*, vol. 11, pp. 108-111, Feb. 2004.

[107] R. Lorenz and S. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Processing*, vol. 53, pp. 1684-1696, May 2005.

[108] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2003.

[109] S. Cui, Z.-Q. Luo, and Z. Ding, "Robust blind multiuser detection against CDMA signature mismatch," in *Proc. ICASSP*, Salt Lake City, UT, May 2001, pp. 2297-2300.

[110] S. Shahbazpanahi and A. B. Gershman, "Robust blind multiuser detection for synchronous CDMA systems using worst-case performance optimization," *IEEE Trans. Wireless Communications*, vol. 3, pp. 2232-2245, Nov. 2004.

[111] K. Zarifi, S. Shahbazpanahi, A.B. Gershman, and Z.-Q. Luo, "Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver," *IEEE Trans. Signal Processing*, vol 53, pp. 295-305, Jan. 2005.

[112] M. Biguesh, S. Shahbazpanahi, and A. B. Gershman, "Robust power adjustment for transmit beamforming in cellular communication systems," in *Proc. ICASSP*, Hong Kong, April 2003, vol. 5, pp. 105-108.

[113] M. Biguesh, S. Shahbazpanahi, and A. B. Gershman, "Robust downlink power control in wireless cellular systems," *EURASIP J. Wireless Communications and Networking*, special issue on Multiuser MIMO Networks, no. 2, pp. 261-272, Dec. 2004.

[114] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Information Theory*, vol. 48, pp. 1804-1824, July 2002.

[115] M. Gharavi-Alkhansari and A. B. Gershman, "Constellation space invariance of space-time block codes with application to optimal ML decoding," *IEEE Trans. Information Theory*, vol. 51, pp. 331-334, Jan. 2005.

[116] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Application of second-order cone programming," *Linear Algebra Applicat.*, pp. 193-228, Nov. 1998.

[117] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625-653, Aug. 1999.

[118] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772-781, Sep. 1978.

[119] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. New York: Springer-Verlag, 1997.

[120] A. Prékopa, *Stochastic Programming*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1995.

[121] D. P. Bertsekas, *Nonlinear Programming*. Belmont, Athena Scientific, 1995.

[122] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Information Theory*, vol. 49, pp. 951-963, April 2003.

[123] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, Sep. 22-24, 1999.

[124] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. 3rd Edition, New York: McGraw-Hill, 1991.

[125] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, pp. 2381-2401, Sep. 2003.

[126] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988-2006, July 1999.

[127] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalizations," *IEEE Trans. Signal Processing*, vol. 51, pp. 2410-2423, Sep. 2003.

[128] S. S. Chan, T. N. Davidson, and K. M. Wong, "Asymptotically minimum BER linear block precoders for MMSE equalization," *IEE Proc. Communications*, vol. 151, pp. 297-304, Aug. 2004.

[129] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Information Theory*, vol. 45, pp. 1019-1032, April 1999.

[130] Z. Liu, Y. Xin, and G. B. Giannakis, "Linear constellation precoding for OFDM with maximum multipath diversity and coding gains," *IEEE Trans. Communications*, vol. 51, pp. 416-427, March 2003.

[131] Y.-P. Lin and S.-M. Phoong, "BER minimized OFDM systems with channel independent precoders," *IEEE Trans. Signal Processing*, vol. 51, pp. 2369-2380, Sep. 2003.

[132] B. Yang, Z. Cao, and K.B. Letaief, "Analysis of low-complexity windowed DFT-based MMSE channel estimator for OFDM systems," *IEEE Trans. Communications*, vol. 49, pp. 1977-1987, Nov. 2001.

[133] S. G. Wilson, *Digital Modulation and Coding*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[134] A. O. Hero and T. L. Marzetta, "Cutoff rate and signal design for the quasi-static Rayleigh-fading space-time channel," *IEEE Trans. Information Theory*, vol. 47, pp. 2400-2416, Sep. 2001.

[135] V. Aue, G. P. Fettweis, and R. Valenzuela, "A comparison of the performance of linearly equalized single carrier and coded OFDM over frequency selective fading channels using the random coding technique," *Proc. ICC*, Atlanta, Georgia, USA, June 1998, pp. 753-757.

[136] M. Debbah, "Linear precoders for OFDM wireless communications," Ph.D. dissertation, Ecole Normale Superieure de Cachan, France, 2002.

[137] R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Cut-off rate based transmit optimization for spatial multiplexing on general MIMO channels," *Proc. ICASSP*, Hong Kong, April 2003, vol. 5, pp. 61-64.

[138] Y. Xin and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Communications*, vol. 2, pp. 2465-2476, March 2003.

[139] G. L. Turin, "The characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199-201, 1960.

[140] B. Sklar, *Digital Communications: Fundamentals and Applications*. 2nd. Edition, Upper Saddle River, NJ: Prentice-Hall, 2001.

[141] *Universal Mobile Telecommunications System (UMTS); Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03 Version 3.2.0)*, ETSI Standard, Document TR 101 112 V3.2.0, April 1998.

[142] *Channel Model for HIPERLAN/2 in Different Indoor Scenarios*, ETSI Standard, Document 3ERI085, March 1998.

[143] Z. Wang, S. Zhou, and G. B. Giannakis, "Joint coding-precoding with low-complexity turbo-decoding," *IEEE Trans. Wireless Communications*, vol. 3, pp. 832-842, May 2004.

[144] R. F. H. Fischer and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. GLOBECOM*, London, U.K., Nov. 1996, pp. 724-728.

[145] A. Czylwik, "Adaptive OFDM for wideband radio channels," in *Proc. GLOBECOM*, London, U.K., Nov. 1996, pp. 713-718.

[146] Y. Xue and A. B. Gershman, "Alamouti-type wireless communication system with one-bit feedback," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Sitges, Spain, June 2004.

[147] Y. Sun and M. L. Honig, "Minimum feedback rates for multi-carrier transmission with correlated frequency-selective fading," in *Proc. GLOBECOM*, San Francisco, USA, Dec. 2003, pp. 1628-1632.

[148] A. Kamikura, N. Okada, and M. Nakagawa, "OFDM subcarrier selection for real-time wireless communication," in *Proc. Int. Conf. Commun. Syst.*, Singapore, Nov. 2002, pp. 52-56.

[149] H. Schmidt and K-D. Kammeyer, "Reducing the peak to average power ratio of multicarrier signals by adaptive subcarrier selection," in *Proc. IEEE Int. Conf. Univ. Personal Comm.*, Florence, Italy, Oct. 1998, pp. 933-937.

[150] M.-S. Alouini and A. J. Goldsmith, "A unified approach for calculating error rates of linearly modulated signals over generalized fading channels," *IEEE Trans. Communications*, vol. 47, pp. 1324-1334, Sep. 1999.

[151] P. Ligdas and N. Farvardin, "Optimizing the transmit power for slow fading channels," *IEEE Trans. Information Theory*, vol. 46, pp. 565-576, March 2000.

[152] S. Thoen, L. Van der Perre, B. Gyselinckx, and M. Engels, "Performance analysis of combined transmit-SC/receive-MRC," *IEEE Trans. Communications*, vol. 49, pp. 5-8, Jan. 2001.

[153] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. New York: John Wiley & Sons, 2000.

[154] A. F. Molisch, *Wideband Wireless Digital Communications*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[155] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Signal Processing*, vol. 53, pp. 1697-1710, May 2005.

[156] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Processing*, vol. 53, pp. 2678-2689, Aug. 2005.

[157] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multiaccess MIMO systems with imperfect channel state information," *IEEE Trans. Signal Processing*, vol. 53, pp. 3081-3090, Aug. 2005.

[158] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Linear block precoding for OFDM systems based on maximization of mean cutoff rate," *IEEE Trans. Signal Processing*, vol. 53, Dec. 2005.

[159] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Adaptive OFDM techniques with one-bit-per-subcarrier channel state feedback," *IEEE Trans. Communications*, accepted subject to revisions.

[160] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation using time-varying user power loading and parallel factor analysis," in *Proc. 58th IEEE Vehicular Technology Conference (VTC)*, Orlando, USA, Oct. 4-9, 2003, vol. 1, pp. 79-83.

[161] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Deterministic Cramér-Rao bound for symmetric PARAFAC model with application to blind spatial signature estimation," in *Proc. 3rd IEEE Int. Symposium on Signal Processing and Information Technology (ISSPIT)*, Darmstadt, Germany, Dec. 14-17, 2003, pp. 411-414.

[162] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "A robust linear receiver for multi-access space-time block coded MIMO systems based on probability constrained optimization," in *Proc. 59th IEEE Vehicular Technology Conference (VTC)*, Milan, Italy, May 17-19, 2004, vol. 1, pp. 118-122.

[163] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multiple-access MIMO wireless systems," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Quebec, Canada, May 17-21, 2004, vol. 2, pp. 9-12.

[164] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models based on linear programming," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Quebec, Canada, May 17-21, 2004, vol. 2, pp. 113-116.

[165] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust fitting of multilinear models with application to blind multiuser receivers: Iterative weighted median filtering approach," in *Proc. 5th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Lisbon, Portugal, July 11-14, 2004, pp. 478-482.

[166] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "A robust linear receiver for uplink multi-user MIMO systems based on probability-constrained optimization and second-order cone programming," in *Proc. 3rd IEEE Signal Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, July 18-21, 2004, pp. 153-157.

[167] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Exploiting the structure of OS-TBC's to improve the robustness of worst-case optimization based linear multi-user MIMO receivers," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, USA, March 19-23, 2005, vol. 4, pp. 781-784.

[168] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receiver design for multi-access space-time block coded MIMO systems using stochastic optimization," in *Proc. 13th IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, July 17-20, 2005.

[169] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Linear OFDM precoder design for multiuser wireless communications using cutoff rate optimization," in *Proc. 12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, Sep. 6-10, 2004, pp. 2071-2074.

[170] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Combining error-correction coding and cutoff rate maximization based precoding", in *Proc. Int. ITG/IEEE Workshop on Smart Antennas (WSA)*, Duisburg, Germany, April 4-5, 2005.

[171] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "On average one bit per subcarrier channel state information feedback in OFDM wireless communication systems,"

in *Proc. IEEE Global Telecommunication Conference (GLOBECOM)*, Dallas, Texas, USA, Nov. 29 - Dec. 3, 2004, vol. 6, pp. 4011-4015.

[172] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "The impact of imperfect one bit per subcarrier channel state information feedback on adaptive OFDM wireless communication systems," in *Proc. 60th IEEE Vehicular Technology Conference (VTC)*, Los Angeles, CA, USA, Sep. 26-29, 2004, vol. 1, pp. 626-630.

Wissenschaftlicher Werdegang

Education

- Sep. 1995 - Jul. 1999, Shanghai Jiaotong University, Shanghai, China. Obtained the Bachelor of Engineering degree in Electrical Engineering
- Mar. 1997 - Jul. 1999, Shanghai Jiaotong University, Shanghai, China. Obtained the Bachelor of Engineering degree in Computer Engineering
- Sep. 2000 - Oct. 2002, University of Duisburg-Essen, Duisburg, Germany. Obtained the M.Sc. degree in Computer Science and Communication Engineering
- Nov. 2002 - Mar. 2005, University of Duisburg-Essen, Duisburg, Germany. Ph.D. study in Electrical Engineering
- Apr. 2005 - Nov. 2005, Darmstadt University of Technology, Darmstadt, Germany. Ph.D. study in Electrical Engineering

Research Experience

- Dec. 1999 - Jun. 2000, research associate with the Department of Electrical Machinery at Shanghai Jiaotong University
- Dec. 2002 - Mar. 2005, research associate with the Department of Communication Systems at University of Duisburg-Essen
- Apr. 2005 - present, research associate with the Institute of Telecommunications at Darmstadt University of Technology

Working Experience

- Apr. 2001 - Apr. 2002, Division of Wireless Chips and Systems at Fraunhofer Institute of Microelectronic Circuits and Systems, Duisburg, Germany
- Oct. 2001 - Mar. 2002, Division of Application-Specific Integrated Circuit (ASIC) design at Nokia Ltd., Bochum, Germany