



TECHNISCHE UNIVERSITÄT
ILMENAU

Research Reports from the Communications Research Laboratory
at Ilmenau University of Technology

Massive MIMO and Multi-hop Mobile Communication Systems

Sepideh Gherekhloo

Technische Universität Ilmenau
Fakultät für Elektrotechnik und Informationstechnik
Institut für Informationstechnik
Fachgebiet Nachrichtentechnik



Massive MIMO and Multi-hop Mobile Communication Systems

Sepideh Gherekhloo

geboren am 23.01.1991 in Mashhad (Iran)

Dissertation

zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.)

Anfertigung im: Fachgebiet Nachrichtentechnik
Institut für Informationstechnik
Fakultät für Elektrotechnik und Informationstechnik

Gutachter: Univ.-Prof. Dr.-Ing. Martin Haardt
Prof. André L. F. de Almeida
Prof. Dr.-Ing. Marius Pesavento

Tag der Einrichtung: 01.03.2024
Tag der wissenschaftlichen Aussprache: 13.05.2024

DOI: 10.22032/dbt.61615
URN: urn:nbn:de:gbv:ilm1-2024000236

Acknowledgement

This thesis is a result of my work at the Communication Research Laboratory (CRL), Ilmenau University of Technology. While my name may be alone on the front cover of this thesis, I am by no means its sole contributor. Rather, there are a number of people behind this piece of work who without their gracious support it would not have been possible.

Firstly, I want to express my great gratitude to Prof. Martin Haardt for receiving me at the CRL, discovering my potential, and giving me the opportunity to develop myself further. Moreover, I would sincerely thank him for his continuous support of my scientific work and research, his patience, motivation, enthusiasm, and immense knowledge. I am also grateful to thank Prof. André L. F. de Almeida and Prof. Marius Pesavento for volunteering to review the thesis. In addition, my special thank goes to Prof. André L. F. de Almeida for his cooperation as well as his inspiring suggestions and ideas during this thesis.

I want to give special thanks to my beloved husband, Dr. Khaled Nafez Rauf Ardah, for his invaluable guidance, for countless technical discussions, and for sharing his knowledge and skills. I cannot find the words to describe the beauty and love that you brought to my life. Thanks for making my life so colorful, interesting, and rich in many perspectives. Also, I would like to use this opportunity to express my gratitude to my friends and colleagues at CRL for their support and help throughout the course of this journey. My working time at the CRL will remind me of a very pleasant working atmosphere and many moments that will be remembered. In particular, I would like to thank Dr. Mike Wolf, Dr. Marjan Maleki, Dr. Liana Khamidullina, Alla Manina, Gerald Chetachi Nwalozie, Marko Hennhöfer, Joseph Kibugi Chege, Mikus Grasis, and Damir Rakhimov.

None of what I have achieved would have been possible without the enormous support of my family. I would like to thank my parents Shahin Lari and Alireza Gharakhloo for all they sacrificed in their lives to make a better life for me, for their unconditional love and support in my life. I would like to thank my brother and sister and their families for always being there for me when I need them. I am very fortunate to have you in my life. There are many more families and friends that I was lucky enough to have in my life with bright minds and wonderful personalities. God bless you all.

Abstract

Since the late 1990s, massive multiple-input multiple-output (MIMO) has been suggested to improve the achievable data rate in wireless communication systems [AHK21]. To overcome the high path losses in the high frequency bands, the use of massive MIMO will be a must rather than an option in future wireless communication systems. At the same time, due to the high cost and high energy consumption of the traditional fully digital beamforming architecture, a new beamforming architecture is required. Among the proposed solutions, the hybrid analog digital (HAD) beamforming architecture has received considerable attention. The promising massive MIMO gains heavily rely on the availability of accurate channel state information (CSI). This thesis considers a wideband massive MIMO orthogonal frequency division multiplexing (OFDM) system. We propose a channel estimation method called sequential alternating least squares approximation (SALSA) by exploiting a hidden tensor structure in the uplink measurement matrix considering the HAD architecture at the base station (BS). Then, the channel matrix is estimated sequentially using an alternating least squares (ALS) method. The proposed SALSA algorithm is applicable for any massive MIMO system regardless of the channel characteristics. Detailed simulation results show that SALSA offers flexible control of the complexity-accuracy trade-off compared to the classical least squares (LS) and minimum mean squared error (MMSE) methods. Then, we propose a HAD beamforming with a low-complexity design applicable to any hybrid architecture, where we decouple the designing of the baseband and analog beamforming matrices. We show that the analog beamforming matrix design can be written as a series of convex sub-problems that are updated iteratively until convergence is obtained. Compared to the reference algorithms, our framework either achieves a comparable performance or significantly outperforms them depending on the HAD beamforming architecture. Our proposed HAD beamforming design has a low-complexity and is applicable for any hybrid architecture.

Exploring new less-congested frequency bands, such as millimeter wave (mmWave) frequencies, i.e., 30 - 300 GHz, has been proposed as a promising solution to increase network capacity. However, mmWave and high-frequency bands suffer from a high path loss compared to the sub-6 GHz band, which renders them applicable only for short-range and indoor communication scenarios. Reconfigurable intelligent surfaces (RISs) have been introduced as a low-cost and energy-efficient green solution to improve the communication range and overcome blockage issues in future wireless communication networks. This thesis considers an RIS-aided mmWave MIMO system, where the RIS is constituted of passive reflecting elements. We propose two novel channel estimation techniques for Single RIS (S-RIS)-aided systems. By exploiting the low-rank nature of mmWave channels in the angular domains, we propose a non-iterative Two-stage RIS-aided Channel Estimation (TRICE) framework, where every stage is formulated as a multi-dimensional direction of arrival (DoA) estimation problem. As a result, our TRICE framework is very general in the sense that any efficient multi-dimensional DoA estimation solution can be readily used in every stage to estimate the associated channel parameters. Then, we extend our TRICE framework by exploiting the tensor structure of the received signals, which admit a canonical polyadic decomposition (CPD), which is also known as Parallel Factor (PARAFAC) analysis. This extension is called Tensor-based RIS Channel Estimation (TenRICE), in which the tensor factor matrices are estimated via an ALS method. Numerical simulations evaluate the resulting system

performance and show that both methods require a lower training overhead and a lower computational complexity, compared to the benchmark solutions, while TenRICE has a superior performance approaching the Cramér Rao lower Bound (CRB). After that, we propose a heuristic non-iterative two-step method to design the RIS reflection coefficients, where the RIS reflection vector is obtained in a closed-form using the Frobenius-Norm Maximization (FroMax) strategy. Our simulation results show that FroMax achieves a comparable performance to benchmark methods with a lower complexity.

Since the performance of the S-RIS-aided systems depends on the distance of the RIS to the transmitter and the receiver, multi-RIS-aided systems have attracted more attention in the recent years. This thesis considers a flat-fading Double RIS (D-RIS)-aided MIMO system and proposes channel estimation techniques when (i) both RISs have the same training overhead and (ii) RISs have different training overheads. For case (i), we propose an ALS-based channel estimation method, called channel estimation for joint training (CEJOINT), by exploiting the Tucker2 tensor structure of the received signals. We show that if the reflective elements of the S-RIS system are carefully distributed between the two RISs in a D-RIS system, the training overhead in the D-RIS system can be reduced and the estimation accuracy can also be increased compared to the S-RIS system. For case (ii), we show that the received signal can be represented as a 4-way tensor satisfying a nested PARAFAC decomposition model. Exploiting such a structure, we propose a closed-form Khatri-Rao Factorization (KRF)-based and an iterative ALS-based channel estimation method, which are called KRF-based for separate training (KRF-SEPT) and ALS-based for separate training (ALS-SEPT), respectively. The numerical results show that both proposed methods have a comparable performance as long as the identifiability conditions of the KRF-SEPT are satisfied, while the ALS-SEPT method can achieve a satisfactory performance with less training overhead. Moreover, the performance of the proposed ALS-SEPT method can further be improved by using KRF-SEPT as an initialization.

Zusammenfassung

Seit den späten 1990er Jahren wurden massive MIMO (multiple-input multiple-output) Systeme vorgeschlagen, um die erreichbaren Datenraten in drahtlosen Kommunikationssystemen zu verbessern [AHK21]. Um die hohen Pfadverluste in den Hochfrequenzbändern zu überwinden, wird die Verwendung von massive MIMO in zukünftigen drahtlosen Kommunikationssystemen eher ein Muss als eine Option sein. Gleichzeitig ist aufgrund der hohen Kosten und des hohen Energieverbrauchs der herkömmlichen volldigitalen Beamforming-Architektur eine neue Beamforming-Architektur erforderlich. Unter den vorgeschlagenen Lösungen hat eine hybride analog digitale (HAD) Beamforming-Architektur erhebliche Aufmerksamkeit erhalten. Die vielversprechenden Gewinne durch den Einsatz von massive MIMO hängen stark von der Verfügbarkeit von genauer Kanalzustandsinformation (CSI) ab. Diese Arbeit betrachtet ein breitbandiges OFDM-System mit massive MIMO. Wir schlagen eine Kanalschätzmethode namens SALSA (sequential alternating least squares approximation) vor, die eine versteckte Tensorstruktur in der gemessenen Uplink-Matrix unter Berücksichtigung der HAD-Architektur an der Basisstation (BS) ausnutzt. Anschließend wird die Kanalmatrix sequentiell mit einer ALS (alternating least squares)-Methode geschätzt. Der vorgeschlagene SALSA-Algorithmus ist unabhängig von den Kanaleigenschaften auf jedes massive MIMO-System anwendbar. Detaillierte Simulationsergebnisse zeigen, dass SALSA im Vergleich zu den klassischen LS (least squares)- und MMSE (minimum mean squared error)-Verfahren einen flexiblen Abtausch zwischen Komplexität und Genauigkeit aufweist. Anschließend schlagen wir einen Entwurf für die Beamforming-Matrix mit geringer Komplexität vor, der auf jede hybride Architektur anwendbar ist und bei dem wir den Entwurf der Basisband- und der analogen Beamforming-Matrizen entkoppeln. Wir zeigen, dass das Design der analogen Beamforming-Matrix als eine Reihe von konvexen Teilproblemen geschrieben werden kann, die iterativ aktualisiert werden, bis Konvergenz erreicht ist. Im Vergleich zu den Referenzalgorithmen erreicht unser Algorithmus entweder eine vergleichbare Leistung oder übertrifft sie erheblich, je nach der HAD Beamforming Architektur. Der von uns vorgeschlagene Entwurf für die Beamforming-Architektur weist eine geringe Komplexität auf und ist auf jede Hybridarchitektur anwendbar.

Die Erkundung neuer, weniger überlasteter Frequenzbänder, wie der Frequenzen im Millimeterwellenbereich (mmWave), d. h. 30 - 300 GHz, wurde als vielversprechende Lösung zur Erhöhung der Netzkapazität vorgeschlagen. Diese Frequenzbänder weisen jedoch im Vergleich zum Sub-6-GHz-Band hohe Pfadverluste auf, so dass sie nur für Kurzstrecken- und Innenraumkommunikationsszenarien geeignet sind. Rekonfigurierbare intelligente Oberflächen (RISs - Reconfigurable intelligent surfaces) wurde als kostengünstige und umweltfreundliche Lösung zur Verbesserung der Kommunikationsreichweite und zur Überwindung von Blockierungsproblemen in zukünftigen drahtlosen Kommunikationsnetzwerken eingeführt. Diese Arbeit betrachtet ein RIS-basiertes mmWave MIMO-System, bei dem die RIS aus passiven reflektierenden Elementen besteht. Wir schlagen zwei neue Kanalschätzmethoden für Single RIS (S-RIS)-basierte Systeme vor. Unter Ausnutzung der geringen Rangstruktur von mmWave-Kanälen in der räumlichen Auflösung schlagen wir einen nichtiterativen TRICE (Two-stage RIS-aided Channel Estimation)-Algorithmus vor, bei dem jede Stufe als mehrdimensionales Richtungs

(DoA)-Schätzproblem formuliert wird. Infolgedessen ist unser TRICE-Algorithmus sehr allgemein in dem Sinne, dass jede effiziente mehrdimensionale DoA-Schätzlösung ohne weiteres in jeder Stufe zur Schätzung der zugehörigen Kanalparameter verwendet werden kann. Wir erweitern dann unseren TRICE-Algorithmus, indem wir die Tensorstruktur der empfangenen Signale ausnutzen, die eine CPD (canonical polyadic decomposition)-Analyse, auch bekannt als PARAFAC (Parallel Factor)-Analyse, ermöglicht. Diese Erweiterung wird als TenRICE (Tensor-based RIS Channel Estimation) bezeichnet, wobei die Tensorfaktormatrizen mit einer ALS-Methode geschätzt werden. Numerische Simulationen bewerten die resultierende Systemleistung und zeigen, dass beide Methoden im Vergleich zu den Benchmark-Lösungen weniger Trainingsaufwand und Rechenkomplexität erfordern, während TenRICE eine überlegene Leistung aufweist, die der Cramér Rao Schranke CRB nahekommt. Anschließend schlagen wir ein heuristisches, nicht-iteratives zweistufiges Verfahren zum Entwurf der RIS-Reflexionskoeffizienten vor, bei dem der RIS-Reflexionsvektor in geschlossener Form unter Verwendung der FroMax (Frobenius-Norm Maximization)-Strategie ermittelt wird. Unsere Simulationsergebnisse zeigen, dass FroMax eine vergleichbare Leistung wie Benchmark-Methoden mit geringerer Komplexität erreicht.

Da die Leistung von S-RIS-gestützten Systemen von der Entfernung der RIS zum Sender und Empfänger abhängt, haben Systeme mit mehreren RISs in den letzten Jahren mehr Aufmerksamkeit erhalten. In dieser Arbeit wird ein Double RIS (D-RIS)-basiertes MIMO-System mit flachem Fading betrachtet und es werden Methoden zur Kanalschätzung vorgeschlagen, wenn (i) beide RISs den gleichen Trainingsaufwand haben und (ii) die RISs unterschiedliche Trainingsaufwände haben. Für den Fall (i) schlagen wir eine ALS-basierte Kanalschätzungsmethode, genannt CEJOINT (channel estimation for joint training), vor, die die Tucker2-Tensorstruktur der empfangenen Signale nutzt. Wir zeigen, dass, wenn die reflektierenden Elemente des S-RIS-Systems sorgfältig zwischen den beiden RISs in einem D-RIS-System verteilt werden, der Trainingsaufwand im D-RIS-System reduziert werden kann und die Schätzgenauigkeit im Vergleich zum S-RIS-System ebenfalls erhöht werden kann. Für den Fall (ii) zeigen wir, dass das empfangene Signal als ein 4-dimensionaler-Tensor dargestellt werden kann, der ein nested PARAFAC-Zerlegungsmodell erfüllt. Unter Ausnutzung einer solchen Struktur schlagen wir eine geschlossene Form für eine KRF (Khatri-Rao Factorization)-basierte und eine iterative ALS-basierte Kanalschätzungsmethode vor, die als KRF-SEPT (KRF-based for separate training) bzw. ALS-SEPT SEPT (ALS-based for separate training) bezeichnet werden. Die numerischen Ergebnisse zeigen, dass beide vorgeschlagenen Methoden eine vergleichbare Leistung haben, solange die Identifizierbarkeitsbedingungen der KRF-SEPT-Methode erfüllt sind, während die ALS-SEPT-Methode eine zufriedenstellende Leistung mit weniger Trainingsaufwand erreichen kann. Darüber hinaus kann die Leistung der vorgeschlagenen ALS-SEPT-Methode durch die Verwendung von KRF-SEPT als Initialisierung weiter verbessert werden.

Contents

Acknowledgements	i
Abstract	iii
Zusammenfassung	vi
List of Figures	xiv
List of Tables	xvii
List of Algorithms	xix
Acronyms	xxiv
1 Introduction	1
1.1 Motivation	1
1.1.1 Part I: Massive MIMO Wireless Communication Systems	1
1.1.2 Part II: RIS-aided Wireless Communication Systems	5
1.2 Major Contributions	8
1.2.1 Part I: Massive MIMO Wireless Communication Systems	8
1.2.2 Part II: RIS-aided Wireless Communication Systems	9
1.2.3 Scientific Production	12
1.3 Preliminary Definitions and Concepts	13
1.3.1 Notation	13
1.3.2 Matrix Properties	13
1.3.3 Tensor Algebra	15
1.3.3.1 Definitions and Properties	15
1.3.3.2 Tensor Decompositions	17
I Massive MIMO Wireless Communication Systems	19
2 Channel Estimation	21
2.1 Introduction and State of the Art	21
2.1.1 Chapter Contributions	22
2.1.2 Chapter Organization	23
2.2 System Configuration	23
2.2.1 System Model	23
2.2.2 Channel Model	23
2.2.3 Signal Model	24
2.3 Algorithms	26

2.3.1	The Baseline LS-based Channel Estimation Algorithm	26
2.3.2	The Baseline LMMSE-based Channel Estimation Algorithm	26
2.3.3	The Proposed SALSA Channel Estimation Algorithm	27
2.4	Simulation Results	30
2.5	Chapter Conclusion	42
3	Hybrid Beamforming Design	43
3.1	Introduction and State of the Art	43
3.1.1	Chapter Contributions	45
3.1.2	Chapter Organization	45
3.2	System Model	45
3.3	Beamforming Design	46
3.3.1	Baseband Beamforming Design	47
3.3.2	Analog Beamforming Design	48
3.3.2.1	Baseline Method	49
3.3.2.2	Proposed Analog Beamforming Design	50
3.4	Simulation Results	52
3.5	Chapter Conclusion	55
II	RIS-Aided Wireless Communication Systems	57
4	Single-RIS-Aided Systems	59
4.1	Introduction and State of the Art	59
4.1.1	Fundamental Performance Metrics	60
4.1.2	Channel Estimation	60
4.1.3	RIS Beamforming Design	62
4.1.4	Chapter Contribution	64
4.1.5	Chapter Organization	64
4.2	System Configuration	64
4.2.1	System Model	64
4.2.2	Channel Model	67
4.3	Algorithms for Channel Estimation Phase	69
4.3.1	Baseline Channel Estimation Algorithm	69
4.3.2	Proposed TRICE Algorithm	69
4.3.3	Proposed TenRICE Algorithm	72
4.4	Algorithms for Data Transmission Phase	75
4.4.1	Beamforming Design	75
4.4.2	Reflection Matrix Design	77
4.4.2.1	FroMax-1	77
4.4.2.2	FroMax-2	77
4.5	Discussion	79
4.5.1	Uniqueness and Identifiability Conditions	79
4.5.2	Computational Complexity	80

4.6	Simulation Results	81
4.6.1	Channel Estimation Phase	82
4.6.2	Data Transmission Phase	85
4.7	Chapter Conclusion	87
5	Double RIS-Aided Systems	89
5.1	Introduction and State of the Art	89
5.1.1	Chapter Contribution	90
5.1.2	Chapter Organization	90
5.2	System Model	90
5.3	Algorithms	92
5.3.1	Joint Channel Training of RISs	92
5.3.2	Separate Channel Training of RISs	95
5.3.2.1	KRF-SEPT Algorithm	96
5.3.2.2	ALS-SEPT Algorithm	97
5.4	Discussion	98
5.4.1	Comparison with S-RIS-aided systems	98
5.4.2	Uniqueness and Identifiability Conditions	99
5.4.3	Computational Complexity	99
5.4.4	Scaling Ambiguity	100
5.5	Simulation Results	101
5.5.1	D-RIS-aided System versus S-RIS-aided System	102
5.5.2	D-RIS-aided System with Separable Training Overheads	104
5.6	Chapter Conclusion	106
6	Conclusions and Future Works	109
6.1	Conclusions	109
6.2	Future Works	111
A	Appendices	115
A.1	Proof of (1.26)	115
A.2	Jensen’s Inequality	117
A.3	Design of Analog Decoding Matrix in Chapter 3	119
A.4	Derivation of Upper-bound in (3.18)	121
A.5	Proof of Proposition 3.3.1	122
A.6	Derivation of CRB to 4-way CP Tensor in (4.50)	123
A.7	Nested PARAFAC Models	127
A.8	Symbols and Notations	128

List of Figures

1.2	3GPP release timeline [3GP16a; 3GP16b; 3GP20b; 22; 21b; 21a].	4
1.3	Atmospheric absorption variation due to several factors [Tri+21].	6
1.4	RIS applications and use cases in wireless communication systems [Hon23].	7
1.5	Structure of the thesis.	9
1.6	Graphical representation of the forward unfolding of a 3-way tensor.	16
2.1	Uplink single-user massive MIMO OFDM communication system.	24
2.2	The transmission time interval (TTI) division for uplink transmission.	25
2.3	MSE vs. the sum of Kronecker products R	33
2.4	Average singular values of Case 1, Case 2, and Case 3 channel matrices.	34
2.5	Convergence of the SALS method in Algorithm 2.2.	35
2.6	NMSE vs. SNR for different Q and P division scenarios, with $T_{BS} = 12$	37
2.7	NMSE vs. SNR for different Q and P division scenarios, with $T_{BS} = 16$	38
2.8	Decision making procedure in SALS to select the dimensions of the factor matrices B_r and C_r depending on the training beamforming matrix A	39
2.9	NMSE vs. SNR for different Q and P division scenarios and antenna array structures, with $T_{BS} = 12$	40
2.10	NMSE vs. SNR for different R and β	40
2.11	NMSE vs. SNR for different M_{BS}	41
3.1	The FC and the PC HAD architectures at the BS.	46
3.2	Convergence of Algorithm 3.3.	53
3.3	System SE versus SNR for $M_{BS} = 64$, $M_{UE} = 8$, $U = 4$, $N_{UE} = 2$, $N_{BS} = 8$, and $K = 16$	54
3.4	System SE versus number of antennas at the BS M_{BS} for $M_{UE} = 8$, $U = 4$, $N_{UE} = 2$, $N_{BS} = 8$, and $K = 16$	55
3.5	System SE versus number of sub-carriers K for $M_{BS} = 64$, $M_{UE} = 8$, $U = 4$, $N_{UE} = 2$, and $N_{BS} = 8$	56
3.6	System SE versus number of groups G for $M_{BS} = 64$, $M_{UE} = 8$, $U = 4$, $N_{UE} = 2$, $N_{BS} = 8$, and $K = 16$	56
4.1	An RIS-aided mmWave MIMO communication system.	65
4.2	One channel coherence block.	65
4.3	Definition of azimuth ($\theta_{T,l}^h$) and co-elevation ($\theta_{T,l}^v$) angles of an impinging planar wavefront on an RIS with URA structure.	68
4.4	Compare proposed methods in terms of NMSE vs. SNR.	81
4.5	Compare proposed methods in terms of MSE vs. SNR.	83
4.6	Effect of grid resolution on the channel estimation accuracy.	84
4.7	Effect of the number of multi path on the channel estimation accuracy.	84
4.8	Convergence of the TenRICE method in Algorithm 4.5.	85
4.9	SE vs. SNR.	86

4.10	Dominant singular values of a perfect effective channel $\mathbf{H}_e = \mathbf{H}_R \text{diag}\{\boldsymbol{\omega}\} \mathbf{H}_T$	86
5.1	A D-RIS aided MIMO Communication System.	91
5.2	Transmission blocks for joint channel training of RISs.	92
5.3	Graphical 3D representation of the 4-way tensor \mathcal{Y}	95
5.4	Transmission blocks for separate channel training of RISs.	96
5.5	Number of channel coefficients ($\ell_{X\text{-RIS}}$) with $X \in \{\text{S}, \text{D}\}$ with $M_S = 40$ and $M_{S2} = M_S - M_{S1}$	102
5.6	Minimum training overhead (S and Q) with $M_S = 40$ and $M_{S2} = M_S - M_{S1}$	103
5.7	NMSE versus SNR comparing D-RIS against the S-RIS systems.	104
5.8	NMSE versus SNR for different training overhead ($M_{S1} = 30$ and $M_{S1} = 20$).	105
5.9	Effect of initialization on ALS-SEPT for $I = 15, J = 20$ ($M_{S1} = 30$ and $M_{S1} = 20$).	105
5.10	Effect of initialization on ALS-SEPT for $I = 15, J = 25$ ($M_{S1} = 30$ and $M_{S1} = 20$).	106
A.1	Using Jensen's inequality for a secant line of a convex function	117

List of Tables

2.1	System Parameters to generate CDL channel tensor $\mathcal{H} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}} \times N_{\text{taps}}}$	30
2.2	Dimension of the factor matrices in the Kronecker product Q_1, Q_2, P_1 , and P_2 for $Q = Q_1 Q_2$ and $P = P_1 P_2$	36
4.1	Training overhead analysis	80
4.2	Computational complexity analysis	81
5.1	Identifiability condition of algorithms.	100
5.2	Order of the computational Complexity of algorithms.	101

List of Algorithms

2.1	Sequential Kronecker Factorization	28
2.2	SALSA For MIMO-OFDM Channel Estimation.	29
3.3	Proposed Analog Beamforming Design	51
4.4	Two-Stage RIS-Aided MIMO Channel Estimation (TRICE)	72
4.5	Tensor-based RIS-aided CE (TenRICE)	75
4.6	FroMax-based methods for RIS reflection design.	79
5.7	Channel estimation for joint training (CEJOINT).	94
5.8	Khatri-Rao Factorization (KRF).	97
5.9	KRF-based for separate training (KRF-SEPT).....	97
5.10	ALS-based for separate training (ALS-SEPT).....	98
5.11	ALS method for channel estimation in S-RIS MIMO systems [AAB21].	99
1.12	Proposed algorithm in [Ard+18].....	120

Acronyms

1G first generation

3GPP 3rd Generation Partnership Project

ADC analog-to-digital converter

AF amplify-and-forward

ALS alternating least squares

ALS-SEPT ALS-based for separate training

AltMax alternating maximization

AR augmented reality

AWGN Additive White Gaussian Noise

BD block diagonalization

BPSK binary phase shift keying

BS base station

CDL clustered delay line

CP CANDECOMP/PARAFAC

CPD canonical polyadic decomposition

CRB Cramér Rao lower Bound

CS Compressed Sensing

CSI channel state information

D-RIS Double RIS

DAC digital-to-analog converter

DET dominant eigenmode transmission

DF decode-and-forward

Acronyms

DFT Discrete Fourier Transform

DoA direction of arrival

DoD direction of departure

ESPRIT Estimation of Signal Parameters Via Rotational Invariance Techniques

EX Exabyte

FC fully connected

FD full-duplex

FFT Fast Fourier Transform

FIM Fisher information matrix

FLOPS Floating-Point Operations Per Second

FroMax Frobenius-Norm Maximization

HAD hybrid analog digital

HD half-duplex

HetNets heterogeneous networks

ICI inter carrier interference

IEEE Institute of Electrical and Electronics Engineers

IoT Internet of Things

IRS intelligent reflecting surface

ISI inter symbol interference

KKT Karush-Kuhn-Tucker

KRF Khatri-Rao Factorization

KRF-SEPT KRF-based for separate training

LIS large intelligent surface

LMMSE linear minimum mean squared error

LoS Line of Sight

LS least squares

LSKRF LS Khatri-Rao Factorization

LTE Long Term Evolution

MEMS micro electromechanical systems

MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MMSE	minimum mean squared error
mmWave	millimeter wave
MRT	maximum ratio transmission
MSE	mean squared error
NMSE	normalized mean squared error
NOMP	Newtonized OMP
NR	New Radio
OFDM	orthogonal frequency division multiplexing
OMP	orthogonal matching pursuit
PARAFAC	Parallel Factor
PBIT	passive beamforming and information transfer
PC	partially connected
PS	phase shifter
QAM	quadrature amplitude modulation
RF	radio frequency
RFID	radio frequency identification
RIS	reconfigurable intelligent surface
RSs	reference signals
RX	receiver
S-RIS	Single RIS
SALSA	sequential alternating least squares approximation
SCM	software-controlled metasurface
SDR	semidefinite relaxation
SE	spectral efficiency
SIC	successive interference cancellation
SIMO	single-input multiple-output
SINR	signal-leakage-plus-noise-ratio

Acronyms

SISO single-input single-output

SNR signal to noise ratio

SRSs sounding reference signals

SVD singular value decomposition

TDD time division duplexing

TenRICE Tensor-based RIS Channel Estimation

CEJOINT channel estimation for joint training

THz Terahertz

TRICE Two-stage RIS-aided Channel Estimation

TTI transmission time interval

TX transmitter

UDN ultra-dense network

UE user equipment

ULA uniform linear array

URA uniform rectangular array

VR virtual reality

ZF zero forcing

Introduction

1.1. Motivation

Guglielmo Marconi's successful attempt in 1895 to transmit a three-dot Morse code over a distance of three kilometers using electromagnetic waves brought the first glimmer of hope to the idea of wireless communications [SSS09]. The evolution of wireless communication systems started using the cellular infrastructure from the first generation (1G) operating with analog systems in the 1980s. Since the introduction of 2G in the 1990s, wireless communication systems have been operating fully digitally.

Every decade, a new generation of wireless communication systems emerges to enable higher achievable data rates with lower latency through the use of higher frequencies, larger bandwidths, and multiple antennas. According to the studies [Akh+20; Cho+20], an explosive increase in mobile data traffic is expected in the next decades due to new applications such as augmented reality (AR), virtual reality (VR), and Internet of Things (IoT) [Shi+18] (from 7.462 Exabyte (EX)/month in 2010 to 5016 EX/month in 2030 [Cho+20]). These new applications and their new service requirements cannot be supported by the current 4G wireless communication systems and will put significant pressure on the existing systems. Therefore, the researchers in both academia and industrial communities are actively proposing new technologies and solutions to meet the requirements of current and new applications and use cases. The next generations of wireless communication network, i.e., 5G and beyond, should provide some key performance indicators such as supporting an order of magnitude increase in the peak data rate with more area spectral efficiency (SE) and energy-efficient networks [And+14]. An overview of the development of wireless communication systems from 1G to 6G with the timeline in terms of applications, network characteristics, technology, and key indicator factors is given in Fig. 1.1 [Akh+20].

In the following, we give an overview of some key enabling technologies that are the main focus of this thesis and have been proposed for 5G and beyond wireless communication networks to meet the aforementioned performance requirements.

1.1.1. Part I: Massive MIMO Wireless Communication Systems

The multiple-input multiple-output (MIMO) technology is a key element of current 4G and future wireless communication standards such as IEEE 802.11n, Long Term Evolution (LTE), and 5G new radio (NR). By taking advantage of the rich scattering environments, i.e., multipath propagation phenomena, MIMO technology is able to provide higher data rates and more robust and reliable communications. Assuming a single-user MIMO system, as its simplest implementation, two different types of transmission modes can be realized, i.e., (i) transmit/spatial diversity, where multiple antennas transmit the same data simultaneously to boost the signal to noise ratio (SNR) and communication reliability and (ii) spatial multiplexing, where independent data streams, which are also called layers in the 3rd Generation Partnership Project (3GPP) specifications, are sent on each antenna to increase the capacity and data throughput [Guo+17; Tel00; Gol+03; ZT03; Ala98; TJC99; Ges+03; Pau+04]. While both techniques

can be used together in MIMO systems to achieve a balance of reliability and high data rates, spatial multiplexing is one of the best known MIMO methods. The same approach is applied for the multi-user MIMO setting by splitting the layers among the users. From the introduction of spatial multiplexing in 3GPP Release 7 up to Release 12, the same antenna array structure is assumed, where the antennas are built in columns and each column is controlled separately for spatial multiplexing in the azimuth direction. Release 13 brought a paradigm shift in the antenna arrays, allowing networks to provide larger and more direct arrays, referred to as full-dimension MIMO by controlling antenna elements in both azimuth and elevation for 3D beamforming [DH23]. Release 13's achievements in separating physical antennas from logical elements has paved the way to larger arrays, also known as massive MIMO technology, which increases the number of antennas in the existing MIMO systems by orders of magnitude [Lar+14]. By increasing the number of antenna elements, e.g., to a few hundreds, we can transmit a more significant number of data streams and serve many dozens of terminals in the same time-frequency resource, which basically offers all the benefits of traditional MIMO but on a much larger scale. In Fig. 1.2, you can find the 3GPP timeline and some of the features [3GP16a; 3GP16b; 3GP20b; 22; 21b; 21a].

The almost worldwide deployed 4G wireless communication networks operate mainly in the sub-6 GHz frequency band. Since most of the sub-6 GHz frequency bands are already occupied and 4G has reached its capacity limits due to the spectrum scarcity, a switch to higher frequency bands, i.e., millimeter wave (mmWave) and Terahertz (THz), is widely considered in 5G and future wireless networks [Rap+13]. This can be explained by the large amount of unused spectrum, which enables extremely high data rates and thus a significant increase in capacity. However, during the last few years, several studies have been published about channel measurements at high-frequency bands [Mac+15; Han+15b; Rub+19; Rub+23; Dup+20] showing that their propagation channel suffers from a higher free space path loss compared to the sub-6 GHz band [Dup+18; Dup+19a; Rub+23]. In addition, as shown in Fig. 1.3, some high-frequency bands experience more attenuation due to atmospheric absorption, rain, and snow effects [Tri+21]. These issues make higher-frequency bands applicable only for short-range or indoor communications. In other words, the high-frequency bands are only realizable in outdoor communication applications by either increasing the transmission power or by using high-gain, high-directional antennas. While an increase in transmit power is limited and must be within the range established by regulations, the high beamforming gain in large array systems can combat the path losses and provide links with adequate SNR. Therefore, using massive MIMO systems becomes a must rather than an option in the future wireless generation. Thanks to the small wavelength of high-frequency bands, packing a large number of antennas into small form factors is realizable and practically possible, which suits the requirements of massive MIMO in 5G and beyond.

However, both high-frequency bands and massive MIMO technologies make signal processing different and more complicated from conventional wireless communication systems [Gao+19]. First, new hardware constraints are required to enable a practical implementation of massive MIMO systems such as power consumption and circuit technology [WDV23; Don+16]. Traditionally, the amplitude and phase of signals are adjusted using a digital precoder in the baseband of conventional MIMO systems. This kind of beamforming architecture is known as fully digital, in which a dedicated radio frequency (RF) chain including a signal mixture and an analog-to-digital converter (ADC) are required for each antenna element. Due to their high cost and power consumption, implementing a fully digital precoding architecture in a high-frequency band-based massive MIMO system is currently very challenging, if

1. Introduction

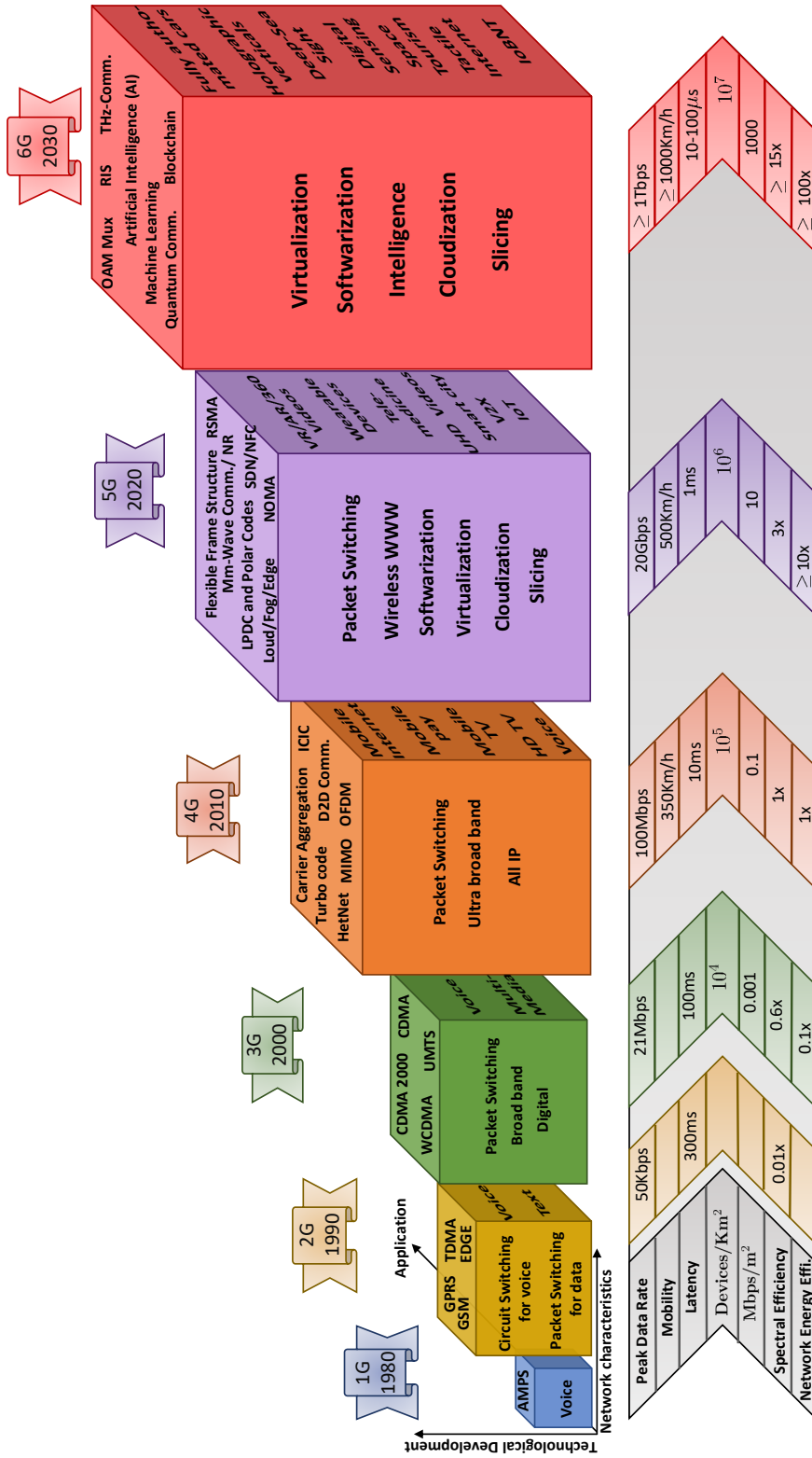


Figure. 1.1: Development of wireless communication systems with the timeline 1G to 6G [Akh+20].

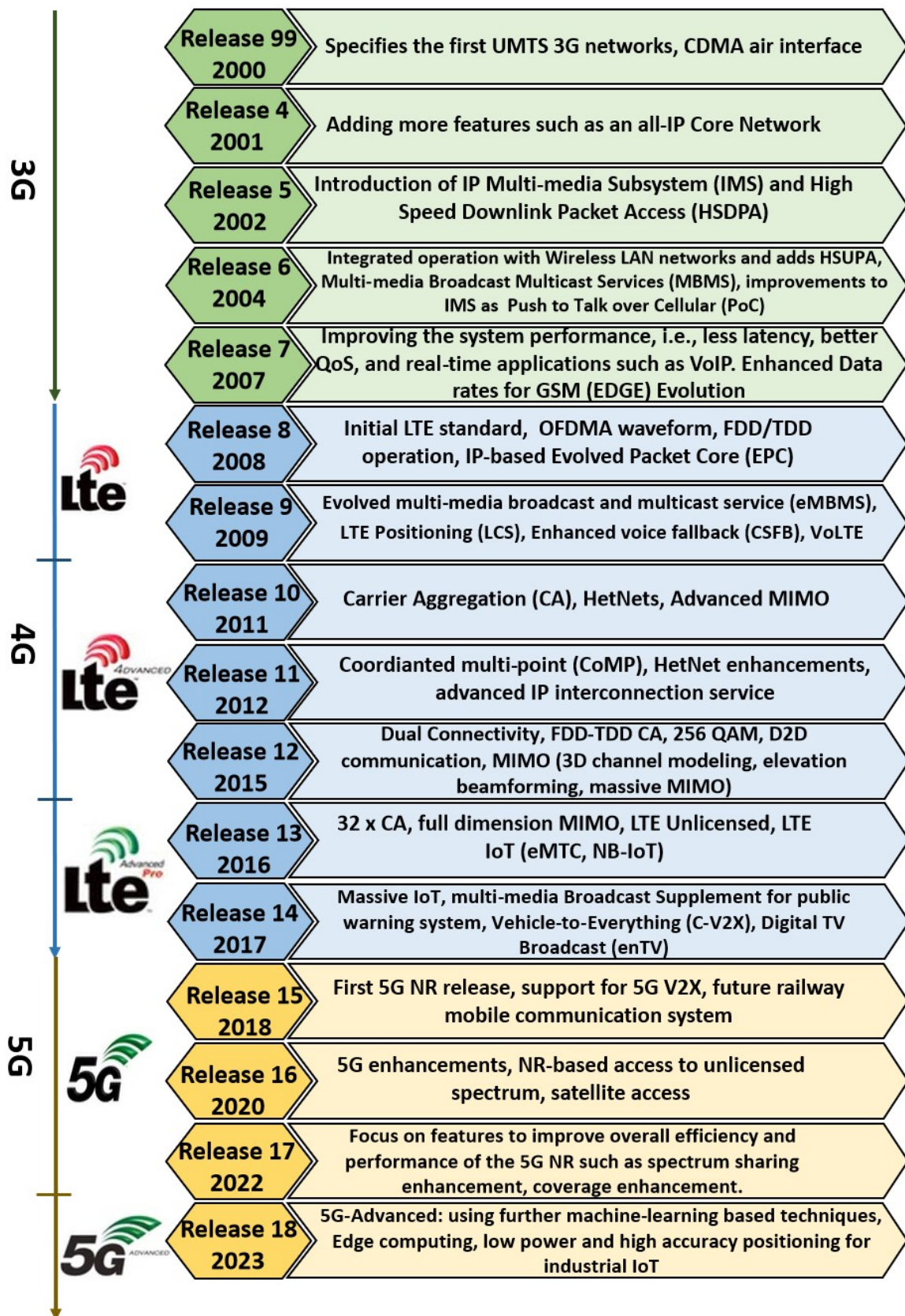


Figure 1.2: 3GPP release timeline [3GP16a; 3GP16b; 3GP20b; 22; 21b; 21a].

not impossible. Therefore, other kinds of beamforming architectures are proposed for massive MIMO systems such as analog and hybrid analog digital (HAD) beamforming architectures. On the one hand, analog beamforming is a very simple and cost-effective architecture, where the digital output of the RF chain is mapped to a large number of antennas via a network of phase shifters (PSs) or switches [Wan+09; BBS13; Hur+13; Son+15]. However, the main idea of MIMO systems, i.e., spatial multiplexing, cannot be realized with analog beamforming alone. To overcome this issue, HAD beamforming architectures have been proposed, which split the signal processing operations between the analog and the digital domains to reduce the number of RF chains as compared to the number of antenna elements [GAH20; Aya+14; Ard+18; HMP19]. Compared to the fully digital counterparts, the HAD beamforming architectures have received a lot of attention recently [Alk+14; LW20; XXX17; MH14; BBS13] and are widely considered for future wireless communication systems, since they provide a better compromise between cost, size, complexity, energy-consumption, and SE.

To exploit the potential of massive MIMO systems, the knowledge of propagation channel characteristics, i.e., channel state information (CSI) at the transmitter and the receiver is critical for wireless communication system design, simulation, and modeling. Since the multipath propagation results in rapid changes in the phase and amplitude of the transmitted signal, the receiver needs to know the accurate CSI in order to mitigate their impact and decode the received signal successfully [Lar+14]. On the other hand, the transmitter needs to know the accurate CSI in order to efficiently design the transmit precoding matrices so that the received signals add constructively at the desired receivers and destructively at the unwanted receivers. Therefore, there is a need to develop new channel estimation algorithms that provide accurate channel estimation with low training overhead and with a low SNR, while considering (i) the different channel models in the sub-6 GHz and high-frequency bands, i.e., high-rank versus low-rank channels and (ii) the corresponding HAD beamforming architecture. Moreover, HAD beamforming problems are inherently non-convex and difficult to design in contrast to their fully digital counterparts, mainly due to some additional non-convex constraints, e.g., constant modulus constraints imposed on the analog domain part of the designed beamforming matrices. **In Part I of this thesis, we address these problems and propose new and efficient channel estimation and HAD beamforming design methods for massive MIMO systems, while taking into account the above design and practical constraints.**

1.1.2. Part II: RIS-aided Wireless Communication Systems

Achieving high data rates while maintaining the system's energy efficiency remains a major concern in future 6G wireless communication networks [WZ19]. An ultra-dense network (UDN) with dense small cells has been proposed to boost the achievable data rate [Boc+14]. However, because of (i) the energy consumption of the circuits and cooling of the base station (BS)s and (ii) the increasing number of the active components such as BSs and relays, the energy consumption scales up almost linearly with the number of active components in the network [Zha+17]. Cognitive radio [Lav+17] is also becoming increasingly important in future wireless communication networks, where a flexible spectrum allocation is proposed to further improve the resource management.

On the other hand, the propagation problems in the high-frequency bands, i.e., short communication coverage due to high path loss and blockage, remain a challenge for providing reliable outdoor wireless communications. Therefore, an energy-efficient technology is needed that improves the communication coverage and overcomes the propagation blockages in the high-frequency bands. Numer-

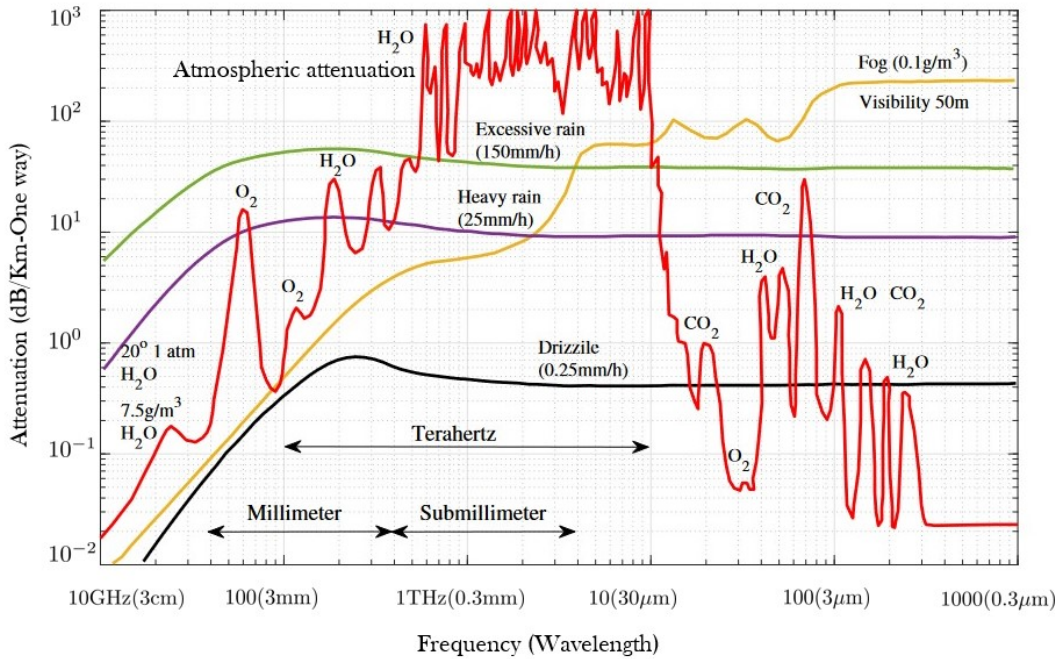


Figure 1.3: Atmospheric absorption variation due to several factors [Tri+21].

ous techniques already exist to overcome the above issues, such as amplify-and-forward (AF) relays [GAH21], backscatter communications, and massive MIMO [Lia+19; Bas+19; WZ19; Yan+20b]. However, these techniques have some disadvantages. For example, the AF relay requires a transmitter module to amplify and regenerate the signal. Moreover, an AF relay usually operates in half-duplex (HD) mode, since the full-duplex (FD) mode requires additional effective interference cancellation techniques to eliminate the strong self-interference. On the other hand, backscatter communications are based on communications between a radio frequency identification (RFID) tag and a reader, where the RFID tag reflects the signal sent by the reader. Particularly, the RFID sends its own information back to the reader. Obviously, the backscatter communication technique cannot be used to improve the communication link. Therefore, other techniques are required for wireless communication systems not only to further improve the data rate but also to keep the energy and hardware costs low.

During the last few years, the reconfigurable intelligent surface (RIS) technology has attracted much attention as one of the key enabling technologies for improving future wireless communication systems. An RIS, also known as an intelligent reflecting surface (IRS), software-controlled metasurface (SCM), and large intelligent surface (LIS), is a man-made surface from electromagnetic materials that can be controlled electronically by adjusting their phase, amplitude, and/or polarization to influence and to tune the wireless communication channels in real-time. There are different implementations of RISs such as liquid crystal surfaces, reflected arrays, and software-defined meta surfaces [The+22; Lia+18b; Lav+18; Liu+18; Hon23]. The peculiarity of the RIS technology, which distinguish it from other technologies used in the current 4G and 5G wireless communication networks design principles, is the ability to create a controllable and smart environment by shaping and fully controlling the elec-

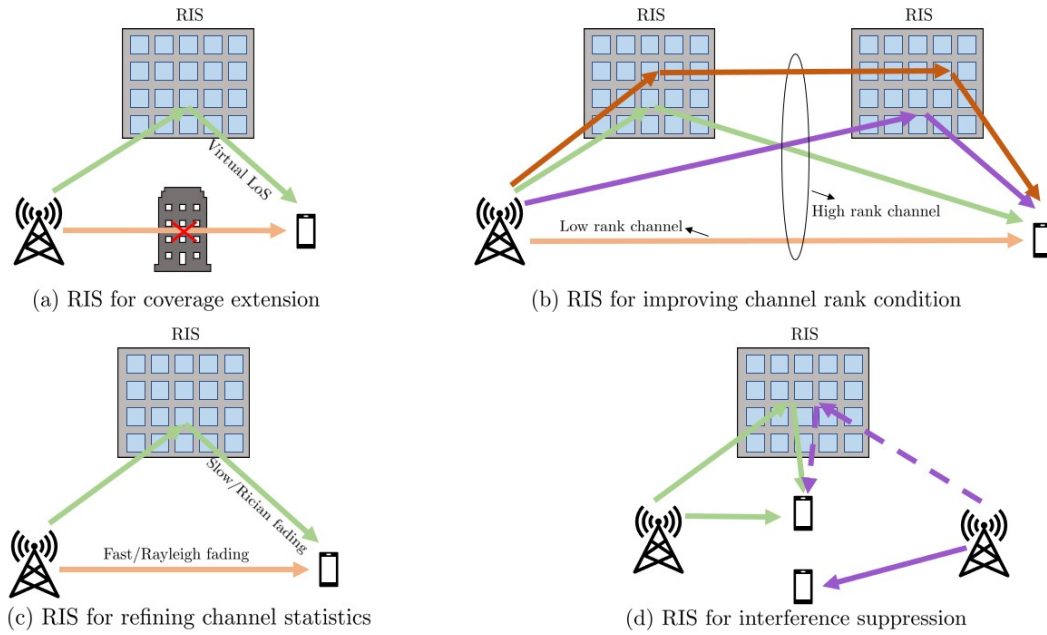


Figure 1.4: RIS applications and use cases in wireless communication systems [Hon23].

tromagnetic response of environmental objects distributed in the network [Lia+18a]. Moreover, most of the RIS implementation architectures are made of passive, inexpensive, and thin electromagnetic materials, which make them easy to deploy on environment structures such as building facades and indoor walls. Due to their unique characteristics, the RIS technology has found a wide range of applications and use cases [Hon23], as shown in Fig. 1.4. In Fig. 1.4 (a), the RIS creates a virtual LoS link, which avoids the obstacle between the transmitter and the receiver via a smart reflection. Moreover, adding more RISs provides more signal paths, resulting in a higher channel rank as shown in Fig. 1.4 (b). Due to the obstacles, the channel between the transmitter and the receiver can be modeled as Rayleigh/fast fading, while the RIS-aided link provides a Rician/slow fading channel. The coexistence of these two channels improves the channel statistics/distribution, which is crucial for ultra high-reliability communication applications as shown in Fig. 1.4 (c). Fig. 1.4 (d) illustrates the possibility of cancelling the inter-cell interference or suppressing the co-channel interference by smartly adjusting the RIS reflection coefficients.

Comparing RISs with the AF relays [GAH21; BÖL20; Yil+21; Wan+22], (i) the power consumption for the smart controller to adjust the reflection matrix of an RIS is practically negligible compared to the transmit power of AF relays [Hua+19], (ii) the RIS operates in FD mode, which is spectrally more efficient. Moreover, in RIS-aided systems, the Line of Sight (LoS) and the reflected paths with the same information signals can be designed to add coherently to strengthen the received signal, unlike the backscatter. On the other hand, massive MIMO is completely different from RISs in terms of array architecture, i.e., active antenna elements versus passive/semi-passive reflecting elements, and operating mechanism, i.e., transmit versus reflect [WZ19]. Moreover, the energy consumption in massive MIMO and mmWave technologies cannot be arbitrarily reduced due to the use of costly RF chains and also the complex signal processing, which is required to achieve high-performance communications [Wu+17].

Although the RISs have many promising unique potentials, from a signal processing perspective, they face us with new challenges to efficiently integrate them into wireless communication networks, such as reflection coefficient design and optimization, channel estimation, and deployment from a communication design perspective. In contrast to conventional wireless communication systems, in which all communication nodes use only active elements, RIS-based systems are deployed with semi-passive or completely passive components, to reduce the RIS cost and complexity. This in turn complicates the channel estimation methods, since the propagation channel can only be sensed and estimated at the receiver side. Moreover, the large number of channel coefficients to be estimated limits the feasibility of CSI acquisition within a practical coherence time, since an RIS is expected to have a massive number of reflecting components. On the other hand, RIS reflection matrix design problems are inherently non-convex and difficult to solve, mainly due to the massive number of reflecting components and some additional non-convex constraints, e.g., constant modulus constraints imposed due to the RIS hardware requirements. **In Part II of this thesis, we address these problems to propose new and efficient channel estimation and RIS reflection design methods for RIS-aided systems, while taking into account the above design and practical constraints.**

1.2. Major Contributions

The content of this thesis is organized in two separate parts, which can be read independently of each other. Also, each chapter is self-contained, so one can read the chapters in any order. The common algebraic tools used in both parts are outlined in Subsection 1.3. The structure of the thesis is visualized in Fig. 1.5. In the following, we give a summary of the main contributions.

1.2.1. Part I: Massive MIMO Wireless Communication Systems

The first part of the thesis is devoted to massive MIMO wireless communication systems. Our main focus in this part is on channel estimation and beamforming design for massive MIMO orthogonal frequency division multiplexing (OFDM) systems with a large number of antennas.

Chapter 2 considers a wideband massive MIMO OFDM system operating in a sub-6 GHz band where a single-user with a fully digital beamforming structure communicates with a BS having a HAD beamforming architecture. In contrast to earlier works on channel estimation for massive MIMO systems, which mainly considered the low-rank channel characteristics of the mmWave frequency bands, we propose a novel channel estimation method called sequential alternating least squares approximation (SALSA) by exploiting a hidden tensor structure in the uplink measurement matrix. Specifically, by showing that any MIMO channel matrix can be approximately decomposed into a summation of R factor matrices having a Kronecker structure, the uplink measurement matrix can be reshaped into a 3-way tensor admitting a Tucker decomposition [HRD08]. Exploiting such a tensor representation, the MIMO channel matrix can be estimated sequentially using the alternating least squares (ALS) method, where one of the three factor matrices that represents the analog training beams is known [CLA09]. Detailed simulation results are provided showing that the SALSA-based approach can achieve a more accurate channel estimation in difficult scenarios as compared to the classical least squares (LS)-based and minimum mean squared error (MMSE) approaches. Here, note that the proposed SALSA method is not limited to the single-user case and can easily be applied for multi-user scenarios using orthogonal pilot signals. The content of this chapter has been published in

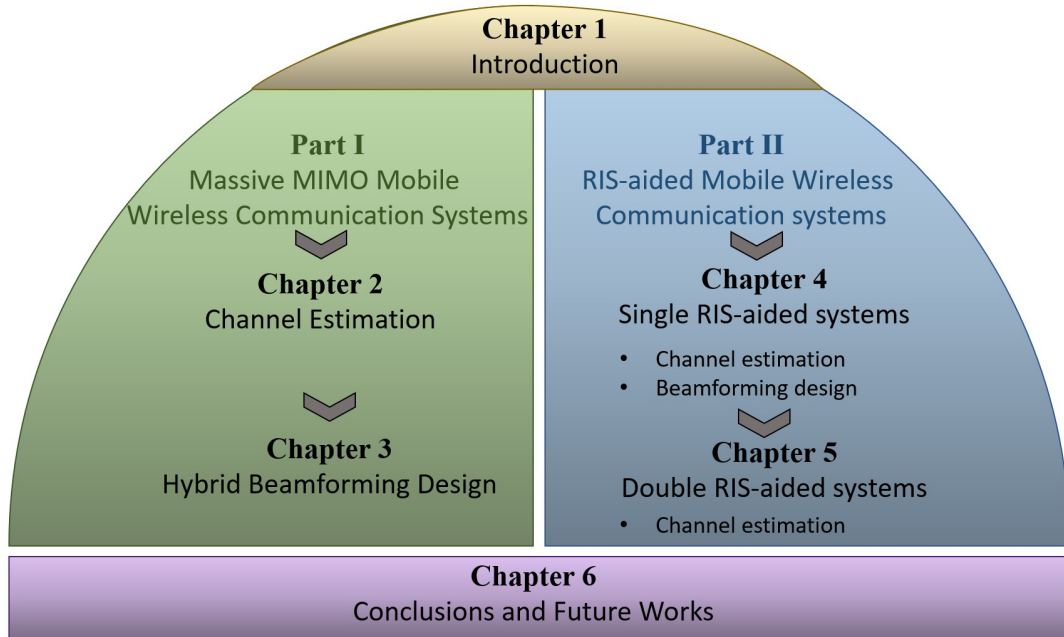


Figure 1.5: Structure of the thesis.

[GAH23] S. Gherekhloo, K. Ardah, and M. Haardt, "SALSA: A Sequential Alternating Least Squares Approximation Method for MIMO Channel Estimation," *IEEE Transactions on Vehicular Technology* (2023), doi: 10.1109/TVT.2023.3347290.

In Chapter 3, we consider a frequency-selective downlink multi-user MIMO OFDM system and propose a hybrid beamforming design approach with a low complexity that is applicable for any hybrid beamforming architecture. We assume that the transmit baseband beamforming matrices are designed using the classical zero forcing (ZF) method, while the receive baseband beamforming matrices are designed using the MMSE approach [Chr+08]. Moreover, we derive an approximation of the system's SE upper-bound, which is used as the objective function of the transmit HAD matrix optimization. To obtain a solution, we divide the problem into a series of convex sub-problems that are updated iteratively until convergence is obtained. Our proposed solution has either a comparable or a slightly better performance compared to the benchmark method, depending on the beamforming architecture. The content of this chapter has been published in

[GAH20] S. Gherekhloo, K. Ardah, and M. Haardt, "Hybrid Beamforming Design for Downlink MU-MIMO-OFDM Millimeter-Wave Systems," in *Proc. IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Hangzhou, China, 2020, pp. 1-5.

1.2.2. Part II: RIS-aided Wireless Communication Systems

In this part, our focus is on RIS-aided wireless communication systems. The main motivation for considering RISs is to improve the system coverage and overcome the obstacles in the high-frequency

bands.

Chapter 4 considers the channel estimation, the beamforming, and the RIS reflection design problems in point-to-point Single RIS (S-RIS)-aided MIMO systems using the mmWave channel model. We assume that the RIS contains passive reflecting elements and the direct link between the transmitter and the receiver is blocked or pre-estimated by turning the RIS elements off, as in [MJ19]. We propose two channel estimation methods. First, we use a structured channel training procedure to propose a Two-stage RIS-aided Channel Estimation (TRICE) framework for single-user S-RIS-aided mmWave MIMO communication systems. In the first stage, the directions of departure (DoDs) of the dominant paths of the transmitter-to-RIS channel and the directions of arrival (DoAs) of the dominant paths of the RIS-to-receiver channel are estimated. In the second stage, by using the estimated channel parameters of the first stage, the effective azimuth and elevation angles of the cascaded transmitter-to-RIS-to-receiver channel at the RIS are estimated, one-by-one, including the effective complex path gains. In both stages, we show that the parameter estimation can be carried out *via* a multi-dimensional DoA estimation scheme, for which several solutions exist as in [HRD08; ZH17a; ZH17b; SPP18; AAH19a; MRM16; Cao+18; AH21], among many others. For instance, we estimate the DoAs and DoDs in a closed-form using the Discrete Fourier Transform (DFT)-beamspace Estimation of Signal Parameters Via Rotational Invariance Techniques (ESPRIT) methods of [ZH17a; ZH17b]. Moreover, we use the Compressed Sensing (CS) techniques [SPP18; AAH19a; MRM16; Cao+18; AH21] to further improve the estimation performance in the case of difficult scenarios such as high noise power or a small number of measurement vectors. The TRICE method is compared to the LS and the on-grid CS technique proposed in [Wan+20], where all channel parameters are estimated jointly. It is shown that the proposed TRICE method has a lower training overhead and a lower computation complexity, compared to the LS estimate and the proposed method in [Wan+20].

After that, we extend our TRICE framework and propose a canonical polyadic decomposition (CPD)-based channel estimation method called Tensor-based RIS Channel Estimation (TenRICE), by jointly exploiting the tensor structure of the received signals and the low-rank nature of the mmWave channels. Note that the CPD is also known as Parallel Factor (PARAFAC) analysis. Using the TenRICE method, the transmitter-to-RIS and the RIS-to-receiver channels can be estimated separately using ALS, up to a trivial permutation and scaling factor. The TenRICE method has a superior performance compared to the TRICE framework, approaching the Cramér Rao lower Bound (CRB). This superior performance of the TenRICE compared to the TRICE is due to the exploitation of the tensor structure of the received signal in addition to considering the low-rank characteristics of the channel. In the end, we formulate the beamforming and the RIS reflection design as a SE maximization problem. Due to its non-convexity, we propose a heuristic non-iterative two-step solution, where the RIS reflection vector is obtained, different from [ZZ20a], in a closed-form using a Frobenius-norm Maximization strategy, termed FroMax. Our proposed FroMax solution achieves a comparable performance to benchmark methods with a lower complexity. The content of this chapter has been published in

[Ard+21] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "TRICE: A Channel Estimation Framework for RIS-Aided Millimeter-Wave MIMO Systems," *IEEE Signal Processing Letters*, vol. 28, pp. 513-517, 2021, doi: 10.1109/LSP.2021.3059363 .

[Ghe+21] S. Gherekhloo, K. Ardah, A. L. F. de Almeida, and M. Haardt, "Tensor-Based Channel Estimation and Reflection Design for RIS-Aided Millimeter-Wave MIMO Communication Systems," in *Proc. 55th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2021, pp. 1683-1689, doi: 10.1109/IEEECONF53345.2021.9723362.

Chapter 5 considers a Double RIS (D-RIS)-aided MIMO system. First, we propose an efficient channel estimation method by exploiting the tensor structure of the received signal, called channel estimation for joint training (CEJOINT). Specifically, we show that the received signals in flat-fading D-RIS-aided MIMO systems can be arranged in a 3-way tensor that admits Tucker2 decomposition. Accordingly, an ALS-based method is proposed, where the transmitter-to-RIS-1 channel, the RIS-1-to-RIS-2 channel, and the RIS-2-to-receiver channel can be estimated separately, up to trivial scaling factors. We compare an ALS method for S-RIS systems proposed in [AA20; AAB21] with CEJOINT in terms of the minimum training overhead and the estimation accuracy. It is shown that if the RIS elements in the S-RIS systems are distributed carefully between the two RISs in the D-RIS systems, the training overhead for CEJOINT can be reduced and the corresponding channel estimation accuracy can also be improved.

Then, different from CEJOINT, we assume that the training overhead assigned to each RIS in D-RIS-aided MIMO systems can be adjusted separately. Consequently, we show that the received signals in flat-fading D-RIS aided MIMO systems can be represented as a 4-way tensor that satisfies a nested PARAFAC decomposition model. Exploiting such a structure, we propose a non-iterative three-step channel estimation method, where one of the three channel matrices in each step can be determined in closed-form using a low-complexity Khatri-Rao Factorization (KRF) technique. To enhance the channel estimation accuracy, we further propose an ALS-based channel estimation method where we estimate one channel matrix assuming that the other two are fixed. We call the closed-form and the iterative channel estimation methods KRF-based for separate training (KRF-SEPT) and ALS-based for separate training (ALS-SEPT), respectively. The ALS-SEPT has more relaxed constraints on the training overhead than the KRF-SEPT method. Here, and different from the CEJOINT, we propose to initialize two channel matrices in ALS-SEPT using the closed-form KRF-SEPT method, even if the constraints of KRF-SEPT are not satisfied. Via computer simulations, we show that the ALS-SEPT method using KRF-SEPT initialization does not only have a faster convergence rate but has also a better estimation accuracy as compared to the ALS-SEPT method with random initializations. Moreover, we show that both proposed methods have a comparable estimation accuracy if the identifiability constraints of the KRF-SEPT method are satisfied. However, if less training overhead is used, the ALS-SEPT method outperforms the closed-form KRF-SEPT method with less training overhead. Last but not least, both proposed channel estimation methods can be extended to a multi-user scenario, where the channel estimation can be performed separately for each user with orthogonal training sequences in the time, frequency, and/or space domains without any multi-user interference. The content of this chapter has been published in

[Ard+22] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt, "Double-RIS Versus Single-RIS Aided Systems: Tensor-Based MIMO Channel Estimation and Design Perspectives," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, 2022, pp. 5183-5187, doi: 10.1109/ICASSP43922.2022.9746287.

[Ghe+23] S. Gherekhloo, K. Ardah, A. L. F. de Almeida, M. Maleki, and M. Haardt, "Nested PAR-AFAC Tensor-Based Channel Estimation Method for Double RIS-Aided MIMO Communication Systems," in *Proc. 31st European Signal Processing Conference (EUSIPCO)*, Helsinki, Finland, 2023, pp. 1674-1678, doi: 10.23919/EUSIPCO58844.2023.10290107.

In Chapter 6, we present the main conclusions along with some future research directions. We refer the readers to Appendix A for the proofs and the derivations in this thesis.

1.2.3. Scientific Production

The list of publications from which the content and the contribution of this work are taken is given below.

[GAH23] **S. Gherekhloo**, K. Ardah, and M. Haardt, "SALSA: A Sequential Alternating Least Squares Approximation Method for MIMO Channel Estimation," *IEEE Transactions on Vehicular Technology* (2023), doi: 10.1109/TVT.2023.3347290.

[Ard+21] K. Ardah, **S. Gherekhloo**, A. L. F. de Almeida, and M. Haardt, "TRICE: A Channel Estimation Framework for RIS-Aided Millimeter-Wave MIMO Systems," *IEEE Signal Processing Letters*, vol. 28, pp. 513-517, 2021, doi: 10.1109/LSP.2021.3059363 .

[GAH20] **S. Gherekhloo**, K. Ardah, and M. Haardt, "Hybrid Beamforming Design for Downlink MU-MIMO-OFDM Millimeter-Wave Systems," in *Proc. IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Hangzhou, China, 2020, pp. 1-5.

[GAH21] **S. Gherekhloo**, K. Ardah, and M. Haardt, "Fully Digital and Hybrid Beamforming Design for Millimeter-Wave MIMO-OFDM Two-Way Relaying Systems," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, Rio de Janeiro, Brazil, 2021, pp. 296-300.

[Ghe+21] **S. Gherekhloo**, K. Ardah, A. L. F. de Almeida, and M. Haardt, "Tensor-Based Channel Estimation and Reflection Design for RIS-Aided Millimeter-Wave MIMO Communication Systems," in *Proc. 55th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2021, pp. 1683-1689, doi: 10.1109/IEEECONF53345.2021.9723362 .

[Ard+22] K. Ardah, **S. Gherekhloo**, A. L. F. de Almeida, and M. Haardt, "Double-RIS Versus Single-RIS Aided Systems: Tensor-Based MIMO Channel Estimation and Design Perspectives," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, Singapore, 2022, pp. 5183-5187, doi: 10.1109/ICASSP43922.2022.9746287.

[Ghe+23] **S. Gherekhloo**, K. Ardah, A. L. F. de Almeida, M. Maleki, and M. Haardt, "Nested PARAFAC Tensor-Based Channel Estimation Method for Double RIS-Aided MIMO Communication Systems," in *Proc. 31st European Signal Processing Conference (EUSIPCO)*, Helsinki, Finland, 2023, pp. 1674-1678, doi: 10.23919/EUSIPCO58844.2023.10290107 .

1.3. Preliminary Definitions and Concepts

This section defines the notation and some necessary basic concepts used in this thesis. All properties/definitions required to understand the following chapters are summarized in a compact and systematic form. First, the general notation used throughout the work is defined. Then, an overview of some of the matrix and tensor properties/concepts used in the rest of the thesis is given.

1.3.1. Notation

We present a scalar, column vector, and matrix as a , \mathbf{a} , and \mathbf{A} , respectively. The (i, j) th element of the matrix \mathbf{A} is denoted by $[\mathbf{A}]_{[i,j]}$. For the vector case, the i th element of \mathbf{a} is denoted by $[\mathbf{a}]_{[i]}$. We use the superscripts $(\cdot)^{-1}$, $(\cdot)^+$, $(\cdot)^*$, $(\cdot)^\top$, $(\cdot)^\text{H}$, and $\angle(\cdot)$ to denote the matrix inverse, Moore-Penrose pseudo inverse, complex conjugate, transposition, conjugate transposition, and angle, respectively. Moreover, $\mathbb{E}\{\cdot\}$, $\text{trace}[\cdot]$, $\|\cdot\|_F$, and $\|\cdot\|_2$ denote the expectation, trace, Frobenius norm, and 2-norm operators, respectively. We use $\lceil x \rceil$ to denote the ceiling function. We define $\mathbf{1}_N$ as the all ones column vector of length N , \mathbf{I}_N as the $N \times N$ identity matrix, $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ as the circularly symmetric complex Gaussian distribution with zero mean and covariance matrix \mathbf{R} .

1.3.2. Matrix Properties

For the two matrices $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{B} \in \mathbb{C}^{P \times Q}$, the Kronecker product between them, i.e., $\mathbf{A} \otimes \mathbf{B}$ is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} [\mathbf{A}]_{[1,1]} \mathbf{B} & \cdots & [\mathbf{A}]_{[1,M]} \mathbf{B} \\ \vdots & \ddots & \vdots \\ [\mathbf{A}]_{[N,1]} \mathbf{B} & \cdots & [\mathbf{A}]_{[N,M]} \mathbf{B} \end{bmatrix} \in \mathbb{C}^{NP \times MQ} \quad (1.1)$$

The Khatri-Rao product is another important operator. It can be obtained by calculating the column-wise Kronecker product of the matrices, which requires that \mathbf{A} and \mathbf{B} have the same number of columns, i.e., $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{B} \in \mathbb{C}^{P \times M}$, where the Khatri-Rao product between them is defined as

$$\mathbf{A} \diamond \mathbf{B} = [[\mathbf{a}]_{[:,1]} \otimes [\mathbf{b}]_{[:,1]}, \cdots, [\mathbf{a}]_{[:,m]} \otimes [\mathbf{b}]_{[:,m]}, \cdots, [\mathbf{a}]_{[:,M]} \otimes [\mathbf{b}]_{[:,M}]] \in \mathbb{C}^{NP \times M} \quad (1.2)$$

To perform the element-wise product between matrices, the Hadamard product is defined. Note that the two matrices \mathbf{A} and \mathbf{B} must have the same dimensions, i.e., $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{B} \in \mathbb{C}^{N \times M}$, where the Hadamard product between them is defined as

$$\mathbf{Z} = \mathbf{A} \odot \mathbf{B} \in \mathbb{C}^{N \times M}, \quad (1.3)$$

$[\mathbf{Z}]_{[a,b]} = [\mathbf{A}]_{[a,b]} [\mathbf{B}]_{[a,b]}$ for $a = \{1, \dots, N\}$, and $b = \{1, \dots, M\}$. In the following, we summarize some of the important related properties of the products discussed above as [Bre78; PP06]

$$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A} \quad (1.4)$$

$$(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top \quad (1.5)$$

$$(\mathbf{A} \otimes \mathbf{B})^\text{H} = \mathbf{A}^\text{H} \otimes \mathbf{B}^\text{H} \quad (1.6)$$

$$(\mathbf{A} \otimes \mathbf{B})^+ = \mathbf{A}^+ \otimes \mathbf{B}^+ \quad (1.7)$$

$$\|\mathbf{A} \otimes \mathbf{B}\|_F = \|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_F \quad (1.8)$$

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \quad (1.9)$$

$$(\mathbf{A} \cdot \mathbf{C}) \otimes (\mathbf{B} \cdot \mathbf{D}) = (\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \otimes \mathbf{D}) \quad (1.10)$$

$$(\mathbf{A} \cdot \mathbf{C}) \diamond (\mathbf{B} \cdot \mathbf{D}) = (\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \diamond \mathbf{D}) \quad (1.11)$$

$$(\mathbf{A} \diamond \mathbf{B})^\top \cdot (\mathbf{A} \diamond \mathbf{B}) = (\mathbf{A}^\top \cdot \mathbf{A}) \odot (\mathbf{B}^\top \cdot \mathbf{B}) \quad (1.12)$$

$$\text{rank}(\mathbf{A} \otimes \mathbf{B}) = \text{rank}(\mathbf{A})\text{rank}(\mathbf{B}). \quad (1.13)$$

Assuming a matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M] \in \mathbb{C}^{N \times M}$, where $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M$ are its columns, we use the $\text{vec}\{\cdot\}$ operator to vectorize \mathbf{A} as

$$\text{vec}\{\mathbf{A}\} = [\mathbf{a}_1^\top, \mathbf{a}_2^\top, \dots, \mathbf{a}_M^\top]^\top \in \mathbb{C}^{NM \times 1}. \quad (1.14)$$

We use $\text{unvec}\{\cdot\}$ as the inverse of the vec operator, i.e.,

$$\text{unvec}_{N \times M}\{[\mathbf{a}_1^\top, \mathbf{a}_2^\top, \dots, \mathbf{a}_M^\top]^\top\} = \mathbf{A} \in \mathbb{C}^{N \times M}. \quad (1.15)$$

The $\text{diag}\{\cdot\}$ operator is used to create a diagonal matrix as

$$\text{diag}\{\mathbf{a}\} = \begin{bmatrix} [\mathbf{a}]_{[1]} & 0 & \cdots & 0 \\ 0 & [\mathbf{a}]_{[2]} & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & [\mathbf{a}]_{[M]} \end{bmatrix} \in \mathbb{C}^{M \times M} \quad (1.16)$$

where $\mathbf{a} \in \mathbb{C}^M$. Moreover, $\text{undia}\{\cdot\}$ is the reverse of the diag operator, i.e.,

$$\text{undia}\left\{ \begin{bmatrix} [\mathbf{a}]_{[1]} & 0 & \cdots & 0 \\ 0 & [\mathbf{a}]_{[2]} & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & [\mathbf{a}]_{[M]} \end{bmatrix} \right\} = \mathbf{a} \in \mathbb{C}^M, \quad (1.17)$$

it extracts the diagonal of the square matrix.

In the following, we list some of the $\text{vec}\{\cdot\}$ properties [Bre78; PP06; Neu69]

$$\text{vec}\{\mathbf{A} \cdot \mathbf{X} \cdot \mathbf{B}\} = (\mathbf{B}^\top \otimes \mathbf{A}) \cdot \text{vec}\{\mathbf{X}\} \quad (1.18)$$

$$\text{vec}\{\mathbf{A} \cdot \mathbf{B}\} = (\mathbf{I}_P \otimes \mathbf{A}) \cdot \text{vec}\{\mathbf{B}\} = (\mathbf{B}^\top \otimes \mathbf{I}_M) \cdot \text{vec}\{\mathbf{A}\} \quad (1.19)$$

$$\text{vec}\{\mathbf{A} \cdot \text{diag}\{\mathbf{x}\} \cdot \mathbf{B}\} = (\mathbf{B}^\top \diamond \mathbf{A}) \cdot \mathbf{x} \quad (1.20)$$

where $\mathbf{A} \in \mathbb{C}^{N \times M}$, $\mathbf{B} \in \mathbb{C}^{P \times Q}$, $\mathbf{X} \in \mathbb{C}^{M \times P}$, and $\mathbf{x} \in \mathbb{C}^M$. Note that, in (1.19) and (1.20), we assume that $M = P$. In addition, we use the trace operator with the following properties assuming that the multiplication are well defined [Bre78; PP06; Neu69]

$$\text{trace}[\mathbf{A} \cdot \mathbf{B}] = \text{trace}[\mathbf{B} \cdot \mathbf{A}] = \text{vec}\{\mathbf{A}^\top\}^\top \cdot \text{vec}\{\mathbf{B}\} \quad (1.21)$$

$$\text{trace}[\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C} \cdot \mathbf{D}] = \text{trace}[\mathbf{B} \cdot \mathbf{C} \cdot \mathbf{D} \cdot \mathbf{A}] = \text{trace}[\mathbf{C} \cdot \mathbf{D} \cdot \mathbf{A} \cdot \mathbf{B}] = \text{trace}[\mathbf{D} \cdot \mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}] \quad (1.22)$$

$$\text{trace}[\mathbf{A} \otimes \mathbf{B}] = \text{trace}[\mathbf{B}] \cdot \text{trace}[\mathbf{A}] \quad (1.23)$$

$$\text{trace}[\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}] = \text{vec}\{\mathbf{A}^\top\}^\top \cdot (\mathbf{I}_N \otimes \mathbf{B}) \cdot \text{vec}\{\mathbf{C}\} \quad (1.24)$$

$$\text{trace}[\mathbf{B} \cdot \mathbf{Y} \cdot \mathbf{B}^\mathbf{H}] = \text{vec}\{\mathbf{B}\}^\mathbf{H} \cdot (\mathbf{Y}^\top \otimes \mathbf{I}_P) \cdot \text{vec}\{\mathbf{B}\}, \quad (1.25)$$

where $\mathbf{Y} \in \mathbb{C}^{Q \times Q}$.

Lemma 1.3.1. *Let $\mathbf{A}_1 \in \mathbb{C}^{J_1 \times L_1}$ and $\mathbf{A}_2 \in \mathbb{C}^{J_2 \times L_2}$. Then*

$$\mathbf{A}_1 \otimes \mathbf{A}_2 = \mathbf{A}_1 \boldsymbol{\Omega}_1 \diamond \mathbf{A}_2 \boldsymbol{\Omega}_2, \quad (1.26)$$

where $\boldsymbol{\Omega}_1 = \mathbf{I}_{L_1} \otimes \mathbf{1}_{L_2}^\top$ and $\boldsymbol{\Omega}_2 = \mathbf{1}_{L_1}^\top \otimes \mathbf{I}_{L_2}$ so that $\boldsymbol{\Omega}_1 \diamond \boldsymbol{\Omega}_2 = \mathbf{I}_{L_1 L_2}$.

Proof: Please refer to Appendix A.1.

1.3.3. Tensor Algebra

1.3.3.1 Definitions and Properties

Tensor algebra is a mathematical discipline that has become very popular in various scientific fields from psychometrics and chemometrics [KB09] to array signal processing [Lim+17], communications, biomedical signal processing, etc. [Cic+15]. Tensors are used throughout this thesis, so we give a brief introduction in the hope that it may help the non-familiar reader to understand the basic concepts of tensor algebra.

In [KB09], a tensor is defined as an element of the tensor product of vector spaces. In other words, a tensor can be represented as a multi-dimensional array defined for any coordinate system. In this thesis, we consider the canonical coordinate system [Com14]. The number of dimensions, also called modes, of the associated multi-dimensional array defines the tensor order. Therefore, $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_D}$ represents an D -way tensor with M_d denoting the size along the d th dimension (mode) for $d \in \{1, 2, \dots, D\}$. Obviously, the tensors with $D = 2$ are matrices. The (m_1, m_2, \dots, m_D) th element of \mathcal{A} is denoted by $[\mathcal{A}]_{[m_1, m_2, \dots, m_D]}$ for $m_d \in \{1, \dots, M_d\}$ and $d \in \{1, \dots, D\}$. The total number of elements of \mathcal{A} is $M = \prod_{d=1}^D M_d$. By fixing all indices except for the d th index, we refer to the d -mode vectors of a tensor, which are the generalizations of row and column vectors of matrices. A matrix $[\mathcal{A}]_{(d)}$ with the dimension of $M_d \times \frac{M}{M_d}$ is called the d -mode unfolding of \mathcal{A} , which contains all d -mode vectors as its columns. Depending on the ordering of the columns in the d -mode unfolding, the arrangement of the other $d - 1$ indices is defined. Below we present the two most popular ordering options from the multitude of existing ones.

- Reverse cyclical: In [DDV00], the authors introduce this way of ordering, which has become the standard practice in the community. The ordering starts with the $(d - 1)$ th index, then the $(d - 2)$ th, and it continues up to the 1st, then it starts over with the D th index, the $(D - 1)$ th, and continues up to the $(d + 1)$ th index.
- Forward column ordering: This d -mode unfolding is obtained by starting at the first index and going up to the D th index by omitting the d th index. This can be implemented in MATLAB using the "permute" and "reshape" commands. In Fig. 1.6, a 3-way tensor $\mathcal{A} \in \mathbb{C}^{5 \times 4 \times 3}$ and its forward unfoldings are shown. Throughout this thesis, we use the forward column ordering.

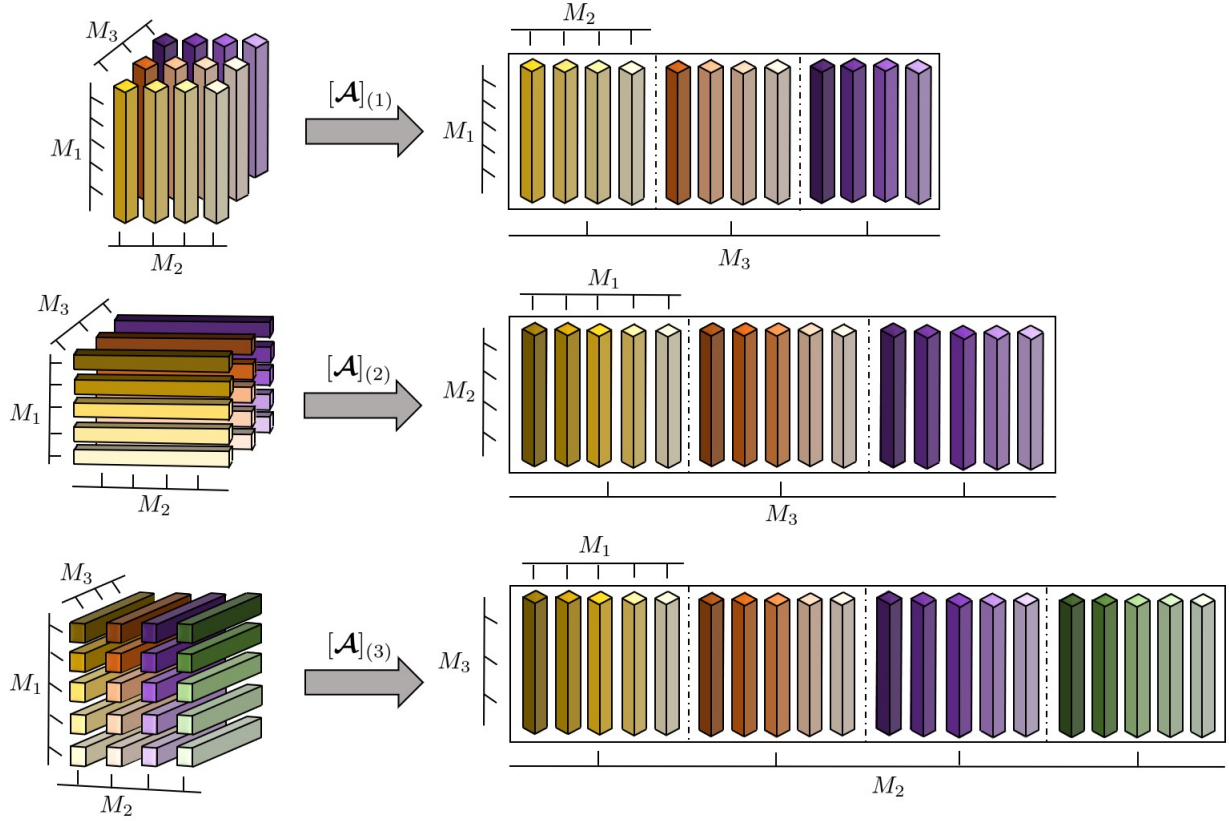


Figure 1.6: Graphical representation of the forward unfolding of a 3-way tensor.

The d -mode product of a tensor \mathcal{A} with a matrix $\mathbf{B} \in \mathbb{C}^{N \times M_d}$ is defined as

$$\mathbf{C} = \mathcal{A} \times_d \mathbf{B} \Rightarrow [\mathbf{C}]_{(d)} = \mathbf{B} \cdot [\mathcal{A}]_{(d)}, \quad (1.27)$$

where $\mathbf{C} \in \mathbb{C}^{M_1 \times \dots \times N \times \dots \times M_D}$. Let the tensor \mathbf{C} be defined as

$$\mathbf{C} = \mathcal{A} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \dots \times_D \mathbf{X}_D. \quad (1.28)$$

Then, the d -mode unfolding can be expressed as

$$[\mathbf{C}]_{(d)} = \mathbf{X}_d \cdot [\mathcal{A}]_{(d)} \cdot \left(\mathbf{X}_D \otimes \mathbf{X}_{D-1} \otimes \dots \otimes \mathbf{X}_{d+1} \otimes \mathbf{X}_{d-1} \otimes \mathbf{X}_{d-2} \otimes \dots \otimes \mathbf{X}_1 \right)^\top, \quad (1.29)$$

where $\mathbf{X}_d \in \mathbb{C}^{N_d \times M_d}$, $d \in \{1, 2, \dots, D\}$, are the factor matrices. Note that the order of the matrices in the Kronecker product is a direct result of the forward column ordering that was chosen for the d -mode unfolding. Some more properties of the d -mode product and d -mode unfoldings are listed below.

$$\mathcal{A} \times_d \mathbf{X}_d \times_p \mathbf{X}_p = \mathcal{A} \times_p \mathbf{X}_p \times_d \mathbf{X}_d \quad \text{where } d \neq p \quad (1.30)$$

$$\mathcal{A} \times_d \mathbf{X}_d \times_d \mathbf{Y}_d = \mathcal{A} \times_d (\mathbf{Y}_d \cdot \mathbf{X}_d), \quad (1.31)$$

where $d, p \in \{1, 2, \dots, D\}$ and $\mathbf{Y}_d \in \mathbb{C}^{P_d \times M_d}$.

The d -rank of the tensor \mathcal{A} refers to the dimension of the space spanned by the d -mode vectors which is also known as d -space of \mathcal{A} . In other words, the d -rank is the rank of the d -mode unfolding, i.e., $[\mathcal{A}]_{(d)}$. The d -rank of a tensor is also called the multi-linear rank and denoted as $d\text{-rank}(\mathcal{A}), \forall d \in \{1, 2, \dots, D\}$. Note that all $d\text{-rank}(\mathcal{A})$ can be different. Similar to a matrix that can be constructed as the sum of R rank-one matrices, a tensor can be constructed as the sum of R rank-one tensors. Therefore, the smallest possible number of rank-one tensors that are summed to construct the tensor is defined as the tensor rank R . It is important to mention that the tensor rank is not directly related to the d -ranks, but it only provides an upper bound, i.e., $\text{rank}(\mathcal{A}) \geq d\text{-rank}(\mathcal{A}), \forall d$.

In the following, we introduce the vectorization operation for tensors

$$\text{vec}\{\mathcal{A}\} = \text{vec}\{[\mathcal{A}]_{(1)}\} = \mathbf{P}_{M_1, M_2, \dots, M_D}^d \cdot \text{vec}\{[\mathcal{A}]_{(d)}\}, \quad (1.32)$$

where $\mathbf{P}_{M_1, M_2, \dots, M_D}^d$ is a unique set of permutation matrices for $d \in \{1, 2, \dots, D\}$. Note that the equation (1.32) shows the fact that the vectorized versions of different unfoldings contain the same elements in a different order. Therefore, the permutation matrices with correct ordering can be defined as in [Roe13]. The vec -operation applied to a tensor \mathcal{A} can be performed by stacking it in an increasing order, i.e., start from the first index and going up to the D th index. Below, we show the vectorization of a 3-way tensor $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$ as

$$\begin{aligned} \text{vec}\{\mathcal{A}\} &= \text{vec}\{[\mathcal{A}]_{[:, :, 1]} \sqcup_3 [\mathcal{A}]_{[:, :, 2]} \sqcup_3 \dots \sqcup_3 [\mathcal{A}]_{[:, :, M_3]}\} \\ &= \begin{bmatrix} \text{vec}\{[\mathcal{A}]_{[:, :, 1]}\}^T & \text{vec}\{[\mathcal{A}]_{[:, :, 2]}\}^T & \dots & \text{vec}\{[\mathcal{A}]_{[:, :, M_3]}\}^T \end{bmatrix}^T, \end{aligned} \quad (1.33)$$

where $[\mathcal{A}]_{[:, :, j]} \in \mathbb{C}^{M_1 \times M_2}$ is the j th slice of \mathcal{A} for $j \in \{1, 2, \dots, M_3\}$ and \sqcup_d denotes the concatenation along the d -mode.

The higher order norm $\|\cdot\|_{\text{H}}$ or Frobenius norm of a tensor is defined as

$$\|\mathcal{A}\|_{\text{H}} = \|[\mathcal{A}]_{(d)}\|_{\text{F}} = \|\text{vec}\{[\mathcal{A}]_{(d)}\}\|_2 \quad \forall d \in \{1, 2, \dots, D\}, \quad (1.34)$$

which is the higher-order extension of the Frobenius norm.

1.3.3.2 Tensor Decompositions

In the past two decades, several works have been written on tensor decompositions because it is one of the most appealing features of multi-linear algebra for signal processing [Sid+17]. Particularly, the uniqueness conditions of the decompositions have many benefits in signal processing applications, such as blind detection and identification schemes [Alm+06] and relaxed parameter estimation identifiability conditions [MGB08]. Here, we focus on the two most popular tensor factorization methods, i.e., CPD (CANDECOMP/PARAFAC Decomposition) and Tucker decomposition.

CPD (CANDECOMP/ PARAFAC Decomposition) In the same manner as a matrix that can be expressed by the summation of its rank-one components, e.g., in the singular value decomposition (SVD),

tensors of rank R can be constructed by summing its rank-one tensor components. Assume that the D -way tensor \mathcal{A} admits a rank R , then its CPD is given as

$$\begin{aligned}\mathcal{A} &= \sum_{r=1}^R [\mathbf{A}_1]_{[:,r]} \circ \cdots \circ [\mathbf{A}_D]_{[:,r]} \\ &= \mathcal{I}_{D,R} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \cdots \times_D \mathbf{A}_D,\end{aligned}\quad (1.35)$$

where $\mathbf{A}_d \in \mathbb{C}^{M_d \times R}$, $d = \{1, \dots, D\}$, are the factor matrices, $\mathcal{I}_{D,R} \in \mathbb{R}^{R \times R \times \cdots \times R}$ is the super-diagonal tensor of the order D , where there are ones on the super-diagonal and the rest of the elements is zeros, and \circ denotes the outer product between the columns of the factor matrices \mathbf{A}_d . The d -mode unfolding of \mathcal{A} , i.e., $[\mathcal{A}]_{(d)}$, is expressed as

$$[\mathcal{A}]_{(d)} = \mathbf{A}_d (\mathbf{A}_D \diamond \cdots \diamond \mathbf{A}_{d+1} \diamond \mathbf{A}_{d-1} \diamond \cdots \mathbf{A}_1)^\top \in \mathbb{C}^{M_d \times M_1 \cdots M_{d-1} M_{d+1} \cdots M_D}. \quad (1.36)$$

Note that the factor matrices in the CPD are not unitary, which means that the D underlying components are not necessarily orthogonal.

For the CPD, it has been shown that the factor matrices are unique up to scaling and permutation ambiguities, which renders the CPD a very practical tool for a wide range of applications. For a matrix \mathbf{A} , its Kruskal rank, i.e., $k\text{-rank}(\cdot)$, represents the largest k number of columns, which are linearly independent. The sufficient condition for uniqueness of the CPD can also be derived using the Kruskal rank [Kru77; Kru89; Com14], e.g., for a 3-way tensor $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$ the sufficient condition for the uniqueness is given as

$$k\text{-rank}(\mathbf{A}_1) + k\text{-rank}(\mathbf{A}_2) + k\text{-rank}(\mathbf{A}_3) \geq 2R + 2, \quad (1.37)$$

where $\mathbf{A}_d \in \mathbb{C}^{R \times M_d}$ is the d th factor matrix for $d \in \{1, 2, 3\}$.

This guarantee enables us to determine the conditions required to recover column-scaled and column-permuted versions of the factor matrices \mathbf{A}_d by decomposing \mathcal{A} . Further discussions on the sufficient conditions for the CPD are outside the scope of this thesis, so for interested readers, we refer to [KB09].

Tucker Decomposition The Tucker decomposition was first proposed in [Tuc66]. Assuming a D -way tensor $\mathcal{A} \in \mathbb{C}^{M_1 \times \cdots \times M_D}$, the Tucker decomposition is expressed as

$$\begin{aligned}\mathcal{A} &= \sum_{r_1=1}^{R_1} \cdots \sum_{r_D=1}^{R_D} [\mathcal{G}]_{[r_1, \dots, r_D]} \left([\mathbf{A}_1]_{[:,r_1]} \circ \cdots \circ [\mathbf{A}_D]_{[:,r_D]} \right), \\ &= \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \cdots \times_D \mathbf{A}_D,\end{aligned}\quad (1.38)$$

where $\mathbf{A}_d \in \mathbb{C}^{M_d \times R_d}$ is the d th factor matrix for $d = \{1, \dots, D\}$ and $\mathcal{G} \in \mathbb{C}^{R_1 \times \cdots \times R_D}$ is the core tensor. The d -mode unfolding of \mathcal{A} , i.e., $[\mathcal{A}]_{(d)}$, can be expressed as

$$[\mathcal{A}]_{(d)} = \mathbf{A}_d [\mathcal{G}]_{(d)} \left(\mathbf{A}_D \otimes \cdots \otimes \mathbf{A}_{d+1} \otimes \mathbf{A}_{d-1} \otimes \cdots \mathbf{A}_1 \right)^\top. \quad (1.39)$$

Part I

Massive MIMO Wireless Communication Systems

This part of the thesis is devoted to the wideband massive multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) wireless communication systems. We consider both the channel estimation problem and the hybrid analog digital (HAD) beamforming design problem. In particular, in Chapter 2, we propose a novel channel estimation algorithm called sequential alternating least squares approximation (SALSA), which exploits the hidden tensor structure of the received signals. In Chapter 3, on the other hand, we propose a hybrid beamforming design algorithm that is applicable to any HAD beamforming architecture. The effectiveness of the proposed algorithms is illustrated by numerical results.

Channel Estimation

2.1. Introduction and State of the Art

Massive MIMO [Mar+16] is one of the key enabling technologies of 5G NR wireless communication networks [DPS20] and it shall remain relevant in future 6G wireless systems. By employing a large number of antennas at the BS relative to the number of scheduled users, massive MIMO systems increase the data throughput relative to legacy systems by providing a large beamforming gain and an improved multi-user interference suppression owing to its high spatial resolution [Bjö+15]. Recently, massive MIMO communications have received a special attention with the introduction of mmWave-based wireless communications [Hea+16], since the use of massive MIMO in such systems becomes a requirement rather than an option to compensate the high path loss encountered in the wireless communications at higher-frequency bands. On the other hand, it is well-known that the promised theoretical gains of massive MIMO systems heavily rely on the availability of accurate CSI and the considered beamforming structure. However, the CSI acquisition task in massive MIMO systems has been recognized as very challenging, due to the high dimensionality of the channel matrices, the overwhelming training overhead, and the prohibitive computational complexity [Xie+18].

Massive MIMO systems can either be deployed in time division duplexing (TDD) mode or frequency division duplexing (FDD) mode. Using the TDD mode, both the uplink and the downlink communications occur in the same frequency band, but at different time slots. However, in the FDD mode, the uplink and the downlink communications operate simultaneously at the same time, but on different frequency bands. Operating on different frequency bands for the uplink and downlink in FDD-based communication systems means that the downlink channel is neither the same nor can it be derived from the uplink channel. In this case, the conventional downlink channel strategy used in the TDD mode, in which the receiver estimates its own downlink channel and sends it to the transmitter for subsequent signal transmission and resource allocation, is no longer applicable. In addition, downlink channel estimation in Massive MIMO systems has other drawbacks, such as an enormous training effort, which reduces system efficiency. Therefore, much attention has been paid to the TDD mode, wherein the channel reciprocity feature occurs due to the use of the same frequency band for the uplink and downlink communications, which can be exploited to obtain the downlink channel from its uplink counterpart. Therefore, the training overhead for channel estimation can be reduced significantly [Mar+16; Xie+18; Geb+21].

In conventional pilot-based channel estimation methods, the transmitter emits a sequence of known bits called pilots or reference signals (RSs) [BG06], wherein the receiver uses the measurements and known pilots/RSs to estimate the propagation channel. In practical systems, such as 3GPP 5G NR, the CSI-RSs are used to estimate the channel in the downlink direction, i.e., from a BS to a user equipment (UE), and the sounding reference signals (SRSs) are used to estimate the channel in the uplink direction, i.e., from a UE to a BS. Pilot-based channel estimation methods are very well established in the literature [BG06; Geb+21]. However, since the transmission of pilot signals consumes some system resources, e.g.,

time, frequency, and power, they usually incur some efficiency loss, which increases with an increasing pilot-transmission overhead. In order to reduce the impact of the pilot transmission overhead, some alternative techniques have been proposed such as semi-blind and blind channel estimation methods [Keh+19; ZN04; MCD02; HLL13]. Specifically, semi-blind techniques utilize not only the known pilot symbols but also the unknown data symbols to reduce the amount of pilot data to achieve the required estimation accuracy. Thereby the SE and channel throughput are improved. On the other hand, blind methods jointly perform channel estimation and data detection without the use of pilot symbols, at the expense of a higher computational complexity. Moreover, blind methods suffer from certain ambiguities that need to be resolved, which add to the complexity of the communication.

In mmWave systems, the channel estimation problem is often transformed into a multi-dimensional DoA estimation problem [ALH15; Ard+20b; ZH17a], thanks to the low-rank (sparse) nature of mmWave MIMO channels [Hea+16], where several techniques, e.g., CS [ALH15; Ard+20b; PTV23] and ESPRIT [ZH17a] can be readily employed to obtain a high CSI estimation accuracy while requiring only a small training overhead. Differently, in sub-6 GHz-based systems, the MIMO channels often experience a high-rank nature, which makes most, if not all, mmWave-based MIMO channel estimation methods unfeasible. To this end, classical channel estimation techniques, e.g., LS and linear minimum mean squared error (LMMSE) methods [BLM03; BG06] can be used to estimate MIMO channels in the sub-6 GHz band. These methods were originally developed for single-antenna and small-scale MIMO systems. On the one hand, the LS technique is very simple, but it suffers from a severe performance degradation in difficult scenarios, e.g., with a small number of training snapshots and/or a low SNR [ZP06]. On the other hand, the LMMSE-based method offers a more robust performance in dealing with noise than LS at the cost of higher computational complexity [Sri+04]. Moreover, the LMMSE-based method requires some prior estimation of the statistical channel correlation matrix and noise variance, which makes it challenging to realize LMMSE in practice [AIU06; UV18]. Since sub-6 GHz massive MIMO communications are, and will remain, an integral part of current 4G and 5G and future wireless communication systems, more efficient channel estimation techniques than the classical methods are required.

2.1.1. Chapter Contributions

In this chapter, we consider the channel estimation problem in sub-6 GHz uplink wideband MIMO-OFDM communication systems. It is assumed that a UE equipped with multiple antennas and a fully digital beamforming architecture communicates with a BS that is equipped with a massive antenna array and a HAD beamforming architecture. A generalized HAD beamforming architecture, as shown in Fig. 2.1 is assumed at the BS, which comprises the classical fully connected (FC) and partially connected (PC) architectures as special cases. **As the main contribution**, we propose a novel channel estimation method called SALSA by exploiting a *hidden tensor structure* in the uplink measurement matrix. Specifically, by showing that any MIMO channel matrix can be approximately decomposed into a summation of R factor matrices having a Kronecker structure. In this case, the uplink measurement matrix can be reshaped into a 3-way tensor admitting a Tucker decomposition [HRD08]. Exploiting such a tensor representation, the MIMO channel matrix can be estimated sequentially using the ALS method [CLA09].

In the literature, we note that several tensor-based channel estimation methods have been proposed, e.g., [CLA09; Alm07; dFR18; Lin+20; Qia+18; Ghe+21; HRD08], and references therein. However, these methods exploit simplified design assumptions on the antenna array structure and channel matrices,

which generally lead to a relatively straightforward tensor modeling. Nonetheless, practical channel models, e.g., the 3GPP clustered delay line (CDL) channel model [3GP20a; RDT22], do not necessarily comply with these assumptions, and therefore, the existing methods, in many cases, cannot be employed. On the other hand, the proposed SALSA method does not require these assumptions on the channel models, which renders it very general and appealing for practical implementations. The content of this chapter has been published in [GAH23].

2.1.2. Chapter Organization

In this chapter, we start with the system configurations in Section 2.2, where the channel model, beamforming architectures, and the system model are discussed in detail. Then, the baseline LS and LMMSE methods and the proposed SALSA algorithm are presented in Section 2.3. To examine the proposed channel estimation algorithm performance, we generate the channel matrix as described in Subsection 2.2.2 using the realistic and well-known 3GPP TR 38.901 CDL channel simulator. The benefits of the SALSA method compared to the baseline LS and LMMSE methods are demonstrated through numerical simulation results in Section 2.4. Finally, the conclusion is provided in Section 2.5.

2.2. System Configuration

2.2.1. System Model

We consider an uplink single-user wideband MIMO-OFDM communication system, as depicted in Fig. 2.1, where a UE with M_{UE} antennas is communicating with a BS with M_{BS} antennas over K subcarriers. The UE has a fully digital beamforming architecture while the BS has a HAD beamforming architecture with $N_{\text{BS}} \ll M_{\text{BS}}$ RF chains. We assume that the M_{BS} antennas and the N_{BS} RF chains are divided *equally*¹ into $G \geq 1$ groups, where each group has $\dot{M}_{\text{BS}} = \frac{M_{\text{BS}}}{G}$ antennas and $\dot{N}_{\text{BS}} = \frac{N_{\text{BS}}}{G}$ RF chains (i.e., $M_{\text{BS}} = \dot{M}_{\text{BS}} \cdot G$ and $N_{\text{BS}} = \dot{N}_{\text{BS}} \cdot G$) and the RF chains in every group are connected with every antenna element in the same group.

2.2.2. Channel Model

We assume that the considered MIMO OFDM system operates on a perfectly tuned TDD channel, which enables the assumption of channel reciprocity. Moreover, a cyclic-prefix OFDM-based multi-carrier modulation scheme is used, which avoids inter symbol interference (ISI) by choosing the cyclic-prefix larger than the number of channel taps to combat the multipath effect as well as enables zero inter carrier interference (ICI). The system bandwidth B with the center frequency f_c is divided into K OFDM subcarriers each with bandwidth Δf , i.e., $B = K \cdot \Delta f$.

The communication between the BS and the UE takes place over the time-variant frequency-selective MIMO channel $\mathbf{H}(\tau, t) \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}$, where t and τ represent the continuous time- and delay-domain, respectively. The channel matrix $\mathbf{H}(\tau, t)$ with L significant multipath components can be expressed as

$$\mathbf{H}(\tau, t) = \sum_{\ell=1}^L \mathbf{H}^{\ell}(t) \cdot \delta(\tau - \tau_{\ell}), \quad (2.1)$$

¹To simplify the exposition, we assume that M_{BS} , N_{BS} , and G are selected so that \dot{M}_{BS} and \dot{N}_{BS} are integer numbers, without loss of generality.

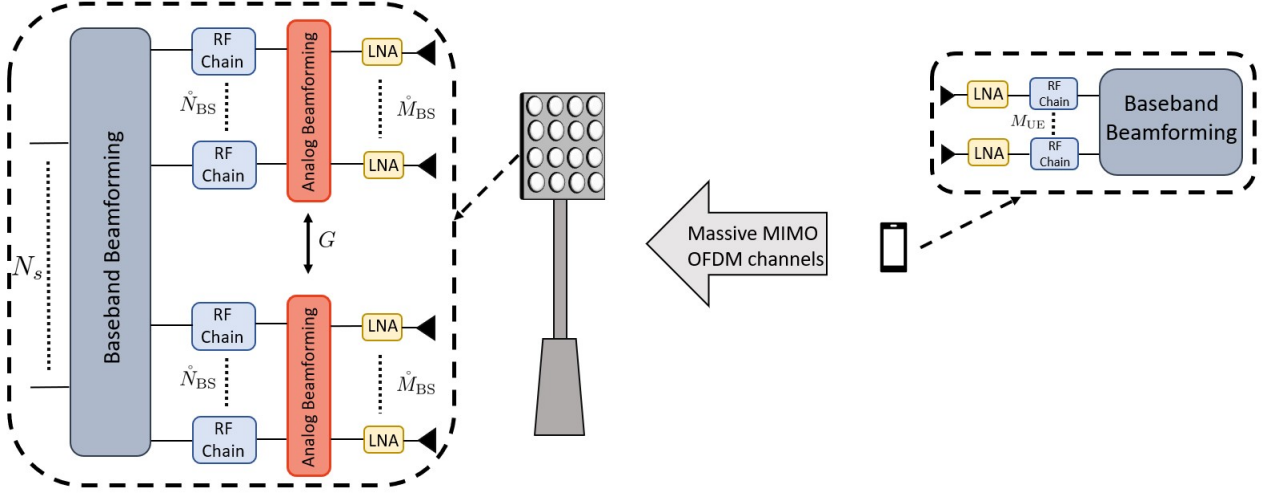


Figure 2.1: Uplink single-user massive MIMO OFDM communication system.

where $\mathbf{H}^\ell(t) \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}$ denotes the ℓ th multipath channel matrix, $\tau_\ell \in [0, \text{DS}]$ represents the corresponding delay with the maximum delay DS, and $\delta(\cdot)$ is the Dirac function.

The frequency-domain channel response can be obtained by applying the Fourier transform as

$$\mathbf{H}(f, t) = \sum_{\ell=1}^L \mathbf{H}^\ell(t) \cdot e^{-j2\pi f \tau_\ell} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}. \quad (2.2)$$

Assuming that T_{sym} is the OFDM symbol period, then the discrete-time representation of (2.2) on the k th sub-carrier and during the n th symbol period can be written as

$$\mathbf{H}_k[n] = \mathbf{H}(k\Delta f, nT_{\text{sym}}) = \sum_{\ell=1}^L e^{-j2\pi(k\Delta f)\tau_\ell} \cdot \mathbf{H}^\ell[n] \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}, \quad (2.3)$$

where $\mathbf{H}^\ell[n] = \mathbf{H}^\ell(nT_{\text{sym}}) \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}$ represents the discrete-time representation of the ℓ th multipath channel matrix.

2.2.3. Signal Model

We assume a block-fading channel model as shown in Fig. 2.2, where the channel coherence time T_C is divided into $T_{\text{BS}} \cdot T_{\text{UE}}$ transmission time intervals (TTIs), i.e., $T_C = T_{\text{BS}} \cdot T_{\text{UE}}$, where every block has T_{UE} snapshots. The analog combining matrix at the i th block at the BS is denoted as $\bar{\mathbf{A}}_i \in \mathbb{C}^{M_{\text{BS}} \times N_{\text{BS}}}$, which

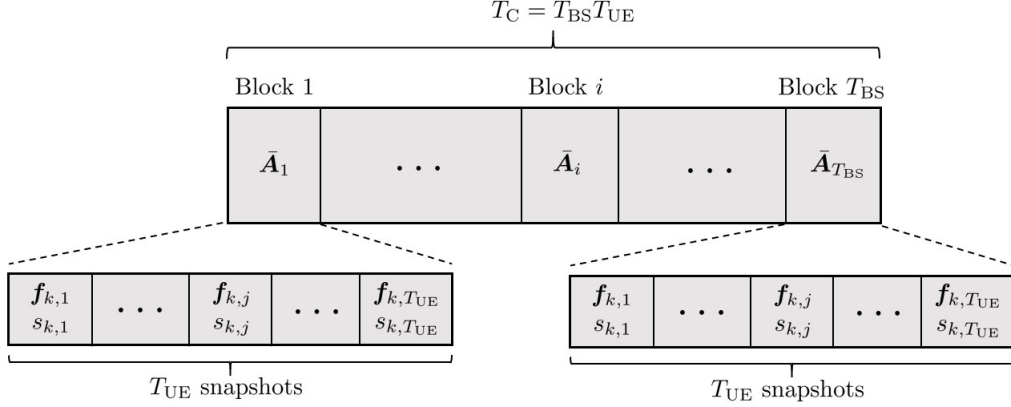


Figure 2.2: The transmission time interval (TTI) division for uplink transmission.

has a block-diagonal structure given as ²

$$\bar{\mathbf{A}}_i = \begin{bmatrix} \bar{\mathbf{A}}_{i,1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \bar{\mathbf{A}}_{i,G} \end{bmatrix} = \text{blkdiag}[\bar{\mathbf{A}}_{i,1}, \dots, \bar{\mathbf{A}}_{i,G}] \in \mathbb{C}^{M_{BS} \times N_{BS}}, \quad (2.4)$$

where $\bar{\mathbf{A}}_{i,g} \in \mathbb{C}^{M_{BS} \times N_{BS}}$ is the g th block-matrix with constant modulus entries, i.e., $|\bar{\mathbf{A}}_{i,g}[r,c]| = \frac{1}{\sqrt{M_{BS}}}$, where $\bar{\mathbf{A}}_{i,g}[r,c]$ is the (r, c) th entry of $\bar{\mathbf{A}}_{i,g}$.

The received signal by the BS in the (i, j) th TTI over the k th sub-carrier, with $i \in \{1, \dots, T_{BS}\}$, $j \in \{1, \dots, T_{UE}\}$, and $k \in \{1, \dots, K\}$, can be expressed as

$$\bar{\mathbf{y}}_{k,i,j} = \bar{\mathbf{A}}_i^H \mathbf{H}_k \mathbf{f}_{k,j} s_{k,j} + \bar{\mathbf{A}}_i^H \bar{\mathbf{n}}_{k,i,j} \in \mathbb{C}^{N_{BS}}, \quad (2.5)$$

where $s_{k,j} \in \mathbb{C}$ is the training symbol, $\mathbf{f}_{k,j} \in \mathbb{C}^{M_{UE}}$ is the (k, j) th precoding vector, $\bar{\mathbf{n}}_{k,i,j} \in \mathbb{C}^{M_{BS}}$ is the BS AWGN with zero-mean and variance σ_n^2 , and $\mathbf{H}_k \in \mathbb{C}^{M_{BS} \times M_{UE}}$ is the k th sub-carrier frequency-domain MIMO channel matrix. Initially, we collect the measurement vectors $\{\bar{\mathbf{y}}_{k,i,j}\}_{j=1}^{T_{UE}}$ next to each other as

$$\bar{\mathbf{Y}}_{k,i} = [\bar{\mathbf{y}}_{k,i,1}, \bar{\mathbf{y}}_{k,i,2}, \dots, \bar{\mathbf{y}}_{k,i,T_{UE}}], \quad (2.6)$$

which can be written as

$$\bar{\mathbf{Y}}_{k,i} = \bar{\mathbf{A}}_i^H \mathbf{H}_k \mathbf{F}_k + \bar{\mathbf{A}}_i^H \bar{\mathbf{N}}_{k,i} \in \mathbb{C}^{N_{BS} \times T_{UE}}, \quad (2.7)$$

where $\mathbf{F}_k = [\mathbf{f}_{k,1} s_{k,1}, \dots, \mathbf{f}_{k,T_{UE}} s_{k,T_{UE}}] \in \mathbb{C}^{M_{UE} \times T_{UE}}$ and $\bar{\mathbf{N}}_{k,i} = [\bar{\mathbf{n}}_{k,i,1}, \dots, \bar{\mathbf{n}}_{k,i,T_{UE}}] \in \mathbb{C}^{M_{BS} \times T_{UE}}$. We

²Note that if the number of groups $G = 1$, the above analog structure coincides with the fully connected (FC) analog structure, where every RF chain is connected to every antenna element. On the other hand, if $G = N_{BS}$, the above analog structure coincides with the partially connected (PC) analog structure, where every RF chain is connected to a unique subset of antenna elements [Ard+18].

assume that the $\mathbf{F}_k, \forall k$, are designed with orthogonal rows, i.e., $\mathbf{F}_k \mathbf{F}_k^H = \mathbf{I}_{M_{\text{UE}}}, \forall k$, and $T_{\text{UE}} \geq M_{\text{UE}}$. After applying the right filtering to (2.7) we obtain

$$\mathbf{Y}_{k,i} = \bar{\mathbf{Y}}_{k,i} \mathbf{F}_k^H = \bar{\mathbf{A}}_i^H \mathbf{H}_k + \mathbf{N}_{k,i} \in \mathbb{C}^{N_{\text{BS}} \times M_{\text{UE}}}, \quad (2.8)$$

where $\mathbf{N}_{k,i} = \bar{\mathbf{A}}_i^H \bar{\mathbf{N}}_{k,i} \mathbf{F}_k^H$. Next, we collect the measurement matrices $\{\mathbf{Y}_{k,i}\}_{i=1}^{T_{\text{BS}}}$ on the top of each other as

$$\mathbf{Y}_k = [\mathbf{Y}_{k,1}^T, \dots, \mathbf{Y}_{k,T_{\text{BS}}}^T]^T, \quad (2.9)$$

which can be written as

$$\mathbf{Y}_k = \mathbf{A} \mathbf{H}_k + \mathbf{N}_k \in \mathbb{C}^{I \times M_{\text{UE}}}, \quad (2.10)$$

where $I = T_{\text{BS}} N_{\text{BS}}$, $\mathbf{A} = [\bar{\mathbf{A}}_1, \dots, \bar{\mathbf{A}}_{T_{\text{BS}}}]^H \in \mathbb{C}^{I \times M_{\text{BS}}}$, and $\mathbf{N}_k = [\mathbf{N}_{k,1}^T, \dots, \mathbf{N}_{k,T_{\text{BS}}}^T]^T \in \mathbb{C}^{I \times M_{\text{UE}}}$. After that, we collect the measurement matrices $\{\mathbf{Y}_k\}_{k=1}^K$ next to each other as

$$\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_K], \quad (2.11)$$

which can be expressed as

$$\mathbf{Y} = \mathbf{A} \mathbf{H} + \mathbf{N} \in \mathbb{C}^{I \times K M_{\text{UE}}}, \quad (2.12)$$

where $\mathbf{N} = [\mathbf{N}_1, \dots, \mathbf{N}_K]$ and $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_K] \in \mathbb{C}^{M_{\text{BS}} \times K M_{\text{UE}}}$ is the total MIMO OFDM channel matrix. In the following, our aim is to estimate the channel matrix \mathbf{H} from the measurements in (2.12).

2.3. Algorithms

2.3.1. The Baseline LS-based Channel Estimation Algorithm

Given the measurement matrix in (2.12), a LS-based method can be used to obtain an estimate of the total MIMO channel matrix as [BG06]

$$\hat{\mathbf{H}}_{\text{LS}} = \mathbf{A}^+ \mathbf{Y} = [\hat{\mathbf{H}}_1, \dots, \hat{\mathbf{H}}_K] \in \mathbb{C}^{M_{\text{BS}} \times K M_{\text{UE}}}, \quad (2.13)$$

where $\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ is the Moore-Penrose pseudo-inverse of \mathbf{A} . Note that, due to the left filtering, the LS-based method requires that $I \geq M_{\text{BS}}$, i.e., $T_{\text{BS}} \geq \frac{M_{\text{BS}}}{N_{\text{BS}}}$ to provide an accurate channel estimate, which becomes impractical for massive MIMO systems.

2.3.2. The Baseline LMMSE-based Channel Estimation Algorithm

The LMMSE estimator of \mathbf{H} can be written as [BG06]

$$\hat{\mathbf{H}}_{\text{LMMSE}} = \mathbf{R}_H \mathbf{A}^H (\mathbf{A} \mathbf{R}_H \mathbf{A}^H + \sigma_n^2 \mathbf{I}_I)^{-1} \mathbf{Y} \quad (2.14)$$

where $\mathbf{R}_H = \mathbb{E}\{\mathbf{H}\mathbf{H}^H\}$ is the statistical channel correlation matrix of \mathbf{H} [AIU06; UV18], where \mathbf{H} is the concatenation of \mathbf{H}_k over all sub-carriers.

2.3.3. The Proposed SALSA Channel Estimation Algorithm

To obtain a more accurate channel estimate while reducing the training overhead, we propose here a novel channel estimation method called SALSA, which is derived by exploiting a hidden tensor structure in the measurement matrix in (2.12). The tensor representation not only contributes to a more accurate channel estimation but also helps to reduce the training overhead. Before showing the SALSA approach, we first recall the following propositions from [VP93; Wu+16; GMN18].

Proposition 2.3.1. Let $\mathbf{X} \in \mathbb{C}^{Q \times P}$ be a matrix given as

$$\mathbf{X} = \mathbf{X}_1 \otimes \mathbf{X}_2 = \begin{bmatrix} \mathbf{X}_{1,1} & \cdots & \mathbf{X}_{1,P_1} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{Q_1,1} & \cdots & \mathbf{X}_{Q_1,P_1} \end{bmatrix} \in \mathbb{C}^{Q \times P}, \quad (2.15)$$

where $\mathbf{X}_1 \in \mathbb{C}^{Q_1 \times P_1}$, $\mathbf{X}_2 \in \mathbb{C}^{Q_2 \times P_2}$, $Q = Q_1 Q_2$, $P = P_1 P_2$, and $\mathbf{X}_{a,b} = [\mathbf{X}_1]_{[a,b]} \mathbf{X}_2$ is the (a, b) th block-matrix of \mathbf{X} . Let $\mathbf{K} \in \mathbb{C}^{Q_1 P_1 \times Q_2 P_2}$ be a rank-one matrix given as

$$\mathbf{K} = \begin{bmatrix} \mathbf{x}_{1,1}^\top \\ \cdots \\ \mathbf{x}_{Q_1,1}^\top \\ \cdots \\ \mathbf{x}_{1,P_1}^\top \\ \cdots \\ \mathbf{x}_{Q_1,P_1}^\top \end{bmatrix} = \mathbf{x}_1 \mathbf{x}_2^\top, \quad (2.16)$$

where $\mathbf{x}_{a,b} = \text{vec}\{\mathbf{X}_{a,b}\} = [\mathbf{X}_1]_{[a,b]} \text{vec}\{\mathbf{X}_2\}$, $\mathbf{x}_1 = \text{vec}\{\mathbf{X}_1\} \in \mathbb{C}^{Q_1 P_1}$, and $\mathbf{x}_2 = \text{vec}\{\mathbf{X}_2\} \in \mathbb{C}^{Q_2 P_2}$. Since \mathbf{K} is rank-one, its SVD can be written as

$$\mathbf{K} = \sigma \mathbf{u} \mathbf{v}^H \quad (2.17)$$

where $\mathbf{u} \in \mathbb{C}^{Q_1 P_1}$ and $\mathbf{v} \in \mathbb{C}^{Q_2 P_2}$ are the left singular vector and the right singular vector of \mathbf{K} , respectively, associated with only the non-zero singular value σ . Then, an optimal solution to

$$\underset{\mathbf{X}_1, \mathbf{X}_2}{\text{minimize}} \|\mathbf{X} - (\mathbf{X}_1 \otimes \mathbf{X}_2)\|_{\text{F}}^2 \quad (2.18)$$

can be obtained as

$$\mathbf{X}_1 = \sqrt{\sigma} \cdot \text{unvec}_{Q_1 \times P_1} \{\mathbf{u}\} \in \mathbb{C}^{Q_1 \times P_1} \quad (2.19)$$

$$\mathbf{X}_2 = \sqrt{\sigma} \cdot \text{unvec}_{Q_2 \times P_2} \{\mathbf{v}^*\} \in \mathbb{C}^{Q_2 \times P_2}. \quad (2.20)$$

Proof: For more details, please refer to [Wu+16].

Proposition 2.3.2. For any given $Q \times P$ matrix \mathbf{X} , it can be approximately written as a summation of $R \geq 1$

factor matrices as

$$\mathbf{X} \approx \sum_{r=1}^R \mathbf{X}_r = \sum_{r=1}^R \mathbf{X}_{1,r} \otimes \mathbf{X}_{2,r}, \quad (2.21)$$

where $\mathbf{X}_r = \mathbf{X}_{1,r} \otimes \mathbf{X}_{2,r}$, $\mathbf{X}_{1,r} \in \mathbb{C}^{Q_1 \times P_1}$, and $\mathbf{X}_{2,r} \in \mathbb{C}^{Q_2 \times P_2}$ with $Q = Q_1 Q_2$ and $P = P_1 P_2$.

Proof: The proof follows directly by applying Proposition 2.3.1 sequentially [GMN18]. In Algorithm 2.1, we summarize how to find the corresponding factor matrices \mathbf{X}_1 and \mathbf{X}_2 as described in [GMN18].

Algorithm 2.1 Sequential Kronecker Factorization

- 1: **Input:** A matrix $\mathbf{X} \in \mathbb{C}^{Q \times P}$
 - 2: Select R, Q_1, P_1, Q_2, P_2 such that $Q = Q_1 Q_2$ and $P = P_1 P_2$
 - 3: **for** $r = 1$ to R **do**
 - 4: Get $\mathbf{X}_r = \mathbf{X} - \sum_{r'=1}^{r-1} \mathbf{X}_{1,r'} \otimes \mathbf{X}_{2,r'}$
 - 5: Given \mathbf{X}_r , get $\mathbf{X}_{1,r}$ and $\mathbf{X}_{2,r}$ using (2.19) and (2.20), respectively
 - 6: **end for**
 - 7: **Output:** $\hat{\mathbf{X}} = \sum_{r=1}^R \mathbf{X}_{1,r} \otimes \mathbf{X}_{2,r} \in \mathbb{C}^{Q \times P}$
-

Next, we turn our attention back to estimating the MIMO OFDM channel. Let $Q = M_{\text{BS}}$ and $P = K M_{\text{UE}}$. Then, from Proposition 2.3.2, the total frequency-domain massive MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ in (2.12) can be *approximately* written as

$$\mathbf{H} \approx \sum_{r=1}^R \mathbf{C}_r \otimes \mathbf{B}_r \in \mathbb{C}^{Q \times P} \quad (2.22)$$

where $\mathbf{B}_r \in \mathbb{C}^{Q_1 \times P_1}$, $\mathbf{C}_r \in \mathbb{C}^{Q_2 \times P_2}$, $Q = Q_1 Q_2$, and $P = P_1 P_2$. The exact choice of the parameters, i.e., P_1, P_2, Q_1, Q_2 , and R , will be discussed later in Section 2.4. Let $I = T_{\text{BS}} N_{\text{BS}}$. Then, by substituting (2.22) into (2.12) and assuming that R is sufficiently large, we can write

$$\mathbf{Y} = \mathbf{A} \left(\sum_{r=1}^R \mathbf{C}_r \otimes \mathbf{B}_r \right) + \mathbf{N} = \sum_{r=1}^R \mathbf{A}(\mathbf{C}_r \otimes \mathbf{B}_r) + \mathbf{N} = \sum_{r=1}^R \mathbf{Y}_r + \mathbf{N} \in \mathbb{C}^{I \times P}, \quad (2.23)$$

where $\mathbf{Y}_r = \mathbf{A}(\mathbf{C}_r \otimes \mathbf{B}_r) \in \mathbb{C}^{I \times P}$. Note that \mathbf{Y}_r can be seen as the 1-mode unfolding of a 3-way Tucker tensor given as

$$\mathbf{y}_r = \mathcal{S} \times_1 \mathbf{A} \times_2 \mathbf{B}_r^\top \times_3 \mathbf{C}_r^\top \in \mathbb{C}^{I \times P_1 \times P_2}, \quad (2.24)$$

where $\mathcal{S} \in \mathbb{Z}^{Q \times Q_1 \times Q_2}$ denotes the core-tensor with the 1-mode unfolding given as $[\mathcal{S}]_{(1)} \stackrel{\text{def}}{=} \mathbf{I}_Q$. The ℓ -mode unfoldings of $\mathbf{y}_r, \ell = \{1, 2, 3\}$, can be written as

$$[\mathbf{y}_r]_{(1)} = \mathbf{A}[\mathcal{S}]_{(1)}(\mathbf{C}_r \otimes \mathbf{B}_r) \in \mathbb{C}^{I \times P}, \quad (2.25)$$

$$[\mathbf{y}_r]_{(2)} = \mathbf{B}_r^\top [\mathcal{S}]_{(2)}(\mathbf{C}_r \otimes \mathbf{A}^\top) \in \mathbb{C}^{P_1 \times I P_2}, \quad (2.26)$$

$$[\mathbf{y}_r]_{(3)} = \mathbf{C}_r^\top [\mathcal{S}]_{(3)}(\mathbf{B}_r \otimes \mathbf{A}^\top) \in \mathbb{C}^{P_2 \times I P_1}. \quad (2.27)$$

2. Channel Estimation

From (2.24), the 3-way Tucker tensor form of (2.23) can be expressed as

$$\mathcal{Y} = \sum_{r=1}^R \mathcal{Y}_r + \mathcal{N} \in \mathbb{C}^{I \times P_1 \times P_2}, \quad (2.28)$$

where \mathcal{N} denotes the 3-way tensor representation of the noise matrix N . This latter formulation suggests that the factor matrices $\{\mathbf{B}_r, \mathbf{C}_r\}_{r=1}^R$ can be estimated sequentially as follows. Let \mathcal{Y}_r be the tensor obtained at the r th sequential step as

$$\mathcal{Y}_r = \mathcal{Y} - \sum_{r'=1}^{r-1} \mathcal{Y}_{r'} \in \mathbb{C}^{I \times P_1 \times P_2}. \quad (2.29)$$

Then, by exploiting the 2-mode and the 3-mode unfoldings of \mathcal{Y}_r , the r th factor matrices \mathbf{B}_r and \mathbf{C}_r can be obtained sequentially using, e.g., the ALS method [CLA09], where one factor matrix is assumed to be fixed when solving for the other. Remember that \mathbf{A} represents the beamforming at the BS which is known, then \mathbf{B}_r and \mathbf{C}_r can be obtained as

$$\mathbf{B}_r^\top = [\mathcal{Y}_r]_{(2)} \Psi_2^+ = [\mathcal{Y}_r]_{(2)} \Psi_2^H (\Psi_2 \Psi_2^H)^{-1} \quad (2.30)$$

$$\mathbf{C}_r^\top = [\mathcal{Y}_r]_{(3)} \Psi_3^+ = [\mathcal{Y}_r]_{(3)} \Psi_3^H (\Psi_3 \Psi_3^H)^{-1}, \quad (2.31)$$

where Ψ_2 and Ψ_3 are given as

$$\Psi_2 = [\mathcal{S}]_{(2)} (\mathbf{C}_r \otimes \mathbf{A}^\top) \in \mathbb{C}^{Q_1 \times IP_2} \quad (2.32)$$

$$\Psi_3 = [\mathcal{S}]_{(3)} (\mathbf{B}_r \otimes \mathbf{A}^\top) \in \mathbb{C}^{Q_2 \times IP_1}. \quad (2.33)$$

A summary of the proposed SALSA method for estimating the total massive MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ is given in Algorithm 2.2. In the next Section 2.4, we show some simulation results evaluating the convergence behavior of Algorithm 2.2.

Algorithm 2.2 SALSA For MIMO-OFDM Channel Estimation.

- 1: **Input:** Measurement matrix $\mathbf{Y} \in \mathbb{C}^{I \times P}$ as in (2.12)
 - 2: Select $R \geq 1$, $N_{\max\text{-iter}} \geq 1$, Q_1, Q_2, P_1 , and P_2 such that $Q = Q_1 Q_2 = M_{\text{BS}}$ and $P = P_1 P_2 = M_{\text{UE}} K$
 - 3: Obtain the 3-way Tucker tensor \mathcal{Y} in (2.28) from \mathbf{Y}
 - 4: **for** $r = 1$ to R **do**
 - 5: Get $\mathcal{Y}_r = \mathcal{Y} - \sum_{r'=1}^{r-1} \hat{\mathcal{Y}}_{r'}$
 - 6: Initialize $\mathbf{C}_r^{(0)} \in \mathbb{C}^{Q_2 \times P_2}$, e.g., randomly
 - 7: **for** $n = 1$ to $N_{\max\text{-iter}}$ **do**
 - 8: Get $\mathbf{B}_r^{(n)}$ using (2.30) for given $\mathbf{C}_r^{(n-1)}$
 - 9: Get $\mathbf{C}_r^{(n)}$ using (2.31) for given $\mathbf{B}_r^{(n)}$
 - 10: **end for**
 - 11: Set $\hat{\mathbf{B}}_r = \mathbf{B}_r^{(N_{\max\text{-iter}})}$ and $\hat{\mathbf{C}}_r = \mathbf{C}_r^{(N_{\max\text{-iter}})}$
 - 12: Get $\hat{\mathcal{Y}}_r = \mathcal{S} \times_1 \mathbf{A} \times_2 \hat{\mathbf{B}}_r^\top \times_3 \hat{\mathbf{C}}_r^\top$, go back to Step (5)
 - 13: **end for**
 - 14: **Output:** $\hat{\mathbf{H}}_{\text{SALSA}} = \sum_{r=1}^R \hat{\mathbf{C}}_r \otimes \hat{\mathbf{B}}_r \in \mathbb{C}^{Q \times P}$
-

Note that, due to the right filtering in (2.30) and (2.31), the SALSA method in Algorithm 2.2 requires that (C1) $Q_1 \leq IP_2$ and (C2) $Q_2 \leq IP_1$, where $I = T_{\text{BS}}N_{\text{BS}}$, i.e., $T_{\text{BS}} \geq \min\{\frac{Q_1}{N_{\text{BS}}P_2}, \frac{Q_2}{N_{\text{BS}}P_1}\}$ to provide an accurate channel estimation. Therefore, under practical settings, the SALSA method in Algorithm 2.2 requires less training overhead than the LS method in (2.13). On the other hand, assuming that the complexity of calculating the Moore-Penrose pseudo-inverse of an $a \times b$ matrix is on the order of $\mathcal{O}(\min\{a, b\}^3)$, where \mathcal{O} denotes the Big O notation, then the complexity of the LS method in (2.13) is on the order of $\mathcal{O}(\min\{I, P\}^3)$, while for the SALSA method in Algorithm 2.2 the complexity is on the order of $\mathcal{O}(R \cdot N_{\text{max-iter}} \cdot Q_1^3 \cdot Q_2^3)$, assuming that the conditions (C1) and (C2) are satisfied. Moreover, for LMMSE in (2.14), the complexity is on the order of $\mathcal{O}(I^3)$, assuming that the complexity of calculating the inverse of an $a \times a$ matrix is on the order of $\mathcal{O}(a^3)$.

2.4. Simulation Results

Table 2.1: System Parameters to generate CDL channel tensor $\mathcal{H} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}} \times N_{\text{taps}}}$.

System Parameter	Value
Scenario	UMi
Cell radius	100 m
BS height	10 m
UE height	1.5 m
Carrier frequency f_c	4 GHz
Sampling frequency f_s	30.72 MHz
No. of sub-carriers K	1 or 16
No. of antennas at BS M_{BS}	16 or 64 ($M_{\text{BS}}^v \times M_{\text{BS}}^h$)
No. of antennas at UE M_{UE}	4 or 64 ($M_{\text{UE}}^v \times M_{\text{UE}}^h$)
Polarization	Single

Performance of the Sequential Kronecker Factorization in Algorithm 2.1

Initially, we show simulation results in Fig. 2.3 analyzing the performance and the convergence behavior of the Sequential Kronecker Factorization in Algorithm 2.1. This analysis in Fig. 2.3 and the observations are important to understand and explain the obtained results of the proposed SALSA method in Algorithm 2.2. As it is discussed in Section 2.3, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$

2. Channel Estimation

can be *approximated* using the sum of the Kronecker products as given in (2.22). To gain more insights, we consider different cases, where in each case we assume that the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ is modeled differently as described in the sequel.

- **Case 1 (Random):** In this case, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ with $Q = P = 64$ is generated randomly from a circularly-symmetric complex Gaussian distribution with unit-variance and zero-mean, i.e., $\mathbf{H}_{[n,m]} \sim \mathcal{CN}(0, 1), \forall n, m$.
- **Case 2 (DFT):** Here, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ with $Q = P = 64$ is simply given by the normalized $[64 \times 64]$ DFT matrix, i.e., $\mathbf{H} = \left(\frac{\omega^{mn}}{\sqrt{64}}\right), m, n \in \{0, \dots, 63\}, \omega = e^{-2\pi j/Q}$, and $j^2 = -1$.
- **Case 3 (A multi-subcarrier 3GPP CDL channel model with URAs):** In this case, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ is generated according to the 3GPP clustered delay line (CDL) channel model described in TR 38.901, which is given in the Subsection 2.2.2. More details on the step-by-step instructions, as well as the Matlab scripts for channel generation, are available in [3GP20a]. Specifically, in our simulation, we first generate a time-domain channel tensor $\mathcal{H} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}} \times N_{\text{taps}}}$, where N_{taps} represents the number of time-domain channel taps calculated according to [RDT22] and using the system parameters shown in Table 2.1. For this case, we assume $M_{\text{BS}} = 64$ and $M_{\text{UE}} = 4$ both with uniform rectangular arrays (URAs), i.e., $M_{\text{BS}}^{\text{v}} = 8$ and $M_{\text{BS}}^{\text{h}} = 8$ such that $M_{\text{BS}} = M_{\text{BS}}^{\text{v}} M_{\text{BS}}^{\text{h}}$, $M_{\text{UE}}^{\text{v}} = 2$ and $M_{\text{UE}}^{\text{h}} = 2$ such that $M_{\text{UE}} = M_{\text{UE}}^{\text{v}} M_{\text{UE}}^{\text{h}}$. Then, we perform a K -point FFT operation, with $K = 16$ in this case, along the third dimension for each receive-transmit antenna pair to obtain the frequency-domain channel tensor $\mathcal{H} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}} \times K}$, where the k th slice matrix, i.e., $\mathbf{H}_k = [\mathcal{H}]_{[:, :, k]} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}$ represents the k th subcarrier frequency-domain MIMO channel matrix. Eventually, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ is constructed by stacking the k th slice matrices of $\mathbf{H}_k \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}}}$ next to each other as $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_K] = [\mathcal{H}]_{(1)} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}} K}$, i.e., $Q = M_{\text{BS}} = 64$ and $P = M_{\text{UE}} K = 64$.
- **Case 4 (A single-subcarrier 3GPP CDL channel model with URAs):** In this case, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ is generated as described in Case 3, but assuming $M_{\text{UE}} = 64$ with $M_{\text{UE}}^{\text{v}} = 8$, $M_{\text{UE}}^{\text{h}} = 8$, $K = 1$, and the other parameters are the same (i.e., a single-subcarrier channel with $Q = M_{\text{BS}} = 64$, $P = M_{\text{UE}} = 64$).
- **Case 5 (A multi-subcarrier 3GPP CDL channel model with ULAs):** In this case, the total MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{P \times Q}$ is generated as described in Case 3, but assuming $M_{\text{BS}}^{\text{v}} = 1$, $M_{\text{BS}}^{\text{h}} = 64$, and the other parameters are the same (i.e., the BS has a ULA).

In Fig. 2.3, the performance of the Kronecker factorization method with each of the above cases is evaluated in terms of the mean squared error (MSE) that is defined as $\text{MSE} = \mathbb{E}\left\{\left\|\mathbf{H} - \sum_{r=1}^R \mathbf{C}_r \otimes \mathbf{B}_r\right\|_{\text{F}}^2\right\}$. From Fig. 2.3, we have the following observations.

- **Observation 1:** The Sequential Kronecker Factorization in Algorithm 2.1 can approximate *any* channel matrix and the approximation becomes tighter as the number of channel factor matrices R increases.
- **Observation 2:** The optimal number of Kronecker products, denoted in the figure by R_{opt} , required to approximate a channel matrix depends on the selected scenario of Q_1, Q_2, P_1 , and P_2 , where

$R_{\text{opt}} \approx \min\{P_1 Q_1, P_2 Q_2\}$. In other words, reducing the dimensions of one of the channel factor matrices, i.e., $\mathbf{B}_r \in \mathbb{C}^{Q_1 \times P_1}$ or $\mathbf{C}_r \in \mathbb{C}^{Q_2 \times P_2}$, reduces the value of R_{opt} .

- Observation 3: For Case 3 and Case 4, the convergence behavior as a function of the Kronecker products R (i.e., the convergence rate or slope) of Algorithm 2.1 depends on the dimensionality of the factor matrices $\mathbf{B}_r \in \mathbb{C}^{Q_1 \times P_1}$ and $\mathbf{C}_r \in \mathbb{C}^{Q_2 \times P_2}$. Specifically, *when the factor matrices \mathbf{C}_r have a smaller dimensionality than the factor matrices \mathbf{B}_r* , Algorithm 2.1 requires less Kronecker products R and converges with a faster rate. However, this is not true for Case 1 and Case 2, where the dimensionality of the factor matrices \mathbf{B}_r and \mathbf{C}_r can be swapped without affecting the convergence behavior of Algorithm 2.1. To illustrate this, consider, for example, the blue-colored solid and dashed lines, where solid lines represent the scenario $[Q_1, Q_2, P_1, P_2] = [8, 8, 64, 1]$, i.e., $\mathbf{B}_r \in \mathbb{C}^{8 \times 64}$ and $\mathbf{C}_r \in \mathbb{C}^{8 \times 1}, \forall r$, and dashed lines represent the scenario $[Q_1, Q_2, P_1, P_2] = [8, 8, 1, 64]$, i.e., $\mathbf{B}_r \in \mathbb{C}^{8 \times 1}$ and $\mathbf{C}_r \in \mathbb{C}^{8 \times 64}, \forall r$. For Case 1 and Case 2, both solid and dashed blue-colored lines have the same convergence behavior. However, for Case 3 and Case 4, the solid blue-colored lines converge faster than that of the dashed blue-colored lines. Note that this is also true for the green-colored and the black-colored scenarios.
- Observation 4: The convergence behavior as a function of the Kronecker products R (i.e., the convergence rate or slope) of Algorithm 2.1 depends on the input channel matrix characteristics. For example, for Case 1 and Case 2 (i.e., in case of a Random and DFT-based channel model), we can see that MSE stays constant as R increases until it reaches R_{opt} , after which the MSE approaches *almost* zero. However, in case of the other 3GPP CDL-based channel models (i.e., Case 3 to Case 5), we can see that MSE becomes smaller as R increases before it reaches R_{opt} . To illustrate this, we plot in Fig. 2.4 the singular values of the channel matrices used in Case 1, Case 2, and Case 3. From Fig. 2.4, we can see that for Case 1 (Random-based channel case), most of the singular values of the input channel matrices are above 1, i.e., most of the singular values are significant and contribute highly to the channel matrix. This is also true for Case 2 (DFT-based channel case), where all the 64 singular values of the 64×64 DFT matrix are equal to one. With these two cases, Algorithm 2.1 requires R to be almost equal to R_{opt} to have an accurate Kronecker factorization approximation. On the other hand, for Case 3 (3GPP CDL-based channel model), we can see that most of the singular values of the input channel matrices are below 1, except for the first 10 singular values, and the singular values decrease almost exponentially. This is different from the singular values in Case 1, where they decrease almost linearly. Therefore, for Case 3, Algorithm 2.1 requires less Kronecker products R to approximate the channel matrices to an acceptable accuracy such that the MSE is below 10^{-5} .
- Observation 5: The convergence behavior of Algorithm 2.1 depends, as well, on the antenna array structure. This can be seen by comparing, for example, Case 3 and Case 5, where the only difference between these two cases is that in Case 3 the BS has a URA composed of $M_{\text{BS}}^v = 8$ and $M_{\text{BS}}^h = 8$ elements, while in Case 5, the BS has a ULA composed of $M_{\text{BS}}^v = 1$ and $M_{\text{BS}}^h = 64$ elements. Clearly, the above observations in Case 3 for the parameters $[Q_1, Q_2, P_1, P_2] = [8, 8, 64, 1]$ ($\mathbf{B}_r \in \mathbb{C}^{8 \times 64}$ and $\mathbf{C}_r \in \mathbb{C}^{8 \times 1}$) are reversed in Case 5 for the parameters $[Q_1, Q_2, P_1, P_2] = [8, 8, 1, 64]$ ($\mathbf{B}_r \in \mathbb{C}^{8 \times 1}$ and $\mathbf{C}_r \in \mathbb{C}^{8 \times 64}$), i.e., in Case 5, Algorithm 2.1 requires the sum of less Kronecker products R and converges with a faster rate *when the factor matrices \mathbf{B}_r have a smaller dimensionality than the factor matrices \mathbf{C}_r* .

2. Channel Estimation

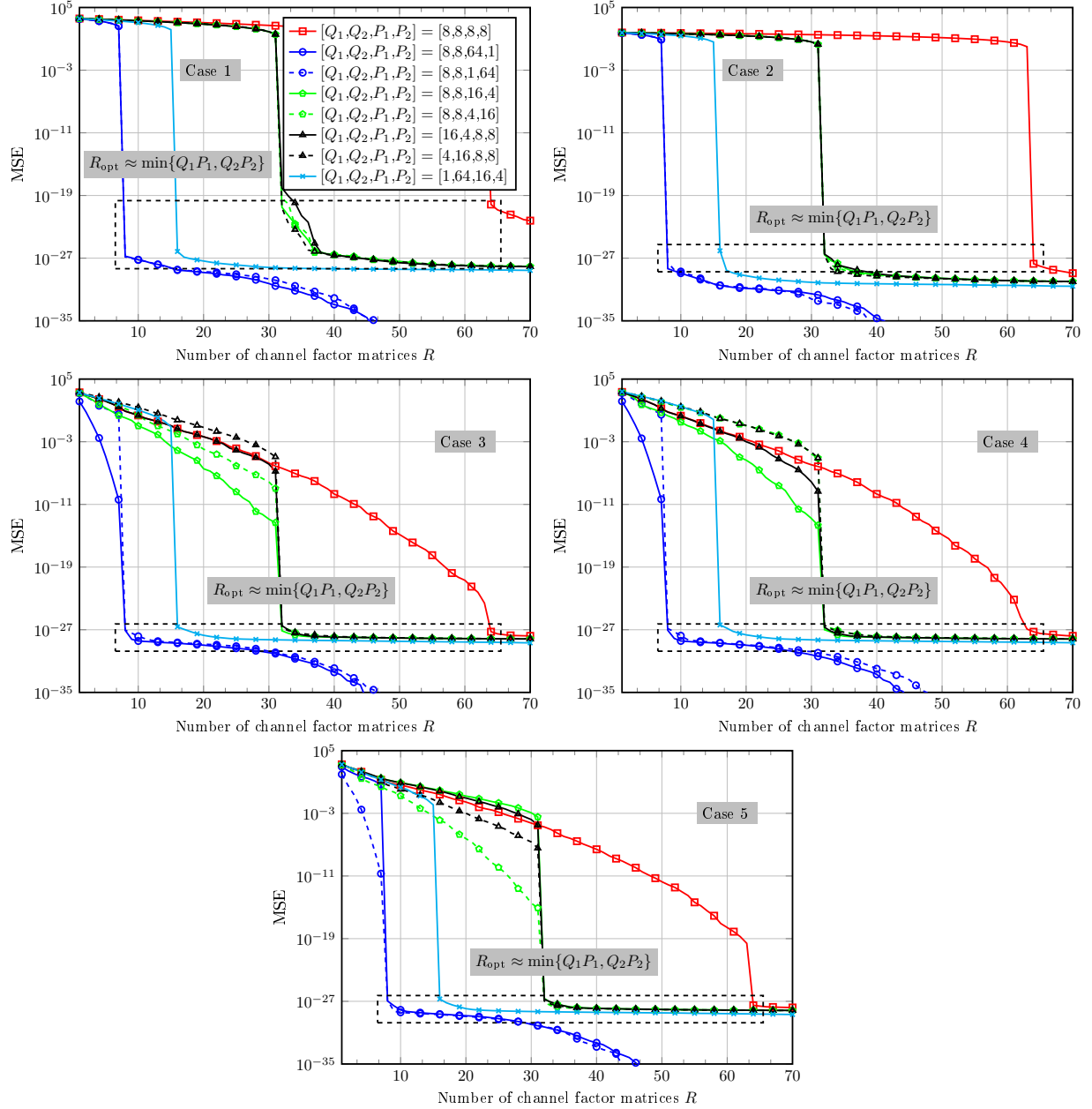


Figure 2.3: MSE vs. the sum of Kronecker products R .

From the above observations, we can expect that the performance of the proposed SALSA method depends not only on the dimension of the factor matrices, i.e., Q_1, Q_2, P_1, P_2 , and the number of the Kronecker products R , but also on the considered channel model and the antenna array structure. Unfortunately, the 3GPP CDL channel model described in TR 38.901 [3GP20a] is too complex to investigate a clear reason for the above listed observations. This, however, might be possible with a more simplified channel model, which we leave for future work.

In the remainder of this section, we study the performance of the proposed SALSA method com-

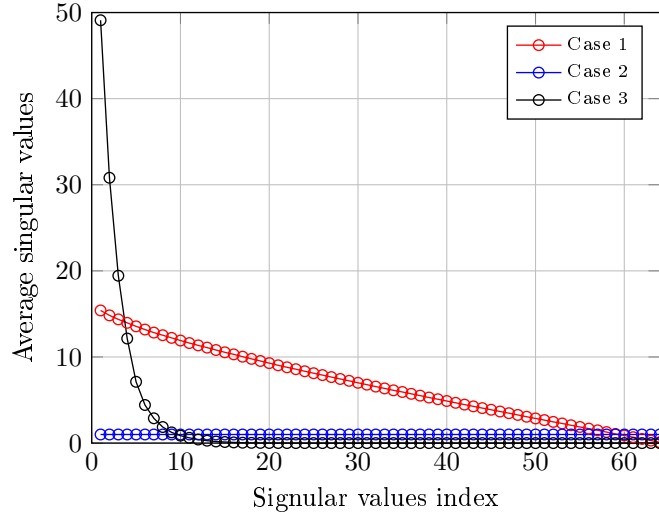


Figure 2.4: Average singular values of Case 1, Case 2, and Case 3 channel matrices.

pared to the LS and LMMSE methods assuming that the total MIMO channel is generated as illustrated in Case 3 above, unless stated otherwise. In all simulation scenarios, we set $M_{\text{UE}} = 4$, $K = 16$, $T_{\text{UE}} = M_{\text{UE}}$, $N_{\text{BS}} = 4$, $G = 2$. The analog decoding matrix $\mathbf{A} \in \mathbb{C}^{T_{\text{BS}} N_{\text{BS}} \times M_{\text{BS}}}$ is assumed to be generated randomly, where every non-zero entry is obtained as $a = 1/\sqrt{M_{\text{BS}}} \cdot e^{j\phi}$, where $\phi \in \mathcal{U}(0, 2\pi)$ and \mathcal{U} denotes the uniform distribution from 0 to 2π . We show the simulation results in terms of the normalized mean squared error (NMSE) that is defined as

$$\text{NMSE} = \frac{\mathbb{E}\{\|\mathbf{H} - \hat{\mathbf{H}}_{\text{X}}\|_{\text{F}}^2\}}{\mathbb{E}\{\|\mathbf{H}\|_{\text{F}}^2\}}, \quad (2.34)$$

where $\text{X} \in \{\text{LS}, \text{LMMSE}, \text{SALSA}\}$. Moreover, the SNR is defined as

$$\text{SNR} = \frac{\mathbb{E}\{\|\mathcal{Y} - \mathcal{N}\|_{\text{F}}^2\}}{\mathbb{E}\{\|\mathcal{N}\|_{\text{F}}^2\}}. \quad (2.35)$$

For the LMMSE method, we assume that the statistical channel correlation matrix, i.e., \mathbf{R}_{H} , is obtained as

$$\mathbf{R}_{\text{H}} = \mathbf{R}_{\text{H}}^{\text{true}} + \beta \cdot \mathbf{R}_{\text{H}}^{\text{noise}} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{BS}}}, \quad (2.36)$$

where $\mathbf{R}_{\text{H}}^{\text{true}}$ is the true statistical channel correlation matrix obtained as an average of $\Gamma = 1000$ channel realizations, i.e.,

$$\mathbf{R}_{\text{H}}^{\text{true}} = \frac{1}{\Gamma} \sum_{\gamma=1}^{\Gamma=1000} \mathbf{H}_{\gamma} \mathbf{H}_{\gamma}^{\text{H}} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{BS}}}, \quad (2.37)$$

with $\mathbf{H}_{\gamma} = [\mathbf{H}_{\gamma,1}, \dots, \mathbf{H}_{\gamma,K}] \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{UE}} K}$ being the γ th channel realization. Moreover, $\mathbf{R}_{\text{H}}^{\text{noise}}$ is an additive-noise covariance matrix and β is a noise power scaling parameter obtained such that

$$\beta = \frac{\|\mathbf{R}_{\text{H}} - \mathbf{R}_{\text{H}}^{\text{true}}\|_{\text{F}}^2}{\|\mathbf{R}_{\text{H}}^{\text{true}}\|_{\text{F}}^2}. \quad (2.38)$$

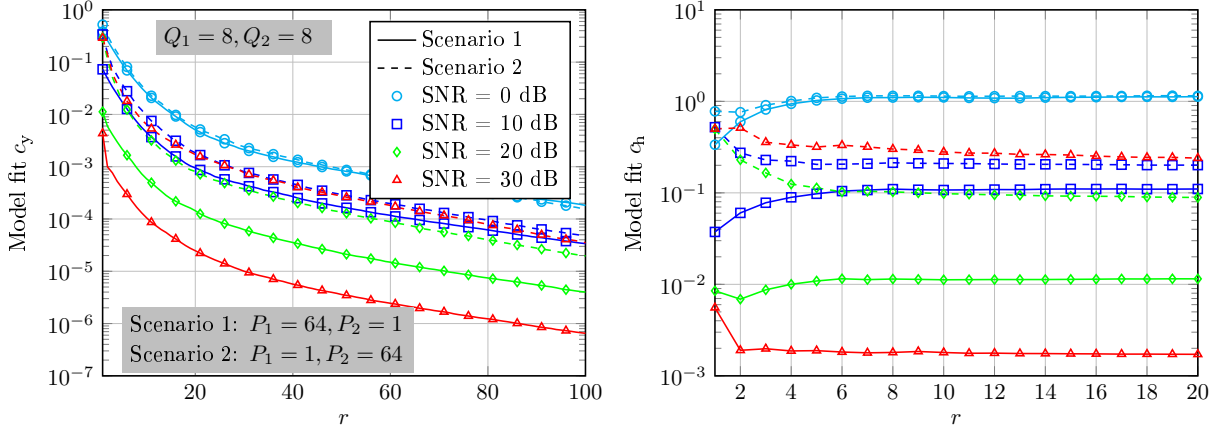


Figure 2.5: Convergence of the SALS method in Algorithm 2.2.

Note that, $\beta = 0$ can be considered as the ideal scenario, which is used in [AIU06] as a baseline approach for some \mathbf{R}_H estimation methods. On the other hand, for scenarios with $\beta > 0$, we demonstrate the impact of the estimation accuracy of \mathbf{R}_H on the estimation accuracy of the LMMSE method.

Convergence Behavior of Proposed SALS Method in Algorithm 2.2

As can be seen from Algorithm 2.2, SALS estimates a MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{Q \times P}$ sequentially from an input measurement tensor \mathcal{Y} in (2.28) by decomposing it into the sum of R Kronecker products. To this end, we need to run a maximum of R sequential steps. In the r th step, the factor matrices $\mathbf{B}_r \in \mathbb{C}^{Q_1 \times P_1}$ and $\mathbf{C}_r \in \mathbb{C}^{Q_2 \times P_2}$ are estimated using a bi-linear ALS method, which is designed to run up to a maximum of $N_{\max\text{-iter}}$ iterations, which we choose as $N_{\max\text{-iter}} = 100$. In Fig. 2.5, we show simulation results investigating the convergence behavior of Algorithm 2.2. We define the *model fit* functions c_y and c_h at the r th step as

$$c_y(r) = \frac{\|\mathcal{Y}_{(1)} - \mathbf{A} \sum_{r'=1}^r (\mathbf{C}_{r'} \otimes \mathbf{B}_{r'})\|_{\mathbb{F}}^2}{\|\mathcal{Y}_{(1)}\|_{\mathbb{F}}^2}, \quad (2.39)$$

$$c_h(r) = \frac{\|\mathbf{H} - \sum_{r'=1}^r (\mathbf{C}_{r'} \otimes \mathbf{B}_{r'})\|_{\mathbb{F}}^2}{\|\mathbf{H}\|_{\mathbb{F}}^2}. \quad (2.40)$$

From the left hand side of Fig. 2.5, we can see that Algorithm 2.2 seems to converge *monotonically* to at least a local optimal solution. Our observations show that these results hold true for all other choices of the dimension variables Q_1, Q_2, P_1 , and P_2 . Note that, Algorithm 2.2 has a faster convergence rate with Scenario 1 than it has with Scenario 2, which is, in fact, expected from our listed observations of Fig. 2.3. Moreover, it is worth mentioning that in a different implementation of Algorithm 2.2, the *model fit* function c_y can be used as a convergence criterion, e.g., if $c_y(r)$ or $|c_y(r) - c_y(r-1)|$ becomes below a certain threshold. On the other hand, we can see that the *model fit* function c_h does not necessarily have a monotonically decreasing behavior. In fact, we can observe that the *optimal or best* R depends on the SNR level, which increases by increasing the SNR level. This observation will be discussed in more

details in conjunction with Fig. 2.10 next.

Table 2.2: Dimension of the factor matrices in the Kronecker product $Q_1, Q_2, P_1,$ and P_2 for $Q = Q_1Q_2$ and $P = P_1P_2$.

Scenario No.	Q_1 and Q_2 values	P_1 and P_2 values
Scenario 1	$[Q_1, Q_2] = [64, 1]$	$[P_1, P_2] = [64, 1]$
\vdots		\vdots
Scenario 7		$[P_1, P_2] = [1, 64]$
\vdots	\vdots	\vdots
Scenario 43	$[Q_1, Q_2] = [1, 64]$	$[P_1, P_2] = [64, 1]$
\vdots		\vdots
Scenario 49		$[P_1, P_2] = [1, 64]$

Effect of Q and P Division Scenarios

Here, we show simulation results investigating the best parameter setting for Q and P with the constraints of $Q = Q_1Q_2, P = P_1P_2, Q_x \geq 1, P_x \geq 1,$ and Q_x, P_x are natural numbers, where $x \in \{1, 2\}$. Recall that $Q = M_{\text{BS}}$ and $P = M_{\text{UE}}K$. Therefore, we have $Q = P = 64$ and the candidate numbers of Q_x and P_x are 1, 2, 4, 8, 16, 32, and 64. Therefore, we have in total 49 possible scenarios as illustrated in Table 2.2, e.g., Scenario 1: $\{[Q_1, Q_2, P_1, P_2] = [64, 1, 64, 1]\}$ and Scenario 2: $\{[Q_1, Q_2, P_1, P_2] = [64, 1, 32, 2]\}$. We have simulated the SALSA algorithm using all the 49 possible scenarios. In Figs. 2.6 and 2.7, we show the NMSE versus SNR results for the different Q and P division scenarios as illustrated in Table 2.2.

From Fig. 2.6, when $T_{\text{BS}} = 12$, i.e., $I = T_{\text{BS}}N_{\text{BS}} = 48 < M_{\text{BS}}$, the analog training matrix $\mathbf{A} \in \mathbb{C}^{I \times M_{\text{BS}}}$, i.e., the first factor matrix of the measurement tensor in (2.24), is left non-invertible, i.e., $\mathbf{A}^+ \mathbf{A} \neq \mathbf{I}_{M_{\text{BS}}}$. Therefore, the LS-based algorithm has a very bad channel estimation accuracy. Differently, we can see that the LMMSE method provides the best channel estimation accuracy, thanks to the prior knowledge of the statistical channel correlation matrix \mathbf{R}_H and the noise variance σ_n^2 . On the other hand, we can see that the best NMSE of the SALSA method is achieved when $Q_1 = 8, Q_2 = 8, P_1 = 64,$ and $P_2 = 1$, i.e., when $\mathbf{B}_r \in \mathbb{C}^{8 \times 64}$ and $\mathbf{C}_r \in \mathbb{C}^{8 \times 1}, \forall r$. The main reason is that by dividing $Q = 64$ equally between Q_1 and Q_2 , i.e., $Q_1 = Q_2 = 8$, SALSA reduces the impact of the non-invertibility of \mathbf{A} by distributing it between the second (i.e., \mathbf{B}_r) and the third (i.e., \mathbf{C}_r) factor matrices of the measurement tensor, which leads to a better channel estimation accuracy. Furthermore, by setting $P_1 = 64$ and $P_2 = 1$, the required number of channel factor matrices R reduces as compared to the other division scenarios, as we have illustrated in Fig. 2.3 and the listed observations. Specifically, as we have noted in Observation 3, when the factor matrices \mathbf{C}_r have a smaller P than the factor matrices \mathbf{B}_r , i.e., $P_2 < P_1$, less Kronecker products R are required in the sum to obtain an acceptable approximation accuracy with a faster convergence rate. Given that the best dimension of Q_1 and Q_2 is 8, i.e., $Q_1 = 8, Q_2 = 8$, then

2. Channel Estimation

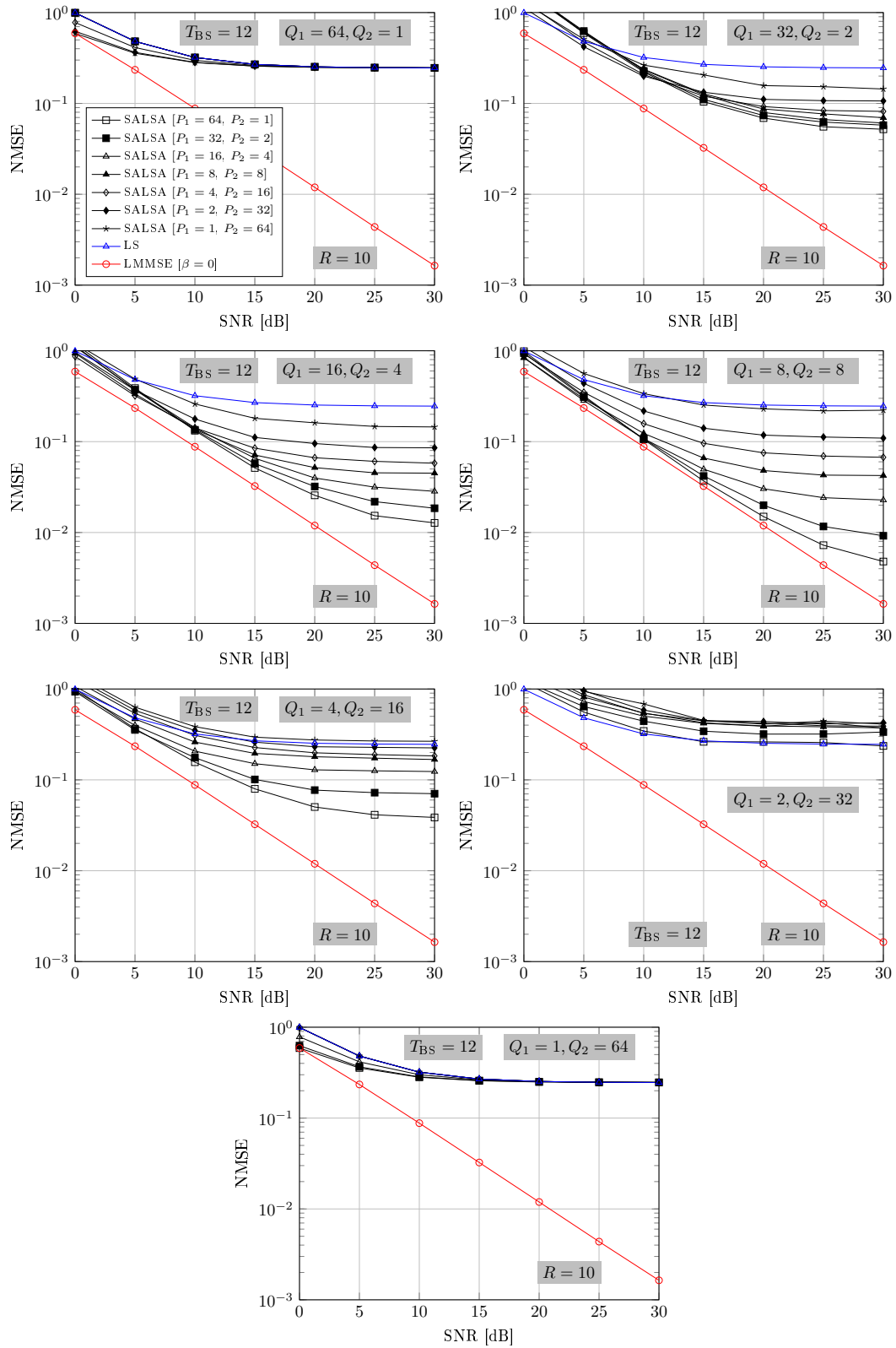


Figure 2.6: NMSE vs. SNR for different Q and P division scenarios, with $T_{BS} = 12$.

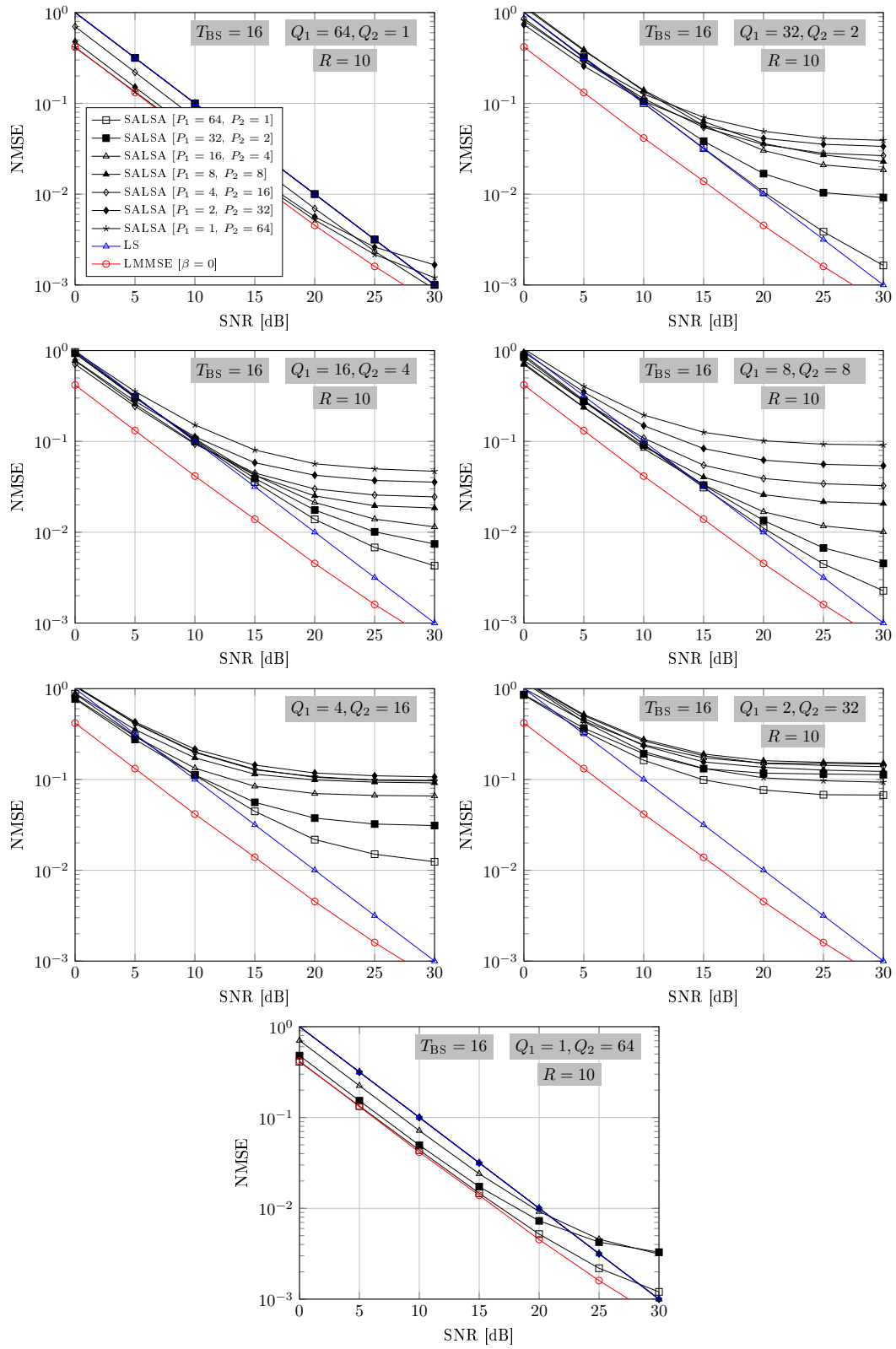


Figure 2.7: NMSE vs. SNR for different Q and P division scenarios, with $T_{BS} = 16$.

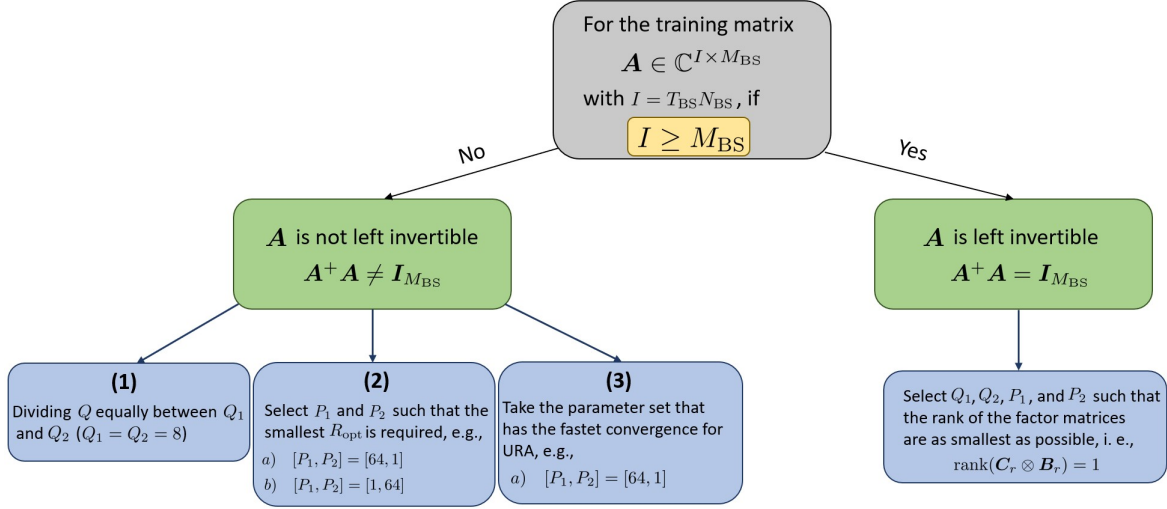


Figure 2.8: Decision making procedure in SALSA to select the dimensions of the factor matrices B_r and C_r depending on the training beamforming matrix A .

the best dimensions of P_1 and P_2 are $P_1 = 64$, and $P_2 = 1$. Therefore, the C_r become column vectors, i.e., $C_r \in \mathbb{C}^{8 \times 1}$, and the blocks B_r are as largest as possible, i.e., $B_r \in \mathbb{C}^{8 \times 64}$. Please refer to the blue-colored lines of Case 3 in Fig. 2.3.

Differently, in Fig. 2.7, when $T_{BS} = 16$, i.e., $I = M_{BS}$, the analog training matrix A is left-invertible, i.e., $A^+A = I_{M_{BS}}$. Therefore, the LS-based method has a high channel estimation accuracy. Moreover, we can see that the LMMSE-based method maintains its superiority, providing the best channel estimation accuracy. For SALSA, we have noticed that the best channel estimation accuracy is obtained with the scenario of $[Q_1, Q_2, P_1, P_2] = [1, 64, 64, 1]$, i.e., when $B_r \in \mathbb{C}^{1 \times 64}$ and $C_r \in \mathbb{C}^{64 \times 1}$ or with scenario of $[Q_1, Q_2, P_1, P_2] = [64, 1, 1, 64]$, i.e., when $B_r \in \mathbb{C}^{64 \times 1}$ and $C_r \in \mathbb{C}^{1 \times 64}$. With both of these scenarios, the channel matrix $H \in \mathbb{C}^{64 \times 64}$ in (2.22) is decomposed into a summation of R factor matrices $C_r \otimes B_r \in \mathbb{C}^{64 \times 64}$, each having rank-one, i.e.,

$$\text{rank}\{C_r \otimes B_r\} = \text{rank}\{C_r\} \cdot \text{rank}\{B_r\} = 1, \forall r,$$

according to (1.13), which leads to a better estimation accuracy and approaches the LMMSE-based method. The summary of the decision making procedure in SALSA by selecting the dimensional of the factor matrices B_r and C_r are illustrated in Fig. 2.8.

Next, in Fig. 2.9, we show the simulation results of the NMSE and the SNR to evaluate the SALSA performance considering different antenna array structures. In particular, we compare the performance of SALSA when the entire MIMO channel matrix is generated as explained in Case 3 and Case 5, where BS has a URA and a ULA, respectively. We see that when the BS has a ULA (case 5), SALSA performs best by assigning the factor matrices B_r a lower dimensionality than the factor matrices C_r . This was expected from Observation 5, since a smaller number of factor matrices R is required to approximate the channel matrix and the algorithm 2.1 has a faster convergence rate. However, if the BS is a URA

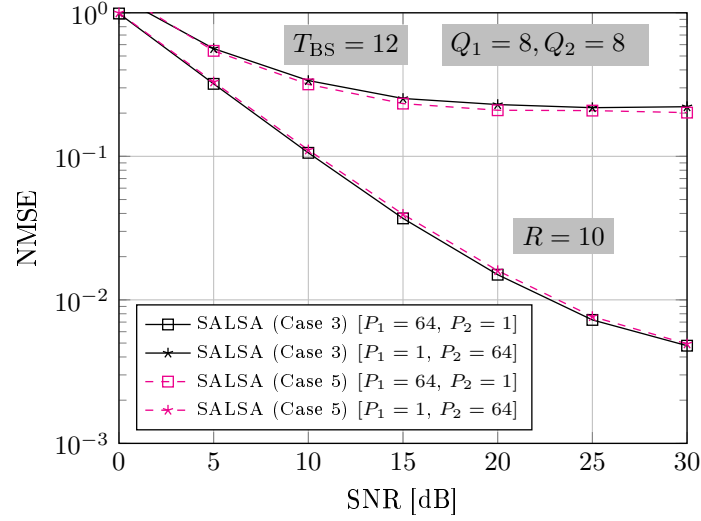


Figure 2.9: NMSE vs. SNR for different Q and P division scenarios and antenna array structures, with $T_{BS} = 12$.

(case 3), the best dimensionality mapping is reversed, i.e., the factor matrices C_r should have a smaller dimension than the factor matrices B_r . Fig. 2.9 clearly shows that SALSAs perform comparably in both cases, but with opposite dimensions for P_1 and P_2 variables.

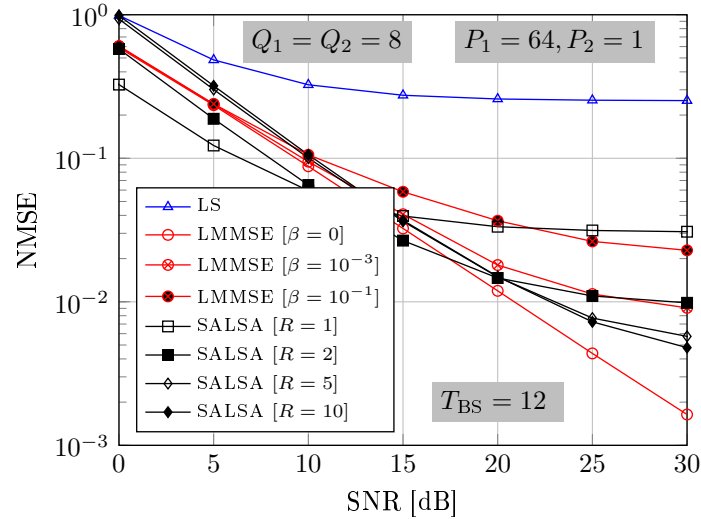


Figure 2.10: NMSE vs. SNR for different R and β .

Effect of the Number of Factor Matrices

In Fig. 2.10, we show the NMSE versus SNR results obtained by varying the number of channel factor matrices R and the noise power scaling parameter of \mathbf{R}_H , i.e., β . Clearly, Fig. 2.10 shows that the

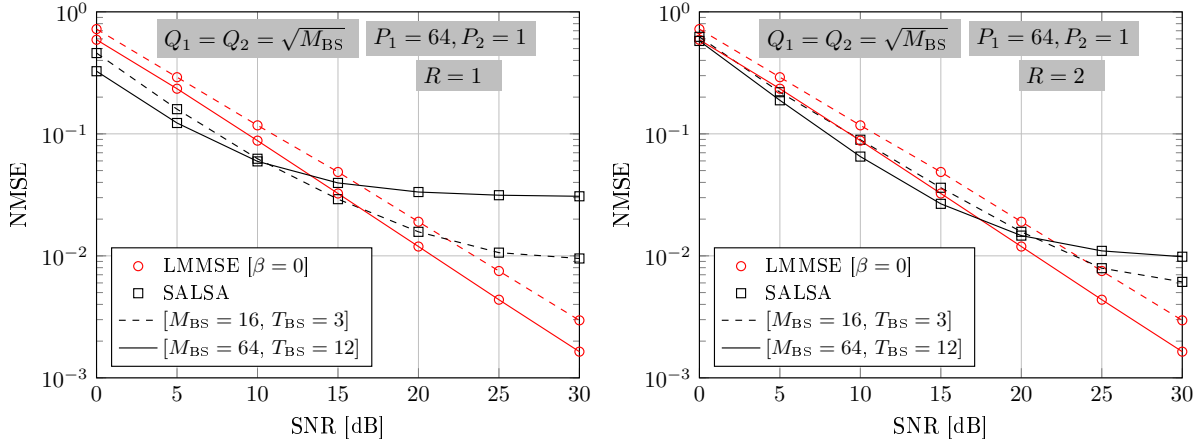


Figure 2.11: NMSE vs. SNR for different M_{BS} .

channel estimation accuracy of LMMSE decreases as β increases, since an increasing β imitates a decrease of the estimation accuracy of \mathbf{R}_H . On the other hand, Fig. 2.10 shows that the channel estimation accuracy of SALSAs increases with an increasing R in the high SNR region (i.e., $\text{SNR} \geq 15$ dB) while it decreases with an increasing R in the low SNR region (i.e., $\text{SNR} < 15$ dB). The main reason is that, in the high SNR region, where the impact of the noise is minimal, by increasing R we successively estimate more factor matrices belonging to the subspace of the true channel matrix. This leads to a better channel estimation accuracy, as we have illustrated in Fig. 2.3 for noiseless channel matrices. On the other hand, the channel measurement tensor is noise-limited in the low SNR region. Therefore, the sequentially estimated channel factor matrices become very noisy above a certain R as they mainly belong to the noise subspace. In other words, in this case, as R increases, the influence of the noise increases and the overall estimation accuracy decreases. Clearly, for every SNR level, there is an optimal R value (i.e., R_{opt}) wherein the channel estimation accuracy is maximized. We conjecture that R_{opt} is a function of the noise variance σ_n^2 .

Effect of the Number of Antennas

Finally, in Fig. 2.11, we show the NMSE versus SNR simulation results by varying the number of antenna elements at the BS, i.e., M_{BS} . For a fair comparison, we adjust the number of training blocks T_{BS} so that both considered scenarios have the same ratio of $\frac{I}{M_{BS}} = 0.75$ (recall that $I = T_{BS}N_{BS}$). From Fig. 2.11, we can observe that the SALSAs method has a similar behavior for both M_{BS} scenarios. In the low SNR region, the SALSAs method has a better channel estimation accuracy than LMMSE, while LMMSE has a better channel estimation accuracy than SALSAs in the high SNR region. However, the main difference is that with the smaller M_{BS} value (i.e., $M_{BS} = 16$ in the figure), the SALSAs method, with fixed R value, maintains its estimation accuracy superiority over LMMSE for a higher SNR level as compared to the scenario with the larger M_{BS} value (i.e., $M_{BS} = 64$ in the figure). This can be explained with reference to Fig. 2.3, which shows that the optimal R value (i.e., R_{opt}) is smaller with smaller factor matrices (i.e., when $M_{BS} = 16$, we have $[Q_1, Q_2] = [4, 4]$ and, therefore, $R_{\text{opt}} \approx 4$, while when $M_{BS} = 64$, we have $[Q_1, Q_2] = [8, 8]$ and, therefore, $R_{\text{opt}} \approx 8$. Note that $[P_1, P_2] = [64, 1]$ for both scenarios). It should be remembered that the performance of SALSAs in the high SNR region can be further improved

by increasing R , as it is also shown in Fig. 2.11.

2.5. Chapter Conclusion

In this chapter, we have proposed a novel channel estimation method for MIMO-OFDM sub-6 GHz wireless communication systems called SALSAs. We have shown that an accurate channel estimation can be obtained with a small training overhead by exploiting a hidden tensor structure in the received measurement matrix, which estimates the channel matrix sequentially using an ALS-based method. Our results show that the SALSAs method outperforms the conventional LS-based method, especially in the low training overhead. Moreover, it is shown that SALSAs, under some settings, outperforms LMMSE-based methods with an ideal knowledge of the statistical channel correlation matrix, in the low SNR region, while it approaches it in the high SNR region. Given that SALSAs does not require the prior knowledge of the statistical channel correlation matrix, as compared to LMMSE, renders it more appealing for practical implementations.

Hybrid Beamforming Design

3.1. Introduction and State of the Art

Precoding/beamforming is a ubiquitous and essential technique to reap the full potential of massive MIMO systems by focusing the radiated energy in a specific direction to maximize the SNR of the desired signal while reducing the inter-user and inter-beam interference. More antennas means the beam will be narrower, the beamforming gain will be greater, and the interference will be smaller. However, the conventional fully digital beamforming architectures, where one RF chain (ADC/DAC, data converters, mixers, etc.) is dedicated per antenna, becomes impractical due to their high hardware cost and energy-consumption [AKS17]. Therefore, other beamforming architectures have been proposed for massive MIMO systems such as analog and HAD beamforming architectures. On the one hand, analog beamforming is a very simple and cost-effective architecture, where the digital output of the RF chain is mapped to a large number of antennas via a network of PSs or switches [Wan+09; BBS13; Hur+13; Son+15]. However, the main idea of MIMO systems, i.e., spatial multiplexing, cannot be realized. To overcome this issue, HAD beamforming architectures have gained more interest, which split the beamforming operations between the analog and the digital domains to reduce the number of RF chains as compared to the number of antenna elements [GAH20; Aya+14; Ard+18; HMP20]. The HAD architectures are motivated by the fact that the number of RF chains is lower bounded by the number of transmitted data streams, while the beamforming gain and the diversity order depend on the number of antennas, considering a suitable RF beamforming design [ZMK05; Pal+06]. Depending on the mapping from the RF chains to the antenna elements, the hybrid beamforming architectures are categorized into mainly two types, (i) fully connected (FC) and (ii) partially connected (PC), where the mapping can be achieved using a network of PSs and/or switches [GAH20; Ard+18]. Specifically, in FC HAD architectures, each RF chain is connected to all antenna elements. Differently, each RF chain is connected to a subset of antenna elements in PC HAD architectures. Compared to the FC, PC HAD architectures sacrifice beamforming gain for less cost, size, complexity, and energy-consumption.

In [SY16], it is shown that if the number of RF chains is twice the number of data streams, the FC HAD architectures can provide the exact same performance of any fully digital beamforming scheme. Therefore, the objective of any hybrid beamforming architecture is to reduce the number of RF chains, e.g., to be equal to the number of data streams, while providing a performance that is close to the fully digital beamforming counterpart. However, the HAD beamforming design is a non-convex optimization problem and, in general, more complicated than the fully digital counterparts. This is mainly because of the joint design of the digital and analog beamforming matrices and the constant modulus and/or binary constraints owing to the use of PSs and/or switches in the analog domain [Aya+14; Ard+18]. Therefore, sub-optimal and approximate solutions have been widely adopted in the literature, e.g., by decoupling the optimization of the analog and digital parts and treating them separately [Aya+12; ZMK05; Ard+20a; LXD14; ND16; SY16; WLY18; Mén+16; Zha+18; Ard+18; Aya+14; SY16].

A widely used sub-optimal approach is the matrix factorization, where a known or a pre-computed

fully digital beamforming matrix is factorized into a hybrid analog and a digital matrix, e.g., by minimizing the Euclidean distance between them as proposed in [Aya+14; ZMK05]. However, this solution approach increases the computational complexity as it requires the calculation of the fully digital beamforming matrix beforehand. Another solution approach is a codebook-based analog beamforming design [SR15; KKS13], where the analog beamforming vectors are selected from a known codebook. Such an approach simplifies the analog beamforming design, since the codebook candidate vectors can be selected or designed while satisfying the analog domain hardware constraints, e.g., the constant modulus constraints. For example, the array response vectors of the MIMO channel matrix [Yu+16; LL14; KL15] or the DFT matrix [BBS13; Rus+15] can be considered for the codebook design, since they satisfy the constant modulus constraints. While the codebook-based approach simplifies the analog beamforming design, a trade-off between design complexity and the achievable performance arises, since increasing the number of candidate vectors not only improves the performance and the beamforming gain, but also increases the complexity of the selection of analog vectors. Finally, an alternating optimization technique is another widely used sub-optimal approach for HAD beamforming design, where the analog beamforming matrix is designed assuming that the digital beamforming matrix is fixed, and vice-versa, as proposed in [ZMK05; Ard+20a; LXD14; ND16; SY16; WLY18; Mén+16; Zha+18; Ard+18; Aya+14; SY16]. In general, a number of iterations, which is required to fine tune the HAD beamformers, increases the computational complexity.

Most of the initial works on HAD beamforming, e.g., [Aya+12; LXD14; ND16; SY16; WLY18; Mén+16; Zha+18] consider narrowband frequency-flat single-carrier massive MIMO systems, where their extension to wideband frequency-selective massive MIMO OFDM systems with multiple carriers is generally not straightforward, especially for PC architectures. For example, in [Ard+18], a unified analog beamforming design algorithm is proposed that is valid for both PS-based and switch-based architectures, which updates the analog matrices such that the equivalent channel's capacity is maximized. On the other hand, classical beamforming design methods like block diagonalization (BD) and ZF, for the multi-user scenarios, and maximum ratio transmission (MRT), for the single-user scenarios, are used to compute the digital part considering the resulting equivalent channels. Later on, the HAD beamforming design considering massive MIMO OFDM systems was considered in [KHY15; Gen+13; LW20; Zha+19; ZWH16; KCS17; SY17]. For example, in [LW20], a downlink multi-user multiple-input single-output (MISO) system was assumed, where the baseband beamforming matrix is designed using the classical ZF method, while the analog beamforming matrix is designed such that the approximated upper-bound of the system's SE is maximized. The solution proposed in [LW20] is mainly intended for FC architectures, while its simple adaptation to the PC architecture leads to a high-performance loss. In contrast, the authors in [Zha+19] propose an HAD beamforming design approach considering a FC downlink multi-user MIMO OFDM system, where the analog beamforming matrices are designed using a subspace matching approach exploiting the tensor structure of the propagation channels, while the baseband beamforming matrices are designed afterward as a solution to a signal-to-leakage-plus-noise ratio optimization problem. In [Dai+15], an alternating optimization-based HAD beamforming design approach called successive interference cancellation (SIC) is proposed, assuming a PC architecture. The baseband precoding matrix has a diagonal structure, with each diagonal element representing the power associated with the corresponding data stream. The main drawback of the proposed solutions in [Dai+15; Han+15a] is that the analog beamformer is mainly contributing to the beamforming gain, neglecting the inter-beam interference, which is obviously a sub-optimal strategy.

3.1.1. Chapter Contributions

In this chapter, we consider a frequency-selective downlink multi-user massive MIMO OFDM system. We propose a HAD beamforming design approach that is applicable to both FC and PC architectures. To this end, we assume that the baseband precoding matrices are designed using the classical ZF technique, while the baseband decoding matrices are designed using the MMSE approach [Chr+08]. Moreover, we derive an approximation of system's SE upper bound, which is used as the objective function of the analog precoding matrix optimization. To obtain a solution, we divide the problem into a series of convex sub-problems that are updated iteratively until convergence is obtained. The content of this chapter has been published in [GAH20].

3.1.2. Chapter Organization

We begin this chapter with a brief introduction of the system model of the downlink multi-user massive MIMO OFDM system and the problem formulation in Section 3.2. Then, we discuss the HAD beamforming design in Section 3.3, which is divided into two subsections. In Subsection 3.3.1, we focus on the design of baseband beamforming matrices. Assuming that the baseband beamforming matrices are designed as described in Subsection 3.3.1, we propose an analog beamforming design suitable for any HAD beamforming architecture in Subsection 3.3.2. Moreover, we explain in detail a baseline method for analog beamforming design in Subsection 3.3.2.1. The performance of the proposed beamforming algorithm is provided in Section 3.4. In the end, we provide a conclusion in Section 3.5.

3.2. System Model

We consider a downlink multi-user massive MIMO OFDM system, where a single BS communicates with U UEs. The BS is equipped with M_{BS} antennas and $N_{\text{BS}} \leq M_{\text{BS}}$ RF chains, while each UE is equipped with M_{UE} antennas, $N_{\text{UE}} \leq M_{\text{UE}}$ RF chains, and receives N_s data streams. In contrast to the previous chapter, we have HAD beamforming on both sides. A cyclic-prefix OFDM-based multi-carrier modulation scheme with K sub-carriers is applied to combat the multipath effect. Assuming that the cyclic-prefix length has the same length as the maximum excess delay of the channel, such that the ISI is avoided, the received signal on the k th sub-carrier, $k \in \{1, \dots, K\}$, at the u th UE, $u \in \{1, \dots, U\}$, is given as

$$\hat{\mathbf{s}}_{u,k} = \mathbf{W}_{u,k}^H \mathbf{H}_{u,k} \mathbf{F}_{u,k} \mathbf{s}_{u,k} + \mathbf{W}_{u,k}^H \mathbf{H}_{u,k} \sum_{u'=1, u' \neq u}^U \mathbf{F}_{u',k} \mathbf{s}_{u',k} + \mathbf{W}_{u,k}^H \mathbf{n}_{u,k} \in \mathbb{C}^{N_s}, \quad (3.1)$$

Where $\mathbf{H}_{u,k} \in \mathbb{C}^{M_{\text{UE}} \times M_{\text{BS}}}$ is the massive MIMO frequency domain channel, $\mathbf{F}_{u,k} \in \mathbb{C}^{M_{\text{BS}} \times N_s}$ denotes the precoding matrix, $\mathbf{W}_{u,k} \in \mathbb{C}^{M_{\text{UE}} \times N_s}$ is the decoding matrix, $\mathbf{s}_{u,k} \in \mathbb{C}^{N_s}$ is the data vector with $\mathbb{E}\{\mathbf{s}_{u,k} \mathbf{s}_{u,k}^H\} = \mathbf{I}_{N_s}$, and $\mathbf{n}_{u,k} \sim \mathcal{CN}(\mathbf{0}_{N_s}, \sigma_n^2 \mathbf{I}_{N_s})$ is the AWGN vector with variance σ_n^2 . We consider a HAD beamforming architecture at the BS and each UE such that $\mathbf{F}_{u,k} \stackrel{\text{def}}{=} \mathbf{F}_u^{\text{RF}} \mathbf{F}_{u,k}^{\text{BB}}$ and $\mathbf{W}_{u,k} \stackrel{\text{def}}{=} \mathbf{W}_u^{\text{RF}} \mathbf{W}_{u,k}^{\text{BB}}$ respectively. Here, $\mathbf{F}_u^{\text{RF}} \in \mathbb{C}^{M_{\text{BS}} \times N_{\text{BS}}}$ and $\mathbf{W}_u^{\text{RF}} \in \mathbb{C}^{M_{\text{UE}} \times N_{\text{UE}}}$ represent the analog beamforming matrices, while $\mathbf{F}_{u,k}^{\text{BB}} \in \mathbb{C}^{N_{\text{BS}} \times N_s}$ and $\mathbf{W}_{u,k}^{\text{BB}} \in \mathbb{C}^{N_{\text{UE}} \times N_s}$ denote the baseband beamforming matrices. Without loss of generality, we study the two extreme cases of the generalized HAD beamforming architectures, which are the FC and the PC architectures, which are illustrated in Fig. 3.1. In the FC architecture, each RF chain is connected to all antennas using a networks of PSs. Therefore, $\mathbf{F}^{\text{RF}} =$

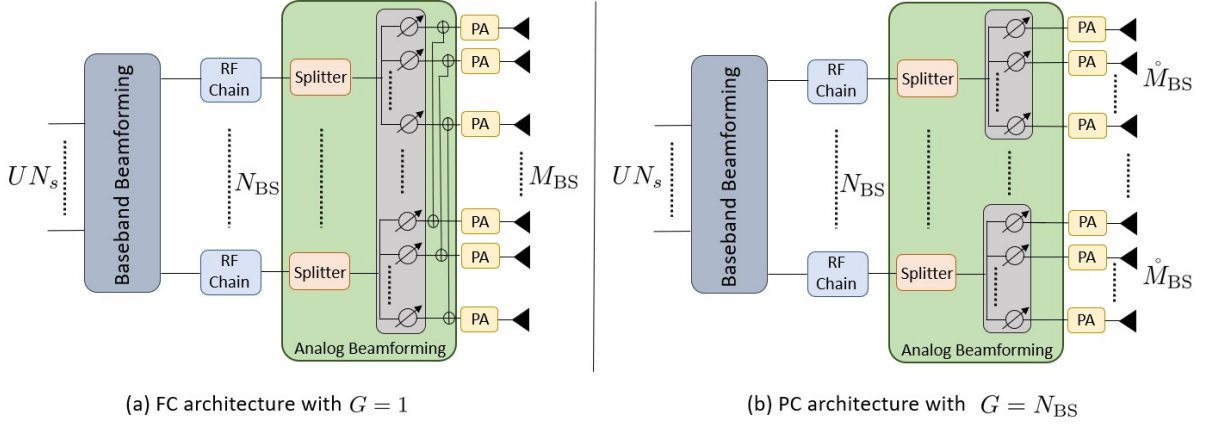


Figure 3.1: The FC and the PC HAD architectures at the BS.

$[\mathbf{f}_1^{\text{RF}}, \dots, \mathbf{f}_{N_{BS}}^{\text{RF}}], \mathbf{f}_a^{\text{RF}} = [\mathbf{F}^{\text{RF}}]_{[:,a]} \in \mathbb{C}^{M_{BS}}, |[\mathbf{f}_a^{\text{RF}}]_{[b]}| = \frac{1}{\sqrt{M_{BS}}}, \mathbf{W}_u^{\text{RF}} = [\mathbf{w}_{u,1}^{\text{RF}}, \dots, \mathbf{w}_{u,N_{UE}}^{\text{RF}}], \mathbf{w}_{u,a}^{\text{RF}} = [\mathbf{W}_u^{\text{RF}}]_{[:,a]} \in \mathbb{C}^{M_{UE}}$, and $|[\mathbf{w}_{u,a}^{\text{RF}}]_{[b]}| = \frac{1}{\sqrt{M_{UE}}}$. In the PC architecture, each RF chain is connected to a subset of the antennas using a network of PSs. Therefore, $\mathbf{F}^{\text{RF}} = \text{blkdiag}\{\mathbf{f}_1^{\text{RF}}, \dots, \mathbf{f}_{N_{BS}}^{\text{RF}}\}, \mathbf{f}_a^{\text{RF}} \in \mathbb{C}^{\dot{M}_{BS}}, |[\mathbf{f}_a^{\text{RF}}]_{[b]}| = \frac{1}{\sqrt{\dot{M}_{BS}}}$ with $\dot{M}_{BS} = \frac{M_{BS}}{N_{BS}}, \mathbf{W}_u^{\text{RF}} = \text{blkdiag}\{\mathbf{w}_{u,1}^{\text{RF}}, \dots, \mathbf{w}_{u,N_{UE}}^{\text{RF}}\}, \mathbf{w}_{u,a}^{\text{RF}} \in \mathbb{C}^{\dot{M}_{UE}}$, and $|[\mathbf{w}_{u,a}^{\text{RF}}]_{[b]}| = \frac{1}{\sqrt{\dot{M}_{UE}}}$ with $\dot{M}_{UE} = \frac{M_{UE}}{N_{UE}}$. The achievable SE at the u th UE on the k th sub-carrier can be expressed as

$$\text{SE}_{u,k} = \log_2 |\mathbf{I}_{N_s} + \Phi_{u,k}^{-1} \mathbf{W}_{u,k}^{\text{H}} \mathbf{H}_{u,k} \mathbf{F}_{u,k} \mathbf{F}_{u,k}^{\text{H}} \mathbf{H}_{u,k}^{\text{H}} \mathbf{W}_{u,k}|, \quad (3.2)$$

where $\Phi_{u,k} = \mathbf{W}_{u,k}^{\text{H}} \left(\mathbf{H}_{u,k} \left(\sum_{u'=1, u' \neq u}^U \mathbf{F}_{u',k} \mathbf{F}_{u',k}^{\text{H}} \right) \mathbf{H}_{u',k}^{\text{H}} + \sigma_n^2 \mathbf{I}_{M_{UE}} \right) \mathbf{W}_{u,k}$ is the covariance matrix of the multi-user interference plus noise. Let $\mathbf{F}_k = [\mathbf{F}_{1,k}, \dots, \mathbf{F}_{U,k}] \in \mathbb{C}^{M_{BS} \times U M_{UE}}$. Then, we enforce $\|\mathbf{F}_k\|_{\text{F}}^2 = P_k$, where $P_k \stackrel{\text{def}}{=} \frac{P}{K}$ and P is the total transmit power.

Assuming that perfect and instantaneous CSI is available at both the BS and every u th UE on the k th sub-carrier, the objective is to design the HAD beamforming matrices $\mathbf{F}_{u,k}$ and $\mathbf{W}_{u,k}$ to maximize the total SE of the system, i.e.,

$$\begin{aligned} \max_{\{\mathbf{F}_{u,k}, \mathbf{W}_{u,k}\}} \text{SE} &= \sum_{u=1}^U \sum_{k=1}^K \text{SE}_{u,k} \\ \text{s.t. } \|\mathbf{F}_k\|_{\text{F}}^2 &= P_k, \forall k, \mathbf{F}^{\text{RF}} \in \mathcal{F}, \mathbf{W}_u^{\text{RF}} \in \mathcal{W}, \forall u, \end{aligned} \quad (3.3)$$

where \mathcal{F} and \mathcal{W} represent the sets of analog beamforming matrices satisfying the constraints that are associated with each architecture as specified above.

3.3. Beamforming Design

The joint design of the analog and baseband beamforming matrices on the one hand and the non-convex constraints in $\mathbf{W}_u^{\text{RF}} \in \mathcal{W}, \forall u$, and $\mathbf{F}^{\text{RF}} \in \mathcal{F}$, on the other hand, make (3.3) a non-convex optimization problem. Therefore, a sub-optimal strategy of decoupling the beamforming design has been widely

used in the literature, where the analog and baseband beamforming matrices are designed separately. We show how to obtain the baseband beamforming matrices based on an equivalent channel matrix formed by fixing the analog beamforming matrices together with the channel matrix in Subsection 3.3.1. In Subsection 3.3.2, we address the design of the analog beamforming subproblems to obtain \mathbf{F}^{RF} and \mathbf{W}_u^{RF} . The steps of designing the beamforming matrices are as follows

- Analog decoding matrix \mathbf{W}_u^{RF} (as described in Subsection 3.3.2)
- Analog precoding matrix \mathbf{F}^{RF} (as described in Subsection 3.3.2)
- Baseband precoding matrix $\mathbf{F}_{u,k}^{\text{BB}}$ using ZF (as described in Subsection 3.3.1)
- Baseband decoding matrix $\mathbf{W}_{u,k}^{\text{BB}}$ using MMSE (as described in Subsection 3.3.1)

Note that the analog beamforming matrices \mathbf{F}^{RF} and \mathbf{W}_u^{RF} are identical for all sub-carriers while the baseband beamforming matrices $\mathbf{F}_{u,k}^{\text{BB}}$ and $\mathbf{W}_{u,k}^{\text{BB}}$ have to be optimized for different sub-carriers k in broadband systems.

3.3.1. Baseband Beamforming Design

For given and fixed analog beamforming matrices \mathbf{F}^{RF} and $\mathbf{W}_u^{\text{RF}}, \forall u$, the BS designs its transmit baseband beamforming matrices using the conventional linear ZF scheme to eliminate the multi-user interference. We assume that UE u uses a digital noise whitening filter after applying its receive analog beamforming \mathbf{W}_u^{RF} , i.e.,

$$\Upsilon_u = \left((\mathbf{W}_u^{\text{RF}})^H \mathbf{W}_u^{\text{RF}} \right)^{-\frac{1}{2}} \in \mathbb{C}^{N_{\text{UE}} \times N_{\text{UE}}}, \quad (3.4)$$

Let $\mathring{\mathbf{H}}_{u,k} \stackrel{\text{def}}{=} \Upsilon_u^H (\mathbf{W}_u^{\text{RF}})^H \mathbf{H}_{u,k} \in \mathbb{C}^{N_{\text{UE}} \times M_{\text{BS}}}$, $\mathring{\mathbf{H}}_k \stackrel{\text{def}}{=} [\mathring{\mathbf{H}}_{1,k}^T, \dots, \mathring{\mathbf{H}}_{U,k}^T]^T \in \mathbb{C}^{U N_{\text{UE}} \times M_{\text{BS}}}$, and $\bar{\mathbf{H}}_k \stackrel{\text{def}}{=} \mathring{\mathbf{H}}_k \mathbf{F}^{\text{RF}} \in \mathbb{C}^{U N_{\text{UE}} \times N_{\text{BS}}}$ for the k th sub-carrier. Let us define $\bar{\mathbf{F}}_k^{\text{BB}}$ at the k th sub-carrier as

$$\bar{\mathbf{F}}_k^{\text{BB}} \stackrel{\text{def}}{=} [\bar{\mathbf{F}}_{1,k}^{\text{BB}}, \dots, \bar{\mathbf{F}}_{U,k}^{\text{BB}}] \in \mathbb{C}^{N_{\text{BS}} \times U N_s}, \quad (3.5)$$

where $\bar{\mathbf{F}}_{u,k}^{\text{BB}} \in \mathbb{C}^{N_{\text{BS}} \times N_s}$ is the u th sub-matrix of $\bar{\mathbf{F}}_k^{\text{BB}}$ related to the u th UE. The un-normalized ZF precoding matrix for the k th sub-carrier can be obtained as

$$\bar{\mathbf{F}}_k^{\text{BB}} = \bar{\mathbf{H}}_k^H (\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^H)^{-1}, \quad (3.6)$$

which is normalized as given below

$$\mathbf{F}_k^{\text{BB}} = \sqrt{P_k} (\bar{\mathbf{F}}_k^{\text{BB}} / \|\mathbf{F}^{\text{RF}} \bar{\mathbf{F}}_k^{\text{BB}}\|_{\text{F}}) \in \mathbb{C}^{N_{\text{BS}} \times U N_s}. \quad (3.7)$$

where $\mathbf{F}_k^{\text{BB}} \stackrel{\text{def}}{=} [\mathbf{F}_{1,k}^{\text{BB}}, \dots, \mathbf{F}_{U,k}^{\text{BB}}]$. On the other hand, the baseband decoding matrix of the u th UE on the k th sub-carrier, $\mathbf{W}_{u,k}^{\text{BB}} \in \mathbb{C}^{N_{\text{UE}} \times N_s}, \forall k, \forall u$, is computed such that the MSE of the received signal is minimized. In other words, it is calculated using the MMSE scheme as

$$\mathbf{W}_{u,k}^{\text{BB}} = \min_{\mathbf{W}_{u,k}^{\text{BB}}, \forall u, \forall k} \mathbb{E}\{\|\hat{\mathbf{s}}_{u,k} - \mathbf{s}_{u,k}\|^2\}, \quad (3.8)$$

where its solution is given by [ASC17]

$$\mathbf{W}_{u,k}^{\text{BB}} = \boldsymbol{\Omega}_{u,k}^{-1} (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{H}_{u,k} \mathbf{F}_{u,k}, \quad (3.9)$$

where

$$\boldsymbol{\Omega}_{u,k} = \mathbf{S}_{u,u,k} \mathbf{S}_{u,u,k}^{\text{H}} + \sum_{u'=1, u' \neq u}^U \mathbf{S}_{u,u',k} \mathbf{S}_{u,u',k}^{\text{H}} + \sigma_n^2 \boldsymbol{\Upsilon}_u^{\text{H}} (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{W}_u^{\text{RF}} \boldsymbol{\Upsilon}_u \in \mathbb{C}^{N_{\text{UE}} \times N_{\text{UE}}}, \quad (3.10)$$

where $\mathbf{S}_{u,x,k} = \boldsymbol{\Upsilon}_u^{\text{H}} (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{H}_{u,k} \mathbf{F}_{x,k}$ and $\mathbf{F}_{x,k}^{\text{BB}}, x \in \{u, u'\}$ are calculated as in (3.7). Note that the first term in (3.10) represents the desired signal covariance matrix, the second term denotes the inter-user interference covariance matrix, and the third term is the whitened decoded noise covariance matrix. Due to the ZF at the BS and applying the noise whitening filter $\boldsymbol{\Upsilon}_u$, i.e., $\sigma_n^2 \boldsymbol{\Upsilon}_u^{\text{H}} (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{W}_u^{\text{RF}} \boldsymbol{\Upsilon}_u = \sigma_n^2 \mathbf{I}_{N_{\text{UE}}}$, we can simplify (3.10) as

$$\boldsymbol{\Omega}_{u,k} = \mathbf{S}_{u,u,k} \mathbf{S}_{u,u,k}^{\text{H}} + \sigma_n^2 \mathbf{I}_{N_{\text{UE}}}. \quad (3.11)$$

3.3.2. Analog Beamforming Design

Since the baseband beamforming matrices are designed using the ZF method at the BS, the multi-user interference can be neglected during the design of the analog beamforming matrices, where the problem can be formulated such that the capacity of each effective channel that takes into account the analog beamforming matrices, i.e.,

$$\mathbf{H}_{u,k}^{\text{eff}} \stackrel{\text{def}}{=} (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{H}_{u,k} \mathbf{F}^{\text{RF}}, \quad (3.12)$$

is maximized. In the following, we decouple the analog beamforming matrices design between \mathbf{W}_u^{RF} and \mathbf{F}^{RF} by treating them independently. First, we design the analog decoding matrix \mathbf{W}_u^{RF} for the u th UE using the following approach

$$\max_{\mathbf{W}_u^{\text{RF}} \in \mathcal{W}} \eta_u = \sum_{k=1}^K \log_2 |\mathbf{I}_{N_{\text{UE}}} + (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{H}_{u,k} \mathbf{H}_{u,k}^{\text{H}} \mathbf{W}_u^{\text{RF}}|, \quad (3.13)$$

where we use the approximation $\mathbf{F}^{\text{RF}} (\mathbf{F}^{\text{RF}})^{\text{H}} \approx \mathbf{I}_{M_{\text{BS}}}$, which is valid for large antenna arrays [SY16]. To obtain a solution of problem (3.13), we first rewrite its objective function in terms of its upper-bound as

$$\eta_u < \log_2 |\mathbf{I}_{N_{\text{UE}}} + (\mathbf{W}_u^{\text{RF}})^{\text{H}} \sum_{k=1}^K \mathbf{H}_{u,k} \mathbf{H}_{u,k}^{\text{H}} \mathbf{W}_u^{\text{RF}}|, \quad (3.14)$$

which follows from Jensen's inequality as described in Appendix A.2 and $\mathbf{D}_u = \sum_{k=1}^K \mathbf{H}_{u,k} \mathbf{H}_{u,k}^{\text{H}} \in \mathbb{C}^{M_{\text{UE}} \times M_{\text{UE}}}$. Therefore, the upper-bound of (3.13) reduces to

$$\max_{\mathbf{W}_u^{\text{RF}} \in \mathcal{W}} \log_2 |\mathbf{I}_{N_{\text{UE}}} + (\mathbf{W}_u^{\text{RF}})^{\text{H}} \mathbf{D}_u \mathbf{W}_u^{\text{RF}}|. \quad (3.15)$$

Problem (3.15) has been addressed in [Gao+16] considering a PC analog beamforming architecture. Meanwhile, the authors in [Ard+18] proposed a unifying design approach for the flat-fading case that is applicable for both analog beamforming architectures, i.e., the FC and the PC ones. Here, we assume

3. Hybrid Beamforming Design

that the analog decoding matrices $\mathbf{W}_u^{\text{RF}}, \forall u$, are updated using the proposed method in [Ard+18] as described in Appendix A.3.

Assuming that $\mathbf{F}_{u,k}^{\text{BB}}$ and $\mathbf{W}_{u,k}^{\text{BB}}, \forall u, k$, are computed by (3.7) and (3.9), respectively. Then, due to the noise whitening filter and the ZF filter at the BS, $\Phi_{u,k}^{-1} = \frac{1}{\sigma_n^2} \mathbf{I}_{N_s}$, which simplifies (3.2) to [Chr+08]

$$\text{SE}_{u,k} = \log_2 |\mathbf{I}_{N_s} + \frac{1}{\sigma_n^2} (\mathbf{F}_{u,k}^{\text{BB}})^{\text{H}} (\mathbf{F}^{\text{RF}})^{\text{H}} \mathring{\mathbf{H}}_{u,k}^{\text{H}} \mathring{\mathbf{H}}_{u,k} \mathbf{F}^{\text{RF}} \mathbf{F}_{u,k}^{\text{BB}}|. \quad (3.16)$$

Therefore, the objective function of (3.3) simplifies to

$$\text{SE} = \sum_{u=1}^U \sum_{k=1}^K \text{SE}_{u,k} = \sum_{k=1}^K \log_2 |\mathbf{I}_{UN_s} + \frac{1}{\sigma_n^2} (\mathbf{F}_k^{\text{BB}})^{\text{H}} \bar{\mathbf{H}}_k^{\text{H}} \bar{\mathbf{H}}_k \mathbf{F}_k^{\text{BB}}|, \quad (3.17)$$

which can be upper-bounded as

$$\text{SE} \leq UN_s \log_2 \left(1 + \frac{P_k}{\sigma_n^2 (UN_{\text{UE}})^2} \text{trace}[(\mathbf{F}^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{F}^{\text{RF}}] \right), \quad (3.18)$$

where

$$\mathbf{D} = \sum_{k=1}^K \mathring{\mathbf{H}}_k^{\text{H}} \mathring{\mathbf{H}}_k \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{BS}}}. \quad (3.19)$$

The derivation of the upper-bound in (3.18) is provided in Appendix A.4. Therefore, \mathbf{F}^{RF} can be designed so that (3.18) is maximized, i.e.,

$$\max_{\mathbf{F}^{\text{RF}} \in \mathbb{C}^{M_{\text{BS}} \times N_{\text{BS}}}} \text{trace}[(\mathbf{F}^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{F}^{\text{RF}}] \quad \text{s.t. } \mathbf{F}^{\text{RF}} \in \mathcal{F}, \text{rank}(\mathbf{F}^{\text{RF}}) = N_{\text{BS}}. \quad (3.20)$$

3.3.2.1 Baseline Method

For FC architectures, a direct solution of (3.20) can be calculated by applying the eigenvalue decomposition technique on the equivalent channel covariance matrix \mathbf{D} given in (3.19), as

$$\mathbf{D} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{\text{H}}, \quad (3.21)$$

where $\mathbf{Q} \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{BS}}}$ contains the eigenvectors of \mathbf{D} in its columns, while $\mathbf{\Lambda} = \text{diag}\{\boldsymbol{\lambda}\} \in \mathbb{R}^{M_{\text{BS}} \times M_{\text{BS}}}$ holds the associated eigenvalues on its diagonal elements with $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_{M_{\text{UE}}}]^{\text{T}}$ and $\lambda_1 \geq \dots \geq \lambda_{M_{\text{BS}}}$. We take the phase-angles of the eigenvectors corresponding to the N_{BS} largest eigenvalues as

$$\mathbf{F}^{\text{RF}} = \frac{1}{\sqrt{M_{\text{UE}}}} e^{j\angle \bar{\mathbf{Q}}} \in \mathbb{C}^{M_{\text{BS}} \times N_{\text{BS}}}, \quad (3.22)$$

where $\bar{\mathbf{Q}} = [\mathbf{Q}]_{[:,1:N_{\text{BS}}]} \in \mathbb{C}^{M_{\text{BS}} \times N_{\text{BS}}}$ and $\angle(\cdot)$ is a function that extracts the phase angles.

For the PC architecture, the authors in [LW20] divide the analog beamforming design into N_{BS} sub-problems. Let $\tilde{\mathbf{D}}_c \in \mathbb{C}^{\tilde{M}_{\text{BS}} \times \tilde{M}_{\text{BS}}}$ be the channel covariance matrix of all antennas connected to the c th RF chain with $\tilde{M}_{\text{BS}} = \frac{M_{\text{BS}}}{N_{\text{BS}}}$ and $c \in \{1, \dots, N_{\text{BS}}\}$, i.e., $\mathbf{D} = \text{blkdiag}\{\tilde{\mathbf{D}}_1, \dots, \tilde{\mathbf{D}}_{N_{\text{BS}}}\}$. The eigenvalue decomposition of $\tilde{\mathbf{D}}_c$ is given as

$$\tilde{\mathbf{D}}_c = \mathbf{Q}_c \mathbf{\Lambda}_c \mathbf{Q}_c^{\text{H}}, \quad (3.23)$$

where $\mathbf{Q}_c \in \mathbb{C}^{\tilde{M}_{\text{BS}} \times \tilde{M}_{\text{BS}}}$ contains the eigenvectors of the covariance matrix \mathbf{D}_c in its columns, while

$\Lambda_c = \text{diag}\{\lambda_c\} \in \mathbb{R}^{\dot{M}_{\text{BS}} \times \dot{M}_{\text{BS}}}$ holds the associated eigenvalues on its diagonal elements with $\lambda_c = [\lambda_{1,c}, \dots, \lambda_{\dot{M}_{\text{BS}},c}]^T$ and $\lambda_{1,c} \geq \dots \geq \lambda_{\dot{M}_{\text{BS}},c}$. Then, \mathbf{f}_c^{RF} can be designed by taking the phase-angles of the eigenvector corresponding to the dominant eigenvalue of $\tilde{\mathbf{D}}_c$ as

$$\mathbf{f}_c^{\text{RF}} = \frac{1}{\sqrt{\dot{M}_{\text{UE}}}} e^{j\angle \bar{\mathbf{q}}_c} \in \mathbb{C}^{\dot{M}_{\text{BS}}}, \quad (3.24)$$

where $\bar{\mathbf{q}}_c = [\mathbf{Q}_c]_{[:,1]} \in \mathbb{C}^{\dot{M}_{\text{BS}}}$ denotes the eigenvector associated with the largest eigenvalue, i.e., $\lambda_{1,c}$.

Although the method proposed in [LW20] is very simple to implement for the PC architecture, it has a poor performance because it ignores the interference between the columns of \mathbf{F}^{RF} .

3.3.2.2 Proposed Analog Beamforming Design

Here, we want to present our proposed solution, which is not only suitable for the FC architecture but also for the PC architecture. To this end, by taking a closer look at the structure of $(\mathbf{F}^{\text{RF}})^H \mathbf{D} \mathbf{F}^{\text{RF}}$, which can be written as

$$(\mathbf{F}^{\text{RF}})^H \mathbf{D} \mathbf{F}^{\text{RF}} = \begin{bmatrix} (\mathbf{f}_1^{\text{RF}})^H \mathbf{D} \mathbf{f}_1^{\text{RF}} & (\mathbf{f}_1^{\text{RF}})^H \mathbf{D} \mathbf{f}_2^{\text{RF}} & \dots & (\mathbf{f}_1^{\text{RF}})^H \mathbf{D} \mathbf{f}_{N_{\text{BS}}}^{\text{RF}} \\ (\mathbf{f}_2^{\text{RF}})^H \mathbf{D} \mathbf{f}_1^{\text{RF}} & (\mathbf{f}_2^{\text{RF}})^H \mathbf{D} \mathbf{f}_2^{\text{RF}} & \dots & (\mathbf{f}_2^{\text{RF}})^H \mathbf{D} \mathbf{f}_{N_{\text{BS}}}^{\text{RF}} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{f}_{N_{\text{BS}}}^{\text{RF}})^H \mathbf{D} \mathbf{f}_1^{\text{RF}} & (\mathbf{f}_{N_{\text{BS}}}^{\text{RF}})^H \mathbf{D} \mathbf{f}_2^{\text{RF}} & \dots & (\mathbf{f}_{N_{\text{BS}}}^{\text{RF}})^H \mathbf{D} \mathbf{f}_{N_{\text{BS}}}^{\text{RF}} \end{bmatrix} \in \mathbb{C}^{N_{\text{BS}} \times N_{\text{BS}}}, \quad (3.25)$$

where $\mathbf{f}_c^{\text{RF}} \in \mathbb{C}^{\dot{M}_{\text{BS}}}$ is the c th column of $\mathbf{F}^{\text{RF}} \in \mathbb{C}^{\dot{M}_{\text{BS}} \times N_{\text{BS}}}$ with $c \in \{1, \dots, N_{\text{BS}}\}$ and \mathbf{D} is given in (3.19). From (3.25), the objective function of (3.20), i.e., $\text{trace}[(\mathbf{F}^{\text{RF}})^H \mathbf{D} \mathbf{F}^{\text{RF}}]$, can be expressed as

$$\text{trace}[(\mathbf{F}^{\text{RF}})^H \mathbf{D} \mathbf{F}^{\text{RF}}] = (\mathbf{f}_1^{\text{RF}})^H \mathbf{D} \mathbf{f}_1^{\text{RF}} + \dots + (\mathbf{f}_{N_{\text{BS}}}^{\text{RF}})^H \mathbf{D} \mathbf{f}_{N_{\text{BS}}}^{\text{RF}} = \sum_{c=1}^{N_{\text{BS}}} (\mathbf{f}_c^{\text{RF}})^H \mathbf{D} \mathbf{f}_c^{\text{RF}}. \quad (3.26)$$

Accordingly, (3.20) can be equivalently written as

$$\max_{\mathbf{f}_c^{\text{RF}}} \sum_{c=1}^{N_{\text{BS}}} (\mathbf{f}_c^{\text{RF}})^H \mathbf{D} \mathbf{f}_c^{\text{RF}} \quad \text{s.t.} \quad \mathbf{f}_c^{\text{RF}} \in \mathcal{F}, \forall c, \text{rank}(\mathbf{F}^{\text{RF}}) = N_{\text{BS}}. \quad (3.27)$$

Note that (3.27) is non-convex optimization problem. To obtain a solution, we propose to approach (3.27) heuristically by optimizing

$$\max_{\mathbf{f}_c^{\text{RF}}} \sum_{c=1}^{N_{\text{BS}}} \left((\mathbf{f}_c^{\text{RF}})^H \mathbf{D} \mathbf{f}_c^{\text{RF}} - \Upsilon_c \right) \quad \text{s.t.} \quad \mathbf{f}_c^{\text{RF}} \in \mathcal{F}, \forall c, \quad (3.28)$$

where $\Upsilon_c = \sum_{b=1, b \neq c}^{N_{\text{BS}}} |(\mathbf{f}_b^{\text{RF}})^H \mathbf{D} \mathbf{f}_c^{\text{RF}}|^2$, which represents the total *interference-leakage* caused by \mathbf{f}_c^{RF} . We conjecture that the subtraction of Υ_c may ensure that the resulting matrix \mathbf{F}^{RF} has linearly independent columns. This might be explained by noting that the best solution of (3.28) is achieved when Υ_c is minimized, while the first term, i.e., $(\mathbf{f}_c^{\text{RF}})^H \mathbf{D} \mathbf{f}_c^{\text{RF}}$, is maximized. Therefore, the best solution can only be achieved when the columns of \mathbf{F}^{RF} are linearly independent. Note that the quadratic form is used for Υ_c so that the objective function in (3.28) falls into a classical format as in [Par+13]. In the following,

3. Hybrid Beamforming Design

Algorithm 3.3 Proposed Analog Beamforming Design

Input: $\mathbf{D} = \sum_{k=1}^K \hat{\mathbf{H}}_k^H \hat{\mathbf{H}}_k \in \mathbb{C}^{M_{\text{BS}} \times M_{\text{BS}}}$ and $\mathbf{F}^{\text{RF}}[0] \in \mathcal{F}$ (random matrix)

- 1: Compute $\mathbf{d}_c[0] \stackrel{\text{def}}{=} \mathbf{D} \mathbf{f}_c^{\text{RF}}[0] \in \mathbb{C}^{M_{\text{BS}}}, \forall c$, and set iteration index $p = 1$
- 2: **while** not converged **do**
- 3: **for** $c = 1$ to N_{BS} **do**
- 4: Compute $\mathbf{f}_{c,\text{opt}}^{\text{RF}}$ using (3.31)
- 5: Set $\mathbf{f}_c^{\text{RF}}[p] = 1/\sqrt{X_{\text{BS}}}\mathbf{e}^{j\angle(\mathbf{f}_{c,\text{opt}} \odot \phi_c)}$
- 6: Set $\mathbf{d}_c[p] \stackrel{\text{def}}{=} \mathbf{D} \mathbf{f}_c^{\text{RF}}[p] \in \mathbb{C}^{M_{\text{BS}}}$
- 7: **end for**
- 8: $p = p + 1$
- 9: **end while**

Output: $\mathbf{F}^{\text{RF}}[p]$

we propose a solution to (3.28) in an alternating fashion. Let $\mathbf{d}_c \stackrel{\text{def}}{=} \mathbf{D} \mathbf{f}_c^{\text{RF}} \in \mathbb{C}^{M_{\text{BS}}}, c \in \{1, \dots, N_{\text{BS}}\}$. Then, (3.28) can be rewritten as

$$\max_{\mathbf{f}_c^{\text{RF}}} \sum_{c=1}^{N_{\text{BS}}} (\mathbf{d}_c^H \mathbf{f}_c^{\text{RF}} - \sum_{b=1, b \neq c}^{N_{\text{BS}}} |\mathbf{d}_b^H \mathbf{f}_c^{\text{RF}}|^2) \quad \text{s.t. } \mathbf{f}_c^{\text{RF}} \in \mathcal{F}, \forall c. \quad (3.29)$$

Problem (3.29) is non-convex, due to the constant modulus constrains. Note that when $\mathbf{f}_c^{\text{RF}} \in \mathcal{F}$, then $\|\mathbf{f}_c^{\text{RF}}\|^2 = 1$. Therefore, we propose to relax (3.29) and write it as

$$\max_{\mathbf{f}_c^{\text{RF}}} \sum_{c=1}^{N_{\text{BS}}} (\mathbf{d}_c^H \mathbf{f}_c^{\text{RF}} - \sum_{b=1, b \neq c}^{N_{\text{BS}}} |\mathbf{d}_b^H \mathbf{f}_c^{\text{RF}}|^2) \quad \text{s.t. } \|\mathbf{f}_c^{\text{RF}}\|^2 = 1, \forall c, \quad (3.30)$$

which is a convex optimization problem. The optimal solution to \mathbf{f}_c^{RF} satisfying the Karush-Kuhn-Tucker (KKT) conditions is given by *Proposition 3.3.1*.

Proposition 3.3.1. *The optimal solution to (3.30) with respect to \mathbf{f}_c^{RF} is given as*

$$\mathbf{f}_{c,\text{opt}}^{\text{RF}} = \mu_c \left(\sum_{b=1, b \neq c}^{N_{\text{BS}}} \mathbf{d}_b \mathbf{d}_b^H + \mathbf{I}_{M_{\text{BS}}} \right)^{-1} \mathbf{d}_c, \quad (3.31)$$

where $\mu_c = 1/\left\| \left(\sum_{b=1, b \neq c}^{N_{\text{BS}}} \mathbf{d}_b \mathbf{d}_b^H + \mathbf{I}_{M_{\text{BS}}}^{-1} \right) \mathbf{d}_c \right\|$.

Proof: Please refer to Appendix A.5.

A summary of the proposed analog beamforming approach for updating \mathbf{F}^{RF} can be found in Algorithm 3.3, where the entries of $\mathbf{F}^{\text{RF}}[0]$ have random phases and amplitude one and in step 5, we enforce the constant modulus constraints of $\mathbf{f}_{c,\text{opt}} \in \mathcal{F}$. Here, $\phi_c \in \mathbb{B}^{M_{\text{BS}}}$ is a binary vector. For a FC architecture, $X_{\text{BS}} \stackrel{\text{def}}{=} M_{\text{BS}}$ and $\phi_c \stackrel{\text{def}}{=} \mathbf{1}_{M_{\text{BS}}}, \forall c$, i.e., an all ones vector. For a PC architecture, $X_{\text{BS}} \stackrel{\text{def}}{=} M_{\text{BS}}/N_{\text{BS}}$

and ϕ_c contains ones at indices $[(c-1)\overset{\circ}{M}_{\text{BS}} + 1, \dots, c\overset{\circ}{M}_{\text{BS}}]$ and zeros elsewhere as shown below.

$$\phi_c = \left[\underbrace{\mathbf{0}^T \dots \mathbf{0}^T}_{(c-1)\overset{\circ}{M}_{\text{BS}}} \mathbf{1}_{\overset{\circ}{M}_{\text{BS}}}^T \underbrace{\mathbf{0}^T \dots \mathbf{0}^T}_{(1-c)\overset{\circ}{M}_{\text{BS}}} \right]^T. \quad (3.32)$$

Note that Algorithm 3.3 updates \mathbf{F}^{RF} until reaching a convergence threshold by checking the stop criterion given as

$$|c(p) - c(p-1)| < \epsilon, \quad (3.33)$$

where ϵ is the convergence threshold and $c(p)$ is defined as

$$c(p) = \log_2 |(\mathbf{F}^{\text{RF}}[p])^H \mathbf{D} \mathbf{F}^{\text{RF}}[p]|. \quad (3.34)$$

3.4. Simulation Results

Here, we evaluate the performance of the proposed algorithm by means of numerical results. We generate the MIMO OFDM channel matrices according to the 3GPP CDL channel model given in Subsection 2.2.2. We define the SNR as $\text{SNR} = 10 \log_{10} \frac{P}{\sigma_n^2}$ dB. Remember that the FC architecture has $G = 1$ group, while the PC architecture at the BS (UE) has $G = N_{\text{BS}}$ ($G = N_{\text{UE}}$) groups. For comparison, we show simulation results of the proposed HAD methods in [Gao+16]¹ and [Zha+19]², and the fully digital ZF-based approach³.

Convergence Behavior of Proposed Analog Beamforming Design Method in Algorithm 3.3

As explained above, Algorithm 3.3 updates the analog beamforming matrix \mathbf{F}^{RF} iteratively until reaching a convergence threshold by checking the stop criterion given in (3.33). In Fig. 3.2, we show some simulation results evaluating the convergence behavior of Algorithm 3.3. We assume that $M_{\text{BS}} = 64$, $M_{\text{UE}} = 8$, $U = 4$, $N_s = 2$, $N_{\text{UE}} = N_s$, $N_{\text{BS}} = UN_s$, and $K = 16$. From Fig. 3.2, we notice that Algorithm 3.3 always converges to a point where the cost function stops increasing or decreasing, although not monotonically. Moreover, we notice that Algorithm 3.3 has a fast convergence rate, requiring approximately 5 to 10 iterations to reach an acceptable convergence threshold of $\epsilon = 10^{-5}$.

Sum Rate Performance

Here, we want to evaluate the performance of the proposed hybrid beamforming design in terms of the SE. We assume that $M_{\text{BS}} = 64$, $M_{\text{UE}} = 8$, $U = 4$, $N_s = 2$, $N_{\text{UE}} = N_s$, $N_{\text{BS}} = UN_s$, and $K = 16$. From Fig. 3.3, we can see that all the HAD algorithms using the FC architecture achieve a comparable performance to each other with some performance gap compared to the fully digital architecture.

¹In [Gao+16], the authors consider the PC architecture. Its extension to a FC architecture coincides with the proposed method in [LW20].

²In [Zha+19], the authors proposed a hybrid beamforming design for a FC architecture. Analogous to [LW20], the proposed solution can be simply extended to the PC architecture by updating the c th block of the analog beamforming matrix considering the unfolding of the channel tensor of all antennas connected to the c th RF chain.

³To have a fair comparison, we assume that the fully digital ZF transmit precoding matrix is calculated using (3.7), while assuming that $\mathbf{F}^{\text{RF}} = \mathbf{I}_{M_{\text{BS}}}$ and $\mathbf{W}_u^{\text{RF}} = \mathbf{U}_u \in \mathbb{C}^{M_{\text{UE}} \times N_{\text{UE}}}$, where \mathbf{U}_u consists of the N_{UE} dominant eigenvectors of the channel covariance matrix \mathbf{D}_u given by (3.14).

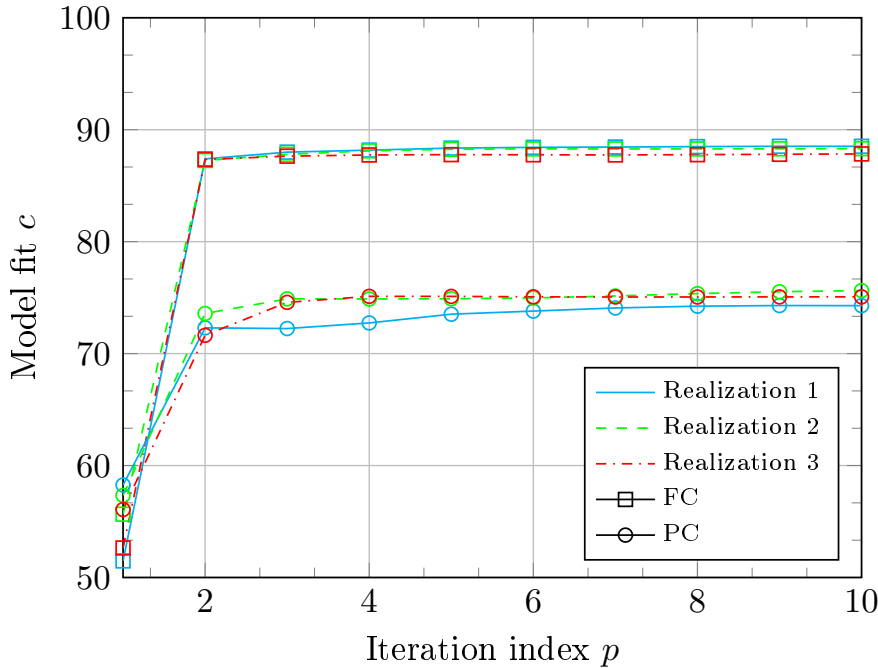


Figure 3.2: Convergence of Algorithm 3.3.

Clearly, the PC architecture has a large performance loss compared to the FC and the fully digital counterparts due to the smaller number of antennas connected to each RF chain. Nonetheless, we can see that our proposed PC HAD approach outperforms the other PC HAD methods in [Gao+16; Zha+19]. The proposed PC HAD beamforming method in [Zha+19] simplifies the analog beamforming design by completely decoupling the optimization between the analog beamforming blocks. This results in a significant performance degradation of the system. Differently, the proposed PC HAD beamforming method in [Gao+16] partially decouples the optimization between the analog beamforming blocks by sequentially updating them, where in each step, the previously designed blocks are taken into account when designing the current block. On the other hand, the proposed method in this work updates the analog beamforming matrix sequentially and iteratively with the aim of finding a good balance between maximizing the received signal in each block and minimizing the interference leakage into the other blocks.

Effect of Number of Transmit Antennas on SE

Next, we investigate the effect of the number of transmit antennas on the SE for a fixed SNR, i.e., $\text{SNR} = 15$ dB. We assume that $M_{\text{BS}} = [16, 32, 64, 128, 256]$ and the rest of the simulation parameters remain the same as above. As expected, increasing the number of antennas at the BS improves the SE of the system. From Fig. 3.4, we can see that the FC HAD architectures achieve a comparable performance to each other with some performance gap compared to the fully digital architecture. Moreover, we can see that our proposed PC HAD beamforming architecture outperforms the other PC HAD beamforming algorithms in [Zha+19] and [Gao+16].

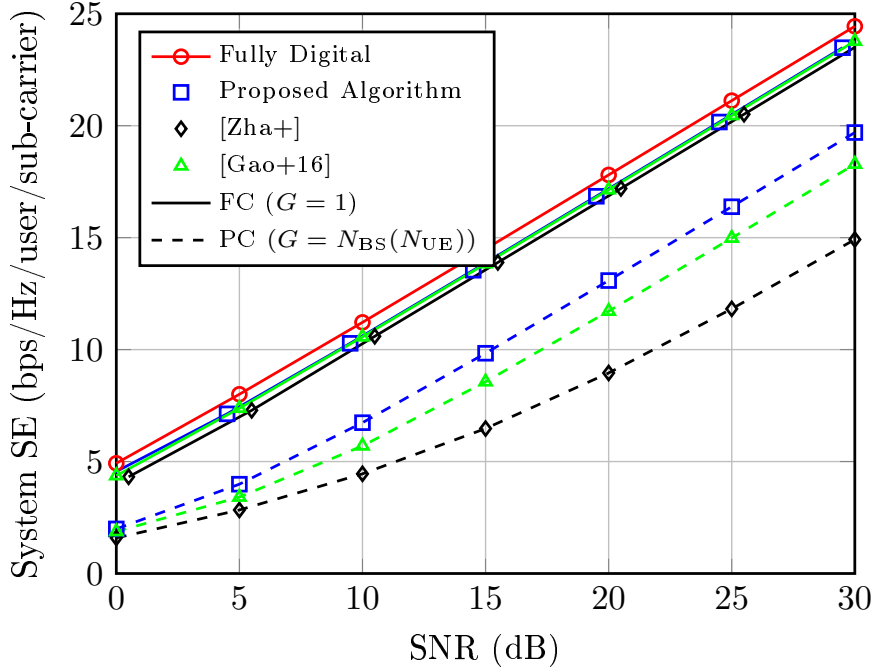


Figure 3.3: System SE versus SNR for $M_{\text{BS}} = 64$, $M_{\text{UE}} = 8$, $U = 4$, $N_{\text{UE}} = 2$, $N_{\text{BS}} = 8$, and $K = 16$.

Effect of Number of Sub-carriers on SE

In Fig. 3.5, we investigate the effect of the number of sub-carriers K on the SE for a fixed SNR, i.e., $\text{SNR} = 15$ dB. We assume that $K = [1, 4, 8, 16, 32, 64]$ and the rest of the simulation parameters remain the same as in Fig. 3.3. We can see from Fig. 3.5 that the SE per sub-carrier per UE decreases when the number of sub-carriers K increases. This can be explained by the following two reasons. The first reason is due to the assumed power allocation approach (recall that $P_k \stackrel{\text{def}}{=} \frac{P}{K}$). Therefore, when K increases, the power allocation per sub-carrier P_k decreases. The second reason, and more importantly, is that the capacity of every equivalent channel $\mathbf{H}_{u,k}^{\text{eff}}$ decreases when K increases. This is due to the fact that the common analog precoding matrix \mathbf{F}^{RF} is computed from \mathbf{D} that is given by (3.18). Therefore, when K increases, \mathbf{F}^{RF} will span a larger subspace that is formed by the average of K covariance matrices, which results in a capacity reduction of each $\mathbf{H}_{u,k}^{\text{eff}}, \forall u, \forall k$. Note that the un-normalized SE increases with increasing K , which can be checked by multiplying each point in Fig. 3.5 by $U = 4$ and the respective value of K .

Different Beamforming Architecture

As already explained in Chapter 2, the hybrid beamforming architecture can be described by a generalized PC formulation. Therefore, we want to check the performance of the proposed algorithm considering different hybrid beamforming architectures. To this end, we assume that the number of groups can vary from $G = 1$ up to $G = 8$. The other simulation parameters remain the same as in Fig. 3.3. In

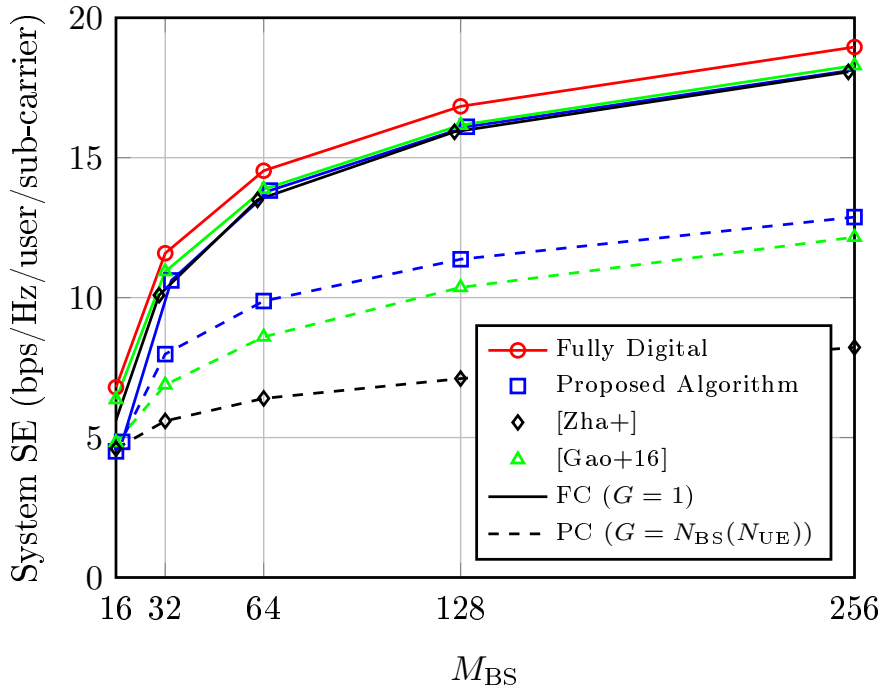


Figure 3.4: System SE versus number of antennas at the BS M_{BS} for $M_{UE} = 8$, $U = 4$, $N_{UE} = 2$, $N_{BS} = 8$, and $K = 16$.

Fig. 3.6, the SE versus G for a fixed SNR, i.e., SNR = 15 dB, is illustrated. As a reminder, $G = 1$ and $G = 8$ correspond to the FC and the PC architecture, respectively. As expected, the SE of the system decreases as the number of groups increases. It can be seen that the generalized PC formulation offers more freedom to assign the antennas and the RF chains to each other than the extreme cases of the FC and PC architectures. Therefore, a trade-off between system complexity and system performance can be taken into account.

3.5. Chapter Conclusion

In this chapter, a new HAD design method is proposed for multi-user massive MIMO OFDM mmWave systems, which updates the analog beamforming matrices sequentially and iteratively with the aim of finding a good balance between maximizing the received signal in each block and minimizing the interference leakage to the other blocks. Simulation results reveal that the proposed method has a comparable performance with the benchmark methods in the FC HAD architectures, while it significantly outperforms them in the PC and other generalized PC cases.

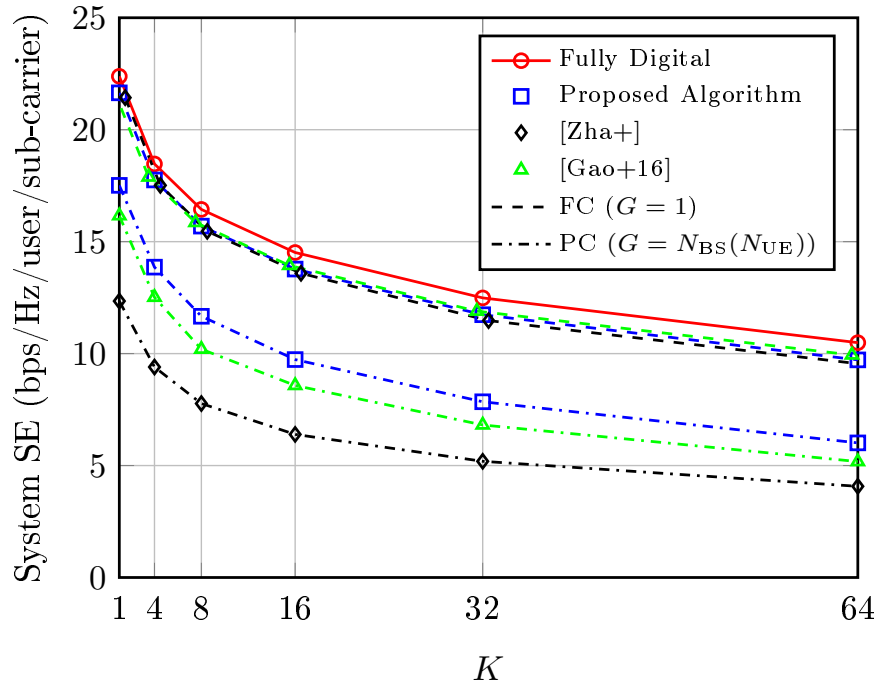


Figure 3.5: System SE versus number of sub-carriers K for $M_{\text{BS}} = 64$, $M_{\text{UE}} = 8$, $U = 4$, $N_{\text{UE}} = 2$, and $N_{\text{BS}} = 8$.

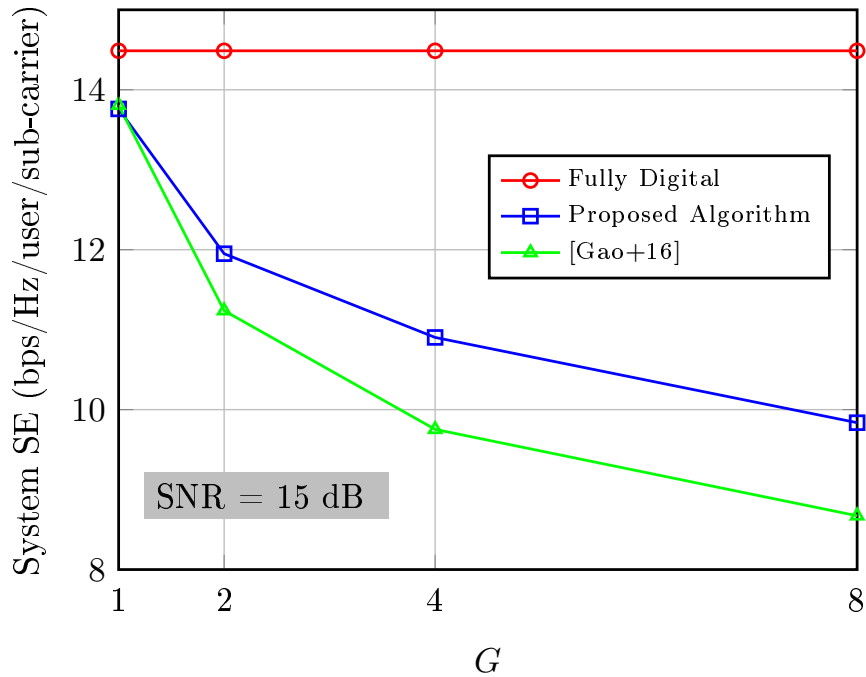


Figure 3.6: System SE versus number of groups G for $M_{\text{BS}} = 64$, $M_{\text{UE}} = 8$, $U = 4$, $N_{\text{UE}} = 2$, $N_{\text{BS}} = 8$, and $K = 16$.

Part II

RIS-Aided Wireless Communication Systems

This part of the thesis is devoted to reconfigurable intelligent surface (RIS)-aided mmWave MIMO wireless communication systems. We consider both the channel estimation problem and the design of the reflection coefficients. In particular, in Chapter 4, we consider Single RIS (S-RIS)-aided MIMO systems. First, we propose a novel channel estimation algorithm called Two-stage RIS-aided Channel Estimation (TRICE), which exploits the low-rank characteristics of the channel in the mmWave band. Our second proposed channel estimation method is called Tensor-based RIS Channel Estimation (Ten-RICE) with the aim of improving the channel estimation performance in terms of the training overhead and the channel estimation accuracy by exploiting the tensor structure of the measurement signals. Moreover, we propose a heuristic non-iterative two-step solution to obtain the RIS reflection vector in a closed-form termed as Frobenius-Norm Maximization (FroMax). In Chapter 5, we turn our focus to Double RIS (D-RIS)-aided systems. We show that the channel estimation in D-RIS-aided systems can outperform the S-RIS-aided one in terms of training overhead and channel estimation accuracy by carefully distributing the reflection elements between RIS-1 and RIS-2. Moreover, we propose two channel estimation methods by showing that the received signal admits a nested PARAFAC model and can be written as a 4-way tensor. The effectiveness of the proposed channel estimation and beamforming algorithm is illustrated by the numerical results.

Single-RIS-Aided Systems

4.1. Introduction and State of the Art

The wireless communication channel between a transmitter and a receiver has generally been regarded as uncontrollable in past wireless communication systems. Needless to say, having a favorable wireless propagation channel between the communicating ends, e.g., a strong LoS would provide a high communication performance, e.g., a higher SE and a lower transmit power. With every generation, several wireless communication technologies and solutions have been proposed to increase the probability of having a favorable wireless propagation channel. For example, heterogeneous networks (HetNets) or network densification technologies [Rom+15] have been proposed, in which a number of low-power nodes, e.g., pico nodes, are deployed within the coverage area of a high-power node, e.g., a macro node, to further improve system capacity and efficiency to overcome the coverage holes. In other words, HetNets densify the area of interest by activating several serving network nodes, which can ultimately increase the probability of having a favorable LoS wireless propagation channel. Wireless relaying techniques, e.g., using either amplify-and-forward (AF) or decode-and-forward (DF) schemes were also proposed [GAH21; BÖL20; Yil+21; Wan+22], which can be used to extend the wireless communication range, especially when the direct communication channel between the communicating ends have unfavorable or weak channel conditions. Although these technologies are valuable solutions and can improve the performance of wireless communications, they come with a high cost, complexity, and energy consumption, which is contrary to the goals of future 6G wireless communication systems that require more affordable and energy-friendly wireless communication networks.

As a solution to the above issues, RISs, also know as reconfigurable metasurfaces [Ren+19], intelligent reflecting surfaces (IRSs), software-controlled metasurfaces (SCMs), and large intelligent surfaces (LISs) [Wu+21; Lia+18a], have been proposed recently for future 6G wireless communication networks [Wu+21; Lia+18a; Lia+18c]. The RISs have a potential not only for overcoming the unfavorable channel conditions, but also for enabling unprecedented system performance while reducing the energy consumption. Generally speaking, an RIS is a 2D surface equipped with a large number of tunable units that can be realized using, e.g., inexpensive reflected arrays, software-defined metamaterials such as graphene and liquid crystals, or varactor diodes [AZY20; CTY16; Jia+22; Lia+18a]. Programmable electronic circuits can be integrated into planar structures fabricated using lithography and nano-printing methods [Jia+22]. It is worth noting that reflecting surfaces are already a known technology and have been used in different applications, e.g., radar systems, remote sensing, and satellite or deep-space communications. However, traditional reflecting surfaces are equipped with passive and fixed PSs, which make them unsuitable for dynamic mobile wireless communication networks.

Thanks to the recent developments in RF micro electromechanical systems (MEMS) and metamaterials, the reconfigurability and controllability of an RIS unit in real-time become possible [Cui+14]. Moreover, most of the RIS implementation architectures are made of passive, inexpensive, and thin electromagnetic materials, which make them easy to deploy on environment structures such as build-

ing facades and indoor walls. Due to their unique characteristics, RIS technologies have found a wide range of applications and use cases [Hon23; Wu+21; Lia+18c; CTY16], as shown in Fig. 1.4. For example, by deploying an RIS between a transmitter and a receiver, the phase, amplitude, and/or polarization of the RIS units can be adjusted in real-time to reflect the impinging signals to meet a desired objective, e.g., to add constructively at the intended receiver and/or to add destructively at unintended receivers. In addition to the reflection, there are other operation modes for RISs such as transmission, sensing, absorption, and simultaneous transmission and reflection [Jia+22].

Moreover, from a hardware implementation point of view, the RISs can be divided into two major groups, i.e., passive and semi-passive RISs. In contrast to passive RISs, semi-passive RISs include a few active elements that are connected with RF chains [AV20]. Clearly, the active elements increase the cost and power consumption of semi-passive RISs. However, these active elements can be used to improve the performance of RIS-aided systems and enable pilot-based channel estimation methods, as described in [AV20]. Note that the general idea of RIS technology resembles, to some extent, the AF-based relays [GAH21; BÖL20; Yil+21; Wan+22]. For example, both technologies can be used to extend the wireless communication range by deploying an RIS or a relay node between a transmitter and a receiver. However, RISs have some unique features that renders them more appealing for future 6G wireless networks, e.g.,

- (i) They have low power consumption, cost, and complexity.
- (ii) They are able to operate in the more spectrally efficient full-duplex (FD) mode, unlike AF relays that usually operates in half-duplex (HD) mode.
- (iii) They are easy to deploy on environment structures such as building facades and indoor walls, due to their thin structure.

In the following, we give a brief overview and summarize some achievements in the state-of-the-art.

4.1.1. Fundamental Performance Metrics

In the earlier works on RIS-aided systems, some fundamental performance metrics have been derived, showing their potential in achieving the promising gains mentioned above [Yan+20c; Yan+20a; BA20; Fer+20]. In [Yan+20c], the secrecy outage probability of an RIS-aided system in the presence of a LoS and an eavesdropper was derived using the central limit theorem, making the results valid only for a large number of reflecting elements. In [Yan+20a], accurate approximations for the channel distributions and other performance metrics of RIS-aided system were derived for Rayleigh fading channels. Following the same system assumptions as in [Yan+20a], the authors in [BA20] presented closed-form expressions for the outage probability, symbol error rate, and ergodic capacity. A closed-form expression for the bit error rate is provided in [Fer+20] for BPSK and M -QAM modulation techniques. In [Yan+21], the exact analysis of the outage probability and a closed-form expression for the asymptotic sum rate is derived, which shows that the capacity scaling law in multi-RIS-aided systems depends strongly on the number of RISs as well as the number of reflecting elements.

4.1.2. Channel Estimation

The channel acquisition in RIS-aided systems faces several challenges. For instance, assuming a passive RIS implementation to reduce the RIS cost and complexity, the propagation channel can only be sensed

and estimated at the receiver. Furthermore, the large number of channel coefficients to be estimated limits the feasibility of CSI acquisition within a practical coherence time, since an RIS is expected to have a massive number of passive reflecting elements. Therefore, efficient channel estimation methods for RIS-aided systems are required to guarantee successful communications and reap the RIS potentials. Recently, channel estimation methods for RIS-aided systems have been proposed, e.g., using LS-based methods as in [MJ19; JD20; ZZ20b; AA20], or MMSE-based methods as in [Nad+20]. An overview of the LS and the MMSE estimator for full illumination is provided in [JGU22]. In [MJ19; Yan+20d], the authors propose an on-off-based channel estimation technique for single-user RIS-aided systems where only one reflecting element is turned on at a time. However, the sub-optimality of the on-off method was shown in [JD20]. On the other hand, a DFT-based channel estimation method is considered in [ZZ20b; JD20], where an LS-based channel estimation is proposed. Despite the same training overhead of the DFT-based and on-off methods, the DFT-based method outperforms the on-off method since all RIS elements are always on. The optimization of discrete phase shifts in [YZZ20] and joint pilot and phase optimization are investigated in [Kan20]. In [NYS22], the authors proposed training signal design considering the MSE of the channel estimation and approaching it using CRB. They have shown that in mmWave channels, the channel parameters, i.e., path gains and path angles, can be derived in a closed-form by exploiting the sparsity of the channel under the following assumptions (i) Bayesian parameter, where the path gain and path angles are modeled randomly, and (ii) hybrid parameter, where the path gains are modeled as random parameters and the path angles are assumed to be unknown deterministic parameters. From their simulation results, it was shown that less number of training symbols are required in mmWave RIS-aided channels than the number of reflecting elements in the RIS. In [PD21], the authors proposed a channel estimation method to reduce the training overhead by considering the correlation of the channels between the adjacent elements of RIS. In [HY20], the authors formulated a single-user channel estimation problem as a combined sparse matrix factorization and matrix completion problem and used a CS-based technique to solve it. A multi-user scenario was considered in [WLC20], where it was pointed out that the training overhead can be reduced since the channel between the BS and the RIS is common for all users. However, most of these works require the number of training sub-frames to be, at least, equal to the number of RISs reflecting elements, which increases the training overhead and complexity.

Considering the low-rank structure of the mmWave channels, the channel training overhead and complexity can be reduced significantly, as it has been shown in [TAA21; WGA20; He+20a; Che+23; Wan+20]. In these works, every channel matrix is modeled as a summation of L paths, where L is much smaller than the number of transmit and receive antennas, and every path is completely characterized by a DoD, a DoA, and a complex path gain. Therefore, the channel estimation is formulated as a sparse recovery problem, for which CS techniques [Don06] can be used to efficiently recover the channel parameters using a small training overhead. In [TAA21], the above problem is facilitated by assuming that the RIS has a few active elements, which, however, increases the deployment cost and the energy consumption of RIS-aided systems. In [WGA20], the authors assumed that the BS-to-RIS channel is perfectly known, while in [He+20a], the cascade channel matrix is assumed to have a single path, i.e., $L = 1$. Differently, the authors in [Wan+20] proposed a general sparse recovery formulation for $L \geq 1$ scenarios. In most of these works, however, the channel parameters are assumed to fall perfectly on a grid, which may never be true in practice. Therefore, there exists a trade-off between the estimation accuracy and the complexity, where both increase as a function of the grid resolution. Due to the

multi-dimensionality of the cascaded channel, a 4D sensing matrix is required by the method proposed in [Wan+20], which makes it computationally prohibitive even with low grid resolutions. In [Wan+20; JD20], the authors provide a sparse representation of the cascaded channel in mmWave RIS-aided MISO systems. In [Fan+16], an iterative refinement method was proposed to estimate the channel parameter under the assumption that all available channels have only LoS links. A codebook-based scheme is studied in [He+20b] to design the beamforming and reflection matrices jointly without estimating the MIMO channel parameters. In [HWJ21], the atomic norm minimization is used to estimate the channel parameters sequentially. In [Che+23], a two-step multi-user joint channel estimation is proposed for an RIS-aided MISO system, where it is assumed that the channel between the BS and the RIS is common among all users. In [WSD21], motivated by the assumption that the angular cascaded channels associated with different users enjoy completely common non-zero rows and partially non-zero columns, the authors proposed a double-structured OMP channel estimation method. There exist some other strategies to reduce the pilot overhead, e.g., an element grouping strategy [Yan+20d], a two-stage channel estimation method exploiting correlations in the cascaded channel [Zho+22].

On the other hand, tensor-based signal modeling and processing methods offer essential benefits over their bi-linear (matrix) counterparts, since they can improve the parameter identifiability owing to the strong uniqueness features of tensor decomposition [FA14]. In [AAB21], it is demonstrated that the received signals in RIS-aided MIMO wireless communication systems can be expressed as a 3-way tensor admitting a CPD. However, the proposed method in [AAB21] assumes sub-6 GHz systems and, therefore, needs a large number of training sub-frames, similarly to [MJ19; JD20; Nad+20; Zha+20].

4.1.3. RIS Beamforming Design

Beamforming is a fundamental technique in MIMO systems, which enhances the possibility of the signal reception at users and increases the system capacity by making the transmitted signals more directional instead of radiating randomly. This is achieved by adjusting the phase and amplitude of every transmit antenna. An RIS node can be regarded, to some extent, as a MIMO node, where the phase and amplitude of every RIS element can be adjusted to create a desired beam shape and direction, which is the key guarantee for RISs to improve the system performance. In contrast to the conventional system in which beamforming is done only at the transmitter and/or receiver, in RIS-aided communication systems, we jointly optimize the active beamforming at the transmitter and/or receiver and the reflection coefficients at the RIS, providing a new degree of freedom to improve the communication performance. However, the involved optimization problem is, in general, non-convex, due to its joint optimization and some additional non-convex constraints, e.g., constant modulus constraints imposed due to the RIS hardware requirements. Moreover, RISs are expected to have a large number of reflecting elements, which calls for low-complexity RIS beamforming/reflection design methods to make RIS-aided systems practically possible. Therefore, relaxation and alternating optimization techniques are commonly used in the literature to obtain, at least, a locally optimal solution.

Several works have been established for improving the transmission rate by using RIS-aided wireless systems [ZZ20a]. RIS reflection design, in particular, has been extensively investigated under various setups and objectives [ZZ20a; WZ18; WZ19; Nad+19]. It was shown in [WZ19; WZ18] that an RIS-aided system with N reflecting elements asymptotically increases the received signal power or SNR on the order of $\mathcal{O}(M_S^2)$ when M_S is large enough. Such a squared power gain compared to the constant power gain in massive MIMO systems without RISs [NLM13] is the result of combining both

received and transmit arrays for energy harvesting and reflect beamforming. Moreover, it was shown in [WZ19] that a UE located in the coverage region of the RIS can tolerate more interference compared to the one outside that region. The authors in [ZZ20a] considered the capacity maximization and proposed an alternating optimization approach to find a locally optimal solution by iteratively optimizing the transmit covariance matrix and one of the RIS reflection coefficients at a time. However, such an approach increases the computational complexity and becomes a limiting factor in practice, especially in a massive RIS setup. In [JS19], the authors proposed an alternating difference-of-convex programming algorithm to solve the non-convex bi-quadratic problem for over-the-air communication scenarios. In [Nad+19], the reflection elements are obtained for rank-one MIMO channels. In [WZ18; WZ19], the authors proposed a semidefinite relaxation (SDR) approach to overcome the unit modulus constraints of RIS elements. However, the proposed algorithms obtain an approximate solution with no guarantee of optimality. A simple phase extraction approach is proposed in [Han+19] obtained from an approximation of the ergodic capacity.

A joint design of the precoding matrix at the transmitter and the reflecting elements at the RIS is proposed in [WZ19]. According to the simulation results of [WZ19], the RIS-aided system achieves a significantly larger transmission rate than its counterpart without RIS. In [Hua+18b], an extension of [WZ19] for multi-user scenarios is studied. The performance of the RIS-aided systems is studied in [Hua+19; Hua+18a] in terms of their energy efficiency. It was shown that the RIS-aided system has 300% more energy efficiency compared to a regular multi-antenna AF relaying system. In [WZ18], the reflecting matrix is designed to maximize the total received power at the UE in single-user RIS-aided MISO system. In [Hua+19], two computationally efficient algorithms based on alternating maximization are proposed to assign the transmit power at the BS and the reflector values of the RIS such that the energy efficiency of the system is maximized. Similarly, the authors in [Hua+18a] proposed design methods for power allocation and induced phases to either maximize the energy efficiency or SE of the system in multi-user MISO systems. However, the proposed designs are impractical due to the requirement of the availability of global CSI, i.e., from all UEs, at the RIS, which makes the signal exchange overhead enormously high. The authors in [Nad+19] proposed a precoding and beamforming designing technique for multi-user MISO systems, where the channel between the BS and the RIS is LoS and the channel between the RIS and the UEs are Rayleigh correlated channels. Then, the proposed beamforming and precoding matrices are designed considering the impact of both the rank structure of the LoS channel matrix and the spatial correlation between the reflection elements at the RIS.

Furthermore, a beamforming design for MIMO systems with a maximization of the minimum SINR criterion was proposed in [ZZ20a]. In [ZZ20a], the authors derived the channel capacities of the RIS-aided MIMO channel in the asymptotically low and high SNR region. In addition, an algorithm to design beamforming matrices so that the transmission rate is maximized was proposed in [ZZ20a]. According to the simulation results, it was shown that the transmission rate of the system is significantly improved in systems with RISs compared to the ones without RISs. Apart from the above works, numerous other works have been reported to improve the performance of wireless communication networks via RISs by not only designing proper passive beamforming but also transmitting information, which is called passive beamforming and information transfer (PBIT) [YYK20]. The main idea of PBIT is to enable RISs to transmit their private data to the receiver. The private data of an RIS can be its environmental monitoring data, the wireless control link, etc. Another application is to enhance the secrecy rate of the RIS-aided system by designing the precoding and reflecting beamforming such that

the signal is directed to the legitimate UE while degrading its reception at the eavesdroppers [GWZ20; CZZ19; She+19; YXS19].

4.1.4. Chapter Contribution

In this thesis, we consider the channel estimation problem in a single-user S-RIS-aided mmWave MIMO communication system, similarly to [Wan+20], where the RIS has passive reflecting elements and the direct link between the transmitter and the receiver is assumed to be blocked or pre-estimated by turning the RIS elements off, as in [MJ19] and illustrated in Fig. 4.1.

First, we propose a Two-stage RIS-aided Channel Estimation (TRICE) framework using a structured channel training procedure. In the first stage, the DoDs of the transmitter-to-RIS channel and the DoAs of the RIS-to-receiver channel are estimated. In the second stage, by using the estimated channel parameters in the first stage, the effective azimuth and elevation angles of the cascaded transmitter-to-RIS-to-receiver channel at the RIS are estimated, one-by-one, including the effective complex path gains. In both stages, we show that the parameter estimation can be carried out via a multi-dimensional DoA estimation scheme, for which several solutions exist as in [ZH17a; ZH17b; SPP18; AAH19a; MRM16; Cao+18], among many others. Then, we extend our TRICE framework and propose a CP tensor decomposition method, termed Tensor-based RIS Channel Estimation (TenRICE), by jointly exploiting the tensor structure of the received signals and the low-rank nature of mmWave channels. Using the TenRICE method, the transmitter-to-RIS and the RIS-to-receiver channels can be estimated separately, up to a trivial scaling factor.

After that, we formulate the beamforming and the RIS reflection design as a SE maximization problem. Due to its non-convexity, we propose a heuristic non-iterative two-step solution, where the RIS reflection vector is obtained, in contrast to [ZZ20a], in a closed-form using a Frobenius-norm Maximization strategy, termed FroMax. The content of this chapter has been published in [Ard+21; Ghe+21].

4.1.5. Chapter Organization

In Section 4.2, we describe first the system model with two main phases, e.g., the channel estimation phase and the data transmission phase. In the last part of Section 4.2, we give a brief introduction to the classical Saleh-Valenzuela channel model [SV87]. Then, our proposed channel estimation approaches are presented in Section 4.3, where we first explain the proposed TRICE method and then the improved TenRICE method. In Section 4.4, we propose our FroMax beamforming algorithm. In Section 4.5, we discuss the performance of our proposed algorithms in terms of identifiability conditions and computational complexity. The numerical results are provided in Section 4.6 to evaluate the performance of the proposed algorithms. Finally, we provide the conclusion in Section 4.7.

4.2. System Configuration

4.2.1. System Model

We consider a single-user mmWave MIMO communication system as depicted in Fig. 4.1, where a transmitter with M_T antennas is communicating with a receiver with M_R antennas via an RIS-aided

MIMO channel. It is assumed that the direct channel between the transmitter and the receiver is poor or unavailable due to blockage.

The RIS is composed of M_S inexpensive reflecting elements arranged uniformly with an inter-element spacing of a half-wavelength on a rectangular surface with M_S^v vertical and M_S^h horizontal elements such that $M_S = M_S^v \cdot M_S^h$.

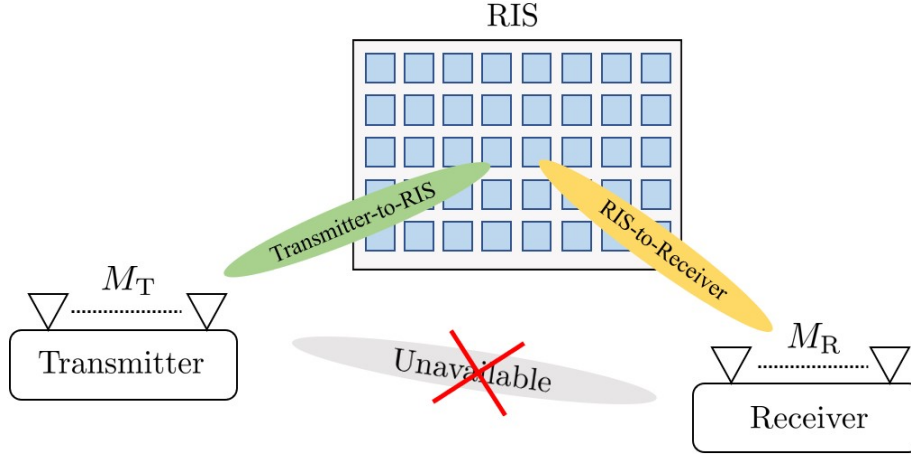


Figure 4.1: An RIS-aided mmWave MIMO communication system.

Let $\mathbf{H}_T \in \mathbb{C}^{M_S \times M_T}$ be the transmitter-to-RIS channel and $\mathbf{H}_R \in \mathbb{C}^{M_R \times M_S}$ be the RIS-to-receiver channel with $\mathbb{E}\{\|\mathbf{H}_T\|_F^2\} = M_S M_T$ and $\mathbb{E}\{\|\mathbf{H}_R\|_F^2\} = M_S M_R$, as illustrated in Fig. 4.2. We assume a block-fading channel scenario, where \mathbf{H}_T and \mathbf{H}_R remain constant during every channel coherence block and change from block to block. We assume that every block is divided into two sub-blocks, i.e., one for channel estimation and another for data transmission, see Fig. 4.2.

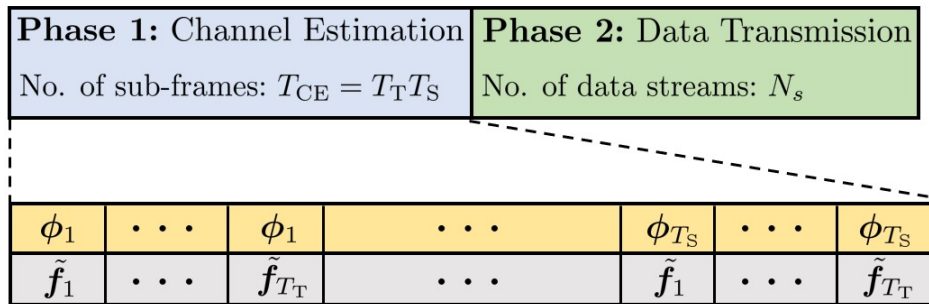


Figure 4.2: One channel coherence block.

Channel Estimation Phase:

As shown in Fig. 4.2, we conduct a channel training procedure at the beginning of each block, which comprises T_{CE} frames divided into $T_{\text{T}} \cdot T_{\text{S}}$ sub-frames, i.e., $T_{\text{CE}} = T_{\text{T}} \cdot T_{\text{S}}$, where T_{T} denotes the number of training beams at the transmitter and T_{S} represents the number of training beams at the RIS. The received signal at the receiver in the (s, t) th sub-frame is given as

$$\mathbf{y}_{s,t} = \mathbf{W}^{\text{H}} \mathbf{H}_{\text{R}} \text{diag}\{\phi_s\} \mathbf{H}_{\text{T}} \tilde{\mathbf{f}}_t s_t + \mathbf{W}^{\text{H}} \mathbf{n}_{s,t} \in \mathbb{C}^{N_{\text{R}}}, \quad (4.1)$$

where $\mathbf{W} \in \mathbb{C}^{M_{\text{R}} \times N_{\text{R}}}$ is a fixed training decoding matrix with N_{R} beams, $\tilde{\mathbf{f}}_t \in \mathbb{C}^{M_{\text{T}}}$ is the t th training vector of the transmitter with $\|\tilde{\mathbf{f}}_t\|_2^2 = 1$, $t \in \{1, \dots, T_{\text{T}}\}$, $\phi_s \in \mathbb{C}^{M_{\text{S}}}$ is the s th training vector of the RIS with $|\phi_s[i]| = \frac{1}{\sqrt{M_{\text{S}}}}, \forall i, s \in \{1, \dots, T_{\text{S}}\}$, $s_t \in \mathbb{C}$ is the pilot symbol, and $\mathbf{n}_{s,t} \in \mathbb{C}^{M_{\text{R}}}$ is the AWGN vector having zero-mean circularly symmetric complex-valued entries with variance σ_n^2 . Note that a single training vector in each sub-frame means that for the transmitter with a HAD beamforming architecture, a single RF chain can be used during the channel training to reduce the energy consumption.

By stacking the measurement vectors $\{\mathbf{y}_{s,t}\}_{t=1}^{T_{\text{T}}}$ next to each other as

$$\mathbf{Y}_s = [\mathbf{y}_{s,1}, \dots, \mathbf{y}_{s,T_{\text{T}}}] \in \mathbb{C}^{N_{\text{R}} \times T_{\text{T}}}. \quad (4.2)$$

Applying (1.20), we vectorize \mathbf{Y}_s as

$$\mathbf{y}_s = (\mathbf{F}^{\text{T}} \mathbf{H}_{\text{T}}^{\text{T}} \diamond \mathbf{W}^{\text{H}} \mathbf{H}_{\text{R}}) \phi_s + \mathbf{n}_s \in \mathbb{C}^{N_{\text{R}} T_{\text{T}}}, \quad (4.3)$$

where $\mathbf{n}_s = [(\mathbf{W}^{\text{H}} \mathbf{n}_{s,1})^{\text{T}}, \dots, (\mathbf{W}^{\text{H}} \mathbf{n}_{s,T_{\text{T}}})^{\text{T}}]^{\text{T}}$ and $\mathbf{F} = [\tilde{\mathbf{f}}_1 s_{11}, \dots, \tilde{\mathbf{f}}_{T_{\text{T}}} s_{T_{\text{T}}}] \in \mathbb{C}^{M_{\text{T}} \times T_{\text{T}}}$. Then, we write the measurement matrix $\mathbf{Y} \in \mathbb{C}^{N_{\text{R}} T_{\text{T}} \times T_{\text{S}}}$ by stacking $\{\mathbf{y}_s\}_{s=1}^{T_{\text{S}}}$ next to each other and applying (1.11) as

$$\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_{T_{\text{S}}}] \quad (4.4)$$

which can be written as

$$\begin{aligned} \mathbf{Y} &= (\mathbf{F}^{\text{T}} \mathbf{H}_{\text{T}}^{\text{T}} \diamond \mathbf{W}^{\text{H}} \mathbf{H}_{\text{R}}) \Phi + \mathbf{N} \\ &= (\mathbf{F}^{\text{T}} \otimes \mathbf{W}^{\text{H}}) (\mathbf{H}_{\text{T}}^{\text{T}} \diamond \mathbf{H}_{\text{R}}) \Phi + \mathbf{N} \\ &= (\mathbf{F}^{\text{T}} \otimes \mathbf{W}^{\text{H}}) \mathbf{H}_{\text{c}} \Phi + \mathbf{N}, \end{aligned} \quad (4.5)$$

where

$$\mathbf{H}_{\text{c}} = \mathbf{H}_{\text{T}}^{\text{T}} \diamond \mathbf{H}_{\text{R}} \in \mathbb{C}^{M_{\text{T}} M_{\text{R}} \times M_{\text{S}}} \quad (4.6)$$

denotes the cascaded channel matrix, $\Phi = [\phi_1, \dots, \phi_{T_{\text{S}}}] \in \mathbb{C}^{M_{\text{S}} \times T_{\text{S}}}$ holds the T_{S} phase shift vectors of the RIS, and $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_{T_{\text{S}}}]$.

Data Transmission Phase:

Given the estimated channels \mathbf{H}_{T} and \mathbf{H}_{R} , the transmitter first designs the precoding matrix $\mathbf{P} \in \mathbb{C}^{M_{\text{T}} \times N_{\text{s}}}$, the RIS reflection vector $\boldsymbol{\omega} \in \mathbb{C}^{M_{\text{S}}}$ with $|\omega[i]| = \frac{1}{\sqrt{M_{\text{S}}}}, \forall i$, and the decoding matrix $\mathbf{Q} \in \mathbb{C}^{M_{\text{R}} \times N_{\text{s}}}$ to transmit the data vector $\mathbf{s} \in \mathbb{C}^{N_{\text{s}}}$ of N_{s} data symbols with $\mathbb{E}\{\mathbf{s}\mathbf{s}^{\text{H}}\} = \mathbf{I}_{N_{\text{s}}}$ to the receiver. Therefore,

the received signal vector at the receiver is given as

$$\hat{\mathbf{s}} = \mathbf{Q}^H \mathbf{H}_R \text{diag}\{\boldsymbol{\omega}\} \mathbf{H}_T \mathbf{P} \mathbf{s} + \mathbf{Q}^H \mathbf{n} \in \mathbb{C}^{N_s}, \quad (4.7)$$

where $\mathbf{n} \in \mathbb{C}^{M_R}$ denotes the AWGN with $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_{M_R}$. The system SE can be expressed as

$$\text{SE} = \log_2 |\mathbf{I}_{N_s} + \mathbf{R}^{-1} \mathbf{Q}^H \mathbf{H}_R \text{diag}\{\boldsymbol{\omega}\} \mathbf{H}_T \mathbf{P} \mathbf{P}^H \mathbf{H}_T^H \text{diag}\{\boldsymbol{\omega}^*\} \mathbf{H}_R^H \mathbf{Q}|, \quad (4.8)$$

where

$$\mathbf{R} = \sigma_n^2 \mathbf{Q}^H \mathbf{Q} \in \mathbb{C}^{N_s \times N_s} \quad (4.9)$$

is the noise covariance matrix. The objective is to design the beamforming matrices \mathbf{P} and \mathbf{Q} and the RIS reflection vector $\boldsymbol{\omega}$ such that it maximizes the achievable SE of the system, which can be expressed as

$$\begin{aligned} \max_{\mathbf{P}, \mathbf{Q}, \boldsymbol{\omega}} \quad & \log_2 |\mathbf{I}_{N_s} + \mathbf{R}^{-1} \mathbf{Q}^H \mathbf{H}_R \text{diag}\{\boldsymbol{\omega}\} \mathbf{H}_T \mathbf{P} \mathbf{P}^H \mathbf{H}_T^H \text{diag}\{\boldsymbol{\omega}^*\} \mathbf{H}_R^H \mathbf{Q}| \\ \text{s.t.} \quad & \|\mathbf{P}\|_F^2 \leq P_{\max} \quad \text{and} \quad |[\boldsymbol{\omega}]_{[i]}| = \frac{1}{\sqrt{M_S}}, \forall i, \end{aligned} \quad (4.10)$$

where P_{\max} is the transmit power at the transmitter.

4.2.2. Channel Model

Here, we consider mmWave MIMO wireless communication systems [SV87]. We assume that the transmitter and the receiver employ uniform linear arrays (ULAs)¹. It is assumed that there are L_T and L_R paths for \mathbf{H}_T and \mathbf{H}_R , respectively. In the mmWave bands, it has been observed that the number of paths is much smaller compared to the number of antennas, i.e., $\text{rank}\{\mathbf{H}_T\} \leq L_T$ and $\text{rank}\{\mathbf{H}_R\} \leq L_R$. Similarly to [Wan+20], \mathbf{H}_T and \mathbf{H}_R are modeled according to the Saleh-Valenzuela model [SV87] as

$$\mathbf{H}_T = \sum_{l=1}^{L_T} g_{T,l} \boldsymbol{\nu}_{2D}(\mu_{T,l}^v, \mu_{T,l}^h) \boldsymbol{\nu}_{1D}(\psi_{T,l})^T \quad (4.11)$$

$$\mathbf{H}_R = \sum_{l=1}^{L_R} g_{R,l} \boldsymbol{\nu}_{1D}(\psi_{R,l}) \boldsymbol{\nu}_{2D}(\mu_{R,l}^v, \mu_{R,l}^h)^T, \quad (4.12)$$

where $g_{X,l} \sim \mathcal{CN}(0, 1)$, $X \in \{T, R\}$, denotes the l th path gain, $\psi_{T,l} \in [0, 2\pi]$ is the l th DoD spatial frequency at the transmitter, $\psi_{R,l} \in [0, 2\pi]$ is the l th DoA spatial frequency at the receiver, $\mu_{T,l}^h \in [0, 2\pi]$ and $\mu_{T,l}^v \in [0, \pi]$ are the l th horizontal and vertical DoD spatial frequencies and $\mu_{R,l}^h \in [0, 2\pi]$ and $\mu_{R,l}^v \in [0, \pi]$ are the l th horizontal and vertical DoA spatial frequencies at the RIS. Let Δ denote the antenna spacing and λ be the signal wavelength. Then, the spatial frequencies are defined as

$$\psi_{X,l} = 2\pi \frac{\Delta}{\lambda} \cos(\theta_{X,l}) \quad (4.13)$$

$$\mu_{X,l}^v = 2\pi \frac{\Delta}{\lambda} \cos(\theta_{X,l}^h) \sin(\theta_{X,l}^v) \quad (4.14)$$

¹The extension of the proposed channel estimation frameworks to scenarios where the transmitter and/or the receiver are equipped with uniform rectangular arrays (URAs) is straightforward.

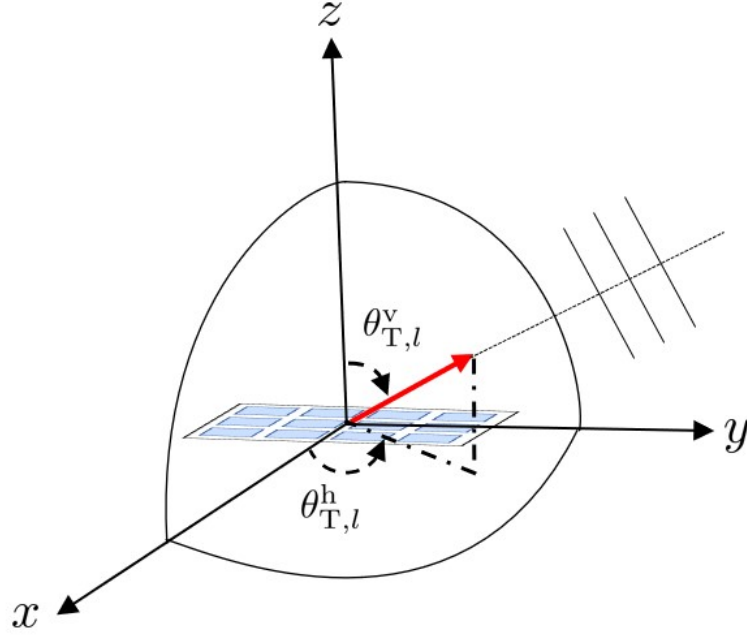


Figure 4.3: Definition of azimuth ($\theta_{T,l}^h$) and co-elevation ($\theta_{T,l}^v$) angles of an impinging planar wavefront on an RIS with URA structure.

$$\mu_{X,l}^h = 2\pi \frac{\Delta}{\lambda} \sin(\theta_{X,l}^h) \sin(\theta_{X,l}^v) \quad (4.15)$$

where $\theta_{X,l} \in [-180^\circ, 180^\circ]$ is the l th path angle in the angular domain, while $\theta_{X,l}^h \in [-180^\circ, 180^\circ]$ and $\theta_{X,l}^v \in [-90^\circ, 90^\circ]$ are the l th path azimuth and elevation angle at the RIS in the angular domain, respectively, with $X \in \{T, R\}$ as illustrated in 4.3. The 2D array steering vector $\boldsymbol{\nu}_{2D}(\mu_{X,l}^v, \mu_{X,l}^h)$ can be expressed as a function of the 1D array steering vectors, $\boldsymbol{\nu}_{1D}(\mu_{X,l}^v)$ and $\boldsymbol{\nu}_{1D}(\mu_{X,l}^h)$, as

$$\boldsymbol{\nu}_{2D}(\mu_{X,l}^v, \mu_{X,l}^h) = \boldsymbol{\nu}_{1D}(\mu_{X,l}^v) \diamond \boldsymbol{\nu}_{1D}(\mu_{X,l}^h). \quad (4.16)$$

with $X \in \{T, R\}$. For a given spatial frequency α , the steering vector $\boldsymbol{\nu}_{1D}(\alpha)$ is given as

$$\boldsymbol{\nu}_{1D}(\alpha) = [1, e^{j\alpha}, \dots, e^{j(M-1)\alpha}]^T \in \mathbb{C}^M. \quad (4.17)$$

Then, the channel matrices \mathbf{H}_T and \mathbf{H}_R in (4.11) and (4.12) can be written in a compact form as

$$\mathbf{H}_T = \mathbf{B}_T \mathbf{G}_T \mathbf{A}_T^T = (\mathbf{B}_T^v \diamond \mathbf{B}_T^h) \mathbf{G}_T \mathbf{A}_T^T \quad (4.18)$$

$$\mathbf{H}_R = \mathbf{A}_R \mathbf{G}_R \mathbf{B}_R^T = \mathbf{A}_R \mathbf{G}_R (\mathbf{B}_R^v \diamond \mathbf{B}_R^h)^T, \quad (4.19)$$

where \mathbf{A}_X and \mathbf{B}_X^y denote the transmit/receive array steering matrices and \mathbf{G}_X contains the path gains,

which are given as

$$\mathbf{A}_X = [\boldsymbol{\nu}_{1D}(\psi_{X,1}), \dots, \boldsymbol{\nu}_{1D}(\psi_{X,L_X})] \in \mathbb{C}^{M_X \times L_X} \quad (4.20)$$

$$\mathbf{B}_X^y = [\boldsymbol{\nu}_{1D}(\mu_{X,1}^y), \dots, \boldsymbol{\nu}_{1D}(\mu_{X,L_X}^y)] \in \mathbb{C}^{M_s^y \times L_X} \quad (4.21)$$

$$\mathbf{G}_X = \text{diag}\{g_{X,1}, \dots, g_{X,L_X}\}, \quad (4.22)$$

for $X \in \{\text{T}, \text{R}\}$ and $y \in \{\text{v}, \text{h}\}$.

4.3. Algorithms for Channel Estimation Phase

4.3.1. Baseline Channel Estimation Algorithm

One direct solution is to use the LS-based method. By applying (1.18), the vectorized form of (4.5) can be written as

$$\mathbf{y} = \text{vec}\{\mathbf{Y}\} = \boldsymbol{\Upsilon} \mathbf{h}_c + \mathbf{n}, \quad (4.23)$$

where $\boldsymbol{\Upsilon} = \boldsymbol{\Phi}^\top \otimes \mathbf{F}^\top \otimes \mathbf{W}^\text{H} \in \mathbb{C}^{T_s T_T N_R \times M_s M_T M_R}$, $\mathbf{h}_c = \text{vec}\{\mathbf{H}_c\} \in \mathbb{C}^{M_T M_R M_s}$, and $\mathbf{n} = \text{vec}\{\mathbf{N}\}$. Therefore, an estimate to the channel vector \mathbf{h}_c can be obtained as

$$\hat{\mathbf{h}}_{c, \text{LS}} = \boldsymbol{\Upsilon}^+ \mathbf{y}. \quad (4.24)$$

In spite of the simplicity of the LS scheme, accurate channel estimation can only be achieved if $T_{\text{CE}} = T_T T_s \geq \frac{M_T M_R M_s}{N_R}$. In massive MIMO systems, the LS-based scheme becomes impractical since it requires a large number of training sub-frames T_{CE} and a long channel coherence time. In the LS estimation the composite channel matrix is explicitly estimated, i.e., each entry of \mathbf{H}_c instead of only a few model parameters according to (4.5). Therefore, we propose two channel estimation methods that exploit the channel structures and directly estimate the channel matrices \mathbf{H}_T and \mathbf{H}_R .

4.3.2. Proposed TRICE Algorithm

We are interested in estimating the channel matrices \mathbf{H}_T and \mathbf{H}_R by exploiting their structures. From (4.18) and (4.19), the cascaded channel matrix \mathbf{H}_c can be written as

$$\mathbf{H}_c = (\mathbf{A}_T \mathbf{G}_T \mathbf{B}_T^\top \diamond \mathbf{A}_R \mathbf{G}_R \mathbf{B}_R^\top) \stackrel{(a)}{=} (\mathbf{A}_T \otimes \mathbf{A}_R) \mathbf{G} \mathbf{B}, \quad (4.25)$$

where $\mathbf{G} = (\mathbf{G}_T \otimes \mathbf{G}_R) \in \mathbb{C}^{L \times L}$, $\mathbf{B} = (\mathbf{B}_T^\top \diamond \mathbf{B}_R^\top) \in \mathbb{C}^{L \times M_s}$, $L = L_T L_R$, and $\stackrel{(a)}{=}$ is obtained from (1.11). Inserting (4.25) into (4.5) yields

$$\begin{aligned} \mathbf{Y} &= (\mathbf{F}^\top \otimes \mathbf{W}^\text{H}) (\mathbf{A}_T \otimes \mathbf{A}_R) \mathbf{G} \mathbf{B} \boldsymbol{\Phi} + \mathbf{N} \\ &\stackrel{(b)}{=} (\mathbf{F}^\top \mathbf{A}_T \otimes \mathbf{W}^\text{H} \mathbf{A}_R) \mathbf{G} \mathbf{B} \boldsymbol{\Phi} + \mathbf{N}, \end{aligned} \quad (4.26)$$

where (b) is obtained from (1.10). Defining $\mathbf{A} = (\mathbf{F}^\top \mathbf{A}_T \otimes \mathbf{W}^\text{H} \mathbf{A}_R) \in \mathbb{C}^{T_T N_R \times L_T L_R}$ and $\mathbf{X} = \mathbf{G} \mathbf{B} \boldsymbol{\Phi} \in \mathbb{C}^{L_T L_R \times T_s}$, the measurement matrix \mathbf{Y} can be expressed as

$$\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{N} \in \mathbb{C}^{N_R T_T \times T_s}. \quad (4.27)$$

Observing (4.27), we can see that \mathbf{A} is completely characterized by the frequency vectors at the transmitter and receiver defined as

$$\boldsymbol{\psi}_T = [\psi_{T,1}, \dots, \psi_{T,L_T}]^\top, \quad (4.28)$$

$$\boldsymbol{\psi}_R = [\psi_{R,1}, \dots, \psi_{R,L_R}]^\top. \quad (4.29)$$

Therefore, estimating $\boldsymbol{\psi}_T$ and $\boldsymbol{\psi}_R$ from (4.27) is, in fact, a 2D DoA estimation problem, where several methods exist in the literature, such as in [ZH17a; ZH17b; SPP18; AAH19a; MRM16; Cao+18], among many others. For instance, the DFT-beamspace ESPRIT methods of [ZH17a; ZH17b; ZRH21] can be readily applied to estimate $\boldsymbol{\psi}_T$ and $\boldsymbol{\psi}_R$ in a closed-form with guaranteed automatic pairing [HN98]. While subspace-based methods perform asymptotically optimal, they suffer from a performance degradation in the case of difficult scenarios, such as high noise power and small number of measurement vectors. Alternatively, CS techniques [SPP18; AAH19a; MRM16; Cao+18] have been shown to provide an attractive alternative to subspace-based methods, yielding a good estimation performance even in difficult scenarios. To show this, we note that (4.27) can be written in a sparse form as

$$\mathbf{Y} \approx (\mathbf{F}^\top \bar{\mathbf{A}}_T \otimes \mathbf{W}^\text{H} \bar{\mathbf{A}}_R) \bar{\mathbf{X}} + \mathbf{N} \in \mathbb{C}^{N_R T_T \times T_S}, \quad (4.30)$$

where $\bar{\mathbf{A}}_T \in \mathbb{C}^{M_T \times \bar{L}_T}$ and $\bar{\mathbf{A}}_R \in \mathbb{C}^{M_R \times \bar{L}_R}$ represent two dictionary matrices, in which $\bar{L}_T \gg L_T$ and $\bar{L}_R \gg L_R$ define the number of grid points or, in other words, the grid resolution, while $\bar{\mathbf{X}} \in \mathbb{C}^{\bar{L}_T \bar{L}_R \times T_S}$ is an L row-sparse matrix [SPP18]. Here, (4.30) can be written with equality if, and only if, the true angles $\boldsymbol{\psi}_T$ and $\boldsymbol{\psi}_R$ fall perfectly on the grid points. In this latter case, the l th non-zero row of $\bar{\mathbf{X}}$ is equal to the l th row of \mathbf{X} . Note that (4.30) corresponds to a sparse recovery problem. Therefore, known CS techniques, e.g., [SPP18; AAH19a; MRM16; Cao+18], including the OMP method [SC12] can readily be applied to estimate $\bar{\mathbf{X}}$, as well as, $\boldsymbol{\psi}_T$ and $\boldsymbol{\psi}_R$, with automatic pairing.²

To proceed, let $\hat{\boldsymbol{\psi}}_T$ and $\hat{\boldsymbol{\psi}}_R$ denote the estimated frequency vectors of $\boldsymbol{\psi}_T$ and $\boldsymbol{\psi}_R$. Then, we construct $\hat{\mathbf{A}}_T$, $\hat{\mathbf{A}}_R$, and $\hat{\mathbf{A}} = (\mathbf{F}^\top \hat{\mathbf{A}}_T \otimes \mathbf{W}^\text{H} \hat{\mathbf{A}}_R)$. Therefore, to estimate \mathbf{H}_c in (4.25), an estimate of \mathbf{G} and \mathbf{B} is required. Let us assume that $\boldsymbol{\psi}_T$ and $\boldsymbol{\psi}_R$ are estimated perfectly and that the $\text{rank}\{\hat{\mathbf{A}}\} \geq L$. Then, multiplying (4.27) by $\hat{\mathbf{A}}^+$ from the left-hand-side we get

$$\tilde{\mathbf{Y}} = \hat{\mathbf{A}}^+ \mathbf{Y} = \mathbf{G} \mathbf{B} \boldsymbol{\Phi} + \tilde{\mathbf{N}} \in \mathbb{C}^{L \times T_S}, \quad (4.31)$$

where $\tilde{\mathbf{N}} = \hat{\mathbf{A}}^+ \mathbf{N} \in \mathbb{C}^{L \times T_S}$ is the filtered noise. Since $\mathbf{G} = \mathbf{G}^\top$, due to its diagonal structure, we can write $\tilde{\mathbf{Y}}^\top$ as

$$\tilde{\mathbf{Y}}^\top = \boldsymbol{\Phi}^\top \mathbf{B}^\top \mathbf{G} + \tilde{\mathbf{N}}^\top \in \mathbb{C}^{T_S \times L}. \quad (4.32)$$

Note that, $\mathbf{B}^\top \in \mathbb{C}^{M_S \times L}$ can be written as

$$\begin{aligned} \mathbf{B}^\top &= [(\mathbf{b}_{T,1}^\top \diamond \mathbf{b}_{R,1}^\top)^\top, \dots, (\mathbf{b}_{T,1}^\top \diamond \mathbf{b}_{R,L_R}^\top)^\top, \dots, (\mathbf{b}_{T,L_T}^\top \diamond \mathbf{b}_{R,L_R}^\top)^\top] \\ &= [(\mathbf{b}_{T,1} \odot \mathbf{b}_{R,1}), \dots, (\mathbf{b}_{T,1} \odot \mathbf{b}_{R,L_R}), \dots, (\mathbf{b}_{T,L_T} \odot \mathbf{b}_{R,L_R})], \end{aligned} \quad (4.33)$$

where $\mathbf{b}_{T,l} = \nu_{1D}(\mu_{T,l}^\vee) \diamond \nu_{1D}(\mu_{T,l}^\text{h})$ and $\mathbf{b}_{R,l'} = \nu_{1D}(\mu_{R,l'}^\vee) \diamond \nu_{1D}(\mu_{R,l'}^\text{h})$ are the l th and the l' th column vectors

²Note that the recoverability guarantee of $\bar{\mathbf{X}}$ in (4.30) can be improved by designing the sensing matrix $\bar{\mathbf{A}} = (\mathbf{F}^\top \bar{\mathbf{A}}_T \otimes \mathbf{W}^\text{H} \bar{\mathbf{A}}_R)$, as in [Ard+20b; APH19], which is out of the scope of this thesis.

of \mathbf{B}_T and \mathbf{B}_R , respectively, $l \in \{1, \dots, L_T\}$, $l' \in \{1, \dots, L_R\}$, i.e.,

$$\mathbf{b}_{T,l} = \begin{bmatrix} 1 \cdot \nu_{1D}(\mu_{T,l}^h) \\ e^{j\mu_{T,l}^v} \cdot \nu_{1D}(\mu_{T,l}^h) \\ \vdots \\ e^{j(M_S^v-1)\mu_{T,l}^v} \cdot \nu_{1D}(\mu_{T,l}^h) \end{bmatrix} \in \mathbb{C}^{M_S}, \quad (4.34)$$

$$\mathbf{b}_{R,l'} = \begin{bmatrix} 1 \cdot \nu_{1D}(\mu_{R,l'}^h) \\ e^{j\mu_{R,l'}^v} \cdot \nu_{1D}(\mu_{R,l'}^h) \\ \vdots \\ e^{j(M_S^v-1)\mu_{R,l'}^v} \cdot \nu_{1D}(\mu_{R,l'}^h) \end{bmatrix} \in \mathbb{C}^{M_S}. \quad (4.35)$$

Therefore, the c th column of \mathbf{B}^T , i.e., $\mathbf{b}_c = (\mathbf{b}_{T,l} \odot \mathbf{b}_{R,l'})$ has a Khatri-Rao structure given as

$$\mathbf{b}_c = \begin{bmatrix} 1 & \cdot & \nu_{1D}(\mu_{T,l}^h + \mu_{R,l'}^h) \\ e^{j(\mu_{T,l}^v + \mu_{R,l'}^v)} & \cdot & \nu_{1D}(\mu_{T,l}^h + \mu_{R,l'}^h) \\ \vdots & & \vdots \\ e^{j(M_S^v-1)(\mu_{T,l}^v + \mu_{R,l'}^v)} & \cdot & \nu_{1D}(\mu_{T,l}^h + \mu_{R,l'}^h) \end{bmatrix}, \quad (4.36)$$

where $c = (l-1) \cdot L_R + l' \in \{1, \dots, L\}$. Let $\mu_c^v = \mu_{T,l}^v + \mu_{R,l'}^v$ and $\mu_c^h = \mu_{T,l}^h + \mu_{R,l'}^h$. Then, we have

$$\mathbf{b}_c = \nu_{1D}(\mu_c^v) \diamond \nu_{1D}(\mu_c^h) \in \mathbb{C}^{M_S}, \quad (4.37)$$

where $\nu_{1D}(\mu_c^v) \in \mathbb{C}^{M_S^v}$ and $\nu_{1D}(\mu_c^h) \in \mathbb{C}^{M_S^h}$. Accordingly, $\mathbf{B}^T = (\mathbf{B}^v \diamond \mathbf{B}^h)$, in which

$$\mathbf{B}^v = [\nu_{1D}(\mu_1^v), \dots, \nu_{1D}(\mu_L^v)], \quad (4.38)$$

$$\mathbf{B}^h = [\nu_{1D}(\mu_1^h), \dots, \nu_{1D}(\mu_L^h)]. \quad (4.39)$$

Assuming a Kronecker structure of the RIS, the reflection coefficients matrix Φ can be expressed as

$$\Phi = \Phi_v \otimes \Phi_h \in \mathbb{C}^{M_S \times T_S}, \quad (4.40)$$

where $\Phi_v \in \mathbb{C}^{M_S^v \times T_S^v}$ and $\Phi_h \in \mathbb{C}^{M_S^h \times T_S^h}$, and $T_S = T_S^v \cdot T_S^h$. We rewrite (4.32) as

$$\tilde{\mathbf{Y}}^T \stackrel{(c)}{=} (\Phi_v^T \mathbf{B}^v \diamond \Phi_h^T \mathbf{B}^h) \mathbf{G} + \tilde{\mathbf{N}}^T \in \mathbb{C}^{T_S \times L}, \quad (4.41)$$

where $\stackrel{(c)}{=}$ is obtained by utilizing the structure of Φ in (4.40) and (1.11). Let $\boldsymbol{\mu}^v$ and $\boldsymbol{\mu}^h$ be the frequency vectors, which are given as

$$\boldsymbol{\mu}^v = [\mu_1^v, \dots, \mu_L^v] \quad (4.42)$$

and

$$\boldsymbol{\mu}^h = [\mu_1^h, \dots, \mu_L^h]. \quad (4.43)$$

Similarly to (4.30), the first term on the right-hand-side of (4.41) is completely characterized by $\boldsymbol{\mu}^v$

Algorithm 4.4 Two-Stage RIS-Aided MIMO Channel Estimation (TRICE)

Input: Measurement matrix \mathbf{Y} in (4.30)

- 1: **Stage 1:** Get $\hat{\psi}_T, \hat{\psi}_R$ using, e.g., OMP or method in [ZH17a]
- 2: **Stage 2:** Assuming knowledge of $\hat{\psi}_T$ and $\hat{\psi}_R$ **do**
- 3: Get $\tilde{\mathbf{Y}}^T = [\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_L] \in \mathbb{C}^{M_s \times L}$ from (4.31)
- 4: **for** $c = 1$ to L **do**
- 5: Get $\hat{\mu}_c^h$ and $\hat{\mu}_c^v$ using, e.g., OMP or method in [ZH17a]
- 6: Get c th diagonal entry of $\hat{\mathbf{G}}$, i.e., \hat{g}_c using (4.45)
- 7: **end for**
- 8: Construct $\hat{\mathbf{H}} = (\hat{\mathbf{A}}_T \otimes \hat{\mathbf{A}}_R) \hat{\mathbf{G}} \hat{\mathbf{B}}$ (according to (4.25))
- 9: Estimate $\hat{\mathbf{H}}_T$, and $\hat{\mathbf{H}}_R$ from $\hat{\mathbf{H}}$ using [AA20]

and μ^h . Therefore, μ^v and μ^h can be estimated using the same methods discussed above. However, it should be noted that the joint estimation of μ^v and μ^h does not guarantee the automatic pairing with the pre-estimated frequency vectors ψ_T and ψ_R . To overcome this issue, we utilize the diagonal structure of the \mathbf{G} matrix in (4.41) and propose to estimate μ^v and μ^h sequentially, where the c th entries μ_c^v and μ_c^h can be jointly estimated from the c th column vector of $\tilde{\mathbf{Y}}^T$ in (4.41), i.e., $\tilde{\mathbf{y}}_c$ that is given as

$$\tilde{\mathbf{y}}_c = (\Phi_v^T \nu_{1D}(\mu_c^v) \diamond \Phi_h^T \nu_{1D}(\mu_c^h)) g_c + \tilde{\mathbf{n}}_c \in \mathbb{C}^{T_s}, \quad (4.44)$$

where g_c is the c th diagonal entry of \mathbf{G} and $\tilde{\mathbf{n}}_c$ is the c th column vector of $\tilde{\mathbf{N}}^T$. Note that, due to the Kronecker structure of Φ in (4.40), it is possible to apply the DFT-beamspace ESPRIT method of [ZH17b] on (4.44) to obtain a closed-form estimates of μ_c^v and μ_c^h . Next, for given $\hat{\mu}_c^v$ and $\hat{\mu}_c^h$, the c th path gain g_c can be estimated from (4.44) using LS as

$$\hat{g}_c = (\Phi_v^T \nu_{1D}(\hat{\mu}_c^v) \diamond \Phi_h^T \nu_{1D}(\hat{\mu}_c^h))^+ \tilde{\mathbf{y}}_c. \quad (4.45)$$

Finally, the \mathbf{B} matrix in (4.25) can be reconstructed as $\hat{\mathbf{B}} = (\hat{\mathbf{B}}^v \diamond \hat{\mathbf{B}}^h)^T \in \mathbb{C}^{L \times M_s}$. In summary, the proposed TRICE framework is given by Algorithm 4.4, where in Step 9, an estimate of \mathbf{H}_T and \mathbf{H}_R , up to trivial scaling factors, can be obtained from \mathbf{H}_c using the LS Khatri-Rao Factorization (LSKRF) algorithm proposed in [RH10]. Please note that Algorithm 4.4 is very general in the sense that any other efficient 2D parameter estimation method can be readily used in Step 1 and 5, e.g., the methods proposed in [SPP18; AAH19a; MRM16; Cao+18].

4.3.3. Proposed TenRICE Algorithm

In Subsection 4.3.2, by exploiting the low-rank nature of the mmWave channels, we have proposed the TRICE framework, which formulates the channel estimation in RIS-aided mmWave MIMO systems as a two-stage multi-dimensional sparse-recovery problem. Here, we extend our TRICE framework and propose a CP Tensor decomposition method for RIS-aided channel estimation in mmWave MIMO systems, termed TenRICE, by jointly exploiting the tensor structure of received signals and the low-rank nature of mmWave channels. To this end, we rewrite the measurement matrix \mathbf{Y} given in (4.26) by

considering the Khatri-Rao structure of B as

$$Y = (F^T A_T \otimes W^H A_R) G (B^v \diamond B^h)^T \Phi + N. \quad (4.46)$$

The vectorized form of Y , i.e., $\mathbf{y} = \text{vec}\{Y\} \in \mathbb{C}^{N_R T_T T_S}$, can be written as

$$\begin{aligned} \mathbf{y} &\stackrel{(d)}{=} \text{vec}\{(F^T A_T \Omega_T \diamond W^H A_R \Omega_R) G (B^v \diamond B^h)^T \Phi\} + \mathbf{n} \\ &\stackrel{(e)}{=} (\Phi_v^T B^v \diamond \Phi_h^T B^h \diamond F^T A_T \Omega_T \diamond W^H A_R \Omega_R) \mathbf{g} + \mathbf{n}, \end{aligned} \quad (4.47)$$

where $\mathbf{n} = \text{vec}\{N\}$, $\mathbf{g} = \text{undia}\{G\}$, and Φ_v and Φ_h are given in (4.40). Moreover, $\stackrel{(d)}{=}$ is obtained by using (1.26) to enable writing the received signal utilizing the CPD model, where $\Omega_T \stackrel{\text{def}}{=} I_{L_T} \otimes \mathbf{1}_{L_R}^T$ and $\Omega_R \stackrel{\text{def}}{=} \mathbf{1}_{L_T}^T \otimes I_{L_R}$, and $\stackrel{(e)}{=}$ is obtained by applying (1.18). From (4.47), we observe that \mathbf{y} is the vectorized form of the transposed 4-mode unfolding of a 4-way tensor $\mathcal{Y} \in \mathbb{C}^{N_R \times T_T \times T_S^h \times T_S^v}$, i.e., $\mathbf{y} = \text{vec}\{[\mathcal{Y}]_{(4)}^T\}^3$ that admits a constrained CP decomposition as [FA14; AFM08]

$$\mathcal{Y} = \mathcal{I}_{4,L} \times_1 \underline{A}_R \Omega_R \times_2 \underline{A}_T \Omega_T \times_3 \underline{B}^h \times_4 \underline{B}^v + \mathcal{N}, \quad (4.50)$$

where \mathcal{N} is the noise tensor, $\mathcal{I}_{4,L} \in \mathbb{Z}^{L \times L \times L \times L}$ is a super-diagonal tensor with ones on the super-diagonal, and

$$\underline{A}_R = W^H A_R = W^H [\nu_{1D}(\psi_{R,1}), \dots, \nu_{1D}(\psi_{R,L_R})] \in \mathbb{C}^{N_R \times L}, \quad (4.51)$$

$$\underline{A}_T = F^T A_T = F^T [\nu_{1D}(\psi_{T,1}), \dots, \nu_{1D}(\psi_{T,L_T})] \in \mathbb{C}^{T_T \times L}, \quad (4.52)$$

$$\underline{B}^h = \Phi_h^T B^h = \Phi_h^T [\nu_{1D}(\mu_1^h), \dots, \nu_{1D}(\mu_L^h)] \in \mathbb{C}^{T_S^h \times L}, \quad (4.53)$$

$$\underline{B}^v = \Phi_v^T B^v G = \Phi_v^T [\nu_{1D}(\mu_1^v), \dots, \nu_{1D}(\mu_L^v)] G \in \mathbb{C}^{T_S^v \times L}. \quad (4.54)$$

The n -mode unfolding of the tensor \mathcal{Y} , for $n \in \{1, 2, 3, 4\}$ can be expressed as

$$[\mathcal{Y}]_{(1)} = \underline{A}_R \Omega_R (\underline{B}^v \diamond \underline{B}^h \diamond \underline{A}_T \Omega_T)^T + [\mathcal{N}]_{(1)} \in \mathbb{C}^{N_R \times T_T T_S^h T_S^v}, \quad (4.55)$$

$$[\mathcal{Y}]_{(2)} = \underline{A}_T \Omega_T (\underline{B}^v \diamond \underline{B}^h \diamond \underline{A}_R \Omega_R)^T + [\mathcal{N}]_{(2)} \in \mathbb{C}^{T_T \times N_R T_S^h T_S^v}, \quad (4.56)$$

$$[\mathcal{Y}]_{(3)} = \underline{B}^h (\underline{B}^v \diamond \underline{A}_T \Omega_T \diamond \underline{A}_R \Omega_R)^T + [\mathcal{N}]_{(3)} \in \mathbb{C}^{T_S^h \times N_R T_T T_S^v}, \quad (4.57)$$

$$[\mathcal{Y}]_{(4)} = \underline{B}^v (\underline{B}^h \diamond \underline{A}_T \Omega_T \diamond \underline{A}_R \Omega_R)^T + [\mathcal{N}]_{(4)} \in \mathbb{C}^{T_S^v \times N_R T_T T_S^h}. \quad (4.58)$$

Given \mathcal{Y} , the channel estimation boils down to first estimating the tensor factor matrices. Several techniques have been proposed to achieve this end, e.g., in [RSH12; De 06; AAH19b]. One of these techniques is ALS [CLA09], which alternatively minimizes the data fitting error with respect to one of the factor matrices, with the other three being fixed. For example, to estimate \underline{A}_R , assuming that \underline{A}_T ,

³We substitute $\underline{B}^v = \Phi_v^T B^v G$ in (4.58) and its transpose can be written as

$$[\mathcal{Y}]_{(4)}^T = (\underline{B}^h \diamond \underline{A}_T \Omega_T \diamond \underline{A}_R \Omega_R) G (\Phi_v^T B^v)^T + [\mathcal{N}]_{(4)}^T, \quad (4.48)$$

where $G^T = G$ since G is a diagonal matrix. Using (1.20), we can vectorize (4.48) as

$$\text{vec}\{[\mathcal{Y}]_{(4)}^T\} = (\Phi_v^T B^v \diamond \underline{B}^h \diamond \underline{A}_T \Omega_T \diamond \underline{A}_R \Omega_R) \mathbf{g} + \text{vec}\{[\mathcal{N}]_{(4)}^T\}. \quad (4.49)$$

$\underline{\mathbf{B}}^h$, and $\underline{\mathbf{B}}^v$ are fixed, the problem can be formulated as

$$\underline{\mathbf{A}}_R = \underset{\underline{\mathbf{A}}_R}{\operatorname{argmin}} \|\mathcal{Y}_{(1)} - \underline{\mathbf{A}}_R \Omega_R (\underline{\mathbf{B}}^v \diamond \underline{\mathbf{B}}^h \diamond \underline{\mathbf{A}}_T \Omega_T)^\top\|_F^2, \quad (4.59)$$

which is a convex problem and can be solved using the LS method. Similarly, we can formulate the problem to estimate the other three matrices as

$$\underline{\mathbf{A}}_T = \underset{\underline{\mathbf{A}}_T}{\operatorname{argmin}} \|\mathcal{Y}_{(2)} - \underline{\mathbf{A}}_T \Omega_T (\underline{\mathbf{B}}^v \diamond \underline{\mathbf{B}}^h \diamond \underline{\mathbf{A}}_R \Omega_R)^\top\|_F^2, \quad (4.60)$$

$$\underline{\mathbf{B}}^h = \underset{\underline{\mathbf{B}}^h}{\operatorname{argmin}} \|\mathcal{Y}_{(3)} - \underline{\mathbf{B}}^h (\underline{\mathbf{B}}^v \diamond \underline{\mathbf{A}}_T \Omega_T \diamond \underline{\mathbf{A}}_R \Omega_R)^\top\|_F^2, \quad (4.61)$$

$$\underline{\mathbf{B}}^v = \underset{\underline{\mathbf{B}}^v}{\operatorname{argmin}} \|\mathcal{Y}_{(4)} - \underline{\mathbf{B}}^v (\underline{\mathbf{B}}^h \diamond \underline{\mathbf{A}}_T \Omega_T \diamond \underline{\mathbf{A}}_R \Omega_R)^\top\|_F^2, \quad (4.62)$$

the solutions of (4.59), (4.60), (4.61), and (4.62) are given, respectively, as

$$\hat{\underline{\mathbf{A}}}_R = [\mathcal{Y}]_{(1)} \left[\Omega_R (\underline{\mathbf{B}}^v \diamond \underline{\mathbf{B}}^h \diamond \underline{\mathbf{A}}_T \Omega_T)^\top \right]^+, \quad (4.63)$$

$$\hat{\underline{\mathbf{A}}}_T = [\mathcal{Y}]_{(2)} \left[\Omega_T (\underline{\mathbf{B}}^v \diamond \underline{\mathbf{B}}^h \diamond \underline{\mathbf{A}}_R \Omega_R)^\top \right]^+, \quad (4.64)$$

$$\hat{\underline{\mathbf{B}}}^h = [\mathcal{Y}]_{(3)} \left[(\underline{\mathbf{B}}^v \diamond \underline{\mathbf{A}}_T \Omega_T \diamond \underline{\mathbf{A}}_R \Omega_R)^\top \right]^+, \quad (4.65)$$

$$\hat{\underline{\mathbf{B}}}^v = [\mathcal{Y}]_{(4)} \left[(\underline{\mathbf{B}}^h \diamond \underline{\mathbf{A}}_T \Omega_T \diamond \underline{\mathbf{A}}_R \Omega_R)^\top \right]^+. \quad (4.66)$$

The estimation of these factor matrices using the ALS method is summarized in Algorithm 4.5 (from Step 5 to Step 8), which is guaranteed to converge monotonically to a local optimum point [CLA09].

Let $\hat{\underline{\mathbf{A}}}_R$, $\hat{\underline{\mathbf{A}}}_T$, $\hat{\underline{\mathbf{B}}}^h$, and $\hat{\underline{\mathbf{B}}}^v$ denote the estimated factor matrices at the convergence of the iterative steps of Algorithm 4.5. Then, the parameters associated with each factor matrix can be recovered, e.g., via a simple correlation-based scheme, which can be optimized to solve the following problems

$$\hat{\psi}_{R,c} = \underset{\psi_{R,c} \in [0, 2\pi]}{\operatorname{argmax}} \frac{|\hat{\underline{\mathbf{a}}}_{R,c}^H \mathbf{W}^H \boldsymbol{\nu}_{1D}(\psi_{R,c})|}{\|\hat{\underline{\mathbf{a}}}_{R,c}\| \|\mathbf{W}^H \boldsymbol{\nu}_{1D}(\psi_{R,c})\|}, \quad (4.67)$$

$$\hat{\psi}_{T,c} = \underset{\psi_{T,c} \in [0, 2\pi]}{\operatorname{argmax}} \frac{|\hat{\underline{\mathbf{a}}}_{T,c}^H \mathbf{F}^T \boldsymbol{\nu}_{1D}(\psi_{T,c})|}{\|\hat{\underline{\mathbf{a}}}_{T,c}\| \|\mathbf{F}^T \boldsymbol{\nu}_{1D}(\psi_{T,c})\|}, \quad (4.68)$$

$$\hat{\mu}_c^h = \underset{\mu_c^h \in [0, 2\pi]}{\operatorname{argmax}} \frac{|(\hat{\underline{\mathbf{b}}}_c^h)^H \Phi_h^T \boldsymbol{\nu}_{1D}(\mu_c^h)|}{\|\hat{\underline{\mathbf{b}}}_c^h\| \|\Phi_h^T \boldsymbol{\nu}_{1D}(\mu_c^h)\|}, \quad (4.69)$$

$$\hat{\mu}_c^v = \underset{\mu_c^v \in [0, \pi]}{\operatorname{argmax}} \frac{|(\hat{\underline{\mathbf{b}}}_c^v)^H \Phi_v^T \boldsymbol{\nu}_{1D}(\mu_c^v)|}{\|\hat{\underline{\mathbf{b}}}_c^v\| \|\Phi_v^T \boldsymbol{\nu}_{1D}(\mu_c^v)\|}, \quad (4.70)$$

where the c th entry of $\boldsymbol{\psi}_X$ and $\boldsymbol{\mu}^y$, i.e., $\psi_{X,c}$ and μ_c^y , are associated with the c th column vector of $\hat{\underline{\mathbf{A}}}_X$ and $\hat{\underline{\mathbf{B}}}^y$, i.e., $\hat{\underline{\mathbf{a}}}_{X,c}$ and $\hat{\underline{\mathbf{b}}}_c^y$, respectively, with $X \in \{R, T\}$ and $y \in \{h, v\}$. The optimal solutions to (4.67), (4.68), (4.69), and (4.70) can be efficiently implemented using exhaustive search by first employing a

coarse grid and then gradually refining it around the possible grid points. Alternatively, (4.67), (4.68), (4.69), and (4.70) can be interpreted as off-grid sparse recovery problems, where efficient methods like, Newtonized OMP (NOMP) [MRM16] and Gradient OMP [AH21] can be readily applied to recover $\hat{\psi}_{R,c}$, $\hat{\psi}_{T,c}$, $\hat{\mu}_{c'}^h$, and $\hat{\mu}_{c'}^v$ with high accuracy and low complexity.

Algorithm 4.5 Tensor-based RIS-aided CE (TenRICE)

- 1: Input: Measurement tensor $\mathcal{Y} \in \mathbb{C}^{N_R \times T_T \times T_S^h \times T_S^v}$, training matrices \mathbf{W} , \mathbf{F} , Φ_h , and Φ_v , and I_{\max}
 - 2: Output: Estimated channels $\hat{\mathbf{H}}_T$ and $\hat{\mathbf{H}}_R$
 - 3: Initialization: $(\hat{\mathbf{B}}^v)^{(0)}$, $(\hat{\mathbf{B}}^h)^{(0)}$, and $\hat{\mathbf{A}}_T^{(0)}$, e.g., randomly
 - 4: **while** not converged or $i < I_{\max}$ **do**
 - 5: $\hat{\mathbf{A}}_R^{(i)} = [\mathcal{Y}]_{(1)}[\Omega_R((\hat{\mathbf{B}}^v)^{(i-1)} \diamond (\hat{\mathbf{B}}^h)^{(i-1)} \diamond \hat{\mathbf{A}}_T^{(i-1)} \Omega_T)^T]^+$
 - 6: $\hat{\mathbf{A}}_T^{(i)} = [\mathcal{Y}]_{(2)}[\Omega_T((\hat{\mathbf{B}}^v)^{(i-1)} \diamond (\hat{\mathbf{B}}^h)^{(i-1)} \diamond \hat{\mathbf{A}}_R^{(i)} \Omega_R)^T]^+$
 - 7: $(\hat{\mathbf{B}}^h)^{(i)} = [\mathcal{Y}]_{(3)}[(\hat{\mathbf{B}}^v)^{(i-1)} \diamond \hat{\mathbf{A}}_T^{(i)} \Omega_T \diamond \hat{\mathbf{A}}_R^{(i)} \Omega_R]^+$
 - 8: $(\hat{\mathbf{B}}^v)^{(i)} = [\mathcal{Y}]_{(4)}[(\hat{\mathbf{B}}^h)^{(i)} \diamond \hat{\mathbf{A}}_T^{(i)} \Omega_T \diamond \hat{\mathbf{A}}_R^{(i)} \Omega_R]^+$
 - 9: **end while**
 - 10: Recover $\hat{\psi}_R$, $\hat{\psi}_T$, $\hat{\mu}^h$, $\hat{\mu}^v$ using (4.67), (4.68), (4.69), and (4.70) or NOMP [MRM16]
 - 11: Compute $\hat{\mathbf{g}} = [\Phi_v^T \hat{\mathbf{B}}^v \diamond \Phi_h^T \hat{\mathbf{B}}^h \diamond \mathbf{F}^T \hat{\mathbf{A}}_T \Omega_T \diamond \mathbf{W}^H \hat{\mathbf{A}}_R \Omega_R]^+ \mathbf{y}$
 - 12: Reconstruct $\hat{\mathbf{H}}_c = (\hat{\mathbf{A}}_R \otimes \hat{\mathbf{A}}_T) \text{diag}\{\hat{\mathbf{g}}\} (\hat{\mathbf{B}}^v \diamond \hat{\mathbf{B}}^h)^T$
 - 13: Estimate $\hat{\mathbf{H}}_T$ and $\hat{\mathbf{H}}_R$ from $\hat{\mathbf{H}}$ using [AA20]
-

Next, using the estimated vectors $\hat{\psi}_R$, $\hat{\psi}_T$, $\hat{\mu}^h$, and $\hat{\mu}^v$, given in (4.29), (4.28), (4.43), and (4.43), respectively, in Step 10, we reconstruct $\hat{\mathbf{A}}_R$, $\hat{\mathbf{A}}_T$, $\hat{\mathbf{B}}^h$, and $\hat{\mathbf{B}}^v$. Then, the path gain vector \mathbf{g} can be estimated from (4.47) (or $[\mathcal{Y}]_{(4)}^T$) using a LS method as shown by Step 11. Finally, the cascaded channel matrix $\hat{\mathbf{H}}_c$ in (4.6) can be reconstructed as in Step 12, which can be used to estimate $\hat{\mathbf{H}}_T$ and $\hat{\mathbf{H}}_R$, up to trivial scaling factors, using the LSKRF method [AA20].

4.4. Algorithms for Data Transmission Phase

Here, we assume that the channel matrices \mathbf{H}_T and \mathbf{H}_R are known and we design the transmit and the receive beamforming matrices as well as the RIS reflection vector. Note that (4.10) is non-convex, since the objective function is non-concave over ω and the constant modulus constraints are non-convex functions. Moreover, \mathbf{P} , \mathbf{Q} , and ω are coupled together, which makes (4.10) a difficult problem to solve. In the following, we propose a non-iterative solution to (4.10), which has a comparable performance to that of [ZZ20a], but with a much lower complexity.

4.4.1. Beamforming Design

Given the estimated channels $\hat{\mathbf{H}}_T$ and $\hat{\mathbf{H}}_R$, we define the effective channel matrix as

$$\hat{\mathbf{H}}_e = \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T \in \mathbb{C}^{M_R \times M_T}. \quad (4.71)$$

Then, we can rewrite (4.10) as

$$\begin{aligned} & \max_{\mathbf{P}, \mathbf{Q}} \log_2 |\mathbf{I}_{N_s} + \mathbf{R}^{-1} \mathbf{Q}^H \hat{\mathbf{H}}_e \mathbf{P} \mathbf{P}^H \hat{\mathbf{H}}_e^H \mathbf{Q}| \\ & \text{s.t. } \|\mathbf{P}\|_F^2 \leq P_{\max}, \end{aligned} \quad (4.72)$$

where \mathbf{R} is the noise covariance matrix and given in (4.9). Initially, it is not hard to see that for any given ω , (4.72) reduces to a single-user multi-stream MIMO communication system. The SVD⁴ of the effective channel \mathbf{H}_e is given by

$$\hat{\mathbf{H}}_e = \mathbf{U}_{\hat{\mathbf{H}}_e} \boldsymbol{\Sigma}_{\hat{\mathbf{H}}_e} \mathbf{V}_{\hat{\mathbf{H}}_e}^H. \quad (4.73)$$

Then, the optimal fully digital⁵ solutions to \mathbf{P} and \mathbf{Q} , for fixed ω , are given as

$$\mathbf{P} = \mathbf{V}_s \text{diag}\{\sqrt{p_1}, \dots, \sqrt{p_{N_s}}\} \quad \text{and} \quad \mathbf{Q} = \mathbf{U}_s, \quad (4.74)$$

where $\mathbf{U}_s = [\mathbf{U}_{\hat{\mathbf{H}}_e}]_{[:,1:N_s]} \in \mathbb{C}^{M_R \times N_s}$, $\mathbf{V}_s = [\mathbf{V}_{\hat{\mathbf{H}}_e}]_{[:,1:N_s]} \in \mathbb{C}^{M_T \times N_s}$, and $\{p_i\}_{i=1}^{N_s}$ are the power allocations found using the water-filling method [PF05] such that $\sum_{i=1}^{N_s} p_i = P_{\max}$. Let

$$\begin{aligned} \boldsymbol{\Sigma}_s &= \mathbf{U}_s^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T \mathbf{V}_s \\ &= \begin{bmatrix} [\mathbf{U}_s]_{[:,1]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,1]} & \dots & [\mathbf{U}_s]_{[:,1]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,N_s]} \\ \vdots & \dots & \vdots \\ [\mathbf{U}_s]_{[:,N_s]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,1]} & \dots & [\mathbf{U}_s]_{[:,N_s]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,N_s]} \end{bmatrix} \\ &\stackrel{(f)}{=} \begin{bmatrix} [\mathbf{U}_s]_{[:,1]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,1]} & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & [\mathbf{U}_s]_{[:,N_s]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,N_s]} \end{bmatrix} \\ &\stackrel{(g)}{=} \begin{bmatrix} ([\mathbf{V}_s]_{[:,1]}^T \hat{\mathbf{H}}_T^T \diamond [\mathbf{U}_s]_{[:,1]}^H \hat{\mathbf{H}}_R) \omega & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & ([\mathbf{V}_s]_{[:,N_s]}^T \hat{\mathbf{H}}_T^T \diamond [\mathbf{U}_s]_{[:,N_s]}^H \hat{\mathbf{H}}_R) \omega \end{bmatrix} \\ &= \text{diag}\{\alpha_1, \dots, \alpha_{N_s}\} \in \mathbb{C}^{N_s \times N_s}, \end{aligned} \quad (4.75)$$

where $\alpha_i = [\mathbf{U}_s]_{[:,i]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,i]} = ([\mathbf{V}_s]_{[:,i]}^T \hat{\mathbf{H}}_T^T \diamond [\mathbf{U}_s]_{[:,i]}^H \hat{\mathbf{H}}_R) \omega$. Note that, the third equality, i.e., $\stackrel{(f)}{=}$ is obtained from (4.71) and (4.73) since $[\mathbf{U}_s]_{[:,i]}^H \hat{\mathbf{H}}_R \text{diag}\{\omega\} \hat{\mathbf{H}}_T [\mathbf{V}_s]_{[:,j]} = 0, \forall i \neq j$, while the fourth equality, i.e., $\stackrel{(g)}{=}$ is obtained using property (1.20). Note that $\boldsymbol{\Sigma}_s$ contains the dominant singular values of $\hat{\mathbf{H}}_e$, i.e., $\boldsymbol{\Sigma}_s = [\boldsymbol{\Sigma}_{\hat{\mathbf{H}}_e}]_{[1:N_s, 1:N_s]}$. Consequently, the SE expression in (4.72) simplifies to

$$\begin{aligned} \text{SE} &= \log_2 |\mathbf{I}_{N_s} + \frac{1}{\sigma_n^2} \mathbf{Q}^H \hat{\mathbf{H}}_e \mathbf{P} \mathbf{P}^H \hat{\mathbf{H}}_e^H \mathbf{Q}| \\ &= \log_2 |\mathbf{I}_{N_s} + \frac{1}{\sigma_n^2} \mathbf{U}_s^H \hat{\mathbf{H}}_e \mathbf{V}_s \text{diag}\{p_1, \dots, p_{N_s}\} \mathbf{V}_s^H \hat{\mathbf{H}}_e^H \mathbf{U}_s| \end{aligned}$$

⁴Throughout this thesis, we assume that the singular values of a given diagonal matrix are arranged in a decreasing order.

⁵Although the HAD beamforming architecture is generally assumed for mmWave-based wireless communication systems [GAH20; Ard+20a; Ard+18], we assume fully digital beamforming architectures here at the transmitter and the receiver, to simplify the exposition.

$$\begin{aligned}
 &= \log_2 |\mathbf{I}_{N_s} + \frac{1}{\sigma_n^2} \boldsymbol{\Sigma}_s \text{diag}\{p_1, \dots, p_{N_s}\} \boldsymbol{\Sigma}_s| \\
 &= \sum_{i=1}^{N_s} \log_2 \left(1 + \frac{1}{\sigma_n^2} \alpha_i^2 p_i \right),
 \end{aligned} \tag{4.76}$$

where α_i is the i th dominant singular value in $\boldsymbol{\Sigma}_{\hat{\mathbf{H}}_e}$.

4.4.2. Reflection Matrix Design

In the following, we turn our attention to the RIS reflection design and propose two non-iterative solutions for $\boldsymbol{\omega}$ based on a Frobenius-Norm Maximization (FroMax) design strategy.

4.4.2.1 FroMax-1

As a baseline method, we consider the Frobenius norm of the channel, which is motivated by the fact that the maximization Frobenius norm of the channel in a point-to-point MIMO system yields the SNR-optimal strategy dominant eigenmode transmission (DET), as shown in [PNG08]. Therefore, the RIS reflection vector is found as a solution to

$$\begin{aligned}
 \boldsymbol{\omega} &= \underset{\boldsymbol{\omega}}{\text{argmax}} \|\hat{\mathbf{H}}_R \text{diag}\{\boldsymbol{\omega}\} \hat{\mathbf{H}}_T\|_F^2 = \underset{\boldsymbol{\omega}}{\text{argmax}} \|\mathbf{K}\boldsymbol{\omega}\|_2^2 \\
 \text{s.t. } &|[\boldsymbol{\omega}]_{[i]}| = \frac{1}{\sqrt{M_S}}, \forall i,
 \end{aligned} \tag{4.77}$$

where $\mathbf{K} \stackrel{\text{def}}{=} \hat{\mathbf{H}}_T^T \diamond \hat{\mathbf{H}}_R \in \mathbb{C}^{M_T M_R \times M_S}$ obtained by applying (1.18). Note that (4.77) is non-convex due to the constant modulus constraints. Therefore, we first seek a solution to the following relaxed and convex version of (4.77) given as

$$\hat{\boldsymbol{\omega}} = \underset{\hat{\boldsymbol{\omega}}}{\text{argmax}} \|\mathbf{K}\hat{\boldsymbol{\omega}}\|_2^2, \quad \text{s.t. } \|\hat{\boldsymbol{\omega}}\|_2 = 1. \tag{4.78}$$

The SVD of \mathbf{K} is given as

$$\mathbf{K} = \mathbf{U}_K \boldsymbol{\Sigma}_K \mathbf{V}_K^H. \tag{4.79}$$

Then, the optimal solution to (4.78) is given as

$$\hat{\boldsymbol{\omega}} = [\mathbf{V}_K]_{[:,1]} \tag{4.80}$$

To satisfy the constant modulus constraints of (4.77), we use a simple projection function, where the i th entry of $\boldsymbol{\omega}$ is given as

$$[\boldsymbol{\omega}^{\text{FroMax-1}}]_{[i]} = \frac{1}{\sqrt{M_S}} \cdot \left(\frac{[\hat{\boldsymbol{\omega}}]_{[i]}}{\|[\hat{\boldsymbol{\omega}}]_{[i]}\|} \right), \forall i. \tag{4.81}$$

4.4.2.2 FroMax-2

Norm maximization, used as the objective function for FroMax-1 in (4.77), is designed to concentrate energy mainly on the dominant eigenmode of the effective channel matrix $\hat{\mathbf{H}}_e$, resulting in a rank-one matrix $\hat{\mathbf{H}}_e$ in most cases. This makes FroMax-1 undesirable in multi-stream scenarios, where the goal is to transmit as many data streams as possible to get a larger SE. To this end, we turn our attention back

to (4.76). From (4.76), we can clearly see that ω should be designed so that the singular values α_i are maximized. Thus, we propose to modify (4.77) and find ω as a solution to

$$\begin{aligned} \omega &= \underset{\omega}{\operatorname{argmax}} \|\operatorname{diag}\{\Sigma_s\}\|_2^2 = \underset{\omega}{\operatorname{argmax}} \|D\omega\|_2^2 \\ \text{s.t. } &|[\omega]_{[i]}| = \frac{1}{\sqrt{M_S}}, \forall i, \end{aligned} \quad (4.82)$$

where Σ_s is given in (4.75) and D is a matrix defined as

$$D \stackrel{\text{def}}{=} \begin{bmatrix} [\mathbf{V}_s]_{[:,1]}^T \hat{\mathbf{H}}_T^T \diamond [\mathbf{U}_s]_{[:,1]}^H \hat{\mathbf{H}}_R \\ \vdots \\ [\mathbf{V}_s]_{[:,N_s]}^T \hat{\mathbf{H}}_T^T \diamond [\mathbf{U}_s]_{[:,N_s]}^H \hat{\mathbf{H}}_R \end{bmatrix} \in \mathbb{C}^{N_s \times M_S}, \quad (4.83)$$

which is obtained from the forth equality, i.e., ^(g) of (4.75). Similarly to (4.78), (4.83) can be relaxed to a convex form as

$$\bar{\omega} = \underset{\bar{\omega}}{\operatorname{argmax}} \|D\bar{\omega}\|_2^2, \quad \text{s.t. } \|\bar{\omega}\|_2 = 1. \quad (4.84)$$

Let us define the SVD of D as

$$D = U_D \Sigma_D V_D^H. \quad (4.85)$$

Similarly to FroMax-1, the optimal solution to (4.84) can be obtained from the right singular vector that corresponds to the dominant singular value of D . However, to overcome the limitation of such a solution, which concentrates the energy mainly on the dominant eigenmode of D , we propose in the following a heuristic solution to (4.84) by taking the contributions of the N_s right singular vectors corresponding to the dominant N_s singular values of D as

$$\bar{\omega} = \frac{[\mathbf{V}_D]_{[:,1]} + \cdots + [\mathbf{V}_D]_{[:,N_s]}}{\|[\mathbf{V}_D]_{[:,1]} + \cdots + [\mathbf{V}_D]_{[:,N_s]}\|_2}. \quad (4.86)$$

Using $\bar{\omega}$, the RIS reflection vector ω is obtained as

$$[\omega^{\text{FroMax-2}}]_{[i]} = \frac{1}{\sqrt{M_S}} \cdot ([\bar{\omega}]_{[i]} / |[\bar{\omega}]_{[i]}|), \forall i. \quad (4.87)$$

Both proposed algorithms, FroMax-1 and FroMax-2, are summarized in Algorithm 4.6.

Remark 1: From (4.83), it is clear that the unitary matrices \mathbf{U}_s and \mathbf{V}_s are required to construct D . However, since \mathbf{U}_s and \mathbf{V}_s are coupled with ω , an iterative two-step algorithm will be required, where we update \mathbf{U}_s and \mathbf{V}_s in one step and ω in the other step. However, we found that if \mathbf{U}_s and \mathbf{V}_s are appropriately initialized, then one iteration of such an algorithm is sufficient to have a near SE performance to that obtained by the iterative method of [ZZ20a]. Here, we propose to initialize \mathbf{U}_s and \mathbf{V}_s as follows. Let $\hat{\mathbf{H}}_R = \mathbf{U}_{\hat{\mathbf{H}}_R} \Sigma_{\hat{\mathbf{H}}_R} \mathbf{V}_{\hat{\mathbf{H}}_R}^H$ and $\hat{\mathbf{H}}_T = \mathbf{U}_{\hat{\mathbf{H}}_T} \Sigma_{\hat{\mathbf{H}}_T} \mathbf{V}_{\hat{\mathbf{H}}_T}^H$ be the SVD of $\hat{\mathbf{H}}_R$ and $\hat{\mathbf{H}}_T$, respectively. Then, we assume that \mathbf{U}_s and \mathbf{V}_s in (4.83) are given as $\mathbf{U}_s = [\mathbf{U}_{\hat{\mathbf{H}}_R}]_{[:,1:N_s]}$ and $\mathbf{V}_s = [\mathbf{V}_{\hat{\mathbf{H}}_T}]_{[:,1:N_s]}$.

Algorithm 4.6 FroMax-based methods for RIS reflection design.

-
- 1: Input: \mathbf{H}_T , \mathbf{H}_R , and P_{\max}
 - 2: **if** FroMax-1 based method **then**
 - 3: Construct \mathbf{K} as in (4.77) and get $\tilde{\omega}$ from \mathbf{V}_K
 - 4: Obtain $\omega^* \leftarrow \omega^{\text{FroMax-1}}$ using (4.81)
 - 5: **else if** FroMax-2 based method **then**
 - 6: Initialize \mathbf{U}_s and \mathbf{V}_s as $\mathbf{U}_s = [\mathbf{U}_{\mathbf{H}_R}]_{[:,1:N_s]}$ and $\mathbf{V}_s = [\mathbf{V}_{\mathbf{H}_T}]_{[:,1:N_s]}$
 - 7: Construct \mathbf{D} as in (4.83) and get $\tilde{\omega}$ from \mathbf{V}_D
 - 8: Obtain $\omega^* \leftarrow \omega^{\text{FroMax-2}}$ using (4.86)
 - 9: **end if**
 - 10: For given ω^* , obtain \mathbf{Q} and \mathbf{P} as in (4.74)
-

4.5. Discussion

4.5.1. Uniqueness and Identifiability Conditions

Here, we discuss the range of parameter settings required to ensure the uniqueness and identifiability of the algorithms. In Subsection 4.3.1, the LS-based algorithm is considered where $T_{\text{CE}} = T_T T_S \geq \frac{M_T M_R M_S}{N_R}$. The required training overhead for TRICE varies depending on the 2D parameter estimation method used. If TRICE is used in conjunction with ESPRIT, which we call TRICE-ESP, it requires $T_S \geq L \geq 4$, $N_R \geq L_R + 1$, $T_T \geq L_T + 1$, $(T_T - 1)N_R \geq L$, $(N_R - 1)T_T \geq L$, while the training overhead for TRICE is used in conjunction with CS, which we call TRICE-CS, by taking into account the low-rank nature of mmWave channels requires $N_R T_T \approx \mathcal{O}(L \log(\bar{L}_R \bar{L}_T / L)) \ll M_R M_T$, $T_S \approx \mathcal{O}(\log(\bar{L}_S^v \bar{L}_S^h))$ [Reb+16]. Moreover, we compare our proposed TRICE method in [Ard+21] with the algorithm in [Wan+20]. The proposed method in [Wan+20] estimates the channel parameter jointly using two different CS techniques, i.e., OMP and generalized approximate message passing (GAMP) [VS13]. Therefore, we call it Joint-CS, which requires $N_R T_T T_S \approx \mathcal{O}(L \log(\bar{L}_S^v \bar{L}_S^h \bar{L}_R \bar{L}_T / L))$.

Regarding our second proposed channel estimation method, i.e., the TenRICE method [Ghe+21], it is well known that the CP decomposition is unique up to scaling and permutation ambiguities under mild conditions [CLA09; Zho+17]. In general, the uniqueness of a CP decomposition is guaranteed by Kruskal's condition [Kru77]. However, due to the definitions of Ω_R and Ω_T , the first two factor matrices, i.e., $\mathbf{A}_R \Omega_R = \mathring{\mathbf{A}}_R$ and $\mathbf{A}_T \Omega_T = \mathring{\mathbf{A}}_T$ contain repeated columns, where every column of $\mathring{\mathbf{A}}_R$ is repeated L_T times and every column of $\mathring{\mathbf{A}}_T$ is repeated L_R times. This implies that the k -rank of $\mathring{\mathbf{A}}_R$ and $\mathring{\mathbf{A}}_T$ is equal to one. Therefore, the sufficient condition of [KB09] fails [SA10]. As for Algorithm 4.5, which is an ALS-based algorithm, the identifiability in the LS sense requires that each of the following matrices

$$\mathbf{C}_R = \Omega_R (\mathbf{B}_v \diamond \mathbf{B}_h \diamond \mathbf{A}_T \Omega_T)^T \in \mathbb{C}^{L_R \times J_R} \quad (4.88)$$

$$\mathbf{C}_T = \Omega_T (\mathbf{B}_v \diamond \mathbf{B}_h \diamond \mathbf{A}_R \Omega_R)^T \in \mathbb{C}^{L_T \times J_T} \quad (4.89)$$

$$\mathbf{C}_h = (\mathbf{B}_v \diamond \mathbf{A}_T \Omega_T \diamond \mathbf{A}_R \Omega_R)^T \in \mathbb{C}^{L \times J_S^h} \quad (4.90)$$

$$\mathbf{C}_v = (\mathbf{B}_h \diamond \mathbf{A}_T \Omega_T \diamond \mathbf{A}_R \Omega_R)^T \in \mathbb{C}^{L \times J_S^v} \quad (4.91)$$

must have a unique right Moore-Penrose inverse, i.e., full row-rank, where $J_R = T_T T_S$, $J_T = N_R T_S$, $J_S^h = N_R T_T T_S^v$, $J_S^v = N_R T_T T_S^h$. This requires that $J_R \geq L_R$, $J_T \geq L_T$, $J_S^h \geq L$, and $J_S^v \geq L$, where

$L = L_R \cdot L_T$. Since L_R and L_T are practically very small, i.e., $\max\{L_R, L_T\} \approx 3$ [Rap+17], the above conditions are easily satisfied. For example, assuming that the transmitter has a LoS with the RIS, we have that $L_T = 1$, as it has been assumed in [Nad+20]. The order of the training overhead of the different algorithms is summarized in Table 4.1.

Table 4.1: Training overhead analysis

Method	Training overhead
TRICE-ESP	$T_S \geq L \geq 4, N_R \geq L_R + 1, T_T \geq L_T + 1, (T_T - 1)N_R \geq L, (N_R - 1)T_T \geq L$
TRICE-CS	$N_R T_T \approx \mathcal{O}(L \log(\bar{L}_R \bar{L}_T / L)), T_S \approx \mathcal{O}(\log(\bar{L}_S^v \bar{L}_S^h))$
Joint-CS	$N_R T_T T_S \approx \mathcal{O}(L \log(\bar{L}_R \bar{L}_T / L)), T_S \approx \mathcal{O}(\log(\bar{L}_S^v \bar{L}_S^h))$
TenRICE	$J_R \geq L_R, J_T \geq L_T, J_S^h \geq L, \text{ and } J_S^v \geq L$

4.5.2. Computational Complexity

Channel estimation phase: Here, we derive the computational complexity of the algorithms in terms of Floating-Point Operations Per Second (FLOPS). The proposed TRICE-ESP method has a computational complexity on the order of $\mathcal{O}((N_R T_T)^2 T_S + T_S^3 + 3L^3 + L)$. Moreover, the computational complexity of TRICE-CS is on the order of $\mathcal{O}(L(N_R T_T (\bar{L}_T \bar{L}_R + L + L^2)) + 2L^3 + L T_S \bar{L}_S^v \bar{L}_S^h)$, while the computational complexity of the benchmark Joint-CS method is on the order of $\mathcal{O}(L(N_R T_T T_S (\bar{L}_S^v \bar{L}_S^h \bar{L}_T \bar{L}_R + L + L^2)) + L^3)$. Note that the major difference between TRICE-CS and Joint-CS is that the former decouples the channel parameter estimation into two stages, while the latter jointly estimates them. Therefore, TRICE-CS requires a 2D dictionary in every stage, while Joint-CS requires a single 4D dictionary.

For the proposed TenRICE method, we assume that the complexity of calculating the Moore-Penrose inverse of a $n \times m$ matrix is on the order of $\mathcal{O}(\min\{n, m\}^3)$ [Med04; Pen55]. Then, the complexity of the ALS steps in Algorithm 4.5 is on the order of $\mathcal{O}(I_{\max}(L_R^3 + L_T^3 + 2L^3))$, where I_{\max} denotes the number of ALS iterations. Moreover, assuming that the NOMP method from [MRM16] is used in Step 10 of Algorithm 4.5, then the complexity of recovering the channel parameters is on the order of $\bar{L}(L_R + L_T + 2L)$, where \bar{L} denotes the number of grid points used by NOMP in the sparse-coding stage, i.e., $\bar{L} = \bar{L}_T \bar{L}_R$ and $\bar{L} = \bar{L}_S^v \bar{L}_S^h$. Clearly, TenRICE has a much lower complexity than TRICE and joint-CS. The main reason is that TRICE and Joint-CS require multi-dimensional (xD) dictionaries (2D for TRICE and 4D for Joint-CS), compared to the 1D dictionary required by TenRICE. Moreover, TRICE and Joint-CS require a dictionary orthogonalization operation during the parameter recovery [SC12], which is very complex, especially with large dictionaries. This complex operation is not required by TenRICE thanks to its ALS stage, which decouples the recovery process between the channel parameters. The computational complexity of the simulated algorithms is summarized in Table 4.2.

Data transmission phase: We know that the complexity of calculating the SVD of a $n \times m$ matrix can be reduced by using the *Power Iteration* method [HMT11]. However, to simplify the analysis we assume that the SVD is calculated using the bi-diagonalization and the QR algorithm with a complexity

Table 4.2: Computational complexity analysis

Method	Computational complexity
TRICE-ESP	$\mathcal{O}((N_{\text{R}}T_{\text{T}})^2T_{\text{S}} + T_{\text{S}}^3 + 3L^3 + L)$
TRICE-CS	$\mathcal{O}(L(N_{\text{R}}T_{\text{T}}(\bar{L}_{\text{T}}\bar{L}_{\text{R}} + L + L^2)) + 2L^3 + LT_{\text{S}}\bar{L}_{\text{S}}^{\text{v}}\bar{L}_{\text{S}}^{\text{h}})$
Joint-CS	$\mathcal{O}(L(N_{\text{R}}T_{\text{T}}T_{\text{S}}(\bar{L}_{\text{S}}^{\text{v}}\bar{L}_{\text{S}}^{\text{h}}\bar{L}_{\text{T}}\bar{L}_{\text{R}} + L + L^2)) + L^3)$
TenRICE	$\mathcal{O}(I_{\text{max}}(L_{\text{R}}^3 + L_{\text{T}}^3 + 2L^3)) + \mathcal{O}(\bar{L}(L_{\text{R}} + L_{\text{T}} + 2L))$

on the order of $\mathcal{O}(nm^2)$ [CD13]. Then, the complexity of Algorithm 4.6 Steps 3, 6, 7, and 10 is on the order of $\mathcal{O}(M_{\text{R}}M_{\text{T}}M_{\text{S}}^2)$, $\mathcal{O}(M_{\text{R}}M_{\text{S}}^2 + M_{\text{S}}M_{\text{T}}^2)$, $\mathcal{O}(N_{\text{s}}M_{\text{S}}^2)$, and $\mathcal{O}(M_{\text{R}}M_{\text{T}}^2)$, respectively. Accordingly, the complexity of FroMax-1 is on the order of $\mathcal{O}(M_{\text{R}}M_{\text{T}}M_{\text{S}}^2 + M_{\text{R}}M_{\text{T}}^2)$ and of FroMax-2 is on the order of $\mathcal{O}(M_{\text{R}}M_{\text{S}}^2 + M_{\text{S}}M_{\text{T}}^2 + N_{\text{s}}M_{\text{S}}^2 + M_{\text{R}}M_{\text{T}}^2)$. In comparison, the complexity of the alternating maximization (AltMax) method of [ZZ20a] is on the order of $\mathcal{O}(J_{\text{max}}(M_{\text{S}}(3M_{\text{R}}^3 + 2M_{\text{R}}^2M_{\text{T}} + M_{\text{T}}^2) + M_{\text{R}}M_{\text{T}}^2))$, where J_{max} is the maximum number of iterations.

4.6. Simulation Results

In this section, we evaluate the effectiveness of the proposed methods by means of simulation. We define the SNR as $\text{SNR} = \mathbb{E}\{\|\mathbf{y} - \mathcal{N}\|_{\text{F}}^2\} / \mathbb{E}\{\|\mathcal{N}\|_{\text{F}}^2\}$ and assume that $g_c \sim \mathcal{CN}(0, 1)$.

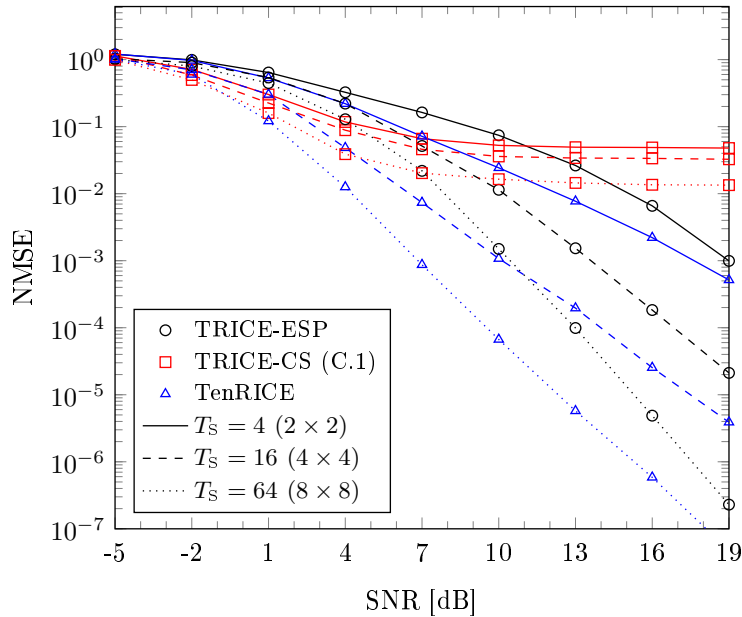


Figure 4.4: Compare proposed methods in terms of NMSE vs. SNR.

4.6.1. Channel Estimation Phase

Here, we show the simulation results to evaluate the effectiveness of the proposed methods, i.e., TRICE and TenRICE. In all simulation results, we assume that $M_T = 32$, $M_R = 16$, $M_S = 256$ [16×16], $N_R = 8$, and $T_T = 8$. The experiment is repeated with 1000 randomly generated noise realizations and Monte-Carlo averaging over the analytical expansions. We assume that the TRICE framework employs at both stages (i) the 2D DFT-beamspace ESPRIT method from [ZH17a] termed as TRICE-ESP, and (ii) the on-grid CS method, denoted as TRICE-CS. For TRICE-CS, the first stage 2D dictionary is formed by using $\beta_T M_T \times \beta_R M_R$ grid points ($\bar{L}_T = \beta_T M_T$, $\bar{L}_R = \beta_R M_R$), while the second stage 2D dictionary is formed by using $\beta_S^v M_S^v \times \beta_S^h M_S^h$ grid points ($\bar{L}_S^v = \beta_S^v M_S^v$, $\bar{L}_S^h = \beta_S^h M_S^h$), where $\{\beta_T, \beta_R, \beta_S^v, \beta_S^h\}$ are integer numbers that denote the ratio of the number of grid points to the number of reflecting elements. The estimation for the TRICE-CS method is performed using the classical OMP technique [SC12] assuming a 2D dictionary in both stages.

In Figs. 4.4 and 4.5, we show simulation results to evaluate the channel estimation accuracy of the above discussed methods, assuming that $L_T = 1$, $L_R = 2$. For TRICE-CS, the grid resolution is defined as (C.1) $\{\beta_T, \beta_R, \beta_S^v, \beta_S^h\} = \{2, 1, 4, 4\}$. To comply with the requirements of the DFT-beamspace ESPRIT method as discussed in [ZH17a], i.e., the spatial frequencies of the channel should fall within a sector-of-interest, we assume that $\psi_{R,\ell} \sim \mathcal{U}(0, 2\pi(N_R - 1)/M_R)$, $\psi_{T,\ell} \sim \mathcal{U}(0, 2\pi(T_T - 1)/M_T)$, $\mu_\ell^h \sim \mathcal{U}(0, 2\pi(T_S^h - 1)/M_S^h)$, $\mu_\ell^v \sim \mathcal{U}(0, 2\pi(T_S^v - 1)/M_S^v)$, where $\mathcal{U}(a, b)$ denotes the uniform distribution from a to b . Moreover, for all methods, the training matrices are chosen as $\mathbf{W}^H = [\mathbf{U}_{M_R}]_{[1:N_R,:]}$, $\mathbf{F}^T = [\mathbf{U}_{M_T}]_{[1:T_T,:]}$, $\Phi_h^T = [\mathbf{U}_{M_S^h}]_{[1:T_S^h,:]}$, $\Phi_v^T = [\mathbf{U}_{M_S^v}]_{[1:T_S^v,:]}$, where \mathbf{U}_M denotes the normalized $M \times M$ DFT-matrix.

In Fig. 4.4, we show simulation results evaluating the impact of the training overhead T_S on the channel estimation accuracy. Clearly, Fig. 4.4 shows that increasing T_S improves the channel estimation accuracy for all methods. For TRICE-CS, we can see that it has a good estimation performance in the low and middle SNR region while its performance does not improve in the high SNR region due to the fixed grid resolution. Differently, TRICE-ESP has a good estimation performance in the medium and high SNR region. On the other hand, TenRICE provides a significantly more accurate channel estimation compared to TRICE-ESP and TRICE-CS. To illustrate this, we show in Fig. 4.5 the corresponding MSE versus SNR results of the $T_S = 16$ case of Fig. 4.4. From Fig. 4.5, we can see that the TenRICE method provides much more accurate parameter estimates, compared to the TRICE-CS and TRICE-ESP, approaching the CRB as the SNR increases. The main reason is that TenRICE does not only exploit the low-rank nature of mmWave channels but also the tensor structure of the measurement signals when estimating the channel parameters. Moreover, TenRICE employs a high-resolution parameter recovery method in NOMP, while the TRICE-CS suffers from quantization errors, due to the on-grid assumption. These advantages enable TenRICE to provide a more accurate channel estimates than TRICE-ESP and TRICE-CS.

Despite the good performance of the TRICE-ESP method, finding the sector of interest is still a difficult task in practical scenarios. Therefore, in Figs. 4.6 and 4.7, we turn our focus only on the TRICE-CS and TenRICE methods. The training matrices, i.e., \mathbf{W} , \mathbf{F} , Φ_h , and Φ_v , are generated randomly such that the (a, b) th entry of \mathbf{W} is given as $[\mathbf{W}]_{[a,b]} = \frac{1}{\sqrt{M_R}} e^{j\theta_{a,b}}$, $\theta_{a,b} \sim \mathcal{U}(0, 2\pi)$, where \mathbf{F} , Φ_h , and Φ_v are generated in a similar manner. Moreover, we assume that the channel spatial frequencies are all distributed between 0 and 2π , i.e., $\psi_{R,\ell}, \psi_{T,\ell}, \mu_\ell^h, \mu_\ell^v \sim \mathcal{U}(0, 2\pi)$. In Fig. 4.6, we show the effect of the selection of $\{\beta_T, \beta_R, \beta_S^v, \beta_S^h\}$ on the TRICE-CS estimation performance, assuming that $L_T = 1$, $L_R =$

4. Single-RIS-Aided Systems

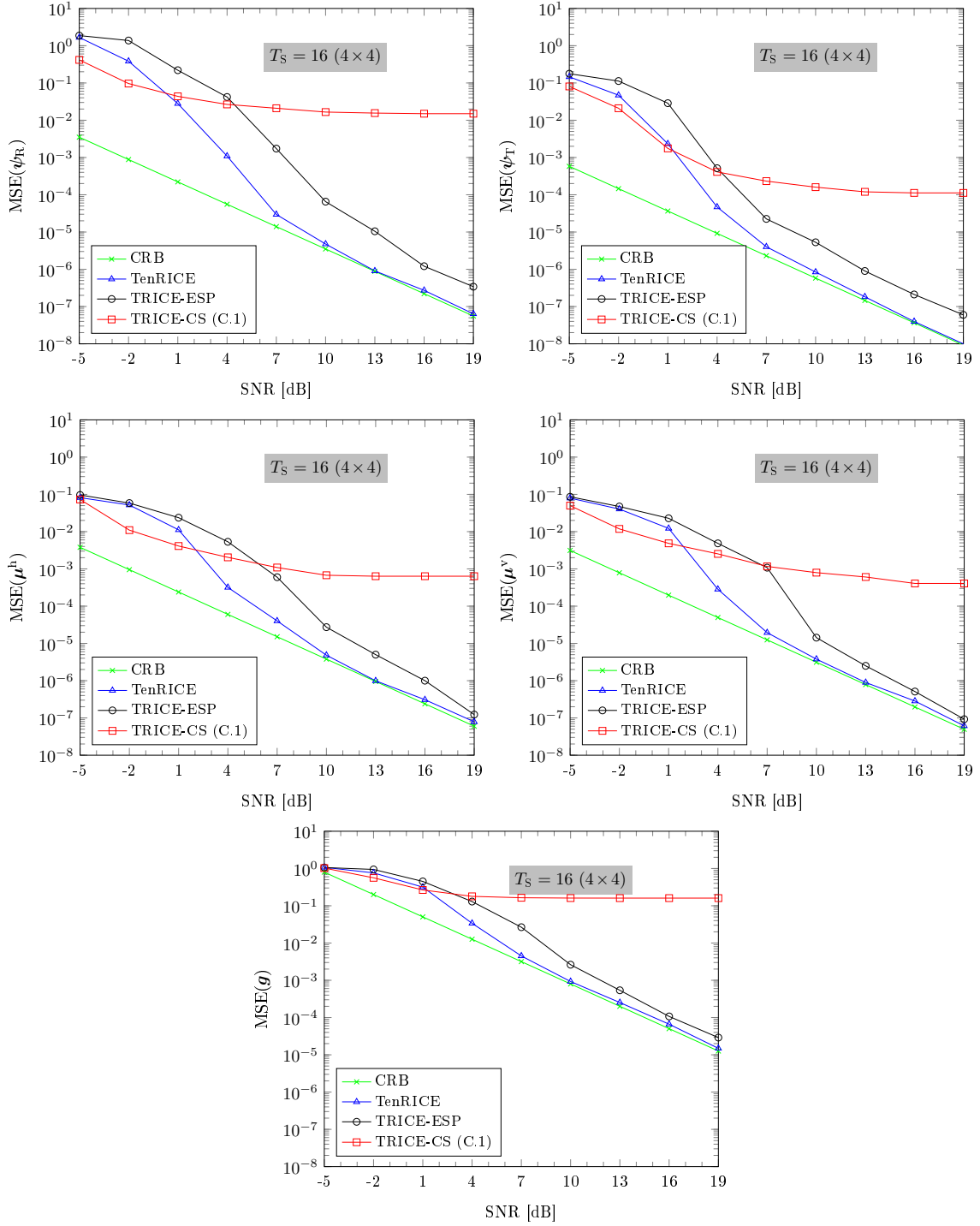


Figure 4.5: Compare proposed methods in terms of MSE vs. SNR.

2, and $T_S = 16$, i.e., $T_S^h = T_S^v = 4$. Clearly, Fig. 4.6 shows that increasing the grid resolution from (C.1) to (C.2), i.e., from $\{\beta_T, \beta_R, \beta_S^v, \beta_S^h\} = \{2, 1, 4, 4\}$ to $\{\beta_T, \beta_R, \beta_S^v, \beta_S^h\} = \{8, 4, 16, 16\}$, improves the channel estimation accuracy of the TRICE-CS method as a result of decreasing the impact of on-grid

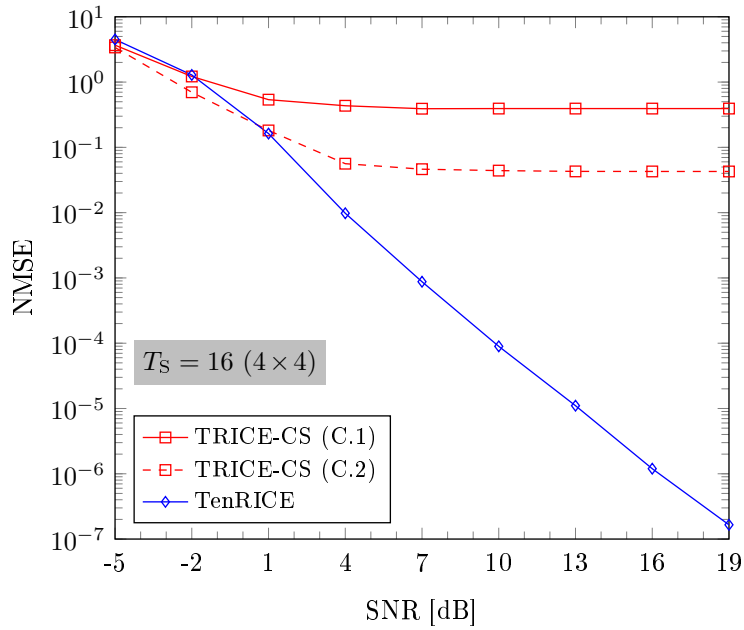


Figure 4.6: Effect of grid resolution on the channel estimation accuracy.

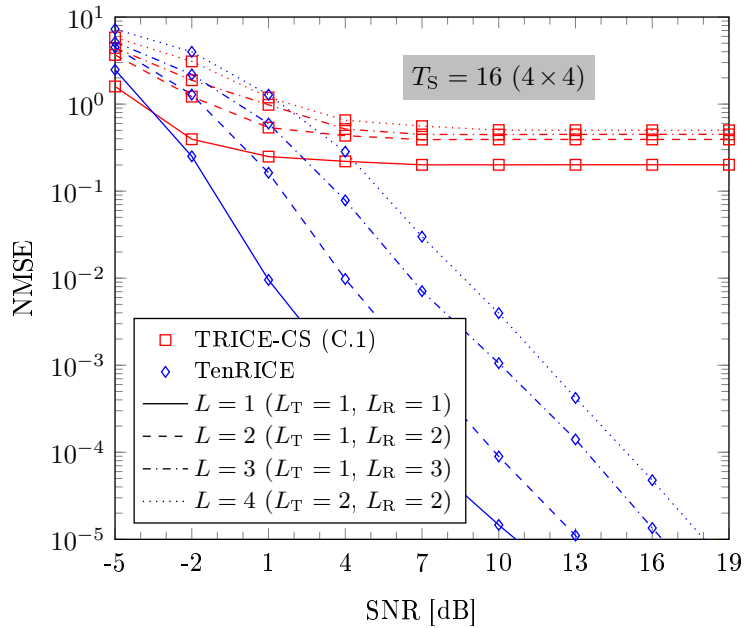


Figure 4.7: Effect of the number of multi path on the channel estimation accuracy.

quantization error. In Fig. 4.7, we show simulation results investigating the effect of the number of paths, i.e., the sparsity level of the channel, on the channel estimation accuracy. It is assumed that $T_S = 16$, i.e., $T_S^h = T_S^v = 4$. We can see that by increasing the number of paths, the channel estimation accuracy degrades. Obviously, the best performance can be obtained in the case of very sparse channels, i.e., $L = 2$. From Figs. 4.6 and 4.7, we can observe that TenRICE provides the best channel estimation

accuracy.

As can be seen from Algorithm 4.5, TenRICE estimates the factor matrices $\underline{\mathbf{A}}_R$, $\underline{\mathbf{A}}_T$, $\underline{\mathbf{B}}^h$, and $\underline{\mathbf{B}}^v$ using a bi-linear ALS method, which is designed to run up to a maximum of I_{\max} iterations, which we choose as $I_{\max} = 100$. In Fig. 4.8, we show simulation results investigating the convergence behavior of Algorithm 4.5. We define the *model fit* function c_{TenRICE} at the i th iteration as

$$c_{\text{TenRICE}}(i) = \frac{\|[\mathcal{Y}]_{(1)} - \underline{\mathbf{A}}_R^{(i)} \underline{\Omega}_R ((\underline{\mathbf{B}}^v)^{(i)} \diamond (\underline{\mathbf{B}}^h)^{(i)} \diamond \underline{\mathbf{A}}_T^{(i)} \underline{\Omega}_T)^T\|_F^2}{\|[\mathcal{Y}]_{(1)}\|_F^2}. \quad (4.92)$$

From Fig. 4.8, we can see that Algorithm 4.5 seems to converge *monotonically* to at least a local optimal solution. Note that, Algorithm 4.5 has a faster convergence rate with a low SNR value (e.g., SNR = 5 dB), as compared to a high SNR values (e.g., SNR = 15 dB). This is also true for the case with a smaller T_S value (i.e., $T_S = 16$) as compared to a larger T_S value (i.e., $T_S = 64$) or in case of smaller L_T and L_R values (i.e., $L_T = 1$ and $L_R = 2$) as compared to larger L_T and L_R values (i.e., $L_T = 2$ and $L_R = 2$). It is worth mentioning that in a different implementation of Algorithm 4.5, the *model fit* function $c_{\text{TenRICE}}(i)$ can be used as a convergence criterion, e.g., we stop the iterations if $|c_{\text{TenRICE}}(i) - c_{\text{TenRICE}}(i-1)| < \epsilon$ with $\epsilon < 10^{-15}$.

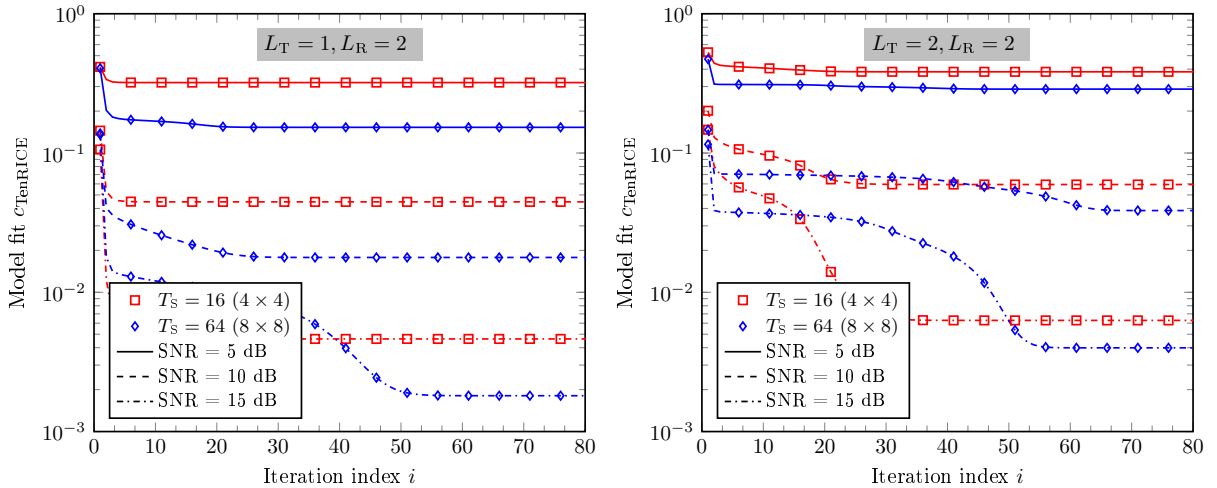


Figure 4.8: Convergence of the TenRICE method in Algorithm 4.5.

4.6.2. Data Transmission Phase

Next, we show simulation results to illustrate the efficiency of the proposed RIS reflection design method, FroMax. For comparison, we include results when the RIS reflection vector ω is designed according to the AltMax method in [ZZ20a] and a Random method, i.e., when the entries of ω are randomly generated such that the m th entry is given as $[\omega]_{[m]} = \frac{1}{\sqrt{M_S}} e^{j\omega_m}$, $\omega_m \sim \mathcal{U}(0, 2\pi)$. We define the SNR as $\text{SNR} = P_{\max}/\sigma^2$.

In Fig. 4.9, we show SE versus SNR results assuming $L_T = L_R = 2$, $M_T = 64$, $M_R = 16$, and $M_S = 256$ [16×16]. All simulation points are averaged over 1000 channel realization. We evaluate

the performance of the beamforming algorithms considering two scenarios, (i) perfect CSI (i.e., where the channel matrix is known perfectly) and (ii) estimated CSI (i.e., the channel matrix is estimated using our proposed TenRICE considering $N_R = T_T = T_S^h = T_S^v = 8$). From Fig. 4.9, we can see that the FroMax-1 has an equal performance to that of FroMax-2 and the AltMax when $N_s = 1$. However, FroMax-1 experiences a performance loss when $N_s = 2$, since it mainly maximizes the dominant singular value, as can be seen from Fig. 4.10 for one particular channel realization. Differently, AltMax and FroMax-2 optimize the dominant N_s singular values of the effective channel in a way that maximizes the system SE. Note that, in the low SNR region below 5 dB, all the simulated methods experience a very low SE performance in case of estimated CSI, due to the channel estimation errors. Therefore, a preprocessing *denoising* step will be required to improve the channel estimation accuracy.

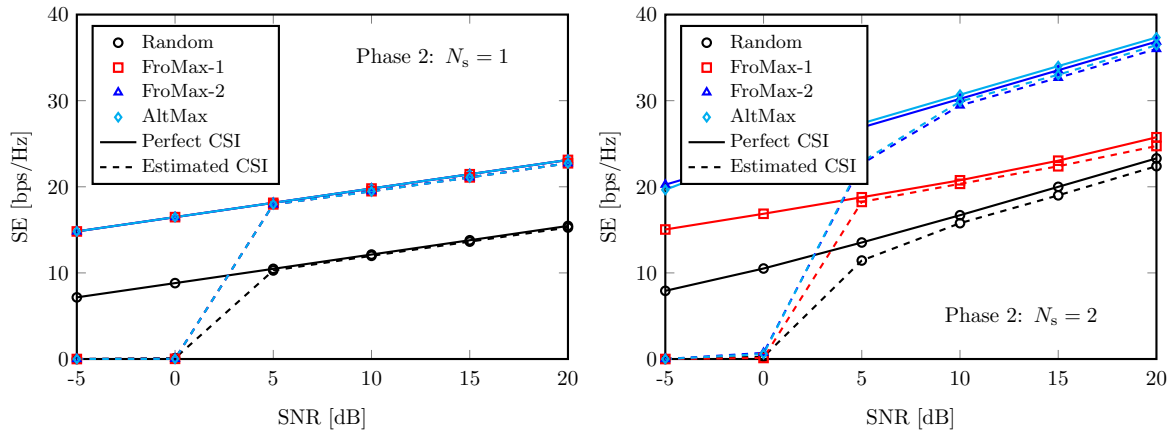


Figure 4.9: SE vs. SNR.

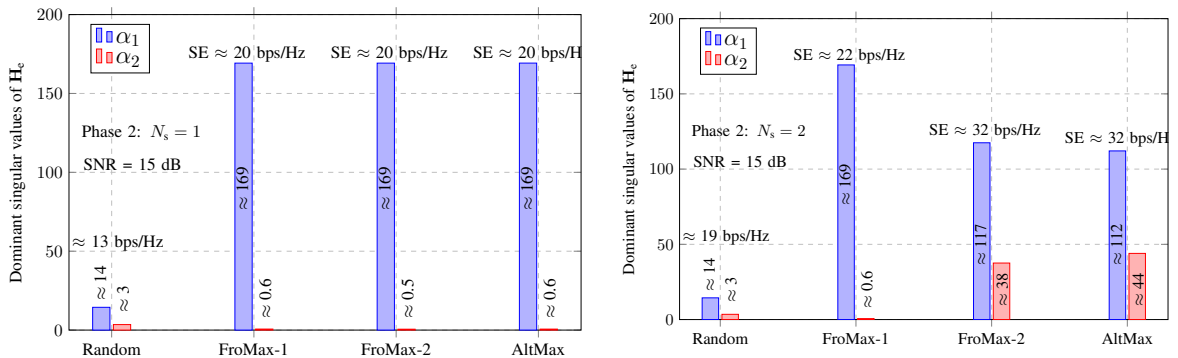


Figure 4.10: Dominant singular values of a perfect effective channel $H_e = H_R \text{diag}\{\omega\} H_T$.

4.7. Chapter Conclusion

In this chapter, we have considered the channel estimation and the RIS reflection design problems in point-to-point S-RIS-aided mmWave MIMO communication systems. First, we have proposed the TRICE framework, which is a two-stage channel parameter estimation scheme. By exploiting the low-rank nature of mmWave channels and by decoupling the channel parameter estimation problem into two stages, we have shown that TRICE not only has a high estimation performance, but only requires a low training overhead and has a low computational complexity, which makes it appealing in practical applications.

Then, we have proposed a CPD tensor-based channel estimation method termed TenRICE, which estimates the transmitter to RIS and the RIS to receiver channels separately, up to a trivial scaling factor. We have shown that by jointly exploiting the low-rank nature of mmWave channels and the tensor structure of received signals, not only the estimation accuracy can be improved, but also the training overhead and the complexity can be reduced. At the end of the chapter, we formulate the beamforming and the RIS reflection design as a SE maximization problem. The proposed non-iterative RIS reflection design method based on a Frobenius-norm maximization (FroMax) design strategy has a comparable performance to a benchmark method but with significantly lower complexity.

Double RIS-Aided Systems

5.1. Introduction and State of the Art

In Chapter 4, we have considered S-RIS-aided systems, where a transmitter communicates with one receiver via a single RIS-aided channel. However, in some application scenarios, such as in urban areas or satellite-to-indoor communications, the transmitter might need a multi-RIS-aided channel to have a successful communication with a receiver. Moreover, it was shown in [BS20] that the RIS in S-RIS-aided systems should be either deployed closer to the transmitter or closer to the receiver to achieve the best performance gain. This fundamental result gives rise to multi-RIS-aided systems, where a transmitter communicates with one or more receivers via multi-RIS-aided channels, which we consider in this chapter.

Among the multi-RIS-aided systems, Double RIS (D-RIS)-aided systems have received considerable attention [Han+20; Han+21; TS21; ZYZ21a]. In [Han+20], the authors have shown that the passive beamforming gain with D-RIS-aided systems is much higher than that with S-RIS-aided systems, i.e., $\mathcal{O}(M_S^4)$ versus $\mathcal{O}(M_S^2)$, where M_S denotes the total number of reflecting elements in both systems. Initial works on D-RIS-aided systems considered reflection design such as [Han+21; TS21; ZYZ21a]. However, the potentials of D-RIS-aided communication systems can only be realized if accurate channel estimation is available. Nonetheless, the channel estimation in D-RIS-aided systems is more problematic compared to S-RIS-aided systems, since the cascaded channel in D-RIS-aided systems contains three parts and not only two parts as in S-RIS-aided systems.

Recently, several channel estimation methods for D-RIS-aided systems have been proposed, such as in [YZZ21; Han+20; ZYZ21c; BX22]. In [Han+20], the authors facilitated the channel estimation task in D-RIS-aided systems by assuming that the two RISs are equipped with receive RF chains, i.e., active RISs. This can be used to enable a sensing capability for the RISs to estimate their channels at the transmitter and receiver, separately. On the other hand, channel estimation in D-RIS-aided systems with fully passive RISs was investigated in [YZZ21], considering only the double-reflection links and ignoring the single-reflection links that are assumed to be too weak. To fulfill this gap, the authors in [YZZ21c] proposed a decoupled channel estimation scheme for D-RIS-aided systems, where the single-reflection channels are estimated successively using an ON-OFF method (i.e., by turning one RIS ON at a time, while keeping the other OFF) and then the double-reflection channel is estimated in a decoupled manner, with the signals canceled over the two estimated single-reflection channels. Furthermore, in [BX22], the authors considered D-RIS-aided single-input multiple-output (SIMO) systems and proposed an LMMSE-based channel estimation approach, where the RISs' training design is optimized via an alternating optimization and projected gradient descent framework with the objective of minimizing the channel estimation sum MSE. Notably, the provided simulation results in [BX22] showed that the proposed RISs for the training design approach provides a better channel estimation accuracy compared to a DFT-based RIS training design approach.

5.1.1. Chapter Contribution

In this chapter, a D-RIS-aided MIMO system is investigated with two different training overhead assumptions. We assume that the communication between the transmitter and receiver is only possible via the double reflection link. The summary of the proposed channel estimation algorithms for each training overhead case is summarized as follows.

- In the first channel estimation approach, we assume that both RISs have the same amount of training overhead, which is called channel estimation for joint training (CEJOINT). We show that the received measurement signals in flat-fading D-RIS-aided MIMO systems can be arranged in a 3-way tensor that admits a Tucker2 decomposition [KB09]. Then, an ALS-based method is proposed to estimate the transmitter-to-RIS-1, the RIS-1-to-RIS-2, and the RIS-2-to-receiver channels separately, up to trivial scaling factors.
- In the second channel estimation approach, we assume that the training overhead assigned to each RIS in D-RIS-aided MIMO systems can be adjusted separately. Consequently, we show that the received signals in flat-fading D-RIS-aided MIMO systems can be represented as a 4-way tensor that satisfies a nested PARAFAC decomposition model [AF13; RH09]. Exploiting such a structure, we propose a non-iterative three-step channel estimation method, where one of the three channel matrices in each step can be determined in closed-form using a low-complexity least squares KRF technique [XFA16]. We call it KRF-based for separate training (KRF-SEPT). To further enhance the channel estimation accuracy, we propose an ALS-based channel estimation method where we estimate one channel matrix assuming that the other two are fixed. We call it ALS-based for separate training (ALS-SEPT). The ALS-SEPT method has more relaxed constraints on the training overhead than the closed-form KRF-SEPT method. Differently from our proposed CEJOINT, we propose to initialize two channel matrices using the KRF-SEPT-based method.

The content of this chapter has been published in [Ard+22] and [Ghe+23].

5.1.2. Chapter Organization

We organize this chapter as follows. We start by presenting the system model in Section 5.2. The proposed channel estimation approaches are presented in Section 5.3 for two types of training overhead scenarios. In Section 5.4, a short overview of the benchmark S-RIS-aided system is given with further discussion on the performance of algorithms in terms of their computational complexities and identifiability conditions. The numerical results to evaluate the performance of the proposed algorithms are given in Section 5.5. Finally, a conclusion is provided in Section 5.6.

5.2. System Model

We consider a D-RIS-aided MIMO communication system as illustrated in Fig. 5.1, where a transmitter with M_T antennas is communicating with a receiver with M_R antennas via a D-RIS-aided channel. As depicted in Fig. 5.1, RIS-1 is assumed to be close to the transmitter and has M_{S1} reflecting elements, while RIS-2 is assumed to be close to the receiver and has M_{S2} reflecting elements. We assume that the other propagation channels, i.e., transmitter-to-receiver, transmitter-to-RIS-2, RIS-1-to-receiver, are

unavailable due to blockage or too weak due to high path loss, which is a valid assumption in some scenarios such as in a tunnel.

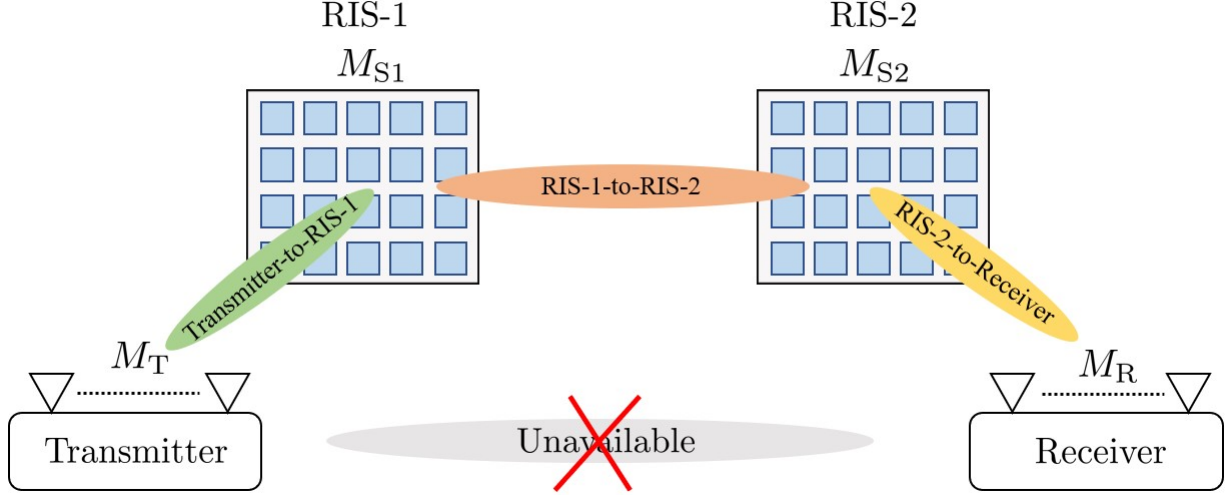


Figure 5.1: A D-RIS aided MIMO Communication System.

Let $\mathbf{H}_T \in \mathbb{C}^{M_{S1} \times M_T}$ be the transmitter-to-RIS-1 channel, $\mathbf{H}_S \in \mathbb{C}^{M_{S2} \times M_{S1}}$ be the RIS-1-to-RIS-2 channel, and $\mathbf{H}_R \in \mathbb{C}^{M_R \times M_{S2}}$ be the RIS-2-to-receiver channel. We assume that we send Q symbols during the training phase, where Q is divided into three transmission blocks as $Q = I \cdot J \cdot P$. The received signal at the (i, j, p) th transmission time, $i \in \{1, \dots, I\}$, $j \in \{1, \dots, J\}$, $p \in \{1, \dots, P\}$, can be expressed as

$$\bar{\mathbf{y}}_{i,j,p} = \mathbf{H}_R \Psi_i \mathbf{H}_S \Phi_j \mathbf{H}_T \mathbf{f}_p s_p + \bar{\mathbf{n}}_{i,j,p} \in \mathbb{C}^{M_R}, \quad (5.1)$$

where $\mathbf{f}_p \in \mathbb{C}^{M_T}$ is the p th training vector at the transmitter with $\|\mathbf{f}_p\| = 1$, $\Phi_j = \text{diag}\{\phi_j\} \in \mathbb{C}^{M_{S1} \times M_{S1}}$ denotes the j th diagonal training matrix of RIS-1, with $\phi_j \in \mathbb{C}^{M_{S1}}$ and $|\phi_j|_{[a]} = \frac{1}{\sqrt{M_{S1}}}$, $\Psi_i = \text{diag}\{\psi_i\} \in \mathbb{C}^{M_{S2} \times M_{S2}}$ is the i th diagonal training matrix of RIS-2, with $\psi_i \in \mathbb{C}^{M_{S2}}$ and $|\psi_i|_{[a]} = \frac{1}{\sqrt{M_{S2}}}$, $s_p \in \mathbb{C}$ is the p th unit-norm training symbol, and $\bar{\mathbf{n}}_{i,j,p} \in \mathbb{C}^{M_R}$ is the AWGN vector having zero-mean circularly symmetric complex-valued entries with variance σ_n^2 .

Let $\Psi = [\psi_1, \dots, \psi_I] \in \mathbb{C}^{M_{S2} \times I}$, $\Phi = [\phi_1, \dots, \phi_J] \in \mathbb{C}^{M_{S1} \times J}$, and $\mathbf{F} = [\mathbf{f}_{1S1}, \dots, \mathbf{f}_{PS1}] \in \mathbb{C}^{M_T \times P}$. Then, by stacking the received measurement signals $\{\bar{\mathbf{y}}_{i,j,p}, \forall p\}$ next to each other as

$$\bar{\mathbf{Y}}_{i,j} = [\bar{\mathbf{y}}_{i,j,1}, \dots, \bar{\mathbf{y}}_{i,j,P}], \quad (5.2)$$

we obtain a measurement matrix $\bar{\mathbf{Y}}_{i,j}$, which can be expressed as

$$\bar{\mathbf{Y}}_{i,j} = \mathbf{H}_R \Psi_i \mathbf{H}_S \Phi_j \mathbf{H}_T \mathbf{F} + \bar{\mathbf{N}}_{i,j} \in \mathbb{C}^{M_R \times P}, \quad (5.3)$$

where $\bar{\mathbf{N}}_{i,j} \in \mathbb{C}^{M_R \times P}$ is defined similarly to $\bar{\mathbf{Y}}_{i,j}$. We assume that the training matrix \mathbf{F} is designed with orthonormal rows, i.e., $\mathbf{F} \mathbf{F}^H = \mathbf{I}_{M_T}$, which implies that $P \geq M_T$. Then, the right filtered measurement

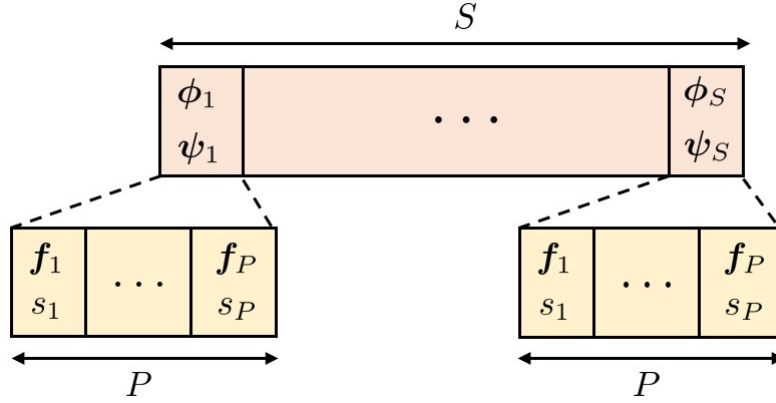


Figure 5.2: Transmission blocks for joint channel training of RISs.

matrix $\mathbf{Y}_{i,j} = \bar{\mathbf{Y}}_{i,j} \mathbf{F}^H$, can be written as

$$\mathbf{Y}_{i,j} = \mathbf{H}_R \Psi_i \mathbf{H}_S \Phi_j \mathbf{H}_T + \mathbf{N}_{i,j} \in \mathbb{C}^{M_R \times M_T}, \quad (5.4)$$

where $\mathbf{N}_{i,j} = \bar{\mathbf{N}}_{i,j} \mathbf{F}^H$.

5.3. Algorithms

Here, we discuss different training assumptions for RISs and propose the corresponding channel estimation techniques to estimate the channel matrices \mathbf{H}_T , \mathbf{H}_S , \mathbf{H}_R from the measurement matrices $\mathbf{Y}_{i,j}$, $\forall i, \forall j$, given in (5.4).

5.3.1. Joint Channel Training of RISs

Assuming that the training RIS matrices Ψ and Φ have the same number of training vectors S , i.e., $S = I = J$ and we vary them together, therefore Q is reduced from $Q = S^2 \cdot P$ to $Q = S \cdot P$, as shown in Fig. 5.2. Accordingly, the right filtered measurement matrix in (5.4) can be rewritten as

$$\mathbf{Y}_s = \mathbf{H}_R \Psi_s \mathbf{H}_S \Phi_s \mathbf{H}_T + \mathbf{N}_s \in \mathbb{C}^{M_R \times M_T}, \quad (5.5)$$

where $s \in \{1, \dots, S\}$. By concatenating $\mathbf{Y}_1, \dots, \mathbf{Y}_S$ in (5.5) behind each other (in the 3-mode), a 3-way tensor can be obtained as

$$\mathcal{Y} = [\mathbf{Y}_1, \sqcup_3, \dots, \sqcup_3, \mathbf{Y}_S] \in \mathbb{C}^{M_R \times M_T \times S}, \quad (5.6)$$

where \mathbf{Y}_s represents its s th frontal slice. Here, we note that the tensor \mathcal{Y} has a Tucker2 representation given as

$$\mathcal{Y} = \mathcal{S} \times_1 \mathbf{H}_R \times_2 \mathbf{H}_T^T + \mathcal{N} \in \mathbb{C}^{M_R \times M_T \times S}, \quad (5.7)$$

where \mathcal{N} is the noise tensor and $\mathcal{S} \in \mathbb{C}^{M_{S2} \times M_{S1} \times S}$ is the core tensor formed by concatenating $\Psi_s \mathbf{H}_S \Phi_s$, $s \in \{1, \dots, S\}$, behind each other in the 3-mode as

$$\mathcal{S} = [\Psi_1 \mathbf{H}_S \Phi_1, \sqcup_3, \dots, \sqcup_3, \Psi_S \mathbf{H}_S \Phi_S] \in \mathbb{C}^{M_{S2} \times M_{S1} \times S}. \quad (5.8)$$

Note that, the vectorized form of the s th frontal slice of the tensor \mathcal{S} , i.e., $\Psi_s \mathbf{H}_S \Phi_s$ can be written as

$$\begin{aligned} \text{vec}\{\Psi_s \mathbf{H}_S \Phi_s\} &= (\Phi_s^\top \otimes \Psi_s) \mathbf{h}_S \\ &\stackrel{(h)}{=} (\Phi_s \otimes \Psi_s) \mathbf{h}_S \\ &\stackrel{(i)}{=} \text{diag}\{(\phi_s \otimes \psi_s)\} \mathbf{h}_S \\ &= \text{diag}\{\mathbf{h}_S\} (\phi_s \otimes \psi_s), \end{aligned} \quad (5.9)$$

where $\mathbf{h}_S = \text{vec}\{\mathbf{H}_S\}$, $\stackrel{(h)}{=}$ is obtained due to the diagonal structure of Φ_s , i.e., $\Phi_s = \Phi_s^\top$, and $\stackrel{(i)}{=}$ is obtained due to the diagonal structures of Φ_s and Ψ_s , i.e., $\Phi_s = \text{diag}\{\phi_s\}$ and $\Psi_s = \text{diag}\{\psi_s\}$ and therefore $(\Phi_s \otimes \Psi_s) = (\text{diag}\{\phi_s\} \otimes \text{diag}\{\psi_s\}) = \text{diag}\{(\phi_s \otimes \psi_s)\}$. From (5.9), we define a matrix $\bar{\mathbf{S}}$ as

$$\begin{aligned} \bar{\mathbf{S}} &= [\text{diag}\{\mathbf{h}_S\}(\phi_1 \otimes \psi_1), \dots, \text{diag}\{\mathbf{h}_S\}(\phi_S \otimes \psi_S)] \\ &= \text{diag}\{\mathbf{h}_S\}[(\phi_1 \otimes \psi_1), \dots, (\phi_S \otimes \psi_S)] \\ &\stackrel{(j)}{=} \text{diag}\{\mathbf{h}_S\}(\Phi \diamond \Psi) \in \mathbb{C}^{M_{S2} M_{S1} \times S}, \end{aligned} \quad (5.10)$$

where $\stackrel{(j)}{=}$ is obtained from (1.2). Here, $\bar{\mathbf{S}}$ in (5.10) can be seen as the transposed version of the 3-mode unfolding of the tensor \mathcal{S} given in (5.8), i.e., $[\mathcal{S}]_{(3)}$

$$[\mathcal{S}]_{(3)} = (\Phi \diamond \Psi)^\top \text{diag}\{\mathbf{h}_S\} \in \mathbb{C}^{S \times M_{S2} M_{S1}}, \quad (5.11)$$

and the n -mode unfolding, $n = \{1, 2\}$ of the tensor \mathcal{S} can be expressed as

$$[\mathcal{S}]_{(1)} = (\Psi^\top \diamond \mathbf{H}_S^\top)^\top \text{diag}\{\text{vec}\{\Phi\}\} \in \mathbb{C}^{M_{S2} \times M_{S1} S}, \quad (5.12)$$

$$[\mathcal{S}]_{(2)} = (\Phi^\top \diamond \mathbf{H}_S)^\top \text{diag}\{\text{vec}\{\Psi\}\} \in \mathbb{C}^{M_{S1} \times M_{S2} S}. \quad (5.13)$$

From (5.7), the channel estimation problem can be formulated as

$$\{\hat{\mathbf{H}}_T, \hat{\mathbf{H}}_S, \hat{\mathbf{H}}_R\} = \underset{\mathbf{H}_T, \mathbf{H}_S, \mathbf{H}_R}{\text{argmin}} \|\mathcal{Y} - \mathcal{S} \times_1 \mathbf{H}_R \times_2 \mathbf{H}_T^\top\|_{\mathbb{F}}^2, \quad (5.14)$$

which is non-convex due to its trilinear structure. To obtain a solution, we resort to an alternating minimization approach, where we solve (5.14) for one variable assuming the other two are fixed. To achieve this end, we exploit the n -mode unfolding of \mathcal{Y} , i.e., $[\mathcal{Y}]_{(n)}$, $n \in \{1, 2, 3\}$, expressed as [KB09; CLA09]

$$[\mathcal{Y}]_{(1)} = \mathbf{H}_R \mathbf{Z}_R(\mathbf{H}_T, \mathbf{H}_S) + [\mathcal{N}]_{(1)} \in \mathbb{C}^{M_R \times M_T S} \quad (5.15)$$

$$[\mathcal{Y}]_{(2)} = \mathbf{H}_T^\top \mathbf{Z}_T(\mathbf{H}_R, \mathbf{H}_S) + [\mathcal{N}]_{(2)} \in \mathbb{C}^{M_T \times M_R S} \quad (5.16)$$

$$[\mathcal{Y}]_{(3)} = [\mathcal{S}]_{(3)} (\mathbf{H}_T^\top \otimes \mathbf{H}_R)^\top + [\mathcal{N}]_{(3)} \in \mathbb{C}^{S \times M_T M_R}, \quad (5.17)$$

where

$$\mathbf{Z}_R(\mathbf{H}_T, \mathbf{H}_S) = [\mathcal{S}]_{(1)} (\mathbf{I}_S \otimes \mathbf{H}_T^\top)^\top \in \mathbb{C}^{M_{S2} \times S M_T}, \quad (5.18)$$

$$\mathbf{Z}_T(\mathbf{H}_R, \mathbf{H}_S) = [\mathcal{S}]_{(2)} (\mathbf{I}_S \otimes \mathbf{H}_R)^\top \in \mathbb{C}^{M_{S1} \times S M_R}. \quad (5.19)$$

Substituting $[\mathcal{S}]_{(3)}$ given in (5.11) into $[\mathcal{Y}]_{(3)}$ given in (5.17), we can rewrite it as

$$[\mathcal{Y}]_{(3)} = (\Phi \diamond \Psi)^\top \text{diag}\{h_S\} (\mathbf{H}_T^\top \otimes \mathbf{H}_R)^\top + [\mathcal{N}]_{(3)}. \quad (5.20)$$

Therefore, the vectorized form of $[\mathcal{Y}]_{(3)}$ using (1.20), i.e, $\mathbf{y}_{(3)} = \text{vec}\{[\mathcal{Y}]_{(3)}\}$ can be expressed as

$$\mathbf{y}_{(3)} = \mathbf{Z}_S(\mathbf{H}_T, \mathbf{H}_R) \mathbf{h}_S + \mathbf{n}_{(3)} \in \mathbb{C}^{S M_T M_R}, \quad (5.21)$$

where

$$\mathbf{Z}_S(\mathbf{H}_T, \mathbf{H}_R) = [(\mathbf{H}_T^\top \otimes \mathbf{H}_R) \diamond (\Phi \diamond \Psi)^\top] \in \mathbb{C}^{S M_T M_R \times M_{S1} M_{S2}}, \quad (5.22)$$

and $\mathbf{n}_{(3)} = \text{vec}\{[\mathcal{N}]_{(3)}\}$. By exploiting (5.15), (5.16), and (5.21), estimates of \mathbf{H}_R , \mathbf{H}_T , and \mathbf{h}_S can be obtained sequentially as

$$\hat{\mathbf{H}}_R = \underset{\mathbf{H}_R}{\text{argmin}} \|\mathcal{Y}_{(1)} - \mathbf{H}_R \mathbf{Z}_R(\mathbf{H}_T, \mathbf{H}_S)\|_F^2 \quad (5.23)$$

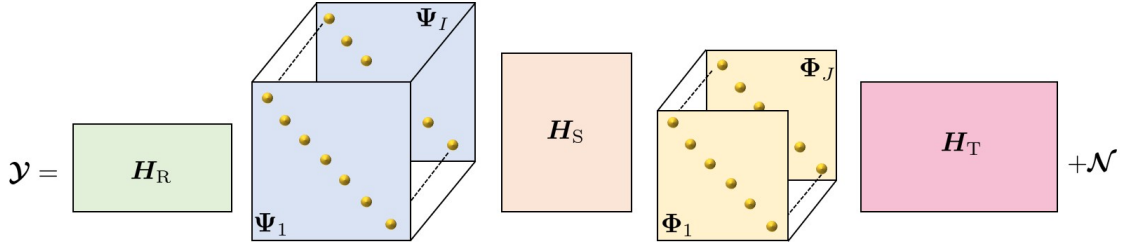
$$\hat{\mathbf{H}}_T = \underset{\mathbf{H}_T}{\text{argmin}} \|\mathcal{Y}_{(2)} - \mathbf{H}_T^\top \mathbf{Z}_T(\mathbf{H}_R, \mathbf{H}_S)\|_F^2 \quad (5.24)$$

$$\hat{\mathbf{h}}_S = \underset{\mathbf{h}_S}{\text{argmin}} \|\mathbf{y}_{(3)} - \mathbf{Z}_S(\mathbf{H}_T, \mathbf{H}_R) \mathbf{h}_S\|_2^2. \quad (5.25)$$

The above problems are convex and can be solved using the ALS method. A summary of the proposed channel estimation for joint training (CEJOINT) is given in Algorithm 5.7.

Algorithm 5.7 Channel estimation for joint training (CEJOINT).

- 1: Input: Measurement tensor $\mathcal{Y} \in \mathbb{C}^{M_R \times M_T \times S}$ as in (5.7)
 - 2: Initialize: $\mathbf{H}_T^{(0)}$ and $\mathbf{H}_S^{(0)}$ and select $N_{\text{max-itr}}$, i.e., the maximum number of iterations.
 - 3: **for** $n = 1$ to $N_{\text{max-itr}}$ **do**
 - 4: $\mathbf{H}_R^{(n)} = [\mathcal{Y}]_{(1)} \{ \mathbf{Z}_R(\mathbf{H}_T^{(n-1)}, \mathbf{H}_S^{(n-1)}) \}^+$
 - 5: $\mathbf{H}_T^{(n)} = ([\mathcal{Y}]_{(2)} \{ \mathbf{Z}_T(\mathbf{H}_R^{(n)}, \mathbf{H}_S^{(n-1)}) \}^+)^\top$
 - 6: $\mathbf{H}_S^{(n)} = \text{unvec}_{M_{S2} \times M_{S1}} \{ \mathbf{Z}_S(\mathbf{H}_T^{(n)}, \mathbf{H}_R^{(n)}) \}^+ \mathbf{y}_{(3)} \}$
 - 7: **end for**
 - 8: Output: $\hat{\mathbf{H}}_R$, $\hat{\mathbf{H}}_T$, and $\hat{\mathbf{H}}_S$
-


 Figure 5.3: Graphical 3D representation of the 4-way tensor \mathcal{Y} .

5.3.2. Separate Channel Training of RISs

Here, we consider a more general scenario in which the training overhead assigned to each RIS in a D-RIS-aided MIMO system can be adjusted separately, i.e., $Q = I \cdot J \cdot P$, as shown in Fig. 5.4. Let $\bar{\mathcal{Y}}_j$ be a 3-way tensor constructed by concatenating the j th block measurement matrices $\mathbf{Y}_{1,j}, \mathbf{Y}_{2,j}, \dots, \mathbf{Y}_{I,j}$ given in (5.4) along the second dimension, where we consider the rows of the matrices as the 1-mode and the columns of the matrices as 3-mode, as

$$\bar{\mathcal{Y}}_j = [\mathbf{Y}_{1,j}, \sqcup_2, \mathbf{Y}_{2,j}, \dots, \sqcup_2, \mathbf{Y}_{I,j}] \in \mathbb{C}^{M_R \times I \times M_T}, \quad (5.26)$$

where \sqcup_n denotes the concatenation along the n th dimension. Moreover, let \mathcal{Y} be a 4-way tensor constructed by concatenating the measurement tensors $\bar{\mathcal{Y}}_1, \bar{\mathcal{Y}}_2, \dots, \bar{\mathcal{Y}}_J$ along the fourth dimension as

$$\mathcal{Y} = [\bar{\mathcal{Y}}_1, \sqcup_4, \bar{\mathcal{Y}}_2, \dots, \sqcup_4, \bar{\mathcal{Y}}_J] \in \mathbb{C}^{M_R \times I \times M_T \times J}. \quad (5.27)$$

Here, we observe that the 4-way tensor \mathcal{Y} can be interpreted as a nested PARAFAC decomposition [AF13], as shown in Fig. 5.3. Please refer to Appendix A.7 for more details. Specifically, let $\bar{\Phi} = \Phi^T \in \mathbb{C}^{J \times M_{S1}}$, $\bar{\Psi} = \Psi^T \in \mathbb{C}^{I \times M_{S2}}$, and define the following two 3-way PARAFAC tensors \mathcal{A} and \mathcal{B} as

$$\mathcal{A} = \mathcal{I}_{3, M_{S1}} \times_1 \mathbf{H}_T^T \times_2 \bar{\Phi} \times_3 \mathbf{H}_S \in \mathbb{C}^{M_T \times J \times M_{S2}}, \quad (5.28)$$

$$\mathcal{B} = \mathcal{I}_{3, M_{S2}} \times_1 \mathbf{H}_R \times_2 \bar{\Psi} \times_3 \mathbf{H}_S^T \in \mathbb{C}^{M_R \times I \times M_{S1}}. \quad (5.29)$$

Then, it can be easily shown that the n -mode unfoldings of the 4-way tensor \mathcal{Y} , $n = \{1, 2, 3, 4\}$, can be expressed as

$$[\mathcal{Y}]_{(1)} = \mathbf{H}_R ([\mathcal{A}]_{(3)}^T \diamond \bar{\Psi})^T + [\mathcal{N}]_{(1)} \in \mathbb{C}^{M_R \times J M_T I}, \quad (5.30)$$

$$[\mathcal{Y}]_{(2)} = \bar{\Psi} ([\mathcal{A}]_{(3)}^T \diamond \mathbf{H}_R)^T + [\mathcal{N}]_{(2)} \in \mathbb{C}^{I \times M_R M_T J}, \quad (5.31)$$

$$[\mathcal{Y}]_{(3)} = \mathbf{H}_T^T (\bar{\Phi} \diamond [\mathcal{B}]_{(3)}^T)^T + [\mathcal{N}]_{(3)} \in \mathbb{C}^{M_T \times J M_R I}, \quad (5.32)$$

$$[\mathcal{Y}]_{(4)} = \bar{\Phi} (\mathbf{H}_T^T \diamond [\mathcal{B}]_{(3)}^T)^T + [\mathcal{N}]_{(4)} \in \mathbb{C}^{J \times M_R I M_T}, \quad (5.33)$$

where $[\mathcal{A}]_{(3)}$ and $[\mathcal{B}]_{(3)}$ are the 3-mode unfolding of \mathcal{A} and \mathcal{B} , respectively, which can be expressed as

$$[\mathcal{A}]_{(3)} = \mathbf{H}_S (\bar{\Phi} \diamond \mathbf{H}_T^T)^T \in \mathbb{C}^{M_{S2} \times J M_T}, \quad (5.34)$$

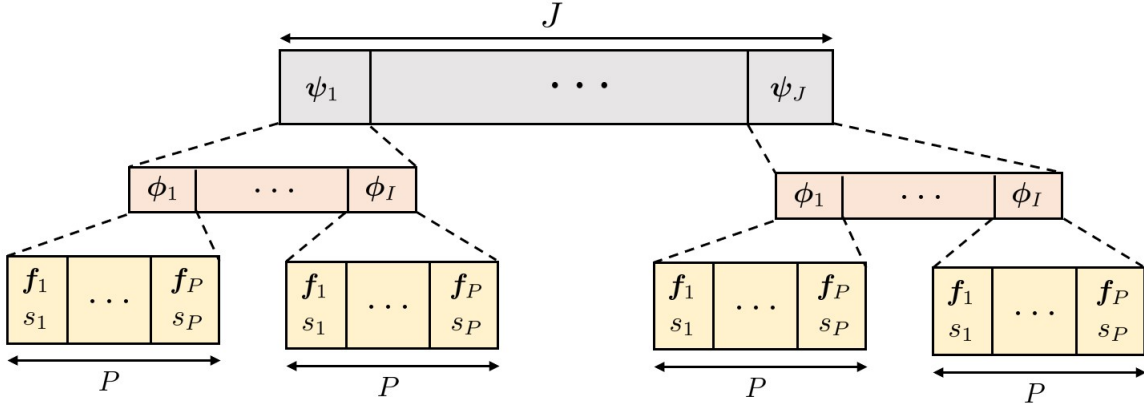


Figure 5.4: Transmission blocks for separate channel training of RISs.

$$[\mathcal{B}]_{(3)} = \mathbf{H}_S^T (\bar{\Psi} \diamond \mathbf{H}_R)^T \in \mathbb{C}^{M_{S1} \times I M_R}. \quad (5.35)$$

In the following, by exploiting the above tensor structure, we propose two solutions for estimating \mathbf{H}_T , \mathbf{H}_S , and \mathbf{H}_R : one is a closed-form solution called KRF-based for separate training (KRF-SEPT) and the other is an iterative solution called ALS-based for separate training (ALS-SEPT).

5.3.2.1 KRF-SEPT Algorithm

Assume that the training matrices $\bar{\Phi}$ and $\bar{\Psi}$ are designed with orthonormal columns, i.e., $\bar{\Phi}^H \bar{\Phi} = \mathbf{I}_J$ and $\bar{\Psi}^H \bar{\Psi} = \mathbf{I}_I$. Then, the left-filtered 2-mode and 4-mode unfolding matrices are given as

$$[\tilde{\mathcal{Y}}]_{(2)} = \bar{\Psi}^H [\mathcal{Y}]_{(2)} = ([\mathcal{A}]_{(3)}^T \diamond \mathbf{H}_R)^T + \bar{\Psi}^H [\mathcal{N}]_{(2)}, \quad (5.36)$$

$$[\tilde{\mathcal{Y}}]_{(4)} = \bar{\Phi}^H [\mathcal{Y}]_{(4)} = (\mathbf{H}_T^T \diamond [\mathcal{B}]_{(3)}^T)^T + \bar{\Phi}^H [\mathcal{N}]_{(4)}. \quad (5.37)$$

Given $[\tilde{\mathcal{Y}}]_{(2)}$ and $[\tilde{\mathcal{Y}}]_{(4)}$ as above, the channel matrices \mathbf{H}_T and \mathbf{H}_R can be obtained as a solution to

$$\hat{\mathbf{H}}_R = \underset{\mathbf{H}_R}{\operatorname{argmin}} \left\| [\tilde{\mathcal{Y}}]_{(2)}^T - ([\mathcal{A}]_{(3)}^T \diamond \mathbf{H}_R) \right\|_F^2 \quad (5.38)$$

$$\hat{\mathbf{H}}_T = \underset{\mathbf{H}_T}{\operatorname{argmin}} \left\| [\tilde{\mathcal{Y}}]_{(4)}^T - (\mathbf{H}_T^T \diamond [\mathcal{B}]_{(3)}^T) \right\|_F^2. \quad (5.39)$$

Clearly, both problems have a similar structure given as

$$[\mathbf{L}, \mathbf{R}] = \underset{\mathbf{L}, \mathbf{R}}{\operatorname{argmin}} \left\| \mathbf{E} - (\mathbf{L} \diamond \mathbf{R}) \right\|_F^2, \quad (5.40)$$

assuming that $\mathbf{E} = (\mathbf{L} \diamond \mathbf{R}) \in \mathbb{C}^{A \times B}$, $\mathbf{L} \in \mathbb{C}^{A_1 \times B}$, $\mathbf{R} \in \mathbb{C}^{A_2 \times B}$, and $A = A_1 A_2$. A closed-form solution to (5.40) can be obtained using the LS Khatri-Rao Factorization (LSKRF) technique [RH09], as summarized in Algorithm 5.8¹.

¹Note that $\mathbf{E}_b = [\mathbf{R}]_{[:,b]} [\mathbf{L}]_{[:,b]}^T \in \mathbb{C}^{A_2 \times A_1}$ in Algorithm 5.8 is a rank one matrix. Its SVD can be efficiently calculated using the Power Iteration method, which has a low computational complexity.

Algorithm 5.8 Khatri-Rao Factorization (KRF).

- 1: **Input:** $\mathbf{E} \in \mathbb{C}^{A \times B}$
 - 2: **for** $b = 1$ to B **do**
 - 3: Get $\mathbf{E}_b = \text{unvec}_{A_2 \times A_1} \{[\mathbf{E}]_{[:,b]}\} \in \mathbb{C}^{A_2 \times A_1}$
 - 4: Compute SVD of \mathbf{E}_b as $\mathbf{E}_b = \mathbf{u}_b \sigma_b \mathbf{v}_b^H \in \mathbb{C}^{A_2 \times A_1}$
 - 5: Set $[\hat{\mathbf{R}}]_{[:,b]} = \sqrt{\sigma_b} \mathbf{u}_b$ and $[\hat{\mathbf{L}}]_{[:,b]} = \sqrt{\sigma_b} \mathbf{v}_b^*$
 - 6: **end for**
 - 7: **Output:** $\hat{\mathbf{L}} \in \mathbb{C}^{A_1 \times B}$ and $\hat{\mathbf{R}} \in \mathbb{C}^{A_2 \times B}$
-

Algorithm 5.9 KRF-based for separate training (KRF-SEPT).

- 1: **Input:** Measurement tensor $\mathcal{Y} \in \mathbb{C}^{M_R \times I \times M_T \times J}$ as in (5.27)
 - 2: Get $\hat{\mathbf{H}}_R$ as a solution to (5.38) via Algorithm 5.8
 - 3: Get $\hat{\mathbf{H}}_T$ as a solution to (5.39) via Algorithm 5.8
 - 4: Given $\hat{\mathbf{H}}_R$ and $\hat{\mathbf{H}}_T$, get $\hat{\mathbf{H}}_S$ using (5.44)
 - 5: Output: $\hat{\mathbf{H}}_R$, $\hat{\mathbf{H}}_T$, and $\hat{\mathbf{H}}_S$
-

On the other hand, to estimate \mathbf{H}_S , we exploit the observation that each n -mode unfolding $[\mathcal{Y}]_{(n)}$ of the 4-way tensor \mathcal{Y} can be seen as a 1-mode unfolding of a 3-way tensor. For instance, the 3-mode unfolding $[\mathcal{Y}]_{(3)}$ is equal to the 1-mode unfolding of a 3-way tensor $\mathcal{Z} \in \mathbb{C}^{M_T \times I M_R \times J}$ given as

$$\mathcal{Z} = \mathcal{I}_{3, M_{S1}} \times_1 \mathbf{H}_T^T \times_2 [\mathcal{B}]_{(3)}^T \times_3 \bar{\Phi} + \mathcal{Q}, \quad (5.41)$$

where $\mathcal{Q} \in \mathbb{C}^{M_T \times I M_R \times J}$ is the 3-way noise tensor representation of the 4-way noise tensor $\mathcal{N} \in \mathbb{C}^{M_R \times I \times M_T \times J}$. Clearly, $[\mathcal{Z}]_{(1)} = [\mathcal{Y}]_{(3)}$. The 2-mode unfolding of \mathcal{Z} is given as

$$\begin{aligned} [\mathcal{Z}]_{(2)} &= [\mathcal{B}]_{(3)}^T (\bar{\Phi} \diamond \mathbf{H}_T^T)^T \\ &= (\bar{\Psi} \diamond \mathbf{H}_R) \mathbf{H}_S (\bar{\Phi} \diamond \mathbf{H}_T^T)^T \in \mathbb{C}^{I M_R \times J M_T}. \end{aligned} \quad (5.42)$$

Let $\mathbf{z}_2 = \text{vec}\{[\mathcal{Z}]_{(2)}\}$, which yields

$$\mathbf{z}_{(2)} = [(\bar{\Phi} \diamond \mathbf{H}_T^T) \otimes (\bar{\Psi} \diamond \mathbf{H}_R)] \mathbf{h}_S = \mathbf{V}_S \mathbf{h}_S, \quad (5.43)$$

where $\mathbf{V}_S = [(\bar{\Phi} \diamond \mathbf{H}_T^T) \otimes (\bar{\Psi} \diamond \mathbf{H}_R)] \in \mathbb{C}^{J I M_T M_R \times M_{S1} M_{S2}}$. Then, \mathbf{H}_S can be estimated as

$$\hat{\mathbf{H}}_S = \text{unvec}_{M_{S2} \times M_{S1}} \{(\mathbf{V}_S)^+ \mathbf{z}_{(2)}\}. \quad (5.44)$$

The summary of the proposed closed-form KRF-SEPT method is given in Algorithm 5.9.

5.3.2.2 ALS-SEPT Algorithm

An ALS-based method can also be used to estimate the channel matrices \mathbf{H}_T , \mathbf{H}_S , and \mathbf{H}_R . Specifically, given $[\mathcal{Y}]_{(1)}$ and $[\mathcal{Y}]_{(3)}$ in (5.30) and (5.32), respectively, an estimate for \mathbf{H}_R and \mathbf{H}_T can be obtained as

$$\hat{\mathbf{H}}_R = [\mathcal{Y}]_{(1)} (([\mathcal{A}]_{(3)}^T \diamond \bar{\Psi})^T)^+ = [\mathcal{Y}]_{(1)} (\mathbf{V}_R)^+ \quad (5.45)$$

Algorithm 5.10 ALS-based for separate training (ALS-SEPT).

- 1: **Input:** Measurement tensor $\mathcal{Y} \in \mathbb{C}^{M_R \times J \times M_T \times I}$ as in (5.27)
 - 2: Initialize $\mathbf{H}_R^{(0)}$ and $\mathbf{H}_T^{(0)}$, e.g., Randomly or as solution to (5.38) and (5.39), respectively.
 - 3: Select $N_{\max\text{-itr}} \geq 1$
 - 4: **for** $n = 1$ to $N_{\max\text{-itr}}$ **do**
 - 5: Given $\mathbf{H}_R^{(n-1)}$ and $\mathbf{H}_T^{(n-1)}$, get $\hat{\mathbf{H}}_S^{(n)}$ using (5.44)
 - 6: Given $\mathbf{H}_T^{(n-1)}$ and $\hat{\mathbf{H}}_S^{(n)}$, get $\hat{\mathbf{H}}_R^{(n)}$ using (5.45)
 - 7: Given $\hat{\mathbf{H}}_R^{(n)}$ and $\hat{\mathbf{H}}_S^{(n)}$, get $\hat{\mathbf{H}}_T^{(n)}$ using (5.46)
 - 8: **end for**
 - 9: Output: $\hat{\mathbf{H}}_R$, $\hat{\mathbf{H}}_T$, and $\hat{\mathbf{H}}_S$
-

$$\hat{\mathbf{H}}_T^\top = [\mathcal{Y}]_{(3)} ((\bar{\Phi} \diamond [\mathcal{B}]_{(3)}^\top)^\top)^\dagger = [\mathcal{Y}]_{(3)} (\mathbf{V}_T)^\dagger, \quad (5.46)$$

where $[\mathcal{A}]_{(3)}$ and $[\mathcal{B}]_{(3)}$ are given by (5.34) and (5.35), respectively, while $\mathbf{V}_R \in \mathbb{C}^{M_{S2} \times JIM_T}$ and $\mathbf{V}_T \in \mathbb{C}^{M_{S1} \times JIM_R}$ are defined as

$$\mathbf{V}_R = ([\mathcal{A}]_{(3)}^\top \diamond \bar{\Psi})^\top \in \mathbb{C}^{M_{S2} \times JIM_T} \quad (5.47)$$

$$\mathbf{V}_T = (\bar{\Phi} \diamond [\mathcal{B}]_{(3)}^\top)^\top \in \mathbb{C}^{M_{S1} \times JIM_R}. \quad (5.48)$$

Then, using (5.44), (5.45), and (5.46), an ALS-based method can be used to estimate \mathbf{H}_S , \mathbf{H}_R , and \mathbf{H}_T , respectively, as summarized in Algorithm 5.10. However, it is known that the convergence rate of ALS-based methods depends on the initialization, where a good initialization strategy generally leads to a fast convergence rate, and therefore a reduced computational complexity. Motivated by the closed-form KRF-SEPT, we propose to initialize \mathbf{H}_R and \mathbf{H}_T using (5.38) and (5.39), respectively.

5.4. Discussion

5.4.1. Comparison with S-RIS-aided systems

We compare the proposed CEJOINT method for D-RIS systems to the ALS-based approach for S-RIS systems proposed in [AA20; AAB21] in terms of the minimum training overhead and the estimation accuracy. In S-RIS-aided system, the communication between the transmitter and the receiver with M_T and M_R antennas, respectively, is aided via a single RIS with M_S reflecting elements, as depicted in Fig. 4.1. Let $\mathbf{G}_T \in \mathbb{C}^{M_S \times M_T}$ be the transmitter-to-RIS channel and $\mathbf{G}_R \in \mathbb{C}^{M_R \times M_S}$ be the RIS-to-receiver channel. Then, it was shown in [AA20; AAB21] that the received signals at the receiver can be arranged in a 3-way tensor admitting the CP decomposition given as

$$\mathcal{X} = \mathcal{I}_{3,N} \times_1 \mathbf{G}_R \times_2 \mathbf{G}_T^\top \times_3 \mathbf{\Omega}^\top + \mathcal{E} \in \mathbb{C}^{M_R \times M_T \times Q}, \quad (5.49)$$

where $\mathcal{I}_{3,N} \in \mathbb{Z}^{M_S \times M_S \times M_S}$ is a super-diagonal tensor, \mathcal{E} is the noise tensor, $\mathbf{\Omega} = [\omega_1, \dots, \omega_Q] \in \mathbb{C}^{M_S \times Q}$ is the RIS training matrix with Q training beams and $|\omega_q|_{[c]} = 1/\sqrt{M_S}$. The n -mode unfolding of \mathcal{X} ,

$n \in \{1, 2\}$ can be expressed as

$$[\mathcal{X}]_{(1)} = \mathbf{G}_R \mathbf{T}_R(\mathbf{G}_T) + [\mathcal{E}]_{(1)} \in \mathbb{C}^{M_R \times Q M_T} \quad (5.50)$$

$$[\mathcal{X}]_{(2)} = \mathbf{G}_T^T \mathbf{T}_T(\mathbf{G}_R) + [\mathcal{E}]_{(2)} \in \mathbb{C}^{M_T \times Q M_R} \quad (5.51)$$

where $\mathbf{T}_R(\mathbf{G}_T) = (\boldsymbol{\Omega}^T \diamond \mathbf{G}_T^T)^T \in \mathbb{C}^{M_S \times Q M_T}$ and $\mathbf{T}_T(\mathbf{G}_R) = (\boldsymbol{\Omega}^T \diamond \mathbf{G}_R)^T \in \mathbb{C}^{M_S \times Q M_R}$. Therefore, an ALS-based method, similarly to CEJOINT in Algorithm 5.7, has been proposed in [AAB21] to obtain an estimate of \mathbf{G}_T and \mathbf{G}_R , as summarized in Algorithm 5.11.

Algorithm 5.11 ALS method for channel estimation in S-RIS MIMO systems [AAB21].

- 1: Input: Measurement tensor $\mathcal{X} \in \mathbb{C}^{M_R \times M_T \times Q}$ as in (5.49)
 - 2: Initialize: $\mathbf{G}_T^{(0)}$ and select $N_{\max\text{-itr}}$.
 - 3: **for** $n = 1$ to $N_{\max\text{-itr}}$ **do**
 - 4: $\mathbf{G}_R^{(n)} = [\mathcal{X}]_{(1)} \{ \mathbf{T}_R(\mathbf{G}_T^{(n-1)}) \}^+$
 - 5: $\mathbf{G}_T^{(n)} = ([\mathcal{X}]_{(2)} \{ \mathbf{T}_T(\mathbf{G}_R^{(n)}) \}^+)^T$
 - 6: **end for**
 - 7: Output: $\hat{\mathbf{G}}_R, \hat{\mathbf{G}}_T$
-

5.4.2. Uniqueness and Identifiability Conditions

Here, we discuss the range of parameter settings required to ensure the uniqueness and identifiability of the algorithms. First, recall (5.50) and (5.51) for S-RIS-aided systems. The identifiability, in the LS sense, can be obtained by noting that \mathbf{T}_R and \mathbf{T}_T must have full column-rank, i.e., $Q M_T \geq M_S$ and $Q M_R \geq M_S$. Similarly, the CEJOINT method in D-RIS-aided systems needs to have full column-rank for \mathbf{Z}_R in (5.15) and \mathbf{Z}_T in (5.16), i.e., $S M_T \geq M_{S2}$ and $S M_R \geq M_{S1}$, respectively, while \mathbf{Z}_S needs to have full row-rank, i.e., $S M_R M_T \geq M_{S1} M_{S2}$ [CLA09]. In the proposed ALS-SEPT method, \mathbf{V}_R and \mathbf{V}_T must have full column-rank and \mathbf{V}_S must have full row-rank. On the other hand, the KRF-SEPT method requires that the constraints (C1) $I \geq M_{S2}$ and (C2) $J \geq M_{S1}$ are satisfied. In Table 5.1, a summary of the identifiability conditions of all methods is given.

5.4.3. Computational Complexity

Here, we derive the computational complexity of the algorithms in terms of FLOPS. Considering a matrix $\mathbf{A} \in \mathbb{C}^{a \times b}$, we know that the computational complexity of its Moore-Penrose inverse is on the order of $\mathcal{O}(\min\{a, b\}^3)$ [Med04; Pen55]. Moreover, the computation of its dominant singular value and singular vectors is on the order of $\mathcal{O}(l \cdot (a^2 b + b^2 a))$, where l is the number of iterations of the power method as shown in [Sun+19]. Since the matrix \mathbf{E} in the equivalent KRF formulation in (5.40) is a rank-one matrix, its SVD can be efficiently calculated with a low complexity. Using the above information, we summarize the computational complexity of the algorithms in Table 5.2.

Table 5.1: Identifiability condition of algorithms.

Algorithm	Identifiability Condition
ALS for S-RIS in 5.4.1	$Q \geq \max\{\lceil \frac{M_S}{M_T} \rceil, \lceil \frac{M_S}{M_R} \rceil\}$
CEJOINT	$S \geq \max\{\lceil \frac{M_{S2}}{M_T} \rceil, \lceil \frac{M_{S1}}{M_R} \rceil, \lceil \frac{M_{S1} M_{S2}}{M_R M_T} \rceil\}$
KRF-SEPT	$I \geq M_{S2}, J \geq M_{S1}, \text{ and } IJ \geq \frac{M_{S1} M_{S2}}{M_R M_T}$
ALS-SEPT	$IJ \geq \max\{\lceil \frac{M_{S2}}{M_T} \rceil, \lceil \frac{M_{S1}}{M_R} \rceil, \lceil \frac{M_{S1} M_{S2}}{M_R M_T} \rceil\}$

5.4.4. Scaling Ambiguity

Assuming that the identifiability conditions of all algorithms given in Table 5.1 are satisfied, the MIMO channel matrices estimated by Algorithm 5.7 for D-RIS (CEJOINT) and Algorithm 5.11 for S-RIS from [AAB21] are unique up to scalar ambiguities. In particular, the estimated channels are related to the perfect (true) channels in the proposed CEJOINT method as

$$\hat{\mathbf{H}}_R \approx \mathbf{H}_R \mathbf{\Delta}_R, \quad (5.52)$$

$$\hat{\mathbf{H}}_T \approx \mathbf{\Delta}_T \mathbf{H}_T, \quad (5.53)$$

$$\hat{\mathbf{H}}_S \approx \mathbf{\Delta}_R^{-1} \mathbf{H}_S \mathbf{\Delta}_T^{-1}, \quad (5.54)$$

where $\mathbf{\Delta}_X, X \in \{T, R\}$ is a diagonal matrix holding the scaling ambiguities per column. Similarly, we can write the relation between the estimated MIMO channel matrices and the true ones estimated by Algorithm 5.11 for S-RIS as

$$\hat{\mathbf{G}}_R \approx \mathbf{G}_R \mathbf{\Lambda}, \quad (5.55)$$

$$\hat{\mathbf{G}}_T \approx \mathbf{\Lambda}^{-1} \mathbf{G}_T, \quad (5.56)$$

where $\mathbf{\Lambda}$ is a diagonal matrix holding the scaling ambiguities. However, these ambiguities disappear when reconstructing an estimate of the effective end-to-end channels, i.e.,

$$\hat{\mathbf{H}}_e = \hat{\mathbf{H}}_R \hat{\mathbf{H}}_S \hat{\mathbf{H}}_T, \quad (5.57)$$

$$\hat{\mathbf{G}}_e = \hat{\mathbf{G}}_R \hat{\mathbf{G}}_T. \quad (5.58)$$

Moreover, note that, due to the knowledge of the RIS reflection matrices $\mathbf{\Psi}$, $\mathbf{\Phi}$ for D-RIS, and $\mathbf{\Omega}$ for S-RIS at the receiver, the permutation ambiguities do not exist [AAB21].

Similarly, we can write the ambiguities of the channels estimated by KRF-SEPT and ALS-SEPT as given in (5.52), (5.53), and (5.54) assuming that the conditions in Table 5.1 are satisfied. Note that these scalar ambiguities have no impact on the active and passive RIS reflection design of $\mathbf{\Phi}^D$ and

Table 5.2: Order of the computational Complexity of algorithms.

Algorithm	Computational Complexity
ALS for S-RIS in 5.4.1	$\mathcal{O}(N_{\max\text{-itr}} \cdot 2M_S^3)$
CEJOINT	$\mathcal{O}(N_{\max\text{-itr}} \cdot (M_{S_2}^3 + M_{S_1}^3 + (M_{S_1}M_{S_2})^3))$
KRF-SEPT	$\mathcal{O}(M_{S_1}^2 + M_{S_2}^2 + (M_{S_1}M_{S_2})^3)$
ALS-SEPT	$\mathcal{O}(N_{\max\text{-itr}} \cdot (M_{S_2}^3 + M_{S_1}^3 + (M_{S_1}M_{S_2})^3))$

Ψ^D , which denote the reflection beamforming matrices of RIS-1 and RIS-2 in the data transmission phase, respectively. Let $\hat{\mathbf{H}}_{\text{cascaded}} = \hat{\mathbf{H}}_R \Psi^D \hat{\mathbf{H}}_S \Phi^D \hat{\mathbf{H}}_T \in \mathbb{C}^{M_R \times M_T}$ be the cascaded channel. Substituting the estimated channels given in (5.52), (5.53), and (5.54) into the cascaded channel $\hat{\mathbf{H}}_{\text{cascaded}}$, it can be seen that the diagonal ambiguity matrices Δ_T and Δ_R commute and they cancel with their inverses. Therefore, the performance of the system in terms of the SE does not depend on the knowledge of each channel \mathbf{H}_T , \mathbf{H}_S , and \mathbf{H}_R separately but only on the cascaded channel [ZYZZ21a].

5.5. Simulation Results

In this section, we evaluate the performance of the proposed algorithms using numerical simulation results. For this purpose, we divide this section into two main parts. In the first part, the D-RIS-aided system with joint channel training overhead for both RISs, i.e., CEJOINT, is compared with the S-RIS-aided system. In the second part, we evaluate the performance of the proposed algorithms, i.e., KRF-SEPT and ALS-SEPT, for the D-RIS system with separate channel training overheads for both RISs.

We assume that the entries of all channel matrices \mathbf{H}_T , \mathbf{H}_S , \mathbf{H}_R , \mathbf{G}_T , and \mathbf{G}_R follow a Rayleigh fading distribution, i.e., they are independent and identically distributed complex Gaussian random variables with zero mean and circular symmetry. The results are shown in terms of the NMSE of the effective channels defined as $\text{NMSE} = \mathbb{E}\{\|\mathbf{H}_e - \hat{\mathbf{H}}_e\|_{\mathbb{F}}^2\} / \{\|\mathbf{H}_e\|_{\mathbb{F}}^2\}$, for the D-RIS-aided system, and $\text{NMSE} = \mathbb{E}\{\|\mathbf{G}_e - \hat{\mathbf{G}}_e\|_{\mathbb{F}}^2\} / \{\|\mathbf{G}_e\|_{\mathbb{F}}^2\}$, for the S-RIS-aided system, where the effective channels $\hat{\mathbf{H}}_e$ and $\hat{\mathbf{G}}_e$ are given in (5.57) and (5.58), respectively. The SNR in D-RIS-aided and S-RIS-aided systems are defined as $\text{SNR} = \mathbb{E}\{\|\mathbf{Y} - \mathcal{N}\|_{\mathbb{F}}^2\} / \{\|\mathcal{N}\|_{\mathbb{F}}^2\}$ and $\text{SNR} = \mathbb{E}\{\|\mathbf{X} - \mathcal{E}\|_{\mathbb{F}}^2\} / \{\|\mathcal{E}\|_{\mathbb{F}}^2\}$, where \mathbf{Y} and \mathbf{X} are given in (5.7)/(5.27) and (5.49), respectively. Moreover, assuming that $S \leq M_{S_1}M_{S_2}$ in the joint channel training for D-RIS and $Q \leq M_S$ for S-RIS-aided systems, the training matrices Φ , Ψ , and Ω are updated using a DFT-based approach as: $\Psi = [\mathbf{W}_{M_{S_2}} \otimes \mathbf{1}_{\bar{S}_2}^T]_{[:,1:S]}$, $\Phi = [\mathbf{1}_{\bar{S}_1}^T \otimes \mathbf{W}_{M_{S_1}}]_{[:,1:S]}$, and $\Omega = [\mathbf{W}_{M_S}]_{[:,1:Q]}$, where $\bar{S}_1 = \lceil \frac{S}{M_{S_1}} \rceil$, $\bar{S}_2 = \lceil \frac{S}{M_{S_2}} \rceil$, and \mathbf{W}_a is the normalized $a \times a$ DFT matrix such that $\Omega^H \Omega = \mathbf{I}_Q$, and $\Upsilon^H \Upsilon = \mathbf{I}_S$, where $\Upsilon \stackrel{\text{def}}{=} \Phi \diamond \Psi$. However, assuming that $J \leq M_{S_1}$ and $I \leq M_{S_2}$, the training matrices Φ and Ψ for the D-RIS-aided system with separate channel training are designed by randomly selecting

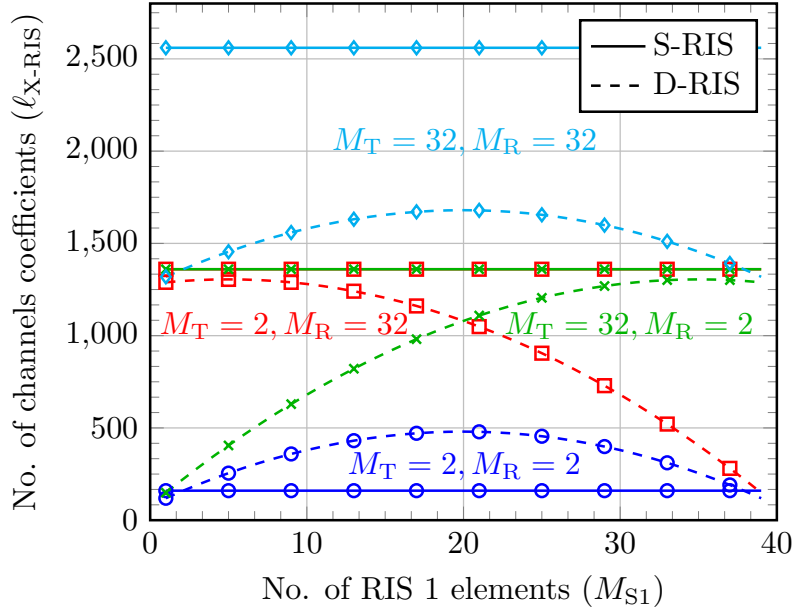


Figure 5.5: Number of channel coefficients (ℓ_{X-RIS}) with $X \in \{S, D\}$ with $M_S = 40$ and $M_{S2} = M_S - M_{S1}$.

J and I rows from normalized M_{S1} -DFT and M_{S2} -DFT matrices, respectively.

5.5.1. D-RIS-aided System versus S-RIS-aided System

Number of Channel Coefficients and Training Overhead

Here, we compare D-RIS-aided systems, i.e., the proposed CEJOINT method, i.e., Algorithm 5.7, to the S-RIS-aided systems, i.e., Algorithm 5.11, according to their required number of channel coefficients that need to be estimated. Let ℓ_{D-RIS} and ℓ_{S-RIS} denote the total number of channel coefficients in the D-RIS and S-RIS communication scenarios, respectively, which are given as

$$\ell_{D-RIS} = M_T M_{S1} + M_{S1} M_{S2} + M_R M_{S2} \quad (5.59)$$

$$\ell_{S-RIS} = M_T M_S + M_R M_S. \quad (5.60)$$

Let us assume that the S-RIS elements M_S are distributed between RIS-1 and RIS-2 in the D-RIS scenario such that $M_S = M_{S1} + M_{S2}$. In Fig. 5.5, we plot the results of (5.59) and (5.60). In Fig. 5.6, we plot the identifiability conditions for S-RIS-aided systems, i.e., Algorithm 5.11, and CEJOINT, i.e., Algorithm 5.7, given in Table 5.1. In Fig. 5.5 and Fig. 5.6, we plot the results for different M_T and M_R values assuming $M_S = 40$. Note that along the x -axis we vary M_{S1} so that $M_{S2} = M_S - M_{S1}$. From Fig. 5.5 and Fig. 5.6, we have the following remarks:

Remark 1: If $M_S \gg \max\{M_R, M_T\}$ and $M_R \approx M_T$, then the S-RIS requires less training overhead compared to the D-RIS, i.e., Algorithm 5.7 (CEJOINT), in most of the M_{S1} and M_{S2} distribution scenarios. This comes from the fact that the number of channel coefficients that the D-RIS needs to estimate, i.e., ℓ_{D-RIS} , is much larger than that of the S-RIS, i.e., ℓ_{S-RIS} .

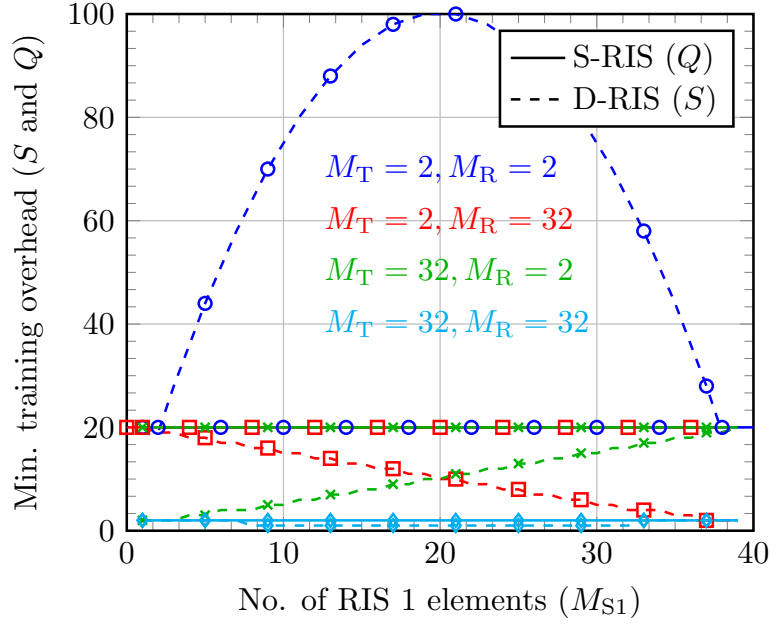


Figure 5.6: Minimum training overhead (S and Q) with $M_S = 40$ and $M_{S2} = M_S - M_{S1}$.

Remark 2: If $M_S \approx \max\{M_R, M_T\}$, then the D-RIS requires less training overhead compared to the S-RIS for the same reason mentioned in Remark 1, i.e., $\ell_{D-RIS} < \ell_{S-RIS}$.

Remark 3: In D-RIS systems, the careful distribution of the M_S elements between RIS-1 and RIS-2 (i.e., M_{S1} and M_{S2}) can reduce the training overhead of Algorithm 5.7 (CEJOINT). From Fig. 5.5, we can note that the best distribution depends on the M_R and M_T values as: if $M_R > M_T$, then it is more beneficial to allocate more elements to RIS-1 than RIS-2, i.e., $M_{S1} > M_{S2}$. This observation is reversed if $M_R < M_T$, i.e., more elements should be allocated to RIS-2 than RIS-1 as $M_{S1} < M_{S2}$.

Channel Estimation Accuracy

Here, we want to compare the performance of the D-RIS-aided systems, i.e., proposed CEJOINT method in Algorithm 5.7, to the S-RIS-aided systems, i.e., Algorithm 5.11, considering the channel estimation accuracy in terms of the NMSE versus SNR. In Fig. 5.7, we fix M_T to 2 and vary the other system parameters. From the left-side figure for $M_R = 8$, we can see that when $M_R = 8$, CEJOINT has a worse NMSE performance compared to the S-RIS, i.e., Algorithm 5.11, especially with the $[M_{S1}, M_{S2}] = [10, 30]$ configuration. This can be explained from Fig. 5.5 and Fig. 5.6 and Remark 1 and Remark 3. Note that in a such system setting, the D-RIS has a larger number of channel coefficients $\ell_{D-RIS} = 560$ compared to the S-RIS, i.e., $\ell_{S-RIS} = 400$. Moreover, as we have highlighted in Remark 3, we can see that the $[M_{S1}, M_{S2}] = [30, 10]$ configuration has a better NMSE performance than $[M_{S1}, M_{S2}] = [10, 30]$, since $M_T < M_R$. On the other hand, when $M_R = 32$, we can see that the D-RIS system has a much better NMSE performance compared to the S-RIS, especially with the $[M_{S1}, M_{S2}] = [30, 10]$ configuration. This can be explained in the same way from Fig. 5.5 and Fig. 5.6 and Remarks 2 and 3. From the figure on the right hand side, we can see that the same observations hold true when we increase M_S from 40 to

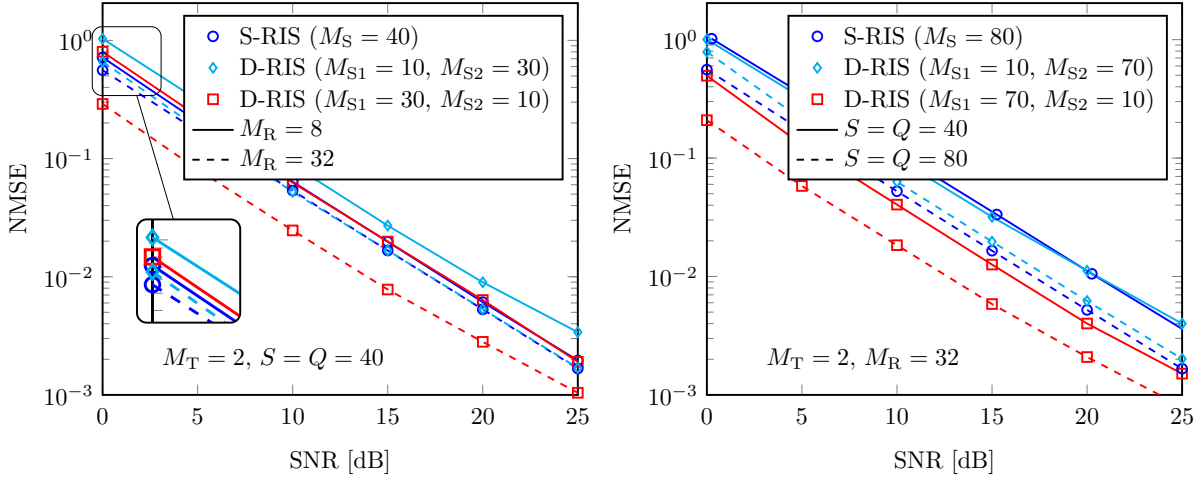


Figure 5.7: NMSE versus SNR comparing D-RIS against the S-RIS systems.

80 or when we increase the training overhead S and Q from 40 to 80.

5.5.2. D-RIS-aided System with Separable Training Overheads

Here, we want to evaluate the performance of the proposed KRF-SEPT and ALS-SEPT channel estimation methods for D-RIS systems in Subsection 5.3.2. In all simulation results, we assume that $M_T = 2$, $M_R = 4$, and $P = M_T$. From the above simulation results, since $M_T < M_R$, we assign more reflective elements to RIS-1 than RIS-2, i.e., $[M_{S1}, M_{S2}] = [30, 20]$.

Compare KRF-SEPT and ALS-SEPT

In Fig. 5.8, we show the NMSE versus the SNR results comparing the KRF-SEPT and ALS-SEPT methods. From Fig. 5.8, we can see that when J and I are selected such that $J = M_{S1}$ and $I = M_{S2}$, both methods achieve the best performance. However, the estimation accuracy of KRF-SEPT degrades significantly if $J < M_{S1}$ and $I < M_{S2}$, since the identifiability constraints for KRF-SEPT are not satisfied. However, ALS-SEPT can still achieve a good performance only after $N_{\max\text{-itr}} = 10$ iterations, at the expense of a higher complexity.

Effect of Initialization on the Performance of ALS-SEPT

Here, we want to check the effect of the initialization on the convergence of the proposed ALS-SEPT method. It is known that the convergence rate of ALS-based methods depend on the initialization, where a good initialization strategy generally leads to a fast convergence rate, and therefore a reduced computational complexity. Motivated by the closed-form KRF-SEPT method, we propose to initialize \mathbf{H}_R and \mathbf{H}_T using (5.45) and (5.46) by using the KRF in (5.38) and (5.39), respectively. In Fig. 5.9 and Fig. 5.10, we show the NMSE versus the SNR results comparing the proposed ALS-SEPT method with KRF-SEPT and the ALS-SEPT with random initialization. As it can be seen, ALS-SEPT initialized with KRF-SEPT does not only enjoy a faster convergence rate as compared to ALS-SEPT with random initial-

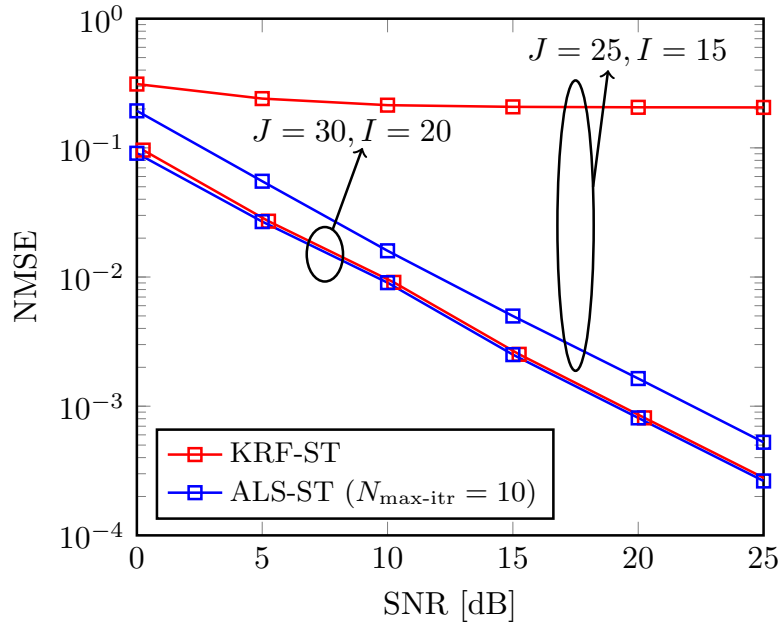


Figure 5.8: NMSE versus SNR for different training overhead ($M_{S1} = 30$ and $M_{S1} = 20$).

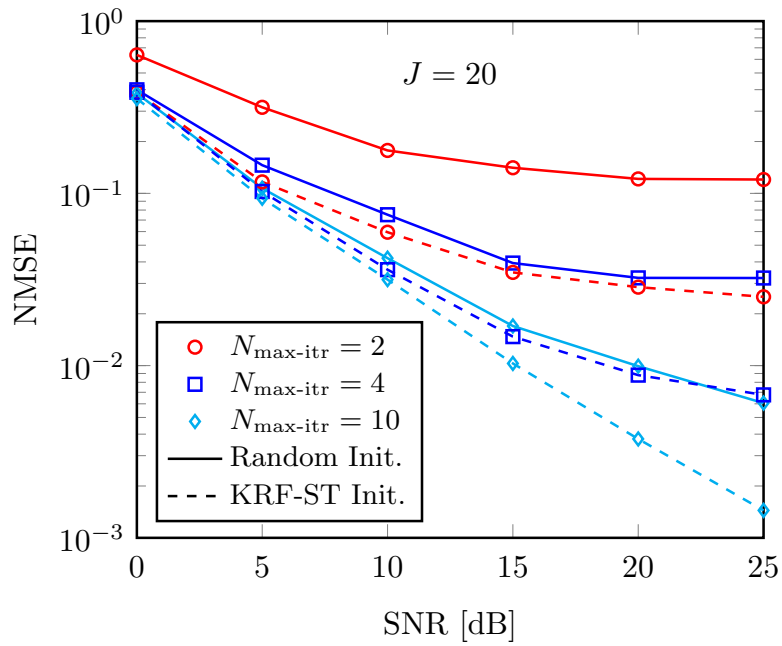


Figure 5.9: Effect of initialization on ALS-SEPT for $I = 15, J = 20$ ($M_{S1} = 30$ and $M_{S1} = 20$).

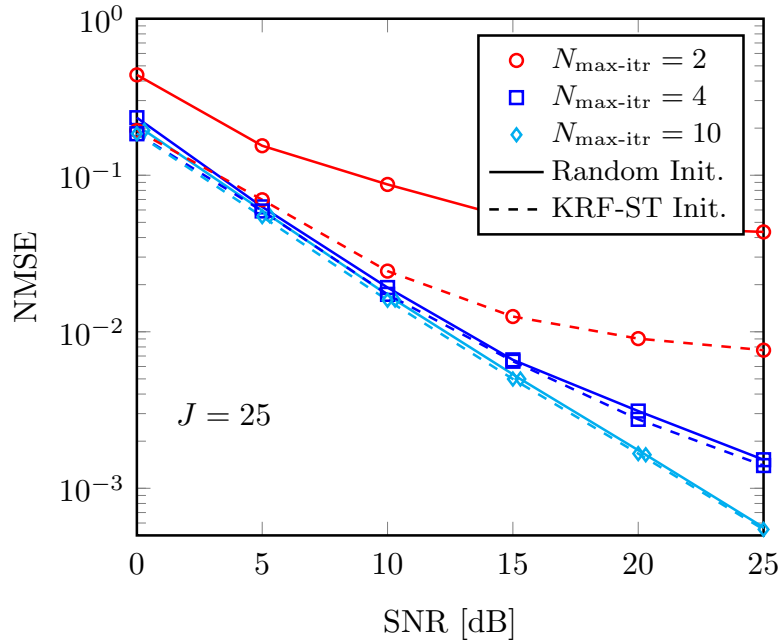


Figure 5.10: Effect of initialization on ALS-SEPT for $I = 15$, $J = 25$ ($M_{S1} = 30$ and $M_{S1} = 20$).

ization, but also achieves a better channel estimation accuracy, especially with a low training overhead as in Fig. 5.9. If we assume that the transmitter in this work represents a BS in a practical scenario, then the required training overhead can be further reduced. The reason is that due to the fixed positions of the BS, RIS-1, and RIS-2, \mathbf{H}_T and \mathbf{H}_S are slowly fading channel matrices, which do not need to be estimated frequently. Since the aforementioned observation is not exploited in this work, the provided results can be seen as the worst-case, i.e., upper-bound, of the channel estimation accuracy.

5.6. Chapter Conclusion

In this chapter, we have considered the channel estimation in the D-RIS MIMO systems. First, we proposed CEJOINT channel estimation method in which both RISs have the same training overhead. We have shown that D-RIS MIMO systems can be used to reduce the training overhead and to improve the channel estimation accuracy compared to S-RIS-aided systems. This comes from the observation that if the RIS elements in the S-RIS system are distributed carefully between the two RISs in the D-RIS system, the number of channel coefficients in the D-RIS system that needs to be estimated reduces significantly compared to the S-RIS system. Therefore, D-RIS systems can be seen as an appealing approach to further increase the coverage, capacity, and efficiency of wireless networks compared to S-RIS systems.

Then, we have proposed a generalized channel estimation technique for D-RIS-aided MIMO systems where the training overhead for each RIS can be selected separately. We have shown that an accurate channel estimation can be obtained with a small training overhead by capitalizing on the nested PARAFAC model of the received training signal. Our results show that we can achieve a good per-

formance with ALS-SEPT channel estimation especially in scenarios with a small training overhead by using the closed-form KRF-SEPT method as an initialization strategy.

Last but not least, both proposed channel estimation methods, i.e., KRF-SEPT and ALS-SEPT, can be extended to multi-user scenarios, where the channel estimation can be performed separately for each user with orthogonal training sequences in the time, frequency, and/or space domains without any multi-user interference. In the case of non-orthogonal pilot sequences, the block PARATUCK model can be used [AAB21], where the associated factor matrices are block matrices. The number of matrix blocks is equal to the number of users. The complexity of the channel estimator would be higher in this case, since we estimate the channels of all users simultaneously.

Conclusions and Future Works

6.1. Conclusions

In this thesis, we have explored advanced algebraic concepts with the objective of finding efficient solutions for future wireless communication systems. The thesis content is organized into two parts, which they can be read independently of each other. In the following, we present the main research findings and contributions of this theses followed by some future works.

The first part of the thesis is devoted to the channel estimation and beamforming design problems in massive MIMO HAD multi-user wireless communication systems. For the channel estimation problem, the objective is to find an efficient channel estimation method for sub-6 GHz systems with a small training overhead, wherein the channel matrix is characterized not only by its high-dimensionality, but also by high-rank nature, unlike the channel characteristics in mmWave-based systems. To overcome these issues while meeting the required objectives, we have turned into multi-dimensional tensor algebra and propose an efficient and general channel estimation method called SALSAs, which is, unlike the state-of-the-art methods, derived by exploiting a *hidden tensor structure* in the frequency-selective measurement matrix without relying on any channel and antenna-array structures or a pre-estimated channel covariance matrix. Specifically, by showing and proving that any MIMO channel matrix can be approximately decomposed into a summation of R factor matrices each having a Kronecker structure, the channel measurement or training matrix can be reshaped into a 3-way tensor admitting a Tucker decomposition. By exploiting such a tensor representation, the SALSAs method estimates the MIMO channel matrix with R sequential steps, wherein a bi-linear ALS method is used to estimate the r th Kronecker factor matrices. Detailed simulation results using a realistic 3GPP CDL channel modeling have shown that the proposed SALSAs method has not only a higher channel estimation accuracy, but also a lower training overhead as compared to the well-known benchmark methods LS and LMMSE.

For the HAD beamforming design problem, the objective was to find an efficient design method for generalized PC beamforming architectures with a low computational complexity. Due to the non-convexity of the formulated problem, we have proposed a two-stage design approach by decoupling the beamforming optimization procedure between the analog and the digital domains. We design the transmit baseband beamforming matrices using the classical ZF method and the receive baseband beamforming matrices using the LMMSE approach. Considering these particular design of the baseband beamforming matrices, we have shown that an approximation of the system's multi-carrier SE upper-bound can be derived and used as the objective function of the transmit analog beamforming matrix optimization. To obtain a solution, we have divided the problem into a series of convex sub-problems that are updated iteratively until a convergence is obtained. For each sub-problem, the objective is to maximize the transmit signal power in each block while minimizing the interference leakage into the other blocks. We have demonstrated numerically that the proposed method has a comparable performance with the benchmark methods in the FC architectures, while it significantly outperforms them in the generalized PC case.

The second part of the thesis is devoted to channel estimation and RIS reflection design problems in RIS-aided mmWave wireless communication systems. First, we have considered a Single RIS (S-RIS)-aided scenarios, wherein a single RIS is deployed between a transmitter and a receiver. For such scenarios, we have proposed two channel estimation methods, TRICE and TenRICE, and an RIS reflection design method FroMax. Specifically, the TRICE method is derived by exploiting the sparse nature (or more accurately, the low-rank nature) of mmWave channels. In every stage, we have shown that the channel parameter estimation can be carried out *via* a multi-dimensional DoA estimation scheme, for which several solutions exist in the literature, such as DFT-beamspace ESPRIT-based methods [ZH17a; ZH17b; ZRH21], CS-based methods [AH21], among many others. To further improve the channel estimation accuracy and to reduce the training overhead, we have proposed the TenRICE method, which not only exploits the low-rank nature of mmWave channels, but also the multi-dimensional tensor structure of the received measurement matrix, which is shown to admit a CP model. Moreover, we derived the Cramér Rao lower Bound (CRB) of the TenRICE method. Detailed simulation results have shown that both TRICE and TenRICE require a smaller training overhead and have a smaller computational complexity as compared to the benchmark methods. Moreover, the TenRICE method outperforms the TRICE method in terms of the channel estimation accuracy, since it takes the advantage of the tensor structure of the measurement matrix.

After that, we have turned our attention to the active beamforming and RIS reflection design problem for SE maximization in S-RIS-aided scenarios. However, the formulated problem is non-convex because of its joint optimization and constant modulus constraints imposed on the RIS reflection matrix. In this case, we have proposed a non-iterative two-stage design method by decoupling the design problem between active beamformers of the transmitter and the receiver and the RIS reflection matrix. Here, we have noted that, for any given RIS reflection matrix, the design of the active beamformers can be simplified to the classical single-user MIMO scenario, for which the optimal design is given by the SVD of the effective channel and water-filling-based power allocation. However, for given active beamformers, the RIS reflection matrix design is still non-convex, due to the constant modulus constraints. To obtain a low-complexity solution, we have proposed heuristic solutions called FroMax-1 and FroMax-2, wherein the problem is formulated as a Frobenius-norm maximization. Specifically, the FroMax-1 strategy designs the RIS reflection matrix by focusing the reflected energy on the dominant eigenmode of the channel, which is only applicable in single-stream cases. To overcome this limitation, the FroMax-2 strategy designs the RIS reflection matrix while taking into account the several dominant eigenmodes selected by the transmitter and, therefore, improves the system's SE. Simulation results have shown that the FroMax-2 strategy has a comparable performance to the benchmark methods but with a significantly lower computational complexity.

Lastly, we have considered the channel estimation problem in D-RIS-aided scenarios, wherein two RISs are deployed between a transmitter and a receiver, which can be used to further extend the coverage of mmWave-based wireless communication systems. For such scenarios, we have proposed two channel estimation methods. In the first method called CEJOINT, we have assumed that the system (or the node controlling the RISs) trains the two RISs jointly, i.e., at the same time. In this case, we have shown that the measurement matrix can be reformulated into a 3-way tensor structure admitting a Tucker2 model, which is then used to derive an ALS-based channel estimation method. On the other hand, in the second method, we have assumed that the system trains each RIS separately, i.e., on different time scales, and have shown that the measurement matrix can be written as a 4-way ten-

sor admitting a nested PARAFAC model, which is then used to derive a closed-form KRF-SEPT and an iterative ALS-SEPT channel estimation technique. Using simulation results, we have shown that by carefully distributing the reflection elements between the two RISs, a more accurate channel estimation with less training overhead can be achieved in the D-RIS-aided scenarios as compared to the S-RIS-aided case. Moreover, we have demonstrated that the ALS-SEPT channel estimation method can achieve a good performance especially in scenarios with a small training overhead by carefully selecting the initialization strategy.

6.2. Future Works

The content of this thesis can be extended in many directions. Some of the possible future directions of our work are listed below.

- The proposed SALSA method in Chapter 2 is derived by approximately decomposing the frequency selective MIMO channel matrix into a summation of R factor matrices each having a Kronecker structure. In the simulation results, we have investigated in detail the effect of the number of factor matrices R on the channel estimation accuracy. It is shown that, for every SNR region/level, there is an optimal R value, where the best channel estimation accuracy is achieved. Therefore, a more profound theoretical understanding of how to select R depending on the SNR could lead to more robust and accurate solutions.

Exploiting the multi-dimensionality of the communication channel in time, frequency, and space becomes more important in massive MIMO systems to obtain a higher channel estimation accuracy with less pilot overhead [Ara+19; Wan+21; Qia+18]. As an example, we can exploit the antenna array structure at the transmitter and receiver as well as the antenna polarization, the channel in the time and frequency domain, which results in a multi-dimensional tensor (up to 7 dimensions). Such high dimensional representation of the signals helps to i) exploit the correlation among multiple dimensions, and ii) mitigate severe error propagation among different dimensions. Therefore, further enhancement of the data structure is an interesting future work to present the signal as a tensor with larger dimensions.

- In Chapter 3, we have proposed a low-complexity HAD beamforming design for frequency-selective downlink multi-user MIMO systems. From the simulation results, we have observed that the Algorithm 3.3 always converges to a point where the cost function stops increasing or decreasing, although not monotonically. Therefore, it is interesting to prove the convergence of the Algorithm 3.3, which can be carried out by interpreting the problem (3.29) from a game theory view-point [NX16; BL17], where the proof can be accomplished by showing its convergence to a Nash equilibrium point.

Moreover, our proposed HAD beamforming algorithm, or any comparable state-of-the-art method for that matter, is only valid for a certain size of the antenna arrays and bandwidth. In other words, extremely large MIMO scenarios cannot be treated only as an enlarged version of a conventional MIMO system. This can be explained by the fact that, in physical reality, if the physical size of the antenna array traveled by the propagation wave exceeds a certain level, then the propagation delay across the array becomes comparable to the symbol period, therefore, the narrowband condition is not anymore true. In this case, the antennas receive different amplitudes and phases

of the same symbol or even distinct symbols at the same sampling time. This effect is already known from radar signal processing and is called spatial-wideband effect [LW10]. In contrast to narrowband signaling with a fixed beam over the antenna manifold, the antenna radiation pattern varies slightly as a function of frequency for the wideband signals. This effect is known as beam squint [Cai+16], which is derived from the spatial-wideband effect. Therefore, pointing the beams toward the same direction for different frequencies is not straightforward anymore [Dup+19b]. Moreover, the reciprocity of the estimated uplink channel is difficult to use to extract the channel parameters in TDD networks [Wan+18]. These effects become more significant with extremely large antenna arrays and much wider bandwidths, for example THz systems. Therefore, it is important to develop an effective HAD beamforming algorithm that considers the beam squint effect and suppresses or exploits it.

- In Chapter 4, we have proposed a beamforming and RIS reflection matrix design method for S-RIS-aided MIMO systems. The numerical results in terms of the SE were illustrated considering perfect and estimated CSI. From these results, we have observed that, in the low SNR region, e.g., below 5 dB, all the simulated methods using the estimated CSI experience a very low SE performance, due to a significant channel estimation error. Therefore, it is interesting to investigate a channel estimation denoising step to improve the channel estimation accuracy in the low SNR region. Moreover, the proposed design method considers only single-user frequency-flat case, wherein the extension to frequency-selective multi-user downlink and uplink scenarios is also an interesting future research topic.
- In Chapter 5, we have shown that the required training overhead in D-RIS-aided systems can be decreased significantly compared to the S-RIS-aided one if the reflecting elements are properly distributed between the RISs. Since the system SE is inversely proportional to the length of the training overhead, we conjecture that there is an optimal distribution of the RIS elements that strike an optimal trade-off between the training overhead and the achievable performance. Further investigation to find a mathematical expression for the exact assignment of the reflecting elements to the RISs is required.

Moreover, we have assumed that the single reflected channels transmitter-to-RIS-1-to-receiver and transmitter-to-RIS-2-to-receiver do not exist. However, in realistic implementations, they could exist and cause significant interference. Therefore, in order to achieve the maximum beamforming gain of D-RIS-aided system, not only the double-reflection link, e.g., transmitter-to-RIS-1-to-RIS-2-to-receiver, must be estimated but also the two single reflection links. The already proposed channel estimation methods in [ZYZ21b; BX22] assumed a multi-user MISO scenario. Therefore, a channel estimation algorithm considering double- and single-reflection links for multi-user D-RIS-aided MIMO system needs to be developed.

Here, a flat-fading narrowband RIS-aided MIMO system is considered. However, in the future, we are interested in wideband multi-carrier RIS-aided MIMO communication systems. It is already shown in [Mao+22] that the amplitude and the phases of the reflecting elements in the RISs are not completely independent in the wideband scenarios. Moreover, in [BS21], it is also shown that depending on the distance of the adjacent reflecting elements to each other, they might experience spatial correlation. Therefore, a channel estimation algorithm for wideband multi-carrier RIS-aided MIMO systems should be developed. Furthermore, beamforming and RIS reflection matrix

6. *Conclusions and Future Works*

design methods for D-RIS-aided MIMO systems were not considered in this thesis, which are an important topic for future research.

Appendix A

Appendices

A.1. Proof of (1.26)

Using (1.11), we can show that

$$\mathbf{A}_1 \Omega_1 \diamond \mathbf{A}_2 \Omega_2 = (\mathbf{A}_1 \otimes \mathbf{A}_2)(\Omega_1 \diamond \Omega_2) = (\mathbf{A}_1 \otimes \mathbf{A}_2). \quad (\text{A.1})$$

where

$$\Omega_1 = (\mathbf{I}_{L_1} \otimes \mathbf{1}_{L_2}^\top) = \begin{bmatrix} \mathbf{1}_{L_2}^\top & \mathbf{0}_{L_2}^\top & \cdots & \mathbf{0}_{L_2}^\top \\ \mathbf{0}_{L_2}^\top & \mathbf{1}_{L_2}^\top & \cdots & \mathbf{0}_{L_2}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{L_2}^\top & \mathbf{0}_{L_2}^\top & \cdots & \mathbf{1}_{L_2}^\top \end{bmatrix} \in \mathbb{C}^{L_1 \times L_1 L_2}, \quad (\text{A.2})$$

and

$$\Omega_2 = (\mathbf{1}_{L_1}^\top \otimes \mathbf{I}_{L_2}) = \begin{bmatrix} \mathbf{I}_{L_2} & \mathbf{I}_{L_2} & \cdots & \mathbf{I}_{L_2} \end{bmatrix} \in \mathbb{C}^{L_2 \times L_1 L_2}, \quad (\text{A.3})$$

where $\mathbf{0}_Q$ is all zeros column vector of dimension $Q \times 1$. First, we show that the right hand side of (A.1) is equal to its left hand side as

$$\begin{aligned} \mathbf{A}_1 \Omega_1 \diamond \mathbf{A}_2 \Omega_2 &= \begin{bmatrix} [\mathbf{A}_1]_{[1,1]} \mathbf{1}_{L_2}^\top & [\mathbf{A}_1]_{[1,2]} \mathbf{1}_{L_2}^\top & \cdots & [\mathbf{A}_1]_{[1,L_1]} \mathbf{1}_{L_2}^\top \\ [\mathbf{A}_1]_{[2,1]} \mathbf{1}_{L_2}^\top & [\mathbf{A}_1]_{[2,2]} \mathbf{1}_{L_2}^\top & \cdots & [\mathbf{A}_1]_{[2,L_1]} \mathbf{1}_{L_2}^\top \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{A}_1]_{[J_1,1]} \mathbf{1}_{L_2}^\top & [\mathbf{A}_1]_{[J_1,2]} \mathbf{1}_{L_2}^\top & \cdots & [\mathbf{A}_1]_{[J_1,L_1]} \mathbf{1}_{L_2}^\top \end{bmatrix} \diamond \begin{bmatrix} \mathbf{A}_2 & \mathbf{A}_2 & \cdots & \mathbf{A}_2 \end{bmatrix} \\ &= \begin{bmatrix} [\mathbf{A}_1]_{[:,1]} \otimes [\mathbf{A}_2]_{[:,1]} & \cdots & [\mathbf{A}_1]_{[:,1]} \otimes [\mathbf{A}_2]_{[:,L_2]} & [\mathbf{A}_1]_{[:,2]} \otimes [\mathbf{A}_2]_{[:,1]} & \cdots & [\mathbf{A}_1]_{[:,L_1]} \otimes [\mathbf{A}_2]_{[:,L_2]} \end{bmatrix} \\ &= \begin{bmatrix} [\mathbf{A}_1]_{[1,1]} \mathbf{A}_2 & [\mathbf{A}_1]_{[1,2]} \mathbf{A}_2 & \cdots & [\mathbf{A}_1]_{[1,L_1]} \mathbf{A}_2 \\ [\mathbf{A}_1]_{[2,1]} \mathbf{A}_2 & [\mathbf{A}_1]_{[2,2]} \mathbf{A}_2 & \cdots & [\mathbf{A}_1]_{[2,L_1]} \mathbf{A}_2 \\ \vdots & \ddots & \vdots & \\ [\mathbf{A}_1]_{[J_1,1]} \mathbf{A}_2 & [\mathbf{A}_1]_{[J_1,2]} \mathbf{A}_2 & \cdots & [\mathbf{A}_1]_{[J_1,L_1]} \mathbf{A}_2 \end{bmatrix} \\ &= \mathbf{A}_1 \otimes \mathbf{A}_2. \end{aligned} \quad (\text{A.4})$$

In the following, we want to prove that for Ω_1 and Ω_2 given in (A.2) and (A.3), respectively, we have

$$\Omega_1 \diamond \Omega_2 = (\mathbf{I}_{L_1} \otimes \mathbf{1}_{L_2}^\top) \diamond (\mathbf{1}_{L_1}^\top \otimes \mathbf{I}_{L_2}) = \mathbf{I}_{L_1 L_2}, \quad (\text{A.5})$$

which can be directly seen by taking the Khatri-Rao product of (A.2) and (A.3). In [MN17; HS81], it was proven that for any given $\mathbf{A} \in \mathbb{C}^{M \times R}$, $\mathbf{B} \in \mathbb{C}^{N \times S}$, $\mathbf{C} \in \mathbb{C}^{P \times R}$ and $\mathbf{D} \in \mathbb{C}^{Q \times S}$ matrices with appropriate

dimensions, the property

$$(\mathbf{A} \otimes \mathbf{B}) \diamond (\mathbf{C} \otimes \mathbf{D}) = \mathbf{P}[(\mathbf{A} \diamond \mathbf{C}) \otimes (\mathbf{B} \diamond \mathbf{D})] \quad (\text{A.6})$$

holds true, where $\mathbf{P} \in \mathbb{C}^{MNPQ \times MNPQ}$ is a permutation matrix that depends only on the dimensions of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and \mathbf{D} and not on their elements. The proof was obtained due to fact that the Kronecker product is associative but not commutative, and therefore

$$([\mathbf{A}]_{[:,r]} \otimes [\mathbf{B}]_{[:,s]}) = \bar{\mathbf{P}}([\mathbf{B}]_{[:,s]} \otimes [\mathbf{A}]_{[:,r]}), \quad (\text{A.7})$$

where $[\mathbf{A}]_{[:,r]}$ and $[\mathbf{B}]_{[:,s]}$ are the r th and the s th columns of \mathbf{A} and \mathbf{B} , respectively, and $\bar{\mathbf{P}}$ is the permutation matrix. However, for our defined matrices, wherein $\mathbf{A} = \mathbf{I}_{L_1}$, $\mathbf{B} = \mathbf{1}_{L_2}^\top$, $\mathbf{C} = \mathbf{1}_{L_1}^\top$ and $\mathbf{D} = \mathbf{I}_{L_2}$, the permutation matrix \mathbf{P} is in fact an identity matrix, i.e., $\mathbf{P} = \mathbf{I}_{L_1 L_2}$, with appropriate dimension, since

$$([\mathbf{I}_{L_1}]_{[:,i]} \otimes [\mathbf{1}_{L_2}^\top]_j) = ([\mathbf{1}_{L_2}^\top]_j \otimes [\mathbf{I}_{L_1}]_{[:,i]}) = [\mathbf{I}_{L_1}]_{[:,i]}, \quad (\text{A.8})$$

where $[\mathbf{I}_{L_1}]_{[:,i]}$ is the i th column of \mathbf{I}_{L_1} and $[\mathbf{1}_{L_2}^\top]_j$ is the j th entry of $\mathbf{1}_{L_2}^\top$, where the latter is always equal to 1, i.e., $[\mathbf{1}_{L_2}^\top]_j = 1, \forall j$. From the above, we have that

$$\boldsymbol{\Omega}_1 \diamond \boldsymbol{\Omega}_2 = (\mathbf{I}_{L_1} \otimes \mathbf{1}_{L_2}^\top) \diamond (\mathbf{1}_{L_1}^\top \otimes \mathbf{I}_{L_2}) \quad (\text{A.9})$$

$$= (\mathbf{I}_{L_1} \diamond \mathbf{1}_{L_1}^\top) \otimes (\mathbf{1}_{L_2}^\top \diamond \mathbf{I}_{L_2}) \quad (\text{A.10})$$

$$= \mathbf{I}_{L_1} \otimes \mathbf{I}_{L_2} \quad (\text{A.11})$$

$$= \mathbf{I}_{L_1 L_2}, \quad (\text{A.12})$$

where (A.10) is obtained from (A.6) and (A.8), while (A.11) is obtained since $\mathbf{I}_{L_1} \diamond \mathbf{1}_{L_1}^\top = \mathbf{I}_{L_1}$ and $\mathbf{1}_{L_2}^\top \diamond \mathbf{I}_{L_2} = \mathbf{I}_{L_2}$, which follow directly from the Khatri-Rao Product definition. This completes the proof.

A.2. Jensen's Inequality

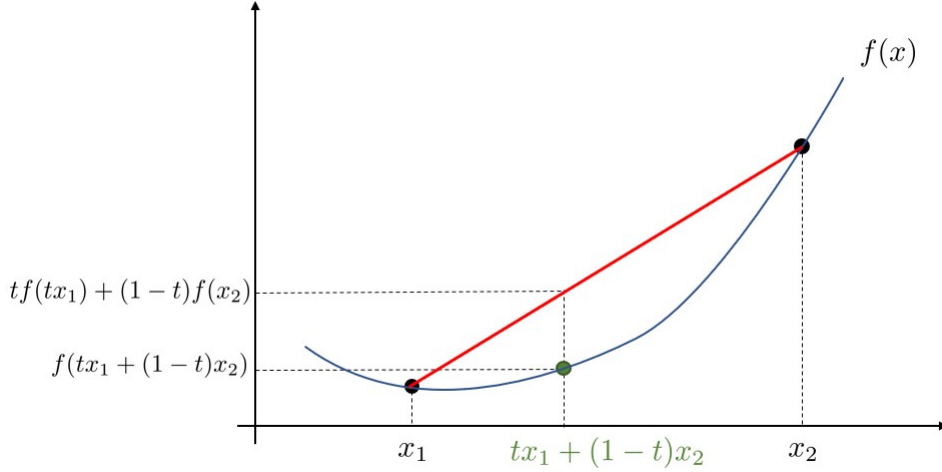


Figure A.1: Using Jensen's inequality for a secant line of a convex function

The danish mathematician Johan Jensen showed the relation between the value of a convex function of an integral to the integral of the convex function, which is also known as Jensen's inequality. For a better understanding, consider the convex function $f(x)$ as illustrated in Fig. A.1. From Fig. A.1, it can be seen that for any $t \in [0, 1]$, the secant line of a convex function lies above it, i.e.,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2). \quad (\text{A.13})$$

In a finite form, Jensen's inequality for any real convex function $f(\cdot)$ can be expressed as [Rud87; Kra99]

$$f\left(\frac{\sum_{i=1}^I a_i x_i}{\sum_{i=1}^I a_i}\right) \leq \frac{\sum_{i=1}^I a_i f(x_i)}{\sum_{i=1}^I a_i}, \quad (\text{A.14})$$

where x_i are numbers in the function $f(\cdot)$ domain and a_i are positive weights with $i \in \{1, \dots, I\}$. For any real concave function $f(\cdot)$, the inequality is reversed, i.e.,

$$f\left(\frac{\sum_{i=1}^I a_i x_i}{\sum_{i=1}^I a_i}\right) \geq \frac{\sum_{i=1}^I a_i f(x_i)}{\sum_{i=1}^I a_i}. \quad (\text{A.15})$$

Note that the equality holds in either of the following two scenarios,

(1) $x_1 = x_2 = \dots = x_I$

(2) $f(\cdot)$ is a linear function in the domain containing $x_i, \forall i$.

Particularly, if a_i s are all equal, (A.14) and (A.15) can be simplified as

$$f\left(\frac{\sum_{i=1}^I x_i}{I}\right) \leq \frac{\sum_{i=1}^I f(x_i)}{I}, \quad (\text{A.16})$$

$$f\left(\frac{\sum_{i=1}^I x_i}{I}\right) \geq \frac{\sum_{i=1}^I f(x_i)}{I}. \quad (\text{A.17})$$

The Jensen's inequality is very general and appears in many forms depending on the context such as the log sum inequality in information theory [CS04]. Knowing that $\log(\cdot)$ is a concave function, then by substituting $f(\cdot) = \log(\cdot)$ in (A.17), we have

$$\log_2\left(\frac{\sum_{i=1}^I x_i}{I}\right) \geq \frac{\sum_{i=1}^I \log_2(x_i)}{I}. \quad (\text{A.18})$$

A.3. Design of Analog Decoding Matrix in Chapter 3

The proposed iterative algorithm in [Ard+18] is outlined in this appendix, which is used in Chapter 3 to obtain the analog decoding matrices \mathbf{W}_u^{RF} at the u th UE. It was shown that (3.15) can be written equivalently as

$$\max_{\mathbf{w}_c^{\text{RF}} \in \mathcal{W}} \log_2 |\mathbf{E}_c| + \log_2 |\mathbf{w}_c^{\text{H}} \mathbf{G}_c \mathbf{w}_c|, \quad (\text{A.19})$$

where $\mathbf{w}_c = [\mathbf{W}_u^{\text{RF}}]_{[:,c]}$, with $c \in \{1, \dots, N_{\text{UE}}\}$, and \mathbf{E}_c and \mathbf{G}_c are defined as

$$\mathbf{E}_c = (\bar{\mathbf{W}}_c)^{\text{H}} \mathbf{D}_u \bar{\mathbf{W}}_c \in \mathbb{C}^{(N_{\text{UE}}-1) \times (N_{\text{UE}}-1)} \quad (\text{A.20})$$

$$\mathbf{G}_c = \mathbf{D}_u - \mathbf{D}_u \bar{\mathbf{W}}_c \mathbf{E}_c (\bar{\mathbf{W}}_c)^{\text{H}} \mathbf{D}_u \in \mathbb{C}^{M_{\text{UE}} \times M_{\text{UE}}}, \quad (\text{A.21})$$

in which \mathbf{D}_u is defines as $\mathbf{D}_u = \sum_{k=1}^K \mathbf{H}_{u,k} \mathbf{H}_{u,k}^{\text{H}} \in \mathbb{C}^{M_{\text{UE}} \times M_{\text{UE}}}$ and $\bar{\mathbf{W}}_c$ is the sub-matrix of \mathbf{W}_u with the c th column removed, i.e.,

$$\bar{\mathbf{W}}_c = [\mathbf{w}_1, \dots, \mathbf{w}_{c-1}, \mathbf{w}_{c+1}, \dots, \mathbf{w}_{N_{\text{UE}}}] \in \mathbb{C}^{M_{\text{UE}} \times (N_{\text{UE}}-1)}. \quad (\text{A.22})$$

From (A.20), it can be seen that the left hand side of (A.19) does not depend on \mathbf{w}_c . Therefore, (A.19) reduces to

$$\max_{\mathbf{w}_c \in \mathcal{W}} \log_2 |\mathbf{w}_c^{\text{H}} \mathbf{G}_c \mathbf{w}_c|. \quad (\text{A.23})$$

Then, the objective term of problem (A.23) is exploited such that the impact of the i th element of vector \mathbf{w}_c , i.e., $[\mathbf{w}_c]_{[i]}$ is explicitly exposed as

$$\begin{aligned} \mathbf{w}_c^{\text{H}} \mathbf{G}_c \mathbf{w}_c &= \sum_{m=1, m \neq i}^{N_{\text{UE}}} \sum_{n=1, n \neq i}^{N_{\text{UE}}} ([\mathbf{w}_c]_{[m]})^* [\mathbf{G}_c]_{[m,n]} [\mathbf{w}_c]_{[n]} \\ &\quad + 2\text{Re} \left(([\mathbf{w}_c]_{[i]})^* \sum_{n=1, n \neq i}^{N_{\text{UE}}} [\mathbf{G}_c]_{[i,n]} [\mathbf{w}_c]_{[n]} \right) \\ &\quad + ([\mathbf{w}_c]_{[i]})^* [\mathbf{G}_c]_{[i,i]} [\mathbf{w}_c]_{[i]}. \end{aligned} \quad (\text{A.24})$$

It can be seen that the i th element of \mathbf{w}_c only appears in the last two terms of the right hand side part of (A.24). Then, the optimal value of $[\mathbf{w}_c]_{[i]}$ can be obtained as

$$[\mathbf{w}_c]_{[i]} = \angle \left(\sum_{n=1}^{N_{\text{UE}}} [\mathbf{G}_c]_{[i,n]} [\mathbf{w}_c]_{[n]} \right). \quad (\text{A.25})$$

The summary of the proposed iterative algorithm in [Ard+18] is given in Algorithm 1.12.

Algorithm 1.12 Proposed algorithm in [Ard+18].

Input: $\mathbf{H}_{u,k}, \forall u, \forall k$.

```

1: for  $u = 1$  to  $U$  do
2:   Compute  $\mathbf{D}_u = \sum_{k=1}^K \mathbf{H}_{u,k} \mathbf{H}_{u,k}^H$ , initialize  $\mathbf{W}_u^{\text{RF}}[0] \in \mathcal{W}$  randomly, and set initialization index
    $p = 1$ 
3:   while not converged do
4:     for  $c$ th column of  $\mathbf{W}_u^{\text{RF}}[p]$ , i.e.,  $\mathbf{w}_{c,c}[p] \in \{1, \dots, N_{\text{UE}}\}$  do
5:       Compute  $\mathbf{E}_c[p]$  and  $\mathbf{G}_c[p]$  as given in (A.20) and (A.21), respectively
6:       for  $i$ th element of  $\mathbf{w}_c[p]$ ,  $i \in \{1, \dots, M_{\text{UE}}\}$  do
7:         Update  $[\mathbf{w}_c]_{[i]}[p]$  as described in (A.25)
8:       end for
9:     end for
10:     $p = p + 1$ 
11:   end while
12: end for
Output:  $\mathbf{W}_u^{\text{RF}}[p]$ 

```

A.4. Derivation of Upper-bound in (3.18)

This appendix contains the derivation of the upper-bound given in (3.18) in Subsection 3.3.2. The difficulty by designing the analog precoding matrix \mathbf{F}^{RF} is that it has to be common for all UEs and all sub-carriers. In Subsection 3.3.2, we assume that the baseband beamforming matrices $\mathbf{F}_k^{\text{BB}}, \forall k$, are calculated using the well-known ZF method, which leads to the fact that

$$(\mathbf{F}_k^{\text{BB}})^{\text{H}} \bar{\mathbf{H}}_k^{\text{H}} \bar{\mathbf{H}}_k \mathbf{F}_k^{\text{BB}} = \mathbf{I}_{UN_s}. \quad (\text{A.26})$$

From (A.26) and by inserting the digital precoding matrix $\bar{\mathbf{F}}_k^{\text{BB}}$ from (3.6) into the SE in (3.17), the total SE of the system can be rewritten as

$$\begin{aligned} \text{SE} &= \sum_{k=1}^K \log_2 |\mathbf{I}_{UN_s} + \frac{1}{\sigma_n^2} (\mathbf{F}_k^{\text{BB}})^{\text{H}} \bar{\mathbf{H}}_k^{\text{H}} \bar{\mathbf{H}}_k \mathbf{F}_k^{\text{BB}}| \\ &= \sum_{k=1}^K \log_2 |\mathbf{I}_{UN_s} + \frac{1}{\sigma_n^2} \frac{P_k}{\|\mathbf{F}^{\text{RF}} \bar{\mathbf{F}}_k^{\text{BB}}\|_{\text{F}}^2} (\bar{\mathbf{F}}_k^{\text{BB}})^{\text{H}} \bar{\mathbf{H}}_k^{\text{H}} \bar{\mathbf{H}}_k \bar{\mathbf{F}}_k^{\text{BB}}| \\ &= UN_s \sum_{k=1}^K \log_2 \left(1 + \frac{P_k}{\sigma_n^2 \|\mathbf{F}^{\text{RF}} \bar{\mathbf{F}}_k^{\text{BB}}\|_{\text{F}}^2} \right). \end{aligned} \quad (\text{A.27})$$

Next, we define the term in the denominator of (A.27) as $\zeta_k \stackrel{\text{def}}{=} \|\mathbf{F}^{\text{RF}} \bar{\mathbf{F}}_k^{\text{BB}}\|_{\text{F}}^2$. Then, we rewrite $\bar{\mathbf{F}}_k^{\text{BB}}$ in the ZF sense as $\bar{\mathbf{F}}_k^{\text{BB}} = \bar{\mathbf{H}}_k^{\text{H}} (\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^{\text{H}})^{-1}$ and substitute it in ζ_k as

$$\begin{aligned} \zeta_k &= \text{trace}[(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^{\text{H}})^{-1} \bar{\mathbf{H}}_k (\mathbf{F}^{\text{RF}})^{\text{H}} \mathbf{F}^{\text{RF}} \bar{\mathbf{H}}_k^{\text{H}} (\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^{\text{H}})^{-1}] \\ &\approx \text{trace}[(\bar{\mathbf{H}}_k \bar{\mathbf{H}}_k^{\text{H}})^{-1}], \end{aligned} \quad (\text{A.28})$$

where the approximation follows from the observation that $(\mathbf{F}^{\text{RF}})^{\text{H}} \mathbf{F}^{\text{RF}}$ can be approximated as $(\mathbf{F}^{\text{RF}})^{\text{H}} \mathbf{F}^{\text{RF}} \approx \mathbf{I}_{N_{\text{BS}}}$, especially when M_{BS} is large [SY16]. From (A.28), we can approximate (A.27) as

$$\begin{aligned} \text{SE} &\approx UN_s \sum_{k=1}^K \log_2 \left(1 + \frac{P_k}{\sigma_n^2 \text{trace}[(\hat{\mathbf{H}}_k \mathbf{F}^{\text{RF}} (\mathbf{F}^{\text{RF}})^{\text{H}} \hat{\mathbf{H}}_k^{\text{H}})^{-1}]} \right), \\ &\stackrel{\text{(a)}}{\leq} UN_s \log_2 \left(1 + \frac{P_k}{\sigma_n^2} \sum_{k=1}^K \frac{1}{\text{trace}[(\hat{\mathbf{H}}_k \mathbf{F}^{\text{RF}} (\mathbf{F}^{\text{RF}})^{\text{H}} \hat{\mathbf{H}}_k^{\text{H}})^{-1}]} \right) \\ &\stackrel{\text{(b)}}{\leq} UN_s \log_2 \left(1 + \frac{P_k}{\sigma_n^2 (UN_{\text{UE}})^2} \sum_{k=1}^K \text{trace}[\hat{\mathbf{H}}_k \mathbf{F}^{\text{RF}} (\mathbf{F}^{\text{RF}})^{\text{H}} \hat{\mathbf{H}}_k^{\text{H}}] \right) \\ &= UN_s \log_2 \left(1 + \frac{P_k}{\sigma_n^2 (UN_{\text{UE}})^2} \text{trace}[(\mathbf{F}^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{F}^{\text{RF}}] \right), \end{aligned} \quad (\text{A.29})$$

where $\hat{\mathbf{H}}_k = [\mathbf{H}_{1,k}^{\text{T}}, \dots, \mathbf{H}_{U,k}^{\text{T}}]^{\text{T}} \in \mathbb{C}^{UN_{\text{UE}} \times M_{\text{BS}}}$, $\mathbf{D} = \sum_{k=1}^K \hat{\mathbf{H}}_k^{\text{H}} \hat{\mathbf{H}}_k$, (a) follows from Jensen's inequality as described in Appendix A.2, and (b) follows from the fact that for any positive definite matrix $\mathbf{X} \in \mathbb{C}^{a \times a}$, we have $\text{trace}[\mathbf{X}^{-1}] \geq a^2 / \text{trace}[\mathbf{X}]$ [Yan00].

A.5. Proof of Proposition 3.3.1

This Appendix contains the proof of Proposition 3.3.1 in Chapter 3. Define the Lagrangian multiplier of (3.30) as

$$\begin{aligned}
\mathcal{L}(\mathbf{f}_c^{\text{RF}}, \lambda_c) &= - \sum_{c=1}^{N_{\text{BS}}} (\mathbf{d}_c^{\text{H}} \mathbf{f}_c^{\text{RF}} - \sum_{q=1, q \neq c}^{N_{\text{BS}}} |\mathbf{d}_q^{\text{H}} \mathbf{f}_c^{\text{RF}}|^2) + \lambda_c (\|\mathbf{f}_c^{\text{RF}}\|^2 - 1) \\
&= - \sum_{c=1}^{N_{\text{BS}}} ((\mathbf{f}_c^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{f}_c^{\text{RF}} - \sum_{q=1, q \neq c}^{N_{\text{BS}}} |(\mathbf{f}_q^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{f}_c^{\text{RF}}|^2) + \lambda_c (\|\mathbf{f}_c^{\text{RF}}\|^2 - 1) \\
&= - \sum_{c=1}^{N_{\text{BS}}} ((\mathbf{f}_c^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{f}_c^{\text{RF}} - \sum_{q=1, q \neq c}^{N_{\text{BS}}} (\mathbf{f}_c^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{f}_q^{\text{RF}} (\mathbf{f}_q^{\text{RF}})^{\text{H}} \mathbf{D} \mathbf{f}_c^{\text{RF}}) + \lambda_c ((\mathbf{f}_c^{\text{RF}})^{\text{H}} \mathbf{f}_c^{\text{RF}} - 1) \quad (\text{A.30})
\end{aligned}$$

where λ_c denotes the Lagrangian multiplier associated with problem (3.30). Taking the (complex) gradient of the Lagrangian $\mathcal{L}(\mathbf{f}_c^{\text{RF}}, \lambda_c)$ with respect to $(\mathbf{f}_c^{\text{RF}})^*$ as

$$\begin{aligned}
\nabla_{(\mathbf{f}_c^{\text{RF}})^*} \mathcal{L}(\mathbf{f}_c^{\text{RF}}, \lambda_c) &= - \mathbf{D} \mathbf{f}_c^{\text{RF}} + \sum_{q=1, q \neq c}^{N_{\text{BS}}} \mathbf{D} \mathbf{f}_q^{\text{RF}} (\mathbf{f}_q^{\text{RF}})^{\text{H}} \mathbf{D}^{\text{H}} \mathbf{f}_c^{\text{RF}} + \lambda_c^* \mathbf{f}_c^{\text{RF}} \\
&= - \mathbf{d}_c + \sum_{q=1, q \neq c}^{N_{\text{BS}}} \mathbf{d}_q \mathbf{d}_q^{\text{H}} \mathbf{f}_c^{\text{RF}} + \lambda_c^* \mathbf{f}_c^{\text{RF}}, \quad (\text{A.31})
\end{aligned}$$

where $\nabla_{(\mathbf{f}_c^{\text{RF}})^*}$ denotes the conjugate Wirtinger gradient. By setting (A.31) equal to zero and solving for \mathbf{f}_c^{RF} , we get

$$\mathbf{f}_{c, \text{opt}}^{\text{RF}} = \left(\sum_{q=1, q \neq c}^{N_{\text{BS}}} \mathbf{d}_q \mathbf{d}_q^{\text{H}} + \lambda_c^* \mathbf{I}_{M_{\text{BS}}} \right)^{-1} \mathbf{d}_c. \quad (\text{A.32})$$

We can simplify update \mathbf{f}_c^{RF} considering that $\|\mathbf{f}_c^{\text{RF}}\|^2 = 1$ as

$$\mathbf{f}_{c, \text{opt}}^{\text{RF}} = \mu_c \left(\sum_{q=1, q \neq c}^{N_{\text{BS}}} \mathbf{d}_q \mathbf{d}_q^{\text{H}} + \mathbf{I}_{M_{\text{BS}}} \right)^{-1} \mathbf{d}_c, \quad (\text{A.33})$$

where $\mu_c = 1 / \left\| \left(\sum_{q=1, q \neq c}^{N_{\text{BS}}} \mathbf{d}_q \mathbf{d}_q^{\text{H}} + \mathbf{I}_{M_{\text{BS}}} \right)^{-1} \mathbf{d}_c \right\|$.

A.6. Derivation of CRB to 4-way CP Tensor in (4.50)

The following is a straightforward extension of the Cramér Rao lower Bound (CRB) derivation in [Zho+17] for a 3-way tensor to our 4-way tensor. Due to brevity, we only provide here the final results. For more details, we advise the readers to refer to [Zho+17].

Let $\mathcal{Y} \in \mathbb{C}^{N_T \times T_T \times T_S^h \times T_S^v}$ be the 4-way tensor given by (4.50), where its four factor matrices are given by (4.55) - (4.58) while $\tilde{\psi}_R \stackrel{\text{def}}{=} [\psi_{R,1}, \dots, \psi_{R,L}]^\top$, $\tilde{\psi}_T \stackrel{\text{def}}{=} [\psi_{T,1}, \dots, \psi_{T,L}]^\top$, $\tilde{\mu}^h \stackrel{\text{def}}{=} [\mu_1^h, \dots, \mu_L^h]^\top$, and $\tilde{\mu}^v \stackrel{\text{def}}{=} [\mu_1^v, \dots, \mu_L^v]^\top$ are their unknown spatial frequency vectors, respectively. Let $\boldsymbol{\eta} \stackrel{\text{def}}{=} [\tilde{\psi}_R^\top \tilde{\psi}_T^\top (\tilde{\mu}^h)^\top (\tilde{\mu}^v)^\top \tilde{\alpha}^\top]^\top$. Let $L(\boldsymbol{\eta})$ define the log-likelihood function of $\boldsymbol{\eta}$, which is given as [Zho+17]

$$L(\boldsymbol{\eta}) = f(\mathcal{Y}; \mathbf{A}_R, \mathbf{A}_T, \mathbf{B}^h, \mathbf{B}^v), \quad (\text{A.34})$$

where

$$\begin{aligned} f(\mathcal{Y}; \mathbf{A}_R, \mathbf{A}_T, \mathbf{B}^h, \mathbf{B}^v) &= c - \frac{1}{\sigma^2} \|\llbracket \mathcal{Y} \rrbracket_{(1)}^\top - \mathbf{C}_R \mathbf{A}_R^\top\|_F^2 \\ &= c - \frac{1}{\sigma^2} \|\llbracket \mathcal{Y} \rrbracket_{(2)}^\top - \mathbf{C}_T \mathbf{A}_T^\top\|_F^2 \\ &= c - \frac{1}{\sigma^2} \|\llbracket \mathcal{Y} \rrbracket_{(3)}^\top - \mathbf{C}^h (\mathbf{B}^h)^\top\|_F^2 \\ &= c - \frac{1}{\sigma^2} \|\llbracket \mathcal{Y} \rrbracket_{(4)}^\top - \mathbf{C}^v (\mathbf{B}^v)^\top\|_F^2 \end{aligned}$$

where $c = -N_R T_T T_S^h T_S^v \ln(\pi \sigma^2)$ and

$$\mathbf{C}_R = (\mathbf{B}^v \diamond \mathbf{B}^h \diamond \mathbf{A}_T) \in \mathbb{C}^{T_T T_S^h T_S^v \times L} \quad (\text{A.35})$$

$$\mathbf{C}_T = (\mathbf{B}^v \diamond \mathbf{B}^h \diamond \mathbf{A}_R) \in \mathbb{C}^{N_R T_S^h T_S^v \times L} \quad (\text{A.36})$$

$$\mathbf{C}^h = (\mathbf{B}^v \diamond \mathbf{A}_T \diamond \mathbf{A}_R) \in \mathbb{C}^{N_R T_T T_S^v \times L} \quad (\text{A.37})$$

$$\mathbf{C}^v = (\mathbf{B}^h \diamond \mathbf{A}_T \diamond \mathbf{A}_R) \in \mathbb{C}^{N_R T_T T_S^h \times L}. \quad (\text{A.38})$$

Assuming that \mathcal{N} entries are zero mean circular symmetric Gaussian random variables with variance σ^2 , i.e.,

$$\mathbb{E}\{[\mathcal{N}]_{[d_1, d_2, d_3, d_4]} [\mathcal{N}]_{[d_1, d_2, d_3, d_4]}^*\} = \begin{cases} \sigma^2 & \text{if } d_i = k_i, \forall i, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.39})$$

Then, the CRB for parameters $\boldsymbol{\eta}$ can be calculated as

$$\text{CRB}(\boldsymbol{\eta}) = \mathcal{D}(\boldsymbol{\eta})^{-1}, \quad (\text{A.40})$$

where $\mathcal{D}(\boldsymbol{\eta})$ is the complex Fisher information matrix (FIM) for $\boldsymbol{\eta}$, which is given by

$$\mathcal{D}(\boldsymbol{\eta}) = \mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)^H \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right) \right\}. \quad (\text{A.41})$$

Note that, $\mathcal{D}(\boldsymbol{\eta})$ is a $5L \times 5L$ matrix, which has a block structure, where each block has a dimension

of $L \times L$. The entries of $\mathcal{D}(\boldsymbol{\eta})$ in its main block diagonal are calculated as

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{R,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{R,q}} \right) \right\} = 2\Re\{[\mathbf{R}_{R,R}]_{[m,n]}\}, \quad (\text{A.42})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{T,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{T,q}} \right) \right\} = 2\Re\{[\mathbf{R}_{T,T}]_{[m,n]}\}, \quad (\text{A.43})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_p^h} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^h} \right) \right\} = 2\Re\{[\mathbf{R}_{h,h}]_{[m,n]}\}, \quad (\text{A.44})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_p^v} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^v} \right) \right\} = 2\Re\{[\mathbf{R}_{h,h}]_{[m,n]}\}, \quad (\text{A.45})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \alpha_p} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \alpha_q} \right) \right\} = [\mathbf{R}_{\alpha,\alpha}]_{[m,n]}^*, \quad (\text{A.46})$$

where $p, q \in \{1, \dots, L\}$, $m = (p-1)L + p$, $n = (q-1)L + q$,

$$\mathbf{R}_{R,R} = \frac{1}{\sigma^2} (\dot{\mathbf{A}}_R^\top \dot{\mathbf{A}}_R^* \otimes \mathbf{C}_R^\top \mathbf{C}_R^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.47})$$

$$\mathbf{R}_{T,T} = \frac{1}{\sigma^2} (\dot{\mathbf{A}}_T^\top \dot{\mathbf{A}}_T^* \otimes \mathbf{C}_T^\top \mathbf{C}_T^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.48})$$

$$\mathbf{R}_{h,h} = \frac{1}{\sigma^2} ((\dot{\mathbf{B}}^h)^\top (\dot{\mathbf{B}}^h)^* \otimes (\mathbf{C}^h)^\top (\mathbf{C}^h)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.49})$$

$$\mathbf{R}_{v,v} = \frac{1}{\sigma^2} ((\dot{\mathbf{B}}^v)^\top (\dot{\mathbf{B}}^v)^* \otimes (\mathbf{C}^v)^\top (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.50})$$

$$\mathbf{R}_{\alpha,\alpha} = \frac{1}{\sigma^2} ((\dot{\mathbf{Q}}^v)^\top (\dot{\mathbf{Q}}^v)^* \otimes (\mathbf{C}^v)^\top (\mathbf{C}^v)^*) \in \mathbb{C}^{L \times L} \quad (\text{A.51})$$

where $\dot{\mathbf{A}}_R = [\dot{\mathbf{a}}_{R,1}, \dots, \dot{\mathbf{a}}_{R,L}] \in \mathbb{C}^{N_R \times L}$, $\dot{\mathbf{A}}_T = [\dot{\mathbf{a}}_{T,1}, \dots, \dot{\mathbf{a}}_{T,L}] \in \mathbb{C}^{T_T \times L}$, $\dot{\mathbf{B}}^h = [\dot{\mathbf{b}}_1^h, \dots, \dot{\mathbf{b}}_L^h] \in \mathbb{C}^{T_S^h \times L}$, $\dot{\mathbf{B}}^v = [\dot{\mathbf{b}}_1^v, \dots, \dot{\mathbf{b}}_L^v] \in \mathbb{C}^{T_S^v \times L}$, and $\dot{\mathbf{Q}}^v = \Phi_v^\top \mathbf{B}^v \in \mathbb{C}^{T_S^v \times L}$, while $\dot{\mathbf{a}}_{R,\ell} = j\mathbf{W}^H \mathbf{D}_{M_R} \mathbf{a}_{1D}(\psi_{R,\ell})$, $\dot{\mathbf{a}}_{R,\ell} = j\mathbf{F}^\top \mathbf{D}_{M_T} \mathbf{a}_{1D}(\psi_{T,\ell})$, $\dot{\mathbf{b}}_\ell^h = j\Phi_h^\top \mathbf{D}_{M_S}^h \mathbf{b}_{1D}(\mu_\ell^h)$, $\dot{\mathbf{b}}_\ell^v = j\Phi_v^\top \mathbf{D}_{M_S}^v \mathbf{b}_{1D}(\mu_\ell^v)$, and $\mathbf{D}_M = \text{diag}\{[0, \dots, (M-1)]\}$.

Moreover, the entries of $\mathcal{D}(\boldsymbol{\eta})$ in its off block diagonals are calculated as

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{R,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{T,q}} \right) \right\} = 2\Re\{[\mathbf{R}_{R,T}]_{[m,n]}\}, \quad (\text{A.52})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{R,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^h} \right) \right\} = 2\Re\{[\mathbf{R}_{R,h}]_{[m,n]}\}, \quad (\text{A.53})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{R,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^v} \right) \right\} = 2\Re\{[\mathbf{R}_{R,v}]_{[m,n]}\}, \quad (\text{A.54})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{R,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \alpha_q} \right) \right\} = [\mathbf{R}_{R,\alpha}]_{[m,n]}^*, \quad (\text{A.55})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{T,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^h} \right) \right\} = 2\Re\{[\mathbf{R}_{T,h}]_{[m,n]}\}, \quad (\text{A.56})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{T,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^v} \right) \right\} = 2\Re\{[\mathbf{R}_{T,v}]_{[m,n]}\}, \quad (\text{A.57})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \psi_{T,p}} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \alpha_q} \right) \right\} = [\mathbf{R}_{T,\alpha}]_{[m,n]}^*, \quad (\text{A.58})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_p^h} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_q^v} \right) \right\} = 2\Re\{[\mathbf{R}_{h,v}]_{[m,n]}^*\}, \quad (\text{A.59})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_p^h} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \alpha_q} \right) \right\} = [\mathbf{R}_{h,\alpha}]_{[m,n]}^*, \quad (\text{A.60})$$

$$\mathbb{E} \left\{ \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \mu_p^v} \right)^* \left(\frac{\partial L(\boldsymbol{\eta})}{\partial \alpha_q} \right) \right\} = [\mathbf{R}_{v,\alpha}]_{[m,n]}^*, \quad (\text{A.61})$$

where

$$\mathbf{R}_{R,T} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_R^\top \otimes \mathbf{C}_R^\top) \mathbf{R}_{n_1, n_2} (\dot{\mathbf{A}}_T^* \otimes \mathbf{C}_T^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.62})$$

$$\mathbf{R}_{R,h} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_R^\top \otimes \mathbf{C}_R^\top) \mathbf{R}_{n_1, n_3} ((\dot{\mathbf{B}}^h)^* \otimes (\mathbf{C}^h)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.63})$$

$$\mathbf{R}_{R,v} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_R^\top \otimes \mathbf{C}_R^\top) \mathbf{R}_{n_1, n_4} ((\dot{\mathbf{B}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.64})$$

$$\mathbf{R}_{R,\alpha} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_R^\top \otimes \mathbf{C}_R^\top) \mathbf{R}_{n_1, n_4} ((\dot{\mathbf{Q}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.65})$$

$$\mathbf{R}_{T,h} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_T^\top \otimes \mathbf{C}_T^\top) \mathbf{R}_{n_2, n_3} ((\dot{\mathbf{B}}^h)^* \otimes (\mathbf{C}^h)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.66})$$

$$\mathbf{R}_{T,v} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_T^\top \otimes \mathbf{C}_T^\top) \mathbf{R}_{n_2, n_4} ((\dot{\mathbf{B}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.67})$$

$$\mathbf{R}_{T,\alpha} = \frac{1}{\sigma^4} (\dot{\mathbf{A}}_T^\top \otimes \mathbf{C}_T^\top) \mathbf{R}_{n_2, n_4} ((\dot{\mathbf{Q}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.68})$$

$$\mathbf{R}_{h,v} = \frac{1}{\sigma^4} ((\dot{\mathbf{B}}^h)^\top \otimes (\mathbf{C}^h)^\top) \mathbf{R}_{n_3, n_4} ((\dot{\mathbf{B}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.69})$$

$$\mathbf{R}_{h,\alpha} = \frac{1}{\sigma^4} ((\dot{\mathbf{B}}^h)^\top \otimes (\mathbf{C}^h)^\top) \mathbf{R}_{n_3, n_4} ((\dot{\mathbf{Q}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.70})$$

$$\mathbf{R}_{v,\alpha} = \frac{1}{\sigma^2} ((\dot{\mathbf{B}}^v)^\top \otimes (\mathbf{C}^v)^\top) ((\dot{\mathbf{Q}}^v)^* \otimes (\mathbf{C}^v)^*) \in \mathbb{C}^{L^2 \times L^2} \quad (\text{A.71})$$

where $\mathbf{R}_{n_a, n_b} = \mathbb{E}\{\mathbf{n}_a \mathbf{n}_b^H\}$ and $\mathbf{n}_b = \text{vec}\{\mathcal{N}\}_{(b)} \in \mathbb{C}^{N_R T_T T_S^h T_S^v}$. Since we assume that \mathcal{N} entries are independent and identically distributed zero mean and circularly symmetric Gaussian random variables with variance σ^2 , every $\mathbf{R}_{n_a n_b}$, $a \neq b$, contains $N_R T_T T_S^h T_S^v$ nonzero entries. Note that, the $[d_1, d_2, d_3, d_4]$ th entry of $\mathcal{N} \in \mathbb{C}^{N_R \times T_T \times T_S^h \times T_S^v}$, i.e., $[\mathcal{N}]_{[d_1, d_2, d_3, d_4]}$ equals to

$$\begin{aligned} [\mathcal{N}]_{[d_1, d_2, d_3, d_4]} &= [[\mathcal{N}]^{(1)}]_{[d_1, d_2 + (d_3 - 1)T_T + (d_4 - 1)T_T T_S^h]} \\ &= [[\mathcal{N}]^{(2)}]_{[d_2, d_1 + (d_3 - 1)N_R + (d_4 - 1)N_R T_S^h]} \\ &= [[\mathcal{N}]^{(3)}]_{[d_3, d_1 + (d_2 - 1)N_R + (d_4 - 1)N_R T_T]} \\ &= [[\mathcal{N}]^{(4)}]_{[d_4, d_1 + (d_2 - 1)N_R + (d_3 - 1)N_R T_T]} \\ &= [\mathbf{n}^{(1)}]_{[d_2 + (d_3 - 1)T_T + (d_4 - 1)T_T T_S^h + (d_1 - 1)T_T T_S^h T_S^v]}^* \\ &= [\mathbf{n}^{(2)}]_{[d_1 + (d_3 - 1)N_R + (d_4 - 1)N_R T_S^h + (d_2 - 1)N_R T_S^h T_S^v]}^* \\ &= [\mathbf{n}^{(3)}]_{[d_1 + (d_2 - 1)N_R + (d_4 - 1)N_R T_T + (d_3 - 1)N_R T_T T_S^v]}^* \\ &= [\mathbf{n}^{(4)}]_{[d_1 + (d_2 - 1)N_R + (d_3 - 1)N_R T_T + (d_4 - 1)N_R T_T T_S^h]}^* \end{aligned} \quad (\text{A.72})$$

Therefore, the corresponding indexes r_1 and r_2 such that $[\mathbf{R}_{\mathbf{n}_1, \mathbf{n}_2}]_{[r_1, r_2]} \neq 0$ are given as

$$r_1 = d_2 + (d_3 - 1)T_{\Gamma} + (d_4 - 1)T_{\Gamma}T_{\mathbb{S}}^{\text{h}} + (d_1 - 1)T_{\Gamma}T_{\mathbb{S}}^{\text{h}}T_{\mathbb{S}}^{\text{v}} \quad (\text{A.73})$$

$$r_2 = d_1 + (d_3 - 1)N_{\mathbb{R}} + (d_4 - 1)N_{\mathbb{R}}T_{\mathbb{S}}^{\text{h}} + (d_1 - 1)N_{\mathbb{R}}T_{\mathbb{S}}^{\text{h}}T_{\mathbb{S}}^{\text{v}}. \quad (\text{A.74})$$

Similarly, the corresponding indexes r_1 and r_2 such that $[\mathbf{R}_{\mathbf{n}_1, \mathbf{n}_3}]_{[r_1, r_2]} \neq 0$, $[\mathbf{R}_{\mathbf{n}_1, \mathbf{n}_4}]_{[r_1, r_2]} \neq 0$, $[\mathbf{R}_{\mathbf{n}_2, \mathbf{n}_3}]_{[r_1, r_2]} \neq 0$, $[\mathbf{R}_{\mathbf{n}_2, \mathbf{n}_4}]_{[r_1, r_2]} \neq 0$, $[\mathbf{R}_{\mathbf{n}_3, \mathbf{n}_4}]_{[r_1, r_2]} \neq 0$ can be obtained from (A.72).

A.7. Nested PARAFAC Models

In this Appendix, we show how 4-way tensor \mathcal{Y} defined in Subsection 5.3.2 can be illustrated as nested PARAFAC decomposition as illustrated in Fig. 5.3. Let the 4-way measurement tensor \mathcal{Y} to be given as

$$\mathcal{Y} = \mathcal{X} + \mathcal{N}, \quad (\text{A.75})$$

where \mathcal{N} is the 4-way noise tensor. Consider the 4-way tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times J_1 \times I_2 \times J_2}$

$$x_{i_1, j_1, i_2, j_2} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} a_{i_1, r_1}^{(1)} \cdot b_{j_1, r_1}^{(1)} \cdot a_{i_2, r_2}^{(2)} \cdot b_{j_2, r_2}^{(2)} \cdot g_{r_1, r_2}. \quad (\text{A.76})$$

where x_{i_1, j_1, i_2, j_2} is the (i_1, j_1, i_2, j_2) th entry of the 4-way tensor \mathcal{X} . As shown in [AF13], the above tensor model can be seen as two nested 3-way tensors admitting PARAFAC models with shared $\mathbf{G} \in \mathbb{C}^{R_1 \times R_2}$ as a common factor matrix. To show this, consider the following two 3-way tensors $\mathcal{Z}^{(1)} \in \mathbb{C}^{I_1 \times J_1 \times R_2}$ and $\mathcal{Z}^{(2)} \in \mathbb{C}^{I_2 \times J_2 \times R_1}$ defined as

$$z_{i_1, j_1, r_2}^{(1)} = \sum_{r_1=1}^{R_1} a_{i_1, r_1}^{(1)} \cdot b_{j_1, r_1}^{(1)} \cdot g_{r_1, r_2} \quad (\text{A.77})$$

$$z_{i_2, j_2, r_1}^{(2)} = \sum_{r_2=1}^{R_2} a_{i_2, r_2}^{(2)} \cdot b_{j_2, r_2}^{(2)} \cdot g_{r_1, r_2}. \quad (\text{A.78})$$

which correspond to PARAFAC decompositions of the $\mathcal{Z}^{(1)}$ and $\mathcal{Z}^{(2)}$ tensors with factor matrices $\{\mathbf{A}^{(1)}, \mathbf{B}^{(1)}, \mathbf{G}^\top\}$ and $\{\mathbf{A}^{(2)}, \mathbf{B}^{(2)}, \mathbf{G}\}$, respectively. We define $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ as

$$\mathbf{D}^{(1)} = (\mathbf{A}^{(1)} \diamond \mathbf{B}^{(1)})\mathbf{G} \in \mathbb{C}^{K_1 \times R_2}, \quad (\text{A.79})$$

$$\mathbf{D}^{(2)} = (\mathbf{A}^{(2)} \diamond \mathbf{B}^{(2)})\mathbf{G}^\top \in \mathbb{C}^{K_2 \times R_1}, \quad (\text{A.80})$$

where $K_n = I_n J_n$. Note that the matrix representations of $\mathcal{Z}^{(1)}$ and $\mathcal{Z}^{(2)}$ correspond to a contraction of $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$. Define the following quantities

$$d_{k_1, r_2}^{(1)} = z_{i_1, j_1, r_2}^{(1)}, \quad (\text{A.81})$$

$$d_{k_2, r_1}^{(2)} = z_{i_2, j_2, r_1}^{(2)}, \quad (\text{A.82})$$

where $k_n = (i_n - 1)J_n + j_n$, while $d_{k_1, r_2}^{(1)}$ and $d_{k_2, r_1}^{(2)}$ are (k_1, r_2) th and (k_2, r_1) th entries of $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, respectively. Therefore, the 3-way tensors $\mathcal{Z}^{(1)}$ and $\mathcal{Z}^{(2)}$ can be expressed as the matrices $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, respectively. Then, we can rewrite (A.76) as two nested PARAFAC models as

$$x_{i_1, j_1, k_2} = \sum_{r_1=1}^{R_1} a_{i_1, r_1}^{(1)} \cdot b_{j_1, r_1}^{(1)} \cdot d_{k_2, r_1}^{(2)} \quad (\text{A.83})$$

$$x_{i_2, j_2, k_1} = \sum_{r_2=1}^{R_2} a_{i_2, r_2}^{(2)} \cdot b_{j_2, r_2}^{(2)} \cdot d_{k_1, r_2}^{(1)}, \quad (\text{A.84})$$

with corresponding factor matrices given as $\{\mathbf{A}^{(1)}, \mathbf{B}^{(1)}, \mathbf{D}^{(2)}\}$ and $\{\mathbf{A}^{(2)}, \mathbf{B}^{(2)}, \mathbf{D}^{(1)}\}$, respectively.

A.8. Symbols and Notations

j	Imaginary unit ($\sqrt{-1}$)
e	Euler number
π	Pi
\mathbb{B}	Set of binary numbers, i.e., zero and one
\mathbb{R}	Set of real numbers
\mathbb{C}	Set of complex numbers
a, b, c	Scalars
$\mathbf{a}, \mathbf{b}, \mathbf{c}$	Column vectors
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	Matrices
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	Tensors
$[\mathbf{a}]_i$	The i th element of the column vector $\mathbf{a} \in \mathbb{C}^I$
$[\mathbf{A}]_{[i,j]}$	The (i, j) th element of the matrix $\mathbf{A} \in \mathbb{C}^{I \times J}$
$[\mathcal{A}]_{[i,j,q]}$	The (i, j, q) th element of the 3-way tensor $\mathcal{A} \in \mathbb{C}^{I \times J \times Q}$
$\mathbf{1}_P$	Column vector of ones of size $P \times 1$
$\mathbf{0}_{P \times Q}$	Matrix of zeros of size $P \times Q$
$\mathbf{1}_{P \times Q}$	Matrix of ones of size $P \times Q$
\mathbf{I}_P	Identity matrix of size $P \times P$
$\mathcal{I}_{D,P}$	Super-diagonal tensor of the order D and the super-diagonal are ones and the rest of the elements is zero
$\angle(z)$	Argument (phase) of a complex variable z
$\text{diag}(\cdot)$	Transforms a vector to a square diagonal matrix.
$\text{undia}(\cdot)$	Extracts the main diagonal of a square matrix and writes its elements to a vector
$\text{blkdiag}\{\mathbf{X}_1, \dots, \mathbf{X}_N\}$	Transforms the matrices $\mathbf{A}_1, \dots, \mathbf{A}_N$ into a block diagonal matrix as $\text{blkdiag}\{\mathbf{A}_1, \dots, \mathbf{A}_N\} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_N \end{bmatrix}$
$\text{vec}\{\cdot\}$	Transforms a matrix or a tensor to a column vector.
$\text{unvec}_{P \times Q}\{\cdot\}$	is the inverse operation of $\text{vec}\{\cdot\}$, which transforms a vector into a matrix of the size $P \times Q$.
$\text{rank}(\cdot)$	Rank of a matrix or a tensor
$d\text{-rank}(\cdot)$	d – rank(\cdot) (multi-linear rank) of a tensor
$k\text{-rank}(\cdot)$	Kruskal rank of a matrix
$\mathbb{E}\{\cdot\}$	Expected value
$ \cdot $	Determinant of a matrix or absolute value of a scalar
$\lceil x \rceil$	Ceiling function

A. Appendices

\sqcup_n	Concatenation along the n th dimension
$\ \cdot\ _2$	Euclidean norm
$\ \cdot\ _F$	Frobenius norm
$\ \cdot\ _H$	Higher-order norm
z^*	Complex conjugate of z
$(\mathbf{A})^T$	Matrix transpose
$(\mathbf{A})^H$	Matrix conjugate transpose
$(\mathbf{A})^{-1}$	Matrix inverse
$(\mathbf{A})^+$	Moore-Penrose pseudo-inverse of matrix
$(\mathbf{A})^{\frac{1}{2}}$	Square root of matrix \mathbf{A} can be computed via its SVD. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ be the SVD of \mathbf{A} , then $(\mathbf{A})^{\frac{1}{2}} = \mathbf{U}(\mathbf{\Sigma})^{\frac{1}{2}}\mathbf{V}^H$. Note that Matlab uses the algorithm in [DHR12] which is based on blocked Schur decomposition to improve the computation speed.
\approx	Approximate
$\mathcal{U}(a, b)$	Uniform distribution from a to b
$\mathcal{N}(\mu, \sigma^2)$	Normal or Gaussian distribution for real valued random variable with mean μ and variance σ^2
$\mathcal{CN}(\mu, \mathbf{R})$	Complex Gaussian distribution with mean μ and covariance matrix \mathbf{R}
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product between $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{B} \in \mathbb{C}^{P \times Q}$ (see Subsection 1.3)
$\mathbf{A} \diamond \mathbf{B}$	Khatri-Rao product between $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{B} \in \mathbb{C}^{P \times M}$ (see Subsection 1.3)
$\mathbf{A} \odot \mathbf{B}$	Hadamard product between $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{B} \in \mathbb{C}^{N \times M}$ (see Subsection 1.3)
$\mathcal{A} \times_n \mathbf{X}$	n -mode product between \mathcal{A} and a matrix \mathbf{X} (see Subsection 1.3)
$[\mathcal{A}]_{(n)}$	n -mode unfolding of tensor \mathcal{A} (see Subsection 1.3)
Big O	Mathematical notation used to approximate the complexity of a given function or an algorithm

Bibliography

- [21a] 5G-Advanced's system architecture begins taking shape at 3GPP. <https://www.nokia.com/blog/5g-advanceds-system-architecture-begins-taking-shape-at-3gpp/>. Nov. 2021.
- [21b] Release 18. <https://www.3gpp.org/release18>. Nov. 2021.
- [22] 5G evolution toward 5G advanced: An overview of 3GPP releases 17 and 18. <https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/5g-evolution-toward-5g-advanced>. Aug. 2022.
- [3GP16a] Portal 3GPP. 3GPP Portal > Specifications. <https://portal.3gpp.org/Specifications.aspx?q=1&releases=189/>. Oct. 2016.
- [3GP16b] Portal 3GPP. 3GPP Portal > Specifications. <https://portal.3gpp.org/Specifications.aspx?q=1&releases=190>. Oct. 2016.
- [3GP20a] 3GPP. *Study on Channel Model for Frequencies from 0.5 to 100 GHz*. 2020.
- [3GP20b] Portal 3GPP. 3GPP Portal > Specifications. <https://portal.3gpp.org/Specifications.aspx?q=1&releases=191>. July 2020.
- [AA20] G. T. de Araújo and A. L. F. de Almeida. "PARAFAC-Based Channel Estimation for Intelligent Reflective Surface Assisted MIMO System". in *Proc. IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. 2020, pp. 1–5. DOI: 10.1109/SAM48682.2020.9104260.
- [AAB21] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer. "Channel Estimation for Intelligent Reflecting Surface Assisted MIMO Systems: A Tensor Modeling Approach". *IEEE Journal of Selected Topics in Signal Processing* 15.3 (2021), pp. 789–802. DOI: 10.1109/JSTSP.2021.3061274.
- [AAH19a] K. Ardah, A. L. F. de Almeida, and M. Haardt. "A Gridless CS Approach for Channel Estimation in Hybrid Massive MIMO Systems". in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 4160–4164. DOI: 10.1109/ICASSP.2019.8683049.
- [AAH19b] K. Ardah, A. L. F. de Almeida, and M. Haardt. "Low-Complexity Millimeter Wave CSI Estimation in MIMO-OFDM Hybrid Beamforming Systems". in *Proc. 23rd International ITG Workshop on Smart Antennas (WSA)*. 2019, pp. 1–5.

-
- [AF13] A. L. F. de Almeida and G. Favier. "Double Khatri–Rao Space-Time-Frequency Coding Using Semi-Blind PARAFAC Based Receiver". *IEEE Signal Processing Letters* 20.5 (2013), pp. 471–474. DOI: 10.1109/LSP.2013.2248149.
- [AFM08] A. L. F. de Almeida, G. Favier, and J. C. M. Mota. "A Constrained Factor Decomposition With Application to MIMO Antenna Systems". *IEEE Trans. Sig. Proc.* 56.6 (June 2008), pp. 2429–2442. ISSN: 1053-587X. DOI: 10.1109/TSP.2008.917026. URL: <https://doi.org/10.1109/TSP.2008.917026>.
- [AH21] K. Ardah and M. Haardt. "Compressed Sensing Constant Modulus Constrained Projection Matrix Design and High-Resolution DoA Estimation Methods". in *Proc. 25th International ITG Workshop on Smart Antennas (WSA)*. 2021, pp. 1–5.
- [AHK21] T. Abood, I. Hburi, and H. F. Khazaal. "Massive MIMO: An Overview, Recent Challenges, and Future Research Directions". in *Proc. International Conference on Advance of Sustainable Engineering and its Application (ICASEA)*. 2021, pp. 43–48. DOI: 10.1109/ICASEA53739.2021.9733081.
- [AIU06] F. A. Dietrich, T. Ivanov, and W. Utschick. "Estimation of channel and noise correlations for MIMO channel estimation". in *Proc. International ITG Workshop on Smart Antennas (WSA), Ulm, Germany*. Oct. 2006.
- [Akh+20] M. W. Akhtar, S. A. Hassan, R. Ghaffar, H. Jung, S. Garg, and M. S. Hossain. "The Shift to 6G Communications: Vision and Requirements". *Hum.-Centric Comput. Inf. Sci.* 10.1 (Dec. 2020). ISSN: 2192-1962. DOI: 10.1186/s13673-020-00258-2.
- [AKS17] I. Ahmed, H. Khammari, and A. Shahid. "Resource Allocation for Transmit Hybrid Beamforming in Decoupled Millimeter Wave Multiuser-MIMO Downlink". *IEEE Access* 5 (2017), pp. 170–182. DOI: 10.1109/ACCESS.2016.2634096.
- [Ala98] S. M. Alamouti. "A simple transmit diversity technique for wireless communications". *IEEE Journal on Selected Areas in Communications* 16.8 (1998), pp. 1451–1458. DOI: 10.1109/49.730453.
- [ALH15] A. Alkhateeb, G. Leus, and R. W. Heath. "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015, pp. 2909–2913. DOI: 10.1109/ICASSP.2015.7178503.
- [Alk+14] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath. "MIMO Precoding and Combining Solutions for Millimeter-Wave Systems". *IEEE Communications Magazine* 52.12 (2014), pp. 122–131. DOI: 10.1109/MCOM.2014.6979963.
- [Alm+06] A. L. F. de Almeida, G. Favier, C. C. Cavalcante, and J. C. M. Mota. "Tensor-Based Space-Time Multiplexing Codes for MIMO-OFDM Systems with Blind Detection". in *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*. 2006, pp. 1–5. DOI: 10.1109/PIMRC.2006.253943.
- [Alm07] A. L. F. de Almeida. "Tensor modeling and signal processing for wireless communication systems". Theses. Université de Nice Sophia Antipolis, Nov. 2007. URL: <https://theses.hal.science/tel-00460157>.

A. Bibliography

- [And+14] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications* 32.6 (2014), pp. 1065–1082. DOI: 10.1109/JSAC.2014.2328098.
- [APH19] K. Ardah, M. Pesavento, and M. Haardt. "A Novel Sensing Matrix Design for Compressed Sensing via Mutual Coherence Minimization". *IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. 2019, pp. 66–70. DOI: 10.1109/CAMSAP45676.2019.9022467.
- [Ara+19] D. C. Araújo, A. L. F. de Almeida, João P. C. L. Da Costa, and Rafael T. de Sousa. "Tensor-Based Channel Estimation for Massive MIMO-OFDM Systems". *IEEE Access* 7 (2019), pp. 42133–42147. DOI: 10.1109/ACCESS.2019.2908207.
- [Ard+18] K. Ardah, G. Fodor, Y. C. B. Silva, W. C. Freitas, and F. R. P. Cavalcanti. "A Unifying Design of Hybrid Beamforming Architectures Employing Phase Shifters or Switches". *IEEE Transactions on Vehicular Technology* 67.11 (2018), pp. 11243–11247. DOI: 10.1109/TVT.2018.2865705.
- [Ard+20a] K. Ardah, G. Fodor, Y. C. B. Silva, W. C. Freitas, and A. L. F. de Almeida. "Hybrid Analog-Digital Beamforming Design for SE and EE Maximization in Massive MIMO Networks". *IEEE Transactions on Vehicular Technology* 69.1 (2020), pp. 377–389. DOI: 10.1109/TVT.2019.2933305.
- [Ard+20b] K. Ardah, B. Sokal, A. L. F. de Almeida, and M. Haardt. "Compressed Sensing Based Channel Estimation and Open-loop Training Design for Hybrid Analog-digital Massive MIMO Systems". in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 4597–4601. DOI: 10.1109/ICASSP40776.2020.9054443.
- [Ard+21] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt. "TRICE: A Channel Estimation Framework for RIS-Aided Millimeter-Wave MIMO Systems". *IEEE Signal Processing Letters* 28 (2021), pp. 513–517. DOI: 10.1109/LSP.2021.3059363.
- [Ard+22] K. Ardah, S. Gherekhloo, A. L. F. de Almeida, and M. Haardt. "Double-RIS Versus Single-RIS Aided Systems: Tensor-Based MIMO Channel Estimation and Design Perspectives". in *proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022, pp. 5183–5187. DOI: 10.1109/ICASSP43922.2022.9746287.
- [ASC17] K. Ardah, Y. Silva, and F. Cavalcanti. "Decentralized Linear Transceiver Design in Multicell MIMO Broadcast Channels". *Journal of Communication and Information Systems* 32.1 (Oct. 2017). DOI: 10.14209/jcis.2017.11. URL: <https://jcis.sbrt.org.br/jcis/article/view/461>.
- [AV20] G. C. Alexandropoulos and E. Vlachos. "A Hardware Architecture For Reconfigurable Intelligent Surfaces with Minimal Active Elements for Explicit Channel Estimation". in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 9175–9179. DOI: 10.1109/ICASSP40776.2020.9053976.
- [Aya+12] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi. "Low complexity precoding for large millimeter wave MIMO systems". in *Proc. IEEE International Conference on Communications (ICC)*. 2012, pp. 3724–3729. DOI: 10.1109/ICC.2012.6363634.

-
- [Aya+14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath. "Spatially Sparse Precoding in Millimeter Wave MIMO Systems". *IEEE Transactions on Wireless Communications* 13.3 (2014), pp. 1499–1513. DOI: 10.1109/TWC.2014.011714.130846.
- [AZY20] S. Abeywickrama, R. Zhang, and C. Yuen. "Intelligent Reflecting Surface: Practical Phase Shift Model and Beamforming Optimization". in *Proc. IEEE International Conference on Communications (ICC)*. 2020, pp. 1–6. DOI: 10.1109/ICC40277.2020.9148961.
- [BA20] A. A. Boulogeorgos and A. Alexiou. "Performance Analysis of Reconfigurable Intelligent Surface-Assisted Wireless Systems and Comparison With Relaying". *IEEE Access* 8 (2020), pp. 94463–94483. DOI: 10.1109/ACCESS.2020.2995435.
- [Bas+19] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang. "Wireless Communications Through Reconfigurable Intelligent Surfaces". *IEEE Access* 7 (2019), pp. 116753–116773. DOI: 10.1109/ACCESS.2019.2935192.
- [BBS13] J. Brady, N. Behdad, and A. M. Sayeed. "Beamspace MIMO for Millimeter-Wave Communications: System Architecture, Modeling, Analysis, and Measurements". *IEEE Transactions on Antennas and Propagation* 61.7 (2013), pp. 3814–3827. DOI: 10.1109/TAP.2013.2254442.
- [BG06] M. Biguesh and A. B. Gershman. "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals". *IEEE Transactions on Signal Processing* 54.3 (2006), pp. 884–893. DOI: 10.1109/TSP.2005.863008.
- [Bjö+15] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah. "Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer?" *IEEE Trans. Wireless Commun.* 14.6 (June 2015), pp. 3059–3075.
- [BL17] B. Basutli and S. Lambotharan. "Game-theoretic beamforming techniques for multiuser multi-cell networks under mixed quality of service constraints". *IET Signal Processing* 11.5 (2017), pp. 631–639. DOI: <https://doi.org/10.1049/iet-spr.2016.0192>.
- [BLM03] I. Barhumi, G. Leus, and M. Moonen. "Optimal training design for MIMO OFDM systems in mobile wireless channels". *IEEE Transactions on Signal Processing* 51.6 (2003), pp. 1615–1624. DOI: 10.1109/TSP.2003.811243.
- [Boc+14] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski. "Five disruptive technology directions for 5G". *IEEE Communications Magazine* 52.2 (2014), pp. 74–80. DOI: 10.1109/MCOM.2014.6736746.
- [BÖL20] E. Björnson, Ö. Özdogan, and E. G. Larsson. "Intelligent Reflecting Surface Versus Decode-and-Forward: How Large Surfaces are Needed to Beat Relaying?" *IEEE Wireless Communications Letters* 9.2 (2020), pp. 244–248. DOI: 10.1109/LWC.2019.2950624.
- [Bre78] J. Brewer. "Kronecker products and matrix calculus in system theory". *IEEE Transactions on Circuits and Systems* 25.9 (1978), pp. 772–781. DOI: 10.1109/TCS.1978.1084534.
- [BS20] E. Björnson and L. Sanguinetti. "Power Scaling Laws and Near-Field Behaviors of Massive MIMO and Intelligent Reflecting Surfaces". *IEEE Open Journal of the Communications Society* 1 (2020), pp. 1306–1324. DOI: 10.1109/OJCOMS.2020.3020925.

A. Bibliography

- [BS21] E. Björnson and L. Sanguinetti. "Rayleigh Fading Modeling and Channel Hardening for Reconfigurable Intelligent Surfaces". *IEEE Wireless Communications Letters* 10.4 (2021), pp. 830–834. DOI: 10.1109/LWC.2020.3046107.
- [BX22] S. Bazzi and W. Xu. "IRS Parameter Optimization for Channel Estimation MSE Minimization in Double-IRS Aided Systems". *IEEE Wireless Communications Letters* 11 (2022), pp. 2170–2174. URL: <https://api.semanticscholar.org/CorpusID:251334059>.
- [Cai+16] M. Cai, K. Gao, D. Nie, B. Hochwald, J. N. Laneman, H. Huang, and K. Liu. "Effect of Wideband Beam Squint on Codebook Design in Phased-Array Wireless Systems". in *Proc. IEEE Global Communications Conference (GLOBECOM)*. 2016, pp. 1–6. DOI: 10.1109/GLOCOM.2016.7841766.
- [Cao+18] M. Cao, X. Mao, X. Long, and L. Huang. "Direction-of-Arrival Estimation for Uniform Rectangular Array: A Multilinear Projection Approach". in *Proc. 26th European Signal Processing Conference (EUSIPCO)*. 2018, pp. 1237–1241. DOI: 10.23919/EUSIPCO.2018.8553326.
- [CD13] A. Cline and I. Dhillon. "Computation of the Singular Value Decomposition". Dec. 2013, pp. 1027–1039. DOI: 10.1201/b16113-68.
- [Che+23] J. Chen, Y. Liang, H. V. Cheng, and W. Yu. "Channel Estimation for Reconfigurable Intelligent Surface Aided Multi-User mmWave MIMO Systems". *IEEE Transactions on Wireless Communications* (2023), pp. 1–1. DOI: 10.1109/TWC.2023.3246264.
- [Cho+20] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang. "6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions". *IEEE Open Journal of the Communications Society* 1 (2020), pp. 957–975. DOI: 10.1109/OJCOMS.2020.3010270.
- [Chr+08] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi. "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design". *IEEE Transactions on Wireless Communications* 7.12 (2008), pp. 4792–4799. DOI: 10.1109/T-WC.2008.070851.
- [Cic+15] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. PHAN. "Tensor Decompositions for Signal Processing Applications: From two-way to multiway component analysis". *IEEE Signal Processing Magazine* 32.2 (2015), pp. 145–163. DOI: 10.1109/MSP.2013.2297439.
- [CLA09] P. Comon, X. Luciani, and A. L. F. de Almeida. "Tensor decompositions, alternating least squares and other tales". *Journal of Chemometrics* 23.7-8 (2009), pp. 393–405. DOI: <https://doi.org/10.1002/cem.1236>.
- [Com14] P. Comon. "Tensors: A brief introduction". *IEEE Signal Processing Magazine* 31.3 (2014), pp. 44–53. DOI: 10.1109/MSP.2014.2298533.
- [CS04] I. Csiszár and P.C. Shields. "Information Theory and Statistics: A Tutorial". *Foundations and Trends® in Communications and Information Theory* 1.4 (2004), pp. 417–528. ISSN: 1567-2190. DOI: 10.1561/01000000004.
- [CTY16] H. Chen, A. J. Taylor, and N. Yu. "A review of metasurfaces: physics and applications". *Reports on Progress in Physics* 79.7 (June 2016), p. 076401. DOI: 10.1088/0034-4885/79/7/076401.

-
- [Cui+14] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng. *Coding Metamaterials, Digital Metamaterials and Programming Metamaterials*. 2014. arXiv: 1407.8442 [physics.optics].
- [CZZ19] M. Cui, G. Zhang, and R. Zhang. "Secure Wireless Communication via Intelligent Reflecting Surface". *IEEE Wireless Communications Letters* 8.5 (2019), pp. 1410–1414. DOI: 10.1109/LWC.2019.2919685.
- [Dai+15] L. Dai, X. Gao, J. Quan, S. Han, and C. I. "Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems". in *Proc. IEEE International Conference on Communications (ICC)*. 2015, pp. 1334–1339. DOI: 10.1109/ICC.2015.7248508.
- [DDV00] L. De Lathauwer, B. De Moor, and J. Vandewalle. "A Multilinear Singular Value Decomposition". *SIAM Journal on Matrix Analysis and Applications* 21.4 (2000), pp. 1253–1278. DOI: 10.1137/S0895479896305696.
- [De 06] L. De Lathauwer. "A Link between the Canonical Decomposition in Multilinear Algebra and Simultaneous Matrix Diagonalization". *SIAM Journal on Matrix Analysis and Applications* 28.3 (2006), pp. 642–666. DOI: 10.1137/040608830.
- [dFR18] M. N. da Costa, G. Favier, and J. M. T. Romano. "Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers". *Signal Processing* 145 (2018), pp. 304–316. ISSN: 0165-1684. DOI: <https://doi.org/10.1016/j.sigpro.2017.12.015>.
- [DH23] R. M. Dreifuerst and R. W. Heath. *Massive MIMO in 5G: How Beamforming, Codebooks, and Feedback Enable Larger Arrays*. 2023. arXiv: 2301.13390 [eess.SP].
- [DHR12] E. Deadman, N. J. Higham, and R. Ralha. "Blocked schur algorithms for computing the matrix square root". *Proceedings of the 11th International Conference on Applied Parallel and Scientific Computing*. PARA'12. Helsinki, Finland: Springer-Verlag, 2012, pp. 171–182. ISBN: 9783642368028. DOI: 10.1007/978-3-642-36803-5_12.
- [Don+16] S. Dong, Y. Wang, L. Jiang, and Y. Chen. "Energy Efficiency Analysis with Circuit Power Consumption in Downlink Large-Scale Multiple Antenna Systems". in *Proc. IEEE 83rd Vehicular Technology Conference (VTC Spring)*. 2016, pp. 1–5. DOI: 10.1109/VTCSpring.2016.7504219.
- [Don06] D.L. Donoho. "Compressed sensing". *IEEE Transactions on Information Theory* 52.4 (2006), pp. 1289–1306. DOI: 10.1109/TIT.2006.871582.
- [DPS20] E. Dahlman, S. Parkvall, and J. Skold. *5G NR: The Next Generation Wireless Access Technology*. Elsevier Science, 2020. ISBN: 9780128223208. URL: <https://books.google.de/books?id=PZH9DwAAQBAJ>.
- [Dup+18] D. Dupleich, N. Iqbal, C. Schneider, S. Häfner, R. Müller, S. Skoblikov, J. Luo, G. Del Galdo, and R. Thomä. "Influence of system aspects on fading at mm-waves". *IET Microwaves, Antennas & Propagation* 12.4 (2018), pp. 516–524. DOI: <https://doi.org/10.1049/iet-map.2017.0601>.
- [Dup+19a] D. Dupleich, R. Müller, M. Landmann, E. Shinwasusin, K. Saito, J. Takada, J. Luo, R. Thomä, and G. Del Galdo. "Multi-Band Propagation and Radio Channel Characterization in Street Canyon Scenarios for 5G and Beyond". *IEEE Access* 7 (2019), pp. 160385–160396. DOI: 10.1109/ACCESS.2019.2948869.
-

- [Dup+19b] D. Dupleich, E. Schäfer, G. Del Galdo, and R. Thomä. “Influence of spatial-wideband effects in evaluation of mm-wave communications”. *IET Microwaves, Antennas & Propagation* 13.13 (2019), pp. 2200–2206. DOI: <https://doi.org/10.1049/iet-map.2018.6160>.
- [Dup+20] D. Dupleich, R. Müller, M. Landmann, J. Luo, G. Del Galdo, and R. S. Thomä. “Multi-band Characterization of Propagation in Industry Scenarios”. in *Proc. 14th European Conference on Antennas and Propagation (EuCAP)*. 2020, pp. 1–5. DOI: [10.23919/EuCAP48036.2020.9135630](https://doi.org/10.23919/EuCAP48036.2020.9135630).
- [FA14] G. Favier and A. L. F. de Almeida. “Overview of constrained PARAFAC models”. *EURASIP Journal on Advances in Signal Processing* 2014.1 (Sept. 2014). DOI: [10.1186/1687-6180-2014-142](https://doi.org/10.1186/1687-6180-2014-142).
- [Fan+16] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum. “Super-Resolution Compressed Sensing for Line Spectral Estimation: An Iterative Reweighted Approach”. *IEEE Transactions on Signal Processing* 64.18 (2016), pp. 4649–4662. DOI: [10.1109/TSP.2016.2572041](https://doi.org/10.1109/TSP.2016.2572041).
- [Fer+20] R. C. Ferreira, M. S. P. Facina, F. A. P. De Figueiredo, G. Fraidenraich, and E. R. De Lima. “Bit Error Probability for Large Intelligent Surfaces Under Double-Nakagami Fading Channels”. *IEEE Open Journal of the Communications Society* 1 (2020), pp. 750–759. DOI: [10.1109/OJCOMS.2020.2996797](https://doi.org/10.1109/OJCOMS.2020.2996797).
- [GAH20] S. Gherekhloo, K. Ardah, and M. Haardt. “Hybrid Beamforming Design for Downlink MU-MIMO-OFDM Millimeter-Wave Systems”. in *Proc. IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. 2020, pp. 1–5. DOI: [10.1109/SAM48682.2020.9104405](https://doi.org/10.1109/SAM48682.2020.9104405).
- [GAH21] S. Gherekhloo, K. Ardah, and M. Haardt. “Fully Digital and Hybrid Beamforming Design for Millimeter-Wave MIMO-OFDM Two-Way Relaying Systems”. in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*. 2021, pp. 296–300. DOI: [10.1109/SSP49050.2021.9513786](https://doi.org/10.1109/SSP49050.2021.9513786).
- [GAH23] S. Gherekhloo, K. Ardah, and M. Haardt. “SALSA: A Sequential Alternating Least Squares Approximation Method for MIMO Channel Estimation”. *IEEE Transactions on Vehicular Technology* (2023), pp. 1–6. DOI: [10.1109/TVT.2023.3347290](https://doi.org/10.1109/TVT.2023.3347290).
- [Gao+16] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath. “Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays”. *IEEE Journal on Selected Areas in Communications* 34.4 (2016), pp. 998–1009. DOI: [10.1109/JSAC.2016.2549418](https://doi.org/10.1109/JSAC.2016.2549418).
- [Gao+19] F. Gao, Z. Tian, E. G. Larsson, M. Pesavento, and S. Jin. “Introduction to the Special Issue on Array Signal Processing for Angular Models in Massive MIMO Communications”. *IEEE Journal of Selected Topics in Signal Processing* 13.5 (2019), pp. 882–885. DOI: [10.1109/JSTSP.2019.2938880](https://doi.org/10.1109/JSTSP.2019.2938880).
- [Geb+21] A. Gebre, J. Shaikh, T. Kebede, and F. Z. Gelete. “Comparative Performance Analysis of Channel Estimation Techniques for Massive MIMO System”. in *Proc. IEEE Indian Conference on Antennas and Propagation (InCAP)*. 2021, pp. 236–239. DOI: [10.1109/InCAP52216.2021.9726282](https://doi.org/10.1109/InCAP52216.2021.9726282).

- [Gen+13] J. Geng, Z. Wei, X. Wang, W. Xiang, and D. Yang. "Multiuser hybrid analog/digital beamforming for relatively large-scale antenna arrays". in *Proc. IEEE Globecom Workshops (GC Wkshps)*. 2013, pp. 123–128. DOI: 10.1109/GLOCOMW.2013.6824973.
- [Ges+03] D. Gesbert, M. Shafi, Da-shan Shiu, P. J. Smith, and A. Naguib. "From theory to practice: an overview of MIMO space-time coded wireless systems". *IEEE Journal on Selected Areas in Communications* 21.3 (2003), pp. 281–302. DOI: 10.1109/JSAC.2003.809458.
- [Ghe+21] S. Gherekhloo, K. Ardah, A. L. F. de Almeida, and M. Haardt. "Tensor-Based Channel Estimation and Reflection Design for RIS-Aided Millimeter-Wave MIMO Communication Systems". in *Proc. 55th Asilomar Conference on Signals, Systems, and Computers*. 2021, pp. 1683–1689. DOI: 10.1109/IEEECONF53345.2021.9723362.
- [Ghe+23] S. Gherekhloo, K. Ardah, A. L. F. de Almeida, M. Maleki, and M. Haardt. "Nested PARAFAC Tensor-Based Channel Estimation Method for Double RIS-Aided MIMO Communication Systems". in *Proc. 31st European Signal Processing Conference (EUSIPCO)*. 2023, pp. 1674–1678. DOI: 10.23919/EUSIPCO58844.2023.10290107.
- [GMN18] C. Garvey, C. Meng, and J. G. Nagy. *Singular Value Decomposition Approximation via Kronecker Summations for Imaging Applications*. 2018. arXiv: 1803.11525 [math.NA].
- [Gol+03] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. "Capacity limits of MIMO channels". *IEEE Journal on Selected Areas in Communications* 21.5 (2003), pp. 684–702. DOI: 10.1109/JSAC.2003.810294.
- [Guo+17] W. Guo, J. Fan, G. Y. Li, Q. Yin, and X. Zhu. "Adaptive SU/MU-MIMO scheduling schemes for LTE-A downlink transmission". *IET Communications* 11.6 (2017), pp. 783–792. DOI: <https://doi.org/10.1049/iet-com.2016.0456>.
- [GWZ20] X. Guan, Q. Wu, and R. Zhang. "Intelligent Reflecting Surface Assisted Secrecy Communication: Is Artificial Noise Helpful or Not?" *IEEE Wireless Communications Letters* 9.6 (2020), pp. 778–782. DOI: 10.1109/LWC.2020.2969629.
- [Han+15a] S. Han, C. I, Z. Xu, and C. Rowell. "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G". *IEEE Communications Magazine* 53.1 (2015), pp. 186–194. DOI: 10.1109/MCOM.2015.7010533.
- [Han+15b] K. Haneda, J. Järveläinen, A. Karttunen, M. Kyrö, and J. Putkonen. "A Statistical Spatio-Temporal Radio Channel Model for Large Indoor Environments at 60 and 70 GHz". *IEEE Transactions on Antennas and Propagation* 63.6 (2015), pp. 2694–2704. DOI: 10.1109/TAP.2015.2412147.
- [Han+19] Y. Han, W. Tang, S. Jin, C. Wen, and X. Ma. "Large Intelligent Surface-Assisted Wireless Communication Exploiting Statistical CSI". *IEEE Transactions on Vehicular Technology* 68.8 (2019), pp. 8238–8242. DOI: 10.1109/TVT.2019.2923997.
- [Han+20] Y. Han, S. Zhang, L. Duan, and R. Zhang. "Cooperative Double-IRS Aided Communication: Beamforming Design and Power Scaling". *IEEE Wireless Communications Letters* 9.8 (2020), pp. 1206–1210. DOI: 10.1109/LWC.2020.2986290.
- [Han+21] Y. Han, S. Zhang, L. Duan, and R. Zhang. *Double-IRS Aided MIMO Communication under LoS Channels: Capacity Maximization and Scaling*. 2021. arXiv: 2102.13537 [cs.IT].

A. Bibliography

- [He+20a] J. He, M. Leinonen, H. Wymeersch, and M. Juntti. "Channel Estimation for RIS-Aided mmWave MIMO Systems". in *Proc. IEEE Global Communications Conference*. 2020, pp. 1–6. DOI: 10.1109/GLOBECOM42002.2020.9348112.
- [He+20b] J. He, H. Wymeersch, T. Sanguanpuak, O. Silven, and M. Juntti. "Adaptive Beamforming Design for mmWave RIS-Aided Joint Localization and Communication". in *Proc. IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. 2020, pp. 1–6. DOI: 10.1109/WCNCW48565.2020.9124848.
- [Hea+16] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed. "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems". *IEEE Journal of Selected Topics in Signal Processing* 10.3 (2016), pp. 436–453. DOI: 10.1109/JSTSP.2016.2523924.
- [HLL13] A. Hu, T. Lv, and Y. Lu. "Subspace-Based Semi-Blind Channel Estimation for Large-Scale Multi-Cell Multiuser MIMO Systems". in *Proc. IEEE 77th Vehicular Technology Conference (VTC Spring)*. 2013, pp. 1–5. DOI: 10.1109/VTCspring.2013.6692772.
- [HMP19] G. Hegde, C. Masouros, and M. Pesavento. "Interference Exploitation-Based Hybrid Precoding With Robustness Against Phase Errors". *IEEE Transactions on Wireless Communications* 18.7 (2019), pp. 3683–3696. DOI: 10.1109/TWC.2019.2917064.
- [HMP20] G. Hegde, C. Masouros, and M. Pesavento. "Robust Hybrid Precoding For Interference Exploitation in Massive MIMO Systems". in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 5290–5294. DOI: 10.1109/ICASSP40776.2020.9052973.
- [HMT11] N. Halko, P. G. Martinsson, and J. A. Tropp. "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions". *SIAM Review* 53.2 (2011), pp. 217–288. DOI: 10.1137/090771806.
- [HN98] M. Haardt and J. A. Nossek. "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems". *IEEE Transactions on Signal Processing* 46.1 (1998), pp. 161–169. DOI: 10.1109/78.651206.
- [Hon23] I. Hong. "Reviews Based on the Reconfigurable Intelligent Surface Technical Issues". *Electronics* 12.21 (2023). ISSN: 2079-9292. DOI: 10.3390/electronics12214489.
- [HRD08] M. Haardt, F. Roemer, and G. Del Galdo. "Higher-Order SVD-Based Subspace Estimation to Improve the Parameter Estimation Accuracy in Multidimensional Harmonic Retrieval Problems". *IEEE Transactions on Signal Processing* 56.7 (2008), pp. 3198–3213. DOI: 10.1109/TSP.2008.917929.
- [HS81] H. V. Henderson and S. R. Searle. "The vec-permutation matrix, the vec operator and Kronecker products: a review". *Linear and Multilinear Algebra* 9.4 (1981), pp. 271–288. DOI: 10.1080/03081088108817379.
- [Hua+18a] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen. "Energy Efficient Multi-User MISO Communication Using Low Resolution Large Intelligent Surfaces". in *Proc. IEEE Globecom Workshops (GC Wkshps)*. 2018, pp. 1–6. DOI: 10.1109/GLOCOMW.2018.8644519.

-
- [Hua+18b] C. Huang, A. Zappone, M. Debbah, and C. Yuen. “Achievable Rate Maximization by Passive Intelligent Mirrors”. in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018, pp. 3714–3718. DOI: 10.1109/ICASSP.2018.8461496.
- [Hua+19] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen. “Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication”. *IEEE Transactions on Wireless Communications* 18.8 (2019), pp. 4157–4170. DOI: 10.1109/TWC.2019.2922609.
- [Hur+13] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh. “Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks”. *IEEE Transactions on Communications* 61.10 (2013), pp. 4391–4403. DOI: 10.1109/TCOMM.2013.090513.120848.
- [HW]21] J. He, H. Wymeersch, and M. Juntti. “Channel Estimation for RIS-Aided mmWave MIMO Systems via Atomic Norm Minimization”. *IEEE Transactions on Wireless Communications* 20.9 (2021), pp. 5786–5797. DOI: 10.1109/TWC.2021.3070064.
- [HY20] Z. He and X. Yuan. “Cascaded Channel Estimation for Large Intelligent Metasurface Assisted Massive MIMO”. *IEEE Wireless Communications Letters* 9.2 (2020), pp. 210–214. DOI: 10.1109/LWC.2019.2948632.
- [JD20] T. L. Jensen and E. De Carvalho. “An Optimal Channel Estimation Scheme for Intelligent Reflecting Surfaces Based on a Minimum Variance Unbiased Estimator”. in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 5000–5004. DOI: 10.1109/ICASSP40776.2020.9053695.
- [JGU22] M. Joham, H. Gao, and W. Utschick. “Estimation Of Channels In Systems With Intelligent Reflecting Surfaces”. in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022, pp. 5368–5372. DOI: 10.1109/ICASSP43922.2022.9747624.
- [Jia+22] M. Jian, G. C. Alexandropoulos, E. Basar, C. Huang, R. Liu, Y. Liu, and C. Yuen. “Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques”. *Intelligent and Converged Networks* 3.1 (2022), pp. 1–32. DOI: 10.23919/ICN.2022.0005.
- [JS19] T. Jiang and Y. Shi. “Over-the-Air Computation via Intelligent Reflecting Surfaces”. in *Proc. IEEE Global Communications Conference (GLOBECOM)*. 2019, pp. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9013643.
- [Kan20] J. Kang. “Intelligent Reflecting Surface: Joint Optimal Training Sequence and Reflection Pattern”. *IEEE Communications Letters* 24.8 (2020), pp. 1784–1788. DOI: 10.1109/LCOMM.2020.2991178.
- [KB09] T. G. Kolda and B. W. Bader. “Tensor Decompositions and Applications”. *SIAM Review* 51.3 (2009), pp. 455–500. DOI: 10.1137/07070111X.
- [KCS17] Y. Kwon, J. Chung, and Y. Sung. “Hybrid beamformer design for mmwave wideband multi-user MIMO-OFDM systems : (Invited paper)”. in *Proc. IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2017, pp. 1–5. DOI: 10.1109/SPAWC.2017.8227744.

A. Bibliography

- [Keh+19] R. Kehrle Miranda, J. P. C. L. Da Costa, B. Guo, A. L. F. de Almeida, G. Del Galdo, and R. T. De Sousa. "Low-Complexity and High-Accuracy Semi-Blind Joint Channel and Symbol Estimation for Massive MIMO-OFDM". *Circuits Syst Signal Process* 38 38.3 (Mar. 2019), pp. 1114–1136. DOI: 10.1007/s00034-018-0898-1.
- [KHY15] L. Kong, S. Han, and C. Yang. "Wideband hybrid precoder for massive MIMO systems". in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. 2015, pp. 305–309. DOI: 10.1109/GlobalSIP.2015.7418206.
- [KKS13] C. Kim, T. Kim, and J. Seol. "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems". in *Proc. IEEE Globecom Workshops (GC Wkshps)*. 2013, pp. 61–65. DOI: 10.1109/GLOCOMW.2013.6824962.
- [KL15] M. Kim and Y. H. Lee. "MSE-Based Hybrid RF/Baseband Processing for Millimeter-Wave Communication Systems in MIMO Interference Channels". *IEEE Transactions on Vehicular Technology* 64.6 (2015), pp. 2714–2720. DOI: 10.1109/TVT.2014.2346400.
- [Kra99] S. G. Krantz. "Handbook of Complex Variables". 1999. URL: <https://api.semanticscholar.org/CorpusID:118408892>.
- [Kru77] J. B. Kruskal. "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics". *Linear Algebra and its Applications* 18.2 (1977), pp. 95–138. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/0024-3795\(77\)90069-6](https://doi.org/10.1016/0024-3795(77)90069-6).
- [Kru89] J. B. Kruskal. "Rank, decomposition, and uniqueness for 3-way and n-way arrays". *Multivariate Data Analysis*. NLD: North-Holland Publishing Co., 1989, pp. 7–18. ISBN: 0444874100.
- [Lar+14] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. "Massive MIMO for next generation wireless systems". *IEEE Communications Magazine* 52.2 (2014), pp. 186–195. DOI: 10.1109/MCOM.2014.6736761.
- [Lav+17] A. Lavrenko, F. Römer, G. Del Galdo, and R. Thomä. "Multiband TDOA estimation from sub-Nyquist samples with distributed wideband sensing nodes". in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. 2017, pp. 96–100. DOI: 10.1109/GlobalSIP.2017.8308611.
- [Lav+18] G. Lavigne, K. Achouri, V. S. Asadchy, S. A. Tretyakov, and C. Caloz. "Susceptibility Derivation and Experimental Demonstration of Refracting Metasurfaces Without Spurious Diffraction". *IEEE Transactions on Antennas and Propagation* 66.3 (2018), pp. 1321–1330. DOI: 10.1109/TAP.2018.2793958.
- [Lia+18a] C. Liaskos, S. Nie, A. Tsioliariidou, A. Pitsillides, S. Ioannidis, and I. Akyildiz. "A New Wireless Communication Paradigm through Software-Controlled Metasurfaces". *IEEE Communications Magazine* 56.9 (2018), pp. 162–169. DOI: 10.1109/MCOM.2018.1700659.
- [Lia+18b] C. Liaskos, S. Nie, A. Tsioliariidou, A. Pitsillides, S. Ioannidis, and I. Akyildiz. "Realizing Wireless Communication Through Software-Defined HyperSurface Environments". in *Proc. IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*. 2018, pp. 14–15. DOI: 10.1109/WoWMoM.2018.8449754.
- [Lia+18c] C. Liaskos, A. Tsioliariidou, A. Pitsillides, S. Ioannidis, and I. Akyildiz. *Using any Surface to Realize a New Paradigm for Wireless Communications*. 2018. arXiv: 1806.04585 [cs.NI].

-
- [Lia+19] Y. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo. "Large Intelligent Surface/Antennas (LISA): Making Reflective Radios Smart". in *Journal of Communications and Information Networks* 4.2 (2019), pp. 40–50. DOI: 10.23919/JCIN.2019.8917871.
- [Lim+17] D. V. de Lima, J. P. C. L. da Costa, F. Antreich, R. K. Miranda, and G. Del Galdo. "Time-Delay estimation via CPD-GEVD applied to tensor-based GNSS arrays with errors". in *Proc. IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. 2017, pp. 1–5. DOI: 10.1109/CAMSAP.2017.8313098.
- [Lin+20] Z. Lin, T. Lv, W. Ni, J. A. Zhang, and R. P. Liu. "Tensor-Based Multi-Dimensional Wideband Channel Estimation for mmWave Hybrid Cylindrical Arrays". *IEEE Transactions on Communications* 68.12 (2020), pp. 7608–7622. DOI: 10.1109/TCOMM.2020.3023934.
- [Liu+18] F. Liu, A. Pitolakis, M. S. Mirmoosa, O. Tsilipakos, X. Wang, A. C. Tasolamprou, S. Abadal, A. Cabellos-Aparicio, E. Alarcón, C. Liaskos, N. V. Kantartzis, M. Kafesaki, E. N. Economou, C. M. Soukoulis, and S. Tretyakov. "Programmable Metasurfaces: State of the Art and Prospects". in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*. 2018, pp. 1–5. DOI: 10.1109/ISCAS.2018.8351817.
- [LL14] J. Lee and Y. Lee. "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing". June 2014, pp. 5838–5842. ISBN: 978-1-4799-2003-7. DOI: 10.1109/ICC.2014.6884253.
- [LW10] W. Liu and S. Weiss. *Wideband Beamforming: Concepts and Techniques*. Wiley Publishing, 2010. ISBN: 0470713925.
- [LW20] Y. Liu and J. Wang. "Low-Complexity OFDM-Based Hybrid Precoding for Multiuser Massive MIMO Systems". *IEEE Wireless Communications Letters* 9.3 (2020), pp. 263–266. DOI: 10.1109/LWC.2019.2929518.
- [LXD14] L. Liang, W. Xu, and X. Dong. "Low-Complexity Hybrid Precoding in Massive Multiuser MIMO Systems". *IEEE Wireless Communications Letters* 3.6 (2014), pp. 653–656. DOI: 10.1109/LWC.2014.2363831.
- [Mac+15] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng. "Indoor Office Wideband Millimeter-Wave Propagation Measurements and Channel Models at 28 and 73 GHz for Ultra-Dense 5G Wireless Networks". *IEEE Access* 3 (2015), pp. 2388–2424. DOI: 10.1109/ACCESS.2015.2486778.
- [Mao+22] Z. Mao, W. Wang, Q. Xia, C. Zhong, X. Pan, and Z. Ye. "Element-Grouping Intelligent Reflecting Surface: Electromagnetic-Compliant Model and Geometry-Based Optimization". *IEEE Transactions on Wireless Communications* 21.7 (2022), pp. 5362–5376. DOI: 10.1109/TWC.2021.3139611.
- [Mar+16] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo. *Fundamentals of Massive MIMO*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316799895.
- [MCD02] B. Muquet, M. de Courville, and P. Duhamel. "Subspace-based blind and semi-blind channel estimation for OFDM systems". *IEEE Transactions on Signal Processing* 50.7 (2002), pp. 1699–1712. DOI: 10.1109/TSP.2002.1011210.

A. Bibliography

- [Med04] A. R. Medhat. "On the Moore–Penrose generalized inverse matrix". *Applied Mathematics and Computation* 158.1 (2004), pp. 185–200. ISSN: 0096-3003. DOI: <https://doi.org/10.1016/j.amc.2003.09.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0096300303009998>.
- [Mén+16] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath. "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?" *IEEE Access* 4 (2016), pp. 247–267. DOI: 10.1109/ACCESS.2015.2514261.
- [MGB08] S. Miron, X. Guo, and D. Brie. "DOA estimation for polarized sources on a vector-sensor array by PARAFAC decomposition of the fourth-order covariance tensor". in *Proc. 16th European Signal Processing Conference*. 2008, pp. 1–5.
- [MH14] J. Mo and R. W. Heath. "High SNR capacity of millimeter wave MIMO systems with one-bit quantization". in *Proc. Information Theory and Applications Workshop (ITA)*. 2014, pp. 1–5. DOI: 10.1109/ITA.2014.6804238.
- [MJ19] D. Mishra and H. Johansson. "Channel Estimation and Low-complexity Beamforming Design for Passive Intelligent Surface Assisted MISO Wireless Energy Transfer". in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 4659–4663. DOI: 10.1109/ICASSP.2019.8683663.
- [MN17] B. Masiero and V. H. Nascimento. "Revisiting the Kronecker Array Transform". *IEEE Signal Processing Letters* 24.5 (2017), pp. 525–529. DOI: 10.1109/LSP.2017.2674969.
- [MRM16] B. Mamandipoor, D. Ramasamy, and U. Madhoo. "Newtonized Orthogonal Matching Pursuit: Frequency Estimation Over the Continuum". *IEEE Transactions on Signal Processing* 64.19 (2016), pp. 5066–5081. DOI: 10.1109/TSP.2016.2580523.
- [Nad+19] Q. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini. *Asymptotic Max-Min SINR Analysis of Reconfigurable Intelligent Surface Assisted MISO Systems*. 2019. arXiv: 1903.08127 [cs.IT].
- [Nad+20] Q. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini. "Intelligent Reflecting Surface-Assisted Multi-User MISO Communication: Channel Estimation and Beamforming Design". *IEEE Open Journal of the Communications Society* 1 (2020), pp. 661–680. DOI: 10.1109/OJCOMS.2020.2992791.
- [ND16] W. Ni and X. Dong. "Hybrid Block Diagonalization for Massive Multiuser MIMO Systems". *IEEE Transactions on Communications* 64.1 (2016), pp. 201–211. DOI: 10.1109/TCOMM.2015.2502954.
- [Neu69] H. Neudecker. "Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products". *Journal of the American Statistical Association* 64.327 (1969), pp. 953–963. DOI: 10.1080/01621459.1969.10501027.
- [NLM13] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems". *IEEE Transactions on Communications* 61.4 (2013), pp. 1436–1449. DOI: 10.1109/TCOMM.2013.020413.110848.

- [NX16] J. Ni and H. Xiao. "Game theoretic approach for joint transmit beamforming and power control in cognitive radio MIMO broadcast channels". *EURASIP Journal on Wireless Communications and Networking* (2016), pp. 1–10. URL: <https://api.semanticscholar.org/CorpusID:36933494>.
- [NYS22] S. Noh, H. Yu, and Y. Sung. "Training Signal Design for Sparse Channel Estimation in Intelligent Reflecting Surface-Assisted Millimeter-Wave Communication". *IEEE Transactions on Wireless Communications* 21.4 (2022), pp. 2399–2413. DOI: 10.1109/TWC.2021.3112173.
- [Pal+06] S. Pallav, N. B. Mehta, A. F. Molisch, and J. Zhang. "Channel Statistics-Based RF Pre-Processing with Antenna Selection". *IEEE Transactions on Wireless Communications* 5.12 (2006), pp. 3501–3511. DOI: 10.1109/TWC.2006.256973.
- [Par+13] J. Park, G. Lee, Y. Sung, and M. Yukawa. "Coordinated Beamforming With Relaxed Zero Forcing: The Sequential Orthogonal Projection Combining Method and Rate Control". *IEEE Transactions on Signal Processing* 61.12 (2013), pp. 3100–3112. DOI: 10.1109/TSP.2013.2258343.
- [Pau+04] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolckei. "An overview of MIMO communications - a key to gigabit wireless". *Proceedings of the IEEE* 92.2 (2004), pp. 198–218. DOI: 10.1109/JPROC.2003.821915.
- [PD21] Y. Pan and Z. Deng. "Channel Estimation For Wireless Communication Systems Aided By Large Intelligent Reflecting Surface". in *Proc. IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. 2021, pp. 637–643. DOI: 10.1109/ICBAIE52039.2021.9389923.
- [Pen55] R. Penrose. "A generalized inverse for matrices". *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3 (1955), pp. 406–413. DOI: 10.1017/S0305004100030401.
- [PF05] D. P. Palomar and J. R. Fonollosa. "Practical algorithms for a family of waterfilling solutions". *IEEE Transactions on Signal Processing* 53.2 (2005), pp. 686–695. DOI: 10.1109/TSP.2004.840816.
- [PNG08] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. 1st. USA: Cambridge University Press, 2008. ISBN: 0521065933.
- [PP06] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. English. Version 20051003. Technical University of Denmark, 2006.
- [PTV23] M. Pesavento, M. Trinh-Hoang, and M. Viberg. "Three More Decades in Array Signal Processing Research: An optimization and structure exploitation perspective". *IEEE Signal Processing Magazine* 40.4 (2023), pp. 92–106. DOI: 10.1109/MSP.2023.3255558.
- [Qia+18] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang. "Tensor-Based Channel Estimation for Dual-Polarized Massive MIMO Systems". *IEEE Transactions on Signal Processing* 66.24 (2018), pp. 6390–6403. DOI: 10.1109/TSP.2018.2873506.
- [Rap+13] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez. "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *IEEE Access* 1 (2013), pp. 335–349. DOI: 10.1109/ACCESS.2013.2260813.

- [Rap+17] T. S. Rappaport, Y. Xing, G. R. Mac Cartney, A. F. Molisch, E. Mellios, and J. Zhang. “Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks—With a Focus on Propagation Models”. *IEEE Transactions on Antennas and Propagation* 65.12 (2017), pp. 6213–6230. DOI: 10.1109/TAP.2017.2734243.
- [RDT22] D. G. Riviello, F. Di Stasio, and R. Tuninato. “Performance Analysis of Multi-User MIMO Schemes under Realistic 3GPP 3-D Channel Model for 5G mmWave Cellular Networks”. *Electronics* 11.3 (2022). ISSN: 2079-9292. DOI: 10.3390/electronics11030330.
- [Reb+16] H. Rebered, F. Renna, R. Calderbank, and M. R. D. Rodrigues. “Bounds on the Number of Measurements for Reliable Compressive Classification”. *IEEE Transactions on Signal Processing* 64.22 (Nov. 2016), pp. 5778–5793. DOI: 10.1109/tsp.2016.2599496.
- [Ren+19] M. Di Renzo, M. Debbah, D. Phan-Huy, A. Zappone, M. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. de Rosny, A. Bounceu, G. Lerosey, and M. Fink. *Smart Radio Environments Empowered by AI Reconfigurable Meta-Surfaces: An Idea Whose Time Has Come*. 2019. arXiv: 1903.08925 [cs.IT].
- [RH09] F. Roemer and M. Haardt. “Tensor-based channel estimation (TENCE) for two-way relaying with multiple antennas and spatial reuse”. in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009, pp. 3641–3644. DOI: 10.1109/ICASSP.2009.4960415.
- [RH10] F. Roemer and M. Haardt. “Tensor-Based Channel Estimation and Iterative Refinements for Two-Way Relaying With Multiple Antennas and Spatial Reuse”. *IEEE Transactions on Signal Processing* 58.11 (2010), pp. 5720–5735. DOI: 10.1109/TSP.2010.2062179.
- [Roe13] F. Roemer. “Advanced Algebraic Concepts for Efficient Multi-Channel Signal Processing”. PhD thesis. Ilmenau University of Technology, Communication Research Laboratory, Germany, 2013.
- [Rom+15] B. Romanous, N. Bitar, A. Imran, and H. Refai. “Network densification: Challenges and opportunities in enabling 5G”. in *Proc. IEEE 20th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*. 2015, pp. 129–134. DOI: 10.1109/CAMAD.2015.7390494.
- [RSH12] F. Roemer, C. Schroeter, and M. Haardt. “A semi-algebraic framework for approximate CP decompositions via joint matrix diagonalization and generalized unfoldings”. in *Proc. Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. 2012, pp. 2023–2027. DOI: 10.1109/ACSSC.2012.6489396.
- [Rub+19] L. Rubio, Vicent M. Rodrigo P., J. Molina-García-Pardo, L. Juan-Llácer, J. Pascual-García, J. Reig, and C. Sanchis-Borrás. “Millimeter Wave Channel Measurements in an Intra-Wagon Environment”. *IEEE Transactions on Vehicular Technology* 68.12 (2019), pp. 12427–12431. DOI: 10.1109/TVT.2019.2947205.
- [Rub+23] L. Rubio, V. M. R. Peñarrocha, M. Cabedo-Fabres, B. Bernardo-Clemente, J. Reig, H. Fernández, J. R. Pérez, R. P. Torres, L. Valle, and Ó. Fernández. “Millimeter-Wave Channel Measurements and Path Loss Characterization in a Typical Indoor Office Environment”. *Electronics* 12.4 (2023). ISSN: 2079-9292.
- [Rud87] W. Rudin. *Real and complex analysis, 3rd ed.* USA: McGraw-Hill, Inc., 1987. ISBN: 0070542341.

- [Rus+15] C. Rusu, R. Méndez-Rial, N. González-Prelcicy, and R. W. Heath. “Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems”. in *Proc. IEEE International Conference on Communications (ICC)*. 2015, pp. 1340–1345. DOI: 10.1109/ICC.2015.7248509.
- [SA10] A. Stegeman and A. L. F. de Almeida. “Uniqueness Conditions for Constrained Three-Way Factor Decompositions with Linearly Dependent Loadings”. *SIAM Journal on Matrix Analysis and Applications* 31.3 (2010), pp. 1469–1490. DOI: 10.1137/080743354.
- [SC12] B. L. Sturm and M. G. Christensen. “Comparison of orthogonal matching pursuit implementations”. in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. 2012, pp. 220–224.
- [She+19] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao. “Secrecy Rate Maximization for Intelligent Reflecting Surface Assisted Multi-Antenna Communications”. *IEEE Communications Letters* 23.9 (2019), pp. 1488–1492. DOI: 10.1109/LCOMM.2019.2924214.
- [Shi+18] M. Shirichian, S. Chamaani, A. Akbarpour, and G. Del Galdo. “Analysis and Design of Broadband Simultaneous Wireless Information and Power Transfer (SWIPT) System Considering Rectifier Effect”. *Energies* 11.9 (2018). ISSN: 1996-1073. DOI: 10.3390/en11092387.
- [Sid+17] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. “Tensor Decomposition for Signal Processing and Machine Learning”. *IEEE Transactions on Signal Processing* 65.13 (2017), pp. 3551–3582. DOI: 10.1109/TSP.2017.2690524.
- [Son+15] J. Song, J. Choi, S. G. Larew, D. J. Love, T. A. Thomas, and A. A. Ghosh. “Adaptive Millimeter Wave Beam Alignment for Dual-Polarized MIMO Systems”. *IEEE Transactions on Wireless Communications* 14.11 (2015), pp. 6283–6296. DOI: 10.1109/TWC.2015.2452263.
- [SPP18] C. Steffens, M. Pesavento, and M. E. Pfetsch. “A Compact Formulation for the $\ell_{2,1}$ Mixed-Norm Minimization Problem”. *IEEE Transactions on Signal Processing* 66.6 (2018), pp. 1483–1497. DOI: 10.1109/TSP.2017.2788431.
- [SR15] J. Singh and S. Ramakrishna. “On the Feasibility of Codebook-Based Beamforming in Millimeter Wave Systems With Multiple Antenna Arrays”. *IEEE Transactions on Wireless Communications* 14.5 (2015), pp. 2670–2683. DOI: 10.1109/TWC.2015.2390637.
- [Sri+04] V. Srivastava, C. Keong Ho, P. Ho Wang Fung, and S. Sun. “Robust MMSE channel estimation in OFDM systems with practical timing synchronization”. in *Proc. IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733)*. Vol. 2. 2004, 711–716 Vol.2. DOI: 10.1109/WCNC.2004.1311273.
- [SSS09] M. Salazar-Palma, T. K. Sarkar, and D. Sengupta. “A brief chronology of the origin and developments of wireless communication and supporting electronics”. in *Proc. IEEE Applied Electromagnetics Conference (AEMC)*. 2009, pp. 1–4. DOI: 10.1109/AEMC.2009.5430591.
- [Sun+19] W. Sun, L. Huang, H.C. So, and J. Wang. “Orthogonal tubal rank-1 tensor pursuit for tensor completion”. *Signal Processing* 157 (2019), pp. 213–224. ISSN: 0165-1684. DOI: <https://doi.org/10.1016/j.sigpro.2018.11.015>.
- [SV87] A. A. M. Saleh and R. Valenzuela. “A Statistical Model for Indoor Multipath Propagation”. *IEEE Journal on Selected Areas in Communications* 5.2 (1987), pp. 128–137. DOI: 10.1109/JSAC.1987.1146527.

A. Bibliography

- [SY16] F. Sohrabi and W. Yu. "Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays". *IEEE Journal of Selected Topics in Signal Processing* 10.3 (2016), pp. 501–513. DOI: 10.1109/JSTSP.2016.2520912.
- [SY17] F. Sohrabi and W. Yu. "Hybrid Analog and Digital Beamforming for mmWave OFDM Large-Scale Antenna Arrays". *IEEE Journal on Selected Areas in Communications* 35.7 (2017), pp. 1432–1443. DOI: 10.1109/JSAC.2017.2698958.
- [TAA21] A. Taha, M. Alrabeiah, and A. Alkhateeb. "Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning". *IEEE Access* 9 (2021), pp. 44304–44321. DOI: 10.1109/ACCESS.2021.3064073.
- [Tel00] I. Telatar. "Capacity of Multi-Antenna Gaussian Channels". 10 (Apr. 2000).
- [The+22] P. I. Theoharis, R. Raad, F. Tubbal, M. U. A. Khan, and A. Jamalipour. "Wideband Reflectarrays for 5G/6G: A Survey". *IEEE Open Journal of Antennas and Propagation* 3 (2022), pp. 871–901. DOI: 10.1109/OJAP.2022.3195665.
- [TJC99] V. Tarokh, H. Jafarkhani, and A.R. Calderbank. "Space-time block codes from orthogonal designs". *IEEE Transactions on Information Theory* 45.5 (1999), pp. 1456–1467. DOI: 10.1109/18.771146.
- [Tri+21] S. Tripathi, N. V. Sabu, A. K. Gupta, and H. S. Dhillon. "Millimeter-Wave and Terahertz Spectrum for 6G Wireless". *6G Mobile Wireless Networks*. Cham: Springer International Publishing, 2021, pp. 83–121.
- [TS21] G. Tian and R. Song. "Cooperative beamforming for a double-IRS-assisted wireless communication system". *EURASIP Journal on Advances in Signal Processing* 2021 (Aug. 2021). DOI: 10.1186/s13634-021-00780-0.
- [Tuc66] L. R. Tucker. "Some mathematical notes on three-mode factor analysis". *Psychometrika* 31 (1966), pp. 279–311. URL: <https://api.semanticscholar.org/CorpusID:44301099>.
- [UV18] K. Upadhyya and S. A. Vorobyov. "Covariance Matrix Estimation for Massive MIMO". *IEEE Signal Processing Letters* 25.4 (2018), pp. 546–550. DOI: 10.1109/LSP.2018.2805725.
- [VP93] C. F. Van Loan and N. Pitsianis. *Approximation with Kronecker Products*. Ed. by Marc S. Moonen, Gene H. Golub, and Bart L. R. De Moor. Dordrecht: Springer Netherlands, 1993, pp. 293–314. ISBN: 978-94-015-8196-7.
- [VS13] J. P. Vila and P. Schniter. "Expectation-Maximization Gaussian-Mixture Approximate Message Passing". *IEEE Transactions on Signal Processing* 61.19 (2013), pp. 4658–4672. DOI: 10.1109/TSP.2013.2272287.
- [Wan+09] J. Wang, Z. Lan, C. Pyo, T. Baykas, C. Sum, M.A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato. "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems". *IEEE Journal on Selected Areas in Communications* 27.8 (2009), pp. 1390–1399. DOI: 10.1109/JSAC.2009.091009.
- [Wan+18] B. Wang, F. Gao, S. Jin, H. Lin, G. Ye Li, S. Sun, and T. S. Rappaport. "Spatial-Wideband Effect in Massive MIMO with Application in mmWave Systems". *IEEE Communications Magazine* 56.12 (2018), pp. 134–141. DOI: 10.1109/MCOM.2018.1701051.

-
- [Wan+20] P. Wang, J. Fang, H. Duan, and H. Li. "Compressed Channel Estimation for Intelligent Reflecting Surface-Assisted Millimeter Wave Systems". *IEEE Signal Processing Letters* 27 (2020), pp. 905–909. DOI: 10.1109/LSP.2020.2998357.
- [Wan+21] J. Wang, W. Zhang, J. Sun, and C. Wang. "Tensor-Based Channel Estimation for 3D mmWave Massive MIMO Systems". in *Proc. IEEE International Conference on Communications in China (ICCC)*. 2021, pp. 517–522. DOI: 10.1109/ICCC52777.2021.9580338.
- [Wan+22] M. Wang, W. Duan, G. Zhang, M. Wen, J. Choi, and P. Ho. "On the Achievable Capacity of Cooperative NOMA Networks: RIS or Relay?" *IEEE Wireless Communications Letters* 11.8 (2022), pp. 1624–1628. DOI: 10.1109/LWC.2022.3169806.
- [WDV23] S. Wesemann, J. Du, and H. Viswanathan. "Energy Efficient Extreme MIMO: Design Goals and Directions". *IEEE Communications Magazine* 61.10 (Oct. 2023), pp. 132–138. ISSN: 1558-1896. DOI: 10.1109/mcom.004.2200958.
- [WGA20] Z. Wan, Z. Gao, and M. Alouini. "Broadband Channel Estimation for Intelligent Reflecting Surface Aided mmWave Massive MIMO Systems". in *Proc. IEEE International Conference on Communications (ICC)*. 2020, pp. 1–6. DOI: 10.1109/ICC40277.2020.9149146.
- [WLC20] Z. Wang, L. Liu, and S. Cui. "Channel Estimation for Intelligent Reflecting Surface Assisted Multiuser Communications: Framework, Algorithms, and Analysis". *IEEE Transactions on Wireless Communications* 19.10 (2020), pp. 6607–6620. DOI: 10.1109/TWC.2020.3004330.
- [WLY18] X. Wu, D. Liu, and F. Yin. "Hybrid Beamforming for Multi-User Massive MIMO Systems". *IEEE Transactions on Communications* 66.9 (2018), pp. 3879–3891. DOI: 10.1109/TCOMM.2018.2829511.
- [WSD21] X. Wei, D. Shen, and L. Dai. "Channel Estimation for RIS Assisted Wireless Communications—Part II: An Improved Solution Based on Double-Structured Sparsity". *IEEE Communications Letters* 25.5 (2021), pp. 1403–1407. DOI: 10.1109/LCOMM.2021.3052787.
- [Wu+16] K. K. Wu, Y. Yam, H. Meng, and M. Mesbahi. "Kronecker product approximation with multiple factor matrices via the tensor product algorithm". in *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2016, pp. 004277–004282. DOI: 10.1109/SMC.2016.7844903.
- [Wu+17] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober. "An Overview of Sustainable Green 5G Networks". *IEEE Wireless Communications* 24.4 (2017), pp. 72–80. DOI: 10.1109/MWC.2017.1600343.
- [Wu+21] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang. "Intelligent Reflecting Surface-Aided Wireless Communications: A Tutorial". *IEEE Transactions on Communications* 69.5 (2021), pp. 3313–3351. DOI: 10.1109/TCOMM.2021.3051897.
- [WZ18] Q. Wu and R. Zhang. "Intelligent Reflecting Surface Enhanced Wireless Network: Joint Active and Passive Beamforming Design". in *Proc. IEEE Global Communications Conference (GLOBECOM)*. 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647620.
- [WZ19] Q. Wu and R. Zhang. "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming". *IEEE Transactions on Wireless Communications* 18.11 (2019), pp. 5394–5409. DOI: 10.1109/TWC.2019.2936025.
-

A. Bibliography

- [XFA16] L. R. Ximenes, G. Favier, and A. L. F. de Almeida. "Closed-Form Semi-Blind Receiver For MIMO Relay Systems Using Double Khatri–Rao Space-Time Coding". *IEEE Signal Processing Letters* 23.3 (2016), pp. 316–320. DOI: 10.1109/LSP.2016.2518699.
- [Xie+18] H. Xie, F. Gao, S. Jin, J. Fang, and Y. Liang. "Channel Estimation for TDD/FDD Massive MIMO Systems With Channel Covariance Computing". *IEEE Transactions on Wireless Communications* 17.6 (2018), pp. 4206–4218. DOI: 10.1109/TWC.2018.2821667.
- [XXX17] Z. Xiao, P. Xia, and X. Xia. "Codebook Design for Millimeter-Wave Channel Estimation With Hybrid Precoding Structure". *IEEE Transactions on Wireless Communications* 16.1 (2017), pp. 141–153. DOI: 10.1109/TWC.2016.2619705.
- [Yan+20a] L. Yang, F. Meng, Q. Wu, D. B. da Costa, and M. Alouini. "Accurate Closed-Form Approximations to Channel Distributions of RIS-Aided Wireless Systems". *IEEE Wireless Communications Letters* 9.11 (2020), pp. 1985–1989. DOI: 10.1109/LWC.2020.3010512.
- [Yan+20b] L. Yang, X. Yan, D. B. da Costa, T. A. Tsiftsis, H. Yang, and M. Alouini. "Indoor Mixed Dual-Hop VLC/RF Systems Through Reconfigurable Intelligent Surfaces". *IEEE Wireless Communications Letters* 9.11 (2020), pp. 1995–1999. DOI: 10.1109/LWC.2020.3010809.
- [Yan+20c] L. Yang, J. Yang, W. Xie, M. O. Hasna, T. Tsiftsis, and M. D. Renzo. "Secrecy Performance Analysis of RIS-Aided Wireless Communication Systems". *IEEE Transactions on Vehicular Technology* 69.10 (2020), pp. 12296–12300. DOI: 10.1109/TVT.2020.3007521.
- [Yan+20d] Y. Yang, B. Zheng, S. Zhang, and R. Zhang. "Intelligent Reflecting Surface Meets OFDM: Protocol Design and Rate Maximization". *IEEE Transactions on Communications* 68.7 (2020), pp. 4522–4535. DOI: 10.1109/TCOMM.2020.2981458.
- [Yan+21] L. Yang, Y. Yang, D. B. da Costa, and I. Trigui. "Outage Probability and Capacity Scaling Law of Multiple RIS-Aided Networks". *IEEE Wireless Communications Letters* 10.2 (2021), pp. 256–260. DOI: 10.1109/LWC.2020.3026712.
- [Yan00] Xiaojing Yang. "A Matrix Trace Inequality". *Journal of Mathematical Analysis and Applications* 250.1 (2000), pp. 372–374. ISSN: 0022-247X. DOI: <https://doi.org/10.1006/jmaa.2000.7068>. URL: <https://www.sciencedirect.com/science/article/pii/S0022247X00970689>.
- [Yil+21] I. Yildirim, F. Kilinc, E. Basar, and G. C. Alexandropoulos. "Hybrid RIS-Empowered Reflection and Decode-and-Forward Relaying for Coverage Extension". *IEEE Communications Letters* 25.5 (2021), pp. 1692–1696. DOI: 10.1109/LCOMM.2021.3054819.
- [Yu+16] X. Yu, J. Shen, J. Zhang, and K. B. Letaief. "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems". *IEEE Journal of Selected Topics in Signal Processing* 10.3 (2016), pp. 485–500. DOI: 10.1109/JSTSP.2016.2523903.
- [YXS19] X. Yu, D. Xu, and R. Schober. "Enabling Secure Wireless Communications via Intelligent Reflecting Surfaces". in *Proc. IEEE Global Communications Conference (GLOBECOM)*. 2019, pp. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9014322.
- [YYK20] W. Yan, X. Yuan, and X. Kuai. "Passive Beamforming and Information Transfer via Large Intelligent Surface". *IEEE Wireless Communications Letters* 9.4 (2020), pp. 533–537. DOI: 10.1109/LWC.2019.2961670.

-
- [YZZ20] C. You, B. Zheng, and R. Zhang. "Channel Estimation and Passive Beamforming for Intelligent Reflecting Surface: Discrete Phase Shift and Progressive Refinement". *IEEE Journal on Selected Areas in Communications* 38.11 (2020), pp. 2604–2620. DOI: 10.1109/JSAC.2020.3007056.
- [YZZ21] C. You, B. Zheng, and R. Zhang. "Wireless Communication via Double IRS: Channel Estimation and Passive Beamforming Designs". *IEEE Wireless Communications Letters* 10.2 (2021), pp. 431–435. DOI: 10.1109/LWC.2020.3034388.
- [ZH17a] J. Zhang and M. Haardt. "Channel estimation and training design for hybrid multi-carrier MmWave massive MIMO systems: The beamspace ESPRIT approach". in *Proc. 25th European Signal Processing Conference (EUSIPCO)*. 2017, pp. 385–389. DOI: 10.23919/EUSIPCO.2017.8081234.
- [ZH17b] J. Zhang and M. Haardt. "Channel estimation for hybrid multi-carrier mmwave MIMO systems using three-dimensional unitary ESPRIT in DFT beamspace". in *Proc. IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. 2017, pp. 1–5. DOI: 10.1109/CAMSAP.2017.8313174.
- [Zha+17] S. Zhang, Q. Wu, S. Xu, and G. Y. Li. "Fundamental Green Tradeoffs: Progresses, Challenges, and Impacts on 5G Networks". *IEEE Communications Surveys and Tutorials* 19.1 (2017), pp. 33–56. DOI: 10.1109/COMST.2016.2594120.
- [Zha+18] D. Zhang, Y. Wang, X. Li, and W. Xiang. "Hybridly Connected Structure for Hybrid Beamforming in mmWave Massive MIMO Systems". *IEEE Transactions on Communications* 66.2 (2018), pp. 662–674. DOI: 10.1109/TCOMM.2017.2756882.
- [Zha+19] D. Zhang, Y. Wang, X. Li, and W. Xiang. "Hybrid beamforming for downlink multiuser millimeter wave MIMO-OFDM systems". *IET Communications* 13.11 (2019), pp. 1557–1564.
- [Zha+20] J. Zhang, C. Qi, P. Li, and P. Lu. "Channel Estimation for Reconfigurable Intelligent Surface Aided Massive MIMO System". in *Proc. IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2020, pp. 1–5. DOI: 10.1109/SPAWC48557.2020.9154276.
- [Zho+17] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum. "Low-Rank Tensor Decomposition-Aided Channel Estimation for Millimeter Wave MIMO-OFDM Systems". *IEEE Journal on Selected Areas in Communications* 35.7 (2017), pp. 1524–1538. DOI: 10.1109/JSAC.2017.2699338.
- [Zho+22] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst. "Channel Estimation for RIS-Aided Multiuser Millimeter-Wave Systems". *IEEE Transactions on Signal Processing* 70 (2022), pp. 1478–1492. DOI: 10.1109/TSP.2022.3158024.
- [ZMK05] X. Zhang, A. F. Molisch, and S. Kung. "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection". *IEEE Transactions on Signal Processing* 53.11 (2005), pp. 4091–4103. DOI: 10.1109/TSP.2005.857024.
- [ZN04] Y. Zeng and T. Ng. "A Semi-Blind Channel Estimation Method for Multiuser Multiantenna OFDM Systems". *IEEE Trans. Sig. Proc.* 52.5 (May 2004), pp. 1419–1429. ISSN: 1053-587X. DOI: 10.1109/TSP.2004.826183. URL: <https://doi.org/10.1109/TSP.2004.826183>.
-

A. Bibliography

- [ZP06] H. Zamiri-Jafarian and S. Pasupathy. "MIMO-OFDM Channel Estimation Using Improved LS Algorithm". in *Proc. IEEE Vehicular Technology Conference*. 2006, pp. 1–5. DOI: 10.1109/VTCF.2006.122.
- [ZRH21] J. Zhang, D. Rakhimov, and M. Haardt. "Gridless Channel Estimation for Hybrid mmWave MIMO Systems via Tensor-ESPRIT Algorithms in DFT Beamspace". *IEEE Journal of Selected Topics in Signal Processing* 15.3 (2021), pp. 816–831. DOI: 10.1109/JSTSP.2021.3063908.
- [ZT03] L. Zheng and D.N.C. Tse. "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels". *IEEE Transactions on Information Theory* 49.5 (2003), pp. 1073–1096. DOI: 10.1109/TIT.2003.810646.
- [ZWH16] J. Zhang, A. Wiesel, and M. Haardt. "Low rank approximation based hybrid precoding schemes for multi-carrier single-user massive MIMO systems". in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2016), pp. 3281–3285.
- [ZYZ21a] B. Zheng, C. You, and R. Zhang. "Double-IRS Assisted Multi-User MIMO: Cooperative Passive Beamforming Design". *IEEE Transactions on Wireless Communications* 20.7 (June 2021), pp. 4513–4526. DOI: 10.1109/twc.2021.3059945.
- [ZYZ21b] B. Zheng, C. You, and R. Zhang. *Efficient Channel Estimation for Double-IRS Aided Multi-User MIMO System*. 2021. arXiv: 2011.00738 [cs.IT].
- [ZYZ21c] B. Zheng, C. You, and R. Zhang. "Uplink Channel Estimation for Double-IRS Assisted Multi-User MIMO". in *Proc. IEEE International Conference on Communications (ICC)*. 2021, pp. 1–6. DOI: 10.1109/ICC42927.2021.9501057.
- [ZZ20a] S. Zhang and R. Zhang. "Capacity Characterization for Intelligent Reflecting Surface Aided MIMO Communication". *IEEE Journal on Selected Areas in Communications* 38.8 (2020), pp. 1823–1838. DOI: 10.1109/JSAC.2020.3000814.
- [ZZ20b] B. Zheng and R. Zhang. "Intelligent Reflecting Surface-Enhanced OFDM: Channel Estimation and Reflection Optimization". *IEEE Wireless Communications Letters* 9.4 (2020), pp. 518–522. DOI: 10.1109/LWC.2019.2961357.