

Hybrid Beamforming Techniques for Massive MIMO Full Duplex Radio Systems

Dissertation

submitted to

Sorbonne Université

*in partial fulfillment of the requirements for the degree of
Doctor of Philosophy*

Author:

Chandan Kumar Sheemar

Supervisor: Prof. Dirk Slock

Scheduled for defense on the 28 March, 2022, before a committee composed of:

Reviewers

Prof.	Tharmalingam Ratnarajah	University of Edinburgh, United Kingdom
Prof.	Taneli Riihonen	Tampere University, Finland

Examiners

Prof	Constantinos B. Papadias	The American College of Greece, Greece
Prof.	Laura Cottatellucci	Friedrich Alexander University, Germany
Prof.	George C. Alexandropoulos	University of Athens, Greece
Prof.	Marios Kountouris	EURECOM, France
Prof.	Dirk Slock	EURECOM, France

Techniques de Beamforming Hybride pour les Systèmes Radio Massivement MIMO en Full Duplex

Dissertation

submitted to

Sorbonne Université

pour l'obtention du Grade de Docteur

Author:

Chandan Kumar Sheemar

Thèse dirigée par Dirk Slock, Professeur, EURECOM, France

Scheduled for defense on the 28 March, 2022, before a committee composed of:

Rapporteur

Prof.	Tharmalingam Ratnarajah	University of Edinburgh, United Kingdom
Prof.	Taneli Riihonen	Tampere University, Finland

Jury

Prof.	Constantinos B. Papadias	The American College of Greece, Greece
Prof.	Laura Cottatellucci	Friedrich Alexander University, Germany
Prof.	George C. Alexandropoulos	University of Athens, Greece
Prof.	Marios Kountouris	EURECOM, France
Prof.	Dirk Slock	EURECOM, France

Abstract

Full Duplex (FD) radio has emerged as a promising solution to increase the data rates by up to a factor of two via simultaneous transmission and reception in the same frequency band. Self-interference (SI) is a significant challenge to deal with to achieve an ideal FD operation, which could be 90 – 100 dB higher than the received signal of interest. SI cancellation (SIC) techniques can mitigate the SI signal in the propagation domain or perform active cancellation by subtracting a copy of the transmitted signal on the receive side. Beamforming is also a potent tool for FD systems, which can mitigate the SI signal while meeting the data traffic requirements. It can be distinguished into two main categories: *fully digital beamforming* or *hybrid beamforming (HYBF)*. The former is helpful for the traditional multiple-input-multiple-output (MIMO) FD systems with a limited number of antennas, typically deployed in sub-6 GHz. The latter can be used for the massive MIMO (mMIMO) systems in millimeter wave (mmWave) band, such that they can be built cost-efficiently with a fewer number of radio-frequency (RF) chains. This thesis aims to present several digital and HYBF designs for FD systems, starting from very simple and then covering the most challenging scenarios for FD systems. Moreover, a novel and scalable SI architecture is also presented, promising for the mMIMO FD systems. One of the major drawbacks of the literature on multi-user FD communication systems is that it does not contain any distributed solution. As the FD paradigm shifts towards the mmWave band, several new challenges arise, which can make the implementation of the centralized HYBF designs infeasible, especially in a large or/and dense multi-cell network, for which a mathematical analysis is also presented. To make FD feasible in large or/and dense networks, we present the concept of per-link parallel and distributed (P&D) HYBF by introducing the first-ever P&D algorithm for mmWave, which enables parallel optimization of the beamformers at the multi-processor FD base stations and has very low complexity. Intelligent reflecting surfaces (IRSs) are prominent for the next generation of wireless communication systems. In the final part of this thesis, the concept of near-field IRSs for the mmWave FD systems is introduced to leverage the full potential of FD operation while drastically reducing the hardware cost and minimizing power consumption. The multi-cell dynamic time division duplexing (DTDD) systems are a particular case of the multi-cell FD systems. This thesis concludes with a novel contribution to DTDD by proposing a fully distributed and low complexity solution for the joint scheduling and power allocation problem, which in practice results in being NP-hard.

Abrégé

La radio Full Duplex (FD) est apparue comme une solution prometteuse pour augmenter les débits de données jusqu'à un facteur de deux via une transmission et une réception simultanées dans la même bande de fréquences. L'auto-interférence (SI) est un défi important à relever pour obtenir un fonctionnement FD idéal, qui pourrait être supérieur de 90 à 100 dB au signal d'intérêt reçu. Les techniques d'annulation SI atténuent le signal SI dans le domaine de propagation ou le soustraient du côté réception en utilisant une copie du signal transmis. Beamforming est également un outil puissant pour les systèmes FD, qui peut atténuer le signal SI tout en répondant aux exigences du trafic de données. Il peut être distingué en *beamforming entièrement digitale* ou *beamforming hybride (HYBF)*. Le premier est utile pour les systèmes FD traditionnels à entrées multiples et sorties multiples (MIMO) avec un nombre limité d'antennes, généralement déployées à moins de 6 GHz. Ce dernier peut être utilisé pour les systèmes MIMO massifs (mMIMO) dans la bande des ondes millimétriques (mmWave), de sorte qu'ils peuvent être construits de manière rentable avec un nombre réduit de chaînes radiofréquence (RF). Cette thèse vise à présenter plusieurs conceptions numériques et HYBF pour les systèmes FD, en partant de très simples et en couvrant ensuite les scénarios les plus difficiles pour les systèmes FD. De plus, une nouvelle architecture SI évolutive est également présentée, prometteuse pour les systèmes mMIMO FD. L'un des inconvénients majeurs de la littérature sur les systèmes de communication FD multi-utilisateurs est qu'elle ne contient aucune solution distribuée. Alors que le paradigme FD évolue vers la bande mmWave, plusieurs nouveaux défis se posent, ce qui peut rendre impossible la mise en œuvre des conceptions HYBF centralisées, en particulier dans un réseau multicellulaire large ou/et dense, pour lequel une analyse mathématique est également présentée. Pour rendre la FD réalisable dans les réseaux étendus ou/et denses, nous présentons le concept de HYBF parallèle et distribué par lien (P&D) en introduisant le tout premier algorithme P&D pour mmWave, qui permet optimisation parallèle des formateurs de faisceaux au niveau des stations de base FD multiprocesseurs et a une très faible complexité. Les surfaces réfléchissantes intelligentes (IRS) sont importantes pour la prochaine génération de systèmes de communication sans fil. Dans la dernière partie de cette thèse, le concept d'IRS en champ proche pour les systèmes MIMO FD est introduit, ce qui permet d'exploiter au maximum le plein potentiel des systèmes FD tout en réduisant considérablement le coût du matériel et en minimisant la consommation d'énergie. Les systèmes multicellulaires à dynamic time division duplexing (DTDD) sont un cas particulier des systèmes FD multicellulaires. Cette thèse se termine par une nouvelle contribution à DTDD en proposant une solution

entièrement distribuée et de faible complexité pour le problème conjoint d'ordonnancement et d'allocation de puissance, qui se traduit en pratique par être NP-difficile.

Acknowledgements

In this world, people have their own fights and challenges, and not everybody gets the opportunity to achieve their life goals. I feel humbled and honoured that I am able to achieve this target in my life. Growing up I had no intention to go to university for higher education but somehow ended up doing a PhD. It is because of some amazing Professors that I had the opportunity to learn from, who made the learning process so much fun. The list of Professors that I would like to thank for their teaching, advice and inspiration is very long. Given that low complexity is desirable in engineering, I opt to keep things simple with very low complexity.

First of all, I would like to thank my teacher during high school, Mr. Mauro Giacon, for introducing me to the field of telecommunication. Then, during the five and half years that I spent at the University of Padua, Italy, first as a bachelor and then as a master student, I had no intention to pursue a career in research. However, somehow I got the opportunity to do my master thesis with Prof. Stefano Tomasin. I remember spending a lot of time with him in his office and discussing how to solve the problems that I had formulated for my master thesis. That was the exact period when I started becoming interested in doing research. He guided me during the first steps in research and has always been very supportive. Because of his and my girlfriend Sunny's motivation, I dedicated to look for a PhD opportunity. I must say, I got fortunate that I received the opportunity to do a PhD at EURECOM, under the supervision of Prof. Dirk Slock. I want to thank him not only for the skills that he has taught me but also for the random discussions we had, for teaching me the importance of being independent in research, for always being there for me, for sharpening my vision as a researcher and for teaching me that no research problem is impossible to solve.

I want to also thank many friends that I met at EURECOM, during these three year and especially my friend Christo for always helping me with every single thing. When I joined EURECOM, in the first year he spent a lot of time in teaching me so many thing and I am very thankful of that. I would like to also thank my parents for their continuous support. Finally, the main emotional and moral support I had during this PhD was from amazing girlfriend Sunny, who always supported and motivated me no matter what, and I dedicate this thesis to her.

Contents

Abstract	i
Abrégé [Français]	iii
Acknowledgements	v
Contents	vii
List of Figures	x
List of Tables	xv
Acronyms	xvii
Notations	xix
1 Introduction	1
1.1 Cellular Networks	1
1.1.1 Half Duplex Systems	2
1.1.2 Full Duplex Systems	3
1.2 Contributions and Thesis Outline	5
1.2.1 Journals	6
1.2.2 Conferences	7
1.2.3 Patents	7
2 On the SISO and MIMO Full Duplex Systems in Sub-6 GHz	9
2.1 Signal Reconstruction of Saturated Signals in SISO FD	9
2.1.1 State-of-the-Art and Motivation	9
2.1.2 System Model	10
2.1.3 Simulation Results	12
2.1.4 Large System Analysis of the MSE	14
2.1.5 Conclusions	17
2.2 Beamforming for MIMO Full Duplex Under the Joint Sum-Power and Per-Antenna Power Constraints	17
2.2.1 System Model	18
2.2.2 Digital Beamforming	20
2.2.3 Convergence Proof	21
2.2.4 Simulation Results	22

2.2.5	Conclusions	23
2.3	Towards Massive MIMO In-Band Full Duplex Radio	23
2.3.1	MIMO D2RF HSIC Architecture	24
2.3.2	System Model	25
2.3.3	MIMO SI Channel Estimation	25
2.3.4	Zero Forcing HSIC Solution	26
2.3.5	Self-Interference Nulling in MIMO Full Duplex	27
2.3.6	RF Calibration and LoS Near Field MIMO Channel Model	27
2.3.7	Reduced Complexity Analog SI Nulling	29
2.3.8	Simulation and Measurement Results	31
2.3.9	RF Calibration Results	31
2.3.10	Reduced Complexity SIN Results	32
2.3.11	Conclusions	35
3	Hybrid Beamforming for Point-to-Point and IAB Millimeter Wave Full Duplex	37
3.1	Full Duplex Relay for IAB in Millimeter Wave	37
3.1.1	System Model	38
3.1.2	Problem Simplification	40
3.1.3	Hybrid Beamforming	41
3.1.4	Convergence	43
3.1.5	Simulation Results	43
3.2	Multi-Stage Hybrid Beamforming for Point-to-Point OFDM MIMO Full Duplex	46
3.2.1	System Model	46
3.2.2	Optimization of the Beamformers	48
3.2.3	Two stage Transmit Beamforming	50
3.2.4	Simulation Results	53
3.2.5	Conclusion	55
3.3	Generalization of the point-to-point mMIMO Full-Duplex Communication in Millimeter Wave	55
3.3.1	System Model	56
3.3.2	Minorization-Maximization	58
3.3.3	Hybrid Beamforming and Combining	59
3.3.4	Simulation Results	64
3.3.5	Conclusions	64
4	Hybrid Beamforming for Single-Cell Massive MIMO Millimeter Full Duplex	67
4.1	Introduction and Motivation	67
4.1.1	Main Contributions	68
4.2	System Model	69

4.2.1	Problem Formulation	72
4.3	Minorization-Maximization	74
4.3.1	Concave Reformulation	76
4.4	Hybrid Beamforming and Combining	76
4.4.1	Digital Beamforming	77
4.4.2	Analog Beamforming	77
4.4.3	Analog Combining	79
4.4.4	Optimal Power Allocation	80
4.4.5	Convergence	82
4.4.6	Complexity Analysis	84
4.5	Simulation Results	85
4.6	Conclusions	90
5	Hybrid Beamforming for Millimeter Wave Multi-Cell Massive MIMO Full Duplex	91
5.1	Introduction and Motivation	91
5.1.1	Main Contributions	92
5.2	System Model	93
5.2.1	Channel Modelling	94
5.2.2	Problem Formulation	95
5.3	Minorization-Maximization	98
5.4	Centralized Hybrid Beamforming	100
5.4.1	Digital Beamforming	100
5.4.2	Analog Beamforming	101
5.4.3	Analog Combining	102
5.4.4	Optimal Power Allocation	103
5.4.5	Convergence of C-HYBF	105
5.5	Parallel and Distributed Implementation	107
5.5.1	Per-Link Independent Sub-Problems in UL	109
5.5.2	Per-Link Independent Sub-Problems in DL	111
5.5.3	On the Convergence of P&D-HYBF	114
5.5.4	Computational Complexity Analysis	116
5.6	Simulation Results	118
5.7	Conclusions	124
6	Near-Field Intelligent Reflecting Surfaces for Millimeter Wave MIMO Full Duplex	125
6.1	System Model	126
6.1.1	Channel Modelling	127
6.1.2	Problem Formulation	130
6.2	Beamforming	131
6.2.1	Active Digital Beamforming	132

6.2.2	Passive Beamforming with the IRSs	133
6.2.3	Convergence	136
6.2.4	Simulation Results	137
6.3	Conclusions	141
7	Dynamic TDD: A Special Case of Multi-Cell Full Duplex Systems	143
7.1	Introduction and Motivation	143
7.2	System Model	144
7.2.1	Scheduling and Power Allocation Problem	145
7.3	Scheduling and Power Allocation Decoupling	146
7.3.1	Average Interference Power	147
7.3.2	Power Allocation Sub-problem	148
7.3.3	Approximated WSRs	148
7.4	Game-theoretic Scheduling	148
7.4.1	Mixed Strategy Selection	149
7.4.2	Impact Of The Payoff Model	150
7.5	Numerical Results	150
7.6	Conclusions	152
8	Conclusions and Future Work	153
	Appendices	157
A	Gradient Derivation for OFDM FD System	159
B	Proofs for Chapter 5	161
B.1	Proof of Theorem 5	163
C	Proofs for Chapter 5	167
C.1	Gradients	167
C.2	Proof of Theorem 6	169
C.3	Proof of Theorem 7	170

List of Figures

1.1	A basic cellular network with each base station serving its users.	1
2.1	Fraction of available samples as a function of n	14
2.2	Reconstruction SNR for 8, 10 and 12 bit ADCs as a function of n	15
2.3	Average WSR as a function of SNR with $k = \beta = -40$ dB, $N_1 = N_2 = 10$ and $M_1 = M_2 = 5$	23
2.4	2×2 MIMO FD architecture with multiple HSIC and DSIC.	24
2.5	Estimation phase - Transmitted pilot signal on each transmitter.	26
2.6	Estimation phase - Received signal on each receiver.	26
2.7	MIMO HSIC Performance.	27
2.8	Transmitter and Receiver Antenna Layout	32
2.9	SVP of the SI MIMO channel for configuration 3 with $P = 47$ cm and with transmitter and receiver in orientation H and \mathbb{I} , respectively.	33
2.10	SVP of the SI MIMO channel for configuration 3 with $P = 37.1$ cm and with transmitter and receiver in orientation H.	34
2.11	Results obtained with the SI nulling algorithm for configuration 1 and 3 for $\mathbf{C}(6 \times 4)$, $\mathbf{N}(4 \times 6)$ and $\mathbf{N}(3 \times 6)$	34
3.1	Average WSR as function of $k = \beta$ dB, with $SNR = -5$ dB.	45
3.2	Average WSR as function of SNR with different levels of LDR noise.	45
3.3	Bidirectional FD MIMO OFDM System with Multi-Stage/Hybrid BF. Only a single node is shown for simplicity in the figure.	46
3.4	Sum Rate comparisons for, Single Carrier, $N_t^i = N_r^i = 8, M_t^i = M_r^i =$ $4, d_i = 1, \forall i, L = 4$ paths.	54
3.5	Sum Rate comparisons for, OFDM, $N_s = 4, N_t^i = N_r^i = 8, M_t^i = M_r^i =$ $4, d_i = 1, \forall i, L = 4$ paths.	54
3.6	mMIMO interference channel in the mmWave consisting of Hybrid nodes with separate transmit (TX) and receive (RX) array.	56
3.7	Average WSR as a function of SNR with transmit and receive antennas $N_a^t = N_a^r = 100$ and RF chains $M_a^t = M_a^r = 32$ or $16, \forall a \in \mathcal{L}$ or \mathcal{R}	65
3.8	Average WSR as a function of SNR with transmit and receive antennas $N_a^t = N_a^r = 64$ and RF chains $M_a^t = M_a^r = 32$ or $16, \forall a \in \mathcal{L}$ or \mathcal{R}	65

4.1	FD in mmWave with HYBF to serve multi-antenna users. Tx and Rx denote transmit and receive, respectively.	69
4.2	(a) All phase shifters are unit-modulus. (b) With amplitude control. . . .	78
4.3	Typical convergence behaviour of the proposed HYBF for mmWave mMIMO FD system.	84
4.5	Average WSR as a function of the LDR noise with SNR = 40 dB.	87
4.4	Average WSR as a function of the LDR noise with SNR = 0 dB.	87
4.6	Average WSR as a function of the SNR with LDR noise $k_0 = -80$ dB. . .	88
4.7	Average WSR as a function of the SNR with LDR noise $k_0 = -60$ dB. . .	88
4.8	Average WSR as function of the SNR with LDR noise $k_0 = -40$ dB. . . .	89
4.9	Average WSR as a function of the RF chains with LDR noise $k_0 = -80$ dB and $k_0 = -40$ dB at SNR= 20 dB.	90
5.1	Multi-cell mMIMO HD and FD systems with a significant difference in the interference contributions and therefore required amount of CSI. . . .	92
5.2	Decomposition of the WSR maximization problem in UL into per-link independent optimization sub-problems. The sub-problems are solved from the bottom to the top.	109
5.3	Decomposition of the WSR optimization problem in DL into per-link independent optimization sub-problems. The sub-problems are solved from the bottom to the top.	112
5.4	Typical Convergence behaviour of the proposed HYBF designs with 32 RF chains in comparison with the fully digital beamforming.	116
5.5	Execution time for the C-HYBF and the P&D HYBF schemes with 32-RF chains.	119
5.6	Average WSR as a function of the LDR noise at SNR= 20 dB with 32 or 16 RF chains and 10 or 4 bit phase-resolution.	120
5.7	Average WSR as a function of the LDR noise at SNR= 20 dB with 12 or 10 RF chains and 10 or 4 bit phase-resolution.	121
5.8	Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -80$ dB, with 32 or 16 RF chains and 10 or 4 bit phase-resolution. .	121
5.9	Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -80$ dB, with 12 or 10 RF chains and 10 or 4 bit phase-resolution. .	122
5.10	Average WSR as function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -40$ dB, with 32 or 16 RF chains and 10 or 4 bit phase-resolution. .	123
5.11	Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -40$ dB, with 12 or 10 RF chains and 10 or 4 bit phase-resolution. .	123
6.1	Near field IRSs assisted MIMO point-to-point FD communication in mmWave.	126
6.2	Field regions of the radiating element of size D	128
6.3	Typical convergence behaviour of the proposed joint active and passive beamforming design.	138

6.4	Average WSR as a function of SNR with NF-IRSs of size 30×30 , 20×20 and 10×10 , placed at distance of 3m.	139
6.5	Average WSR as a function of SNR with NF-IRSs of size 30×30 , 20×20 and 10×10 , placed at distance of 30m.	140
6.6	Average WSR as a function of the distance between the FD nodes and the NF-IRSs at SNR= 0 dB.	140
6.7	Average WSR as a function of the distance between the FD nodes and the NF-IRSs at SNR= 30 dB.	141
7.1	Multi-cell DTDD and FD systems, where both are affected by the CI. . .	144
7.2	Average network payoff of a DTDD system for OPT, SIP, APP, and STDD, as a function of c , with fixed 15 users per cell.	151
7.3	a) CDF for the WSR with $c = 0.3$ and 15 users/cell. b) Average network payoff of a DTDD system for OPT, SIP, APP, and STDD, as a function of the number of users per cell.	152

List of Tables

4.1	Notations	72
4.2	Simulation parameters to simulate the multi-user mmWave FD system. .	86
5.1	Notations	96
5.2	Simulation parameters to simulate a multi-cell mMIMO mmWave FD system.	118
6.1	Field regions boundries for an IRS of size $D = 100\lambda$	129

Acronyms and Abbreviations

The acronyms and abbreviations used throughout the manuscript are specified in the following. They are presented here in their singular form, and their plural forms are constructed by adding and *s*, e.g. TX (transmitter) and TXs (transmitters). The meaning of an acronym is also indicated the first time that it is used.

5G	5-th Generation
AoA	Angle(s) of Arrival
AoD	Angle(s) of Departure
ADC	Analog-to-Digital-Converters
AGC	Automatic Gain Control
AR	Autoregressive
BS	Base Station
BS-to-BS	Base Station to Base Station
CDF	Cumulative Distribution Function
CI	Cross-Interference
CSI	Channel State Information
DAC	Digital-to-Analog-Converters
DC	Difference of Convex
DL	Downlink
DL-to-UL	Downlink to Uplink
DFT	Discrete Fourier Transform
DSIC	Digital Self Interference Cancellation
DTDD	Dynamic Time Division Duplexing
EIRP	Effective Isotropic Radiated Power
EM	Electromagnetic
FD	Full Duplex
FDD	Frequency Division Duplexing
HD	Half Duplex
IAB	Integrated Access and Backhaul
IFFT	Inverse Fast Fourier Transform
IRS	Intelligent Reflecting Surfaces
LDR	Limited Dynamic Range
LMMSE	Linear Minimum Mean Squared error

LS	Least Square
MIMO	Multiple-Input-Multiple-Output
MINC	Mixed-Integer and Non-Convex
mMIMO	Massive Multiple-Input-Multiple-Output
MMSE	Minimum Mean Squared Error
NF	Near-Field
NF-IRS	Near-Field Intelligent Reflecting Surfaces
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplex
PA	Power Amplifier
RF	Radio Frequency
RSNR	Reconstruction SNR
SDR	Software-Defined Radio
SI	Self Interference
SIC	Self Interference Cancellation
SISO	Single-Input Single-Output
SISR	Self-Interference-to-Signal-Ratio
SI	Self-interference Nulling
SNR	Signal-to-Noise Ratio
SVP	Singular Value Profile
TDD	Time Division Duplex
TX	Transmit
RX	Receive
UL	Uplink
UL-to-DL	Uplink to Downlink
ULA	Uniform Linear Array
URA	Uniform Rectangular Array
QN	Quantization Noise
WMMSE	Weighted MMSE
WR	Weighted Rate
w.r.t	With Respect To
WSMMSE	Weighted Sum MMSE
WSR	Weighted Sum-Rate
ZF	Zero Forcing

Mathematical Notation

x	Scalar
\mathbf{x}	Vector
\mathbf{X}	Matrix
$\mathcal{CN}(\mathbf{x}, \mathbf{X})$	Circularly symmetric complex Gaussian distribution with mean \mathbf{x} and covariance matrix \mathbf{X}
$\text{COV}(\cdot)$	Covariance matrix
\mathbb{E}	Expectation
$\mathbb{E}_{\mathbf{x}}$	Expectation with respect to the random variable \mathbf{x}
$\text{Tr}(\cdot)$	Trace
$(\cdot)^H$	Hermitian transpose
$(\cdot)^T$	Transpose
\mathbf{I}	Identity matrix
\mathbb{P}	Probability
\mathbf{D}_d	Selection of d dominant generalized eigenvectors
$\mathbf{0}_{M \times 1}$	Vector of zeros of size M
$\text{SVD}(\mathbf{X})$	Singular value decomposition of \mathbf{X}
$\mathbb{C}^{M_l \times N_m}$	Complex
$\text{diag}(\mathbf{X})$	Returns a diagonal matrix made of the diagonal elements of \mathbf{X}
$(\mathbf{X})^+$	Returns the maximum of $\mathbf{0}$ or \mathbf{X}
$(x)^+$	Returns the maximum of 0 or scalar x

Chapter 1

Introduction

The rapid and radical evolution in information and communication technology in the last decades has changed completely the way we communicate and live. From the first communication devices to the latest cellular networks, the complexity, number of users/devices, and amount of data traffic sent and received have experienced exponential growth. The exponentially growing data demands show no sign of slowing down, which urge the requirement for engineers and researchers to prepare innovative technologies to cope with the forthcoming challenges. This thesis aims to present novel contributions to evolve one of the most promising wireless transmission technology, which will be discussed next.

1.1 Cellular Networks

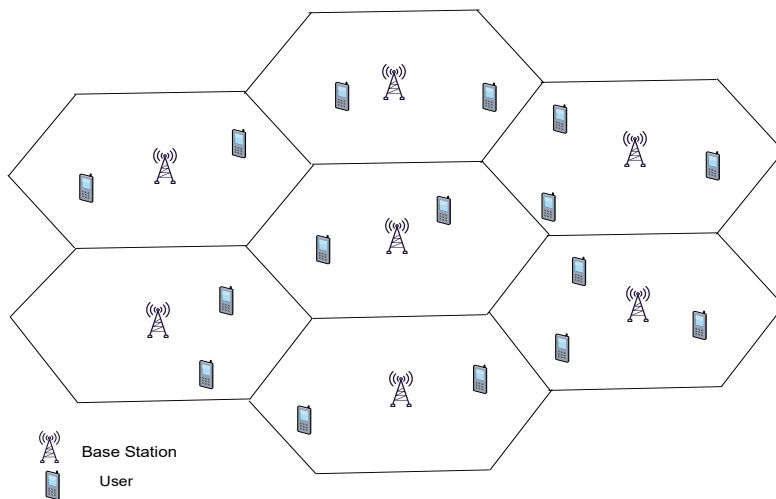


Figure 1.1: A basic cellular network with each base station serving its users.

The wireless networks follow a cellular paradigm by dividing the geographical area into cells as shown in Fig. 1.1. The users inside each cell are served by its base station (BS).

Intra-cell and Inter-cell interference are the critical challenges for serving multiple users in different cells simultaneously. However, significant advances in signal processing for the multi-antenna BSs is leading to a remarkable evolution in the interference management techniques [1].

1.1.1 Half Duplex Systems

Contemporary wireless communication systems are based on the half duplex (HD) technology, which allows the BSs to serve its users by splitting the transmission and reception operation. They can operate either by using *time division duplexing (TDD)* [2], or *frequency division duplexing (FDD)* [3, 4].

FDD

The FDD systems divide the resources by allocating different frequency bands for uplink (UL) and downlink (DL) transmissions [5]. The UL and DL scheduling modes are separated into two adjacent bands separated by a guard band, which results in minimal interference contributions. However, since the UL and DL are on different frequency bands that are far apart, the channels are generally uncorrelated and have different frequency responses. Moreover, the UL and DL separation requires a guard band, clearly a waste of the available resources. To isolate the UL and DL modes in the radio frequency (RF) circuitry, each link requires a separate oscillator of different carrier frequencies, whereas each terminal requires an expensive duplexer and a sharp RF filter [3]. Consequently, a very high hardware cost is associated with the FDD systems.

TDD

In TDD systems, the BSs divide the time into frames, and the UL and DL mode is scheduled in a synchronized manner for the whole network, depending on the global data traffic demand [1]. As the UL and DL transmission occurs in the same frequency band, more flexibility is available for spectrum utilization by simply deciding the time dedicated to UL and DL. Moreover, TDD systems also benefit from channel reciprocity. Namely, by estimating the channels based on the received signal, the same channels can be used as the channel state information (CSI) at the transmit side. Also, they offer simpler transmit parameter optimization for resource allocation and beamforming for multiple-input-multiple-output (MIMO) BSs. In contrast to the FDD systems, TDD systems require less hardware cost as a single oscillator can be shared in UL and DL at the BSs. Moreover, a duplexer is not required as the signal for both links do not appear simultaneously at the terminals.

Traditional TDD systems suffer from the drawback that all the BSs can operate only in a synchronized manner based on the global traffic demands. However, due to highly dynamically variable data demands arising from cell-to-cell due to newly emerging technologies, it is desirable that each BS could choose its scheduling mode, i.e., UL or DL, to satisfy its local traffic demand. Dynamic TDD (DTDD) has emerged as a promising solution for the next generation of wireless cellular networks, which enables each BS to

select its scheduling mode independently [6, 7]. In the final part of this thesis, we will present a novel contribution also for the DTDD systems and discuss how they are closely related to the FD systems.

1.1.2 Full Duplex Systems

The revolution in wireless communications has resulted in ever-increasing data demands and services on our limited wireless spectrum, which urge the demand for spectrally efficient communication systems to accommodate the forthcoming data growth. *Full duplex (FD)* is a promising technology that can change wireless communications radically and is the main topic of this thesis. It can enable simultaneous transmission and reception in the same frequency band, theoretically doubling the spectral efficiency. In contrast to the TDD and FDD, it avoids using two independent channels for bi-directional communication by providing more flexibility in spectrum utilization, improving data security, reducing the air interface latency/delay issues [8, 9, 10, 11, 12] and can also enable advanced joint communications and sensing [13].

Self-Interference (SI) is a major challenge to deal with to achieve an ideal FD operation, which can be 90-100 dB higher than the received signal of interest [14, 15]. Several SI cancellation (SIC) techniques have been developed in the literature to mitigate SI. They can be divided into two main categories: *passive suppression*, and *active cancellation*. In passive suppression, the SI signal is suppressed in the propagation domain before the receiver circuitry processes it. Types of passive suppression strategies are: 1) Cross-polarization, 2) Antenna separation, 3) Directional antenna, and 4) Shielding. The SI signal is mitigated in active cancellation by subtracting a processed copy of the transmitted signal from the received signal. Active cancellation techniques could be divided into digital and analog techniques based on the signal domain (digital domain or analog domain) where the SI signal is subtracted. In digital active cancellation, SIC is performed in the baseband by using the information about the transmitted signal. It can be performed without requiring additional hardware. In analog active cancellation, SIC is performed in the analog domain at the antenna level to suppress the SI signal sufficiently, such that it does not overwhelm the analog-to-digital-converters (ADCs). In contrast to digital cancellation, analog cancellation schemes require additional hardware. The amount of hardware required depends on the domain from which the information about the transmitted signal is used, which could be the digital or analog domain in the transmit chains.

A major challenge for the FD systems is to build efficient SIC architecture to suppress the SI in the analog domain as it requires additional hardware. Several SI architectures are available in the literature for sub-6 GHz for the classical single-input-single-output (SISO) and MIMO FD system, see e.g. [16, 17, 18, 19, 20, 21, 22, 23]. In [16], experimental data-driven characterization of the SISO FD systems is studied by proposing a novel SI architecture. In [17], a novel SIC architecture for the SISO OFDM FD systems is presented, in which auxiliary chains are built to accommodate an OFDM signal. Moreover, the feasibility of the proposed design is also shown for a real-time 64 sub-carriers 10 MHz FD SISO OFDM system. In [18], the first WiFi-based MIMO FD radio SIC architecture

is presented. The complexity of the proposed design scales almost linearly with the number of antennas. Further, a novel digital estimation and cancellation algorithms that eliminate almost all interference and achieve the same performance as a single antenna FD SISO system is also presented. A prototype for the proposed design is also presented with personalized analog circuit boards and integrated with a WiFi compatible standard WARP software radio implementation. In [19], we presented a novel SIC architecture which exploits the information of the transmitted signal from the baseband and uses it to cancel the SI in the analog domain. For complexity reduction, the non-linear cancellation burden is moved to the digital domain. The complexity of the proposed architecture scales only linearly as a function of the number of receive antennas. In [20], a novel SIC architecture for the analog canceller consisting of the reduced number of taps and simple multiplexers is presented. The values of each tap and configurations of the multiplexers are jointly designed with the digital beamforming filters. In [21], a novel SIC architecture which comprises auxiliary transmit chains and novel use of the demultiplexers, which allows flexible routing of the signal, is presented. The flexible signal routing mechanism reduces the number of auxiliary transmit chains, resulting in significant cost reduction. In [23], a novel joint analog and digital SIC architecture for an OFDM MIMO FD system is studied. Research towards the deployment of FD systems in the mmWave (mmWave) has recently started, and having efficient SIC architecture is still one of the major challenges for the mmWave FD systems. Due to the massive number of antennas in the mmWave FD systems, existing SIC architectures in the literature cannot be adopted due to their dependency on the number of antennas. However, the conditions for the SI signal in the mmWave FD systems are much more favourable than the traditional MIMO FD systems in sub-6 GHz. Namely, the SI signal is subject to much more path loss due to propagation challenges. Moreover, as the number of transmit antennas become massive, the total power per antenna reduces significantly, which can result in a reduced burden for the analog SIC stage. Two exiting SIC architectures for the mmWave MIMO FD systems are available in [24, 25], in which the analog SIC stage is designed between the transmit and receive RF chains, which results in significant hardware cost reduction.

Once the SI is sufficiently mitigated with the aid of the SIC techniques, beamforming can result to be a very powerful tool for the MIMO FD systems with separate transmit and receive antenna arrays. It can perform joint mitigation of the residual SI and interference while serving multiple users in an FD network. It can be divided into two main categories: 1) Digital beamforming and 2) Hybrid beamforming (HYBF). Digital beamforming is feasible for the traditional MIMO FD systems. It becomes infeasible for the massive MIMO (mMIMO) FD systems due to the high hardware cost associated with the required number of RF chains. HYBF is a promising and cost-efficient solution for the mMIMO FD systems as it enables designing the FD transceivers with fewer RF chains than the number of antennas. This thesis will mainly focus on novel HYBF techniques for the mmWave mMIMO FD systems. We remark that the FD operation can also be enabled for communication systems with only one transmit antenna array used simultaneously for transmission and reception. However, such systems are outside the scope of this thesis. Herein, only the case of FD systems with separate transmit and receive arrays will be considered.

1.2 Contributions and Thesis Outline

This thesis aims to considerably evolve the state-of-the-art of FD systems from the signal processing perspective. To this end, we first present several novel contributions for the classical SISO and MIMO FD systems in sub-6 GHz and then novel approaches for HYBF in mMIMO FD systems.

The contributions for the sub-6GHz FD systems are presented in Chapter 2. We first tackle the quantization noise in the FD systems, which is its most dominant noise. To reduce it, we propose to upscale the total signal by letting it saturate and reconstruct the missing sample using estimation theory. Then a novel digital beamforming design for the MIMO FD system taking into account the joint sum-power and the practical per-antenna power constraints, is presented. The chapter concludes with a novel SIC architecture for a single-carrier MIMO FD system, and a novel RF calibration algorithm and experimental measurements for the MIMO OFDM FD system.

The contributions for HYBF in mmWave mMIMO FD systems are presented in Chapter 3, Chapter 4 and Chapter 5. Chapter 3 firstly presents a novel HYBF design for the mmWave FD integrated access and backhaul (IAB) with multi-antenna users. Then, a novel HYBF/multi-stage beamforming design for a mmWave point-to-point OFDM mMIMO FD system is presented. It concludes with another novel HYBF design which generalizes the point-to-point mmWave FD system to the K-pairs links. Chapter 4 consists in developing a novel centralized HYBF (C-HYBF) algorithm for a single-cell mmWave mMIMO FD system to serve multiple multi-antenna users with multiple streams. The proposed algorithm considers many practical aspects and relies on the optimization method minorization-maximization (MM) [26]. Chapter 5 first studies a novel C-HYBF design for a multi-cell mmWave mMIMO FD system by extending the work for a single-cell presented in Chapter 4. However, as well-explained later, C-HYBF is not feasible in a multi-cell FD network for several reasons not discussed in the literature so far. To overcome the drawbacks of centralized implementation, we introduce the concept of per-link parallel and distributed (P&D) HYBF for mmWave by proposing a novel P&D-HYBF algorithm for a multi-cell mmWave FD system. The proposed P&D-HYBF design exhibits similar performance as the C-HYBF but has very low complexity, scaling only linearly as the network size grows, and can be implemented in a fully distributed fashion in parallel by the multi-processors FD BSs.

HYBF, even though very desirable due to less hardware cost, has many drawbacks. To replace it, Chapter 6 introduces the concept of near-field (NF) intelligent reflecting surfaces (IRSs) for MIMO FD systems. Such an idea allows removing the analog stage of the mmWave (or THz) FD systems as it has many drawbacks, assists in shaping the SI channel, reduces the power consumption and significantly improves the performance. This thesis concludes with Chapter 7, which is dedicated to the DTDD systems which suffer from cross-interference (CI) as the multi-cell FD systems. In such systems, a special case of the FD systems, CI can be mitigated with intelligent scheduling of different cells. However, the CI contribution is strictly dictated by the power allocation of the neighbouring cells and the joint scheduling and power allocation optimization problem results in being NP-hard. We map this problem into a game-theoretic framework of

static games and solve it with the concept of mixed/probabilistic Nash equilibrium in a fully distributed fashion. Apart from the aforementioned research contributions, three patent applications are under preparation, which will lead to mass deployment of the FD systems with significantly lower hardware costs.

In summary, the contributions of thesis are the following:

- A novel approach for reconstructing the saturated signals in the FD systems while reducing the quantization noise, which is the most dominant noise.
- A novel digital beamforming design, SIC architecture, RF calibration algorithm for MIMO FD system and experimental measurements for the MIMO OFDM FD system in sub-6 GHz.
- Novel HYBF designs for mmWave IAB, point-to-point OFDM, and generalization of the point-to-point mMIMO FD systems to K-pair links in mmWave.
- A novel HYBF design for mmWave mMIMO FD system to serve multi-antenna users under several practical aspects.
- A novel C-HYBF scheme for the multi-cell mMIMO mmWave FD system and identification of the fact that any C-HYBF is infeasible for a real-time multi-cell mmWave FD network.
- A novel P&D-HYBF design for the multi-cell mMIMO mmWave FD systems.
- Introduction to the concept of NF-IRSs for FD systems, leading to very less power consumption and hardware cost.
- A fully distributed and very low-complexity solution for the joint scheduling and power allocation problem for DTDD, which in practice is NP-hard.

The content of this thesis is part of the publications listed below.

1.2.1 Journals

- [J1] **Chandan Kumar Sheemar**, Leonardo Badia and Stefano Tomasin, "Game-Theoretic Mode Scheduling for Dynamic TDD in 5G Systems", in IEEE Communications Letters, Vol. 25, no. 7, pp. 2425 - 2429, Jul. 2021.
- [J2] **Chandan Kumar Sheemar**, Christo Kurisummoottil Thomas and Dirk Slock, "Practical Hybrid Beamforming for Millimeter Wave Massive MIMO Full Duplex with Limited Dynamic Range", in IEEE Open Journal of the Communications Society, vol. 3, p. 127-143, Jan. 2022.
- [J3] **Chandan Kumar Sheemar** and Dirk Slock, "Per-Link Parallel and Distributed Hybrid Beamforming for Multi-Cell Massive MIMO Millimeter Wave Full Duplex", submitted to IEEE Transactions on Wireless Communications. [Online]. Available: <https://arxiv.org/abs/2112.02335>.

- [J4] **Chandan Kumar Sheemar** and Dirk Slock, "Near-Field Intelligent Reflecting Surfaces for MIMO Full Duplex with Channel Modelling and Beamforming for Millimeter Wave", to be submitted to IEEE Transactions on Wireless Communications.

1.2.2 Conferences

- [C1] **Chandan Kumar Sheemar** and Dirk Slock, "Hybrid Beamforming and Combining for Millimeter Wave Full Duplex Massive MIMO Interference Channel", in IEEE Globecom, Dec. 2021. [Online]. Available: <https://arxiv.org/abs/2108.00465>.
- [C2] **Chandan Kumar Sheemar** and Dirk Slock, "Hybrid Beamforming for Bidirectional Massive MIMO Full Duplex Under Practical Considerations," in IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Apr. 2021, pp. 1-5.
- [C3] **Chandan Kumar Sheemar** and Dirk Slock, "Massive MIMO mmWave Full Duplex Relay for IAB with Limited Dynamic Range," in the IEEE 11-th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Apr. 2021, pp. 1-5.
- [C4] **Chandan Kumar Sheemar** and Dirk Slock, "Beamforming for Bidirectional MIMO Full Duplex Under the Joint Sum Power and Per Antenna Power Constraints," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Jun. 2021, pp. 4800-4804.
- [C5] **Chandan Kumar Sheemar** and Dirk Slock, "Receiver Design and AGC optimization with Self Interference Induced Saturation," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020, pp. 5595-5599.
- [C6] Patrick Rosson, **Chandan Kumar Sheemar**, Neharika Valecha and Dirk Slock, "Towards Massive MIMO In-Band Full Duplex Radio," in IEEE 16th International Symposium on Wireless Communication Systems (ISWCS), Aug. 2019, pp. 69-74.
- [C7] Christo Kurisummoottil Thomas, **Chandan Kumar Sheemar** and Dirk Slock, "Multi-Stage/Hybrid BF under Limited Dynamic Range for OFDM FD Backhaul with MIMO SI Nulling," in IEEE 16th International Symposium on Wireless Communication Systems (ISWCS), Aug. 2019, pp. 96-101.

1.2.3 Patents

Three patent applications about the MIMO FD systems are currently under preparation and will be submitted for approval.

Chapter 2

On the SISO and MIMO Full Duplex Systems in Sub-6 GHz

This chapter presents novel contributions for the single antenna and multi-antenna FD systems in sub-6 GHz. We present a novel transceiver design for the SISO FD systems by introducing the idea of reconstructing the received signal saturated at the ADCs due to high amount of SI power. Then a novel beamforming design under the joint sum-power and the practical per-antenna power constraints is also presented for a point-to-point MIMO FD system. The final part of this chapter presents a novel SIC architecture for mMIMO FD, a novel RF calibration algorithm for MIMO FD and some experiment results which motivate how should be the transmit and receive arrays at the FD node must be placed to achieve a low rank SI channel. The experimental measurements were taken at EURECOM, using the software-defined radio (SDR) USRP N310. The content of this chapter is available in the research papers [C4]-[C6].

2.1 Signal Reconstruction of Saturated Signals in SISO FD

2.1.1 State-of-the-Art and Motivation

To achieve FD operation, SI signal needs to be subtracted from the total receive signal to allow proper reception. SI signal power is around 110 dB higher compared to the receive signal of interest and its cancellation is not an easy task. This is mainly due to the nonlinearities present in the transmit and receive chains, which lead to inaccurate SI channel estimate and hence limit the SIC capabilities. Besides the nonlinearities, also the limited dynamic range (LDR) of the ADCs can significantly limit the performance of the FD system. Saturation of the converters can limit correct adaption of digital SIC (DSIC) stage to mimic the SI signal with opposite sign. Moreover, we also lose the receive signal of interest for the duration of saturation. In [27, 28], the authors claim that saturation of the ADCs is a major bottleneck for FD systems, which is preventing us to benefit from their full potential. The first analysis of the residual SI for a FD MIMO-OFDM system which took into account the LDR model was presented in [29].

Though all the other critical challenges to achieve FD operation have been well tackled by the researchers, the possibility to reconstruct saturated/missing samples has not been yet taken into account. So the work we present here considers for the first time ever this possibility to achieve higher gains with the FD system. To avoid saturation, one existing solution is the automatic gain control (AGC) which scales down the total signal to fit it into the LDR of converters. But, this solution preserves only few quantization levels for the Rx signal of interest and hence increases the quantization noise (QN). Recently, an alternative FD transceiver structure to discard the saturated samples based on non uniform sampling and zero crossing of the SI signal for an OFDM system has been proposed [30]. In contrast to the approach [30], the approach presented herein is much more appealing as it doesn't require additional hardware as in [30] and applicable to FD systems equipped with classical ADCs having uniform samplers.

We propose to deliberately scaling up the total Rx signal before the ADC, which of course leads to have more saturated samples but reduces QN on the available samples as well. Missing samples are then reconstructed according to their linear minimum mean square error (LMMSE) estimate by fixed lag Kalman smoothing. For the state of the art on the LMMSE estimation of missing samples, we refer to [31]. In this work, we deal only with the case of real signals but in principle our approach should be applied to both I and Q branches. Here we assume that the available samples are SI free, but ideally the joint optimization of DSIC stage and reconstructing of missing samples should be considered. Classical LMMSE estimation techniques requiring matrix inversions, may not be feasible when the dimensions increase. Our fixed lag Kalman smoothing approach to reconstruct missing samples in a FD system is more appealing, being computationally very efficient, as it is a recursive approach.

2.1.2 System Model

We consider a simple scenario consisting of a SISO FD system with only one single-antenna user. Let x and y denote the signal transmitted from the user and the received signal, respectively. We assume that x is generated according to an autoregressive (AR) process of order M . Therefore, its sample at time k can be written as

$$x_k = -a_1x_{k-1} - a_2x_{k-2} - \dots - a_Mx_{k-M} + z. \quad (2.1)$$

The scalar $z \sim \mathcal{CN}(0, \sigma_z^2)$ is an independent noise term driving the AR process. We further assume that x is a bandlimited signal with a known low pass spectrum $S_{x,x}(f)$ and it is upsampled of factor $1/\alpha$, where α is a rational number. We assume perfect SIC for the fraction of available samples β . At the receiver side, the measurement equation for non saturated sample of signal x , at time k , can be written as

$$y_k = x_k + v_k. \quad (2.2)$$

The scalar v_k denotes the granular QN, which is uniformly distributed in $[-\delta/2, \delta/2]$ with variance σ_v^2 . Let δ and b denote the quantization step size, and the number of bits of the ADC deployed in the receive chain. Let $\pm\Delta$ denote the saturation points of the ADC. The quantity δ is connected to Δ as $\delta = \Delta/2^{(b-1)}$, where b denotes the number of bits

used for quantization of the received signal at the ADCs. Missing samples are denoted with y'_k , and their measurement equation can be written as

$$y'_k = x_k + v_k + s_k, \quad (2.3)$$

where s_k denotes saturation noise. For y'_k , s_k dominates and has variance $\sigma_{s_k}^2 \gg \sigma_{v_k}^2$. It is easy to identify the positions of missing samples as $y'_k = \pm\Delta$. As these positions are well-known, we consider discarding the signal amplitudes in these positions and replace it with zeros. Our received signal obeys the AR model (2.1), and the coefficients of this AR process can be estimated by using the following equation

$$\begin{bmatrix} r_{x,x}(0) & r_{x,x}(1) & \dots & r_{x,x}(M) \\ r_{x,x}(1) & \ddots & \dots & \vdots \\ \vdots & \ddots & \dots & r_{x,x}(1) \\ r_{x,x}(M) & \dots & \dots & r_{x,x}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} \sigma_z^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (2.4)$$

where $r_{x,x}(\cdot)$ is the correlation sequence of x obtained by inverse discrete Fourier transform (DFT) of the low pass spectrum $\mathbf{S}_{xx}(f)$, which is assumed to be known at the receive side. The system of equations (2.4) is known as Yule-Walker equations and by scaling all the elements by σ_z^2 , it can be rewritten as

$$\mathbf{R}_{M+1} \mathbf{a} = \mathbf{e}_1. \quad (2.5)$$

The matrix \mathbf{R}_{M+1} denotes the matrix of correlation sequence and the vectors \mathbf{a} and \mathbf{e}_1 denote the AR coefficients and first standard basis vector, respectively. The vector \mathbf{a} and noise σ_z^2 can be also estimated with the well-known Levinson's algorithm by solving (2.4), which is a recursive approach and has complexity $\mathcal{O}(M^2)$.

Kalman's filter equations are very appealing to the case in which unknown parameters that we wish to estimate follow a state space model. These unknowns can be estimated by minimizing the mean squared error (MSE). The complete procedure consists in following a series of steps at each time instant to predict the future state by performing LMMSE prediction, based on the so-called Gauss-Markov model. The Kalman gain is then evaluated based on the estimation error and measurement covariance matrices and its objective is to correct the priori estimate. Correction of the state estimate at time k consists in summing to the priori estimate, a factor consisting of the Kalman gain multiplying innovation, which is a difference between the measured value and the priori estimate. Kalman gain varies from $[0, 1]$ and it is 0 when the measurements are very noisy and innovation is not taken into account at all. It assumes value 1 when the measurement can provide perfect information to update the priori estimate and it summed directly the innovation. Finally, LMMSE is achieved with minimum prediction MSE multiplying $(1 - k_k)$, where k_k denotes Kalman gain for one scalar measurement. For theoretical background on Kalman Filtering and the state space models, we refer the reader to [32].

Once we have the estimates of the missing samples, we also wish to smooth these estimates based on the information of the neighbouring samples. To do so, we leverage fixed lag Kalman smoothing with lag $L \ll N$, where N is the total number of samples

and to estimate the sample at time k , measurements up to time $k + L$ are taken into account. The state space model for the fixed lag Kalman smoothing can be written as

$$\mathbf{x}_k = \mathbf{H}\mathbf{x}_{k-1} + \mathbf{e}_1 z, \quad (2.6)$$

$$\mathbf{y}_k = \mathbf{A}(\mathbf{x}_k + \mathbf{v}_k + \mathbf{s}_k), \quad (2.7)$$

where \mathbf{x}_k is the $L \times 1$ state vector at time k , \mathbf{v}_k is the granular quantization noise state vector, \mathbf{A} is the $L \times L$ state observation matrix, \mathbf{y}_k is the observed state at time k , $\mathbf{e}_1 z$ is noise of the AR process, \mathbf{s}_k is the saturation noise state vector with non zero elements only for saturated samples, \mathbf{H} is the $L \times L$ state space matrix with its first row equal to $[-a_1, \dots, -a_M \ \mathbf{0}]$. The left lower block of \mathbf{H} is an identity matrix of size $L - 1$ denoted with \mathbf{I}_{L-1} and its L -th column is made of all zeros. We denote with $\hat{\mathbf{x}}_k^-$ and $\hat{\mathbf{x}}_k^+$ the priori and posteriori estimate of the state \mathbf{x}_k . The a priori and posteriori error covariance matrices are denoted with $\mathbf{R}_{\tilde{x}\tilde{x}}^{k-}$ and $\mathbf{R}_{\tilde{x}\tilde{x}}^{k+}$, respectively, and \mathbf{Q}_k denotes the measurement covariance matrix. The Kalman gain for the state space model at time k is denoted with \mathbf{K}_k .

Let N denote the total samples of the received signal x . By adapting the well-established fixed lag Kalman smoothing technique to the case of saturated samples in FD system, the LMMSE estimates of the missing samples can be obtained by iterating Algorithm 1. The smoothed LMMSE estimate of sample x_k can be found in the L -th

Algorithm 1 Fixed lag Kalman smoothing

Initialize $\mathbf{x}_0^+ = E[\mathbf{x}_0]$ and $\mathbf{R}_{\tilde{x}\tilde{x}}^{0+} = E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)(\mathbf{x}_0 - \hat{\mathbf{x}}_0^+)^T]$
for $k = 1, \dots, N$

1. $\mathbf{R}_{\tilde{x}\tilde{x}}^{k-} = \mathbf{H}\mathbf{R}_{\tilde{x}\tilde{x}}^{k-1+}\mathbf{H}^T + \mathbf{Q}_{k-1}\mathbf{e}_1^T$
2. $\mathbf{K}_k = \mathbf{R}_{\tilde{x}\tilde{x}}^{k-}\mathbf{A}^T(\mathbf{A}\mathbf{R}_{\tilde{x}\tilde{x}}^{k-}\mathbf{A}^T + \mathbf{R}_k)$
3. $\hat{\mathbf{x}}_k^- = \mathbf{H}\hat{\mathbf{x}}_{k-1}^-$
4. $\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{A}\hat{\mathbf{x}}_k^-)$
5. $\mathbf{R}_{\tilde{x}\tilde{x}}^{k+} = (\mathbf{I}_L - \mathbf{K}_k\mathbf{A})\mathbf{R}_{\tilde{x}\tilde{x}}^{k-}$

end

element of the state vector \mathbf{x}_{k+L} and the minimum MSE (MMSE) is contained in $\mathbf{R}_{\tilde{x}\tilde{x}}^{(k+L)+}(L, L)$. Tuning of the AGC and the trade-off between the saturated samples and QN is discussed in the following.

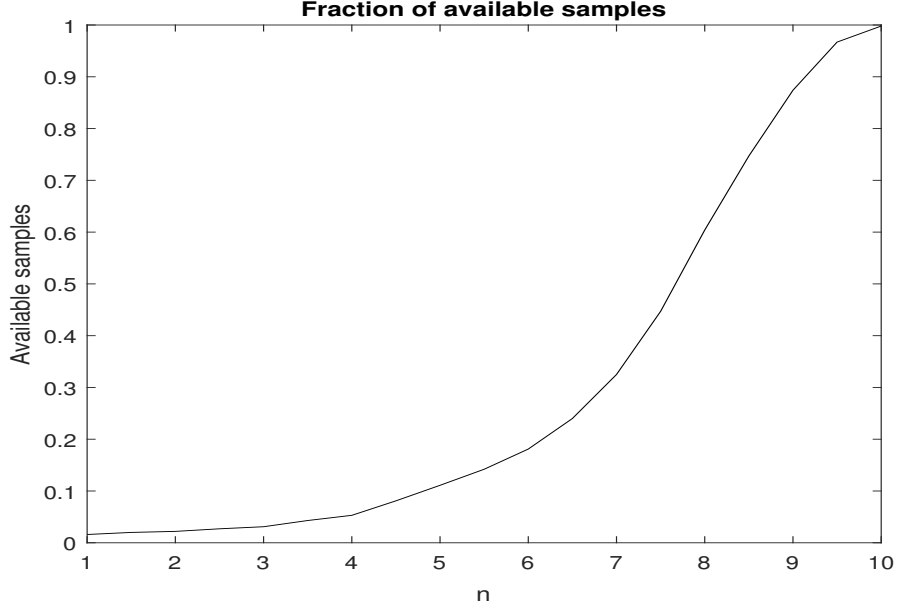
2.1.3 Simulation Results

In this section, we present simulation results to evaluate the performance of fixed lag Kalman smoothing to reconstruct missing/saturated samples and investigate the trade-off between the saturated samples and QN. For evaluation purpose, we consider ADCs in

the receive chain of the SISO FD system with resolution of 8, 10 or 12 bits. We assume that the SI signal has the same characteristics of x , except the power. We assume also that after the analog SIC stage a residual SI of 50 dB is left, to be taken care of in the baseband. At the input of the ADC, we deliberately scale up the signal consisting of residual SI and x , by letting it saturate with an AGC which dictates the fraction of available samples β . The fraction β is assumed to be varying as a function of the scalar n , control parameter of the AGC, which is linked to the saturation point as $\Delta = 2^n \sigma_x$. We also assume perfect SIC for the non saturated samples. We further assume that received signal is upsampled of factor $1/\alpha = 4$ and its spectrum $\mathbf{S}_{x,x}(f)$ at the receiver side to be known.

The fixed lag Kalman smoothing is initialized at time instant $L + 1$ with $L = 20$, and we assume that \mathbf{x}_L is known in a noisy form (only QN). We generate the receive signal of length $N = 1000$ with variance $\sigma_x^2 = 1$ according to an AR process of order $M = 10$, which has a perfect lowpass spectrum $\mathbf{S}_{x,x}(f)$. At the receiver side, as $\mathbf{S}_{x,x}(f)$ is known, its inverse DFT is calculated to get the correlation sequence of size $M + 1$, which appears in (2.4). Then, the AR coefficients \mathbf{a} and noise variance σ_z^2 driving the process are estimated by using the Levinson's algorithm. For the SI signal, we assume that it has the same spectrum, same length and generated using the same AR coefficients of x , but with different variance σ_s^2 .

Fig. 2.1 shows the variation in the fraction of available samples β as a function of the control parameter of the AGC n , varying with step size 0.5. We define the signal-to-noise-ratio (SNR) for the SISO FD system as $\text{SNR} = \sigma_x^2 / \sigma_v^2$, where σ_v^2 is the granular QN variance. To evaluate the reconstruction performance, we define the reconstruction SNR (RSNR) as $\text{RSNR} = \sigma_x^2 / \text{MSE}$. For fixed lag Kalman smoothing with lag L , the MSE for each sample at time k is obtained by averaging over the (L, L) -th element of the posteriori error covariance matrix $\mathbf{R}_{\hat{x}\hat{x}}^{(k+L)+}$. Fig. 2.2 shows the reconstruction performance of the missing samples with different bit resolutions. For $n > 6$, when more than 25% (upsampling of factor 4) becomes available, more and more innovations are taken into account to improve the performance. However, by increasing the fraction of available samples, the QN, different for ADCs with different resolution, starts becoming non-negligible. Such an increase has a direct impact on the reconstruction performance. It may be worth mentioning that because of the initialization condition, state at time L to be perfectly known, the LMMSE estimates of missing samples with very small β tends still to give a positive RSNR (in dB). This is due to the fact that the saturated samples from time $L + 1$ onwards are estimated according to the AR model (2.1), and for $n \in [1, 6]$, only very few innovations are taken into account to update the a priori estimates. This leads to having almost the same reconstruction performance for $n \in [1, 6]$, even with ADCs with different resolution as the available samples occupy the whole dynamic range, and the effect of QN is negligible. We also remark that the initial condition for which we suppose the state at time L to be known represents the case in which saturation may occur due to instantaneous SI power increase, but before that, we still had some non-saturated samples available, which we can be used to reconstruct the saturated samples without taking into account the innovation, i.e., solely based on the priori estimate. Initializing the estimation process with no knowledge of available samples can degrade the RSNR for


 Figure 2.1: Fraction of available samples as a function of n .

the three ADCs considered when $n \in [1, 6]$, i.e., when the fraction of available samples is less than $1/4$, and results to be the same for $n > 6$.

It is clearly visible from Fig. 2.2 for the case of 8 bit ADC that RSNR is equal to SNR when only 25% ($1/\alpha = 4$) of the non saturated samples are available. For each ADC considered, there exists a different optimum point leading to the optimum RSNR*. This point represents the optimum compromise between β and QN, which allows to achieve higher gains for FD systems. It is evident from Fig. 2.2 that, as the resolution increases, the QN variance decreases and therefore to achieve RSNR* (higher than SNR) more and more fraction of available samples are needed. The tuning of AGC should be done to make sure that we receive the optimum fraction of available samples β^* at the receiver side to reconstruct the missing samples.

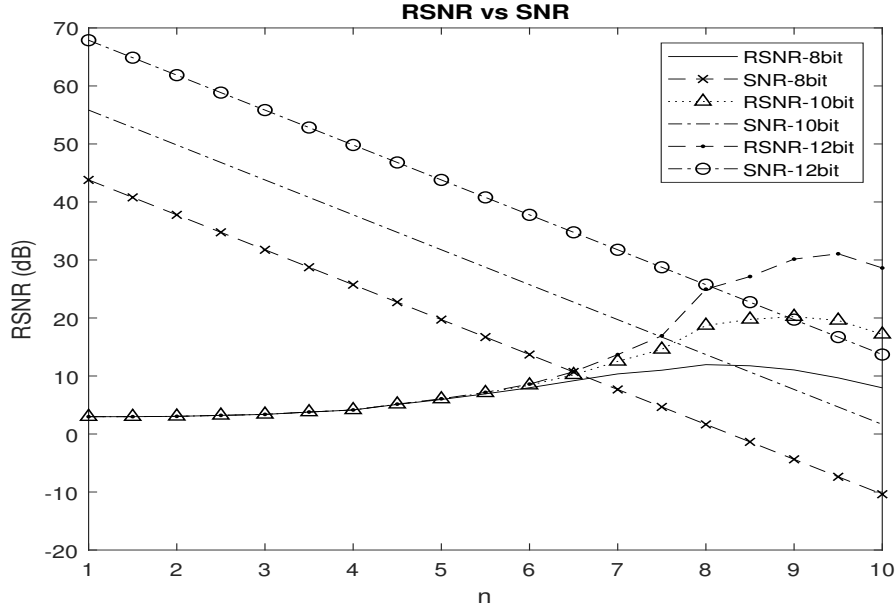
2.1.4 Large System Analysis of the MSE

In this section, we derive analytically the approximate resulting MSE as a function of the fraction of available samples β . Our derivation is based on the properties of trace operator, random matrix theory, cyclic permutation, circulant matrix approximation and partially on the expressions derived in [33]. At the end, we also establish the link between β and saturation point Δ of an ADC.

The error correlation matrix for the LMMSE estimation can be written as

$$\mathbf{R}_{\tilde{x}\tilde{x}} = \mathbf{R}_{xx} - \mathbf{R}_{xx}\mathbf{A}^H(\mathbf{A}\mathbf{R}_{yy}\mathbf{A}^H)^{-1}\mathbf{A}\mathbf{R}_{xx}, \quad (2.8)$$

where \mathbf{A} now becomes a $K \times N$ matrix which selects K non saturated samples from the total of N received samples to achieve $\beta = K/N$, $\mathbf{R}_{yy} = \mathbf{R}_{xx} + \sigma_v^2\mathbf{I}$ and \mathbf{R}_{xx} are the


 Figure 2.2: Reconstruction SNR for 8, 10 and 12 bit ADCs as a function of n .

correlation matrix of the received and transmitted signal. \mathbf{R}_{xx} is a Toeplitz matrix and can be approximated as a circulant matrix \mathbf{R}_{xx}^c , which can be written as

$$\mathbf{R}_{xx}^c = \frac{1}{N} \mathbf{F}^H \mathbf{D} \mathbf{F}, \quad (2.9)$$

where \mathbf{D} is the DFT of the finite correlation sequence of (2.4), \mathbf{F} denotes the DFT of size N and also $\mathbf{F}^{-1} = \mathbf{F}^H$ is true. We assume that \mathbf{A} performs random sub-selection of the samples to achieve β and approximate $\mathbf{A} \mathbf{F}^H \sim \mathbf{S}$, a $K \times N$ matrix with elements $s_{i,j} \sim \mathcal{N}(0, 1)$. This approximation can be introduced because \mathbf{A} performs random subset selection of samples to achieve β . This fraction is strictly related to the commulative distribution of \mathbf{A} as the DFT matrix has unit magnitude and different phase terms. \mathbf{A} selects its K rows randomly and therefore $\mathbf{A} \mathbf{F}^T$ can be seen as a random set of orthonormal columns. By taking the elements pairwise, the average norm of \mathbf{S} (if divided by N) as $N \rightarrow \infty$ goes to one, the inner product between two rows goes to zero and the inner product between two elements goes to 0 as it has independent variables. By taking into account the aforementioned approximations, we can rewrite (2.8) as follows

$$\mathbf{R}_{\tilde{x}\tilde{x}} = \mathbf{F}^{-1} \mathbf{D} \mathbf{F} - \mathbf{F}^{-1} \mathbf{D} \mathbf{S}^H (\mathbf{S} \mathbf{D} \mathbf{S}^H + N \sigma_v^2 \mathbf{I})^{-1} \mathbf{S} \mathbf{D} \mathbf{F}. \quad (2.10)$$

By applying the trace operator and dividing everything by N , we get

$$\frac{\text{Tr}(\mathbf{R}_{\tilde{x}\tilde{x}})}{N} = \text{Tr} \left(\frac{\mathbf{F}^{-1} \mathbf{D} \mathbf{F}}{N} - \frac{\mathbf{F}^{-1} \mathbf{D} \mathbf{S}^H}{N} (\mathbf{S} \mathbf{D} \mathbf{S}^H + N \sigma_v^2 \mathbf{I})^{-1} \mathbf{S} \mathbf{D} \mathbf{F} \right). \quad (2.11)$$

By using the property of trace operator, which allows cyclic permutation, $\mathbf{F} \mathbf{F}^{-1}$ becomes identity in both of the terms of the MSE expression. Hence, (2.11) can be

further simplified as

$$\frac{\text{Tr}(\mathbf{R}_{\tilde{x}\tilde{x}})}{N} = \text{Tr}\left(\frac{\mathbf{D}}{N} - \frac{\mathbf{D}\mathbf{S}^H}{N}(\mathbf{S}\mathbf{D}\mathbf{S}^H + N\sigma_v^2\mathbf{I})^{-1}\mathbf{S}\mathbf{D}\right). \quad (2.12)$$

By applying the matrix inversion lemma, we can simplify the $()^{-1}$ term in (2.12) as

$$\mathbf{S}^H(N\sigma_v^2\mathbf{I} + \mathbf{S}\mathbf{D}\mathbf{S}^H)^{-1} = \mathbf{D}^{-1}\left(\frac{1}{N\sigma_v^2}\mathbf{S}^H\mathbf{S} + \mathbf{D}^{-1}\right)\frac{\mathbf{S}^H}{N\sigma_v^2}, \quad (2.13)$$

which allows us to rewrite (2.12) as

$$\begin{aligned} \frac{MSE}{N} &= \frac{1}{N}\text{Tr}(\mathbf{D}) - \frac{1}{N}\text{Tr}((\mathbf{S}^H\mathbf{S} + N\sigma_v^2\mathbf{D}^{-1})^{-1}\mathbf{S}^H\mathbf{S}\mathbf{D}) \\ &= \sigma_v^2\text{Tr}((\mathbf{S}^H\mathbf{S} + N\sigma_v^2\mathbf{D}^{-1})^{-1}). \end{aligned} \quad (2.14)$$

By using (8)-(14) from [33], for $s_{i,j}$ being iid, (2.14) can be written as

$$\frac{MSE}{N} = e\sigma_v^2 = \frac{1}{N}\text{Tr}\left(\left(\frac{K}{N}\frac{1}{1+e}\mathbf{I}_N + \sigma_v^2\mathbf{D}^{-1}\right)^{-1}\right), \quad (2.15)$$

which yields

$$e\sigma_v^2 = \frac{1}{N}\sum_i^N \frac{1}{\frac{K}{N}\frac{1}{1+e} + \frac{\sigma_v^2}{d_i}}. \quad (2.16)$$

For the ideal low pass bandlimited spectrum case, fraction α of the N d_i values are $d_i = d$. By taking the inverse DFT of the spectrum of x , we get

$$\sigma_x^2 = \frac{1}{N}\alpha Nd, \quad (2.17)$$

and for exact bandlimited spectrum it equals fraction α of the (over)sampling frequency. Therefore, $\frac{1}{\alpha}$ denotes the oversampling factor, which implies $d = \sigma_x^2/\alpha$ and $\rho = \sigma_x^2/\sigma_v^2$ is the SNR. By using the previous results, we can rewrite (2.16) as

$$e\sigma_v^2 = \alpha \frac{1}{\frac{K}{N}\frac{1}{1+e} + \frac{\alpha}{\rho}}. \quad (2.18)$$

It implies

$$\frac{K}{N}\frac{e}{1+e} + \frac{\alpha}{\rho}e = \frac{\alpha}{\sigma_v^2}, \quad (2.19)$$

which can be rearranged as

$$\beta e + \frac{\alpha}{\rho}(e + e^2) - \frac{\alpha}{\sigma_v^2}(1 + e) = 0. \quad (2.20)$$

By multiplying with ρ and dividing by α we get

$$\frac{\rho\beta}{\alpha}e + e + e^2 - \frac{\rho}{\sigma_v^2}(1 + e) = 0, \quad (2.21)$$

which leads to the following expression

$$e^2 + (1 + \frac{\rho\beta}{\alpha} - \frac{\rho}{\sigma_v^2})e - \frac{\rho}{\sigma_v^2} = 0, \quad (2.22)$$

and finally by solving for e we get

$$e = \frac{1}{2} \left[-\left(1 + \frac{\rho\beta}{\alpha} - \frac{\rho}{\sigma_v^2}\right) + \sqrt{\left(1 + \frac{\rho\beta}{\alpha} - \frac{\rho}{\sigma_v^2}\right)^2 + \frac{4\rho}{\sigma_v^2}} \right]. \quad (2.23)$$

Under the assumptions that x_k , v_k and the SI signal follow Gaussian distribution and s_k has the dominant contribution, the fraction of available samples β can be linked to Δ as

$$\begin{aligned} \beta &= P(|y'_k| < \Delta) = 1 - P(|y'_k| \geq \Delta) = 1 - 2P(y'_k \geq \Delta) \\ &= 1 - 2(1 - P(y'_k \leq \Delta)) = 2P(y'_k \leq \Delta) - 1 = 2F'_y(\Delta) - 1 \end{aligned} \quad (2.24)$$

where $F'_y(\cdot)$ is the cumulative distribution function (CDF) of y'_k . Now $F_{y'}(\cdot) = F_{s+x+v}(\cdot) \approx F_{s+x}(\cdot) \approx F_s(\cdot)$ where $F_s(\cdot)$ is the CDF of a Gaussian random variable with zero mean and variance σ_s^2 . So under the assumption that s_k is Gaussian, we can finally conclude that $\beta = 2F_s(\Delta) - 1$.

2.1.5 Conclusions

In this work, for the first time ever, we considered the possibility to reconstruct saturated samples according to their LMMSE estimates by using fixed lag Kalman smoothing. To leverage our approach, we assumed that the signal to be reconstructed follows an AR model, it is upsampled, and its spectrum to be known at the receiver side. We proposed to deliberately provoke saturation of the ADCs, leading to fewer available samples but with less quantization noise. Simulation results show that saturated samples can be reconstructed well if the assumptions mentioned above are met. Also, the optimum compromise between the fraction of available samples and QN is studied, which depends on the resolution of the ADCs.

2.2 Beamforming for MIMO Full Duplex Under the Joint Sum-Power and Per-Antenna Power Constraints

This section presents a novel digital beamforming design for MIMO FD to maximize the weighted sum rate (WSR) under the joint sum-power and per-antenna power constraints. The regulations impose the sum-power constraints at each terminal, limiting the total transmit power. In practice, each transmit antenna is equipped with its power amplifier (PA) [34] and the per-antenna power constraints arise due to the power consumption limits imposed on the physical PAs [35]. The joint sum-power and per-antenna power constraints imposed on each terminal consider both the regulations and physical limits of the PA to optimize the systems' performance.

The joint constraints for FD systems have much more to offer compared to the HD systems. If there is no saturation noise, the most dominant noise contribution comes for

the PAs [36], which introduce additional non-linearities when operating in the non-linear region. Consequently, the residual SI power increases, limiting the maximum achievable gain for an FD system. With the per-antenna power constraints, we can limit the non-linear behaviour of PAs and improve the SI channel estimation while complying with the sum-power constraints naturally imposed by the regulations by equally distributing the transmit power at each antenna. Moreover, the transmit antennas nearest the receive array contribute the most to the SI signal's line-of-sight (LoS) component. As the analog SIC stage has very high energy consumption, we can reduce it by restricting the per-antenna constraints on the transmit antennas nearest to the receive array. Note that restricting the per-antenna power constraints improves the UL rate but can degrade the DL rate. Ideally, an optimal trade-off between the UL and DL rate must be investigated in practice.

2.2.1 System Model

We consider a MIMO FD communication system consisting of two MIMO FD nodes communicating with each other. Let $\mathcal{F} = \{1, 2\}$ contain the indices of the two nodes. Let N_l and M_l denote the number of transmit and receive antennas, respectively, at the FD node $l \in \mathcal{F}$. We consider a multi-stream approach and let $\mathbf{s}_l \in \mathbb{C}^{d_l \times 1}$ denote the d_l white and unitary variance data streams transmitted from node $l \in \mathcal{F}$. Let $\mathbf{V}_l \in \mathbb{C}^{N_l \times d_l}$ denote the digital beamformer at the node $l \in \mathcal{F}$. The channel between transmit array of node $m \in \mathcal{F}$ and receive array of node $l \in \mathcal{F}$, with $m \neq l$, is denoted with $\mathbf{H}_{l,m} \in \mathbb{C}^{M_l \times N_m}$ and the SI channel at node l is denoted with $\mathbf{H}_{l,l} \in \mathbb{C}^{M_l \times N_l}$, $\forall l \in \mathcal{F}$. Let $\mathbf{n}_l \sim \mathcal{CN}(0, \sigma_l^2 \mathbf{I}_{M_l})$ denote the thermal noise vectors at the FD node l . The signal received at the FD node l can be written as

$$\mathbf{y}_l = \mathbf{H}_{l,m}(\mathbf{V}_m \mathbf{s}_m + \mathbf{c}_m) + \mathbf{e}_l + \mathbf{n}_l + \mathbf{H}_{l,l}(\mathbf{V}_l \mathbf{s}_l + \mathbf{c}_l), \quad (2.25)$$

with $l, m \in \mathcal{F}$ and $l \neq m$. Let $\mathbf{T}_l = \mathbf{V}_l \mathbf{V}_l^H$ denote the transmit covariance matrix of node $l \in \mathcal{F}$. The terms \mathbf{c}_l and \mathbf{c}_m are the transmitter and \mathbf{e}_l and \mathbf{e}_m are the receiver noise distortions due to LDR noise at the node l and m , respectively, with $l \neq m$, and can be modelled as [37]

$$\mathbf{c}_l \sim \mathcal{CN}(\mathbf{0}_{N_l \times 1}, k_l \text{diag}(\mathbf{T}_l)), \quad \forall l \in \mathcal{F}, \quad (2.26)$$

$$\mathbf{e}_l \sim \mathcal{CN}(\mathbf{0}_{M_l \times 1}, \beta_l \text{diag}(\mathbf{\Phi}_l)), \quad \forall l \in \mathcal{F}, \quad (2.27)$$

where $k_l \ll 1$, $\beta_l \ll 1$ and $\mathbf{\Phi}_l = \text{Cov}(\mathbf{X}_l)$, where \mathbf{X}_l denotes the undistorted received vector at node l , such that $\mathbf{X}_l = \mathbf{y}_l - \mathbf{e}_l$, $\forall l \in \mathcal{F}$. Let \mathbf{X}_l be the received covariance matrix of the undistorted received signal at node l given by

$$\mathbf{X}_l = \mathbf{H}_{l,m} \mathbf{T}_m \mathbf{H}_{l,m}^H + \mathbf{H}_{l,m} k_m \text{diag}(\mathbf{T}_m) \mathbf{H}_{l,m}^H + \sigma_l^2 \mathbf{I} + \mathbf{H}_{l,l}(\mathbf{T}_l + k_l \text{diag}(\mathbf{T}_l)) \mathbf{H}_{l,l}^H. \quad (2.28)$$

Let $\mathbf{K}_l \triangleq \mathbf{H}_{l,m} \mathbf{T}_m \mathbf{H}_{l,m}^H$ denote the useful received signal covariance part. The received (signal plus) interference and noise covariance matrices received at the FD node $l \in \mathcal{F}$,

denoted with (\mathbf{R}_l) $\mathbf{R}_{\bar{l}}$, can be written as

$$\mathbf{R}_l \approx (\mathbf{X}_l + \beta_l \text{diag}(\mathbf{X}_l)), \quad \mathbf{R}_{\bar{l}} \approx \mathbf{R}_l - \mathbf{K}_l. \quad (2.29)$$

The WSR maximization problem for the bidirectional MIMO FD system, under the joint sum-power and per-antenna power constraints can be formally stated as

$$\max_{\mathbf{V}} \sum_{m \in \mathcal{F}} w_m \ln \det(\mathbf{R}_{\bar{m}}^{-1} \mathbf{R}_m) \quad (2.30a)$$

$$\text{s.t.} \quad \text{diag}(\mathbf{V}_l \mathbf{V}_l^H) \preceq \mathbf{P}_l, \quad \forall l \in \mathcal{F}, \quad (2.30b)$$

$$\text{Tr}(\mathbf{V}_l \mathbf{V}_l^H) \leq p_l, \quad \forall l \in \mathcal{F}. \quad (2.30c)$$

where w_m denote the rate weight for node m and p_l and \mathbf{P}_l (a diagonal matrix) denote the sum-power and per-antenna power constraints for node l , respectively. The collection of digital beamformers is denoted with \mathbf{V} .

Problem (2.30) is non-concave in \mathbf{T}_l due to SI terms at both the FD nodes which leads to finding the global optimum solution very challenging. To find a feasible solution, we construct its minorizer using the MM approach [38] and then optimize the beamformers at each iteration by adopting an alternating optimization process.

The WSR shown in (2.30) can be written as a sum of weighted rate (WR) of nodes l and $m \in \mathcal{F}, m \neq l$, i.e., $WSR = WR_l + WR_m$. Note that WR_l is concave in \mathbf{T}_m and non-concave in \mathbf{T}_l due to SI and WR_m is concave in \mathbf{T}_l and non-concave in \mathbf{T}_m due to SI. Since a linear function is simultaneously convex and concave, difference of convex (DC) programming [38] introduces the first order Taylor series expansion of WR_m in \mathbf{T}_l and WR_l in \mathbf{T}_m around $\hat{\mathbf{T}}$, i.e., all \mathbf{T}_i , as

$$\underline{WR}_l(\mathbf{T}_m, \hat{\mathbf{T}}) = WR_l(\mathbf{T}_m, \hat{\mathbf{T}}) - \text{Tr}((\mathbf{T}_m - \hat{\mathbf{T}})\mathbf{G}_m), \quad (2.31a)$$

$$\underline{WR}_m(\mathbf{T}_l, \hat{\mathbf{T}}) = WR_m(\mathbf{T}_l, \hat{\mathbf{T}}) - \text{Tr}((\mathbf{T}_l - \hat{\mathbf{T}})\mathbf{G}_l), \quad (2.31b)$$

where \mathbf{G}_m and \mathbf{G}_l are the gradients of WR_m and WR_l with respect to $\hat{\mathbf{T}}_l$ and $\hat{\mathbf{T}}_m$, respectively. By using the matrix differentiation properties [39], they can be easily derived and result to be

$$\begin{aligned} \mathbf{G}_l = & w_l (\mathbf{H}_{l,l}^H (\mathbf{R}_{\bar{l}}^{-1} - \mathbf{R}_l^{-1} + \beta_m \text{diag}(\mathbf{R}_{\bar{l}}^{-1} - \mathbf{R}_l^{-1})) \mathbf{H}_{l,l} \\ & + k_l \text{diag}(\mathbf{H}_{l,l}^1 (\mathbf{R}_{\bar{l}}^{-1} - \mathbf{R}_l^{-1}) \mathbf{H}_{l,l}), \end{aligned} \quad (2.32)$$

$$\begin{aligned} \mathbf{G}_m = & w_m (\mathbf{H}_{m,m}^H (\mathbf{R}_{\bar{m}}^{-1} - \mathbf{R}_m^{-1} + \beta_l \text{diag}(\mathbf{R}_{\bar{m}}^{-1} - \mathbf{R}_m^{-1})) \mathbf{H}_{m,m} \\ & + k_m \text{diag}(\mathbf{H}_{m,m}^1 (\mathbf{R}_{\bar{m}}^{-1} - \mathbf{R}_m^{-1}) \mathbf{H}_{m,m}). \end{aligned} \quad (2.33)$$

Note that, the tangent expressions constitutes a touching lower bound for (2.30), hence the DC programming is also a minorization approach, regardless of the reparameterization of the transmit covariance matrices as a function of the beamformers.

Let λ_l and $\Psi_l = \text{diag}(\psi_1^l, \dots, \psi_{N_l}^l)$ be the Lagrange multipliers associated with the sum-power and per-antenna power constraint at node $l \in \mathcal{F}$, respectively. Dropping the

constant terms, reparameterizing \mathbf{T}_l as a function of the beamformers, performing this linearization $\forall l \in \mathcal{F}$, augmenting the WSR cost function with the per-antenna and sum power constraints, yield the following Lagrangian

$$\begin{aligned} \mathcal{L} = \sum_{m \in \mathcal{F}} & (\psi_m p_m + \text{Tr}(\mathbf{\Psi}_m \mathbf{P}_m) + \text{ldet}(\mathbf{I} + \mathbf{V}_m^H \mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m} \mathbf{V}_m) \\ & - \text{Tr}(\mathbf{V}_m^H (\mathbf{G}_m + \lambda_m \mathbf{I} + \mathbf{\Psi}_m) \mathbf{V}_m)). \end{aligned} \quad (2.34)$$

2.2.2 Digital Beamforming

To optimize the digital beamformer, we take the derivative of (2.34) with respect to $\mathbf{V}_m, \forall m \in \mathcal{F}$ and $l \neq m$, which yields the following Karush–Kuhn–Tucker (KKT) condition

$$\mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m} \mathbf{V}_m (\mathbf{I} + \mathbf{V}_m^H \mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m} \mathbf{V}_m)^{-1} - (\mathbf{G}_m + \lambda_m \mathbf{I} + \mathbf{\Psi}_m) \mathbf{V}_m = 0. \quad (2.35)$$

Theorem 1. *The digital beamformer \mathbf{V}_m , fixed \mathbf{V}_l , can be optimized as the generalized dominant eigenvector solution of the pair of the following matrices*

$$\mathbf{V}_m \rightarrow \mathbf{D}_{d_l}(\mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m}, \mathbf{G}_m + \lambda_m \mathbf{I} + \mathbf{\Psi}_m). \quad (2.36)$$

Proof. The results follow directly by applying the proof available in [38] for Proposition 1, by substituting $\mathbf{H}_{i,i}^H \mathbf{R}_i^{-1} \mathbf{H}_{i,i}$ appearing in [38] with $\mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m}$ and $\mathbf{A}_i + \mu_i \mathbf{I}$ appearing in [38] with $\mathbf{G}_m + \lambda_m \mathbf{I} + \mathbf{\Psi}_m$. \square

The generalized dominant eigenvector solution provides the optimized beamforming directions but not the power. Therefore, we consider scaling the columns of the digital beamformers to unit-norm and design a novel power allocation scheme. Solution (2.36) diagonalize the following matrices

$$\mathbf{V}_m^H \mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m} \mathbf{V}_m = \mathbf{\Sigma}_m^{(1)}, \quad (2.37a)$$

$$\mathbf{V}_m^H (\mathbf{G}_m + \lambda_m \mathbf{I} + \mathbf{\Psi}_m) \mathbf{V}_m = \mathbf{\Sigma}_m^{(2)}, \quad (2.37b)$$

at each iteration to maximize the WSR. Once the beamformers are computed, the optimal power allocation can be included while searching for the multipliers, satisfying the joint constraints. Formally, the power optimization problem for a MIMO system can be written as

$$\max_{\mathbf{P}_m} \sum_{l \in \mathcal{F}} w_l \text{ldet}(\mathbf{I} + \mathbf{\Sigma}_m^{(1)} \mathbf{P}_m) - \text{Tr}(\mathbf{\Sigma}_m^{(2)} \mathbf{P}_m), \quad \forall m \in \mathcal{F}. \quad (2.38)$$

Let \mathbf{P} denote the collection of powers. Note that as \mathbf{V}_m is a generalized dominant eigenvector solution for (2.34), it diagonalize $\mathbf{\Sigma}_m^{(2)}$ and $\mathbf{\Sigma}_m^{(1)}$. Multiplying it by a diagonal power matrix, it still yields a generalized dominant eigenvector and therefore the validity of Theorem 1 is preserved. The optimal power allocation, $\forall m \in \mathcal{F}$, can be obtained by solving (2.38), which yields the following optimal power allocation scheme

$$\mathbf{P}_m = (w_l (\mathbf{V}_m^H (\mathbf{G}_m + \lambda_m \mathbf{I} + \mathbf{\Psi}_m) \mathbf{V}_m)^{-1} - (\mathbf{V}_m^H \mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m} \mathbf{V}_m)^{-1})^+, \quad (2.39)$$

where $(\mathbf{X})^+ = \max\{\mathbf{0}, \mathbf{X}\}$. To satisfy the per-antenna and sum power constraints, we consider the following Lagrange dual function

$$\min_{\lambda_l, \Psi_l} \mathcal{L}(\lambda_l, \Psi_l). \quad (2.40)$$

The dual function $\mathcal{L}(\lambda_l, \Psi_l)$ is the pointwise supremum of a family of functions of λ_l, Ψ_l , it is convex [40], and the globally optimal value λ_l, Ψ_l and can be found by using any of the numerous convex optimization techniques. In this work, we adopt the Bisection algorithm for the search of multipliers. Let $\mathcal{L}_l = \{\lambda_l, \psi_l, \dots, \psi_{N_l}\}$ denote the set containing the Lagrange multipliers associated with the joint constraints at the node $l \in \mathcal{F}$. Let $\underline{\mu}_i$, and $\overline{\mu}_i$ denote the upper and lower bounds for the Lagrange multiplier $\mu_i \in \mathcal{L}_l$. The complete procedure to solve (2.30) to the local optimum based on alternating optimization is formally stated in Algorithm 1.

Algorithm 2 Beamforming for BD-FD

Given: The CSI and rate weights.

Initialize: $\mathbf{V}_l, \forall l \in \mathcal{F}$.

Repeat until convergence

for: $\forall m \in \mathcal{F}$.

 Compute $\hat{\mathbf{G}}_m$ with (2.33).

 Compute \mathbf{V}_m with (2.36) and normalize it.

 Set $\underline{\mu}_i = 0$ and $\overline{\mu}_i = \mu_{i_{max}} \forall i \in \mathcal{L}_m$.

for: $\forall \mu_i \in \mathcal{L}_m$

Repeat until convergence

 set $\mu_i = (\underline{\mu}_i + \overline{\mu}_i)/2$.

 Compute \mathbf{P}_m with (2.39),

If constraint for μ_i is violated,

 set $\underline{\mu}_i = \mu_i$, **else** $\overline{\mu}_i = \mu_i$,

 Set $\mathbf{T}_m = \mathbf{V}_m \mathbf{P}_m \mathbf{V}_m^H$

Next m .

2.2.3 Convergence Proof

To prove the convergence of Algorithm 2, the ingredients required are minorization [26], Lagrange duality, saddle point and KKT conditions [40]. Let $WSR(\mathbf{T})$ denote the cost function (2.30) as a function of transmit covariance matrices, and let $\underline{WSR}(\mathbf{T}, \hat{\mathbf{T}})$ denote its minorized version, given as

$$WSR(\mathbf{T}) \geq \underline{WSR}(\mathbf{T}, \hat{\mathbf{T}}) = \sum_{l \in \mathcal{F}} w_l \ln \det(\mathbf{I} + \mathbf{V}_m^H \mathbf{H}_{l,m}^H \mathbf{R}_l^{-1} \mathbf{H}_{l,m} \mathbf{V}_m) - \text{Tr}((\mathbf{T}_m - \hat{\mathbf{T}}) \mathbf{G}_m). \quad (2.41)$$

The minorizer, which becomes concave in \mathbf{T} , has the same gradient of (2.30). Hence, the KKT conditions are not affected. Reparametrizing the transmit covariance matrices \mathbf{T}

as a function of the powers \mathbf{P} and digital beamformers \mathbf{V} , then adding the sum-power and per-antenna power constraints yield the Lagrangian (2.34). During the alternating optimization process, every alternating update of (2.34) leads to an increase in the WSR, ensuring convergence for both the parameters. For the KKT conditions, at the convergence point, the gradients of (2.34) for the digital beamformers yield the same gradients of the original cost function (2.30). For fixed digital beamformers, (2.34) is concave in \mathbf{P} , therefore we have strong duality for the saddle point $\max_{\mathbf{P}} \min_{\lambda, \Psi} \mathcal{L}$. Moreover, at the convergence point the solution to $\min_{\lambda, \Psi} \mathcal{L}$ satisfies the complementary slackness condition i.e.,

$$\lambda_l(p_l - \text{Tr}(\mathbf{V}_l \mathbf{V}_l^H)) = 0, \forall l \in \mathcal{F}, \quad (2.42a)$$

$$\text{Tr}(\Psi_l \mathbf{P}_l - \text{diag}(\mathbf{V}_l \mathbf{V}_l^H)) = 0, \forall l \in \mathcal{F}. \quad (2.43a)$$

2.2.4 Simulation Results

This section presents simulation results for the proposed digital beamforming design for MIMO FD under the joint sum-power and per-antenna power constraints. For comparison, we define the following benchmark schemes: 1) We define an ideal FD (I-FD) system for which there is no LDR noise, i.e. $k, \beta = 0$, and 2) We define an ideal half-duplex (I-HD) system for which is also not affected by the LDR noise, i.e. $k, \beta = 0$, and the split equally the resources in time to serve in uplink and downlink mode. We also compare our design with the weighted minimum MSE (WMMSE) design proposed in [41].

The SNR for the FD node l is defined as $\text{SNR} = p_l / \sigma_l^2$, where p_l and σ_l^2 are the total transmit power and noise variance, respectively. The SI channel is modelled with a Rician fading channel model. We assume a highly dominant LoS component for the SI channel modelled with a Rician factor 10^5 dB [16]. The direct channels are modelled as $\mathcal{CN} \sim (0, 1)$ and the per-antenna power constraints at each node are set as the total sum-power divided by the number of transmit antennas. We assume the same LDR noise level for both the MIMO FD nodes in transmission and reception, i.e., $k = k_l = k_m$ and $\beta = \beta_l = \beta_m$. Both the nodes are assumed to have 10 transmit and 5 receive antennas. The proposed design is labelled as the practical beamforming (P-BF) design and the results are reported by averaging over 100 channel realizations.

Fig. 2.3 shows the performance as a function of the SNR with $k = \beta = -40$ dB. We can observe the impact of LDR noise and how it affects the maximum achievable gain, which tends to deviate from the gain of an I-FD. We can also see that both the communication links, i.e., $\mathbf{H}_{2,1}$ and $\mathbf{H}_{1,2}$, achieve the same WSR on average. The most important result can be seen by comparing our design with the WMMSE design [42], which considers only the sum power constraints. It is visible that when the transmit power is low, then the system with only sum-power constraints performs better as the LDR results to be below the noise level and the beamforming design is less constrained. The LDR noise per-antenna tend to increase as the total transmit power increases. However, setting the per-antenna power constraint equal to sum-power divided by the number of antennas equally distributes the total transmit power on each antenna. Consequently,

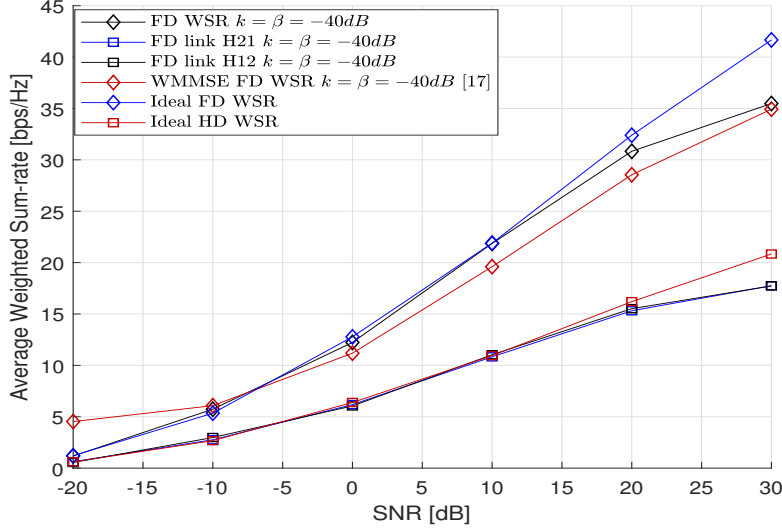


Figure 2.3: Average WSR as a function of SNR with $k = \beta = -40$ dB, $N_1 = N_2 = 10$ and $M_1 = M_2 = 5$.

equally distributing the total transmit power on each antenna results in minimum transmit LDR noise level per antenna, which allows to achieve higher gains at high transmit power.

2.2.5 Conclusions

We studied the problem of WSR maximization for a point-to-point MIMO FD system under the joint-sum power and per-antenna power constraints and under the LDR noise model. A novel alternating optimization based algorithm, which relies on the MM optimization method, is proposed. Results show that when there is LDR noise, it is desirable to design the FD systems under the joint sum-power and per-antenna power constraints, which minimize the transmit LDR noise variance per-antenna, being proportional to its transmit power.

2.3 Towards Massive MIMO In-Band Full Duplex Radio

In this section, we present several contributions to pave the path towards massive MIMO FD systems. We present a novel hybrid SIC (HSIC) architecture, which exploits the information of transmitted signal from baseband and uses it to cancel SI in the analog domain before the ADCs. The complexity of the proposed digital-to-RF (D2RF) HSIC architecture scales only linearly as the number of receive antennas, which is very desirable for massive MIMO FD systems. Correct analysis of the SI channel from the measured data is crucial to investigate the FD systems. However, when measuring the SI channel based on the transmitted signal, the actual channel results to be affected by the RF components available in the transmit and receive chains. We present a novel RF calibration algorithm for MIMO FD systems, such that correct conclusions on the SI channel from the measured data could be drawn. Moreover, we present a novel reduced complexity analog SI nulling

(SIN) algorithm, motivated by the Vandermonde vector structure of the uniform linear array (ULA) response. This section concludes with some experiment results taken at EURECOM, France, which shows how the transmit and receive antenna arrays at the FD node must be arranged to achieve a low-rank SI channel.

2.3.1 MIMO D2RF HSIC Architecture

We propose a novel SI architecture based on the solution proposed by Rice University [16], appropriate for MIMO FD transceivers. Our architecture can regenerate all the SI generated in a MIMO configuration and mitigate it very efficiently. In the MIMO context, SI originates from the interactions between all the transmit and receive antennas, with each antenna having its RF chain. The proposed solution can also be applied to the multi-antenna case in which antennas are shared between the transmitter and receiver via circulators. However, for the exposition here, we shall focus on the case of separate transmit and receive antenna arrays, which is the topic of this thesis. For a $N_{tx} \times N_{rx}$ MIMO configuration, each receiver is affected by N_{tx} SI contributions coming from the N_{tx} transmit antenna. Therefore, the FD transceiver must mitigate $N_{tx}N_{rx}$ SI components. A pure analog solution would need $N_{tx}N_{rx}$ SISO cancellers, which may also need to be frequency-selective.

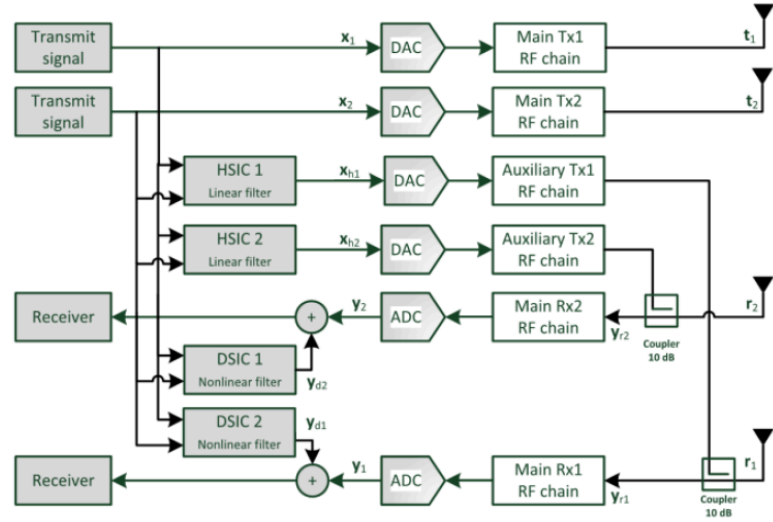


Figure 2.4: 2×2 MIMO FD architecture with multiple HSIC and DSIC.

To reduce the complexity of an analog MIMO RF SIC, we decided to move analog filtering to the digital domain. This is done by using digital filters and auxiliary transmitters as shown in Fig. 2.4. The goal of the auxiliary transmitters is to cancel the SI in the analog domain as much as possible, in any case enough for proper operation of the (ADCs in the) receive chains. The output of each auxiliary transmitter is a combination of signals coming from filtering the desired transmitter signals. In the end, DSIC mitigate the remaining SI as discussed e.g. in [11]. Here we focus on the D2RF HSIC for a MIMO

configuration. Our proposed MIMO D2RF HSIC solution is particularly well adapted for MIMO with a large number of antennas as it only requires N_{rx} HSIC branches, as compared to $N_{tx}N_{rx}$ branches in the case of analog SIC. To follow the Effective Isotropic Radiated Power (EIRP) regulation, the use of many transmit antennas implies power reduction per transmit antenna, and thus reduction of the SI contribution coming from each antenna.

2.3.2 System Model

It is possible to define a $N_{rx} \times (N_{tx} + N_{rx})$ SI MIMO channel between all RF generators (i.e. main transmit chains and auxiliary transmit chains) and the receive chains of the same transceiver. Ignoring non-linearities, and considering only SI, we can write the system equation in baseband in the frequency domain as:

$$\mathbf{y} = [\mathbf{H}_{iac} \quad \mathbf{H}_{ih}] \begin{bmatrix} \mathbf{x} \\ \mathbf{x}_h \end{bmatrix} \quad (2.44)$$

where $\mathbf{y} = [y_1 \cdots y_{N_{rx}}]^T$ is the vector of received signal after HSIC cancellation and conversion (see Fig. 2.4), $\mathbf{x} = [x_1 \cdots x_{N_{tx}}]^T$ is the transmit signal vector in the main transmit chain, $\mathbf{x}_h = [x_{h1} \cdots x_{hN_{rx}}]^T$ is the transmit signal vector for the auxiliary transmit chain, \mathbf{H}_{iac} represents the $N_{rx} \times N_{tx}$ MIMO SI channel for the main transmit chains (accounting for propagation from transmit to receive antenna arrays and electromagnetic (EM) coupling between transmit and receive arrays), whereas \mathbf{H}_{ih} represents the $N_{rx} \times N_{rx}$ MIMO SI channel for the auxiliary transmit chains in the HSIC branches (accounting for the explicit coupling from auxiliary transmit to receive RF arrays and their EM coupling also).

We can also define a $N_{rx} \times N_{tx}$ MIMO digital filter \mathbf{H}_{hsic} . This digital filter massages the actual transmit signal \mathbf{x} into an appropriate input \mathbf{x}_h for the auxiliary transmit chains in the HSIC branches before subtraction in the receive chains and SI cancellation. We have the relation

$$\mathbf{x}_h = \mathbf{H}_{hsic} \mathbf{x} . \quad (2.45)$$

2.3.3 MIMO SI Channel Estimation

This section proposes a simple approach to estimate the SI MIMO channel (i.e. the $N_{rx}(N_{tx} + N_{rx})$ sub-channels). The problem is similar to classical MIMO channel estimation. The main difference is that time and frequency synchronisation is not necessary in this configuration as the same oscillator and the same sampling frequency are used for all the transmitters and all the receivers.

A reference signal is transmitted on each transmitted port sequentially. During this phase, the two synchronised receivers estimate all the sub-channels. The SI MIMO channel is estimated like in TDMA (orthogonal transmission scheme). We follow here the dimensions $N_{rx} = N_{tx} = 2$ of Fig. 2.4. The four transmitted and received complex signals (IQ) versus time are shown in Fig. 2.5 and Fig. 2.6 respectively, exhibiting the time multiplexed pilots. This estimation phase is the most straightforward and was

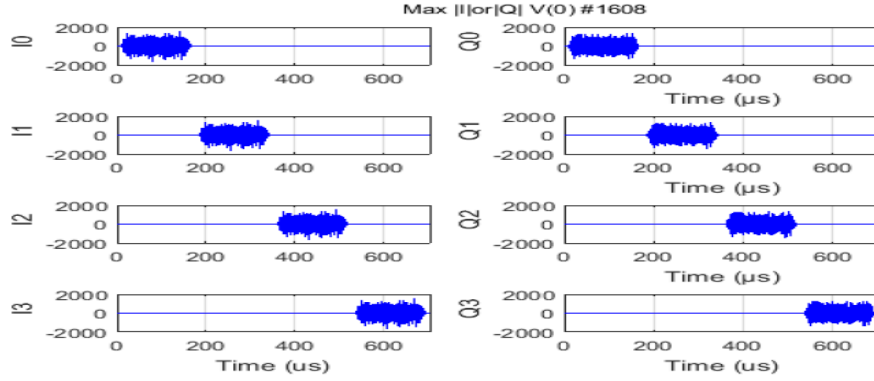


Figure 2.5: Estimation phase - Transmitted pilot signal on each transmitter.

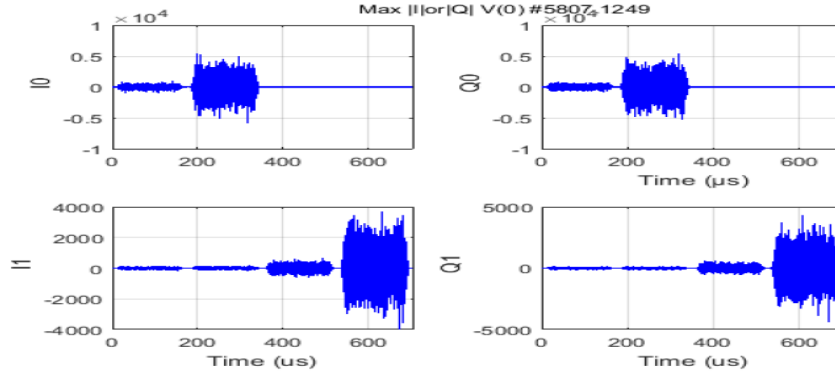


Figure 2.6: Estimation phase - Received signal on each receiver.

implemented in the simulator discussed further, but a non-orthogonal approach can also be used to reduce its duration. Of course, adaptive filtering approaches can also be used to determine \mathbf{H}_{hsic} directly without SI channel estimation (as for the digital SIC).

2.3.4 Zero Forcing HSIC Solution

In the SISO configuration, when the transceiver has only one transmit and one receive antenna, there is only one SI and the HSIC can easily be computed. In the MIMO configuration the problem is a bit more complex. Considering the case of the MIMO configuration shown in Fig. 2.4, we should find the HSIC filters which minimize the SI power of the two received signals. The filtering can be narrowband (i.e. only one complex coefficient per sub-channel) or wideband (i.e. FIR with several coefficients per sub-channel). This depends on delay spread of the MIMO SI channel and the RF bandwidth. Substituting (2.45) into (2.44) gives $\mathbf{y} = (\mathbf{H}_{iac} + \mathbf{H}_{ih} \mathbf{H}_{hsic}) \mathbf{x}$. Hence, in order to zero-force all the SI effect of \mathbf{x} on \mathbf{y} , the HSIC filter obviously needs to be chosen as

$$\mathbf{H}_{hsic} = -\mathbf{H}_{ih}^{-1} \mathbf{H}_{iac} . \quad (2.46)$$

In wideband filtering, to reach good SI mitigation, the delay spread of \mathbf{H}_{hsic} needs to be sufficiently large.

Simulations Results on HSIC

We simulated wideband filtering (up to 4 FIR coefficients per subchannel). The ZF solution is computed based on the 8 MIMO SI subchannel estimations (FIR of length 4). Fig. 2.7 shows the transmitted signals and the SI when the HSIC is not applied (in which case we have antenna isolation only) and when it is applied. The antenna isolation is around 30 dB. HSIC MIMO allows around 30 dB of additional cancellation, up to the noise floor in this scenario. The left and right subfigures correspond to receive chains 1 and 2, respectively.

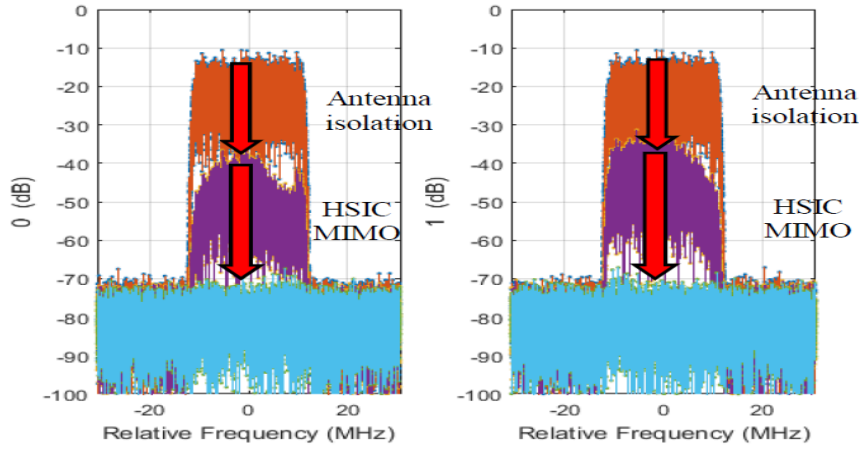


Figure 2.7: MIMO HSIC Performance.

2.3.5 Self-Interference Nulling in MIMO Full Duplex

In the following, ULAs are assumed at the transmit and receive array of the MIMO FD node. We first model the MIMO SI channel between ULAs with the near-field LoS model and propose a new iterative RF calibration algorithm for correct analysis of SI MIMO channel from the measured channel/data. We also present some measurement results, performed with the software-defined radio USRP N310, with different antenna array configurations of the ULAs and our objective would be to achieve a low rank SI MIMO channel, which leads to reduced zero-forcing (ZF) constraints. Finally, a reduced complexity analog SIN algorithm, which exploits the banded nulling structure and motivated by the Vandermonde vector structure of ULA response is proposed.

2.3.6 RF Calibration and LoS Near Field MIMO Channel Model

We consider an Orthogonal frequency-division multiplexing (OFDM) FD system with a total of N_s subcarriers. The measured SI MIMO channel at each subcarrier n can be written as:

$$\mathbf{H}[n] = \mathbf{R}[n]\mathbf{C}[n]\mathbf{T}[n] + \mathbf{V}[n] \quad (2.47)$$

where $\mathbf{C}[n]$ and $\mathbf{V}[n]$ are the internal SI MIMO propagation channel and measurement noise, respectively. $\mathbf{R}[n]$ and $\mathbf{T}[n]$ denote the calibration matrices, which represent the transfer function of RF circuitry in transmit and receive chains. As discussed in [43], these matrices are diagonal, which is also true for our case. We consider the near field line-of-sight (LoS) model for $\mathbf{C}[n]$, with each of its elements given by:

$$c_{i,j}[n] = \frac{\alpha[n]}{d_{i,j}^\gamma} e^{-j\beta d_{i,j}}, \quad (2.48)$$

where $\beta = 2\pi f_n$ is the phase constant with subcarrier frequency f_n , d denotes distance, $\alpha[n]$ is a constant, γ is the pathloss factor and subscript (i, j) denotes the j -th transmit and the i -th receive antennas, respectively. The measured channel $\mathbf{H}[n]$ not only depends on the internal SI MIMO propagation channel, but also on the RF circuitry in the transmit and receive chains. Therefore, to remove the effect of RF circuitry from the measured channel, we present a new iterative algorithm for RF calibration which leads us closer to the actual internal SI MIMO propagation channel. This allows us to draw conclusions which depend only on the antenna configuration and the effects of RF circuitry are removed. As the RF circuitry does not alter the phase of transmitted signals but just their amplitudes, we only need the absolute calibration factors. To obtain these factors in $\mathbf{R}[n]$ and $\mathbf{T}[n]$, we use the least square (LS) fitting criterion at the level of $\mathbf{C}[n]$ instead of $\mathbf{H}[n]$, which reflects the decorrelated Gaussian modeling error on $\mathbf{c}[n]$. By estimating these absolute calibration factors directly from $\mathbf{H}[n]$, we apply them on $\mathbf{c}[n]$. For calibration purposes, we minimize the average LS fit over subcarriers $n - q_s : n + q_s$, i.e. $2q_s + 1$ subcarriers, with $q_s \in [1, n]$. The objective function we consider is:

$$\min_{\alpha[n], \mathbf{T}[n], \mathbf{R}[n]} \sum_{m=\max(1, n-q_s)}^{\min(N_s, n+q_s)} \|\mathbf{R}[n]\mathbf{H}[m]\mathbf{T}'[n] - \mathbf{C}'[m]\|_F^2, \quad (2.49)$$

where $\mathbf{C}[m] = \alpha[m]\mathbf{C}'[m]$ and $\mathbf{T}'[n] = \alpha[n]\mathbf{T}[n]$ is the reparameterization of $\mathbf{T}[n]$ in terms of $\alpha[n]$. Optimization of (7.15) is done at each subcarrier with alternating minimization approach according to the LS fitting w.r.t $\mathbf{T}'[n]$ and $\mathbf{R}[n]$. For the sake of simplicity, optimal solution for $\mathbf{R}[n]$ is also obtained in two steps. Firstly, we solve for its unconstrained solution $\mathbf{R}'[n]$, which is later scaled to recover $\mathbf{R}[n]$. By assuming $\gamma = 1$, the LS cost solution for (7.15) w.r.t $\mathbf{T}'[n]$ and $\mathbf{R}[n]$ (in two steps) at iteration k can be obtained as follows:

$$\mathbf{T}'_k[n] = \text{diag}\left(\sum_m \mathbf{H}^H[m]\mathbf{R}_k^H[n]\mathbf{C}'[m]\right)\left(\text{diag}\left(\sum_m \mathbf{H}^H[m]\mathbf{R}_k^H[n]\mathbf{R}_k[n]\mathbf{H}[m]\right)\right)^{-1}, \quad (2.50)$$

$$\mathbf{R}'_k[n] = \text{diag}\left(\sum_m \mathbf{C}'[m]\mathbf{T}_k'^H[n]\mathbf{H}^H[m]\right)\text{diag}\left(\sum_m \mathbf{H}[m]\mathbf{T}_k'[n]\mathbf{T}_k'^H[n]\mathbf{H}^H[m]\right)^{-1}, \quad (2.51)$$

$$\mathbf{R}_k[n] = \mathbf{R}'_k[n]\sqrt{N_{rx}/\|\mathbf{r}'_k[n]\|_2}, \quad (2.52)$$

where $\mathbf{r}'_k[n] = \text{diag}(\mathbf{R}'_k[n])$, N_{rx} denotes the number of receive antennas and the scaling factor $\sqrt{N_{rx}}/\|\mathbf{r}'[n]\|_2$ forces the values of $\mathbf{R}[n]$ to fluctuate around 1, leading to the true physical RF calibration factors. By alternating between (2.50)-(2.52) at each subcarrier until convergence, we get the optimal $\mathbf{R}[n]^*$ and $\mathbf{T}'[n]^*$. To recover $\alpha[n]^*$, $\mathbf{T}[n]^*$ the following equations can be used:

$$\alpha[n]^* = \sqrt{N_{tx}}/\|\mathbf{T}'[n]^*\|_2, \quad (2.53)$$

$$\mathbf{T}[n]^* = \alpha[n]^* \mathbf{T}'[n]^*, \quad (2.54)$$

where $\mathbf{T}'[n]^* = \text{diag}(\mathbf{T}[n]^*)$ and N_{tx} denotes the number of transmit antennas. We then compute the calibrated internal SI MIMO channel $\hat{\mathbf{C}}[n]$ as:

$$\hat{\mathbf{C}}[n] = \alpha[n]^* \mathbf{C}'[n]. \quad (2.55)$$

The complete iterative procedure for RF calibration at each subcarrier is presented in Algorithm 3.

Algorithm 3 RF Calibration Algorithm

At each subcarrier n , initialize $\mathbf{R}[n]$, $\mathbf{T}'[n] = \mathbf{I}$.
 Initialize $\mathbf{C}'[n]$ according to (2.48) with $\alpha[n]$, $\gamma[n] = 1$.
 Repeat until convergence:

1. Update $\mathbf{T}'_k[n]$ according to (2.50).
2. Update $\mathbf{R}'_k[n]$ using (2.51).
3. Update $\mathbf{R}_k[n]$ from (2.52).

After convergence, from $\mathbf{T}'[n]^*$ get $\alpha[n]^*$, $\mathbf{T}[n]^*$ and then $\hat{\mathbf{C}}[n]$ according to (2.53)-(2.55), respectively.

2.3.7 Reduced Complexity Analog SI Nulling

We now consider a generic FD communication system with ULAs equipped with N_{tx} transmit and N_{rx} receive antennas. The SI MIMO channel is denoted with $\mathbf{C} \in \mathbb{C}^{N_{rx} \times N_{tx}}$ and we assume $N_{tx} < N_{rx}$. For our LoS model, we first assume that the near field effects of amplitude variation with distance and phase variation are negligible. Let $\mathbf{x} = [1, a, a^2, \dots]$ be a Vandermonde vector, which represents the array response of a ULA array. Then, to null \mathbf{x} we can use a Toeplitz matrix $\mathbf{T}([1, -1/a, 0..0])$ with $[1, -1/a, 0..0]$ as its first row, such that $\mathbf{T}([1, -1/a, 0..0])\mathbf{x} = 0$. Under the far field assumption, the channel matrix \mathbf{C} between ULAs, for array response of one array towards the centre of the other array, can be written as a rank 1 product of two Vandermonde vectors. At the receive side, \mathbf{C} looks

like a Vandermonde vector and hence $\mathbf{T}([1, -1/a, \dots, 0])$ in front of \mathbf{C} captures its singular part. Motivated by this idea, for SIN we propose that at the receiver side a SIN filter $\mathbf{N} \in \mathbb{C}^{N_{rx}-2 \times N_{rx}}$ with the Toeplitz structure $\mathbf{T}([1, -1/a, 0, \dots, 0])$ should be applied. If the receive array is a uniform rectangular array (URA), under the far field approximation, \mathbf{C} is still rank 1, as it is a product of the receive and transmit antenna array responses. But the URA response at the receive side should be written as a Kronecker product $\mathbf{v}_V \otimes \mathbf{v}_H$ of two ULA responses of vertical and horizontal dimensions, respectively. Moreover, if we vectorize the URA response $\mathbf{V} = \mathbf{v}_V \mathbf{v}_H^T$ as $\mathbf{v} = \text{vec}(\mathbf{V}) = \mathbf{v}_H \otimes \mathbf{v}_V$, then to null \mathbf{v} we have to use two Toeplitz SIN filters \mathbf{N}_V and \mathbf{N}_H . We can also increase the number of non-zero upper diagonals in one or the other SIN filter, or both, to change the dimensions. Even with URA, we can decide to use only one SIN filter e.g. \mathbf{N}_v , which may also work. The Kronecker Toeplitz SIN structure with two filters given by $(\mathbf{N}_V \otimes \mathbf{N}_H)\mathbf{v} = 0$ leads to nulling the product of two zeros. However, this over nulling could be justified by the fact that the Vandermonde structure we assume is approximate due to near field effects. Therefore, for the actual \mathbf{C} , the Kronecker Toeplitz SIN filter structure with two filters would perform better in nulling the SI for URAs.

Aforementioned reasoning is the motivation for a banded nulling structure that we impose on the SIN filter \mathbf{N} . Nevertheless, in reality we don't impose the Toeplitz structure on it, but we let its elements on the non-zero upper diagonals to be arbitrary complex numbers, which are then optimized for nulling purpose. For our problem formulation, we choose only one SIN filter with two super diagonals to be non zero. The objective function for SIN that we optimize is the following:

$$\min_{a_i, b_i} \|\mathbf{N}\mathbf{C}\|_F^2, \text{ for } a_i, b_i \in \mathbb{C} \text{ and } i = 1, \dots, N_{rx} - 2. \quad (2.56)$$

where a_i and b_i are the elements of the first and the second super diagonals in row i . To solve (7.16), we rewrite the matrix \mathbf{C} as a $(N_{rx} - 2)N_{tx} \times 2(N_{rx} - 2) + 1$ sparse matrix \mathbf{C}_s (2.58) and reformulate (7.16) as a quadratic cost function of complex variables with a linear constraint:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{C}_s \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathbf{e}_1^T \mathbf{x} = 1 \end{aligned} \quad (2.57)$$

where \mathbf{e}_1 is the first standard basis vector of the Euclidean space and the vector \mathbf{x} is defined as $\mathbf{x} = [1, a_1, \dots, a_{N_{rx}-2}, b_1, \dots, b_{N_{rx}-2}]^T$ contains the complex coefficients of \mathbf{N} .

$$\mathbf{C}_s = \begin{bmatrix} c_{1,1} & c_{2,1} & \mathbf{0} & c_{3,1} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{N-2,1} & \mathbf{0} & c_{N-1,1} & \mathbf{0} & c_{N,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{1,M} & c_{2,M} & \vdots & c_{3,M} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{N-2,M} & \mathbf{0} & c_{N-1,M} & \mathbf{0} & c_{N,M} \end{bmatrix} \quad (2.58)$$

The solution for (2.57) can be obtained as:

$$\mathbf{x} = \frac{1}{\mathbf{e}_1^T (\mathbf{C}_s^T \mathbf{C}_s)^{-1} \mathbf{e}_1} (\mathbf{C}_s^T \mathbf{C}_s)^{-1} \mathbf{e}_1. \quad (2.59)$$

The vector \mathbf{x} provides the optimal complex coefficients to be stacked in the two upper diagonals of \mathbf{N} , which null the SI. However, as the dimension of receive signal after SIN cannot be higher than N_{tx} , (2.56) with its reformulation (2.57), should be iterated K times to reduce the size of \mathbf{N} to $L \times N_{rx}$, with $L = N_{rx} - K$ and $L \leq N_{tx}$. At each iteration k for $k = 1, \dots, K$, the row i of \mathbf{N} which produces the highest norm for $\mathbf{A} = \mathbf{N}\mathbf{C}$ should be removed by restructuring $\mathbf{C}_s(k)$ with less columns, which correspond to the removal of coefficients a_i and b_i of \mathbf{N} . By doing so, we suppress the $N_{tx} - L$ dominant dimensions of the SI MIMO channel. The complete reduced complexity analog SIN algorithm for SIN is given in Algorithm 4.

Algorithm 4 Reduced complexity analog SIN algorithm

Initialize $\mathbf{C}_s(0)$ according to (2.58).

for $k = 1, \dots, K$

1. $\mathbf{C}_s(k) = \mathbf{C}_s(k - 1)$.
2. Evaluate $\|\mathbf{N}\mathbf{C}_s(k)\|_F^2$.
3. Update $\mathbf{C}_s(k)$ by removing its columns to remove the row in \mathbf{N} which produces the highest norm for \mathbf{A} .
4. Solve for new $\mathbf{x}(k)$ with updated $\mathbf{C}_s(k)$ in (2.59).
5. Stack $\mathbf{x}(k)$ in the upper diagonals of \mathbf{N}

end

2.3.8 Simulation and Measurement Results

In this section, simulation results are presented for the proposed RF calibration algorithm applied to a 4×4 FD OFDM system with $N_s = 296$ subcarriers implemented with the USRP N310. Then to evaluate the performance of our reduced complexity analog SIN algorithm, we present simulation results for a generic FD system with ULAs equipped with 4 transmit and 6 receive antennas and with one SIN filter \mathbf{N} of size 4×6 and 3×6 .

2.3.9 RF Calibration Results

We considered a 4×4 FD OFDM MIMO communication system. The experiments were performed on a setup consisting of a USRP N310 software defined radio and Huawei patch antenna cards as part of the OAI platform, Eurecom. Each card has four patch antennas shaped like an H. We observed during our experiments that the channel is also affected by the orientation of the antenna cards and thus we perform the channel measurements for two orientations H and \perp of the transmitter and the receiver. All the measurements are taken indoors at the carrier frequency $f_c = 2.585$ GHz.

For SI MIMO channel measurement, we transmit an OFDM frame of bandwidth 5 MHz with 296 subcarriers. The channel measurements are taken for three different

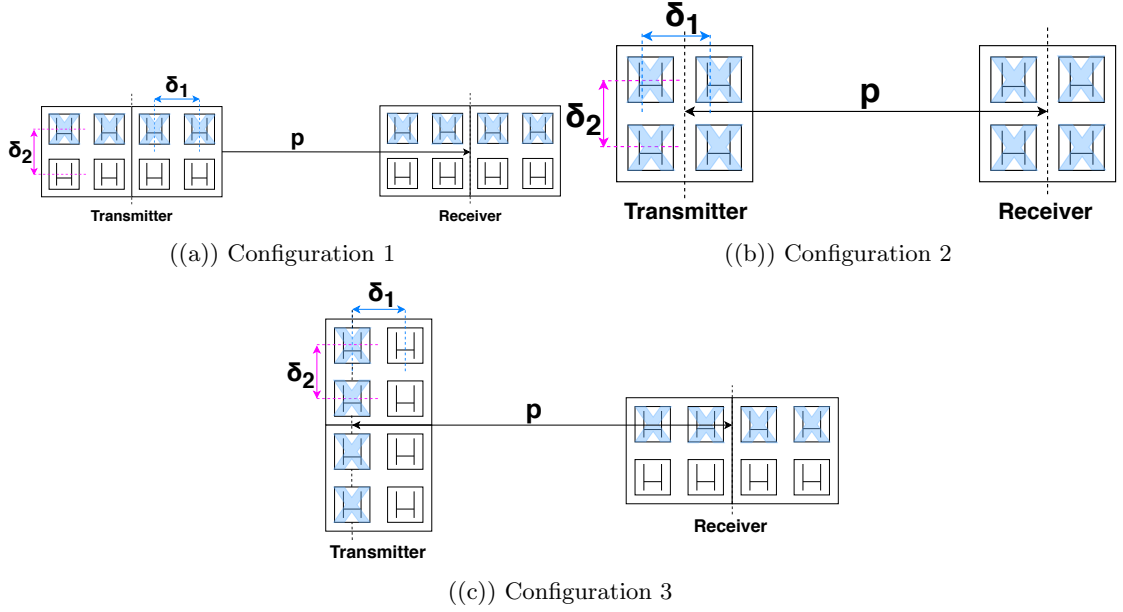


Figure 2.8: Transmitter and Receiver Antenna Layout

configurations of the transmit and receive antennas as shown in Fig. 2.8. The distances d_{ij} are a function of the distance between the transmitter and the receiver P , the horizontal inter-antenna spacing δ_1 and the vertical inter-antenna spacing δ_2 . The measurements are done for different values of P .

Figures 2.9-2.10 show the singular value profile (SVP) for 296 subcarriers of the measured SI MIMO channel \mathbf{H} , the SI MIMO propagation channel \mathbf{C} generated according to the near field LoS model (2.48) and the calibrated SI MIMO channel $\hat{\mathbf{C}}$ obtained by using Algorithm 3. Fig. 2.9 corresponds to the transmitter and receiver in configuration 3 with $P = 47$ cm and transmitter in orientation H and receiver in orientation \perp . Fig. 2.10 shows the SVP of the channels for transmitter and receiver in configuration 3 with $P = 37.1$ cm and transmitter and receiver in orientation H. It can be observed from the Figures that the measured channel can be approximated as rank 2, which is evident after the calibration process in $\hat{\mathbf{C}}$.

We do not include the results for configuration 1 and 2 as the measured (and calibrated) channels for the considered antenna layouts resulted to be full rank.

2.3.10 Reduced Complexity SIN Results

In this section, we present the simulation results for our SIN approach applied to an FD system with ULAs consisting of 4 transmit antennas and 6 receive antennas. The internal SI MIMO propagation channel \mathbf{C} is generated according to the near field LoS model (2.48) with $\gamma, \alpha = 1$. For evaluation of the proposed SIN algorithm, we assume the additive signal model $\mathbf{y} = \mathbf{C}\mathbf{s} + \mathbf{v}$, where \mathbf{s} is the transmit signal and \mathbf{v} is the intended receive signal. We further assume that the signals \mathbf{s} and \mathbf{v} are white. However, in reality

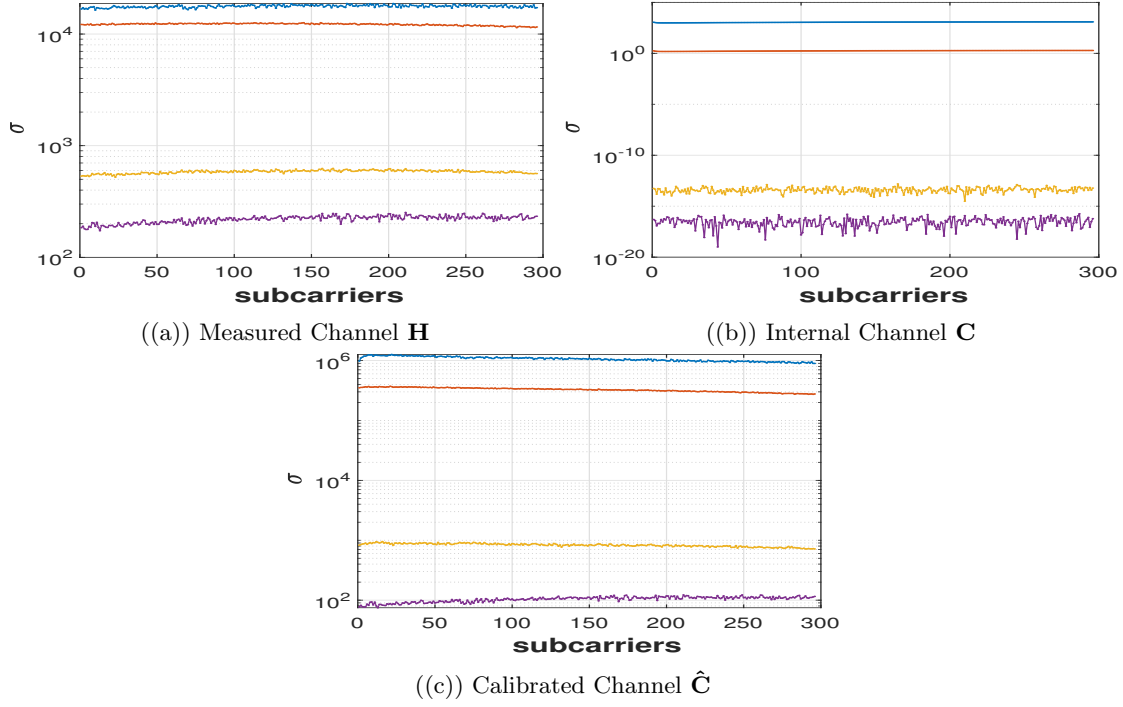


Figure 2.9: SVP of the SI MIMO channel for configuration 3 with $P = 47$ cm and with transmitter and receiver in orientation H and I, respectively.

they are not white but this assumption allows us analyze the upper bounds for the proposed SIN Algorithm 4 for ULAs with different configurations shown in Fig. 2.8. For evaluation purposes, we define the metric SI-to-signal-ratio ($SISR$) as the ratio of effective SI power to effective receive signal power. The $SISR$ before and after the SIN filter \mathbf{N} is denoted with $SISR_i$ and $SISR_o$, respectively. For the assumed signal model above, $SISR_i$ and $SISR_o$ can be calculated as:

$$SISR_i = \frac{\sigma_s^2 \text{Tr}(\mathbf{C}\mathbf{C}^H)}{\sigma_v^2(N_{rx} + 2)} \quad SISR_o = \frac{\sigma_s^2 \text{Tr}(\mathbf{N}\mathbf{C}\mathbf{C}^H\mathbf{N}^H)}{\sigma_v^2 \text{Tr}(\mathbf{N}\mathbf{N}^H)}. \quad (2.60)$$

where σ_s^2 and σ_v^2 denote the variances of \mathbf{s} and \mathbf{v} , respectively.

Fig. 2.11 shows the ratio $SISR_o/SISR_i$ as a function of distance, for \mathbf{C} 6×4 and \mathbf{N} 4×6 and 3×6 . It is evident that the proposed SIN algorithm works better with configuration 1. There is no significant variation in $SISR_o/SISR_i$ for configuration 3 as the distance between transmit and receive antennas increases but for configuration 1 the ratio decays rapidly which demonstrates the nulling of SI. It is also evident in Fig. 2.11 the advantage of reducing the receive signal dimension by 1 with $K = 3$, which removes the most dominant dimension of the SI signal for both configurations. Results are not reported for configuration 2 because we considered only one SIN filter and configuration 2 requires two SIN filters \mathbf{N}_V and \mathbf{N}_H to null SI in vertical and horizontal dimensions to achieve good results.

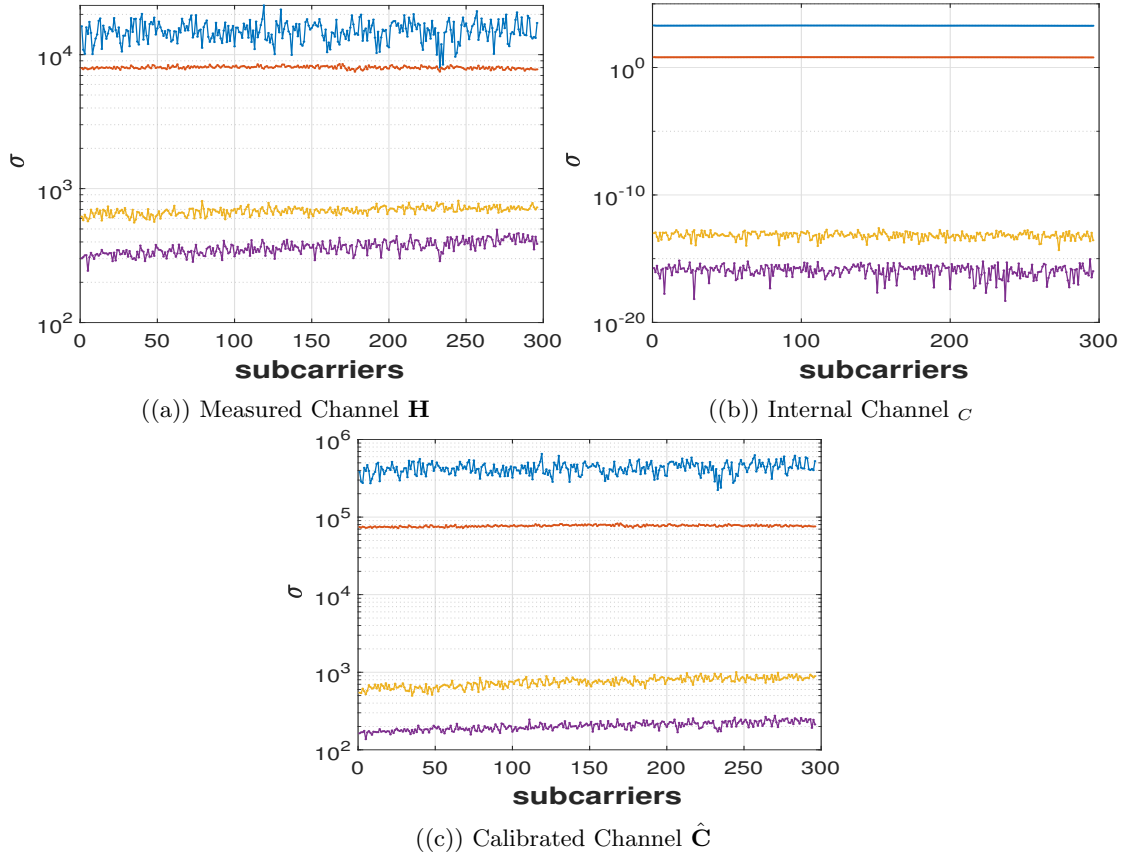


Figure 2.10: SVP of the SI MIMO channel for configuration 3 with $P = 37.1$ cm and with transmitter and receiver in orientation H.

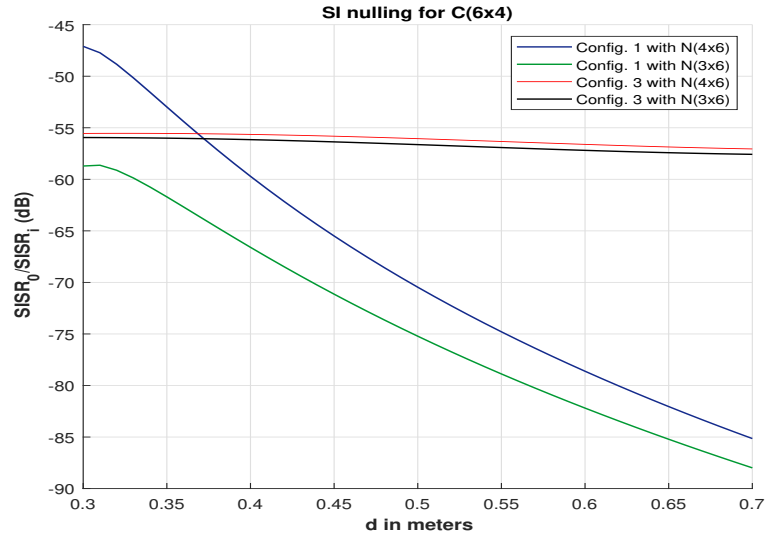


Figure 2.11: Results obtained with the SI nulling algorithm for configuration 1 and 3 for $\mathbf{C}(6 \times 4)$, $\mathbf{N}(4 \times 6)$ and $\mathbf{N}(3 \times 6)$.

2.3.11 Conclusions

This section has presented a new HSIC architecture which is prominent for the upcoming FD massive MIMO communication scenarios, as its complexity at the analog cancelling stage scales linearly only with the number of receive antennas. A new RF calibration algorithm is also proposed, which is later used for the analysis of SI MIMO channel from the measured channel with USRP N310. Results demonstrate that we can achieve a SI MIMO channel of rank 2 with configuration 3 for a 4×4 for a MIMO FD system, which reduces the number of constraints to be imposed to ZF the SI signal. Finally, a reduced complexity analog SIN algorithm is also proposed.

Chapter 3

Hybrid Beamforming for Point-to-Point and IAB Millimeter Wave Full Duplex

HYBF is a potent tool as it enables designing FD systems with a less number of RF chains than the number of antennas. This section first presents a novel HYBF design for Integrated access and backhaul (IAB) mmWave MIMO FD systems serving multi-antenna users. Then a novel HYBF design that generalizes the point-point mMIMO FD communication in the mmWave is presented. The content of this chapter is available in [C1],[C3] and [C7]. The contribution [C2] is not included in this thesis, being a special case of the contribution [C1].

3.1 Full Duplex Relay for IAB in Millimeter Wave

This section considers the case of a FD relay for IAB in mmWave. The literature on the considered scenario is available in [44, 45, 46, 47, 48, 49, 50, 51]. In [44], the authors proposed the first hybrid beamforming design for an amplify and forward FD relay for the mmWave backhaul link. In [45], a novel hybrid beamforming design for a FD relay assisted mmWave macrocell scenario is investigated. In [46], a robust hybrid design for an amplify and forward mmWave FD relay with imperfect channel state information (CSI) serving multiple single antenna users is studied. In [47], a joint beamforming, positioning and power control scheme for a mmWave FD UAV relay is presented. In [48], energy efficient precoding design for FD relay is proposed. The authors of [49] present a two-time scale hybrid beamforming for multiple FD relays scenario. In [50] a robust hybrid beamforming design for a network of FD relays under imperfect CSI is investigated. In [51], hybrid beamforming for a single stream two hop amplify and forward FD relay is proposed. Note that the available hybrid beamforming designs for mmWave are limited to one multi-antenna uplink and downlink user communicating with a FD relay or single antenna multiple downlink users served by a FD relay.

We consider the scenario of one HD BS communicating with multiples multi-antenna

users to be served with multiples streams. The users are assumed to be very far from the BS with no direct link and the communication takes place with a massive MIMO FD relay, which simultaneously receive data from the HD BS and transmits it towards the HD multi-antenna users. The scenario under consideration represents an IAB framework. Compared to the state-of-the-art, we also assume that the transmit and receive chains of the FD relay have LDR which dictates the overall noise level at the FD relay. To the authors best knowledge, our work is first contribution to mmWave FD relay system with LDR range which serves multiple multi-antenna users with multiple streams. We present a novel HYBF and combining design to maximize the WSR from the HD BS to the users aided with the FD relay. The overall optimization problem is separated into two sub-problems under the assumption that the SI at the FD relay is cancelled, for example based on the well known SIC techniques. The global optimization problem is decomposed into two sub-problems and an alternating optimization approach is adopted.

3.1.1 System Model

We consider a mmWave multiuser FD system consisting of one mMIMO HD base station and one mMIMO FD relay serving K multi-antenna users. Let $\mathcal{K} = \{1, ..K\}$ denote the set containing the indices of the users. The BS is assumed to be equipped with B_t transmit antennas and B_r RF chains, respectively. FD relay is assumed to be equipped with N receive and transmit antennas and M receive and transmit RF chains, respectively, and has a SIC scheme to cancel the SI up to a certain level. We assume a narrow-band flat fading radio channel and with no direct link between the BS and the end users due to severe attenuation at the mmWave. Let $\mathbf{s}_k \in \mathbb{C}^{d_k \times 1}$ denote the white unitary data streams intended for user k . Let $\mathbf{F}_k \in \mathbb{C}^{B_r \times d_k}$ and $\mathbf{G}_{RF} \in \mathbb{C}^{B_t \times B_r}$ denote the digital and the analog beamformer for the HD base station. The signal received at the relay can be written as

$$\mathbf{y}_r = \mathbf{H}_{r,b} \mathbf{G}_{RF} \sum_{j \in \mathcal{K}} \mathbf{F}_j \mathbf{s}_j + \mathbf{n}_{SI} + \mathbf{n}_r + \mathbf{e}_r \quad (3.1)$$

where $\mathbf{H}_{r,b} \in \mathbb{C}^{M \times B_t}$ denote the channel from the BS to the relay, $\mathbf{e}_r \sim \mathcal{CN}(0, \sigma_r^2 \mathbf{I})$ denotes the thermal noise vector at the relay and $\mathbf{n}_{SI} \sim \mathcal{CN}(0, \sigma_{SI}^2 \mathbf{I})$ denotes the residual SI. The distortions due to LDR at the receiver side are denoted

$$\mathbf{n}_r \sim \mathcal{CN}(0, \beta \text{diag}(\mathbf{\Phi}_r)) \quad (3.2)$$

where $\beta \ll 1$, $\mathbf{\Phi}_r = \text{Cov}(\mathbf{r})$, and $\mathbf{r} = \mathbf{y}_r - \mathbf{n}_r$ denotes the undistorted received signal. Let $\mathbf{U}_r \in \mathbb{C}^{M \times N}$ and $\mathbf{U}_t \in \mathbb{C}^{N \times M}$ denote the analog combiner and beamformer at the FD relay, respectively. Let $\mathbf{W}_k \in \mathbb{C}^{d_k \times M}$ and $\mathbf{E}_k \in \mathbb{C}^{M \times d_k}$ denote the digital combiner and beamformer for user k at the relay, respectively. We assume that there is no delay in processing and after the hybrid combining, the data streams of user k can be estimated as

$$\hat{\mathbf{s}}_k = \mathbf{W}_k \mathbf{U}_r \mathbf{y}_r. \quad (3.3)$$

Let $\mathbf{Q}_{j,r} = \mathbf{G}_{RF} \mathbf{F}_j \mathbf{F}_j^H \mathbf{G}_{RF}^H$ denote the transmit covariance matrix for user k transmitted to the FD relay by the HD base station. Let \bar{k} denote all the indices in \mathcal{K} , without the

index k . The received (signal plus) interference and noise covariance matrix for user k at the relay is denoted with $(\mathbf{R}_{k,r}) \mathbf{R}_{\bar{k},r}$ and after the hybrid combining are given by

$$\Phi_r = \mathbf{H}_{r,b} \sum_{j \in \mathcal{K}, j \neq k} \mathbf{Q}_{j,r} \mathbf{H}_{r,b}^H + \sigma_{SI}^2 \mathbf{I} + \sigma_r^2 \mathbf{I}, \quad (3.4)$$

$$\mathbf{R}_{\bar{k},r} = \mathbf{W}_k \mathbf{U}_r (\Phi_r + \beta \text{diag}((\Phi_r)) \mathbf{U}_r^H \mathbf{W}_k^H, \quad (3.5)$$

$$\mathbf{R}_{k,r} = \mathbf{W}_k \mathbf{U}_r \mathbf{H}_{r,b} \mathbf{Q}_{k,r} \mathbf{H}_{r,b}^H \mathbf{U}_r^H \mathbf{W}_k^H + \mathbf{R}_{\bar{k},r}, \quad (3.6)$$

where Φ_r denotes the undistorted received covariance matrix. The WSR maximization problem from the HD base station to the FD relay can be formally stated as

$$\max_{\substack{\mathbf{U}_r, \mathbf{W}_k \\ \mathbf{G}_{RF}, \mathbf{F}_k}} \sum_{k \in \mathcal{K}} w_k \ln \det(\mathbf{R}_{\bar{k},r}^{-1} \mathbf{R}_{k,r}) \quad (3.7a)$$

$$\text{s.t.} \quad \text{Tr}\left(\sum_{k \in \mathcal{K}} \mathbf{Q}_{j,r}\right) \preceq p_b, \quad (c_1) \quad (3.7b)$$

$$|\mathbf{U}_r(m, n)|^2 = 1, \quad \text{and} \quad |\mathbf{G}_{RF}(m, n)|^2 = 1 \quad (3.7c)$$

where p_b denotes the total sum-power constraint, (3.7c) denote the unit modulus constraint on the analog part. Once the data streams are detected by the FD relay, they are then forwarded to the multi-antenna users with optimal power allocation based on the channel state information from the relay to the users. The transmit signal from the FD relay can be written as

$$\mathbf{y}_t = \mathbf{U}_t \sum_{j \in \mathcal{K}} \mathbf{E}_k \hat{\mathbf{s}}_j + \mathbf{n}_t \quad (3.8)$$

where \mathbf{n}_t denotes the transmit LDR distortions, which can be modelled as

$$\mathbf{n}_t \sim \mathcal{CN}(0, k \text{diag}(\Phi_t)), \quad (3.9)$$

where $\Phi_t = \text{Cov}(\mathbf{y}_0)$, where $\mathbf{y}_0 = \mathbf{y}_t - \mathbf{n}_t$ denotes the undistorted transmitted signal. Let $\mathbf{H}_k \in \mathbb{C}^{M_k \times N}$ denote the MIMO channel from the FD relay to the HD user k . The received signal at user k can be written as

$$y_k = \mathbf{H}_k (\mathbf{U}_t \sum_{j \in \mathcal{K}} \mathbf{E}_j \hat{\mathbf{s}}_j + \mathbf{n}_t) + \mathbf{n}_k, \quad (3.10)$$

where $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma_k^2 \mathbf{I})$ denotes the noise vector at user k . Let $\mathbf{Q}_k = \mathbf{U}_t \mathbf{E}_k \mathbf{E}_k^H \mathbf{U}_t$ denote the transmit covariance for user k from the FD relay. The received (signal plus) interference and noise covariance at user k is denoted with $(\mathbf{R}_k) \mathbf{R}_{\bar{k}}$ and can be written as

$$\mathbf{R}_{\bar{k}} = \mathbf{H}_k \sum_{j \in \mathcal{K}, j \neq k} \mathbf{Q}_j \mathbf{H}_k^H + k \mathbf{H}_k \text{diag}\left(\sum_{j \in \mathcal{K}} \mathbf{Q}_j\right) \mathbf{H}_k^H + \sigma_k^2 \mathbf{I}, \quad (3.11)$$

$$\mathbf{R}_k = \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^H + \mathbf{R}_{\bar{k}}. \quad (3.12)$$

The WSR maximization problem from FD relay to the HD multi-antenna users can be now formally stated as

$$\max_{\mathbf{U}_t, \mathbf{E}_k} \sum_{k \in \mathcal{K}} w_j \text{Indet}(\mathbf{R}_{\bar{k}}^{-1} \mathbf{R}_k) \quad (3.13a)$$

$$\text{s.t. } \text{Tr}\left(\sum_{k \in \mathcal{K}} \mathbf{Q}_j\right) \preceq p_b, \quad \text{and} \quad |\mathbf{U}_t(m, n)|^2 = 1 \quad (3.13b)$$

where (3.13b) denote the total sum-power constraint and the unit modulus constraint.

3.1.2 Problem Simplification

The problem (3.7) and (3.13) are non-concave due to the interference terms and finding a global optimum is challenging. Hence to render a feasible solution, we consider the minorization maximization approach to solve (3.7) and (3.13) in two steps. Let WSR^r and WSR denote the WSR from the HD base station to the relay and from the FD relay to the HD users, respectively. They can be written as the weighted rate (WR) of user k (WR_k or $WR_{k,r}$) and the WSR of users \bar{k} as

$$WSR_r = WR_{k,r} + WR_{\bar{k},r}, \quad WSR = WR_k + WR_{\bar{k}} \quad (3.14)$$

where $WR_{k,r}$ is concave in $\mathbf{Q}_{k,r}$ and $WR_{\bar{k},r}$ is non-concave in $\mathbf{Q}_{k,r}$. Similarly, WR_k is concave in \mathbf{Q}_k and $WR_{\bar{k}}$ is non-concave in \mathbf{Q}_k . Since a linear function is simultaneously convex and concave, DC programming [38] introduces the first order Taylor series expansion of $WR_{\bar{k},r}$ in $\mathbf{Q}_{k,r}$ around $\hat{\mathbf{Q}}_{k,r}$ and of $WR_{\bar{k}}$ in \mathbf{Q}_k around $\hat{\mathbf{Q}}_k$. Let $\hat{\mathbf{Q}}$ denote the collection of $\hat{\mathbf{Q}}_{k,r}$ and $\hat{\mathbf{Q}}_k$. The new tangent expressios for the WSR can be written as

$$\underline{WR}_{\bar{k},r}(\mathbf{Q}_k, \hat{\mathbf{Q}}) = WR_{k,r}(\mathbf{Q}_{k,r}, \hat{\mathbf{Q}}) - \text{Tr}((\mathbf{Q}_{k,r} - \hat{\mathbf{Q}}) \mathbf{G}_{k,r}), \quad (3.15)$$

$$\underline{WR}_{\bar{k}}(\mathbf{Q}_k, \hat{\mathbf{T}}) = WR_k(\mathbf{Q}_k, \hat{\mathbf{Q}}) - \text{Tr}((\mathbf{Q}_k - \hat{\mathbf{Q}}) \mathbf{G}_k), \quad (3.16)$$

where $\mathbf{G}_{k,r}$ and \mathbf{G}_k denote the gradients of $WR_{\bar{k},r}$ and $WR_{\bar{k}}$ with respect to $\mathbf{Q}_{k,r}$ and \mathbf{Q}_k , respectively.

$$\mathbf{G}_{k,r} = \sum_{j \in \mathcal{K}, j \neq k} w_j (\mathbf{H}_{r,b}^H [\mathbf{U}_j^H \mathbf{W}_j^H (\mathbf{R}_{\bar{j},r}^{-1} - \mathbf{R}_{j,r}^{-1}) \mathbf{W}_j \mathbf{U}_j - \beta \text{diag}(\mathbf{U}_j^H \mathbf{W}_j^H (\mathbf{R}_{\bar{j},r}^{-1} - \mathbf{R}_{j,r}^{-1}) \mathbf{W}_j \mathbf{U}_j)] \mathbf{H}_{r,b}), \quad (3.17)$$

$$\mathbf{G}_k = \sum_{j \in \mathcal{K}, j \neq k} w_j (\mathbf{H}_j^H [\mathbf{R}_{\bar{j}}^{-1} - \mathbf{R}_j^{-1}] \mathbf{H}_j + k \text{diag}(\mathbf{H}_j^H [\mathbf{R}_{\bar{j}}^{-1} - \mathbf{R}_j^{-1}] \mathbf{H}_j)). \quad (3.18)$$

Note that, the tangent expressios constitutes a touching lower bound for (3.7) and (3.13), hence the DC programming approach is also a MM approach [26], regardless of the reparameterization of WSR as a function of beamformers. Let λ_b and λ_r be the Lagrange multipliers associated with the sum-power at the base station and at the FD relay. Let

$$\mathbf{A} \triangleq \mathbf{G}_{RF}^H \mathbf{H}_{r,b}^H \mathbf{U}_r^H \mathbf{W}_k^H \mathbf{R}_{k,FD}^{-1} \mathbf{W}_k \mathbf{U}_r \mathbf{H}_{r,b} \mathbf{G}_{RF}, \quad (3.19a)$$

$$\mathbf{B} \triangleq \mathbf{G}_{RF}^H (\mathbf{G}_k^r + \lambda_b \mathbf{I}) \mathbf{G}_{RF}, \quad (3.19b)$$

$$\mathbf{C} \triangleq \mathbf{U}_t^H \mathbf{H}_k^H \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{U}_t, \quad (3.19c)$$

$$\mathbf{D} \triangleq \mathbf{U}_t^H (\mathbf{G}_k + \lambda_r \mathbf{I}) \mathbf{U}_t, \quad (3.19d)$$

and dropping the constant terms, reparameterizing back the covariance matrices, performing this linearization for all users, augmenting the WSR^r with WSR with their constraints, yields the following Lagrangians

$$\max_{\substack{\mathbf{U}_r, \mathbf{W}_k \\ \mathbf{G}_{RF}, \mathbf{F}_k}} \sum_{k \in \mathcal{K}} w_k \ln \det(\mathbf{I} + \mathbf{F}_k^H \mathbf{A} \mathbf{F}_k) - \text{Tr}(\mathbf{F}_k^H \mathbf{B} \mathbf{F}_k) \quad (\text{P1}) \quad (3.20a)$$

$$\max_{\mathbf{U}_t, \mathbf{E}_k} \sum_{k \in \mathcal{K}} w_k \ln \det(\mathbf{I} + \mathbf{E}_k^H \mathbf{C} \mathbf{E}_k) - \text{Tr}(\mathbf{E}_k^H \mathbf{D} \mathbf{E}_k) \quad (\text{P2}) \quad (3.20b)$$

3.1.3 Hybrid Beamforming

This section present the solution for the problems (P_1) and (P_2) based on the problem simplification (3.20).

Digital Beamforming

To optimize the digital transmit beamformers for the base station and FD relay, we first take the derivative of (3.33) and (3.34) with respect to \mathbf{F}_k and \mathbf{E}_k , which yield the following KKT conditions

$$\mathbf{A} \mathbf{F}_k (\mathbf{I} + \mathbf{F}_k^H \mathbf{A} \mathbf{F}_k)^{-1} - \mathbf{B} \mathbf{F}_k = 0, \quad (3.21)$$

$$\mathbf{C} \mathbf{E}_k (\mathbf{I} + \mathbf{E}_k^H \mathbf{C} \mathbf{E}_k)^{-1} - \mathbf{D} \mathbf{E}_k = 0, \quad (3.22)$$

Theorem 2. *Fixed the other variables, the digital beamformers \mathbf{F}_k and \mathbf{E}_k at each iteration can be optimized as the dominant generalized eigenvector of the pairs*

$$\mathbf{F}_k = \mathbf{D}_{d_k}(\mathbf{A}, \mathbf{B}), \quad \mathbf{E}_k = \mathbf{D}_{d_k}(\mathbf{C}, \mathbf{D}) \quad (3.23)$$

Proof. The result follows similarly as the proof for Proposition 1 [38]. \square

Analog Beamforming

To optimize the analog beamformer, we exploit the knowledge that the optimal fully digital beamformers $\mathbf{F}_{k,dig}$ and $\mathbf{E}_{k,dig}$ are known and computed based on the procedure provided herein. Let $\mathbf{F}^{opt} \in \mathbb{C}^{B_t \times d_1 + \dots + d_K}$ and $\mathbf{E}^{opt} \in \mathbb{C}^{N \times d_1 + \dots + d_K}$ contain the fully digital beamformers at the base station and FD relay, respectively. Let $\mathbf{F}_{RF}^{opt} \in \mathbb{C}^{B_r \times d_1 + \dots + d_K}$ and $\mathbf{E}_{RF}^{opt} \in \mathbb{C}^{M \times d_1 + \dots + d_K}$ contain all the digital beamforming vectors for the matched filter solution of the size of RF chains. Since the analog beamformer is common to all the

users, we optimize the analog beamformer as the one which minimize the error between the fully digital and hybrid solution. Formally, the two unconstrained problem can be stated as

$$\min_{\mathbf{G}_{RF}} \|\mathbf{F}^{opt} - \mathbf{G}_{RF}\mathbf{F}_{RF}^{opt}\|, \quad \min_{\mathbf{U}_t} \|\mathbf{E}^{opt} - \mathbf{U}_t\mathbf{E}_{RF}^{dig}\|, \quad (3.24)$$

and solving it yield the solution

$$\mathbf{G}_{RF} = (\mathbf{F}^{opt}(\mathbf{F}_{RF}^{opt})^H)(\mathbf{F}_{RF}^{opt}(\mathbf{F}_{RF}^{opt})^H)^{-1}, \quad (3.25)$$

$$\mathbf{U}_t = (\mathbf{E}^{opt}(\mathbf{E}_{RF}^{dig})^H)(\mathbf{E}_{RF}^{dig}(\mathbf{E}_{RF}^{dig})^H)^{-1}. \quad (3.26)$$

which it normalized to obtain the unit-modulus solution as $\mathbf{G}_{RF} = \angle \mathbf{G}_{RF}$ and $\mathbf{U}_t = \angle \mathbf{U}_t$.

Digital Combining

To further improve the rate and suppress the interference, we design the combiners based on the MMSE criteria. Namely, let $\mathbf{E}_k = \mathbf{s}_k - \hat{\mathbf{s}}_k$ denote the error vector for user k and the error covariance matrix \mathbf{E}_{rr} at the relay can be written as

$$\mathbf{E}_{rr} = \mathbf{R}_{k,FD} - \mathbf{I} + (\mathbf{W}_k\mathbf{U}_r\mathbf{H}_{r,b}\mathbf{G}_{RF}\mathbf{F}_j + \mathbf{F}_k^H\mathbf{G}_{RF}^H\mathbf{H}_{r,b}\mathbf{U}_r^H)\mathbf{W}_k^H \quad (3.27)$$

and minimizing its trace leads to the solution

$$\mathbf{W}_k = \mathbf{F}_k^H\mathbf{V}^H\mathbf{H}_{r,b}\mathbf{U}_r^h[\Phi_r + \beta\text{diag}(\Phi_r)]^{-1}, \quad \forall k \in \mathcal{K}. \quad (3.28)$$

Analog Combining

As the analog combiner is common to all the users, we adopt the same mechanism as in (3.24) to optimize it. Let $\mathbf{W}^{opt} \in \mathbb{C}^{d_1+\dots+d_K \times N}$ contain the combining vectors to the fully digital solution obtained from (3.28) with $\mathbf{U}_r = \mathbf{I}, \forall k \in \mathcal{K}$. Let $\mathbf{W}_{RF}^{opt} \in \mathbb{C}^{d_1+\dots+d_K \times M}$ contain the combining vectors of correct size with limited number of RF chains as a matched filter solution. The analog combiner can be obtained by solving the following optimization problem

$$\min_{\mathbf{U}_r} \|\mathbf{W}^{opt} - \mathbf{U}_r\mathbf{W}_{RF}^{opt}\|, \quad (3.29)$$

which leads to the solution

$$\mathbf{U}_r = (\mathbf{W}^{opt}(\mathbf{W}_{RF}^{opt})^H)(\mathbf{W}_{RF}^{opt}(\mathbf{W}_{RF}^{opt})^H)^{-1}. \quad (3.30)$$

To meet the unit-modulus constraint, we normalize the amplitudes to one as $\mathbf{U}_r = \angle \mathbf{U}_r$. Note that the digital beamformers (3.23) provide the optimized beamforming directions, but not the optimal power. Therefore, we normalize the columns of the digital beamformers \mathbf{E}_k and \mathbf{F}_k to be unit-norm, which allows designing an optimal power allocation scheme. Let $\Sigma_k^1, \Sigma_k^2, \mathbf{S}_k^1$ and \mathbf{S}_k^2 be defined as

$$\Sigma_k^1 = \mathbf{F}_k^H\mathbf{A}\mathbf{F}_k, \quad \Sigma_k^1 = \mathbf{F}_k^H\mathbf{B}\mathbf{F}_k, \quad (3.31)$$

$$\mathbf{S}_k^1 = \mathbf{E}_k^H \mathbf{C} \mathbf{E}_k, \quad \mathbf{S}_k^1 = \mathbf{E}_k^H \mathbf{D} \mathbf{E}_k. \quad (3.32)$$

and assuming that the optimal digital beamformers are computed at the last based on the results in Theorem 3.3.1, (3.31) and (3.32) are diagonal. The optimal power allocation can be included by solving the following optimization problem

$$\max_{\mathbf{P}_{k,BS}} \sum_{k \in \mathcal{K}} w_k \ln \det(\mathbf{\Sigma}_k^1 \mathbf{P}_{k,BS}) - \text{Tr}(\mathbf{\Sigma}_k^2 \mathbf{P}_{k,BS}) \quad (3.33a)$$

$$\max_{\mathbf{P}_{k,FD}} \sum_{k \in \mathcal{K}} w_k \ln \det(\mathbf{I} + \mathbf{S}_k^1 \mathbf{P}_{k,FD}) - \text{Tr}(\mathbf{S}_k^2 \mathbf{P}_{k,FD}) \quad (3.34a)$$

where $\mathbf{P}_{k,BS}$ and $\mathbf{P}_{k,FD}$ denote the optimal power allocation for the BS and the FD relay. Solving it yield the optimal power allocation

$$\mathbf{P}_{k,BS} = (w_k(\mathbf{\Sigma}_k^2)^{-1} - (\mathbf{\Sigma}_k^1)^{-1})^+ \quad (3.35)$$

$$\mathbf{P}_{k,FD} = (w_k(\mathbf{S}_k^2)^{-1} - (\mathbf{S}_k^1)^{-1})^+ \quad (3.36)$$

where $(\mathbf{X})^+ = \max(\mathbf{0}, \mathbf{X})$. To search for the optimal Lagrange multipliers λ_b and λ_r satisfying the sum power constraints, we adopt the Bisection method in the search range $[0, \lambda_{b,max}]$ and $[0, \lambda_{r,max}]$, respectively. The complete alternating optimization procedure to iteratively optimize the beamformers and combiners is formally stated in Algorithm 5.

3.1.4 Convergence

The convergence can be proved by noticing that the KKT conditions of the simplified problem and the original problem are the same. Each Iteration leads to and increase in the WSR sum-rate which ensures convergence. A formal proof can be stated similarly as the one given Proposition 3 [38].

3.1.5 Simulation Results

In this section, we evaluate the performance of our proposed HYBF and combining design. The users channel and the channel from the BS to the relay are modelled with the path wise channel model, with each channel matrix modelled as

$$\mathbf{H} = \sqrt{\frac{M_0 N_0}{N_c N_p}} \sum_{n_c=1}^{N_c} \sum_{n_p=1}^{N_p} \alpha_k^{(n_p, n_c)} \mathbf{a}_r(\phi_k^{n_p, n_c}) \mathbf{a}_t^T(\theta_k^{n_p, n_c}), \quad (3.37)$$

where N_c and N_p denote the number of clusters and number of rays, respectively, $\alpha_k^{(n_p, n_c)} \sim \mathcal{CN}(0, 1)$ is a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively, and $\mathbf{a}_r(\phi_k^{n_p, n_c})$ and $\mathbf{a}_t^T(\theta_k^{n_p, n_c})$ denote the receive and transmit antenna array response with angle of arrival (AoA) $\phi_k^{n_p, n_c}$ and angle of departure (AoD) $\theta_k^{n_p, n_c}$. The BS and relay are assumed to have uniform linear arrays with $B_t = N = 100$ and RF chains $B_r = M = 32$. The

Algorithm 5 Hybrid Beamforming for IAB FD

Given: The CSI and rate weights.

Initialize: the beamformers $\forall k \in \mathcal{K}$.

Iterate first with $\mathbf{U}_r, \mathbf{V}, \mathbf{U}_t = \mathbf{I}$ to get $\mathbf{F}^{opt}, \mathbf{E}^{opt}$ and \mathbf{W}^{Dig}

Set: $\angle \mathbf{V}, \angle \mathbf{U}_t, \angle \mathbf{U}_r$ with (3.25)-(3.26) and (3.30)

Repeat until convergence

for: $k = 1, \dots, K$.

 Compute $\mathbf{G}_{k,r}$ with (3.17).

 Compute \mathbf{F}_k with (3.23) and normalize it.

 Set $\underline{\lambda}_b = 0$ and $\overline{\lambda}_b = \lambda_{b,max}$.

 Repeat until convergence

 set $\lambda_b = (\underline{\lambda}_b + \overline{\lambda}_b)/2$.

 Compute $\mathbf{P}_{k,BS}$ with (3.35),

 If constraint for μ_i is violated,

 set $\underline{\lambda}_b = \lambda_b$, else $\overline{\lambda}_b = \lambda_b$,

 Set $\mathbf{Q}_{j,FD} = \mathbf{V}\mathbf{F}_j\mathbf{P}_{k,BS}\mathbf{F}_j^H\mathbf{V}^H$

 Update \mathbf{W}_k with (3.28)

 Next k .

Repeat until convergence

for: $k = 1, \dots, K$.

 Compute \mathbf{G}_k with (3.18).

 Compute \mathbf{E}_k with (3.23) and normalize it.

 Set $\underline{\lambda}_r = 0$ and $\overline{\lambda}_r = \lambda_{r,max}$.

 Repeat until convergence

 set $\lambda_r = (\underline{\lambda}_r + \overline{\lambda}_r)/2$.

 Compute $\mathbf{P}_{k,FD}$ with (3.36),

 If constraint for λ_r is violated,

 set $\underline{\lambda}_r = \lambda_r$, else $\overline{\lambda}_r = \lambda_r$,

 Set $\mathbf{Q}_{j,FD} = \mathbf{U}_t\mathbf{E}_k\mathbf{P}_{k,FD}\mathbf{E}_k^H\mathbf{U}_t^H$

 Next k .

SNR for the HD BS and the FD relay is defined as $\text{SNR} = p_b/\sigma_b^2 = p_r/\sigma_r^2$. The maximum transmit power is normalized to one and the thermal noise variance is chosen to meet the desired SNR. We assume that the SI at the FD relay is cancelled up to the LDR noise level, different values of k and β vary the residual SI level. We consider a scenario of two downlink users with 5 antennas to be served with two data streams each. For comparison purpose, we define a fully digital HD system with no LDR noise, and which splits the resource in time and takes two time slots to transmit from the HD BS to the users through the FD relay.

Figure 3.1 shows the average WSR as a function of the overall LDR noise level with $\text{SNR} = -5$ dB. It can be seen that when the LDR noise variance is below the overall noise level, the FD communication system achieves an additional gain of 98% compared to the HD system. However, as the overall LDR noise variance increases, the overall maximum

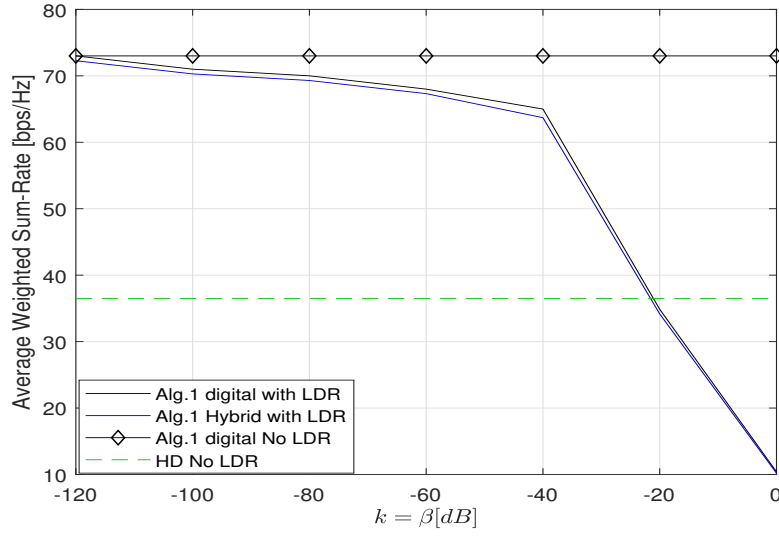


Figure 3.1: Average WSR as function of $k = \beta$ dB, with $SNR = -5$ dB.

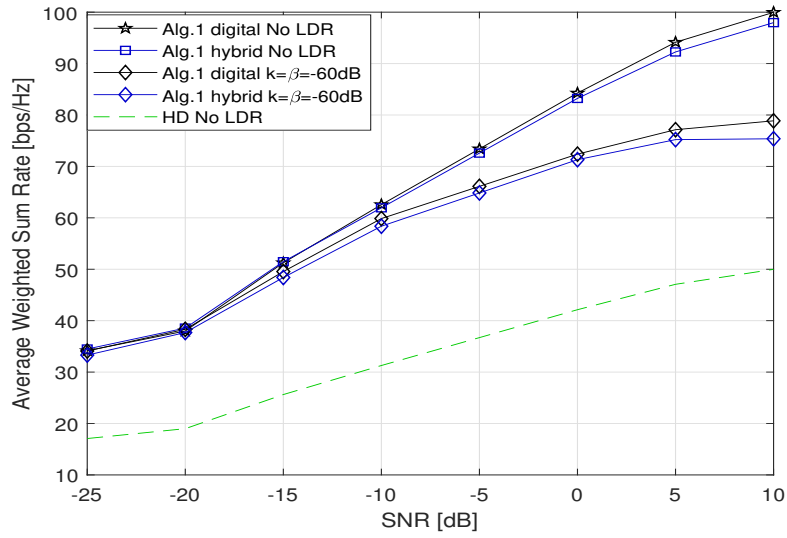


Figure 3.2: Average WSR as function of SNR with different levels of LDR noise.

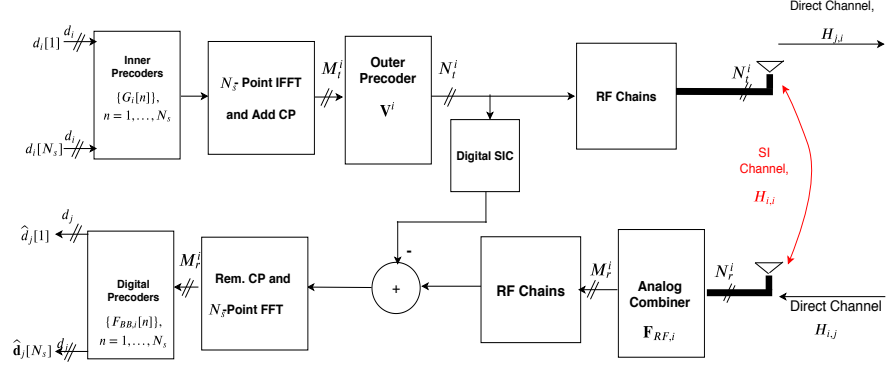


Figure 3.3: Bidirectional FD MIMO OFDM System with Multi-Stage/Hybrid BF. Only a single node is shown for simplicity in the figure.

achievable performance tends closer and closer to the HD system. Figure 3.2 show the achievable WSR as a function of SNR with $k = \beta = 0$ (in linear scale, no LDR noise case) and $k = \beta = -60$ dB. It can be seen that our proposed hybrid scheme performs closer to the fully digital one with 32 RF chains. Moreover, with $k = \beta = -60$ dB, the maximum achievable performance results to be very far from the achievable performance of an ideal FD system. Increasing transmission power also increases the LDR noise variance, which lead to enhanced LDR noise and less gain at high SNR.

3.2 Multi-Stage Hybrid Beamforming for Point-to-Point OFDM MIMO Full Duplex

This section consider the case of a point-to-point mmWave FD system. The literature on such a scenario is available in [52, 53, 54, 55, 56, 57], which is limited to a single carrier FD systems in the mmWave. We present a two-stage beamforming design for a bidirectional point-to-point OFDM MIMO FD system for WSR maximization. The problem is solved by adopting the weighted sum mean square error (WSMSE) approach. At the transmit side, we propose to use a two-stage beamformer at the baseband where the higher dimensional precoder is applied to the time domain signal, which aims to mitigate the SI. Then a lower-dimensional precoder is applied in the OFDM domain that aims to provide spatial multiplexing gain. On the receiver side, we introduce HYBF. The objective of the time domain phase shifter analog beamforming stage is to reduce the SI significantly before the ADCs while preserving the dimension of the desired signal space.

3.2.1 System Model

We consider a point-to-point scenario, consisting of two MIMO FD nodes i and j , communicating with each other. The number of transmit and receive antennas at the both nodes are denoted N_t^i and N_r^j , and N_r^i or N_t^j , respectively. The system is assumed to be an OFDM system with N_s subcarriers. Let $\mathbf{H}_{i,j}$, $i \neq j$ denote the MIMO channel from

node i to node j . Let $\mathbf{H}_{i,i}$ denote the SI channel for node i . Let $\mathbf{d}_j[n]$ of size $d_j \times 1$ denote the intended white and unit-variance data streams transmitted from node j for node i . At the transmit side, a two stage digital beamforming is assumed, composed of an inner beamformer $\mathbf{G}_j \in \mathbb{C}^{M_t^j \times d_j}$ of lower dimension, and an outer beamformer $\mathbf{V}_j \in \mathbb{C}^{N_t^j \times M_t^j}$ of higher dimension. The beamformer \mathbf{V}_j is applied in the time domain signal after the inverse DFT and is common to all the subcarriers, and the inner beamformers \mathbf{G}_j is different for different subcarriers. On the receive side, we assume that node i adopts hybrid combining approach and uses the analog combiner $\mathbf{F}_{RF,i}$ and the digital combiner $\mathbf{F}_{BB,i}$ to combine the received signal. For simplicity, only node i is shown in Fig. 3.3.

Let $\mathbf{y}_i[n]$ denote the signal received by node i after the analog combiner $\mathbf{F}_{RF,i}$, which can be written as

$$\begin{aligned} \mathbf{y}_i[n] = & \mathbf{F}_{RF,i} \mathbf{H}_{i,j}[n] (\mathbf{V}_j \mathbf{G}_j[n] \mathbf{d}_j[n] + \mathbf{c}_j[n]) + \mathbf{F}_{RF,i} \mathbf{H}_{i,i}[n] (\mathbf{V}^i \mathbf{G}_i[n] \mathbf{d}_i[n] + \mathbf{c}_i[n]) \\ & + \mathbf{e}_i[n] + \mathbf{F}_{RF,i} \mathbf{n}_i[n]. \end{aligned} \quad (3.38)$$

The vector $\mathbf{n}_i \sim \mathcal{CN}(0, \mathbf{I}_{N_r^i})$ denotes noise. The vectors \mathbf{c}_i and \mathbf{e}_i denote the transmit and receive LDR noise vectors and can be modelled as

$$\mathbf{c}_i[n] \sim \mathcal{CN}(\mathbf{0}, \frac{\alpha_i}{N_s} \text{diag}(\sum_{n=1}^{N_s} \mathbf{Q}_i[n])), \quad (3.39)$$

$$\mathbf{e}_i[n] \sim \mathcal{CN}(\mathbf{0}, \frac{\beta_i}{N_s} \text{diag}(\mathbf{Z})), \quad (3.40)$$

where $\beta_i \ll 1, \alpha_i \ll 1$ and $\mathbf{Q}_i[n]$ denotes the transmit signal covariance matrix at subcarrier n of node i and can be written as $\mathbf{Q}_i[n] = \mathbf{V}^i \mathbf{G}_i[n] \mathbf{G}_i^H[n] \mathbf{V}^{iH}$. The matrix \mathbf{Z} denotes the sum of the received covariance matrix of the undistorted received signal across all subcarriers [58] assuming the subcarrier signals are decorrelated, and it can be written as $\mathbf{Z} = \sum_{n=1}^{N_s} \mathbb{E}(\mathbf{z}_i[n] \mathbf{z}_i^H[n])$, where $\mathbf{z}_i[n] = \mathbf{y}_i[n] - \mathbf{e}_i[n]$.

We introduce a digital self interference canceller at the base band which subtracts the residual interference signal $\mathbf{H}_{i,i} \mathbf{x}_i$ from the received signal. Assuming that $\mathbf{H}_{i,i}$ is perfectly estimated at the baseband and since \mathbf{x}_i is already known to node i , we can rewrite the received signal at the baseband as,

$$\mathbf{y}_i'[n] = \mathbf{y}_i[n] - \mathbf{F}_{RF,i} \mathbf{H}_{i,i}[n] \mathbf{x}_i[n] = \mathbf{F}_{RF,i} \mathbf{H}_{i,j}[n] \mathbf{x}_j[n] + \mathbf{v}_i[n], \quad (3.41)$$

where $\mathbf{v}_i[n] = \mathbf{F}_{RF,i} \mathbf{H}_{i,j}[n] \mathbf{c}_j[n] + \mathbf{F}_{RF,i} \mathbf{H}_{i,i}[n] \mathbf{c}_i[n] + \mathbf{e}_i[n] + \mathbf{F}_{RF,i} \mathbf{n}_i[n]$ is the unknown interference plus noise component after SI cancellation. We assume that all the channel matrices and scaling factors in (3.38) are known. It is also worth mentioning that the dependence of the signal model (3.41) on the SI power is only through the LDR noise and the beamformers are designed in such a way to minimize it.

Channel Model

In this sub-section, we omit the node indices for simplicity. Considering a delay-d geometric direct channel model for a mmWave propagation environment [59] with L_s

scattering clusters and L_r scatterers or rays in each cluster, we have,

$$\mathbf{H} = \sum_{s=1}^{L_s} \sum_{l=1}^{L_r} \alpha_{s,l} \mathbf{h}_r(\theta_{s,l}) \mathbf{h}_t(\phi_{s,l})^H p(dT_s - \tau_s - \tau_{rl}) \quad (3.42)$$

The scalars $\theta_{s,l}, \phi_{s,l}$ denote the AoA and AoD, respectively, for the l^{th} path in the s^{th} cluster. The vectors $\mathbf{h}_r(\cdot)$ and $\mathbf{h}_t(\cdot)$ denote the antenna array responses at receiver and transmitter, respectively. The scalar $\alpha_{s,l} \sim \mathcal{CN}(0, \frac{N_t N_r}{L_s L_r})$ and $p(\tau)$ denote the complex path gain and the band-limited pulse shaping filter response evaluated at τ seconds, respectively. Each cluster has a time delay $\tau_s \in \mathcal{R}$ and each ray $l \in \{1, \dots, L_r\}$ has a relative time delay τ_{rl} . The (m, n) -th element of the channel response at the subcarrier n can be written as

$$\mathbf{H}[n] = \sum_{d=1}^D \mathbf{H}_d e^{-j2\pi \frac{nd}{N_s}}. \quad (3.43)$$

To represent it in a more compact form, we define the following variables,

$$\mathbf{H}_r = [\mathbf{h}_r(\theta_{1,1}), \dots, \mathbf{h}_r(\theta_{L_s, L_r})], \quad \mathbf{H}_t = [\mathbf{h}_t(\phi_{1,1}), \dots, \mathbf{h}_t(\phi_{L_s, L_r})], \quad (3.44)$$

$$\mathbf{A}_d[n] = \text{diag}(\alpha_{1,1} p(dT_s - \tau_1 - \tau_{r1}), \dots, \alpha_{L_s, L_r} p(dT_s - \tau_{L_s} - \tau_{rL_r})) e^{-j2\pi \frac{nd}{N_s}}, \quad (3.45)$$

which allows to rewrite the channel matrix (3.42) as

$$\mathbf{H}[n] = \mathbf{H}_r \sum_{d=1}^D \mathbf{A}_d[n] \mathbf{H}_t^H. \quad (3.46)$$

The (m, n) -th element of the SI channel can be modelled with

$$(\mathbf{H}_{i,i})_{m,n} = \frac{\rho}{r_{m,n}} e^{-j2\pi \frac{r_{m,n}}{\lambda}}, \quad (3.47)$$

where $r_{m,n}$ is the distance between m -th element of the receive array and n -th element of the transmit array, λ and ρ denote the wavelength and power normalization factor, respectively.

3.2.2 Optimization of the Beamformers

Consider the optimization problem of the two-stage BF with hybrid combining to maximize the WSR maximization as

$$\max_{\substack{\mathbf{V}, \mathbf{G}, \\ \mathbf{F}_{RF}, \mathbf{F}_{BB}}} WSR(\mathbf{G}, \mathbf{V}, \mathbf{F}_{RF}, \mathbf{F}_{BB}) \quad (3.48a)$$

$$\sum_{n=1}^{N_s} \text{Tr}\{ \mathbf{V}^j H \mathbf{V}^j \mathbf{G}_j[n] \mathbf{G}_j^H[n] \} \leq P_j \quad (3.48b)$$

where (3.48b) denote the total sum-power constraint for node j and the u_i denotes the rate weight. Note that the unit-modulus constraint on the analog combiner will

be incorporated later. The collection of beamformers at the two stage beamforming is denotes with \mathbf{G} and \mathbf{V} . For the receiver, the collection of digital and analog combiners is denote with \mathbf{F}_{BB} and \mathbf{F}_{RF} , respectively. The (signal-plus)interference-plus-noise covariance matrices ($\mathbf{R}_i[n]$) $\mathbf{R}_{\bar{i}}[n]$ for (3.48) are considered after the analog combiner, and have the following expressions

$$\mathbf{R}_{\bar{i}}[n] = \mathbf{F}_{RF,i}(\alpha_j \Phi_{i,j}[n] + \alpha_i \Phi_{i,i}[n])\mathbf{F}_{RF,i}^H + \beta_i \text{diag}(\mathbf{F}_{RF,i}(\Theta_{i,j}[n] + \Theta_{i,i}[n])\mathbf{F}_{RF,i}^H) \quad (3.49a)$$

$$\mathbf{R}_i[n] = \mathbf{R}_{\bar{i}}[n] + \mathbf{F}_{RF,i}\Theta_{i,j}[n]\mathbf{F}_{RF,i}^H. \quad (3.49b)$$

The matrices $\Theta_{i,j}[n]$ and $\Phi_{i,j}[n]$ appearing above are defined as

$$\Theta_{i,j}[n] = \mathbf{H}_{i,j}[n]\mathbf{Q}_j[n]\mathbf{H}_{i,j}^H[n], \quad \Phi_{i,j}[n] = \mathbf{H}_{i,j}[n]\text{diag}(\mathbf{Q}_j[n])\mathbf{H}_{i,j}^H[n]. \quad (3.50)$$

After the digital receive combining, we have $\Sigma_{\bar{i}}[n] = \mathbf{F}_{BB,i}[n]\mathbf{R}_{\bar{i}}[n]\mathbf{F}_{BB,i}^H[n]$ and $\Sigma_i[n] = \mathbf{F}_{BB,i}[n]\mathbf{R}_i[n]\mathbf{F}_{BB,i}^H[n]$. Direct maximization of (3.48a), however, requires a joint optimization over the four matrix variables and finding a global optimum solution for similar constrained optimization is very challenging. So we decouple the joint transmitter-receiver optimization and first focus on the design of combiners. We assume that the node i applies the hybrid combiner $\mathbf{F}_i[n] = \mathbf{F}_{BB,i}[n]\mathbf{F}_{RF,i}$ to estimate the signal transmitted from node j . The analog combiner $\mathbf{F}_{RF,i}$ aims to reduce the SI while the digital combiner $\mathbf{F}_{BB,i}$ decouples the streams \mathbf{d}_j intended for user i from j . The estimated signal $\hat{\mathbf{d}}_j[n]$ can be written as

$$\hat{\mathbf{d}}_j[n] = \mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{x}_j[n] + \mathbf{F}_{BB,i}[n]\mathbf{v}_i[n]. \quad (3.51)$$

At the receiver side, maximizing the WSR is equivalent to minimizing the weighted MSE with the MSE weights being chosen as $\mathbf{W}_i[n] = u_i \mathbf{R}_{\hat{\mathbf{d}}_j \hat{\mathbf{d}}_j}[n]^{-1}$ [60, 61]. The error covariance matrix for the detection of \mathbf{d}_j at node i can be written as

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{d}}_j \hat{\mathbf{d}}_j}[n] &= \mathbb{E}\{(\hat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])(\hat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])^H\} \\ &= (\mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}_j\mathbf{G}_j[n] - \mathbf{I})(\mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}_j\mathbf{G}_j[n] - \mathbf{I})^H + \mathbf{F}_{BB,i}\mathbf{R}_{\bar{i}}[n]\mathbf{F}_{BB,i}^H[n]. \end{aligned} \quad (3.52)$$

Fixed the beamformers, the combiners for node i can be alternatively optimized by solving the following optimization problem

$$[\mathbf{F}_{RF,i}, \mathbf{F}_{BB,i}[n], \forall n] = \min_{\mathbf{F}_{RF,i}, \mathbf{F}_{BB,i}[n]} \sum_{n=1}^{N_s} \text{Tr}\{\mathbf{R}_{\hat{\mathbf{d}}_j \hat{\mathbf{d}}_j}[n]\}, \quad (3.53)$$

and solving it leads to the following digital MMSE combiner

$$\mathbf{F}_{BB,i}[n] = \mathbf{G}_j^H[n]\mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H\mathbf{R}_i[n]^{-1}. \quad (3.54)$$

We remark that the optimization of the digital beamformers in (3.53) can be done independently across different subcarriers. To optimize the analog combiner, we define the following variables

$$\mathbf{F}_{BB,i}^H[n]\mathbf{F}_{BB,i}[n] = \mathbf{P}_{B,i}[n] \quad (3.55a)$$

$$\begin{aligned} \mathbf{B}_i = & \sum_{n=1}^{N_s} [(\boldsymbol{\Theta}_{i,j}[n])^T \otimes \mathbf{P}_{B,i}[n] + ((\alpha_j \boldsymbol{\Phi}_{i,j}[n] + \alpha_i \boldsymbol{\Phi}_{i,i}[n])^T \otimes \mathbf{P}_{B,i}[n]) \\ & + (\beta_i (\boldsymbol{\Theta}_{i,j}[n] + \boldsymbol{\Theta}_{i,i}[n])^T \otimes \text{diag}(\mathbf{P}_{B,i}[n]))]. \end{aligned} \quad (3.55b)$$

To derive the unconstrained analog combining matrix, we take the gradient of (3.52) with respect to $\mathbf{F}_{RF,i}$, which yield the following KKT condition

$$\begin{aligned} & \sum_{n=1}^{N_s} \mathbf{P}_{B,i}[n] \mathbf{F}_{RF,i} \boldsymbol{\Theta}_{i,j}[n] - \mathbf{F}_{BB,i}^H[n] \mathbf{G}_j^H[n] \mathbf{V}_j^H \mathbf{H}_{i,j}^H[n] + \mathbf{P}_{B,i}[n] \mathbf{F}_{RF,i} (\alpha_j \boldsymbol{\Phi}_{i,j}[n] + \alpha_i \boldsymbol{\Phi}_{i,i}[n]) \\ & + \beta_i \text{diag}(\mathbf{P}_{B,i}[n]) \mathbf{F}_{RF,i} (\boldsymbol{\Theta}_{i,j}[n] + \boldsymbol{\Theta}_{i,i}[n]) = 0, \\ \text{where } \mathbf{B}_i \text{vec}(\mathbf{F}_{RF,i}) & \stackrel{(a)}{=} \text{vec}(\sum_{n=1}^{N_s} \mathbf{F}_{BB,i}^H[n] \mathbf{G}_j^H[n] \mathbf{V}_j^H \mathbf{H}_{i,j}^H[n]). \end{aligned} \quad (3.56)$$

In (a), we use the result $\text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$ [62]. Given the KKT condition for the analog combiner, it is immediate to conclude that it can be optimized as

$$\text{vec}(\mathbf{F}_{RF,i}) = \mathbf{B}_i^\dagger \text{vec}(\sum_{n=1}^{N_s} \mathbf{F}_{BB,i}^H[n] \mathbf{G}_j^H[n] \mathbf{V}_j^H \mathbf{H}_{i,j}^H[n]), \quad (3.57)$$

where \dagger represents the pseudoinverse. To reshape the analog combiner into correct dimensions, we first apply the operator $\text{unvec}(\cdot)$ and normalize the amplitudes of the analog combiner to unit-modulus as $\mathbf{F}_{RF,i} = \angle \mathbf{F}_{RF,i}$.

3.2.3 Two stage Transmit Beamforming

In this section, we consider the design of the two stage transmit beamformers \mathbf{V}_j and $\mathbf{G}_j[n]$ under the sum power constraint. To facilitate the gradients, we use the result $\frac{\partial \text{Tr}\{\mathbf{A} \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D})\mathbf{B}\}}{\partial \mathbf{X}} = [\mathbf{D} \text{diag}(\mathbf{B}\mathbf{A})\mathbf{C}]^T$, which is straightforward by using the matrix differentiation properties. The beamformers as designed using the weighted sum MSE criteria. For such purpose, the optimization problem can be formulated as

$$\begin{aligned} \min_{\substack{\mathbf{V}_j, \mathbf{G}_j[n] \\ \mathbf{V}_j, \mathbf{G}_j[n]}} & \sum_{n=1}^{N_s} \text{Tr}(\mathbf{W}_i[n] \mathbb{E}(\hat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])(\hat{\mathbf{d}}_j[n] - \mathbf{d}_j[n])^H) \\ & + \text{Tr}(\mathbf{W}_j[n] \mathbb{E}\{(\hat{\mathbf{d}}_i[n] - \mathbf{d}_i[n])(\hat{\mathbf{d}}_i[n] - \mathbf{d}_i[n])^H\}) \end{aligned} \quad (3.58a)$$

$$\text{s.t.} \quad \sum_{n=1}^{N_s} \text{Tr}\{\mathbf{Q}_i[n]\} \leq P_i, \quad \forall i. \quad (3.58b)$$

The matrix $\mathbf{W}_i[n]$ denotes the weight matrix of size $d_i \times d_i$. Augmenting the cost function with the sum-power constraints yield the following Lagrangian function

$$\begin{aligned} \mathcal{L} = & \sum_{n=1}^{N_s} \sum_{i=1}^2 \text{Tr}(\mathbf{W}_i[n](\mathbf{I} - \mathbf{G}_j^H[n]\mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n] - \mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}_j\mathbf{G}_j[n] \\ & + \mathbf{F}_i[n]\mathbf{H}_{i,j}[n]\mathbf{Q}_j\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n] + \mathbf{F}_{BB,i}[n]\mathbf{R}_{\bar{i}}[n]\mathbf{F}_{BB,i}^H[n])) \\ & + (\sum_{i=1}^2 \lambda_i (\sum_{n=1}^{N_s} \text{Tr}\{\mathbf{Q}_i[n]\}) - P_i). \end{aligned} \quad (3.59)$$

Let $\mathbf{A}_j[n]$ and $\hat{\mathbf{A}}_j[n]$ be defined as

$$\mathbf{A}_j[n] = \mathbf{F}_j^H[n]\mathbf{W}_j[n]\mathbf{F}_j[n], \quad \hat{\mathbf{A}}_j[n] = \mathbf{F}_{BB,j}^H[n]\mathbf{W}_j[n]\mathbf{F}_{BB,j}[n]. \quad (3.60)$$

Taking the partial derivative of (3.59) with respect to the inner beamformer $\mathbf{G}_j[n]$, we get

$$\begin{aligned} & \mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}_j\mathbf{G}_j[n] - \mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n] + \lambda_j\mathbf{V}_j^H\mathbf{V}_j\mathbf{G}_j[n] \\ & + \frac{\text{Tr}(\mathbf{F}_{BB,i}[n]^H\mathbf{F}_{BB,i}[n]\partial\mathbf{R}_{\bar{i}}[n])}{\partial\mathbf{G}_j[n]} + \frac{\text{Tr}(\mathbf{F}_{BB,j}[n]^H\mathbf{F}_{BB,j}[n]\partial\mathbf{R}_{\bar{j}}[n])}{\partial\mathbf{G}_j[n]} = \mathbf{0}. \end{aligned} \quad (3.61)$$

Using the expression for $\mathbf{R}_{\bar{i}}[n]$ in (3.49), we can write

$$\begin{aligned} \frac{\text{Tr}\{\mathbf{F}_{BB,i}[n]^H\mathbf{F}_{BB,i}[n]\partial\mathbf{R}_{\bar{i}}[n]\}}{\partial\mathbf{G}_j[n]} &= \alpha_j\mathbf{V}_j^H \text{diag}(\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n])\mathbf{V}_j\mathbf{G}_j[n] \\ &+ \beta_i\mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H \text{diag}(\hat{\mathbf{A}}_i[n])\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{V}_j\mathbf{G}_j[n], \end{aligned} \quad (3.62)$$

$$\begin{aligned} \frac{\text{Tr}\{\mathbf{F}_{BB,j}[n]^H\mathbf{F}_{BB,j}[n]\partial\mathbf{R}_{\bar{j}}[n]\}}{\partial\mathbf{G}_j[n]} &= \alpha_j\mathbf{V}_j^H \text{diag}(\mathbf{H}_{j,j}^H[n]\mathbf{A}_j[n]\mathbf{H}_{j,j}[n])\mathbf{V}_j\mathbf{G}_j[n] \\ &+ \beta_j\mathbf{V}_j^H\mathbf{H}_{j,j}^H[n]\mathbf{F}_{RF,i}^H \text{diag}(\hat{\mathbf{A}}_j[n])\mathbf{F}_{RF,i}\mathbf{H}_{j,j}[n]\mathbf{V}_j\mathbf{G}_j[n], \end{aligned} \quad (3.63)$$

By substituting (3.62)-(3.63) in (3.61), we obtain the the following expression to $\mathbf{G}_j[n]$

$$\mathbf{G}_j[n] = (\mathbf{S}_j[n] + \lambda_j\mathbf{V}_j^H\mathbf{V}_j)^{-1}\mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{F}_i^H[n]\mathbf{W}_i[n]. \quad (3.64)$$

The matrix $\mathbf{S}_j[n]$ can be interpreted as the signal plus interference power seen by the digital beamformer at the transmit side and is expressed as

$$\begin{aligned} \mathbf{S}_j[n] &= \mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n]\mathbf{V}_j + \alpha_j\mathbf{V}_j^H \text{diag}(\mathbf{H}_{i,j}^H[n]\mathbf{A}_i[n]\mathbf{H}_{i,j}[n])\mathbf{V}_j \\ &+ \beta_i\mathbf{V}_j^H\mathbf{H}_{i,j}^H[n]\mathbf{F}_{RF,i}^H \text{diag}(\hat{\mathbf{A}}_i[n])\mathbf{F}_{RF,i}\mathbf{H}_{i,j}[n]\mathbf{V}_j + \alpha_j\mathbf{V}_j^H \text{diag}(\mathbf{H}_{j,j}^H[n]\mathbf{A}_j[n] \\ &\mathbf{H}_{j,j}[n])\mathbf{V}_j + \beta_j\mathbf{V}_j^H\mathbf{H}_{j,j}^H[n]\mathbf{F}_{RF,j}^H \text{diag}(\hat{\mathbf{A}}_j[n])\mathbf{F}_{RF,j}\mathbf{H}_{j,j}[n]\mathbf{V}_j. \end{aligned} \quad (3.65)$$

The values of the Lagrangian multipliers $\lambda_j \geq 0, \forall j$ should be such that the power constraint is met. To compute the powers, we follow a similar approach as in

[63], but extended to two-stage beamformer case applied to FD. Considering the eigen decomposition of the matrices $\mathbf{S}_j[n]$ and $\mathbf{V}_j^H \mathbf{V}_j$ as

$$\mathbf{S}_j[n] = \mathbf{U}_j \mathbf{\Lambda}_j[n] \mathbf{U}_j^H, \quad \mathbf{V}_j^H \mathbf{V}_j = \mathbf{U}_j \mathbf{\Delta}_j \mathbf{U}_j^H. \quad (3.66)$$

Let $\Phi[n] = \mathbf{U}_j^H \mathbf{V}_j^H \mathbf{H}_{i,j}^H[n] \mathbf{F}_i^H[n] \mathbf{W}_i[n] \mathbf{W}_i[n]^H \mathbf{F}_i[n] \mathbf{H}_{i,j}[n] \mathbf{V}_j \mathbf{U}_j$, expanding the power constraint $\sum_{n=1}^{N_s} \text{Tr}\{\mathbf{V}_j \mathbf{G}_j[n](\lambda_j) \mathbf{G}_j^H[n](\lambda_j) \mathbf{V}_j^H\} = P_j$, and solving it for the stream powers we get

$$\sum_{n=1}^{N_s} \sum_{k=1}^{M_t^j} \frac{\Phi[n]_{k,k}(\mathbf{\Delta}_j)_{k,k}}{((\mathbf{\Lambda}_j[n])_{k,k} + \lambda_j(\mathbf{\Delta}_j)_{k,k})^2} = P_j. \quad (3.67)$$

The matrix $\mathbf{X}_{k,k}$ denotes the k -th diagonal element of the matrix \mathbf{X} . Note that $\lambda_j \geq 0$ and the left hand side of (3.67) is a decreasing function of λ_j for $\lambda_j > 0$. Hence we can compute the values of λ_j using one dimensional linear search techniques such as Bisection. In the following, we consider the optimization of the outer beamformer \mathbf{V}_j , assuming the other variables are fixed. Taking the partial derivative of (3.59) with respect to the inner BF \mathbf{V}_j , yields the following KKT condition

$$\begin{aligned} & \mathbf{H}_{i,j}^H[n] \mathbf{F}_i^H[n] \mathbf{W}_i[n] \mathbf{F}_i[n] \mathbf{H}_{i,j}[n] \mathbf{V}_j \mathbf{G}_j[n] \mathbf{G}_j^H[n] - \mathbf{H}_{i,j}^H[n] \mathbf{F}_i^H[n] \mathbf{W}_i[n] \mathbf{G}_j^H[n] \\ & + \frac{\text{Tr}\{\mathbf{F}_{BB,i}[n]^H \mathbf{F}_{BB,i}[n] \partial \mathbf{R}_{\bar{i}}[n]\}}{\partial \mathbf{V}_j[n]} + \frac{\text{Tr}\{\mathbf{F}_{BB,j}[n]^H \mathbf{F}_{BB,j}[n] \partial \mathbf{R}_{\bar{j}}[n]\}}{\partial \mathbf{V}_j[n]} + \\ & \lambda_j \mathbf{V}_j \mathbf{G}_j[n] \mathbf{G}_j^H[n] = \mathbf{0}. \end{aligned} \quad (3.68)$$

For notational convenience, we define $\mathbf{P}_{G,j}[n] = \mathbf{G}_j[n] \mathbf{G}_j^H[n]$. Using the expression for $\mathbf{R}_{\bar{i}}[n]$ in (3.49), we can write,

$$\begin{aligned} \frac{\text{Tr}\{\mathbf{F}_{BB,i}[n]^H \mathbf{F}_{BB,i}[n] \partial \mathbf{R}_{\bar{i}}[n]\}}{\partial \mathbf{V}_j[n]} &= \alpha_j \text{diag}(\mathbf{H}_{i,j}^H[n] \mathbf{A}_i[n] \mathbf{H}_{i,j}[n]) \mathbf{V}_j \mathbf{P}_{G,j}[n] \\ &+ \beta_i \mathbf{H}_{i,j}^H[n] \mathbf{F}_{RF,i}^H \text{diag}(\hat{\mathbf{A}}_i[n]) \mathbf{F}_{RF,i} \mathbf{H}_{i,j}[n] \mathbf{V}_j \mathbf{P}_{G,j}[n], \end{aligned} \quad (3.69)$$

$$\begin{aligned} \frac{\text{Tr}\{\mathbf{F}_{BB,j}[n]^H \mathbf{F}_{BB,j}[n] \partial \mathbf{R}_{\bar{j}}[n]\}}{\partial \mathbf{V}_j[n]} &= \alpha_j \text{diag}(\mathbf{H}_{j,j}^H[n] \mathbf{A}_j[n] \mathbf{H}_{j,j}[n]) \mathbf{V}_j \mathbf{P}_{G,j}[n] \\ &+ \beta_j \mathbf{H}_{j,j}^H[n] \mathbf{F}_{RF,j}^H \text{diag}(\hat{\mathbf{A}}_j[n]) \mathbf{F}_{RF,j} \mathbf{H}_{j,j}[n] \mathbf{V}_j \mathbf{P}_{G,j}[n]. \end{aligned} \quad (3.70)$$

By substituting (3.69)-(3.70) in (3.68) and using the result $\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$, we see that WSR maximizing beamformer \mathbf{V}_j can be computed as

$$\text{vec}(\mathbf{V}_j) = \mathbf{B}_j^H \sum_{n=1}^{N_s} \mathbf{H}_{i,j}^H[n] \mathbf{F}_i^H[n] \mathbf{W}_i[n] \mathbf{G}_j^H[n], \text{ where} \quad (3.71)$$

$$\begin{aligned}
\mathbf{B}_j = & \sum_{n=1}^{N_s} (\mathbf{P}_{G,j}[n] \otimes \mathbf{H}_{i,j}^H[n] \mathbf{A}_i[n] \mathbf{H}_{i,j}[n]) + \alpha_j \mathbf{P}_{G,j}[n] \otimes \text{diag}(\mathbf{H}_{i,j}^H[n] \mathbf{A}_i[n] \mathbf{H}_{i,j}[n]) \\
& + \beta_i \mathbf{P}_{G,j}[n] \otimes \mathbf{H}_{i,j}^H[n] \mathbf{F}_{RF,i}^H \text{diag}(\hat{\mathbf{A}}_i[n]) \mathbf{F}_{RF,i} \mathbf{H}_{i,j}[n] + \alpha_j \mathbf{P}_{G,j}[n] \otimes \text{diag}(\mathbf{H}_{j,j}^H \mathbf{A}_j \\
& \mathbf{H}_{j,j}[n]) + \beta_j (\mathbf{P}_{G,j}[n] \otimes (\mathbf{H}_{j,j}^H[n] \mathbf{F}_{RF,j}^H \text{diag}(\hat{\mathbf{A}}_j[n]) \mathbf{F}_{RF,j} \mathbf{H}_{j,j}[n])).
\end{aligned} \tag{3.72}$$

Alternating optimization based WSR maximizing procedure to maximize the WSR is formally stated in Algorithm 9. It's convergence is straightforward, as each variable is computed to minimize the MSE, which hence lead to a monotonic increase in the WSR.

Algorithm 6 Multi Stage Beamforming Design

Given: $P_i, \mathbf{H}_{i,j}, \mathbf{H}_{i,i}, u_i \forall i, j$.

Initialization: $\mathbf{F}_{RF,i} = e^{j\angle \mathbf{V}_{M^i}(\mathbf{H}_{t,i,j})}$, The \mathbf{G}_i are taken as the ZF precoders for the effective channels $\mathbf{V}^i \mathbf{H}_{j,i}$ with uniform powers.

Iteration (t) :

1. Update the Rx side HYBF, i.e $\mathbf{F}_{BB,i}^{(t)}, \mathbf{F}_{RF,i}^{(t)} \forall i$ using (3.53), (3.57) respectively.
 2. Set $\mathbf{F}_{RF,i} = \angle \mathbf{F}_{RF,i}$.
 3. Update $\mathbf{G}_i^{(t)}[n], \forall i$, from (3.64).
 4. Update $\mathbf{V}^i(t), \forall i$ from (3.71) and λ_i using Bisection method from (3.67).
 5. Check for convergence of the WSR: if not go to step 1).
-

It might be possible that only two-stage beamforming or only hybrid combining might be enough if the transmit or receive antenna dimensions are sufficiently large. However, to conclude that, we present a detailed analysis in the following.

3.2.4 Simulation Results

In this section, we present simulations to evaluate the performance of our proposed HYBF scheme for a bidirectional FD system under LDR noise model. The far-field channels are modelled with the path-wise channel model and their complex path gains for the far-field channels are assumed to be Gaussian with variance distributed according to an exponential profile. For the SI channel, we ignore the near-field effect of amplitude variation with distance and the near field effects in the phase variation. In the Uniform Linear Array (ULA), the AoD or AoA ϕ, θ are assumed to be uniformly distributed in the interval $[0^\circ, 30^\circ]$. For comparison, we consider a fully digital beamforming design. The dimensions of the system simulated to report the results are mentioned below each figure. We consider comparing our design either by using only two stage transmit beamforming at the transmit side or only hybrid combining at the receive side. As using both the two-stage beamformers with hybrid combining could require high hardware cost, we consider comparing them one versus the other to investigate which approach is much more prominent and can be deployed cost-efficiently.

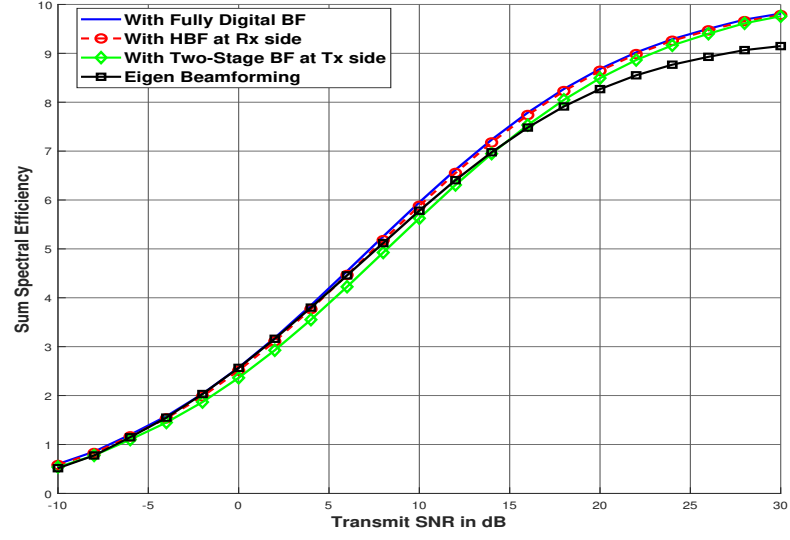


Figure 3.4: Sum Rate comparisons for, Single Carrier, $N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 4$ paths.

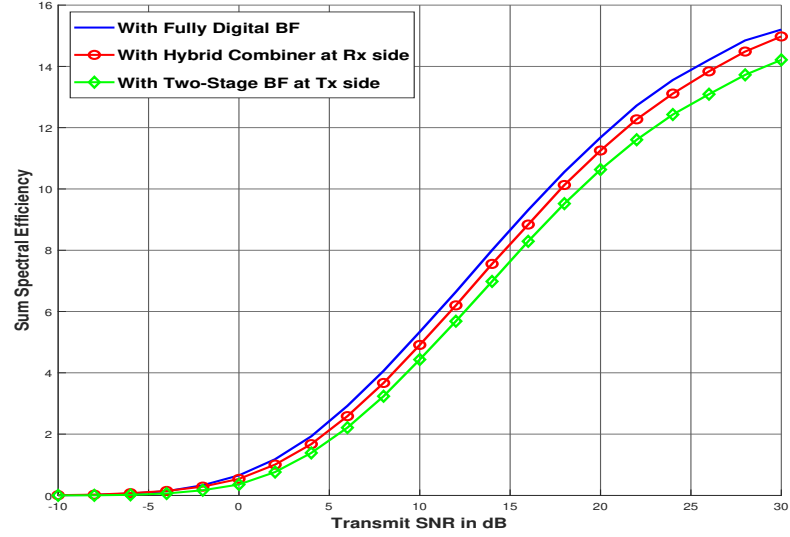


Figure 3.5: Sum Rate comparisons for, OFDM, $N_s = 4, N_t^i = N_r^i = 8, M_t^i = M_r^i = 4, d_i = 1, \forall i, L = 4$ paths.

Figure 3.4-3.5 show the performance of the proposed design either with two-stage beamformers or analog combiner as a function of SNR. We can see that at high SNR, the hybrid combiner achieves better performance than the two-stage beamformers at high-SNR. Figure 3.4, also shows that the eigen beamforming solution, where the left and right singular vectors of the corresponding channels are used as the combiners or beamformers. We can see it perform worse than the proposed scheme.

The dimensions of the two-stage beamforming and hybrid combining are such that the zero forcing capabilities at both sides are comparable. However, the number of LDR noise contributions is equal to the number of antennas at the transmit side, whereas for the analog receive stage, the number of LDR noises is the number of analog combining outputs, which is less. We conjecture that the analog combining reduces the LDR noise to a significant level and this would explain the better performance of the analog stage at receive side (in both figures) compared to the two-stage architecture at transmit side. In Figure 3.4, we compare also compare the eigen beamforming (where the left and right singular vectors of the corresponding channels are used as the Combiner/BF and fully digital) and shows that its performance is inferior compared to our proposed design. Using both the two-stage and hybrid combining can lead to better performance, however it might not be desirable due to hardware-cost.

3.2.5 Conclusion

This section has proposed a two-stage beamforming approach for mmWave MIMO point-to-point OFDM FD system. The design is presented by decoupling the transmission and reception side and an iterative procedure is presented for WSR maximization by minimizing the MSE. Simulation results show that our design achieve significant performance, in terms for WSR, as the WSR is comparable with the WSR of a fully digital solution. Results motivate that using only hybrid analog combiner achieves better performance than using a two-stage beamformer at the transmit side of the OFDM MIMO FD system in mmWave.

3.3 Generalization of the point-to-point mMIMO Full-Duplex Communication in Millimeter Wave

This section considers the general case of a point-to-point mmWave FD system for a single-carrier point-to-point mmWave FD system. The literature on the point-to-point mmWave FD systems is available in [52, 53, 54, 55, 56, 57], which is limited to a single point-to-point FD link. We generalize the point-to-point FD communication to the case of K -pair mMIMO nodes in mmWave. All the nodes are assumed to be equipped with a massive number of antennas and only a limited number of RF chains. The coexistence of multiple point-to-point FD nodes in mmWave leads to an FD mMIMO interference channel in which each node suffers from SI and interference from all the other nodes transmitting simultaneously. To enable FD communication in such a challenging scenario, we present a novel HYBF and combining design based on MM [26] for WSR maximization. Moreover, we present an optimal power allocation scheme for FD mMIMO nodes, which

are both SI and mMIMO interference channel aware. Simulation results show that the proposed HYBF and combining design exhibit significant performance improvement over a fully digital mMIMO HD communication system with only a limited number of RF chains.

3.3.1 System Model

We consider a setup of K -pair mmWave mMIMO FD point-to-point communication links as shown in Figure 3.6. We distinguish the total nodes into two sets, left and right,

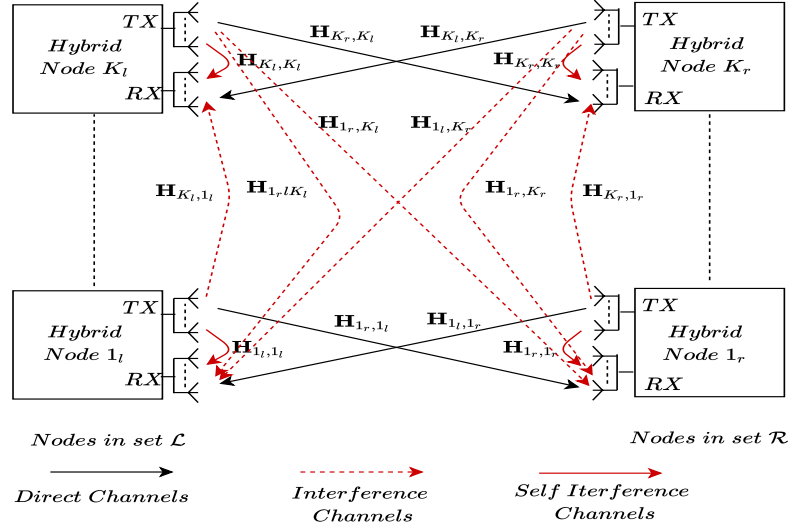


Figure 3.6: mMIMO interference channel in the mmWave consisting of Hybrid nodes with separate transmit (TX) and receive (RX) array.

denoted with $\mathcal{L} = \{1_l, \dots, K_l\}$. and $\mathcal{R} = \{1_r, \dots, K_r\}$, respectively. Nodes involved in the communication link $1 \leq i \leq K$ are denoted with i_a , in which the subscript $a \in \mathcal{L}$ or $\in \mathcal{R}$. We consider a multi-stream approach and let $\mathbf{s}_{i_l} \in \mathbb{C}^{d_{i_l} \times 1}$ denote the white and unitary variance data-streams transmitted from node $i_l \in \mathcal{L}$ intended for node $i_r \in \mathcal{R}$. Let $\mathbf{V}_{i_l} \in \mathbb{C}^{M_{i_l}^t \times d_{i_l}}$ and $\mathbf{G}_{i_l} \in \mathbb{C}^{N_{i_l}^t \times M_{i_l}^t}$ denote the digital and the analog beamformer at node $i_l \in \mathcal{L}$, respectively. Let $\mathbf{F}_{i_l} \in \mathbb{C}^{M_{i_l}^r \times N_{i_l}^r}$ denote the analog combiner at node $i_l \in \mathcal{L}$ for the data streams \mathbf{s}_{i_r} transmitted from node $i_r \in \mathcal{R}$. Let $M_{i_l}^t$ and $M_{i_l}^r$ denote the number of transmit and receive RF chains for node $i_l \in \mathcal{L}$, respectively. Let $N_{i_l}^t$ and $N_{i_l}^r$ denote the total number of transmit and receive antennas for node $i_l \in \mathcal{L}$, respectively. The signal received at node $i_l \in \mathcal{L}$ for the case of mMIMO FD interference channel, after the analog combiner, can be written as

$$\begin{aligned} \mathbf{y}_{i_l} = & \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r} \mathbf{s}_{i_r} + \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l} \mathbf{s}_{i_l} + \mathbf{F}_{i_l} \mathbf{n}_{i_l} + \\ & \sum_{\substack{m_l \in \mathcal{L} \\ m_l \neq i_l}} \mathbf{F}_{i_l} \mathbf{H}_{i_l, m_l} \mathbf{G}_{m_l} \mathbf{V}_{m_l} \mathbf{s}_{m_l} + \sum_{\substack{m_r \in \mathcal{R} \\ m_r \neq i_r}} \mathbf{F}_{i_l} \mathbf{H}_{i_l, m_r} \mathbf{G}_{m_r} \mathbf{V}_{m_r} \mathbf{s}_{m_r}, \end{aligned} \quad (3.73)$$

where $\mathbf{n}_{i_l} \sim \mathcal{CN}(0, \sigma_{i_l}^2 \mathbf{I})$ denote the noise vector at node i , $\mathbf{H}_{i_l, i_r} \in \mathbb{C}^{N_{i_l}^r \times N_{i_r}^t}$, and $\mathbf{H}_{i_l, i_l} \in \mathbb{C}^{N_{i_l}^r \times N_{i_l}^t}$ denote the channel between the node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, and the SI channel for node i_l , respectively. The matrices $\mathbf{H}_{i_l, m_l} \in \mathbb{C}^{N_{i_l}^r \times N_{m_l}^t}$ and $\mathbf{H}_{i_l, m_r} \in \mathbb{C}^{N_{i_l}^r \times N_{m_r}^t}$ denote the interference channels for node i_l from node $m_l \in \mathcal{L}, m_l \neq i_l$ and from node $m_r \in \mathcal{R}, m_r \neq i_r$, respectively. Let $\mathbf{T}_{i_l, i_r} = \mathbf{G}_{i_r} \mathbf{V}_{i_r} \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H$ denote the transmit covariance matrix from node $i_r \in \mathcal{R}$ intended for node $i_l \in \mathcal{L}$. Let $(\mathbf{R}_{i_l}) \mathbf{R}_{i_l}^-$ denote the (signal plus) interference and noise covariance matrix received at node $i_l \in \mathcal{L}$. The covariance matrix \mathbf{R}_{i_l} can be written as

$$\begin{aligned} \mathbf{R}_{i_l} = & \underbrace{\mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{T}_{i_l, i_r} \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H}_{\triangleq \mathbf{S}_{i_l}} + \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_l} \mathbf{T}_{i_r, i_l} \mathbf{H}_{i_l, i_l}^H \mathbf{F}_{i_l}^H + \sigma_{i_l}^2 \mathbf{F}_{i_l} \mathbf{F}_{i_l}^H + \sum_{m_l \in \mathcal{L}, m_l \neq i_l} \mathbf{F}_{i_l} \mathbf{H}_{i_l, m_l} \\ & \mathbf{T}_{m_r, m_l} \mathbf{H}_{i_l, m_l}^H \mathbf{F}_{i_l}^H + \sum_{m_r \in \mathcal{R}, m_r \neq i_r} \mathbf{F}_{i_l} \mathbf{H}_{i_l, m_r} \mathbf{T}_{m_l, m_r} \mathbf{H}_{i_l, m_r}^H \mathbf{F}_{i_l}^H, \end{aligned} \quad (3.74)$$

and $\mathbf{R}_{i_l}^-$ can be written as $\mathbf{R}_{i_l}^- = \mathbf{R}_{i_l} - \mathbf{S}_{i_l}$, where \mathbf{S}_{i_l} denotes the useful signal part for node $i_l \in \mathcal{L}$. In the mmWave, the channel \mathbf{H}_{i_l, i_r} can be modelled as

$$\mathbf{H}_{i_l, i_r} = \sqrt{\frac{N_{i_l}^r N_{i_r}^t}{N_c N_p}} \sum_{n_c=1}^{N_c} \sum_{n_p=1}^{N_p} \alpha_{i_l, i_r}^{n_p, n_c} \mathbf{a}_{i_l}(\phi_{i_l}^{n_p, n_c}) \mathbf{a}_{i_r}^T(\theta_{i_r}^{n_p, n_c}), \quad (3.75)$$

where N_c and N_p denote the number of clusters and number of rays, respectively, $\alpha_{i_l, i_r}^{n_p, n_c} \sim \mathcal{CN}(0, 1)$ is a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively, and $\mathbf{a}_{i_l}(\phi_{i_l}^{n_p, n_c})$ and $\mathbf{a}_{i_r}^T(\theta_{i_r}^{n_p, n_c})$ denote the receive and transmit antenna array response with angle of arrival (AoA) $\phi_{i_l}^{n_p, n_c}$ and angle of departure (AoD) $\theta_{i_r}^{n_p, n_c}$, respectively. Also, the channel matrices \mathbf{H}_{i_l, m_r} and \mathbf{H}_{i_l, m_l} can be modelled as (3.75). The SI channel can be modelled with the Rician fading channel model given as

$$\mathbf{H}_{i_l, i_l} = \sqrt{\frac{\kappa_{i_l}}{\kappa_{i_l} + 1}} \mathbf{H}_{i_l}^{LoS} + \sqrt{\frac{1}{\kappa_{i_l} + 1}} \mathbf{H}_{i_l}^{Ref}, \quad (3.76)$$

where κ_{i_l} , $\mathbf{H}_{i_l}^{LoS}$ and $\mathbf{H}_{i_l}^{Ref}$ denote the Rician factor, the line-of-sight (LoS) component and reflected components matrices, respectively, at node $i_l \in \mathcal{L}$. The channel matrix $\mathbf{H}_{i_l}^{Ref}$ can be modelled as in (3.75). The elements of matrix $\mathbf{H}_{i_l}^{LoS}$ at m -th row and n -th column can be modelled as

$$\mathbf{H}_{i_l}^{LoS}(m, n) = \frac{\rho}{r_{m, n}} e^{-j2\pi \frac{r_{m, n}}{\lambda}} \quad (3.77)$$

where ρ is the power normalization constant which assure that $\mathbb{E}(\|\mathbf{H}_{i_l}^{LoS}(m, n)\|_F^2) = N_{i_l}^r N_{i_l}^t$ and the scalar $r_{m, n}$ depends on the antenna array geometry. The WSR maximization problem for a mMIMO interference channel under the total sum-power constraint with unit-modulus phase shifters over the mmWave FD mMIMO interference channel

can be stated as

$$\max_{\mathbf{G}, \mathbf{V}, \mathbf{F}} \sum_{i_l \in \mathcal{L}} w_{i_l} \ln \det(\mathbf{R}_{i_l}^{-1} \mathbf{R}_{i_l}) + \sum_{i_r \in \mathcal{R}} w_{i_r} \ln \det(\mathbf{R}_{i_r}^{-1} \mathbf{R}_{i_r}). \quad (3.78a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{G}_a \mathbf{V}_a \mathbf{V}_a^H \mathbf{G}_a^H) \preceq p_a, \quad \forall a \in \mathcal{L} \text{ or } \in \mathcal{R}, \quad (3.78b)$$

$$|\mathbf{G}_a(m, n)|^2 = 1, \quad \forall m, n, \quad \forall a \in \mathcal{L} \text{ or } \in \mathcal{R}, \quad (3.78c)$$

$$|\mathbf{F}_a(m, n)|^2 = 1, \quad \forall m, n, \quad \forall a \in \mathcal{L} \text{ or } \in \mathcal{R}. \quad (3.78d)$$

The scalars w_a and p_a denote the rate weight and total sum-power constraint for node $a \in \mathcal{L}$ or $a \in \mathcal{R}$. The constraints (5.12c) and (5.12d) denote the unit-modulus phase-shifter constraint on node a for the analog beamformer and combiner, respectively. In the problem statement, \mathbf{G} , \mathbf{V} , and \mathbf{F} , denote the collection of the analog and digital beamformers and the analog combiners, respectively.

Remark- Note that (3.78), stated as $\ln \det(\cdot)$, is not affected by the digital combiners. They can be chosen to be the well known MMSE combiners and the achieved rate would not be affected (Please see (4)-(9) [64] for more details).

3.3.2 Minorization-Maximization

The problem (3.78) is non-concave due to interference. Given a non-concave structure of the problem due to interference, it makes the searching for its global optima extremely challenging. To render a feasible solution, we leverage the MM [26] approach. Note that the WSR can be written as

$$WSR = WSR_L + WSR_R = \sum_{i_l \in \mathcal{L}} WR_{i_l} + \sum_{i_r \in \mathcal{R}} WR_{i_r} \quad (3.79)$$

where WSR_L and WSR_R denote the WSR in sets \mathcal{L} and \mathcal{R} , respectively, and WR denotes the weighted-rate of one particular user denoted with the subscript. Note that WR_{i_l} is only concave in \mathbf{T}_{i_l, i_r} (point-to-point link between nodes i_l and i_r) and the remaining WRs are non-concave in \mathbf{T}_{i_l, i_r} . Let \bar{i}_l (\bar{i}_r) denote the summation in set \mathcal{L} (\mathcal{R}) without the element i_l (i_r). Since a linear function is simultaneously convex and concave, DC programming introduces the first order Taylor series expansion of $WSR_{\bar{i}_l}$ and WSR_R in \mathbf{T}_{i_l, i_r} , around $\hat{\mathbf{T}}_{i_l, i_r}$, (i.e. all \mathbf{T}_{i_l, i_r}). Let $\hat{\mathbf{T}}$ denote the set containing all such $\hat{\mathbf{T}}_{i_l, i_r}$. The tangent expressios for the FD point-to-point link for the transmit covariance matrices of node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{L}$ can be written by computing the following gradients

$$\hat{\mathbf{A}}_{i_r} = -\frac{\partial WSR_{\bar{i}_l}}{\partial \hat{\mathbf{T}}_{i_l, i_r}} \Big|_{\hat{\mathbf{T}}}, \quad \hat{\mathbf{B}}_{i_r} = -\frac{\partial WSR_R}{\partial \hat{\mathbf{T}}_{i_l, i_r}} \Big|_{\hat{\mathbf{T}}}, \quad (3.80)$$

$$\hat{\mathbf{A}}_{i_l} = -\frac{\partial WSR_{\bar{i}_r}}{\partial \hat{\mathbf{T}}_{i_r, i_l}} \Big|_{\hat{\mathbf{T}}}, \quad \hat{\mathbf{B}}_{i_l} = -\frac{\partial WSR_L}{\partial \hat{\mathbf{T}}_{i_r, i_l}} \Big|_{\hat{\mathbf{T}}}. \quad (3.81)$$

where $\hat{\mathbf{A}}_a$ and $\hat{\mathbf{B}}_a$, for $a \in \mathcal{L}$ or \mathcal{R} , denote the linearization with respect to the same set and the other set, respectively. The gradients can be computed by using the matrix differentiation properties, which leads to the following expressions

$$\hat{\mathbf{A}}_{i_r} = \sum_{m_l \in \mathcal{L}, m_l \neq i_l} \mathbf{H}_{m_l, i_r}^H \mathbf{F}_{m_l}^H (\mathbf{R}_{m_l}^{-1} - \mathbf{R}_{m_l}^{-1}) \mathbf{F}_{m_l} \mathbf{H}_{m_l, i_r} \quad (3.82a)$$

$$\hat{\mathbf{B}}_{i_r} = \sum_{n_r \in \mathcal{R}} \mathbf{H}_{n_r, i_r}^H \mathbf{F}_{n_r}^H (\mathbf{R}_{n_r}^{-1} - \mathbf{R}_{n_r}^{-1}) \mathbf{F}_{n_r} \mathbf{H}_{n_r, i_r}, \quad (3.82b)$$

$$\hat{\mathbf{A}}_{i_l} = \sum_{m_r \in \mathcal{R}, m_r \neq i_r} \mathbf{H}_{m_r, i_l}^H \mathbf{F}_{m_r}^H (\mathbf{R}_{m_r}^{-1} - \mathbf{R}_{m_r}^{-1}) \mathbf{F}_{m_r} \mathbf{H}_{m_r, i_l}, \quad (3.82c)$$

$$\hat{\mathbf{B}}_{i_l} = \sum_{n_l \in \mathcal{L}} \mathbf{H}_{n_l, i_l}^H \mathbf{F}_{n_l}^H (\mathbf{R}_{n_l}^{-1} - \mathbf{R}_{n_l}^{-1}) \mathbf{F}_{n_l} \mathbf{H}_{n_l, i_l}. \quad (3.82d)$$

Note that the rate of node i_r depends on the transmit covariance matrix from node i_l and vice-versa, and the gradients $\hat{\mathbf{B}}_{i_r}$ and $\hat{\mathbf{B}}_{i_l}$ take into account also the SI generated at the node i_l and i_r , respectively. Let λ_{i_l} and λ_{i_r} denote the Lagrange multipliers associated with the sum-power constraint for the nodes i_l and i_r , respectively. Considering the unconstrained analog part, dropping the constant terms and reparametrizing back the transmit covariance matrices as a function of the digital and the analog beamformers, augmenting the minorized WSR cost function with the gradients (3.82) with the sum-power constraints, leads to the following Lagrangian.

$$\begin{aligned} \mathcal{L} = & \sum_{i_l \in \mathcal{L}} \lambda_{i_l} p_{i_l} + \sum_{i_r \in \mathcal{R}} \lambda_{i_r} p_{i_r} + \sum_{i_l \in \mathcal{L}} w_{i_l} \ln \det(\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}) \\ & - \text{Tr}(\mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r} \mathbf{I}) \mathbf{G}_{i_r} \mathbf{V}_{i_r}) + \sum_{i_r \in \mathcal{R}} w_{i_r} \ln \det(\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \\ & \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}) - \text{Tr}(\mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l} \mathbf{V}_{i_l}), \end{aligned} \quad (3.83)$$

Note that (3.83) does not contain the unit-modulus constraints on the analog part and they will be incorporated later.

3.3.3 Hybrid Beamforming and Combining

This section presents a novel HYBF and combining design for solving (3.83).

Digital Beamforming

To optimize the digital for node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, we take the derivative of (3.83) with respect to \mathbf{V}_{i_r} and \mathbf{V}_{i_l} , respectively, which leads to the following KKT conditions

$$\begin{aligned} & \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r} (\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \\ & \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r})^{-1} = \mathbf{G}_{i_r}^H (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r} \mathbf{I}) \mathbf{G}_{i_r} \mathbf{V}_{i_r}, \end{aligned} \quad (3.84a)$$

$$\begin{aligned} & \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l} (\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \\ & \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l})^{-1} = \mathbf{G}_{i_l}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l} \mathbf{V}_{i_l}. \end{aligned} \quad (3.84b)$$

Theorem 3.3.1. *The digital beamformers \mathbf{V}_{i_l} and \mathbf{V}_{i_r} , fixed the remaining variables, can be optimized as a generalized dominant eigenvectors of the pairs of the following matrices*

$$\mathbf{V}_{i_r} = \mathbf{D}_{d_{i_l}} (\mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r}, \mathbf{G}_{i_r}^H (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_r \mathbf{I}) \mathbf{G}_{i_r}), \quad (3.85)$$

$$\mathbf{V}_{i_l} = \mathbf{D}_{d_{i_r}} (\mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l}, \mathbf{G}_{i_l}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l}), \quad (3.86)$$

where $\mathbf{D}_{d_{i_l}}$ ($\mathbf{D}_{d_{i_r}}$) selects the d_{i_l} (d_{i_r}) dominant eigenvectors.

Proof. Given the analog part fixed, we proof the result only for \mathbf{V}_{i_l} . The result for \mathbf{V}_{i_r} is then straightforward based on the proof for \mathbf{V}_{i_l} . The proof relies on simplifying (3.83) with fixed analog part until the Hadamard's inequality applies. The Cholesky decomposition of the matrix $\mathbf{G}_{i_l}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l}$ can be written as $\mathbf{L}_{i_l} \mathbf{L}_{i_l}^H$ where \mathbf{L}_{i_l} is a lower-triangular Cholesky factor. Similarly we do Cholesky decomposition also for $\mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l}$. Given the Cholesky decomposition, we can apply the result provided in Proposition 1 [38] to the matrices $(\mathbf{G}_{i_l}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l})$ and $(\mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l})$. By following similar steps, it follows immediately that the digital beamformer \mathbf{V}_{i_l} can be optimized as a dominant generalized eigenvector of these two matrices. Similar steps can also be followed to prove the result for the digital beamformer \mathbf{V}_{i_r} . \square

Note that the solution stated in Theorem 3.3.1 provides the optimal beamforming directions but not the optimal power allocation. Therefore, to include the optimal power allocation, we normalize the columns of the digital beamformers to unit-norm.

Analog Beamforming

To optimize the analog beamformers for node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, we take the derivative of (3.83) for the analog beamformers \mathbf{G}_{i_l} and \mathbf{G}_{i_r} , which leads to the following KKT conditions

$$\begin{aligned} & \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l} \mathbf{V}_{i_l}^H (\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \\ & \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l})^{-1} = (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l} \mathbf{V}_{i_l} \mathbf{V}_{i_l}^H, \end{aligned} \quad (3.87a)$$

$$\begin{aligned} & \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r} \mathbf{V}_{i_r}^H (\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \\ & \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r})^{-1} = (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r} \mathbf{I}) \mathbf{G}_{i_r} \mathbf{V}_{i_r} \mathbf{V}_{i_r}^H. \end{aligned} \quad (3.87b)$$

To optimize the analog beamformer, the KKT conditions are not resolveable for \mathbf{G}_{i_l} and \mathbf{G}_{i_r} . To do so, we apply the identity $\text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$, which shapes the the KKT conditions as

$$\begin{aligned} & [(\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H (\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r})^{-1})^T \\ & \otimes \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r}] \text{vec}(\mathbf{G}_{i_r}) \\ & = [(\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H)^T \otimes (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_r} \mathbf{I})] \text{vec}(\mathbf{G}_{i_r}), \end{aligned} \quad (3.88a)$$

$$\begin{aligned}
& [(\mathbf{V}_{i_l} \mathbf{V}_{i_l}^H (\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l})^{-1})^T \\
& \otimes \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l}] \text{vec}(\mathbf{G}_{i_l}) \\
& = [(\mathbf{V}_{i_l} \mathbf{V}_{i_l}^H)^T \otimes (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_r} \mathbf{I})] \text{vec}(\mathbf{G}_{i_l}).
\end{aligned} \tag{3.88b}$$

Given the vectorized analog beamformer, by following a similar proof as for the digital beamformers in the Theorem (3.3.1), it can be easily seen that the unconstrained vectorized analog combiners $\text{vec}(\mathbf{G}_{i_r})$ and $\text{vec}(\mathbf{G}_{i_l})$ are given by

$$\begin{aligned}
\text{vec}(\mathbf{G}_{i_r}) = & \mathbf{D}_1((\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H (\mathbf{I} + \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \\
& \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r})^{-1})^T \otimes \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r}, \\
& (\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H)^T \otimes (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_r} \mathbf{I})),
\end{aligned} \tag{3.89a}$$

$$\begin{aligned}
\text{vec}(\mathbf{G}_{i_l}) = & \mathbf{D}_1((\mathbf{V}_{i_l} \mathbf{V}_{i_l}^H (\mathbf{I} + \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \\
& \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l})^{-1})^T \otimes \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l}, \\
& (\mathbf{V}_{i_r} \mathbf{V}_{i_r}^H)^T \otimes (\hat{\mathbf{A}}_{i_r} + \hat{\mathbf{B}}_{i_r} + \lambda_{i_l} \mathbf{I})),
\end{aligned} \tag{3.89b}$$

which needs to be reshaped into correct dimensions with the operation $\text{unvec}(\cdot)$. Furthermore, to meet the unit modulus constraint, we apply the operation $\angle \mathbf{G}_{i_l}$ and $\angle \mathbf{G}_{i_r}$, which scales its amplitude to unit-norm and preserves only the phase part.

Analog Combining

To optimize the analog combiner, we first define the received covariance matrices at the antenna level as $\mathbf{R}_{i_l}^{ant}$ and $\mathbf{R}_{i_r}^{ant}$ (obtained from (3.74) by simply omitting the analog combiner). After the analog combining stage, we have the following expression for the covariance matrices $\mathbf{R}_{i_l} = \mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H$ and $\mathbf{R}_{i_r} = \mathbf{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H$. Note that in the MM approach, we linearize with respect to the WRs for which each transmit covariance matrix generate interference. However, as the combiner do not generate any interference and does not have the sum-power constraint, we can directly consider solving (3.78), which result to be concave for the analog combiners. Namely, we can first write

$$\begin{aligned}
\max_{\mathbf{F}} \sum_{i_l \in \mathcal{L}} w_{i_l} [\ln \det(\mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H) - \ln \det(\mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H)] \\
+ \sum_{i_r \in \mathcal{R}} w_{i_r} [\ln \det(\mathbf{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H) - \ln \det(\mathbf{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H)]
\end{aligned} \tag{3.90}$$

in which both the terms are purely concave, in contrast to minorized version of the WSR, in which the trace terms were only linear. To optimize the analog combiners, we take the derivative of (3.90) with respect to \mathbf{F}_{i_l} and \mathbf{F}_{i_r} , which leads to the following KKT conditions

$$w_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H (\mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H)^{-1} - w_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H (\mathbf{F}_{i_l} \mathbf{R}_{i_l}^{ant} \mathbf{F}_{i_l}^H)^{-1} = 0, \tag{3.91a}$$

$$w_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H (\mathbf{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H)^{-1} - w_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H (\mathbf{F}_{i_r} \mathbf{R}_{i_r}^{ant} \mathbf{F}_{i_r}^H)^{-1} = 0. \tag{3.91b}$$

Given the structure of the KKT conditions, it is immediate to see that the analog combiner can be optimized as a dominant generalized eigenvector solution of

$$\mathbf{F}_{i_l} = \mathbf{D}_{N_{i_l}^r}(\mathbf{R}_{i_l}^{ant}, \mathbf{R}_{i_l}^{ant}), \quad \mathbf{F}_{i_r} = \mathbf{D}_{N_{i_r}^r}(\mathbf{R}_{i_r}^{ant}, \mathbf{R}_{i_r}^{ant}), \quad (3.92)$$

from which we select $M_{i_l}^r$ and $M_{i_r}^r$ rows, respectively, equal to number of RF chains. Given the optimal analog combiner, operation $\angle \mathbf{F}_{i_l}$ and $\angle \mathbf{F}_{i_r}$ is required to meet the unit modulus constraint.

Optimal Power Allocation

Given the optimized beamformers, in this section, we present a novel power allocation scheme for the mMIMO FD interference channel.

Let \mathbf{P}_{i_l} and \mathbf{P}_{i_r} denote the power allocation matrix for the node $i_l \in \mathcal{L}$ and $i_r \in \mathcal{R}$, respectively. Let $\Sigma_{i_l}^{(1)}, \Sigma_{i_l}^{(2)}, \Sigma_{i_r}^{(1)}$ and $\Sigma_{i_r}^{(2)}$ be defined as

$$\Sigma_{i_l}^{(1)} = \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H \mathbf{H}_{i_l, i_r}^H \mathbf{F}_{i_l}^H \mathbf{R}_{i_l}^{-1} \mathbf{F}_{i_l} \mathbf{H}_{i_l, i_r} \mathbf{G}_{i_r} \mathbf{V}_{i_r}, \quad (3.93a)$$

$$\Sigma_{i_l}^{(2)} = \mathbf{V}_{i_r}^H \mathbf{G}_{i_r}^H (\hat{\mathbf{A}}_{i_l} + \hat{\mathbf{B}}_{i_l} + \lambda_{i_r} \mathbf{I}) \mathbf{G}_{i_r} \mathbf{V}_{i_r}, \quad (3.93b)$$

$$\Sigma_{i_r}^{(1)} = \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H \mathbf{H}_{i_r, i_l}^H \mathbf{F}_{i_r}^H \mathbf{R}_{i_r}^{-1} \mathbf{F}_{i_r} \mathbf{H}_{i_r, i_l} \mathbf{G}_{i_l} \mathbf{V}_{i_l}, \quad (3.93c)$$

$$\Sigma_{i_r}^{(2)} = \mathbf{V}_{i_l}^H \mathbf{G}_{i_l}^H (\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + \lambda_{i_l} \mathbf{I}) \mathbf{G}_{i_l} \mathbf{V}_{i_l}. \quad (3.93d)$$

Given (3.93), the problem (3.78) with its minorized version, with respect to the power matrices can be stated as

$$\begin{aligned} \max_{\mathbf{P}_{i_l}} \sum_{i_l \in \mathcal{L}} w_{i_l} \ln \det(\mathbf{I} + \Sigma_{i_l}^{(1)} \mathbf{P}_{i_l}) - \text{Tr}(\Sigma_{i_l}^{(2)} \mathbf{P}_{i_l}), \\ \max_{\mathbf{P}_{i_r}} \sum_{i_r \in \mathcal{R}} w_{i_r} \ln \det(\mathbf{I} + \Sigma_{i_r}^{(1)} \mathbf{P}_{i_r}) - \text{Tr}(\Sigma_{i_r}^{(2)} \mathbf{P}_{i_r}). \end{aligned} \quad (3.94a)$$

To include the optimal power allocation, we take the derivative of (3.94a) for the power matrices \mathbf{P}_{i_l} and \mathbf{P}_{i_r} , and from the KKT conditions it follows that the optimal power allocation matrices can be computed as

$$\mathbf{P}_{i_l} = (w_{i_l} \Sigma_{i_l}^{(2)-1} - \Sigma_{i_l}^{(1)-1})^+, \quad \mathbf{P}_{i_r} = (w_{i_r} \Sigma_{i_r}^{(2)-1} - \Sigma_{i_r}^{(1)-1})^+. \quad (3.95)$$

where $(\mathbf{X})^+ = \max\{\mathbf{0}, \mathbf{X}\}$. To meet the sum-power constraint, we search the optimal multipliers satisfying the power constraints while updating the powers with (3.95). The multipliers λ_{i_l} and λ_{i_r} should be such that the Lagrange dual function (3.83) is finite and the values of the multipliers should be strictly positive. Let $\mathbf{\Lambda}$ and \mathbf{P} denote the collection of multipliers and powers. Formally, the multipliers' search problem can be stated as

$$\min_{\mathbf{\Lambda}} \max_{\mathbf{P}} \mathcal{L}(\mathbf{\Lambda}, \mathbf{P}), \quad \text{s.t.} \quad \mathbf{\Lambda} \succeq \mathbf{0}. \quad (3.96)$$

Algorithm 7 Hybrid Beamforming and Combining

Given: The CSI and rate weights.

Initialize: All the beamformers and combiners.

Repeat until convergence

for $\forall i_l \in \mathcal{L}$ and $\forall i_r \in \mathcal{R}$

 Compute \mathbf{A}_{i_l} and \mathbf{A}_{i_l} with (3.82a) and (3.82c)

 Compute \mathbf{B}_{i_l} and \mathbf{B}_{i_r} with (3.82b) and (3.82d)

 Compute \mathbf{G}_{i_l} with (3.89b), do $\text{unvec}(\mathbf{G}_{i_l})$ and $\angle \mathbf{G}_{i_l}$

 Compute \mathbf{G}_{i_r} with (3.89a), do $\text{unvec}(\mathbf{G}_{i_r})$ and $\angle \mathbf{G}_{i_r}$

 Compute $\mathbf{F}_{i_l}, \mathbf{F}_{i_r}$ with (3.92) and do $\angle \mathbf{F}_{i_l}, \angle \mathbf{F}_{i_r}$

 Compute \mathbf{V}_{i_l} and \mathbf{V}_{i_r} with (3.86) and (3.85)

 Set $\underline{\lambda}_{i_l} = 0$ and $\overline{\lambda}_{i_l} = \mu_{i_{max}} \forall i_l \in \mathcal{L}_l$

Repeat until convergence $\forall i_l \in \mathcal{L}$

 set $\lambda_{i_l} = (\underline{\lambda}_{i_l} + \overline{\lambda}_{i_l})/2$

 Compute $\overline{\mathbf{P}}_{i_l}$ with (3.95),

If constraint for λ_{i_l} is violated,

 set $\underline{\lambda}_{i_l} = \lambda_{i_l}$, **else** $\overline{\lambda}_{i_l} = \lambda_{i_l}$,

Repeat until convergence $\forall i_r \in \mathcal{R}$

 set $\lambda_{i_r} = (\underline{\lambda}_{i_r} + \overline{\lambda}_{i_r})/2$.

 Compute $\overline{\mathbf{P}}_{i_r}$ with (3.95),

If constraint for λ_{i_r} is violated,

 set $\underline{\lambda}_{i_r} = \lambda_{i_r}$, **else** $\overline{\lambda}_{i_r} = \lambda_{i_r}$,

Repeat

The dual function $\max_{\mathbf{P}} \mathcal{L}(\mathbf{A}, \mathbf{P})$ is the pointwise supremum of a family of functions of \mathbf{A} , it is convex [40] and the globally optimal values for \mathbf{A} can be found by using any of the numerous convex optimization techniques. In this work, we adopt the Bisection method. Let $\underline{\lambda}_{i_a}$ and $\overline{\lambda}_{i_a}$, for $i_a \in \mathcal{L}$ or $i_a \in \mathcal{R}$, denote the upper and lower bound for the Lagrange multiplier search for node i_a . Let $[0, \overline{\lambda}_{i_a}^{max}]$ denote the maximum search interval for the multipliers of node i_a . The WSR maximizing HYBF design for the mMIMO interference channel is given in Algorithm 7.

Short Convergence Proof

For the WSR cost function (3.78), we construct its minorizer which restates the WSR maximization as a concave problem, when the remaining variables are fixed in the gradients. The minorizer results to be a touching lower bound for the original WSR problem, therefore we can write

$$\begin{aligned} \underline{WSR} = & \sum_{i_l \in \mathcal{L}} w_{i_l} \ln \det(\mathbf{I} + \Sigma_{i_l}^{(1)} \mathbf{P}_{i_l}) - \text{Tr}(\Sigma_{i_l}^{(2)} \mathbf{P}_{i_l}), \\ & + \sum_{i_r \in \mathcal{R}} w_{i_r} \ln \det(\mathbf{I} + \Sigma_{i_r}^{(1)} \mathbf{P}_{i_r}) - \text{Tr}(\Sigma_{i_r}^{(2)} \mathbf{P}_{i_r}) \end{aligned} \quad (3.97)$$

The minorizer, which is concave in transmit covariance matrices, still has the same gradient of the original WSR and hence the KKT conditions are not affected. Now reparameterizing the transmit covariance matrices in terms of beamformer with the optimal power matrices and adding the power constraints to the minorizer, we get the Lagrangian (3.83). By updating all the beamformers and combiners as a dominant generalized eigenvector solution, leads to a monotonic increase in the WSR [38] at each iteration, thus assuring convergence.

3.3.4 Simulation Results

This section presents the simulation results for the proposed HYBF and combining design for WSR maximization in a mmWave FD mMIMO interference channel.

We consider a scenario of 2 point-to-point communication links consisting of 4 nodes operating in the mmWave. We assume that all the nodes are equipped with uniform linear arrays, with transmit antennas placed at half-wavelength. The angle of arrival $\phi_{i_l}^{n_p, n_c}$ and angle of departure $\theta_{i_l}^{n_p, n_c}$ are chosen to be uniformly distributed in the interval $[-20^\circ, 20^\circ]$. The SI channel is modelled with the Rician factor $\kappa_{i_l} = 10^5$ dB and the distance $r_{m,n}$ is modelled as in [54], with distance between the transmit and receive array of 20 cm and with a relative angle of 90° . The number of clusters and the number of paths is set to be $N_c = 3$ and $N_p = 6$. We define the transmit SNR at node i_l as $\text{SNR}_{i_l} = p_{i_l}/\sigma_{i_l}^2$ and assume it to be the same for all the nodes, denoted as SNR. In the plots, the proposed HYBF scheme will be referred as HYBF. For comparison purposes, we define the following benchmark schemes.

- *Fully Digital FD* - with all the FD nodes having number of RF chains equal to the number of antennas.
- *Fully Digital HD* - with all the nodes having number of RF chains equal to the number of antennas but operating in HD mode by splitting the resources in time.

Figure 3.7 shows the achieved WSR as a function of SNR with a different number of RF chains in comparison with the benchmark schemes. It can be clearly seen that the proposed HYBF and combining has the potential to perform very close to the fully digital FD beamforming design with 32 RF chains. Still the proposed algorithm exhibits some gap compared to the fully digital solution as the analog beamforming stage is constrained and must meet the unit modulus constraint. We can see that also with 16RF chains, the proposed algorithm can considerably outperform the fully digital HD scheme, with operates with 100 transmit and receive RF chains. Figure 3.8 shows the average WSR as a function of SNR with 64 transmit and receive antennas. It can be seen that with the same number of RF chains (32) and less number of transmit and receive antennas, the performance of the proposed HYB algorithm gets strictly close to the fully digital FD scheme.

3.3.5 Conclusions

In this section, we studied the problem of WSR maximization in a mmWave massive MIMO FD interference channel consisting of K point-to-point FD communication links.

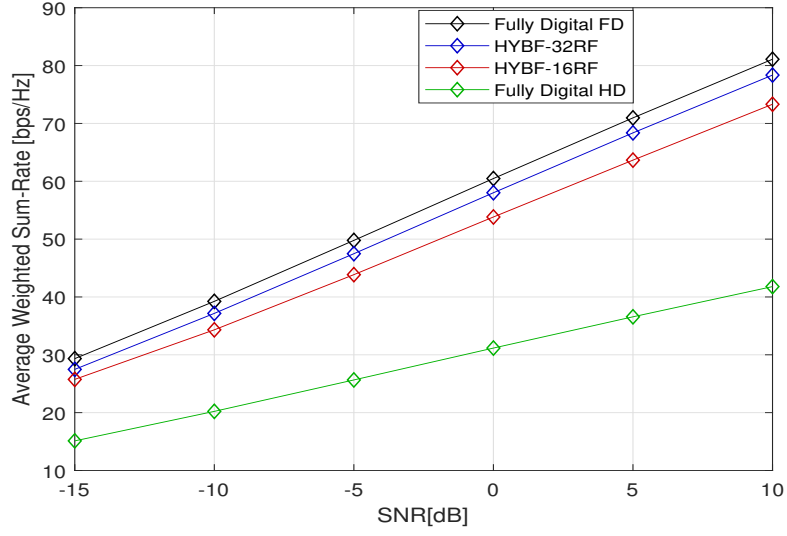


Figure 3.7: Average WSR as a function of SNR with transmit and receive antennas $N_a^t = N_a^r = 100$ and RF chains $M_a^t = M_a^r = 32$ or 16 , $\forall a \in \mathcal{L}$ or \mathcal{R} .

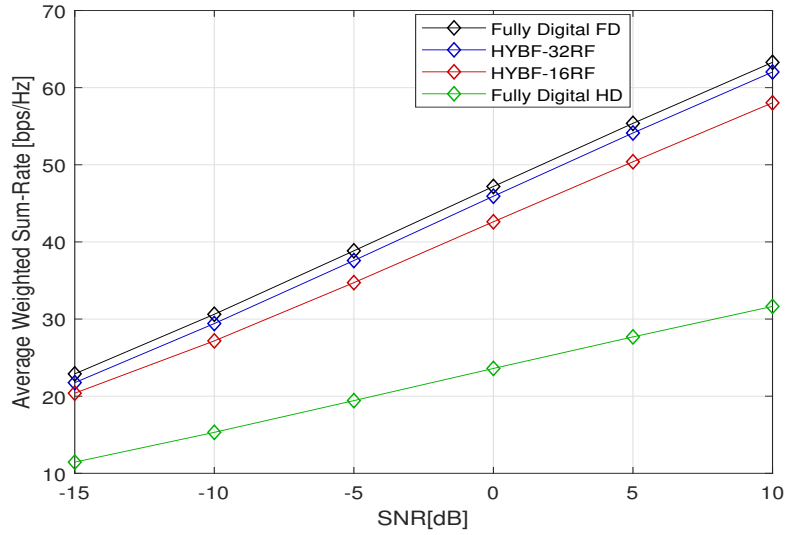


Figure 3.8: Average WSR as a function of SNR with transmit and receive antennas $N_a^t = N_a^r = 64$ and RF chains $M_a^t = M_a^r = 32$ or 16 , $\forall a \in \mathcal{L}$ or \mathcal{R} .

The simultaneous coexistence of multiple FD nodes leads to an extremely challenging communication scenario for which a novel hybrid beamforming and combining scheme is proposed. Simulation results show that the proposed design can perform extremely close to the fully digital FD beamforming design operating with 100 antennas, with only 32 RF chains. Moreover, the proposed design significantly outperforms the fully digital HD system with only 16 RF chains.

Chapter 4

Hybrid Beamforming for Single-Cell Massive MIMO Millimeter Full Duplex

4.1 Introduction and Motivation

This chapter considers the problem of HYBF for a mmWave mMIMO FD system with multi-antenna uplink (UL) and downlink (DL) users. For the literature on HYBF for mmWave FD, we refer the reader to [24, 25, 65, 52, 53, 54, 55, 56, 57, 44, 49, 45, 66, 67, 68, 69, 70, 71, 72]. In [24], a novel HYBF design for the mmWave FD with one UL user while jointly performing radar target sensing on the received side with hybrid combining is proposed. In [65, 52, 53, 54, 55, 56, 57], novel HYBF designs for a point-to-point mmWave massive MIMO (mMIMO) FD system are studied. HYBF schemes of mMIMO FD relays and integrated access and backhaul are presented in [44, 49, 45] and [66], respectively. HYBF designs with single antenna UL and DL users for a single-cell and a multi-cell mmWave FD system are proposed in [67] and [68], respectively. In [69], HYBF for mmWave mMIMO FD with only one UL and one DL multi-antenna user, under the receive LDR is proposed. In [70], HYBF for two fully connected mMIMO FD nodes that approaches SI-free sum-spectral efficiency is proposed. In [25], HYBF for a mmWave FD system equipped with analog SI cancellation stage is presented. In [71], HYBF to generalize the point-to-point mmWave mMIMO FD communication to the case of a K-pair links is presented. Frequency-selective HYBF for a wide-band mmWave FD system is studied in [72].

The literature on multi-antenna multi-user mmWave FD systems is limited only to the case of one UL and one DL user [25, 69, 70, 72]. In [69], the receive side LDR of FD BS is also considered, which is dominated by the quantization noise of the ADCs. However, LDR noise from the transmit side is ignored, which also affects the performance of FD systems significantly [36]. The effect of cross-interference generated from the UL user towards the DL user is also not considered in [69], which can have a major impact on the achievable performance. Cross-interference generated from the neighbouring cells is well investigated in the DTDD networks [73, 7, 74, 75, 76], and it is more harmful to the multi-user FD systems as it occurs in the same cell. For example, consider the case

of a small cell, in which BSs and users are expected to operate with a similar amount of transmit power [76]. Suppose that one FD BS simultaneously serves one UL and one DL user and that both the users are located close to each other and sufficiently far from the BS. In such a case, cross-interference can become as severe as the SI and can completely drown the useful signal intended for the DL user if not considered in the beamforming design. In a multi-user scenario with multiple UL users located near the DL users, each DL user suffers from cross-interference, which is summed over all the UL users' transmit power, with each UL user transmitting with a similar amount of power as the BS. In such a case, cross-interference can become even more severe than the SI if not considered in the design.

4.1.1 Main Contributions

We present a novel HYBF design to maximize the WSR in a single-cell mmWave mMIMO FD system, i.e., for multiple multi-antenna UL and DL users. The users are assumed to have a limited number of antennas and digital processing capability. The FD BS is assumed to have a massive number of antennas and hybrid processing capability. Our design is based on alternating optimization and relies on the mathematical tools offered by MM [26]. The users and BS are assumed to be suffering from the LDR noise due to non-ideal hardware, modelled with the traditional LDR model [37] and by extending it to the case of a hybrid transceiver, respectively. Our work represents the first-ever impairment aware HYBF approach for mmWave FD and its analysis as a function of the LDR noise levels. Extension of the LDR noise model presented herein is applicable to any mmWave FD scenario.

In contrast to the conventional HYBF designs for mmWave FD, in this work, the beamformers are designed under the joint sum-power and the practical per-antenna power constraints. The sum-power constraint at each terminal is imposed by the regulations, which limits its total transmit power. In practice, each transmit antenna is equipped with its PA¹ [34] and the per-antenna power constraints arise due to power consumption limits imposed on the physical PAs [78]. We also present a novel SI, interference, cross-interference and LDR noise aware optimal power allocation scheme to meet the joint constraints.

Compared to the digital part, optimization of the analog stage is more challenging as it must obey the unit-modulus constraint and it is common to all the users. Recently, new transceivers have started to emerge, which with the aid of amplitude modulators (AMs), also allow amplitude control for the analog stage [79, 80, 69]. Such transceivers alleviate the unit-modulus constraint but require additional hardware. Hence, we study both the unit-modulus and AMs cases and investigate when the amplitude control for mmWave FD could be advantageous. In practice, as the analog beamformer and analog combiner can assume only finite values, a quantization constraint is also imposed on them during the optimization process. In our problem formulation, the WSR does not depend on the digital combiners, which are omitted in the design. They must be chosen

¹The mMIMO systems are also expected to be deployed with one PA per-antenna to enable the deployment of very low-cost PAs [77].

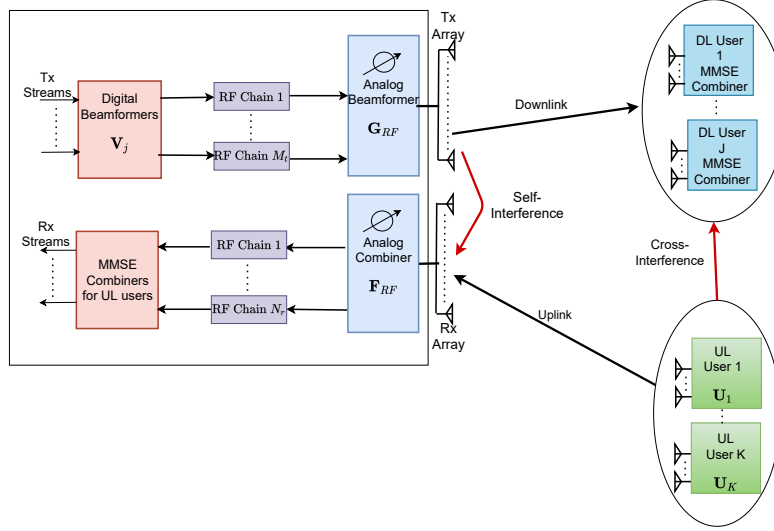


Figure 4.1: FD in mmWave with HYBF to serve multi-antenna users. Tx and Rx denote transmit and receive, respectively.

as the minimum-mean-squared-error (MMSE) combiners after the convergence of the proposed algorithm. By omitting the digital combiners, equal to the sum of the number of UL and DL users, the HYBF design simplifies, and the per-iteration computational complexity reduces significantly.

Simulation results show that our design outperforms a fully digital HD system and can deal with the SI, interference and cross-interference with only a few RF chains. Results are reported with different LDR noise levels, and significant performance gain is observed at any level.

In summary, the contributions of our work are:

- Extension of the LDR noise model for the mmWave band.
- Introduction of the WSR maximization problem formulation for HYBF in a single-cell mmWave mMIMO FD system affected by the LDR noise.
- A novel SI, interference, cross-interference, LDR noise and the practical per-antenna power constraints aware HYBF design.
- Investigation of the achievable WSR in a multi-user mmWave FD system as a function of the LDR noise.
- Optimal interference, SI, LDR noise and the per-antenna power constraints aware power allocation scheme for the hybrid FD BS and UL users.

4.2 System Model

We consider a single-cell mmWave FD system consisting of one hybrid FD BS serving J DL and K UL fully digital multi-antenna users, as shown in Fig. 4.1. We assume

perfect channel state information (CSI)². The FD BS is assumed to have M_t transmit and N_r receive RF chains, and M_0 transmit and N_0 receive antennas. Let $\mathcal{U} = \{1, \dots, K\}$ and $\mathcal{D} = \{1, \dots, J\}$ denote the sets containing the indices of K UL and J DL users, respectively. Let M_k and N_j denote the number of transmit and receive antennas for k -th UL and j -th DL user, respectively. We consider a multi-stream approach and the number of data streams for k -th UL and j -th DL user are denoted as u_k and v_j , respectively. Let $\mathbf{U}_k \in \mathbb{C}^{M_k \times u_k}$ and $\mathbf{V}_j \in \mathbb{C}^{M_j \times v_j}$ denote the precoders for white unitary variance data streams $\mathbf{s}_k \in \mathbb{C}^{u_k \times 1}$ and $\mathbf{s}_j \in \mathbb{C}^{v_j \times 1}$, respectively. Let $\mathbf{G}_{RF} \in \mathbb{C}^{M_0 \times M_t}$ and $\mathbf{F}_{RF} \in \mathbb{C}^{N_0 \times N_r}$ denote the fully connected analog beamformer and combiner at the FD BS, respectively. Let $\mathcal{P} = \{1, e^{i2\pi/n_{ps}}, \dots, e^{i2\pi(n_{ps}-1)/n_{ps}}\}$ denote the set of n_{ps} possible discrete values that the phasors at the analog stage can assume on unit-circle.

For HYBF with the unit-modulus constraint, we define the quantizer function $\mathbb{Q}_P(\cdot)$ to quantize the unit-modulus phasors of analog beamformer \mathbf{G}_{RF} and combiner \mathbf{F}_{RF} such that $\mathbb{Q}_P(\angle \mathbf{G}_{RF}(m, n)) \in \mathcal{P}$ and $\mathbb{Q}_P(\angle \mathbf{F}_{RF}(m, n)) \in \mathcal{P}$, $\forall m, n$. For HYBF with amplitude control, the phase part is still quantized with $\mathbb{Q}_P(\cdot)$ and belongs to \mathcal{P} . Let $\mathcal{A} = \{a_0, \dots, a_{A-1}\}$ denote the set of A possible values that the amplitudes can assume. Let $\mathbb{Q}_A(\cdot)$ denote the quantizer function to quantize the amplitudes of \mathbf{G}_{RF} and \mathbf{F}_{RF} such that $\mathbb{Q}_A(|\mathbf{G}_{RF}(m, n)|) \in \mathcal{A}$ and $\mathbb{Q}_A(|\mathbf{F}_{RF}(m, n)|) \in \mathcal{A}$, $\forall m, n$. A complex number $\mathbf{G}_{RF}(m, n)$ with amplitude in \mathcal{A} and phase part in \mathcal{P} can be written as $\mathbf{G}_{RF}(m, n) = \mathbb{Q}_A(|\mathbf{G}_{RF}(m, n)|) \mathbb{Q}_P(\angle \mathbf{G}_{RF}(m, n))$. The thermal noise vectors for FD BS and j -th DL user are denoted as $\mathbf{n}_0 \sim \mathcal{CN}(0, \sigma_0^2 \mathbf{I}_{N_0})$ and $\mathbf{n}_j \sim \mathcal{CN}(0, \sigma_j^2 \mathbf{I}_{N_j})$, respectively. Let \mathbf{c}_k and \mathbf{e}_j denote the LDR noise vectors for k -th UL and j -th DL user, respectively, which can be modelled as [37]

$$\mathbf{c}_k \sim \mathcal{CN}(\mathbf{0}_{M_k \times 1}, k_k \text{diag}(\mathbf{U}_k \mathbf{U}_k^H)), \quad (4.1)$$

$$\mathbf{e}_j \sim \mathcal{CN}(\mathbf{0}_{N_j \times 1}, \beta_j \text{diag}(\mathbf{\Phi}_j)), \quad (4.2)$$

where $k_k \ll 1$, $\beta_j \ll 1$, $\mathbf{\Phi}_j = \text{Cov}(\mathbf{r}_j)$ and \mathbf{r}_j denotes the undistorted signal received by j -th DL user. Let \mathbf{c}_0 and \mathbf{e}_0 denote the LDR noise vectors in transmission and reception for FD BS, respectively. We model them as

$$\mathbf{c}_0 \sim \mathcal{CN}(\mathbf{0}_{M_0 \times 1}, k_0 \text{diag}(\sum_{n \in \mathcal{D}} \mathbf{G}_{RF} \mathbf{V}_n \mathbf{V}_n^H \mathbf{G}_{RF}^H)), \quad (4.3)$$

$$\mathbf{e}_0 \sim \mathcal{CN}(\mathbf{0}_{N_r \times 1}, \beta_0 \text{diag}(\mathbf{\Phi}_0)), \quad (4.4)$$

where $k_0 \ll 1$, $\beta_0 \ll 1$, $\mathbf{\Phi}_0 = \text{Cov}(\mathbf{r}_0)$ and \mathbf{r}_0 denotes the undistorted signal received by FD BS after the analog combiner \mathbf{F}_{RF} . Note that (4.3) extends the transmit LDR noise model from [37] to the case of a hybrid transmitter. For the hybrid receiver at the mmWave FD BS, ADCs, the most dominant sources of receive LDR noise, are placed after the analog combiner \mathbf{F}_{RF} . Consequently, \mathbf{e}_0 in (4.4) considers the undistorted signal received after the analog combiner. We remark that the extension presented in

²The CSI of the mmWave FD systems can be acquired similarly as in [81] for the mmWave HD system and it is part of the ongoing research [82].

(4.3)-(4.4) is slightly simplified. In practice, as some circuitry might be shared among multiple antennas, it can lead to some correlation.

Let \mathbf{y} and \mathbf{y}_j denote the signals received by the FD BS and j -th DL user, respectively, which can be written as

$$\begin{aligned} \mathbf{y} = & \mathbf{F}_{RF}^H \sum_{k \in \mathcal{U}} \mathbf{H}_k \mathbf{U}_k \mathbf{s}_k + \mathbf{F}_{RF}^H \sum_{k \in \mathcal{U}} \mathbf{H}_k \mathbf{U}_k \mathbf{c}_k + \mathbf{F}_{RF}^H \mathbf{n}_0 \\ & + \mathbf{F}_{RF}^H \mathbf{H}_0 \sum_{j \in \mathcal{D}} \mathbf{G}_{RF} \mathbf{V}_j \mathbf{s}_j + \mathbf{F}_{RF}^H \mathbf{H}_0 \mathbf{c}_0 + \mathbf{e}_0, \end{aligned} \quad (4.5)$$

$$\begin{aligned} \mathbf{y}_j = & \mathbf{H}_j \sum_{n \in \mathcal{D}} \mathbf{G}_{RF} \mathbf{V}_n \mathbf{s}_n + \mathbf{H}_j \sum_{n \in \mathcal{D}} \mathbf{G}_{RF} \mathbf{V}_n \mathbf{c}_0 + \mathbf{e}_j + \mathbf{n}_j \\ & + \sum_{k \in \mathcal{U}} \mathbf{H}_{j,k} \mathbf{U}_k \mathbf{s}_k + \sum_{k \in \mathcal{U}} \mathbf{H}_{j,k} \mathbf{c}_k. \end{aligned} \quad (4.6)$$

The matrices $\mathbf{H}_k \in \mathbb{C}^{N_0 \times M_k}$ and $\mathbf{H}_j \in \mathbb{C}^{N_j \times M_0}$ denote channel response from the k -th UL user to BS and from the BS to j -th DL user, respectively. The matrices $\mathbf{H}_0 \in \mathbb{C}^{N_0 \times M_0}$ and $\mathbf{H}_{j,k} \in \mathbb{C}^{N_j \times M_k}$ denote SI channel response for FD BS and cross-interference channel response between k -th UL and j -th DL users, respectively. At the mmWave, the channel response \mathbf{H}_k can be modelled as [54]

$$\mathbf{H}_k = \sqrt{\frac{M_k N_0}{N_c N_p}} \sum_{n_c=1}^{N_c} \sum_{n_p=1}^{N_p} \alpha_k^{(n_p, n_c)} \mathbf{a}_r(\phi_k^{n_p, n_c}) \mathbf{a}_t^T(\theta_k^{n_p, n_c}), \quad (4.7)$$

where N_c and N_p denote the number of clusters and number of rays (Figure 1 [54]), respectively, and $\alpha_k^{(n_p, n_c)} \sim \mathcal{CN}(0, 1)$ denotes a complex Gaussian random variable whose amplitude and phase are Rayleigh and uniformly distributed, respectively. The vectors $\mathbf{a}_r(\phi_k^{n_p, n_c})$ and $\mathbf{a}_t^T(\theta_k^{n_p, n_c})$ denote the receive and transmit antenna array response with angle of arrival (AoA) $\phi_k^{n_p, n_c}$ and angle of departure (AoD) $\theta_k^{n_p, n_c}$, respectively. The channel matrices \mathbf{H}_j and $\mathbf{H}_{j,k}$ can be modelled similarly as in (4.7). The SI channel can be modelled as [54]

$$\mathbf{H}_0 = \sqrt{\frac{\kappa}{\kappa + 1}} \mathbf{H}_{LoS} + \sqrt{\frac{1}{\kappa + 1}} \mathbf{H}_{ref}, \quad (4.8)$$

where κ denotes the Rician factor, and the matrices \mathbf{H}_{LoS} and \mathbf{H}_{ref} denote the line-of-sight (LoS) and reflected contributions, respectively. The channel matrix \mathbf{H}_{ref} can be modelled as (4.7) and element of \mathbf{H}_{LoS} at the m -th row and n -th column can be modelled as [54]

$$\mathbf{H}_{LoS}(m, n) = \frac{\rho}{r_{m,n}} e^{-i2\pi \frac{r_{m,n}}{\lambda}}. \quad (4.9)$$

where ρ denotes the power normalization constant to assure $\mathbb{E}(\|\mathbf{H}_{LoS}(m, n)\|_F^2) = M_0 N_0$ and λ denotes the wavelength. The scalar $r_{m,n}$ denotes distance between the m -th receive and n -th transmit antenna, which depends on the transmit and receive array geometry (9) [54]. The aforementioned notations are summarized in Table 4.1.

Table 4.1: Notations

M_t	Number of transmit RF chains for the BS
N_r	Number of receive RF chains for the FD BS
M_0	Number of transmit antennas for the BS
N_0	Number of receive antennas for the BS
M_k	Number of transmit antennas for UL user k
N_j	Number of receive antennas for DL user j
\mathbf{U}_k	Digital beamformer for UL user k
\mathbf{V}_j	Digital beamformer for DL user j
\mathbf{G}_{RF}	Analog beamformer for the FD BS
\mathbf{F}_{RF}	Analog combiner for the FD BS
\mathbf{c}_k	Transmit LDR noise from UL user k
\mathbf{c}_0	Transmit LDR noise from the FD BS
\mathbf{e}_0	Receive LDR noise at the FD BS
\mathbf{e}_j	Receive LDR noise at the DL user j
\mathbf{n}_0	Thermal noise at the FD BS
\mathbf{n}_j	Thermal noise at the DL user j
\mathbf{H}_0	SI channel
\mathbf{H}_k	Channel between the BS and UL user k
\mathbf{H}_j	Channel between the BS and DL user j
$\mathbf{H}_{j,k}$	Cross-interference channel between UL user k and DL user j
\preceq	Element-wise inequality

4.2.1 Problem Formulation

Let \bar{k} and \bar{j} denote the indices in sets \mathcal{U} and \mathcal{D} without the elements k and j , respectively. The received (signal plus) interference and noise covariance matrices from UL user $k \in \mathcal{U}$ at the BS and by the DL user $j \in \mathcal{D}$ are denoted as $(\mathbf{R}_k) \mathbf{R}_{\bar{k}}$ and $(\mathbf{R}_j) \mathbf{R}_{\bar{j}}$, respectively. Let \mathbf{T}_k , $\forall k \in \mathcal{U}$, and \mathbf{Q}_j , $\forall j \in \mathcal{D}$, defined as

$$\mathbf{T}_k = \mathbf{U}_k \mathbf{U}_k^H, \quad (4.10a)$$

$$\mathbf{Q}_j = \mathbf{G}_{RF} \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H, \quad (4.10b)$$

denote the transmit covariance matrices from UL user $k \in \mathcal{U}$ and DL user $j \in \mathcal{D}$, respectively. By considering the distortions from non-ideal hardware with the extended LDR noise model, cross-interference, interference and SI, the received covariance matrices at the BS after the analog combiner, i.e., \mathbf{R}_k and $\mathbf{R}_{\bar{k}}$, and at the DL user $j \in \mathcal{D}$, i.e., \mathbf{R}_j and $\mathbf{R}_{\bar{j}}$, can be written as (4.11), shown at the top of the next page. In (4.11), \mathbf{s}_k and \mathbf{s}_j denote the useful received signal covariance matrices from k -th UL user at the FD BS and by j -th DL user, respectively. The undistorted received covariance matrices can be recovered from (4.11) as $\Phi_0 = \mathbf{R}_k$, with $\beta_0 = 0$, and $\Phi_j = \mathbf{R}_j$, with $\beta_j = 0$.

The WSR maximization problem with respect to the digital beamformers, analog beamformer and combiner with amplitudes in \mathcal{A} and phase part in \mathcal{P} , under the joint sum-power and per-antenna power constraints, can be stated as

$$\begin{aligned} \mathbf{R}_k = & \underbrace{\mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{T}_k \mathbf{H}_k^H \mathbf{F}_{RF}}_{\triangleq \mathbf{s}_k} + \sum_{\substack{i \in \mathcal{U} \\ i \neq k}} \mathbf{F}_{RF}^H \mathbf{H}_i \mathbf{T}_i \mathbf{H}_i^H \mathbf{F}_{RF} + \sum_{i \in \mathcal{U}} k_i \mathbf{F}_{RF}^H \mathbf{H}_i \text{diag}(\mathbf{T}_i) \mathbf{H}_i^H \mathbf{F}_{RF} \\ & + \sigma_0^2 \mathbf{I}_{N_0} + \beta_0 \text{diag}(\Phi_0) + \mathbf{F}_{RF}^H \mathbf{H}_0 \left(\sum_{n \in \mathcal{D}} \mathbf{Q}_n + k_0 \text{diag} \left(\sum_{n \in \mathcal{D}} \mathbf{Q}_n \right) \right) \mathbf{H}_0^H \mathbf{F}_{RF}, \end{aligned} \quad (4.11a)$$

$$\begin{aligned} \mathbf{R}_j = & \underbrace{\mathbf{H}_j \mathbf{Q}_j \mathbf{H}_j^H}_{\triangleq \mathbf{s}_j} + \mathbf{H}_j \sum_{\substack{n \in \mathcal{D} \\ n \neq j}} \mathbf{Q}_n \mathbf{H}_j^H + k_0 \mathbf{H}_j \text{diag} \left(\sum_{n \in \mathcal{D}} \mathbf{Q}_n \right) \mathbf{H}_j^H + \sigma_j^2 \mathbf{I}_{N_j} \\ & + \sum_{i \in \mathcal{U}} \mathbf{H}_{j,i} \left(\mathbf{T}_i + k_i \text{diag}(\mathbf{T}_i) \right) \mathbf{H}_{j,i}^H + \beta_j \text{diag}(\Phi_j), \end{aligned} \quad (4.11b)$$

$$\mathbf{R}_{\bar{k}} = \mathbf{R}_k - \mathbf{s}_k, \quad \mathbf{R}_{\bar{j}} = \mathbf{R}_j - \mathbf{s}_j. \quad (4.11c)$$

$$\max_{\substack{\mathbf{U}, \mathbf{V}, \\ \mathbf{G}_{RF} \mathbf{F}_{RF}}} \sum_{k \in \mathcal{U}} w_k \ln \det(\mathbf{R}_{\bar{k}}^{-1} \mathbf{R}_k) + \sum_{j \in \mathcal{D}} w_j \ln \det(\mathbf{R}_{\bar{j}}^{-1} \mathbf{R}_j) \quad (4.12a)$$

$$\text{s.t.} \quad \text{diag}(\mathbf{U}_k \mathbf{U}_k^H) \preceq \mathbf{\Lambda}_k, \quad \forall k \in \mathcal{U}, \quad (4.12b)$$

$$\text{diag} \left(\sum_{j \in \mathcal{D}} \mathbf{G}_{RF} \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H \right) \preceq \mathbf{\Lambda}_0, \quad (4.12c)$$

$$\text{Tr}(\mathbf{U}_k \mathbf{U}_k^H) \leq \alpha_k, \quad \forall k \in \mathcal{U}, \quad (4.12d)$$

$$\text{Tr} \left(\sum_{j \in \mathcal{D}} \mathbf{G}_{RF} \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H \right) \leq \alpha_0. \quad (4.12e)$$

$$\angle \mathbf{G}_{RF}(m, n) \in \mathcal{P}, \text{ and } |\mathbf{G}_{RF}(m, n)| \in \mathcal{A}, \quad \forall m, n, \quad (4.12f)$$

$$\angle \mathbf{F}_{RF}(i, j) \in \mathcal{P}, \text{ and } |\mathbf{F}_{RF}(i, j)| \in \mathcal{A}, \quad \forall i, j. \quad (4.12g)$$

The scalars w_k and w_j denote rate weights for the UL user k and DL user j , respectively. The diagonal matrices $\mathbf{\Lambda}_k$ and $\mathbf{\Lambda}_0$ denote the per-antenna power constraints for the k -th UL user and FD BS, respectively, and the scalars α_k and α_0 denote their sum-power constraint. The collections of digital UL and DL beamformers are denoted as \mathbf{U} and \mathbf{V} , respectively. For unit-modulus HYBF, the constraints in (4.12f) – (4.12g) on the amplitude part become unit-modulus.

Remark 1: Note that the rate achieved with (4.12) is not affected by the digital receivers if they are chosen as the MMSE combiners, see e.g., (4) – (9) [64] for more details. For WSR maximization with the combiners, only the analog combiner has to be considered in the optimization problem as it affects the size of the received covariance matrices from UL users, i.e., the UL rate.

4.3 Minorization-Maximization

Problem (4.12) is non-concave in the transmit covariance matrices \mathbf{T}_k and \mathbf{Q}_j due to the interference terms and searching its globally optimum solution is very challenging. In this section, we present the MM optimization method [26] for solving (4.12) to a local optimum.

The WSR maximization problem (4.12) will be reformulated at each iteration as a concave reformulation with its minorizer, using the DC programming [38] in terms of the variable to be updated, while the other variables will be kept fixed. To proceed, note that the WSR in (4.12) can be written with the weighted-rate (WR) of user $k \in \mathcal{U}$, user $j \in \mathcal{D}$, WSRs for \bar{k} and \bar{j} as

$$\text{WSR} = \underbrace{\text{WR}_k^{UL} + \text{WSR}_{\bar{k}}^{UL}}_{\triangleq \text{WSR}^{UL}} + \underbrace{\text{WR}_j^{DL} + \text{WSR}_{\bar{j}}^{DL}}_{\triangleq \text{WSR}^{DL}}, \quad (4.13)$$

where WSR^{UL} and WSR^{DL} denote the WSR in UL and DL, respectively. Considering the dependence of the transmit covariance matrices, only WR_k^{UL} is concave in \mathbf{T}_k , meanwhile $\text{WSR}_{\bar{k}}^{UL}$ and WSR^{DL} are non-concave in \mathbf{T}_k , when $\mathbf{T}_{\bar{k}}$ and \mathbf{Q}_j , $\forall j \in \mathcal{D}$, are fixed. Similarly, only WSR_j^{DL} is concave in \mathbf{Q}_j and non-concave in $\text{WSR}_{\bar{j}}^{DL}$ and WSR^{UL} , when $\mathbf{Q}_{\bar{j}}$ and \mathbf{T}_k , $\forall k \in \mathcal{U}$, are fixed. Since a linear function is simultaneously convex and concave, DC programming introduces the first order Taylor series expansion of $\text{WSR}_{\bar{k}}^{UL}$ and WSR^{DL} in \mathbf{T}_k , around $\hat{\mathbf{T}}_k$ (i.e. around all \mathbf{T}_k), and of $\text{WSR}_{\bar{j}}^{DL}$ and WSR^{UL} in \mathbf{Q}_j , around $\hat{\mathbf{Q}}_j$ (i.e. around all \mathbf{Q}_j). Let $\hat{\mathbf{T}}$ and $\hat{\mathbf{Q}}$ denote the set containing all such $\hat{\mathbf{T}}_k$ and $\hat{\mathbf{Q}}_j$, respectively. Let $\hat{\mathbf{R}}_k(\hat{\mathbf{T}}, \hat{\mathbf{Q}})$, $\hat{\mathbf{R}}_{\bar{k}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})$, $\hat{\mathbf{R}}_j(\hat{\mathbf{T}}, \hat{\mathbf{Q}})$, and $\hat{\mathbf{R}}_{\bar{j}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})$ denote the covariance matrices $\mathbf{R}_k, \mathbf{R}_{\bar{k}}, \mathbf{R}_j$ and $\mathbf{R}_{\bar{j}}$ as a function of $\hat{\mathbf{T}}$ and $\hat{\mathbf{Q}}$, respectively. The linearized tangent expressions for each communication link by computing the gradients

$$\hat{\mathbf{A}}_k = -\frac{\partial \text{WSR}_{\bar{k}}^{UL}}{\partial \mathbf{T}_k} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad \hat{\mathbf{B}}_k = -\frac{\partial \text{WSR}^{DL}}{\partial \mathbf{T}_k} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad (4.14a)$$

$$\hat{\mathbf{C}}_j = -\frac{\partial \text{WSR}_{\bar{j}}^{DL}}{\partial \mathbf{Q}_j} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad \hat{\mathbf{D}}_j = -\frac{\partial \text{WSR}^{UL}}{\partial \mathbf{Q}_j} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad (4.14b)$$

with respect to the transmit covariance matrices \mathbf{T}_k and \mathbf{Q}_j can be written as

$$\underline{\text{WSR}}_{\bar{k}}^{UL}(\mathbf{T}_k, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \text{WSR}_{\bar{k}}^{UL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}\left((\mathbf{T}_k - \hat{\mathbf{T}}_k) \hat{\mathbf{A}}_k\right), \quad (4.15a)$$

$$\underline{\text{WSR}}^{DL}(\mathbf{T}_k, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \text{WSR}^{DL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}\left((\mathbf{T}_k - \hat{\mathbf{T}}_k) \hat{\mathbf{B}}_k\right), \quad (4.15b)$$

$$\underline{\text{WSR}}_{\bar{j}}^{DL}(\mathbf{Q}_j, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \text{WSR}_{\bar{j}}^{DL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}\left((\mathbf{Q}_j - \hat{\mathbf{Q}}_{\bar{j}}) \hat{\mathbf{C}}_j\right), \quad (4.15c)$$

$$\underline{\text{WSR}}^{UL}(\mathbf{Q}_j, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \text{WSR}^{UL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}\left((\mathbf{Q}_j - \hat{\mathbf{Q}}_j) \hat{\mathbf{D}}_j\right). \quad (4.15d)$$

$$\begin{aligned} \hat{\mathbf{A}}_k = \sum_{i \in \mathcal{U}, i \neq k} w_i & \left(\mathbf{H}_k^H \mathbf{F}_{RF} \left[\hat{\mathbf{R}}_{\bar{i}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_i(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \beta_0 \text{diag} \left(\hat{\mathbf{R}}_{\bar{i}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right. \right. \right. \\ & \left. \left. \left. - \hat{\mathbf{R}}_i(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \right] \mathbf{F}_{RF}^H \mathbf{H}_k - k_i \text{diag} \left(\mathbf{H}_k^H \mathbf{F}_{RF} \left(\hat{\mathbf{R}}_{\bar{i}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_i(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \mathbf{F}_{RF}^H \mathbf{H}_k \right) \right), \end{aligned} \quad (4.16a)$$

$$\begin{aligned} \hat{\mathbf{B}}_k = \sum_{l \in \mathcal{D}} w_l & \left(\mathbf{H}_{l,k}^H \left[\hat{\mathbf{R}}_{\bar{l}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_l(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \beta_j \text{diag} \left(\hat{\mathbf{R}}_{\bar{l}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right. \right. \right. \\ & \left. \left. \left. - \hat{\mathbf{R}}_l(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \right] \mathbf{H}_{l,k} - k_k \text{diag} \left(\mathbf{H}_{l,k}^H \left(\hat{\mathbf{R}}_{\bar{l}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_l(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \mathbf{H}_{l,k} \right) \right), \end{aligned} \quad (4.16b)$$

$$\begin{aligned} \hat{\mathbf{C}}_j = \sum_{n \in \mathcal{D}, n \neq j} w_n & \left(\mathbf{H}_n^H \left[\hat{\mathbf{R}}_{\bar{n}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_n(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \beta_n \text{diag} \left(\hat{\mathbf{R}}_{\bar{n}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right. \right. \right. \\ & \left. \left. \left. - \hat{\mathbf{R}}_n(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \right] \mathbf{H}_n - k_0 \text{diag} \left(\mathbf{H}_n^H \left(\hat{\mathbf{R}}_{\bar{n}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_n(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \mathbf{H}_n \right) \right), \end{aligned} \quad (4.16c)$$

$$\begin{aligned} \hat{\mathbf{D}}_j = \sum_{m \in \mathcal{U}} w_m & \left(\mathbf{H}_0^H \mathbf{F}_{RF} \left[\hat{\mathbf{R}}_{\bar{m}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_m(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \beta_0 \text{diag} \left(\hat{\mathbf{R}}_{\bar{m}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_m(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right. \right. \right. \\ & \left. \left. \left. \right) \right] \mathbf{F}_{RF}^H \mathbf{H}_0 - k_0 \text{diag} \left(\mathbf{H}_0^H \mathbf{F}_{RF} \left(\hat{\mathbf{R}}_{\bar{m}}(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} - \hat{\mathbf{R}}_m(\hat{\mathbf{T}}, \hat{\mathbf{Q}})^{-1} \right) \mathbf{F}_{RF}^H \mathbf{H}_0 \right) \right), \end{aligned} \quad (4.16d)$$

$$\begin{aligned} \max_{\substack{\mathbf{U}, \mathbf{V} \\ \mathbf{G}_{RF} \mathbf{F}_{RF}}} & \sum_{k \in \mathcal{U}} \left[w_k \ln \det \left(\mathbf{I} + \mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_k^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \right) - \text{Tr} \left(\mathbf{U}_k^H \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k \right) \mathbf{U}_k \right) \right] + \\ & \sum_{j \in \mathcal{D}} \left[w_j \ln \det \left(\mathbf{I} + \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \right) - \text{Tr} \left(\mathbf{V}_j^H \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j \right) \mathbf{G}_{RF} \mathbf{V}_j \right) \right] \\ \text{s.t.} & \quad (4.12b) - (4.12g) \end{aligned} \quad (4.17)$$

We remark that the tangent expressions (4.15a)-(4.15d) constitute a touching lower bound for WSR_k^{UL} , WSR_j^{DL} , WSR^{DL} and WSR^{UL} , respectively. Hence, the DC programming approach is also a MM approach, regardless of the restatement of the transmit covariance matrices \mathbf{T}_k and \mathbf{Q}_j as a function of the beamformers.

Theorem 3. *The gradients $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{B}}_k$ which linearize WSR_k^{UL} and WSR^{DL} , respectively, with respect to \mathbf{T}_k , $\forall k \in \mathcal{U}$, and the gradients $\hat{\mathbf{C}}_j$ and $\hat{\mathbf{D}}_j$ which linearize WSR_j^{DL} and WSR^{UL} , respectively, with respect to \mathbf{Q}_j , $\forall j \in \mathcal{D}$, with the first order Taylor series expansion are given in (4.16).*

Proof. Please see Appendix B. □

4.3.1 Concave Reformulation

In this section, we simplify the non-concave WSR maximization problem (4.12). By using the gradients (4.16), (4.12) can be reformulated as (4.17), given at the top of the this page.

Lemma 4.3.1. *The WSR maximization problem (4.12) for a single-cell mmWave FD system with multi-antenna users reformulated at each iteration with its first-order Taylor series expansion as in (4.17) is a concave reformulation for each link.*

Proof. The optimization problem (4.12) restated as in (4.17) for each link is made of a concave part, i.e., $\log(\cdot)$, and a linear part, i.e., $\text{Tr}(\cdot)$. Since a linear function is simultaneously concave and non-concave, (4.17) results to be concave for each link. \square

Remark 2: The problem (4.12) and its reformulated version (4.17) have the same KKT conditions and therefore any sub-optimal (optimal) solution of (4.17) is also sub-optimal (optimal) for (4.12).

Let $\Psi_0 = \text{diag}([\psi_1, \dots, \psi_{M_0}])$ and $\Psi_k = \text{diag}([\psi_{k,1}, \dots, \psi_{k,M_k}])$, denote diagonal matrices containing the Lagrange multipliers associated with per-antenna power constraints for the FD BS and UL user k , respectively. Let l_0 and l_1, \dots, l_K denote the Lagrange multipliers associated with the sum-power constraint for FD BS and K UL users, respectively. Let Ψ denote the collection of Lagrange multipliers associated with the per-antenna power constraints, i.e., Ψ_0 and $\Psi_k, \forall k \in \mathcal{U}$. Let \mathbf{L} denote the collection of Lagrange multipliers associated with the sum-power constraints. Augmenting the linearized WSR maximization problem (4.17) with the sum-power and practical per-antenna power constraints, yields the Lagrangian (4.18), stated in the following.

$$\begin{aligned} \mathcal{L}(\mathbf{U}, \mathbf{V}, \mathbf{G}_{RF}, \mathbf{F}_{RF}, \Psi, \mathbf{L}) = & \sum_{l=0}^K l_l \alpha_l + \text{Tr}(\Psi_0 \Lambda_0) + \sum_{u \in \mathcal{U}} \text{Tr}(\Psi_u \Lambda_u) \\ & + \sum_{k \in \mathcal{U}} \left[w_k \ln \det \left(\mathbf{I} + \mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_k^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \right) \right. \\ & \left. - \text{Tr} \left(\mathbf{U}_k^H \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + l_k \mathbf{I} + \Psi_k \right) \mathbf{U}_k \right) \right] \\ & + \sum_{j \in \mathcal{D}} \left[w_j \ln \det \left(\mathbf{I} + \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \right) \right. \\ & \left. - \text{Tr} \left(\mathbf{V}_j^H \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + l_0 \mathbf{I} + \Psi_0 \right) \mathbf{G}_{RF} \mathbf{V}_j \right) \right] \end{aligned} \quad (4.18)$$

In (4.18), unconstrained analog beamformer and combiner are assumed and their constraints will be incorporated later.

4.4 Hybrid Beamforming and Combining

This section presents a novel HYBF design for a multi-user mmWave mMIMO FD system based on alternating optimization. In the following, optimization of the digital

beamformers, analog beamformer and analog combiner is presented into separate sub-sections. We will assume the other variables to be fixed during the alternating optimization process while updating one variable. Information of the other variables updated during previous iterations will be captured in the gradients.

4.4.1 Digital Beamforming

To optimize the digital beamformers, we take the derivative of (4.18) with respect to the conjugate of \mathbf{U}_k and \mathbf{V}_j , which leads to the following KKT conditions

$$\begin{aligned} & \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_k^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \left(\mathbf{I} + \mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_k^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \right)^{-1} \\ & - \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + \boldsymbol{\Psi}_k + l_k \mathbf{I} \right) \mathbf{U}_k = 0, \end{aligned} \quad (4.19a)$$

$$\begin{aligned} & \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \left(\mathbf{I} + \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \right. \\ & \left. \mathbf{V}_j \right)^{-1} - \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \mathbf{G}_{RF} \mathbf{V}_j = 0. \end{aligned} \quad (4.19b)$$

Given (4.19a)-(4.19b), the digital beamformers can be optimized based on the result stated in the following.

Theorem 4. *The digital beamformers \mathbf{U}_k and \mathbf{V}_j , fixed the other variables, can be optimized as the generalized dominant eigenvector solution of the pair of the following matrices*

$$\mathbf{U}_k = \mathbf{D}_{u_k} \left(\mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_k^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k, \hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + \boldsymbol{\Psi}_k + l_k \mathbf{I} \right) \quad (4.20a)$$

$$\mathbf{V}_j = \mathbf{D}_{v_j} \left(\mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF}, \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \mathbf{G}_{RF} \right), \quad (4.20b)$$

where $\mathbf{D}_d(\cdot)$ selects d generalized dominant eigenvectors.

Proof. Please see Appendix B.1. □

The generalized dominant eigenvector solution provides the optimized beamforming directions but not power [38]. To include the optimal stream power allocation, we normalize the columns of digital beamformers to unit-norm. This operation preserves the optimized beamforming directions and allows to design the optimal power allocation scheme.

4.4.2 Analog Beamforming

This section presents a novel approach to design the analog beamformer for hybrid FD BS in a multi-user scenario to maximize the WSR. The structure of the fully connected analog beamformer \mathbf{G}_{RF} is shown in Figure 2. Assuming the remaining variables to be fixed, we first consider the optimization of unconstrained analog beamformer \mathbf{G}_{RF} as

$$\begin{aligned} & \max_{\mathbf{G}_{RF}} \sum_{j \in \mathcal{D}} \left[w_j \ln \det \left(\mathbf{I} + \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \right) \right. \\ & \left. - \text{Tr} \left(\mathbf{V}_j^H \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + l_0 \mathbf{I} + \boldsymbol{\Psi}_0 \right) \mathbf{G}_{RF} \mathbf{V}_j \right) \right]. \end{aligned} \quad (4.21)$$

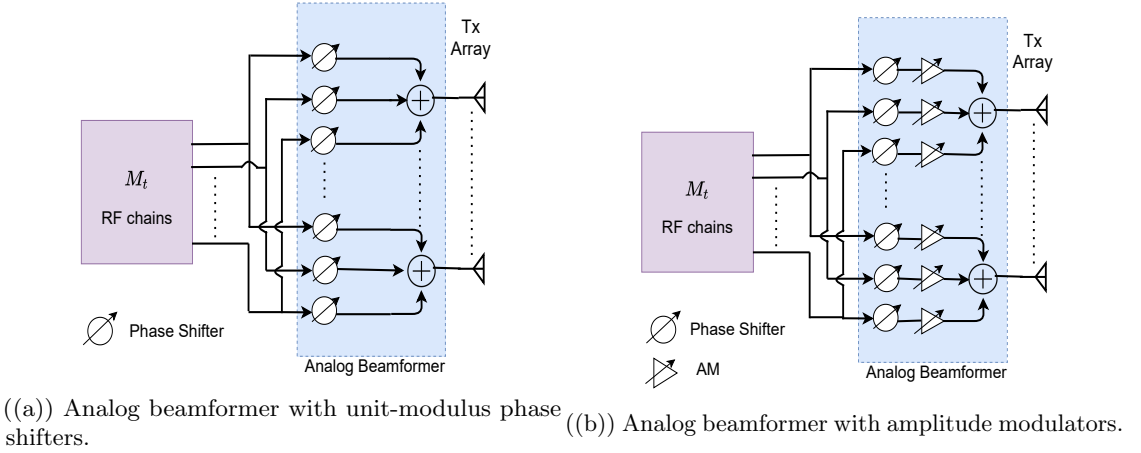


Figure 4.2: (a) All phase shifters are unit-modulus. (b) With amplitude control.

Note that from (4.17) only the terms shown in (4.21) depend on the analog combiner \mathbf{G}_{RF} and information about other variables is captured in gradients $\hat{\mathbf{C}}_j$ and $\hat{\mathbf{D}}_j$. To solve (4.21), we take its derivative with respect to the conjugate of \mathbf{G}_{RF} , which yields the following KKT condition

$$\begin{aligned} & \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \mathbf{V}_j^H \left(\mathbf{I} + \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \right. \\ & \left. \mathbf{G}_{RF} \right)^{-1} - \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \mathbf{G}_{RF} \mathbf{V}_j \mathbf{V}_j^H = 0. \end{aligned} \quad (4.22)$$

Given (4.22), the analog beamformer \mathbf{G}_{RF} for mmWave FD BS can be optimized as stated in the following.

Theorem 5. *The vectorized unconstrained analog beamformer $\text{vec}(\mathbf{G}_{RF})$ can be optimized as one generalized dominant eigenvector solution of the pair of the following matrices*

$$\begin{aligned} \text{vec}(\mathbf{G}_{RF}) = \mathbf{D}_1 \Big(& \sum_{j \in \mathcal{D}} \left(\mathbf{V}_j \mathbf{V}_j^H \left(\mathbf{I} + \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j \mathbf{G}_{RF} \right)^{-1} \right)^T \otimes \mathbf{H}_j^H \mathbf{R}_j^{-1} \mathbf{H}_j, \\ & \sum_{j \in \mathcal{D}} \left(\mathbf{V}_j \mathbf{V}_j^H \right)^T \otimes \left(\hat{\mathbf{c}}_j + \hat{\mathbf{D}}_j + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \Big), \end{aligned} \quad (4.23)$$

where $\mathbf{D}_1(\cdot)$ selects the most dominant generalized eigenvector.

Proof. Please see Appendix B.1. □

Note that Theorem 5 provides the optimized vectorized unconstrained analog beamformer \mathbf{G}_{RF} and we need to reshape it with $\text{unvec}(\text{vec}(\mathbf{G}_{RF}))$. To satisfy the unit-modulus and quantization constraints, we do $\mathbf{G}_{RF}(m, n) = \mathbb{Q}_P(\angle \mathbf{G}_{RF}(m, n))$, $\forall m, n$. For HYBF with AMs, the columns are first scaled to be unit-norm and the quantization constraint is satisfied as $\mathbf{G}_{RF}(m, n) = \mathbb{Q}_A(|\mathbf{G}_{RF}(m, n)|) \mathbb{Q}_P(\angle \mathbf{G}_{RF}(m, n))$, $\forall m, n$.

4.4.3 Analog Combining

This section presents a novel approach to design the analog combiner \mathbf{F}_{RF} for mmWave FD BS to serve multiple UL users. Its design is more straightforward than the analog beamformer. Note that the trace terms appearing in (4.17) have the objective to make beamformers' update aware of the interference generated towards other links. However, \mathbf{F}_{RF} being a combiner, does not generate any interference and therefore does not appear in the trace terms of (4.17). Consequently, to optimize \mathbf{F}_{RF} , we can solve the optimization problem (4.12) instead of using its minorized version (4.17). By considering the unconstrained analog combiner \mathbf{F}_{RF} , from (4.12) we have

$$\max_{\mathbf{F}_{RF}} \sum_{k \in \mathcal{U}} w_k \ln \det(\mathbf{R}_k^{-1} \mathbf{R}_k). \quad (4.24)$$

To solve (4.24), \mathbf{F}_{RF} has to combine the signal received at the antenna level of hybrid FD BS but \mathbf{R}_k and $\mathbf{R}_{\bar{k}}$ represent the received covariance matrices after analog combining. Let (\mathbf{R}_k^{ant}) $\mathbf{R}_{\bar{k}}^{ant}$ denote the (signal-plus) interference and noise covariance matrix received at the antennas of FD BS, which can be obtained from (\mathbf{R}_k) $\mathbf{R}_{\bar{k}}$ given in (4.11) by omitting \mathbf{F}_{RF} . After analog combining, we can recover \mathbf{R}_k and $\mathbf{R}_{\bar{k}}$ as $\mathbf{R}_k = \mathbf{F}_{RF}^H \mathbf{R}_k^{ant} \mathbf{F}_{RF}$ and $\mathbf{R}_{\bar{k}} = \mathbf{F}_{RF}^H \mathbf{R}_{\bar{k}}^{ant} \mathbf{F}_{RF}$, respectively, $\forall k \in \mathcal{U}$. Problem (4.24) can be restated as a function of \mathbf{R}_k^{ant} and $\mathbf{R}_{\bar{k}}^{ant}$ as

$$\max_{\mathbf{F}_{RF}} \sum_{k \in \mathcal{U}} \left[w_k \ln \det(\mathbf{F}_{RF}^H \mathbf{R}_k^{ant} \mathbf{F}_{RF}) - w_k \ln \det(\mathbf{F}_{RF}^H \mathbf{R}_{\bar{k}}^{ant} \mathbf{F}_{RF}) \right]. \quad (4.25)$$

In (4.17), the trace term was only linear, which made the restated optimization problem concave for each link. In (4.25), all the terms are fully concave. To optimize \mathbf{F}_{RF} , we take the derivative with respect to the conjugate of \mathbf{F}_{RF} , which yields the following KKT condition

$$\sum_{k \in \mathcal{U}} w_k \mathbf{R}_k^{ant} \mathbf{F}_{RF} (\mathbf{F}_{RF}^H \mathbf{R}_k^{ant} \mathbf{F}_{RF})^{-1} - \sum_{k \in \mathcal{U}} w_k \mathbf{R}_{\bar{k}}^{ant} \mathbf{F}_{RF} (\mathbf{F}_{RF}^H \mathbf{R}_{\bar{k}}^{ant} \mathbf{F}_{RF})^{-1} = 0. \quad (4.26)$$

It is immediate from (4.26) that the unconstrained analog combiner can be optimized as the generalized dominant eigenvector solution of the pair of sum of the received covariance matrices at the antenna level from all the K UL users, i.e.,

$$\mathbf{F}_{RF} \rightarrow \mathbf{D}_{Nr} \left(\sum_{k \in \mathcal{U}} w_k \mathbf{R}_k^{ant}, \sum_{k \in \mathcal{U}} w_k \mathbf{R}_{\bar{k}}^{ant} \right). \quad (4.27)$$

To satisfy the unit-modulus and quantization constraints for \mathbf{F}_{RF} , we do $\mathbf{F}_{RF}(m, n) = \mathbb{Q}_P(\angle \mathbf{F}_{RF}(m, n)) \in \mathcal{P}$, $\forall m, n$. If AMs are available, the columns are scaled to be unit-norm and quantization constraint is satisfied as

$$\mathbf{F}_{RF}(m, n) = \mathbb{Q}_A(|\mathbf{F}_{RF}(m, n)|) \mathbb{Q}_P(\angle \mathbf{F}_{RF}(m, n)), \forall m, n. \quad (4.28)$$

4.4.4 Optimal Power Allocation

Given the normalized digital beamformers and analog beamformer, optimal power allocation can be included while searching for the Lagrange multipliers satisfying the joint sum-power and practical per-antenna power constraints.

Let $\Sigma_k^{(1)}$ and $\Sigma_k^{(2)}$, $\forall k \in \mathcal{U}$ and $\Sigma_j^{(1)}$ and $\Sigma_j^{(2)}$, $\forall j \in \mathcal{D}$, be defined as

$$\mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_{\bar{k}}^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k = \Sigma_k^{(1)}, \quad (4.29a)$$

$$\mathbf{U}_k^H \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + \boldsymbol{\Psi}_k + l_k \mathbf{I} \right) \mathbf{U}_k = \Sigma_k^{(2)}, \quad (4.29b)$$

$$\mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_{\bar{j}}^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j = \Sigma_j^{(1)}, \quad (4.29c)$$

$$\mathbf{V}_j^H \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \mathbf{G}_{RF} \mathbf{V}_j = \Sigma_j^{(2)}. \quad (4.29d)$$

Given (5.30), the optimal stream power allocation can be included based on the result stated in the following.

Lemma 4.4.1. *Optimal power allocation for the hybrid FD BS and multi-antenna UL users can be obtained by multiplying $\Sigma_j^{(1)}$ and $\Sigma_j^{(2)}$ with the diagonal power matrix \mathbf{P}_j , $\forall j \in \mathcal{D}$ and $\Sigma_k^{(1)}$ and $\Sigma_k^{(2)}$ with the diagonal power matrix \mathbf{P}_k , $\forall k \in \mathcal{U}$, respectively.*

Proof. The beamformers \mathbf{U}_k and \mathbf{V}_k , are computed as the generalized dominant eigenvectors, which make the matrices $\Sigma_k^{(1)}, \Sigma_k^{(2)}, \forall k$ and $\Sigma_j^{(1)}, \Sigma_j^{(2)}, \forall j$ diagonal at each iteration. Multiplying any generalized dominant eigenvector solution matrix with a diagonal matrix still yields a generalized dominant eigenvector solution. Therefore, multiplying $\Sigma_k^{(1)}, \Sigma_k^{(2)}$ with \mathbf{P}_k , $\forall k \in \mathcal{U}$ and $\Sigma_j^{(1)}, \Sigma_j^{(2)}$ with \mathbf{P}_j , $\forall j \in \mathcal{D}$ still preserves the validity of optimized beamforming directions. \square

Given the optimized beamformers and fixed Lagrange multipliers, by using the result stated in Lemma 4.4.1, stream power allocation optimization problems for UL and DL users can be formally stated as

$$\max_{\mathbf{P}_k} \left[w_k \ln \det \left(\mathbf{I} + \Sigma_k^{(1)} \mathbf{P}_k \right) - \text{Tr} \left(\Sigma_k^{(2)} \mathbf{P}_k \right) \right], \quad \forall k \in \mathcal{U}, \quad (4.30a)$$

$$\max_{\mathbf{P}_j} \left[w_j \ln \det \left(\mathbf{I} + \Sigma_j^{(1)} \mathbf{P}_j \right) - \text{Tr} \left(\Sigma_j^{(2)} \mathbf{P}_j \right) \right], \quad \forall j \in \mathcal{D}. \quad (4.30b)$$

Solving (4.30) leads to the following optimal power allocation scheme

$$\mathbf{P}_k = \left(w_k \left(\mathbf{U}_k^H \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + \boldsymbol{\Psi}_k + l_k \mathbf{I} \right) \mathbf{U}_k \right)^{-1} - \left(\mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_{\bar{k}}^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \right)^{-1} \right)^+, \quad (4.31a)$$

$$\mathbf{P}_j = \left(w_j \left(\mathbf{V}_j^H \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \mathbf{G}_{RF} \mathbf{V}_j \right)^{-1} - \left(\mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_{\bar{j}}^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \right)^{-1} \right)^+, \quad (4.31b)$$

where $(\mathbf{X})^+ = \max\{\mathbf{0}, \mathbf{X}\}$. We remark that the proposed power allocation scheme is interference, SI, cross-interference and LDR noise aware as it takes into account their effect in the gradients, which are updated at each iteration. Fixed the beamformers, we can search for multipliers satisfying the joint constraints while doing water-filling for powers. To do so, consider the dependence of Lagrangian (4.18) on multipliers and powers as

$$\begin{aligned} \mathcal{L}(\Psi, \mathbf{L}, \mathbf{P}) &= \sum_{l=0}^K l_l p_l + \text{Tr}(\Psi_0 \Lambda_0) + \sum_{u \in \mathcal{U}} \text{Tr}(\Psi_u \Lambda_u) \\ &+ \sum_{k \in \mathcal{U}} \left[w_k \ln \det(\mathbf{I} + \Sigma_k^{(1)} \mathbf{P}_k) - \text{Tr}(\Sigma_k^{(2)} \mathbf{P}_k) \right] \\ &+ \sum_{j \in \mathcal{D}} \left[w_j \ln \det(\mathbf{I} + \Sigma_j^{(1)} \mathbf{P}_j) - \text{Tr}(\Sigma_j^{(2)} \mathbf{P}_j) \right], \end{aligned} \quad (4.32)$$

where \mathbf{P} is the set of stream powers in UL and DL. The multipliers in Ψ and \mathbf{L} should be such that the Lagrange dual function (5.19) is finite and the values of multipliers should be strictly positive. Formally, Lagrange multipliers' search problem can be stated as

$$\begin{aligned} \min_{\Psi, \mathbf{L}} \max_{\mathbf{P}} \quad & \mathcal{L}(\Psi, \mathbf{L}, \mathbf{P}), \\ \text{s.t.} \quad & \Psi, \mathbf{L} \succeq \mathbf{0}. \end{aligned} \quad (4.33)$$

The dual function $\max_{\mathbf{P}} \mathcal{L}(\Psi, \mathbf{L}, \mathbf{P})$ is the pointwise supremum of a family of functions of Ψ, \mathbf{L} , it is convex [40] and the globally optimal values for Ψ and \mathbf{L} can be obtained by using any of the numerous convex optimization techniques. In this work, we adopt the Bisection algorithm to search the multipliers. Let $\mathcal{M}_0 = \{\lambda_0, \psi_1, \dots, \psi_{M_0}\}$ and $\mathcal{M}_k = \{\lambda_k, \psi_{k,1}, \dots, \psi_{k,M_k}\}$ denote the sets containing Lagrange multipliers associated with the sum-power and practical per-antenna power constraints for FD BS and UL user $k \in \mathcal{U}$, respectively. Let $\underline{\mu}_i$ and $\bar{\mu}_i$ denote the lower and upper bound for the search range of multiplier μ_i , where $\mu_i \in \mathcal{M}_0$ or $\mu_i \in \mathcal{M}_k$. While searching multipliers and performing water-filling for powers, the UL and DL power matrices become non-diagonal. Therefore, we consider the SVD of power matrices to shape them back as diagonal. Namely, let \mathbf{P}_i denote the power matrix for user i , where $i \in \mathcal{U}$ or $i \in \mathcal{D}$. When \mathbf{P}_i becomes non-diagonal, we consider its SVD as

$$[\mathbf{U}_{P_i}, \mathbf{D}_{P_i}, \mathbf{V}_{P_i}] = \text{SVD}(\mathbf{P}_i). \quad (4.34)$$

where \mathbf{U}_{P_i} , \mathbf{D}_{P_i} and \mathbf{V}_{P_i} are the left unitary, diagonal and right unitary matrices, respectively, obtained with the SVD decomposition, and we set $\mathbf{P}_i = \mathbf{D}_{P_i}$ to obtain diagonal power matrices.

For unit-modulus HYBF, the complete alternating optimization based procedure to maximize the WSR based on MM is formally stated in Algorithm 1. For HYBF with AMs, the steps $\angle \mathbf{G}_{RF}$ and $\angle \mathbf{F}_{RF}$ must be omitted and amplitudes of the analog beamformer and combiner must be quantized with $\mathbb{Q}_A(\cdot)$. Once the proposed algorithm converges, all the combiners can be chosen as the MMSE combiners, which will not affect the WSR achieved with Algorithm 1 (4) – (9) [64].

Algorithm 8 Practical Hybrid Beamforming Design

Given: The CSI and rate weights.

Initialize: $\mathbf{G}_{RF}, \mathbf{V}_j, \forall j \in \mathcal{D}$ and $\mathbf{U}_k, \forall k \in \mathcal{U}$.

Set: $\mu_i = 0$ and $\bar{\mu}_i = \mu_{i_{max}} \forall i \in \mathcal{M}_0$ or $\forall i \in \mathcal{M}_k, \forall k \in \mathcal{U}$

Repeat until convergence

 Compute \mathbf{G}_{RF} (4.23), $\text{unvec}(\mathbf{G}_{RF})$ and $\mathbf{G}_{RF} = \angle \mathbf{G}_{RF}$.

 Compute \mathbf{F}_{RF} with (4.27), and do $\mathbf{F}_{RF} = \angle \mathbf{F}_{RF}$.

for: $j = 1 : J$

 Compute $\hat{\mathbf{C}}_j, \hat{\mathbf{D}}_j$ with (4.16)

 Compute \mathbf{V}_j with (4.20b) and normalize it

end

Set: $\mu_0 = 0$ and $\bar{\mu}_0 = \mu_{i_{max}} \forall i \in \mathcal{M}_0$

for: $\forall \mu_0 \in \mathcal{M}_0$

Repeat until convergence

 set $\mu_0 = (\mu_0 + \bar{\mu}_0)/2$

 Compute \mathbf{P}_j with (4.31b) $\forall j$

if constraint for μ_0 is violated

 set $\underline{\mu}_0 = \mu_0,$

else $\bar{\mu}_0 = \mu_0$

$[\mathbf{U}_{P_j}, \mathbf{D}_{P_j}, \mathbf{V}_{P_j}] = \text{SVD}(\mathbf{P}_j), \forall j$

 Set $\mathbf{P}_j = \mathbf{D}_{P_j}$ and $\mathbf{Q}_j = \mathbf{G}_{RF} \mathbf{V}_j \mathbf{P}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H, \forall j$

for: $k = 1 : K$

 Compute $\hat{\mathbf{A}}_k, \hat{\mathbf{B}}_k$ with (4.16)

 Compute \mathbf{U}_k with (4.20a) and normalize it

Set: $\mu_k = 0$ and $\bar{\mu}_k = \mu_{l_{max}}$

for: $\forall \mu_k \in \mathcal{M}_k$

Repeat until convergence

 set $\mu_k = (\mu_k + \bar{\mu}_k)/2$

 Compute \mathbf{P}_k with (4.31a).

if constraint for μ_k is violated

 set $\underline{\mu}_k = \mu_k$

else $\bar{\mu}_k = \mu_k$

$[\mathbf{U}_{P_k}, \mathbf{D}_{P_k}, \mathbf{V}_{P_k}] = \text{SVD}(\mathbf{P}_k)$

 Set $\mathbf{P}_k = \mathbf{D}_{P_k}$ and $\mathbf{T}_k = \mathbf{U}_k \mathbf{P}_k \mathbf{U}_k^H$

Repeat

Quantize $\angle \mathbf{G}_{RF}$ and $\angle \mathbf{F}_{RF}$ ($|\mathbf{G}_{RF}|$ and $|\mathbf{F}_{RF}|$ with AMs)

4.4.5 Convergence

In our context, the ingredients required to prove the convergence are minorization [26], alternating or cyclic optimization [26], Lagrange dual function [40], saddle-point interpretation [40] and KKT conditions [40]. For the WSR cost function (4.12), we construct its minorizer as in (4.15a), (4.15b), (4.15c), (4.15d), which restates the WSR

maximization as a concave problem (4.17) for each link. The minorizer is a touching lower bound for the original WSR problem (4.12), so we can write

$$\text{WSR} \geq \underline{\text{WSR}} = \underline{\text{WR}}_k^{UL} + \underline{\text{WSR}}_k^{UL} + \underline{\text{WR}}_j^{DL} + \underline{\text{WSR}}_j^{DL}. \quad (4.35)$$

The minorizer, which is concave in \mathbf{T}_k and \mathbf{Q}_j , still has the same gradient of the original WSR and hence the KKT conditions are not affected. Reparameterizing \mathbf{T}_k or \mathbf{Q}_j in terms of $\mathbf{U}_k, \forall k \in \mathcal{U}$ and \mathbf{G}_{RF} or $\mathbf{V}_j, \forall j \in \mathcal{D}$, respectively, as in (5.10) with the optimal power matrices and adding the power constraints to the minorizer, we get the Lagrangian (4.18). Every alternating update of \mathcal{L} for $\mathbf{V}_j, \mathbf{G}_{RF}, \mathbf{U}_k, \forall j \in \mathcal{D}, \forall k \in \mathcal{U}$ or for $\mathbf{P}, \mathbf{\Lambda}, \mathbf{\Psi}$ leads to an increase of the WSR, ensuring convergence. For the KKT conditions, at the convergence point, the gradients of \mathcal{L} for $\mathbf{V}_j, \mathbf{G}_{RF}, \mathbf{U}_j$ or \mathbf{P} correspond to the gradients of Lagrangian (4.12), i.e., for the original WSR problem. For fixed analog and digital beamformers, \mathcal{L} is concave in \mathbf{P} , hence we have a strong duality for the saddle point, i.e.

$$\max_{\mathbf{P}} \min_{\mathbf{L}, \mathbf{\Psi}} \mathcal{L}(\mathbf{L}, \mathbf{\Psi}, \mathbf{P}). \quad (4.36)$$

Let \mathbf{X}^* and x^* denote the optimal solution for matrix \mathbf{X} or scalar x at the convergence, respectively. When Algorithm 1 converges, solution of the following optimization problem

$$\min_{\mathbf{\Lambda}, \mathbf{\Psi}} \mathcal{L}(\mathbf{V}^*, \mathbf{G}^*, \mathbf{U}^*, \mathbf{P}^*, \mathbf{L}, \mathbf{\Psi}) \quad (4.37)$$

satisfies the KKT conditions for powers in \mathbf{P} and the complementary slackness conditions

$$l_0^* \left(\alpha_0 - \sum_{j \in \mathcal{D}} \text{Tr} \left(\mathbf{G}_{RF}^* \mathbf{V}_j^* \mathbf{P}_j^* \mathbf{V}_j^{*H} \mathbf{G}_{RF}^{*H} \right) \right) = 0, \quad (4.38a)$$

$$\text{Tr} \left(\mathbf{\Psi}_0^* \left(\mathbf{P}_0 - \sum_{j \in \mathcal{D}} \text{Tr} \left(\mathbf{G}_{RF}^* \mathbf{V}_j^* \mathbf{P}_j^* \mathbf{V}_j^{*H} \mathbf{G}_{RF}^{*H} \right) \right) \right) = 0, \quad (4.38b)$$

$$l_k^* \left(\alpha_k - \text{Tr} \left(\mathbf{U}_k^* \mathbf{P}_k^* \mathbf{U}_k^{*H} \right) \right) = 0, \quad (4.38c)$$

$$\text{Tr} \left(\mathbf{\Psi}_k^* \left(\mathbf{P}_k - \text{Tr} \left(\mathbf{U}_k^* \mathbf{P}_k^* \mathbf{U}_k^{*H} \right) \right) \right) = 0, \quad (4.38d)$$

where all the individual factors in the products are non-negative, and for per-antenna power constraints $\mathbf{\Psi}_0^*$ and $\mathbf{\Psi}_k^*$, the sum of non-negative terms being zero implies all terms result to be zero.

Remark 3: The unit-modulus HYBF scheme converges to a local optimum where $\angle \mathbf{G}_{RF}(m, n), \angle \mathbf{F}_{RF}(m, n) \in \mathcal{P}$ with $|\mathbf{G}_{RF}(m, n)|, |\mathbf{F}_{RF}(m, n)| = 1, \forall m, n$. Unconstrained HYBF with AMs converges to a different local optimum, where $\angle \mathbf{G}_{RF}(m, n), \angle \mathbf{F}_{RF}(m, n) \in \mathcal{P}$ and $|\mathbf{G}_{RF}(m, n)|, |\mathbf{F}_{RF}(m, n)| \in \mathcal{A}, \forall m, n$. Due to quantization, \mathbf{G}_{RF} and \mathbf{F}_{RF} obtained with Algorithm 1 tend to lose their optimality and consequently achieve less WSR compared to their infinite resolution case. For unit-modulus HYBF, the loss in WSR depends only on the resolution of phases. For HYBF with AMs, the loss in WSR depends on the resolution of both the amplitudes and phases.

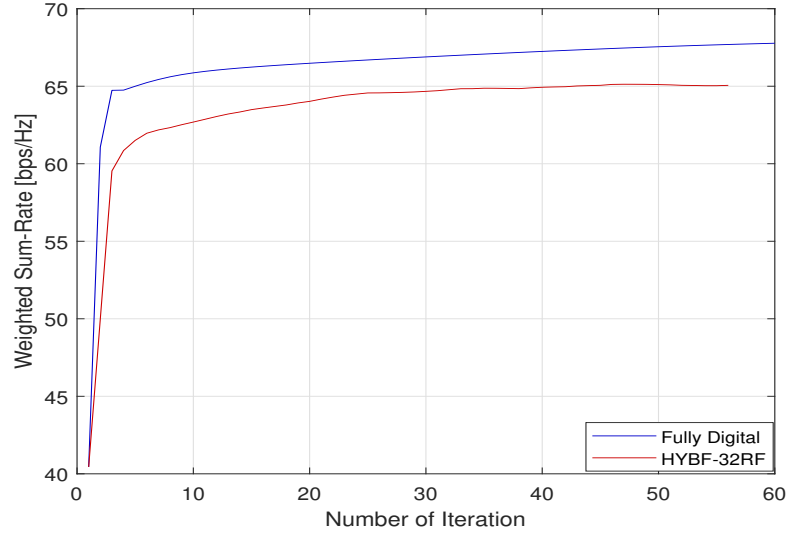


Figure 4.3: Typical convergence behaviour of the proposed HYBF for mmWave mMIMO FD system.

4.4.6 Complexity Analysis

In this section, we analyze the per-iteration computational complexity of Algorithm 1, assuming that the dimensions of antennas get large. Its one iteration consists in updating K and J digital beamformers for the UL and DL user, respectively, and one analog beamformer and combiner for the FD BS. One dominant generalized eigenvector computation to update analog beamformer \mathbf{G}_{RF} from a matrix of size $M_t M_0 \times M_t M_0$ in (4.23), is $\mathcal{O}(M_0^2 M_t^2)$. To update the gradients $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{B}}_k$ for one UL user, the complexity is given by $\mathcal{O}((K-1)N_r^3)$ and $\mathcal{O}(JN_j^3)$, respectively. For the gradient $\hat{\mathbf{C}}_j$ and $\hat{\mathbf{D}}_j$, required to update the beamformer of j -th DL user, computational complexity is $\mathcal{O}((J-1)N_j^3)$ and $\mathcal{O}(KN_r^3)$, respectively. Updating the beamformers of k -th UL and j -th DL users as the generalized dominant eigenvectors adds additional complexity of $\mathcal{O}(u_k M_k^2)$ and $\mathcal{O}(v_j N_j^2)$, respectively. The Lagrange multipliers' update associated with the per-antenna power constraints for FD BS and UL users is linear in the number of antennas M_0 or M_k , respectively. However, as we jointly perform the multipliers' search and power allocation, it adds $\mathcal{O}(v_i^3)$, where $i \in \mathcal{D}$ or $i \in \mathcal{U}$, which can be ignored. Updating the analog combiner \mathbf{F}_{RF} for FD BS is $\mathcal{O}(N_r N_0^2)$. Under the assumption that the dimensions of antennas get large, the per-iteration complexity is

$$\approx \mathcal{O}(K^2 N_r^3 + K J N_j^3 + J^2 N_j^3 + J K N_r^3 + M_0^2 M_t^2 + N_r N_0^2), \quad (4.39)$$

which depends on the size of the antennas and number of UL and DL users served by the mmWave FD BS.

4.5 Simulation Results

This section presents simulation results to evaluate the performance of the proposed HYBF scheme. For comparison, we define the following benchmark schemes:

- a) A *Fully digital HD* scheme with LDR noise, serving the UL and DL users with time division duplexing. Being HD, it is neither affected by the SI nor by the cross-interference.
- b) A *Fully digital FD* scheme with LDR noise. This scheme sets an upper bound for the maximum achievable gain by a hybrid FD system.

Hereafter, HYBF designs with the unit-modulus constraint and with AMs are denoted as HYBF-UM and HYBF-AMs, respectively. We define the signal-to-noise-ratio (SNR) for the mmWave mMIMO FD system as

$$\text{SNR} = \alpha_0 / \sigma_0^2, \quad (4.40)$$

where the scalars α_0 and σ_0^2 denote the total transmit power and thermal noise variance for FD BS, respectively. We set the thermal noise level for DL users to be $\sigma_0^2 = \sigma_j^2, \forall j$, and the transmit power for UL users as $\alpha_0 = \alpha_k, \forall k$. We consider the total transmit power normalized to 1 and choose the noise variance based on desired SNR. To compare the gain of a FD system over a HD system, we define the additional gain in percentage as

$$\text{Gain} = \frac{WSR_{FD} - WSR_{HD}}{WSR_{HD}} \times 100 [\%], \quad (4.41)$$

where WSR_{FD} and WSR_{HD} denote the WSR of a FD and HD system, respectively. To evaluate the performance, we set the per-antenna power constraints for FD BS and UL users as the total transmit power divided by the number of antennas, i.e. $\alpha_0/M_0\mathbf{I}$ and $\alpha_k/M_k\mathbf{I}, \forall k$. The BS and users are assumed to be equipped with a uniform linear array (ULA) with antennas separated by half-wavelength. The transmit and receive antenna array at the BS are assumed to be placed $D = 20$ cm apart, with the relative angle $\Theta = 90^\circ$, and $r_{m,n}$ is modelled as (9) [54]. The Rician factor κ for the SI channel is set to be 1. We assume that the FD BS has $M_0 = 100$ transmit and $N_0 = 50$ receive antennas. It serves two UL and two DL users with $M_k = N_j = 5$ antennas and with 2 data streams for each user. The phases for both designs are quantized in the interval $[0, 2\pi]$ with an 8-bit uniform quantizer $\mathbb{Q}_P(\cdot)$. For HYBF with AMs, the amplitudes are uniformly quantized with a 3-bit uniform quantizer $\mathbb{Q}_A(\cdot)$ in the interval $[0, a_{max}]$, where $a_{max} = \max\{|\max\{\mathbf{G}_{RF}\}|, \max\{|\mathbf{F}_{RF}|\} \}$ is the maximum of the maximum modulus of \mathbf{G}_{RF} or \mathbf{F}_{RF} . We assume the same LDR noise level for the users and FD BS, i.e. $k_0 = \beta_0 = \kappa_k = \beta_j$. The rate weights for the UL and DL users are set to be 1. Aforementioned simulation parameters are summarized in Table 4.2. The digital beamformers are initialized as the dominant eigenvectors of the channel covariance matrices of the intended users. Analog beamformer and combiner are initialized as the dominant eigenvectors of the sum of channel covariance matrices across all the UL and

Table 4.2: Simulation parameters to simulate the multi-user mmWave FD system.

Simulation Parameters		
UL and DL users	K, J	2
Data streams	v_j, u_k	2
Antennas for the BS	M_0, N_0	100, 50
Clusters and Paths	N_c, N_p	3, 3
RF chains (BS)	$M_t = N_r$	8, 10, 16 or 32
User antennas	$M_k = N_j$	5
Rician Factor	κ	1
Tx and Rx array response	$\mathbf{a}_r, \mathbf{a}_t$	ULA, ULA
Angles	$\phi_k, \phi_j, \theta_k, \theta_j$	$\mathcal{U} \sim [-30^\circ, 30^\circ]$
Rate weights	w_k, w_j	1
Uniform Quantizer	$\mathbb{Q}_P(\cdot), \mathbb{Q}_A(\cdot)$	8, 3 bits
Angle between Tx and Rx array (BS)	Θ	90°
Antenna array separation (BS)	D	20 cm
Per-antenna power constraint	$\mathbf{\Lambda}_k, \mathbf{\Lambda}_0$	$\alpha_k/M_k \mathbf{I}, \alpha_0/M_0 \mathbf{I}$

DL users, respectively. Note that as we assume perfect CSI, the SI can be cancelled with HYBF only up to the LDR noise level, which represents the residual SI.

Figure 4.4 shows the achieved average WSR with the proposed HYBF designs as a function of the LDR noise with SNR = 0 dB. The fully digital FD scheme achieves an additional gain of $\sim 97\%$ over a fully digital HD scheme. The impact of different LDR noise levels on the maximum achievable WSR for a mmWave FD system with different number of RF chains is also shown. For $k_0 \leq -40$ dB, HYBF-UM and HYBF-AMs achieve an additional gain of $\sim 85\%, 64\%, 42\%, 3\%$ and $\sim 89\%, 74\%, 60\%, 28\%$ with 32, 16, 10, 8 RF chains, respectively. We can see that as the LDR noise variance increases, achievable WSR for both the hybrid FD and fully digital HD system degrades severely. Figure 4.5 shows the achieved average WSR as a function of the LDR noise with SNR = 40 dB. For $k_0 \leq -80$ dB, HYBF-UM and HYBF-AMs achieve an additional gain of $\sim 65\%, 55\%, 41\%, 15\%$ and $\sim 67\%, 62\%, 55\%, 26\%$ with 32, 16, 10, 8 RF chains, respectively, and increasing the LDR noise variance degrades the achieved average WSR. By comparing Figure 4.4 with Figure 4.5, we can see that at low SNR, HYBF-UM with only 8 RF chains performs close to the fully digital HD scheme. As the SNR increases to 40 dB, HYBF-UM with 8 RF achieves an additional gain of $\sim 15\%$. HYBF-AMs with only 8 RF chains outperforms the fully digital HD scheme for all the SNR levels. Figures 4.4-4.5 also show that HYBF-AMs with 10 RF chains achieves similar average WSR as the HYBF-UM with 16 RF chains. It is interesting to observe that increasing the SNR from 0 dB to 40 dB decreases the thermal noise variance and the LDR noise variance dominates the noise floor already with $k_0 = -80$ dB at SNR = 40 dB. For SNR = 0 dB, the LDR noise variance dominates only for $k_0 > -40$ dB. From this observation, we can conclude that hardware with a low LDR noise is required to benefit from a high SNR in the mmWave FD systems.

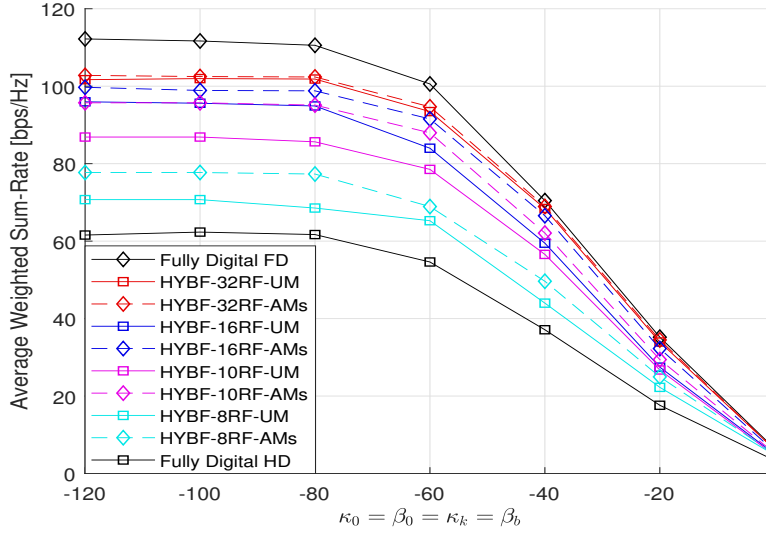


Figure 4.5: Average WSR as a function of the LDR noise with SNR = 40 dB.

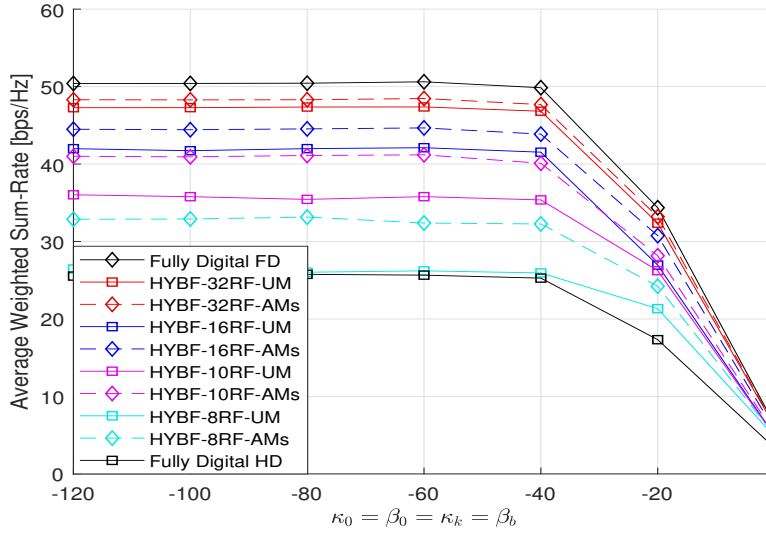


Figure 4.4: Average WSR as a function of the LDR noise with SNR = 0 dB.

Figure 4.6 shows the average WSR with a low LDR noise level $\kappa_0 = -80$ dB with 32, 16, 10 and 8 RF chains as a function of the SNR. Both the proposed designs perform very close to the fully digital FD scheme with 32 RF chains. HYBF-UM and HYBF-AMs outperform the fully digital HD scheme with only 8 RF chains at high SNR and at any SNR level, respectively. It is evident the advantage of AMs, which add additional gain for all the SNR levels when the number RF chains at the FD BS is small. With a high number of RF chains, digital beamforming has enough amplitude manipulation liberty to manage the interference and adding AMs does not bring further improvement. Figure 4.7 shows the average WSR achieved with a moderate LDR noise level $\kappa_0 = -60$ dB. We can

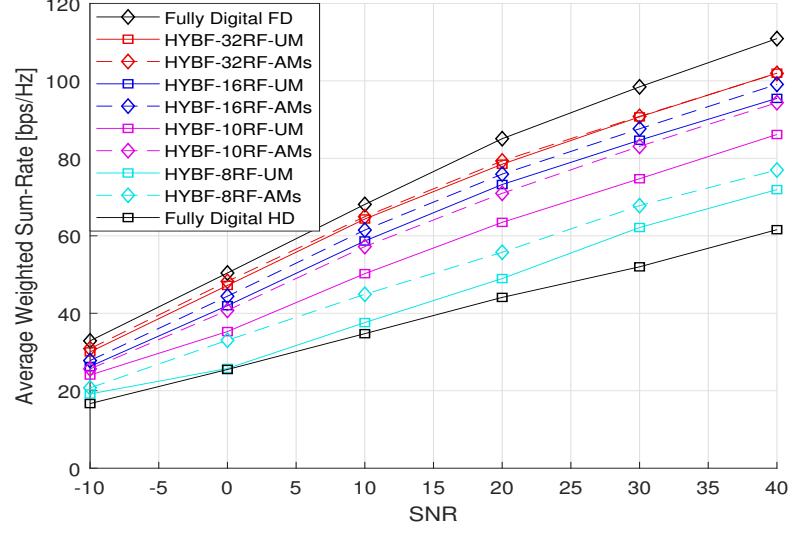


Figure 4.6: Average WSR as a function of the SNR with LDR noise $k_0 = -80$ dB.

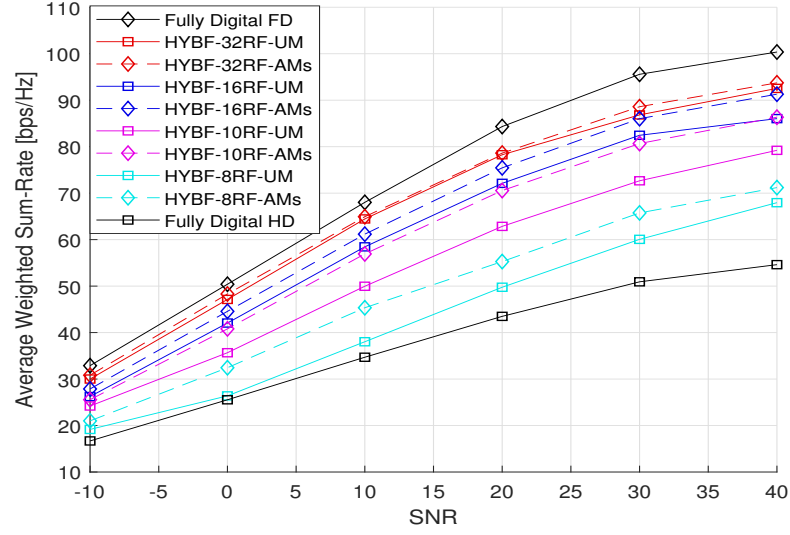
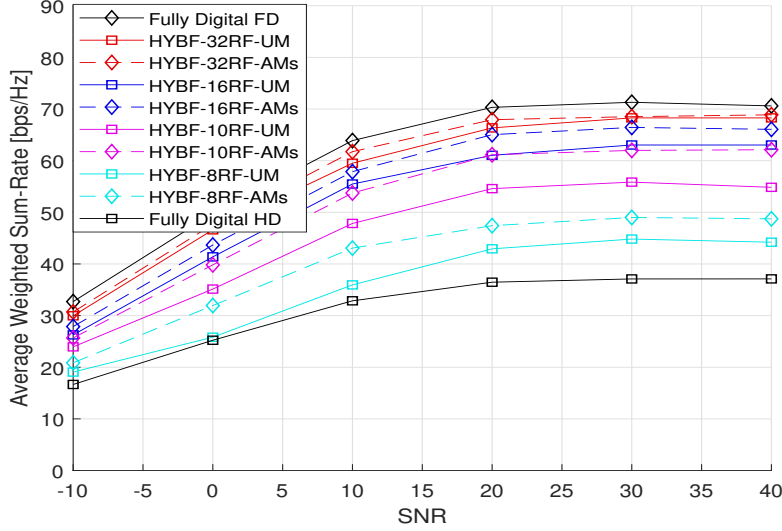


Figure 4.7: Average WSR as a function of the SNR with LDR noise $k_0 = -60$ dB.


 Figure 4.8: Average WSR as function of the SNR with LDR noise $k_0 = -40$ dB.

see that for a low SNR, the achieved average WSR results to be similar as reported in Figure 4.6. At high SNR, the LDR noise variance starts dominating, which leads to less achieved average WSR compared to the case of Figure 4.6. Figure 4.8 shows the achieved WSR as a function of the SNR with a very large LDR noise variance of $\kappa_0 = -40$ dB. By comparing the results reported in Figure 4.8 and Figures 4.6-4.7, we can see that the LDR noise variance dominates for most of the considered SNR range. For a very low SNR, the achieved WSR is similar as reported in Figures 4.6-4.7. However, as the SNR increases, it does not map into higher WSR. It is clear that the maximum achievable WSR with $\kappa_0 = -40$ dB saturates already at SNR= 20 dB for both the HD and FD systems. Further improvement in the SNR does not dictate into higher WSR. When the LDR noise variance dominates, it acts as a ceiling to the effective received-signal-to-LDR-plus-thermal-noise-ratio (RSLTR). The transmit and receive LDR noise variance is proportional to the total transmit power per-antenna and received power per RF chain after the analog combining, respectively. When the LDR noise variance is large, the thermal noise variance has a negligible effect on the effective RSLTR. Consequently, a decrease in the thermal noise variance (increasing SNR) does not dictate a better WSR.

Figure 4.9 shows the achievable performance of HYBF-UM and HYBF-AMs as a function of the RF chains with SNR= 20 dB, in comparison with the benchmark schemes, with very high and very small LDR noise levels. In particular, with very high LDR noise $k_k = -40$ dB and 8 RF chains, HYBF-UM and HYBF-AMs perform close to the fully HD system, and an increase in the number of RF chains improves the performance, which tends towards the achieved WSR by a fully digital FD system with LDR noise level $k_k = -40$ dB. Similar behaviour can be observed for the case of low LDR noise $k_k = -80$ dB. Both the proposed schemes achieve higher WSR with the same number of RF chains in the latter case. We can also see that AMs add additional gain with a low number of RF chains, and as the number of RF chains increase, the gap in the achievable

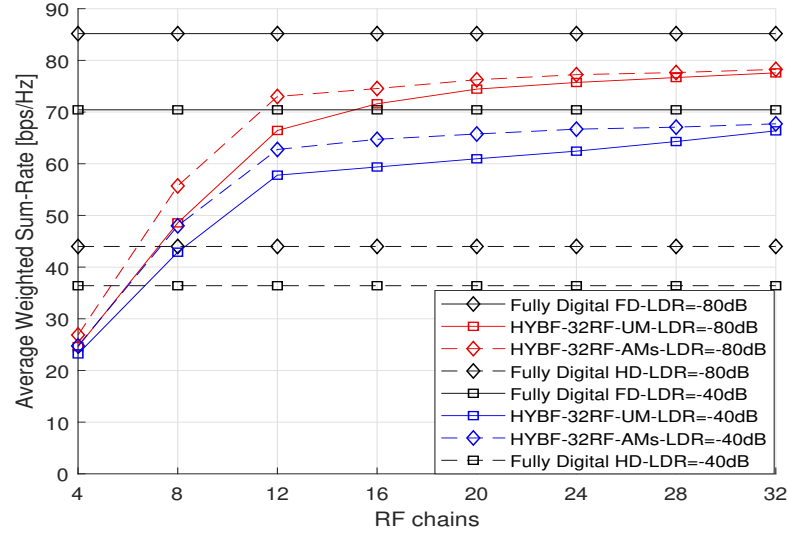


Figure 4.9: Average WSR as a function of the RF chains with LDR noise $k_0 = -80$ dB and $k_0 = -40$ dB at SNR= 20 dB.

WSR with HYBF-AMs and HYBF-UM closes. In particular, with 32 RF chains, the difference in the WSR with or without AMs becomes negligible.

From the results reported in Figures 4.4-4.9, we can conclude that the proposed HYBF schemes achieve significant performance improvement, in terms of average WSR, compared to a fully digital HD system. LDR noise plays a key role in determining the maximum achievable WSR for both the FD and HD systems. Figures 4.4-4.5 shows how an increase in the LDR noise variance degrades the average WSR at low and high SNR levels. Figures 4.6-4.7 shows that with a large to moderate dynamic range, the LDR noise degrades the performance only at very high SNR. Figure 4.8 shows the achieved WSR as a function of a very large LDR noise variance. In that case, it is observed that the WSR saturates at SNR= 20 dB and further improvement in the SNR does not dictate higher WSR. From Figure 4.9, it is clear how the number of RF chains at the mmWave FD BS affects the achievable WSR with different LDR noise levels and with or without the AMs.

4.6 Conclusions

This chapter has presented a novel HYBF design to maximize the WSR in a single-cell mmWave FD system with multi-antenna users and suffering from LDR noise. The beamformers were designed under the joint sum-power and the practical per-antenna power constraints. Simulation results showed that the multi-user mmWave FD systems can outperform the fully digital HD system with only a few RF chains. The advantage of having amplitude control at the analog processing stage is also investigated, and the benefit resulted to be evident with a small number of RF chains. Achievable average WSR with different levels of the LDR noise variance is also investigated, and the proposed HYBF designs outperformed the fully digital HD system at any LDR noise level.

Chapter 5

Hybrid Beamforming for Millimeter Wave Multi-Cell Massive MIMO Full Duplex

5.1 Introduction and Motivation

In wireless communications, the multi-cell FD scenario is the most challenging one as the interference escalates drastically compared to a multi-cell HD system. Besides the interference of an HD system and SI, the DL users in FD also experience UL CI from all the UL users transmitting in-cell, i.e., in the same cell, and out-cell, i.e., in the neighbouring cells. Moreover, the neighbouring BSs generate DL BS-to-BS CI towards the receive antennas of the FD BSs, already affected by the SI and interference. Fig. 5.1 highlights the difference in the interference contributions between the multi-cell HD and FD systems. It is also clear from Fig. 5.1 that the multi-cell FD systems require significant CSI compared to a multi-cell HD system. By shifting the FD paradigm towards the mmWave, the size of direct, SI, interference and CI channel matrices with multi-antenna users becomes massive. To manage interference, C-HYBF schemes will need to transfer a vast amount of CSI to the central node every channel coherence time to optimize the beamformers and combiners, e.g., based on a multi-hop communication if the central node is located far from the FD network. It results in tremendous communication overhead and the central node also requires a very high computational power to optimize many variables jointly, both for the UL and DL users. The optimized variables need to be communicated back to the FD BSs via feedback and each FD BS has to also communicate the optimized beamformers and combiners to its UL and DL users, respectively. For any C-HYBF design, the whole procedure must be executed at the millisecond scale. Given the vast amount of CSI and required computational complexity for FD, it is clearly prohibitive.

Distributed beamforming [83, 84, 85] can eliminate the problem of transferring complete CSI to the central node and reduce computational complexity by decomposing the global optimization problem into per-cell local optimization problems. Distributed

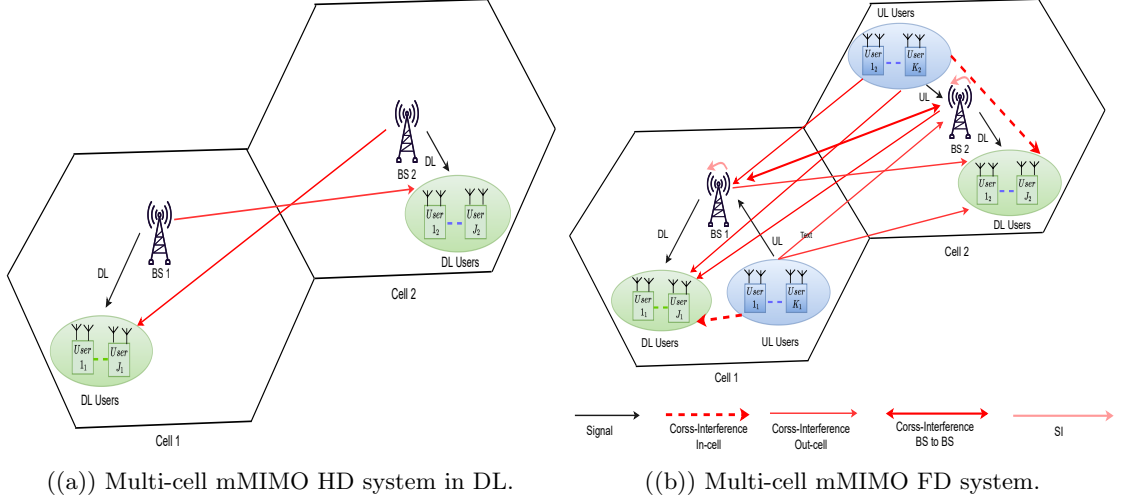


Figure 5.1: Multi-cell mMIMO HD and FD systems with a significant difference in the interference contributions and therefore required amount of CSI.

solutions can be implemented in a non-cooperative fashion [86], or in a cooperative way [87] by exchanging information among the BSs. Per-link parallel and distributed (P&D) [88, 89, 90] beamforming can push further the potential of distributed designs by decomposing the global optimization problem into per-link independent local optimization sub-problems. Therefore, each BS can have the advantage of solving independent sub-problems associated with its communication links separately and simultaneously on different computational processors. Such designs can lead to an exceptionally computationally efficient implementation as the global complexity can be decomposed on multiple processors of each FD BS. If the number of processors available at each BS equals the number of variables to be updated, then all the variables could be optimized in one shot at each iteration, thus escaping the time-consuming alternating optimization. For a detailed discussion about the literature on mmWave FD systems we refer the reader to previous chapter [24, 25, 65, 52, 53, 54, 55, 56, 57, 44, 49, 45, 66, 67, 68, 69, 70, 71, 72]. We remark that the literature does not contain any contribution for the mmWave multi-cell mMIMO FD system with multi-antenna UL and DL users. Moreover, no P&D-HYBF design, neither for FD nor for HD systems, has ever been presented.

5.1.1 Main Contributions

We consider the problem of HYBF for WSR maximization in a multi-cell mMIMO mmWave FD network. The BSs and users are assumed to be operating in the FD and HD mode, respectively, and suffering from the LDR noise due to non-ideal hardware [91]. The FD BSs are assumed to have a massive number of antennas and HYBF capability with fully connected analog beamformers and combiners. Discrete phase-shifters are assumed for the analog stage at the FD BSs. The users are assumed to have a limited number of antennas, fully digital processing capability and are being served with multiple streams by its hybrid FD BS.

Firstly, we present a novel C-HYBF scheme based on alternating optimization, which relies on the mathematical tools offered by the MM optimization technique [26]. However, being centralized, it requires massive computational power to optimize numerous variables jointly. Moreover, complete CSI needs to be transferred to the central node every channel coherence time, and the variables can be optimized only at one computational processor based on alternating optimization. Analysis shows that its complexity scales quadratically as a function of the network size and density, which can limit its scalability. To overcome these drawbacks, we introduce the concept of per-link parallel and distributed HYBF for mmWave and propose a very low-complexity and cooperative P&D-HYBF design for a multi-cell mMIMO FD system. By decomposing the global WSR maximization problem into per-link WSR sub-problems, P&D-HYBF enables each multi-processor FD BS to update multiple variables simultaneously. Analysis shows that its complexity is significantly less than the C-HYBF, which scales only linearly as the network size grows, making it desirable for the next generation of large and dense mmWave FD systems.¹

Simulation results show that the P&D-HYBF scheme achieves similar WSR as the C-HYBF, despite being a distributed design. Both designs converge in a few iterations, which results in a minimal amount of data exchange for the P&D-HYBF. Results are reported with different LDR noise levels, and both designs significantly outperform the fully digital HD system with only a few RF chains at any level.

In summary, the contributions of our work are:

- Introduction of the WSR maximization problem formulation for a multi-cell mMIMO mmWave FD system under LDR.
- Novel C-HYBF scheme for the multi-cell mMIMO mmWave FD system.
- Introduction to the parallel and distributed approach for HYBF in mmWave with the P&D-HYBF design.
- Despite being a distributed design, the P&D-HYBF scheme achieves a similar WSR as the C-HYBF scheme. It possesses all the characteristics to be implemented in a real-time, large and dense multi-cell mMIMO mmWave FD network. It is highly scalable, and converges in a few iterations, thus requiring a minimal communication overhead.
- Simulation results show that both designs outperform the fully digital HD with only a few RF chains at any LDR noise level.

5.2 System Model

Let $\mathcal{B} = \{1, \dots, B\}$ denote the set containing the indices of B FD BSs serving in B cells. Let $\mathcal{D}_b = \{1, \dots, D_b\}$ and $\mathcal{U}_b = \{1, \dots, U_b\}$ denote the sets containing the indices of D_b DL and U_b UL multi-antenna HD users communicating with BS $b \in \mathcal{B}$. The DL user

¹Proposed P&D-HYBF is also applicable to the multi-cell HD or dynamic time division duplexing scenarios [73, 74, 92, 7], which are a special case of the multi-cell FD systems.

$j_b \in \mathcal{D}_b$ and UL user $k_b \in \mathcal{U}_b$ are assumed to have N_{j_b} receive and M_{k_b} transmit antennas, respectively. The BS $b \in \mathcal{B}$ is assumed to have M_b^{RF} and N_b^{RF} transmit and receive RF chains, respectively, and M_b and N_b transmit and receive antennas, respectively. We denote with $\mathbf{V}_{j_b} \in \mathbb{C}^{M_b^{RF} \times d_{j_b}}$ and $\mathbf{U}_{k_b} \in \mathbb{C}^{M_{k_b} \times d_{k_b}}$ the digital beamformers for the white unitary variance data streams $\mathbf{s}_{j_b} \in \mathbb{C}^{d_{j_b} \times 1}$ and $\mathbf{s}_{k_b} \in \mathbb{C}^{d_{k_b} \times 1}$ transmitted for DL user $j_b \in \mathcal{D}_b$ and from UL user $k_b \in \mathcal{U}_b$, respectively. Let $\mathbf{G}_b^{RF} \in \mathbb{C}^{M_b \times M_b^{RF}}$ and $\mathbf{F}_b^{RF} \in \mathbb{C}^{N_b^{RF} \times N_b}$ denote the fully connected analog beamformer and analog combiner for FD BS b , respectively. In practice, as the analog stage can assume only discrete values, let $\mathcal{P}_b = \{1, e^{i2\pi/n_b}, e^{i4\pi/n_b}, \dots, e^{i2\pi n_b - 1/n_b}\}$ denote the set of n_b possible discrete values that the unit-modulus phasors of \mathbf{G}_b^{RF} and \mathbf{F}_b^{RF} can assume. We also define the quantizer function $\mathbb{Q}_b(\cdot)$ to quantize the infinite resolution unit-modulus phasors of $\mathbf{G}_b^{RF}(\mathbf{F}_b^{RF})$ such that $\mathbb{Q}_b(\angle \mathbf{G}_b^{RF})(\mathbb{Q}_b(\angle \mathbf{F}_b^{RF})) \in \mathcal{P}_b$.

The users and BSs are assumed to be suffering from the LDR noise due to non-ideal hardware. The LDR noise for UL user $k_b \in \mathcal{U}_b$ and DL user $j_b \in \mathcal{D}_b$ is denoted as \mathbf{c}_{k_b} and \mathbf{e}_{j_b} , respectively, and modelled as [37]

$$\mathbf{c}_{k_b} \sim \mathcal{CN}(\mathbf{0}_{M_{k_b} \times 1}, k_{k_b} \text{diag}(\mathbf{U}_{k_b} \mathbf{U}_{k_b}^H)), \quad (5.1)$$

$$\mathbf{e}_{j_b} \sim \mathcal{CN}(\mathbf{0}_{N_{j_b} \times 1}, \beta_{j_b} \text{diag}(\mathbf{\Phi}_{j_b})), \quad (5.2)$$

where $k_{k_b} \ll 1$, $\beta_{j_b} \ll 1$, $\mathbf{\Phi}_{j_b} = \text{Cov}(\mathbf{r}_{j_b})$ and \mathbf{r}_{j_b} denotes the undistorted received signal by DL user $j_b \in \mathcal{D}_b$. Let \mathbf{c}_b and \mathbf{e}_b denote the transmit and receive LDR noise for BS b , respectively, which is modelled as [91]

$$\mathbf{c}_b \sim \mathcal{CN}(\mathbf{0}_{M_b \times 1}, k_b \text{diag}(\sum_{n_b \in \mathcal{D}_b} \mathbf{G}_b \mathbf{V}_{n_b} \mathbf{V}_{n_b}^H \mathbf{G}_b^H)), \quad (5.3)$$

$$\mathbf{e}_b \sim \mathcal{CN}(\mathbf{0}_{N_b^{RF} \times 1}, \beta_b \text{diag}(\mathbf{\Phi}_b)), \quad (5.4)$$

where $k_b \ll 1$, $\beta_b \ll 1$, $\mathbf{\Phi}_b = \text{Cov}(\mathbf{r}_b)$ and \mathbf{r}_b denotes the undistorted received signal by BS b after the analog combiner \mathbf{F}_b^{RF} . The thermal noise for BS b and DL user j_b is denoted as $\mathbf{n}_b \sim \mathcal{CN}(\mathbf{0}_{N_b \times 1}, \sigma_b^2 \mathbf{I})$ and $\mathbf{n}_{j_b} \sim \mathcal{CN}(\mathbf{0}_{N_{j_b} \times 1}, \sigma_{j_b}^2 \mathbf{I})$, respectively.

5.2.1 Channel Modelling

Let $\mathbf{H}_{j_b} \in \mathbb{C}^{N_{j_b} \times M_b}$ and $\mathbf{H}_{k_b} \in \mathbb{C}^{N_b \times M_{k_b}}$ denote the direct channels² from BS $b \in \mathcal{B}$ to DL user $j_b \in \mathcal{D}_b$ and from UL user $k_b \in \mathcal{U}_b$ to BS b , respectively. Let $\mathbf{H}_{j_b, k_b} \in \mathbb{C}^{N_{j_b} \times M_{k_b}}$ and $\mathbf{H}_{j_b, k_c} \in \mathbb{C}^{N_{j_b} \times M_{k_c}}$ denote the in-cell UL CI channel for DL user $j_b \in \mathcal{D}_b$ from UL user $k_b \in \mathcal{U}_b$ and the out-cell UL CI channel for DL user $j_b \in \mathcal{D}_b$ from UL user $k_c \in \mathcal{U}_c$, with $b \neq c$, respectively. Let $\mathbf{H}_{j_b, c} \in \mathbb{C}^{N_{j_b} \times M_c}$ and $\mathbf{H}_{b, k_c} \in \mathbb{C}^{N_b \times M_{k_c}}$ denote the interference channels from BS $c \in \mathcal{B}$ to DL user $j_b \in \mathcal{D}_b$ and from UL user $k_c \in \mathcal{U}_c$ to BS b , respectively, with $c \neq b$. Let $\mathbf{H}_{b, c} \in \mathbb{C}^{N_b \times M_c}$ denote the DL BS-to-BS CI channel from BS $c \in \mathcal{B}$ to BS

²We assume perfect CSI, which can be obtained as in [81] and it is a part of the ongoing research for mmWave FD [93].

$b \in \mathcal{B}$. The SI channel for BS b is denoted as $\mathbf{H}_{b,b} \in \mathbb{C}^{N_b \times M_b}$. In the mmWave frequency band, the direct channel \mathbf{H}_{k_b} for UL user $k_b \in \mathcal{U}_b$ can be modelled as [91]

$$\mathbf{H}_{k_b} = \sqrt{\frac{M_{k_b} N_b}{N_{c,b} N_{p,b}}} \sum_{n_{c,b}=1}^{N_{c,b}} \sum_{n_{p,b}=1}^{N_{p,b}} \alpha_{k_b}^{(n_{p,b}, n_{c,b})} \mathbf{a}_{r,b}(\phi_{k_b}^{n_{p,b}, n_{c,b}}) \mathbf{a}_{t,k_b}^T(\theta_{k_b}^{n_{p,b}, n_{c,b}}), \quad (5.5)$$

where $N_{c,b}$ and $N_{p,b}$ denote the number of clusters and number of paths for BS b (Fig. 1 [54]), respectively, and $\alpha_{k_b}^{(n_{p,b}, n_{c,b})} \sim \mathcal{CN}(0, 1)$ is a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively. The vectors $\mathbf{a}_{r,b}(\phi_{k_b}^{n_{p,b}, n_{c,b}})$ and $\mathbf{a}_{t,k_b}(\theta_{k_b}^{n_{p,b}, n_{c,b}})$ denote the receive and transmit antenna array response for BS b and UL user k_b , respectively, with the angle of arrival (AoA) $\phi_{k_b}^{n_{p,b}, n_{c,b}}$ and angle of departure (AoD) $\theta_{k_b}^{n_{p,b}, n_{c,b}}$. The channels \mathbf{H}_{j_b} , \mathbf{H}_{j_b, k_b} , \mathbf{H}_{j_b, k_c} , $\mathbf{H}_{j_b, c}$ and \mathbf{H}_{b, k_c} can be modelled similarly as (5.5). The SI channel $\mathbf{H}_{b,b} \in \mathbb{C}^{N_b \times M_b}$ can be modelled as [91]

$$\mathbf{H}_b = \sqrt{\frac{\kappa_b}{\kappa_b + 1}} \mathbf{H}_b^L + \sqrt{\frac{1}{\kappa_b + 1}} \mathbf{H}_b^R. \quad (5.6)$$

The matrices \mathbf{H}_b^L and \mathbf{H}_b^R denote the line-of-sight (LoS) and reflected components of the SI channel response, respectively, and the scalar κ_b denotes the Rician factor. The channel matrix \mathbf{H}_b^R can be modelled as (5.5) and element of \mathbf{H}_b^L at the m -th row and n -th column can be modelled as [91]

$$\mathbf{H}_b^L(m, n) = \frac{\rho_b}{r_{m,n}} e^{-j2\pi \frac{r_{m,n}}{\lambda}} \quad (5.7)$$

where ρ_b is the power normalization constant to assure $\mathbb{E}(\|\mathbf{H}_b^L(m, n)\|_F^2) = M_b N_b$, and the scalars $r_{m,n}$ and λ denote distance and wavelength, respectively. A summary of the aforementioned notations can be found Table 5.1.

5.2.2 Problem Formulation

Let \mathbf{y}_{j_b} and \mathbf{y}_{k_b} denote the received signal by DL user $j_b \in \mathcal{D}_b$ and from UL user k_b at the FD BS b after the analog combiner \mathbf{F}_b^{RF} , respectively. By using the aforementioned notations, they can be written as

$$\begin{aligned} \mathbf{y}_{j_b} = & \mathbf{H}_{j_b} \left(\sum_{n_b \in \mathcal{D}_b} \mathbf{G}_b^{RF} \mathbf{V}_{n_b} \mathbf{s}_{n_b} + \mathbf{c}_b \right) + \mathbf{e}_{j_b} + \mathbf{n}_{j_b} + \underbrace{\sum_{k_b \in \mathcal{U}_b} \mathbf{H}_{j_b, k_b} (\mathbf{U}_{k_b} \mathbf{s}_{k_b} + \mathbf{c}_{k_b})}_{\text{UL Interference in cell } b} \\ & + \underbrace{\sum_{c \in \mathcal{B}, c \neq b} \mathbf{H}_{j_b, c} \left(\sum_{n_c \in \mathcal{D}_c} \mathbf{G}_c^{RF} \mathbf{V}_{n_c} \mathbf{s}_{n_c} + \mathbf{c}_c \right)}_{\text{DL Interference from neighbouring BSs}} + \underbrace{\sum_{c \in \mathcal{B}, c \neq b} \sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{j_b, k_c} (\mathbf{U}_{k_c} \mathbf{s}_{k_c} + \mathbf{c}_{k_c})}_{\text{UL Interference from neighbouring cells}}, \end{aligned} \quad (5.8)$$

Table 5.1: Notations

M_b^{RF}	# of transmit RF chains for BS b
N_b^{RF}	# of receive RF chains for BS b
M_b	# of transmit antennas for BS b
N_b	# of receive antennas for BS b
M_{k_b}	# of transmit antennas for UL user k_b
N_{j_b}	# of receive antennas for DL user j_b
\mathbf{U}_{k_b}	Digital beamformer for UL user k_b
\mathbf{V}_{j_b}	Digital beamformer for DL user j_b
\mathbf{G}_b^{RF}	Analog beamformer for BS b
\mathbf{F}_b^{RF}	Analog combiner for BS b
\mathbf{c}_{k_b}	Transmit LDR noise from UL user k_b
\mathbf{c}_b	Transmit LDR noise from BS b
\mathbf{e}_b	Receive LDR noise for BS b
\mathbf{e}_j	Receive LDR noise for DL user j_b
\mathbf{n}_b	Thermal noise for BS b
\mathbf{n}_{j_b}	Thermal noise for DL user j_b
\mathbf{H}_b	SI channel for BS b
\mathbf{H}_{k_b}	Channel between BS b and UL user k_b
\mathbf{H}_{j_b}	Channel between BS b and DL user j_b
\mathbf{H}_{j_b, k_b}	CI channel between DL user j_b and user k_b .
\mathbf{H}_{j_b, k_c}	CI channel between DL user j_b and user k_c .
$\mathbf{H}_{b, c}$	CI channel between BS b and BS c .
$\mathbf{H}_{j_b, c}$	Interference channel between DL user j_b and BS c .

$$\begin{aligned}
\mathbf{y}_{k_b} = & \mathbf{F}_b^{RFH} \left(\sum_{k_b \in \mathcal{U}_b} \mathbf{H}_{k_b} (\mathbf{U}_{k_b} \mathbf{s}_{k_b} + \mathbf{c}_{k_b}) + \mathbf{n}_b + \underbrace{\mathbf{H}_{b,b} \left(\sum_{j_b \in \mathcal{D}_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{s}_{j_b} + \mathbf{c}_b \right)}_{\text{SI}} \right) \\
& + \underbrace{\sum_{c \in \mathcal{B}, c \neq b} \mathbf{H}_{b,c} \left(\sum_{j_c \in \mathcal{D}_c} \mathbf{G}_c^{RF} \mathbf{V}_{j_c} \mathbf{s}_{j_c} + \mathbf{c}_c \right)}_{\text{BS to BS interference}} + \underbrace{\sum_{c \in \mathcal{B}, c \neq b} \sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{b,k_c} (\mathbf{U}_{k_c} \mathbf{s}_{k_c} + \mathbf{c}_{k_c})}_{\text{UL Interference from neighbouring cells}} + \mathbf{e}_b.
\end{aligned} \tag{5.9}$$

Let $\overline{k_b}$ and $\overline{j_b}$ denote the indices in sets \mathcal{U}_b and \mathcal{D}_b without the elements k_b and j_b , respectively. Let \overline{b} denote the indices in set \mathcal{B} except the element b . Let \mathbf{T}_{k_b} and \mathbf{Q}_{j_b} denote the transmit covariance matrices transmitted from UL user $k_b \in \mathcal{U}_b$ and for DL user $j_b \in \mathcal{D}_b$ by the BS $b \in \mathcal{B}$, respectively, defined as

$$\mathbf{T}_{k_b} = \mathbf{U}_{k_b} \mathbf{U}_{k_b}^H, \quad \forall k_b \in \mathcal{U}_b, \tag{5.10a}$$

$$\mathbf{Q}_{j_b} = \mathbf{G}_b \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^H, \quad \forall j_b \in \mathcal{D}_b. \tag{5.10b}$$

The received (signal plus) interference and noise covariance matrices by BS $b \in \mathcal{B}$ from the UL user $k_b \in \mathcal{U}_b$ and at the DL user $j_b \in \mathcal{D}_b$ are denoted as (\mathbf{R}_{k_b}) $\mathbf{R}_{\overline{k_b}}$ and (\mathbf{R}_{j_b})

$$\begin{aligned}
\mathbf{R}_{j_b} = & \underbrace{\mathbf{H}_{j_b} \mathbf{Q}_{j_b} \mathbf{H}_{j_b}^H}_{\triangleq \mathbf{S}_{j_b}} + \mathbf{H}_{j_b} \left(\sum_{\substack{n_b \in \mathcal{D}_b \\ n_b \neq j_b}} \mathbf{Q}_{n_b} \right) \mathbf{H}_{j_b}^H + \mathbf{H}_{j_b} k_b \text{diag} \left(\sum_{n_b \in \mathcal{D}_b} \mathbf{Q}_{n_b} \right) \mathbf{H}_{j_b}^H + \sum_{k_b \in \mathcal{U}_b} \mathbf{H}_{j_b, k_b} (\mathbf{T}_{k_b} \\
& + k_{k_b} \text{diag}(\mathbf{T}_{k_b})) \mathbf{H}_{j_b, k_b}^H + \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \mathbf{H}_{j_b, c} \left(\sum_{n_c \in \mathcal{D}_c} \mathbf{Q}_{n_c} + k_c \text{diag}(\mathbf{Q}_{n_c}) \right) \mathbf{H}_{j_b, c}^H \\
& + \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{j_b, k_c} (\mathbf{T}_{k_c} + k_{k_c} \text{diag}(\mathbf{T}_{k_c})) \mathbf{H}_{j_b, k_c}^H + \beta_{j_b} \text{diag}(\Phi_{j_b}) + \sigma_{j_b}^2 \mathbf{I}_{N_{j_b}},
\end{aligned} \tag{5.11a}$$

$$\begin{aligned}
\mathbf{R}_{k_b} = & \mathbf{F}_b^{RFH} \left(\underbrace{\mathbf{H}_{k_b} \mathbf{T}_{k_b} \mathbf{H}_{k_b}^H}_{\triangleq \mathbf{S}_{k_b}} + \sum_{\substack{m_b \in \mathcal{U}_b \\ m_b \neq k_b}} \mathbf{H}_{m_b} \mathbf{T}_{m_b} \mathbf{H}_{m_b}^H + \sum_{m_b \in \mathcal{U}_b} k_{m_b} \mathbf{H}_{m_b} \text{diag}(\mathbf{T}_{m_b}) \mathbf{H}_{m_b}^H + \mathbf{H}_{b, b} \right. \\
& \left. \sum_{j_b \in \mathcal{D}_b} (\mathbf{Q}_{j_b} + k_b \text{diag}(\mathbf{Q}_{j_b})) \mathbf{H}_{b, j_b}^H + \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \mathbf{H}_{b, c} \sum_{j_c \in \mathcal{D}_c} (\mathbf{Q}_{j_c} + k_c \text{diag}(\mathbf{Q}_{j_c})) \mathbf{H}_{b, c}^H \right. \\
& \left. + \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{k_c \in \mathcal{U}_c} \mathbf{H}_{b, k_c} (\mathbf{T}_{k_c} + k_{k_c} \text{diag}(\mathbf{T}_{k_c})) \mathbf{H}_{b, k_c}^H + \sigma_b^2 \mathbf{I}_{n_b} \right) \mathbf{F}_b^{RF} + \beta_b \text{diag}(\Phi_b),
\end{aligned} \tag{5.11b}$$

$$\mathbf{R}_{j_b}^- = \mathbf{R}_{j_b} - \mathbf{S}_{j_b}, \quad \mathbf{R}_{k_b}^- = \mathbf{R}_{k_b} - \mathbf{S}_{k_b}. \tag{5.11c}$$

$\mathbf{R}_{j_b}^-$, respectively, and can be written as (5.11a)-(5.11b), given at the top of the next page. The matrices \mathbf{S}_{j_b} and \mathbf{S}_{k_b} in (5.11a)-(5.11b) denote the useful signal covariance matrices received by the DL user $j_b \in \mathcal{D}_b$ and FD BS $b \in \mathcal{B}$ from the UL user $k_b \in \mathcal{U}_b$, respectively. The undistorted received covariance matrices Φ_{j_b} (5.2) and Φ_b (5.4) can be recovered from (5.11a)-(5.11c) without the receive LDR noise, i.e., with $\beta_{j_b} = 0$ and $\beta_b = 0$, respectively.

The WSR maximization problem with HYBF for a multi-cell mMIMO mmWave FD system, serving multi-antenna $J_b \in \mathcal{D}_b$ DL and $U_b \in \mathcal{U}_b$ UL users, $\forall b \in \mathcal{B}$, under the joint sum-power, unit-modulus and discrete phase-shifters constraints can be stated as

$$\max_{\substack{\mathbf{U}, \mathbf{V}, \\ \mathbf{G}^{RF}, \mathbf{F}^{RF}}} \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} w_{k_b} \ln \det(\mathbf{R}_{k_b}^{-1} \mathbf{R}_{k_b}) + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} w_{j_b} \ln \det(\mathbf{R}_{j_b}^{-1} \mathbf{R}_{j_b}) \tag{5.12a}$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{U}_{k_b} \mathbf{U}_{k_b}^H) \leq p_{k_b}, \quad \forall k_b \in \mathcal{U}_b, \tag{5.12b}$$

$$\text{Tr} \left(\sum_{j \in \mathcal{D}_b} \mathbf{G}_b \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_b^H \right) \leq p_b, \quad \forall b \in \mathcal{B}, \tag{5.12c}$$

$$\mathbf{G}_b^{RF}(m, n) \in \mathcal{P}_b, \quad \forall m, n \text{ \& } \forall b \in \mathcal{B}, \tag{5.12d}$$

$$\mathbf{F}_b^{RF}(i, j) \in \mathcal{P}_b, \quad \forall i, j \text{ \& } \forall b \in \mathcal{B}. \tag{5.12e}$$

The scalars w_{k_b} and w_{j_b} denote rate weights for UL user k_b and DL user j_b , respectively, and the scalars p_{k_b} and p_b denote sum-power constraint for UL user $k_b \in \mathcal{U}_b$ and BS $b \in \mathcal{B}$, respectively. The collections of digital beamformers in UL and DL are denoted as \mathbf{U} and \mathbf{V} , respectively, and the collections of analog beamformers and combiners are denoted as \mathbf{G}^{RF} and \mathbf{F}^{RF} , respectively.

Remark 1: The WSR achieved with (5.12) is not affected with the minimum-mean-squared-error (MMSE) combiners (4) – (9) [64]. Therefore, they can be omitted during the optimization process and can be chosen as the MMSE receivers after solving (5.12). The number of digital combiners would be equal to the number of total users in the multi-cell FD network. By omitting them, the HYBF design simplifies and the per-iteration computational complexity reduces significantly. Moreover, it will significantly reduce the amount of data exchange required among the FD BSs by the P&D-HYBF.

5.3 Minorization-Maximization

Optimization problem (5.12) is non-concave in the transmit covariance matrices $\mathbf{T}_{k_b}, \forall k_b \in \mathcal{U}_b$, and $\mathbf{Q}_{j_b} \in \mathcal{D}_b$, due to the interference generated towards other communication links and finding its global optimum is very challenging. To find its sub-optimal solution, this section presents the MM optimization method [26] to leverage alternating optimization.

The WSR problem (5.12) for the multi-cell mmWave FD system can be reformulated with its minorizer [26] using the DC programming [94, 38]. The WSR in (5.12) can be expressed with the weighted rate (WR) of user $k_b \in \mathcal{U}_b$ ($\text{WR}_{k_b}^{UL}$), user $j_b \in \mathcal{D}_b$ ($\text{WR}_{j_b}^{DL}$), WSR of users \bar{k}_b ($\text{WSR}_{\bar{k}_b}^{UL}$) and \bar{j}_b ($\text{WSR}_{\bar{j}_b}^{DL}$), and WSR of all the UL and DL users in cells different than b , denoted as WSR_b^{UL} and WSR_b^{DL} , respectively. We can express the global network WSR as

$$\text{WSR} = \underbrace{\text{WR}_{k_b}^{UL} + \text{WSR}_{\bar{k}_b}^{UL}}_{\triangleq \text{WSR}_b^{UL}} + \underbrace{\text{WR}_{j_b}^{DL} + \text{WSR}_{\bar{j}_b}^{DL}}_{\triangleq \text{WSR}_b^{DL}} + \text{WSR}_b^{UL} + \text{WSR}_b^{DL}, \quad (5.13)$$

where WSR_b^{UL} and WSR_b^{DL} denote the UL and DL WSR for FD BS b . Considering the dependence on the transmit covariance matrices, only $\text{WSR}_{k_b}^{UL}$ is concave in \mathbf{T}_{k_b} and the remaining terms $\text{WSR}_{\bar{k}_b}^{UL}$, WSR_b^{DL} , $\text{WSR}_{\bar{j}_b}^{DL}$ and WSR_b^{UL} are non concave in \mathbf{T}_{k_b} . Similarly, only $\text{WSR}_{j_b}^{DL}$ is concave in \mathbf{Q}_{j_b} and $\text{WSR}_{\bar{j}_b}^{DL}$, WSR_b^{UL} , $\text{WSR}_{\bar{k}_b}^{UL}$, WSR_b^{DL} are non concave in \mathbf{Q}_{j_b} . As a linear function is simultaneously convex and concave, DC programming introduces the first order Taylor series expansion of $\text{WSR}_{k_b}^{UL}$, WSR_b^{DL} , $\text{WSR}_{\bar{k}_b}^{UL}$ and $\text{WSR}_{\bar{j}_b}^{DL}$ in \mathbf{T}_{k_b} , around $\hat{\mathbf{T}}_{k_b}$ (i.e. around all \mathbf{T}_{k_b}), and for $\text{WSR}_{j_b}^{DL}$, WSR_b^{UL} , $\text{WSR}_{\bar{j}_b}^{UL}$, WSR_b^{DL} around $\hat{\mathbf{Q}}_{j_b}$ (i.e. around all \mathbf{Q}_{j_b}). Let $\hat{\mathbf{T}}$ and $\hat{\mathbf{Q}}$ denote the sets containing all such $\hat{\mathbf{T}}_{k_b}$ and $\hat{\mathbf{Q}}_{j_b}$, respectively. The tangent expressions by computing the gradients with respect to the transmit covariance matrix \mathbf{T}_{k_b} , i.e.,

$$\hat{\mathbf{G}}_{k_b,b}^{UL} = -\frac{\partial \text{WSR}_{k_b}^{UL}}{\partial \mathbf{T}_{k_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad \hat{\mathbf{G}}_{k_b,b}^{DL} = -\frac{\partial \text{WSR}_b^{DL}}{\partial \mathbf{T}_{k_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad (5.14a)$$

$$\begin{aligned}
& \max_{\mathbf{U}, \mathbf{V}, \mathbf{G}^{RF}, \mathbf{F}^{RF}} \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \mathbf{U}_{k_b}) \right. \\
& \quad \left. - \text{Tr}(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{k_b, b}^{UL} + \hat{\mathbf{G}}_{k_b, b}^{DL} + \hat{\mathbf{G}}_{k_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b, \bar{b}}^{DL}) \mathbf{U}_{k_b}) \right] \\
& \quad + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right. \\
& \quad \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b, b}^{UL} + \hat{\mathbf{G}}_{j_b, b}^{DL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right]. \\
& \quad \text{s.t. (5.12b) - (5.12e)}
\end{aligned} \tag{5.18a}$$

$$\tag{5.18b}$$

$$\hat{\mathbf{G}}_{k_b, \bar{b}}^{UL} = - \frac{\partial \text{WSR}_b^{UL}}{\partial \mathbf{T}_{k_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad \hat{\mathbf{G}}_{k_b, \bar{b}}^{DL} = - \frac{\partial \text{WSR}_b^{DL}}{\partial \mathbf{T}_{k_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \tag{5.14b}$$

allow to write the following minorizers

$$\underline{\text{WSR}}_{k_b}^{UL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \text{WSR}_{k_b}^{UL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b}) \hat{\mathbf{G}}_{k_b, b}^{UL}), \tag{5.15a}$$

$$\underline{\text{WSR}}_b^{DL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \underline{\text{WSR}}_b^{DL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b}) \hat{\mathbf{G}}_{k_b, b}^{DL}), \tag{5.15b}$$

$$\underline{\text{WSR}}_{\bar{b}}^{UL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \text{WSR}_{\bar{b}}^{UL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b}) \hat{\mathbf{G}}_{k_b, \bar{b}}^{UL}), \tag{5.15c}$$

$$\underline{\text{WSR}}_{\bar{b}}^{DL}(\mathbf{T}_{k_b}, \hat{\mathbf{T}}, \hat{\mathbf{Q}}) = \text{WSR}_{\bar{b}}^{DL}(\hat{\mathbf{T}}, \hat{\mathbf{Q}}) - \text{Tr}((\mathbf{T}_{k_b} - \hat{\mathbf{T}}_{k_b}) \hat{\mathbf{G}}_{k_b, \bar{b}}^{DL}). \tag{5.15d}$$

Similarly, for the transmit covariance matrix \mathbf{Q}_{j_b} , we have the gradients

$$\hat{\mathbf{G}}_{j_b, b}^{UL} = - \frac{\partial \text{WSR}_b^{UL}}{\partial \mathbf{Q}_{j_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad \hat{\mathbf{G}}_{j_b, b}^{DL} = - \frac{\partial \text{WSR}_{j_b}^{DL}}{\partial \mathbf{Q}_{j_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \tag{5.16a}$$

$$\hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} = - \frac{\partial \text{WSR}_{\bar{b}}^{UL}}{\partial \mathbf{Q}_{j_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \quad \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL} = - \frac{\partial \text{WSR}_{\bar{b}}^{DL}}{\partial \mathbf{Q}_{j_b}} \Big|_{\hat{\mathbf{T}}, \hat{\mathbf{Q}}}, \tag{5.16b}$$

which allow to write the minorizers

$$\underline{\text{WSR}}_b^{UL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \text{WSR}_b^{UL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \text{Tr}((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b}) \hat{\mathbf{G}}_{j_b, b}^{UL}), \tag{5.17a}$$

$$\underline{\text{WSR}}_{j_b}^{DL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \text{WSR}_{j_b}^{DL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \text{Tr}((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b}) \hat{\mathbf{G}}_{j_b, b}^{DL}), \tag{5.17b}$$

$$\underline{\text{WSR}}_{\bar{b}}^{UL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \text{WSR}_{\bar{b}}^{UL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \text{Tr}((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b}) \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL}), \tag{5.17c}$$

$$\underline{\text{WSR}}_{\bar{b}}^{DL}(\mathbf{Q}_{j_b}, \hat{\mathbf{Q}}, \hat{\mathbf{T}}) = \text{WSR}_{\bar{b}}^{DL}(\hat{\mathbf{Q}}, \hat{\mathbf{T}}) - \text{Tr}((\mathbf{Q}_{j_b} - \hat{\mathbf{Q}}_{j_b}) \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL}). \tag{5.17d}$$

The gradients (5.14) for the UL user k_b and (5.16) for the DL user j_b are provided in Appendix C.1.

The tangent expressions (5.15) and (5.17) constitute a touching lower bound for the original WSR. Hence, the DC programming approach is also a MM approach, regardless

$$\begin{aligned}
\mathcal{L} = & \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \mathbf{U}_{k_b}) \right. \\
& \left. - \text{Tr}(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b}) \right] \\
& + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right. \\
& \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right] \\
& + \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \lambda_{k_b} p_{k_b} + \sum_{b \in \mathcal{B}} \psi_b p_b.
\end{aligned} \tag{5.19}$$

of the restatement of the transmit covariance matrices \mathbf{T}_{k_b} and \mathbf{Q}_{j_b} as a function of the beamformers. By using the gradients (5.14) and (5.16), the WSR maximization problem with respect to the beamformers and combiners associated with each communication link can be written as (5.18a), given at the top of this page. The KKT conditions of (5.18) and (5.12) are the same, hence any sub-optimal or optimal solution for (5.18) is also sub-optimal or optimal for (5.12).

Let λ_{k_b} and ψ_b denote the Lagrange multipliers associated with the sum-power constraint for UL user $k_b \in \mathcal{U}_b$ and BS $b \in \mathcal{B}$, respectively. Augmenting the WSR function (5.18a) with the sum-power constraints yield the Lagrangian (5.19). Note that (5.19) does not consider the quantization constraints on the analog beamformers and combiners, which will be incorporated later.

5.4 Centralized Hybrid Beamforming

This section presents a novel C-HYBF design based on alternating optimization by exploiting the tools of MM to solve (5.19) to a local optimum. Hereafter, for the sake of a simplified explanation, we dedicate different sub-sections to optimize different variables. However, while optimizing one variable, the remaining ones are assumed to be fixed and their complete information is saved in the gradients, which will be updated at each iteration.

5.4.1 Digital Beamforming

To optimize the digital beamformers \mathbf{U}_{k_b} and \mathbf{V}_{j_b} , we consider the remaining variables to be fixed. By taking the derivatives of (5.19) with respect to the conjugate of \mathbf{U}_{k_b} and \mathbf{V}_{j_b} , yield the following KKT conditions

$$\begin{aligned}
& \mathbf{H}_{k_b}^H \mathbf{F}_b^{RFH} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RF} \mathbf{H}_{k_b} \mathbf{U}_{k_b} (\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \\
& \mathbf{U}_{k_b})^{-1} - (\hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b} = 0,
\end{aligned} \tag{5.20a}$$

$$\begin{aligned} & \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} (\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^H \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \\ & \mathbf{V}_{j_b})^{-1} - \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} = 0. \end{aligned} \quad (5.20b)$$

Given the structure of the KKT conditions (5.20a)-(5.20b), the digital beamformers can be optimized based on the result provided in the following Theorem.

Theorem 6. *The WSR maximizing digital beamformers \mathbf{U}_{k_b} and \mathbf{V}_{j_b} , fixed the remaining variables, can be obtained as the generalized dominant eigenvector solution of the pair of the following matrices*

$$\mathbf{U}_{k_b} = \mathbf{D}_{d_{k_b}} (\mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b}, \hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL} + \lambda_{k_b} \mathbf{I}), \quad (5.21)$$

$$\mathbf{V}_{j_b} = \mathbf{D}_{d_{j_b}} (\mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF}, \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF}). \quad (5.22)$$

The matrix \mathbf{D}_x selects x dominant generalized eigenvectors, equal to the number of data streams.

Proof. Please see Appendix C.2. □

Digital beamformers, given as the generalized dominant eigenvectors, provide the optimized beamforming directions but not the optimal power allocation. Thus, we scale the columns of the digital beamformers to unit-norm, which will allow designing the optimal power allocation scheme.

5.4.2 Analog Beamforming

To design the analog beamformer \mathbf{G}_b^{RF} for BS $b \in \mathcal{B}$, we assume the remaining variables fixed. By considering only the dependence of the WSR on unconstrained analog beamformer \mathbf{G}_b^{RF} , simplifies (5.18a) into

$$\begin{aligned} & \max_{\mathbf{G}_b^{RF}} \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \ln \det (\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right. \\ & \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right]. \end{aligned} \quad (5.23)$$

To optimize \mathbf{G}_b^{RF} , we take the derivative of (5.23) with respect to the conjugate of \mathbf{G}_b^{RF} , which leads to the following KKT condition

$$\begin{aligned} & \sum_{j_b \in \mathcal{D}} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H (\mathbf{I} + \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF})^{-1} \\ & - (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H = 0. \end{aligned} \quad (5.24)$$

Given the KKT condition (5.24), \mathbf{G}_b^{RF} can be optimized based on the result stated in the following.

Theorem 7. *The vectorized unconstrained analog beamformer \mathbf{G}_b^{RF} , common to all the DL users in set \mathcal{D}_b , can be optimized as one generalized dominant eigenvector solution of the pair of the sum of following matrices*

$$\begin{aligned} \text{vec}(\mathbf{G}_b^{RF}) = \mathbf{D}_1 & \left(\sum_{j_b \in \mathcal{D}_b} (\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H (\mathbf{I} + \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \right. \\ & \left. \mathbf{H}_{j_b} \mathbf{G}_b^{RF})^{-1})^T \otimes \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b}, \sum_{j_b \in \mathcal{D}_b} (\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H)^T \right. \\ & \left. \otimes (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \right), \end{aligned} \quad (5.25)$$

Proof. The proof is provided in Appendix C.3. \square

The result stated in Theorem 7 optimizes the vectorized unconstrained analog beamformer. Operation $\text{unvec}(\text{vec}(\mathbf{G}_b^{RF}))$ is required to reshape it into correct dimensions. To meet the unit-modulus and quantization constraints, we do $\mathbf{G}_b^{RF} = \mathbb{Q}_b(\angle \mathbf{G}_b^{RF}) \in \mathcal{P}_b$.

5.4.3 Analog Combining

Optimization of the analog combiner \mathbf{F}_b^{RF} is more straightforward than the analog beamformer as it does not generate any interference. Note that \mathbf{F}_b^{RF} does not appear in the trace operator of (5.18a). Combiners at the neighbouring FD BSs \bar{b} appear only in the gradients as we take into account the interference generated after the analog combining stage. However, they are fixed while optimizing \mathbf{F}_b^{RF} for BS b during the alternating optimization process.

Given that the analog combiner \mathbf{F}_b^{RF} does not generate any interference, the WSR is purely concave with respect to \mathbf{F}_b^{RF} in the received covariance matrices \mathbf{R}_{k_b} and $\mathbf{R}_{\bar{k}_b}$ for $b \in \mathcal{B}$. Therefore, the original WSR maximization problem (5.12) can be considered to optimize \mathbf{F}_b^{RF} . By considering the dependence of the unconstrained \mathbf{F}_b^{RF} on the WSR, we have the following optimization problem

$$\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} \text{ldet}(\mathbf{R}_{k_b}^{-1} \mathbf{R}_{k_b}). \quad (5.26)$$

The analog combiner \mathbf{F}_b^{RF} has to combine the received covariance matrices at the antenna level. Let $(\mathbf{R}_{k_b}^a) \mathbf{R}_{k_b}^a$ denote the (signal plus) interference and noise covariance matrices received at the antennas of BS $b \in \mathcal{B}$ to be combined with \mathbf{F}_b^{RF} . Given $\mathbf{R}_{k_b}^a$ and $\mathbf{R}_{\bar{k}_b}^a$ at the antenna level, \mathbf{R}_{k_b} and $\mathbf{R}_{\bar{k}_b}$ can be recovered as $\mathbf{R}_{k_b} = \mathbf{F}_b^{RFH} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF}$ and $\mathbf{R}_{\bar{k}_b} = \mathbf{F}_b^{RFH} \mathbf{R}_{\bar{k}_b}^a \mathbf{F}_b^{RF}$. The analog combiner \mathbf{F}_b^{RF} , by writing (5.26) as a function of $\mathbf{R}_{k_b}^a$ and $\mathbf{R}_{\bar{k}_b}^a$ and using the properties of $\log(\cdot)$ function, can be optimized by solving

$$\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} [w_{k_b} \text{ldet}(\mathbf{F}_b^{RFH} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF}) - w_{k_b} \text{ldet}(\mathbf{F}_b^{RFH} \mathbf{R}_{\bar{k}_b}^a \mathbf{F}_b^{RF})]. \quad (5.27)$$

Taking the derivative of (5.27) with respect to \mathbf{F}_b^{RF} leads to the following KKT condition

$$\sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF} (\mathbf{F}_b^{RFH} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF})^{-1} - \sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF} (\mathbf{F}_b^{RFH} \mathbf{R}_{k_b}^a \mathbf{F}_b^{RF})^{-1} = 0. \quad (5.28)$$

From (5.28), it is immediate to see that the WSR maximizing analog combiner \mathbf{F}_b^{RF} can be obtained as the generalized dominant eigenvector solution of the pair of the sum of the received covariance matrices at the antenna level from the UL users in the same cell, i.e.

$$\mathbf{F}_b^{RF} = \mathbf{D}_{N_b^{RF}} \left(\sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a, \sum_{k_b \in \mathcal{U}_b} w_{k_b} \mathbf{R}_{k_b}^a \right). \quad (5.29)$$

The matrix $\mathbf{D}_{N_b^{RF}}$ selects generalized dominant eigenvectors equal to the number of receive RF chains N_b^{RF} at the BS $b \in \mathcal{B}$. As the analog combiner given in (5.29) is unconstrained, we normalize the amplitudes and perform the quantization such that $\mathbf{F}_b^{RF} = \mathbb{Q}_b(\angle \mathbf{F}_b^{RF}) \in \mathcal{P}_b$.

5.4.4 Optimal Power Allocation

This section presents a novel and optimal power allocation scheme for the FD BSs and UL users operating in a mmWave multi-cell FD scenario, given the digital beamformers with unit-norm columns. Let $\mathbf{\Sigma}_{k_b}^{(1)}$, $\mathbf{\Sigma}_{k_b}^{(2)}$, $\mathbf{\Sigma}_{j_b}^{(1)}$ and $\mathbf{\Sigma}_{j_b}^{(2)}$, be defined as

$$\mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \mathbf{U}_{k_b} = \mathbf{\Sigma}_{k_b}^{(1)}, \quad (5.30a)$$

$$\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b} = \mathbf{\Sigma}_{k_b}^{(2)}, \quad (5.30b)$$

$$\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} = \mathbf{\Sigma}_{j_b}^{(1)}, \quad (5.30c)$$

$$\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} = \mathbf{\Sigma}_{j_b}^{(2)}. \quad (5.30d)$$

Let \mathbf{P}_{k_b} and \mathbf{P}_{j_b} denote the power matrices for UL user $k_b \in \mathcal{U}_b$ and DL user $j_b \in \mathcal{D}_b$, respectively. Given the optimized beamformers and fixed Lagrange multipliers, the optimal stream power allocation can be obtained by multiplying (5.30a) and (5.30b) with \mathbf{P}_{k_b} , $\forall k_b$ and (5.30c) and (5.30d) with \mathbf{P}_{j_b} , $\forall j_b$. As the beamformers are given by the generalized dominant eigenvector solution, they diagonalize the matrices $\mathbf{\Sigma}_{k_b}^{(1)}$, $\mathbf{\Sigma}_{k_b}^{(2)}$, $\mathbf{\Sigma}_{j_b}^{(1)}$ and $\mathbf{\Sigma}_{j_b}^{(2)}$. Multiplying (5.30a) and (5.30b) with a diagonal matrix \mathbf{P}_{k_b} or (5.30c) and (5.30d) with a diagonal matrix \mathbf{P}_{j_b} , still yields the generalized eigenvector solution, and thus the optimized beamforming directions are not affected [91].

The stream power allocation optimization problem for UL user k_b and DL user j_b can be formally stated as

$$\max_{\mathbf{P}_{k_b}} \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{\Sigma}_{k_b}^{(1)} \mathbf{P}_{k_b}) - \text{Tr}(\mathbf{\Sigma}_{k_b}^{(2)} \mathbf{P}_{k_b}) \right], \quad (5.31)$$

$$\max_{\mathbf{P}_{j_b}} \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{\Sigma}_{j_b}^{(1)} \mathbf{P}_{j_b}) - \text{Tr}(\mathbf{\Sigma}_{j_b}^{(2)} \mathbf{P}_{j_b}) \right]. \quad (5.32)$$

Solving (5.31) and (5.32) leads to the following optimal power allocation scheme

$$\mathbf{P}_{k_b} = (w_{k_b} (\mathbf{U}_k^H (\hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b})^{-1} - (\mathbf{U}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{H}_{k_b}^H \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RF} \mathbf{H}_{k_b} \mathbf{U}_{k_b})^{-1})^+, \quad (5.33a)$$

$$\mathbf{P}_{j_b} = (w_{j_b} (\mathbf{V}_{j_b}^H \mathbf{G}_b^H (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b \mathbf{V}_{j_b})^{-1} - (\mathbf{V}_{j_b}^H \mathbf{G}_b^H \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b \mathbf{V}_{j_b})^{-1})^+, \quad (5.33b)$$

where $(\mathbf{X})^+ = \max\{\mathbf{0}, \mathbf{X}\}$. Given the optimal stream powers, we can search for the Lagrange multipliers satisfying the total power budget constraint while performing interference, SI, and CI leakage aware water-filling for the powers with (5.33).

Let \mathbf{P}^{DL} and \mathbf{P}^{UL} denote the collection of stream powers in DL and UL, respectively. We define $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ as the collection of multipliers for λ_{k_b} and ψ_b , respectively. Given the optimal stream powers computed with (5.33), consider the dependence of the Lagrangian on the multipliers and powers as

$$\begin{aligned} \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{DL}, \mathbf{P}^{UL}) &= \sum_{b \in \mathcal{B}} \psi_b p_b + \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \lambda_{k_b} p_{k_b} \\ &+ \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \left[w_{k_b} \ln \det(\mathbf{I} + \Sigma_{k_b}^{(1)} \mathbf{P}_{k_b}) - \text{Tr}(\Sigma_{k_b}^{(2)} \mathbf{P}_{k_b}) \right] \\ &+ \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \ln \det(\mathbf{I} + \Sigma_{j_b}^{(1)} \mathbf{P}_{j_b}) - \text{Tr}(\Sigma_{j_b}^{(2)} \mathbf{P}_{j_b}) \right], \end{aligned} \quad (5.34)$$

The multipliers in $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ should be such that the Lagrangian (5.34) is finite and the values of multipliers are strictly positive. The multipliers' search problem can be formally stated as

$$\begin{aligned} \min_{\mathbf{\Psi}, \mathbf{\Lambda}} \max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \quad & \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{DL}, \mathbf{P}^{UL}), \\ \text{s.t.} \quad & \mathbf{\Psi}, \mathbf{\Lambda} \succeq \mathbf{0}. \end{aligned} \quad (5.35)$$

The dual function

$$\max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{DL}, \mathbf{P}^{UL}) \quad (5.36)$$

is the pointwise supremum of a family of functions of $\mathbf{\Psi}, \mathbf{\Lambda}$, it is convex [40] and the globally optimal values for $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ can be found by using any of the numerous convex-optimization techniques. In this work, we adopt the Bisection algorithm to search the multipliers. Let $\underline{\psi}_b, \overline{\psi}_b$ and $\underline{\lambda}_{k_b}, \overline{\lambda}_{k_b}$ denote the upper and lower bound for searching the multipliers ψ_b and λ_{k_b} , respectively. Let $[0, \lambda_{k_b}^{max}]$ and $[0, \psi_b^{max}]$ denote the search range for the multipliers λ_{k_b} and ψ_b , respectively, where $\lambda_{k_b}^{max}$ and ψ_b^{max} denote the maximum values that λ_{k_b} and ψ_b can assume. As the generalized dominant eigenvector solution is computed given the fixed multipliers, doing water-filling for the powers while searching for the multipliers leads to non diagonal power matrices (5.33). Hence, consider a singular value decomposition (SVD) of the power matrices as

$$[\mathbf{L}_{k_b}^{svd}, \mathbf{D}_{k_b}^{svd}, \mathbf{R}_{k_b}^{svd}] = \mathbf{P}_{k_b}, \quad \forall k_b \in \mathcal{U}_b, \quad (5.37a)$$

$$[\mathbf{L}_b^{svd}, \mathbf{D}_b^{svd}, \mathbf{R}_b^{svd}] = \mathbf{P}_{j_b}, \quad \forall j_b \in \mathcal{D}_b, \quad (5.37b)$$

where \mathbf{L}_x^{svd} , \mathbf{D}_x^{svd} and \mathbf{R}_x^{svd} denote the left unitary, diagonal and right unitary matrices obtained via SVD. Given (5.37), the diagonal power matrices can be obtained again as

$$\mathbf{P}_{k_b} = \mathbf{D}_{k_b}, \forall k_b \in \mathcal{U}_b, \quad \mathbf{P}_{j_b} = \mathbf{D}_{j_b}, \forall j_b \in \mathcal{U}_b, \quad (5.38)$$

while searching for the multipliers satisfying the sum-power constraints.

The complete C-HYBF procedure to maximize the WSR in a multi-cell mMIMO mmWave FD system based on the alternating optimization process by using MM is formally stated in Algorithm 9. Once it converges, the combiners can be chosen as the MMSE receivers, which will not affect the WSR achieved after convergence.

5.4.5 Convergence of C-HYBF

The convergence of Algorithm 9 can be proved by using the minorization theory [26], alternating or cyclic optimization [26], Lagrange dual function [40], saddle-point interpretation [40] and KKT conditions [40]. For the WSR cost function (5.12), we construct its minorizer as in (5.15) for the WSR in UL and with (5.17) for the WSR in DL, $\forall b \in \mathcal{B}$. It constructs a touching lower bound for (5.12), hence we can write

$$\text{WSR} \geq \underline{\text{WSR}} = \underline{\text{WR}}_{k_b,b}^{UL} + \underline{\text{WR}}_{k_b,b}^{UL} + \underline{\text{WR}}_{j_b,b}^{DL} + \underline{\text{WR}}_{j_b,b}^{DL} + \underline{\text{WSR}}_b^{DL} + \underline{\text{WSR}}_b^{UL}. \quad (5.39)$$

The minorized WSR, which is concave in \mathbf{T}_{k_b} and \mathbf{Q}_{j_b} , has the same gradient of the original WSR maximization problem (5.12), hence the KKT conditions are not affected. Reparameterizing \mathbf{T}_{k_b} or \mathbf{Q}_{j_b} in terms of \mathbf{G}_b^{RF} , \mathbf{V}_{j_b} , $\forall j_b \in \mathcal{D}_b$ or \mathbf{U}_{k_b} , $\forall k_b \in \mathcal{U}_b$, respectively, including the optimal stream power matrices and augmenting the WSR cost function with the Lagrange multipliers and power constraints leads to (5.34). Alternating update of the Lagrangian \mathcal{L} for the variables \mathbf{V}_{j_b} , \mathbf{G}_b^{RF} , \mathbf{U}_{k_b} , $\forall j_b \in \mathcal{D}_b$, $\forall k_b \in \mathcal{U}_b$, \mathbf{P}_{k_b} , \mathbf{P}_{j_b} , $\mathbf{\Lambda}$, $\mathbf{\Psi}$ leads to a monotonic increase of the WSR, which assures convergence. For the KKT conditions, at the convergence point, the gradients of \mathcal{L} for \mathbf{V}_{j_b} , \mathbf{G}_b^{RF} , \mathbf{U}_{k_b} or \mathbf{P}_{k_b} , \mathbf{P}_{j_b} correspond to the gradients of the Lagrangian of original WSR maximization problem (5.12). For the fixed analog and digital beamformers, \mathcal{L} is concave in powers, hence we have strong duality for the saddle point, i.e.,

$$\max_{\mathbf{P}^{DL}, \mathbf{P}^{UL}} \min_{\mathbf{\Lambda}, \mathbf{\Psi}} \mathcal{L}(\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{P}^{UL}, \mathbf{P}^{DL}). \quad (5.40)$$

Let \mathbf{X}^* and x^* denote the optimal solution for matrix \mathbf{X} or scalar x at the convergence, respectively. As each iteration leads to a monotonic increase in the WSR and the power are updated by satisfying the sum-power constraint, at the convergence point, the solution of the optimization problem

$$\min_{\mathbf{\Lambda}, \mathbf{\Psi}} \mathcal{L}(\mathbf{V}_{j_b}^*, \mathbf{G}_b^{RF*}, \mathbf{F}_b^{RF*}, \mathbf{U}_b^{RF*}, \mathbf{P}^{DL*}, \mathbf{P}^{UL*}, \mathbf{\Lambda}, \mathbf{\Psi}) \quad (5.41)$$

Algorithm 9 Centralized Hybrid Beamforming

Given: The CSI and rate weights.

Initialize: $\mathbf{G}_b^{RF}, \mathbf{F}_b^{RF}, \mathbf{V}_{j_b}, \mathbf{U}_{k_b}, \forall j_b \& \forall k_b$.

Set: $\underline{\lambda}_{k_b} = 0, \overline{\lambda}_{k_b} = \lambda_{k_b}^{max}, \underline{\psi}_b = 0, \overline{\psi}_b = \psi_b^{max}, \forall k_b \& \forall b$

Repeat until convergence

for $b = 1 : B$

 Compute \mathbf{G}_b^{RF} with (5.25), do unvec(\mathbf{G}_b^{RF}) and get $\angle \mathbf{G}_t$

for: $j_b = 1 : D_b$

 Compute $\hat{\mathbf{G}}_{j_b,b}^{UL}, \hat{\mathbf{G}}_{j_b,b}^{DL}, \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL}, \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL}$ (5.16)

 Compute \mathbf{V}_{j_b} with (5.22) and normalize it

 Next j_b

Repeat until convergence

 set $\psi_b = (\underline{\psi}_b + \overline{\psi}_b)/2$

for $j_b = 1 : D_b$

 Compute \mathbf{P}_{j_b} with (5.33b)

$[\mathbf{U}_{\mathbf{P}_{j_b}}, \mathbf{D}_{\mathbf{P}_{j_b}}, \mathbf{V}_{\mathbf{P}_{j_b}}] = SVD(\mathbf{P}_{j_b})$

 Set $\mathbf{P}_{j_b} = \mathbf{D}_{\mathbf{P}_{j_b}}$

 Set $\mathbf{Q}_{j_b} = \mathbf{G}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^H$

 Next j_b

if constraint for ψ_b is violated

 set $\underline{\psi}_b = \psi_b$,

else $\overline{\psi}_b = \psi_b$

for: $k_b = 1 : K_b$

 Compute $\hat{\mathbf{G}}_{k_b,b}^{UL}, \hat{\mathbf{G}}_{k_b,b}^{DL}, \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL}, \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL}$ (5.14)

 Compute \mathbf{U}_{k_b} with (5.21) and normalize it

Repeat until convergence

 set $\lambda_{k_b} = (\underline{\lambda}_{k_b} + \overline{\lambda}_{k_b})/2$

 Compute $\overline{\mathbf{P}}_{k_b}$ with (5.33a).

$[\mathbf{U}_{\mathbf{P}_{k_b}}, \mathbf{D}_{\mathbf{P}_{k_b}}, \mathbf{V}_{\mathbf{P}_{k_b}}] = SVD(\mathbf{P}_{k_b})$

 Set $\mathbf{P}_{k_b} = \mathbf{D}_{\mathbf{P}_{k_b}}$

 Set $\mathbf{T}_b = \mathbf{U}_{k_b} \mathbf{P}_{k_b} \mathbf{U}_{k_b}^H$.

if constraint for λ_{k_b} is violated

 set $\underline{\lambda}_{k_b} = \lambda_{k_b}$

else $\overline{\lambda}_{k_b} = \lambda_{k_b}$

 Next k_b

 Next b

Quantize: \mathbf{G}_b^{RF} and \mathbf{F}_b^{RF} , with $\mathbb{Q}_b(\cdot), \forall b$

satisfies the KKT conditions for the powers in \mathbf{P}^{DL} and \mathbf{P}^{UL} and the complementary slackness conditions

$$\psi_b^* \left(p_b - \sum_{j_b \in \mathcal{D}_b} \text{Tr}(\mathbf{G}_b^{RF*} \mathbf{V}_{j_b}^* \mathbf{P}_{j_b}^* \mathbf{V}_{j_b}^{*H} \mathbf{G}_b^{RF*H}) \right) = 0, \quad (5.42a)$$

$$\lambda_{k_b}^* (p_{k_b} - \text{Tr}(\mathbf{U}_{k_b}^* \mathbf{P}_{k_b}^* \mathbf{U}_{k_b}^{*H})) = 0, \quad (5.42b)$$

where all the individual factors in the products are non-negative.

5.5 Parallel and Distributed Implementation

Algorithm 1 requires enormous communication overhead to transfer full CSI to the central node every channel coherence time and very high computational power to update all the variables jointly. As later shown in Section 5.5.4, it also scales quadratically as a function of the number of users and cells, which limits its scalability in a real-time large and dense multi-cell FD network. To overcome these drawbacks, we introduce the concept of per-link parallel and distributed HYBF of mmWave and propose a very low-complexity P&D-HYBF design based on cooperation. It removes the requirement of transferring full CSI to the central node and allows each FD BS to update the beamformers associated with different users on different computational processors in parallel at each iteration.

To proceed, we first make the following assumptions:

1. There exists a feedback link among the FD BSs and they cooperate among themselves by exchanging information about the digital beamformers, analog beamformers and analog combiners.
2. Local CSI is made accessible for the FD BSs.
3. Each FD BS has multiple computational processors dedicated for UL and DL.
4. The computations take place at the BSs and the optimized beamformers of the UL users are communicated to them afterwards.

Note that the WSR maximization problem (5.12) is decomposed into (5.18a) with MM, in which to update the beamformers for each UL or DL user at each iteration, only the gradients are required. Therefore, they summarize complete information about all the remaining links in the network. From a practical point-of-view, the gradients for each link take into account the interference generated towards all the other links, and hence limit greedy behaviour while updating its beamformer. However, problem (5.18a) is coupled among different links as the covariance matrices of other users directly appear in the gradients. Hence, the update of one beamformer affects the received covariance matrices, and thus the gradients, of all the other users/links. Therefore, (5.18a) can be solved only in a centralized fashion based on alternating optimization.

To decouple the global optimization problem (5.18a) into local per-link independent optimization sub-problems for each FD BS, *we assume that each FD BS has some memory*

to save information about the gradients. For each FD BS, we introduce the following local variables $\forall k_b \in \mathcal{U}_b, \forall j_b \in \mathcal{D}_b$,

$$\overline{\mathbf{L}}_{j_b}^{Out} = \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL}, \quad \overline{\mathbf{L}}_{j_b}^{In} = \hat{\mathbf{G}}_{j_b, b}^{UL} + \hat{\mathbf{G}}_{j_b, b}^{DL}, \quad (5.43a)$$

$$\overline{\mathbf{L}}_{k_b}^{Out} = \hat{\mathbf{G}}_{k_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b, \bar{b}}^{DL}, \quad \overline{\mathbf{L}}_{k_b}^{In} = \hat{\mathbf{G}}_{k_b, b}^{UL} + \hat{\mathbf{G}}_{k_b, b}^{DL}. \quad (5.43b)$$

The variables $\overline{\mathbf{L}}_{k_b}^{In}$ and $\overline{\mathbf{L}}_{k_b}^{Out}$ save information about the overall interference generated inside and outside the cell by the beamformer of UL user k_b , respectively. Similarly, the local variables $\overline{\mathbf{L}}_{j_b}^{In}$ and $\overline{\mathbf{L}}_{j_b}^{Out}$ save information about the interference generated in the same cell and in the neighbouring cells by the FD BS b while serving the DL user j_b , respectively. Note that each BS can update the in-cell local variables $\overline{\mathbf{L}}_{k_b}^{In}$ and $\overline{\mathbf{L}}_{j_b}^{In}$ by itself. A feedback from the neighboring BSs is required only to update $\overline{\mathbf{L}}_{j_b}^{Out}, \forall j_b$ and $\overline{\mathbf{L}}_{k_b}^{Out}, \forall k_b$. To save information about the interference-plus-noise covariance matrices, we define the following local variables

$$\overline{\mathbf{R}}_{k_b} = \mathbf{R}_{k_b}, \quad \overline{\mathbf{R}}_{k_b}^{-1} = \mathbf{R}_{k_b}^{-1}, \quad \forall k_b, \quad (5.44a)$$

$$\overline{\mathbf{R}}_{j_b} = \mathbf{R}_{j_b}, \quad \overline{\mathbf{R}}_{j_b}^{-1} = \mathbf{R}_{j_b}^{-1}, \quad \forall j_b. \quad (5.44b)$$

The analog combiner needs information about the received covariance matrices at the antenna level, for which we define the local variables

$$\overline{\mathbf{R}}_{k_b}^a = \mathbf{R}_{k_b}^a, \quad \overline{\mathbf{R}}_{k_b}^a = \mathbf{R}_{k_b}^a, \quad \forall k_b. \quad (5.45)$$

We remark that all of the aforementioned local variables are fixed and saved in the memory. Replacing the gradients with fixed local variables in the WSR cost function (5.18a), leads to the following optimization problem

$$\begin{aligned} \max_{\substack{\mathbf{U}, \mathbf{V}, \\ \mathbf{G}^{RF}, \mathbf{F}^{RF}}} \quad & \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \overline{\mathbf{R}}_{k_b}^{-1} \mathbf{F}_b^{RFH} \right. \\ & \left. \mathbf{H}_{k_b} \mathbf{U}_{k_b}) - \text{Tr}(\mathbf{U}_{k_b}^H (\overline{\mathbf{L}}_{k_b}^{Out} + \overline{\mathbf{L}}_{k_b}^{In}) \mathbf{U}_{k_b}) \right] \end{aligned} \quad (5.46a)$$

$$\begin{aligned} & + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} w_{j_b} \left[\ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right. \\ & \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\overline{\mathbf{L}}_{j_b}^{Out} + \overline{\mathbf{L}}_{j_b}^{In}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right]. \end{aligned} \quad (5.46b)$$

s.t. (5.12b) – (5.12e) (5.46c)

Note that (5.46a) has the same structure of (5.18a), but by replacing the gradients with the fixed local variables, the global WSR problem decouples into per-link independent optimization sub-problems. Optimization of the analog combiners and analog beamformers is still coupled as they are common to all the UL and DL users in the same cell, respectively. Also, optimization of the digital beamformers for the DL users in the same cell remains coupled as each BS has to satisfy the sum power constraint. Their decoupling and the solution of (5.46) is discussed in the following.

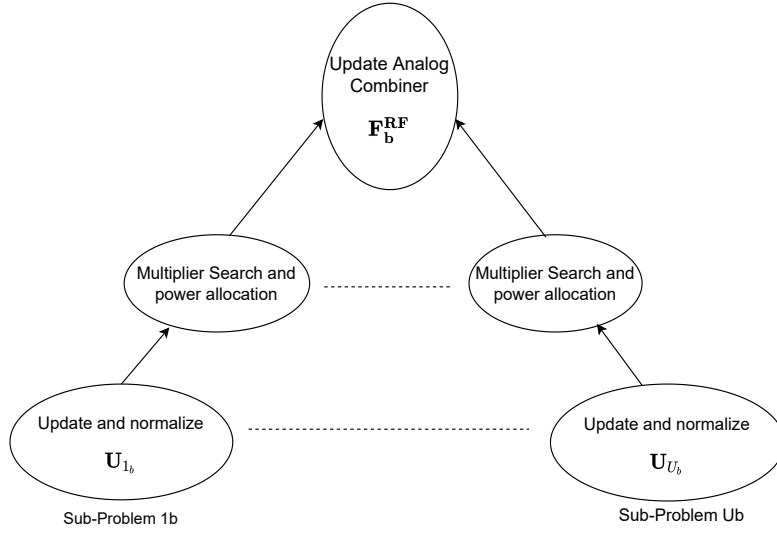


Figure 5.2: Decomposition of the WSR maximization problem in UL into per-link independent optimization sub-problems. The sub-problems are solved from the bottom to the top.

5.5.1 Per-Link Independent Sub-Problems in UL

In the UL setting, each UL user has its own sum-power constraint and only the update of analog combiner is coupled among different UL users in the same cell. *We assume that BS b updates \mathbf{F}_b^{RF} only after having updated all the digital beamformers \mathbf{U}_{k_b} .* Given this assumption and fixed local variables, UL WSR maximization problem for BS b reduces into three layers of sub-problems. At the bottom layer, BS b has to solve independent optimization sub-problems to update \mathbf{U}_{k_b} for the beamforming directions and normalize its column to unit norm, in parallel $\forall k_b$. At the middle layer, BS b has to update the stream power matrix \mathbf{P}_{k_b} while searching the multiplier λ_{k_b} in parallel $\forall k_b$. Finally, at the top layer, once all the U_b two-layer sub-problems are solved by BS b , both for the beamforming directions and powers, only one update of the analog combiner is required. Fig. 5.2 highlights the idea of the proposed decomposition for WSR maximization in UL for BS b and the sub-problems are solved from the bottom to top.

Given the fixed local variables, the local per-link independent optimization problem in UL to optimize \mathbf{U}_{k_b} for user $k_b \in \mathcal{U}_b$ can be stated as

$$\max_{\mathbf{U}_{k_b}} \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \mathbf{U}_{k_b}) - \text{Tr}(\mathbf{U}_{k_b}^H (\overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}}) \mathbf{U}_{k_b}) \right] \quad (5.47a)$$

$$\text{s.t. } \text{Tr}(\mathbf{U}_{k_b} \mathbf{U}_{k_b}^H) \preceq p_{k_b}. \quad (5.47b)$$

The per-link independent Lagrangian for (5.47) $\forall k_b$, becomes

$$\begin{aligned} \mathcal{L}_{k_b} = & w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \mathbf{U}_{k_b}) \\ & - \text{Tr}(\mathbf{U}_{k_b}^H (\overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b}) + \lambda_{k_b} p_{k_b}. \end{aligned} \quad (5.48)$$

To solve it for \mathbf{U}_{k_b} , a derivative of (5.48) can be taken, which leads to a similar KKT condition as (5.20a) with gradients replaced with the local variables. By following a similar proof for (5.21), it can be easily shown that the WSR maximizing beamformer \mathbf{U}_{k_b} can be updated as

$$\mathbf{U}_{k_b} = \mathbf{D}_{d_{k_b}} (\mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b}, \overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}} + \lambda_{k_b} \mathbf{I}), \quad (5.49)$$

which can be computed in parallel on different computational processors $\forall k_b$. To include the optimal power allocation \mathbf{P}_{k_b} in parallel, we consider the normalization of the columns of (5.49) to unit-norm and multiply the $\log \det(\cdot)$ and $\text{Tr}(\cdot)$ terms in (5.48) from the right hand side with the power matrix \mathbf{P}_{k_b} , $\forall k_b$. Power optimization problem for each link can be formally stated similarly as in (5.31), as a function of the local variables $\overline{\mathbf{L}_{k_b}^{In}}$ and $\overline{\mathbf{L}_{k_b}^{Out}}$ instead of gradients. Solving it yields the following parallel and optimal power allocation scheme

$$\mathbf{P}_{k_b} = (w_{k_b} (\mathbf{U}_{k_b}^H (\overline{\mathbf{L}_{k_b}^{In}} + \overline{\mathbf{L}_{k_b}^{Out}} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b})^{-1} - (\mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \overline{\mathbf{R}_{k_b}^{-1}} \mathbf{H}_{k_b} \mathbf{U}_{k_b})^{-1})^+, \quad \forall k_b. \quad (5.50)$$

Power allocation can be included while searching for the multiplier λ_{k_b} associated with the sum-power constraint of UL user k_b . To yield a diagonal power matrix again while searching for λ_{k_b} , we do

$$[\mathbf{L}_{k_b}^{svd}, \mathbf{D}_{k_b}^{svd}, \mathbf{R}_{k_b}^{svd}] = \mathbf{P}_{k_b}, \quad \text{and} \quad \mathbf{P}_{k_b} = \mathbf{D}_{k_b}^{svd}. \quad (5.51)$$

Multiplier λ_{k_b} should be such that (5.48) is finite and the value of λ_{k_b} should be strictly positive. We search it in parallel by solving the following problem $\forall k_b$

$$\begin{aligned} \min_{\lambda_{k_b}} \max_{\mathbf{P}_{k_b}} \quad & \mathcal{L}_{k_b}(\lambda_{k_b}, \mathbf{P}_{k_b}), \\ \text{s.t.} \quad & \lambda_{k_b} \succeq 0. \end{aligned} \quad (5.52)$$

The dual function

$$\max_{\mathbf{P}_{k_b}^{DL}} \mathcal{L}_{k_b}(\lambda_{k_b}, \mathbf{P}_{k_b}) \quad (5.53)$$

is convex [40] and can be solved with the Bisection method. Note that each optimization step stated above is fully decoupled for each communication link in UL as the local variables are fixed.

At the final step, once updated the digital beamformers \mathbf{U}_{k_b} and powers \mathbf{P}_{k_b} , one update of \mathbf{F}_b^{RF} is required, which is common to all the UL users U_b served by BS b . Simultaneous variation of all the UL beamformers \mathbf{U}_{k_b} in parallel vary the received covariance matrices $\overline{\mathbf{R}_{k_b}^a}$ and $\overline{\mathbf{R}_{k_b}^a}$ at the antenna levels, which are required to optimize \mathbf{F}_b^{RF} . However, in the memory, each BS has complete information about the beamformers of the UL users it has just updated at the bottom layer, which can be used to update $\overline{\mathbf{R}_{k_b}^a}$ and $\overline{\mathbf{R}_{k_b}^a}$, $\forall k_b \in \mathcal{U}_b$. Afterwards, for the top layer, optimization of the unconstrained analog beamformer \mathbf{F}_b^{RF} , given the local variables, can be formally stated as

$$\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} \text{ldet}(\overline{\mathbf{R}_{k_b}^{-1}} \overline{\mathbf{R}_{k_b}}). \quad (5.54a)$$

To solve it, we first write it as

$$\max_{\mathbf{F}_b^{RF}} \sum_{k_b \in \mathcal{U}_b} [w_{k_b} \text{ldet}(\mathbf{F}_b^{RFH} \overline{\mathbf{R}_{k_b}^a} \mathbf{F}_b^{RF}) - w_{k_b} \text{ldet}(\mathbf{F}_b^{RFH} \overline{\mathbf{R}_{k_b}^a} \mathbf{F}_b^{RF})], \quad (5.55)$$

and by taking its derivative leads a similar KKT condition as in (5.28), from which it is immediate to see that \mathbf{F}_b^{RF} can be optimized as

$$\mathbf{F}_b^{RF} = \mathbf{D}_{N_b^{RF}} \left(\sum_{k_b \in \mathcal{U}_b} w_{k_b} \overline{\mathbf{R}_{k_b}^a}, \sum_{k_b \in \mathcal{U}_b} w_{k_b} \overline{\mathbf{R}_{k_b}^a} \right) \quad (5.56)$$

which is unconstrained. To meet the unit-modulus and the quantization constraints, we normalize the amplitudes to unit-norm and quantize the phase part as $\mathbf{F}_b^{RF} = \mathbb{Q}_b(\angle \mathbf{F}_b^{RF}) \in \mathcal{P}_b$.

5.5.2 Per-Link Independent Sub-Problems in DL

The DL scenario is much more challenging as the digital beamformers have a coupling total sum-power constraint among the DL users in \mathcal{D}_b . Moreover, \mathbf{G}_b^{RF} is common between all the DL users and affects the transmit covariance matrices, i.e., $\mathbf{Q}_{j_b} = \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH}$, $\forall j_b$, and thus the total transmitted power. To introduce the per-link independent sub-problems in DL, we assume that each BS first updates all the digital beamformers $\mathbf{V}_{j_b}, \forall j_b$, by keeping the Lagrange multiplier ψ_b and the analog beamformer \mathbf{G}_b^{RF} fixed. Also the power matrices \mathbf{P}_{j_b} are included afterwards, while searching for ψ_b . Given this assumption, the WSR maximization problem in DL for each cell decomposes into three layers of sub-problems. At the top layer we have to search for the Lagrange multiplier ψ_b and update the power allocation matrices \mathbf{P}_{j_b} for all the DL users, in parallel. At the middle layer, we have to optimize the analog beamformer \mathbf{G}_b^{RF} . At the bottom layer, we have to update the DL beamformers \mathbf{V}_{j_b} and normalize their columns to unit-norm, in parallel $\forall j_b$. Fig. 5.3 shows the proposed per-link decomposition of the DL WSR maximization problem for FD BS $b \in \mathcal{B}$ and the sub-problems are solved from the bottom to the top.

For the bottom layer, to optimize the digital beamformer \mathbf{V}_{j_b} , each FD BS has to solve the following independent optimization sub-problem in parallel $\forall j_b \in \mathcal{D}_b$

$$\max_{\mathbf{V}_{j_b}} \left[w_{j_b} \text{ldet}(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right] \quad (5.57a)$$

$$\text{Tr} \left(\sum_{j \in \mathcal{D}_b} \mathbf{G}_b \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_b^H \right) \preceq p_b \quad (5.57b)$$

with the coupling constraint (5.57b) among different DL users in the same cell. Augmenting the cost function with the total sum-power constraint yield the Lagrangian

$$\begin{aligned} \mathcal{L}_b^{DL} = & \sum_{j_b \in \mathcal{D}_b} w_{j_b} \text{ldet}(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \\ & - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) + \psi_b p_b. \end{aligned} \quad (5.58)$$

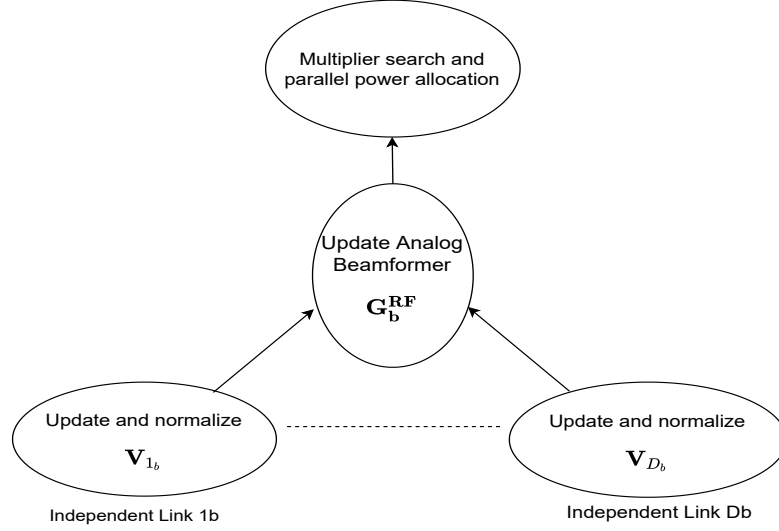


Figure 5.3: Decomposition of the WSR optimization problem in DL into per-link independent optimization sub-problems. The sub-problems are solved from the bottom to the top.

In (5.58), ψ_b and \mathbf{G}_b^{RF} are fixed and will be updated with the power allocation at the top and middle layer, respectively, to meet the total sum-power constraint. Therefore the update of \mathbf{V}_{j_b} at the bottom layer only for the beamforming directions and normalization to unit-norm columns remains decoupled $\forall j_b$. To solve it, a derivative for \mathbf{V}_{j_b} can be taken which leads to a similar KKT condition (5.20b) for the centralized version, with gradients replaced with the fixed local variables. From the KKT conditions, by following a similar proof for the C-HYBF scheme, it can be immediately shown that \mathbf{V}_{j_b} can be updated as

$$\mathbf{V}_{j_b} = \mathbf{D}_{d_{j_b}} (\mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF}, \mathbf{G}_b^H (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \mathbf{G}_b). \quad (5.59)$$

To include the optimal power allocation, we consider the normalization of the columns of \mathbf{V}_{j_b} to unit-norm in parallel. Once all the digital beamformers are optimized, at the middle layer, BS b has to optimize the analog combiner \mathbf{G}_b^{RF} . By considering the unconstrained analog combiner, at the middle layer, each BS has to independently solve the following unconstrained optimization problem

$$\begin{aligned} \max_{\mathbf{G}_b^{RF}} \sum_{j_b \in \mathcal{D}_b} & \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right. \\ & \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right]. \end{aligned} \quad (5.60)$$

Note that each BS has complete information about the digital beamformers optimized at the bottom layer, which can be first used to update $\overline{\mathbf{R}_{j_b}^{-1}}$ and $\overline{\mathbf{L}_{j_b}^{In}}$ appearing in (5.60). By taking the derivative of (5.60) and by following a similar proof of (5.25), it can be

easily shown that \mathbf{G}_b^{RF} can be optimized as

$$\begin{aligned} \text{vec}(\mathbf{G}_b^{RF}) = \mathbf{D}_1 \big(\sum_{j_b \in \mathcal{D}_b} (\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H (\mathbf{I} + \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \\ \mathbf{H}_{j_b} \mathbf{G}_b^{RF})^{-1})^T \otimes \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b}, \sum_{j_b \in \mathcal{D}_b} (\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H)^T \\ \otimes (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \big), \end{aligned} \quad (5.61)$$

with gradients of (5.25) replaced with the local variables. The analog beamformer \mathbf{G}_b^{RF} in (5.61) is unconstrained and vectorized, we do $\text{unvec}(\text{vec}(\mathbf{G}_b^{RF}))$ to shape it into correct dimensions and then set $\mathbf{G}_b^{RF} = \mathbb{Q}(\angle \mathbf{G}_b^{RF}) \in \mathcal{P}_b$ to meet the unit-modulus and quantization constraints. For the top layer, the optimal stream power allocation can be included while searching the multiplier ψ_b to satisfy the sum-power constraint p_b . Assuming the multiplier ψ_b to be fixed, the per-link independent power optimization problem $\forall j_b \in \mathcal{D}_b$ can be stated as

$$\begin{aligned} \max_{\mathbf{P}_{j_b}} \quad & \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}) \right. \\ & \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}) \right]. \end{aligned} \quad (5.62)$$

In (5.62), the update of power matrix $\mathbf{P}_{j_b}, \forall j_b$ remains independent and the multiplier ψ_b must be updated based on the sum of the transmit covariance matrices $\sum_{j_b} \mathbf{G}_b^{RFH} \mathbf{V}_{j_b}^H \mathbf{P}_{j_b} \mathbf{V}_{j_b} \mathbf{G}_b^{RF}$, once all the power matrices \mathbf{P}_{j_b} are updated in parallel. Solving (5.62) in parallel $\forall j_b$ leads to the following optimal power allocation scheme

$$\begin{aligned} \mathbf{P}_{j_b} = & (w_{j_b} (\mathbf{V}_{j_b}^H \mathbf{G}_b^H (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \mathbf{G}_b \mathbf{V}_{j_b})^{-1} \\ & - (\mathbf{V}_{j_b}^H \mathbf{G}_b^H \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b \mathbf{V}_{j_b})^{-1})^+. \end{aligned} \quad (5.63)$$

Finally, the Lagrange multiplier can be searched with the Bisection method and while doing so, the water-filling for the powers for each user in DL in \mathcal{D}_b can be done in parallel with (5.63). Including the optimal power allocation (5.63) in the Lagrangian (5.58) leads to

$$\begin{aligned} \mathcal{L}_b^{DL} = \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \overline{\mathbf{R}_{j_b}^{-1}} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}) \right. \\ \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\overline{\mathbf{L}_{j_b}^{In}} + \overline{\mathbf{L}_{j_b}^{Out}} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b} \mathbf{P}_{j_b}) \right] + \psi_b p_b \end{aligned} \quad (5.64)$$

Multiplier ψ_b should be such that (5.64) is finite and the value of ψ_b should be strictly positive. It can be searched by solving the following problem

$$\begin{aligned} \min_{\psi_b} \max_{\mathbf{P}_b^{DL}} \quad & \mathcal{L}_b^{DL}(\psi_b, \mathbf{P}_b^{DL}), \\ \text{s.t.} \quad & \psi_b \succeq 0. \end{aligned} \quad (5.65)$$

where \mathbf{P}_b^{DL} denotes the collection of powers in DL for BS b . The dual function

$$\max_{\mathbf{P}_b^{DL}} \mathcal{L}_b^{DL}(\psi_b, \mathbf{P}_b^{DL}) \quad (5.66)$$

is convex [40] and can be solved with the Bisection Algorithm. When (5.63) becomes non-diagonal while searching the multiplier ψ_b , we take the SVD of the stream powers in parallel for each DL user and set $\mathbf{P}_{j_b} = \mathbf{D}_{j_b}^{svd}, \forall j_b$. After each iteration, the FD BSs must exchange information such that the local variables could be updated. The complete procedure to execute the cooperative P&D-HYBF for WSR maximization in a multi-cell mMIMO mmWave FD network is formally stated in Algorithm 2. Once it converges, the combiners for the UL and DL users must be chosen as the MMSE combiners, which will not affect the rate achieved at convergence.

Remark 2: Note that each FD cell requires only the local CSI information, i.e., only for the out-cell interfering channels to update the the out-cell local variables $\mathbf{L}_{j_b}^{Out}$ and $\mathbf{L}_{k_b}^{Out}$. As we considered the analog beamformer and combiner in the optimization problem (which leads to the number of digital beamformers equal to the number of user of the size of the RF chains), only one analog beamformer and combiner of bigger dimension and many digital beamformers of minimal dimension need to be shared by each FD BS, regardless of the number of users served in a mMIMO scenario. Moreover, omitting the digital combiners reduces the per-iteration computational complexity, and the communication overhead for P&D-HYBF is also minimized. To further reduce the communication overhead, if the out-cell interference and the CI channels among different cells vary slowly, the FD BSs can consider updating them only when these channels have changed significantly compared to the last feedback stage.

Remark 3: Each BS has full flexibility to reconsider the allocation of processors in UL and DL in a highly asymmetric traffic scenario. For example, suppose that BS b solves the problem early in one direction with fewer users, i.e., in UL or DL. In that case, the idle processors can be reallocated immediately to serve the transmission direction with many users to further distribute the computational burden and achieve faster convergence.

5.5.3 On the Convergence of P&D-HYBF

The convergence proof for P&D-HYBF follows similarly from the proof stated for the C-HYBF scheme. Compared to the centralized version, fixing the local variables leads to a different type of information saved for each communication link while updating its beamformer. As the beamformers are computed as the dominant generalized eigenvectors, they increase the WSR for every link at each iteration. However, the increase is different as the local variables' information differs from the information captured in the gradients. The gradients are updated immediately before updating each beamformer, and the local variables are updated in a synchronized manner in parallel once all the FD BSs have entirely updated their UL and DL beamformers. However, as in P&D-HYBF the BSs share information about the updated variables at each iteration, it makes the beamformers' update aware of the generated interference towards other links, which leads to a monotonic increase in WSR and assures convergence.

Algorithm 10 Parallel and Distributed Hybrid Beamforming

Given: The rate weights, in-cell CSI and interfering channels.

Dedicate: Multiple processors in UL and DL $\forall b$.

Initialize: $\mathbf{G}_b, \mathbf{V}_{j_b}, \mathbf{U}_{k_b}, \forall j_b \in \mathcal{D}_b, \forall k_b \in \mathcal{U}_b$ in each cell.

Repeat until convergence

Cooperation stage ($\forall b \in \mathcal{B}$)

 Share $\mathbf{G}_b^{RF}, \mathbf{F}_b^{RF}$ and $\mathbf{U}_{k_b}, \mathbf{V}_{j_b}, \forall k_b, \forall j_b$

For each cell ($\forall k_b \in \mathcal{U}_b, \forall j_b \in \mathcal{D}_b$.)

 Update $\overline{\mathbf{L}}_{j_b}^{In}, \overline{\mathbf{L}}_{k_b}^{In}$ from the memory.

 Based on the feedback, update $\overline{\mathbf{L}}_{j_b}^{Out}$ and $\overline{\mathbf{L}}_{k_b}^{Out}$

 Update $\overline{\mathbf{R}}_{k_b}, \overline{\mathbf{R}}_{j_b}$ and $\overline{\mathbf{R}}_{j_b}, \overline{\mathbf{R}}_{k_b}$ from the memory.

Solve in parallel $\forall b \in \mathcal{B}$

Parallel DL for BS b

Set: $\underline{\psi}_b = 0, \overline{\psi}_b = \psi_b^{max}$.

 Compute $\mathbf{V}_{j_b}^d$ in parallel with (5.59) and normalize it

 Update $\overline{\mathbf{R}}_{j_b}^{-1}$ and $\overline{\mathbf{L}}_{j_b}^{In}, \forall j_b$ from the memory.

 Compute \mathbf{G}_b^{RF} with (5.61), do unvec and get $\angle \mathbf{G}_b^{RF}$

Repeat until convergence

 set $\psi_b = (\underline{\psi}_b + \overline{\psi}_b)/2$

In parallel $\forall j_b$

 Compute \mathbf{P}_{j_b} with (5.63)

$[\mathbf{X}_{\mathbf{P}_{j_b}}, \mathbf{D}_{\mathbf{P}_{j_b}}, \mathbf{Y}_{\mathbf{P}_{j_b}}] = \text{SVD}(\mathbf{P}_{j_b})$

$\mathbf{P}_{j_b} = \mathbf{D}_{\mathbf{P}_{j_b}}$

$\mathbf{Q}_{j_b} = \mathbf{G}_b \mathbf{V}_{j_b} \mathbf{P}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^H$

if constraint for ψ_b is violated

 set $\underline{\psi}_b = \psi_b$,

else $\overline{\psi}_b = \psi_b$

Parallel UL for BS b

Set: $\underline{\lambda}_{k_b} = 0, \overline{\lambda}_{k_b} = \lambda_{k_b}^{max}, \forall k_b$.

 Compute $\mathbf{U}_{k_b}^d$ (5.49) in parallel and normalize it

Repeat until convergence in parallel $\forall k_b$

 set $\lambda_{k_b} = (\underline{\lambda}_{k_b} + \overline{\lambda}_{k_b})/2$

 Compute \mathbf{P}_{k_b} in parallel with (5.50).

$[\mathbf{X}_{\mathbf{P}_{k_b}}, \mathbf{D}_{\mathbf{P}_{k_b}}, \mathbf{Y}_{\mathbf{P}_{k_b}}] = \text{SVD}(\mathbf{P}_{k_b})$

 Set $\mathbf{P}_{k_b} = \mathbf{D}_{\mathbf{P}_{k_b}}$

 Set $\mathbf{T}_b = \mathbf{U}_{k_b} \mathbf{P}_{k_b} \mathbf{U}_{k_b}^H$

if constraint for λ_{k_b} is violated

 set $\underline{\lambda}_{k_b} = \lambda_{k_b}$

else $\overline{\lambda}_{k_b} = \lambda_{k_b}$

 Update $\overline{\mathbf{R}}_{k_b}^a, \overline{\mathbf{R}}_{k_b}^a, \forall k_b$.

 Compute \mathbf{F}_b^{RF} with (5.56) and get $\angle \mathbf{F}_b^{RF}$.

Repeat

Quantize \mathbf{G}_b^{RF} and \mathbf{F}_b^{RF} , with $\mathbb{Q}_b(\cdot), \forall b$.

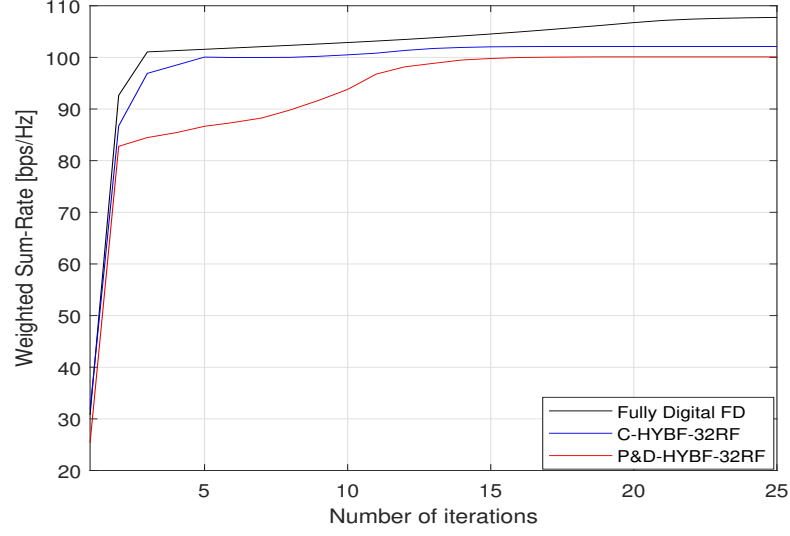


Figure 5.4: Typical Convergence behaviour of the proposed HYBF designs with 32 RF chains in comparison with the fully digital beamforming.

Fig. 5.4 shows a typical convergence behaviour of the proposed C-HYBF and P&D-HYBF schemes in comparison with the centralized fully digital beamforming scheme. It is also visible that the P&D-HYBF requires a similar number of iterations to converge as the C-HYBF and therefore requires a minimal amount of information exchange among different cells. We can also see that the increase in WSR at each iteration is different for both designs as the information captured in the gradients and local variables differs. It is also highlighted that despite being a decentralized design, the P&D-HYBF scheme achieves similar WSR as the C-HYBF.

5.5.4 Computational Complexity Analysis

For complexity analysis, we assume equal number of users in DL and UL in each cell, i.e., $D_b = D$ and $U_b = U$, $\forall b \in \mathcal{B}$. We also assume the same number of antennas in each cell for the BSs, UL and DL users.

Computational Complexity of C-HYBF

One iteration of C-HYBF consists in updating BD DL beamformers \mathbf{V}_{j_b} , BU UL beamformers \mathbf{U}_{k_b} , B analog beamformers \mathbf{G}_b^{RF} and B analog combiners \mathbf{F}_b^{RF} . Assuming that the number of antennas get large and the computations take place at the central node iteratively, the per-iteration computational complexity of the proposed C-HYBF scheme results to be

$$\begin{aligned} \approx \mathcal{O}(B^2 U^2 N_b^{RF^3} + B^2 U D N_{j_b}^3 + B^2 D^2 N_{j_b}^3 + B^2 D U N_b^{RF^3} + B M_b^{RF^2} M_b^2 + B N_b^{RF} N_b^2 \\ + B D d_{j_b} M_b^{RF^2} + B D d_{k_b} N_{k_b}^2). \end{aligned} \quad (5.67)$$

Note that the complexity of C-HYBF (5.67) scales quadratically as a function of the number of cells B (network size) and users U or D (density).

The computational complexity of P&D-HYBF in DL and UL for each FD BS is fully decoupled on different processors. Therefore, in the following, it will be analyzed separately under the assumption that the dimensions of the antennas get large.

Worst-case Computational Complexity in DL for P&D-HYBF

We first assume the number of computational processors dedicated for DL equal to the number of DL users for each FD BS. The worst-case computational complexity for each BS in P&D-HYBF in DL is given for the processor which has to update one digital beamformer and then update also the analog beamformer \mathbf{G}_b^{RF} , which is given by

$$DL \approx \mathcal{O}(BDN_{j_b}^3 + BUN_b^{RF3} + d_{j_b}M_b^{RF2} + M_b^{RF2}M_b^2). \quad (5.68)$$

If the number of processors dedicated for DL is less than the total number of DL users, then each processor may have to update K digital beamformers before updating the analog beamformer \mathbf{G}_b^{RF} . In that case, the worst-case complexity is given for the processor in each cell which updates K digital beamformers and then also the analog beamformer. In such a case, the complexity is given by

$$DL \approx \mathcal{O}(KBDN_{j_b}^3 + KBUN_b^{RF3} + Kd_{j_b}M_b^{RF2} + M_b^{RF2}M_b^2). \quad (5.69)$$

Worst-case Computational Complexity in UL for P&D-HYBF

Assuming the dedicated processors for UL equal to the UL users in each cell, the worst-case complexity in each cell is given for the processor which has to update one analog combiner \mathbf{F}_b^{RF} and one digital beamformer \mathbf{U}_{j_b} . In such a case, the complexity results to be

$$UL \approx \mathcal{O}(BUN_b^{RF3} + BDN_{j_b}^3 + d_{k_b}N_{k_b}^2 + N_b^{RF}N_b^2). \quad (5.70)$$

If the number of dedicated computational processors is less than the UL users, then each processor may have to update N UL beamformers \mathbf{U}_{k_b} before updating the analog combiner \mathbf{F}_b^{RF} . In such a scenario, the worst-case complexity is given by

$$UL \approx \mathcal{O}(NBUN_b^{RF3} + NBDN_{j_b}^3 + Nd_{k_b}N_{k_b}^2 + N_b^{RF}N_b^2). \quad (5.71)$$

From the analysis presented above, it is clear that the complexity of P&D-HYBF distributed over multiple processors at each FD BS is significantly less than the C-HYBF scheme. Namely, the latter is quadratic, and the former is only linear in the number of UL and DL users and number of cells. Intuitively, for every beamformer's update, we have to invert the covariance matrices in the gradients for all the remaining users. As we have to repeat this for all the beamformers, it leads to a quadratic behaviour. Note that any of the centralized HYBF schemes would have a quadratic behaviour. On the other hand, in P&D-HYBF, each processor has to update only one or a very limited number of variables and linearize with the gradients for all the remaining users at each iteration, which leads to only a linear increase in the complexity.

Table 5.2: Simulation parameters to simulate a multi-cell mMIMO mmWave FD system.

Parameters		
Cells	B	2
UL and DL users	U_b, D_b	1, $\forall b$
Data streams	d_{j_b}, d_{k_b}	2, $\forall b$
BSs antennas	M_b, N_b	100, 60
Clusters and Paths	$N_{c,b}, N_{p,b}$	3, 3
RF chains (BSs)	$M_b^{RF} = N_b^{RF}$	10, 12, 16, 32
Rx RF chains	N_r	10, 12, 16, 32
User antennas	$M_{k_b} = N_{j_b}$	5
DL user antennas	N_j	5
Rician Factor	κ_b	1
Array response	$\mathbf{a}_{r,b}, \mathbf{a}_{t,k_b}, \mathbf{a}_{r,j_b}$	ULA, ULA, ULA
AoA	ϕ_{j_b}, ϕ_{k_b}	$\mathcal{U} \sim [-30^\circ, 30^\circ]$
AoD	$\theta_{j_b}, \theta_{k_b}$	$\mathcal{U} \sim [-30^\circ, 30^\circ]$
Rate weights	w_k, w_j	1
Sum Power	p_{k_b}, p_b	1, 1
Uniform Quantizer	$\mathbb{Q}_b(\cdot)$	4 or 10 bits
Relative Angle	Θ_b	90°
Array separation	D_b	20 cm

5.6 Simulation Results

This section presents simulation results to evaluate the performance of the proposed C-HYBF and P&D-HYBF schemes for the multi-cell mMIMO mmWave FD network. For comparison, we consider the following benchmark schemes:

- A centralized *Fully Digital FD* scheme affected by the LDR noise.
- A centralized *Fully Digital HD* scheme with LDR noise, serving the UL and DL users by separating the resources in times. It is neither affected by the SI nor by the CI.

To compare the performance with a fully digital HD system, we define the additional gain in terms of percentage for an FD system over an HD system as

$$Gain = \frac{WSR_{FD} - WSR_{HD}}{WSR_{HD}} \times 100[\%]. \quad (5.72)$$

where WSR_{FD} and WSR_{HD} are the network WSR for the FD and HD system, respectively. We assume the same SNR level for all the FD BSs, defined as

$$SNR = p_b / \sigma_b^2, \quad (5.73)$$

with transmit power p_b and thermal noise variance σ_b^2 . We assume that the UL users and FD BSs transmit with the same amount of power, i.e., $p_{k_b} = p_b, \forall k_b$. The thermal noise level for DL users is set as $\sigma_{j_b}^2 = \sigma_b^2, \forall j_b$. The total transmit power is normalized to 1 and we choose the thermal noise variance to meet the desired SNR. We simulate a multi-cell network consisting of $B = 2$ cells with each FD BS serving one DL and one UL

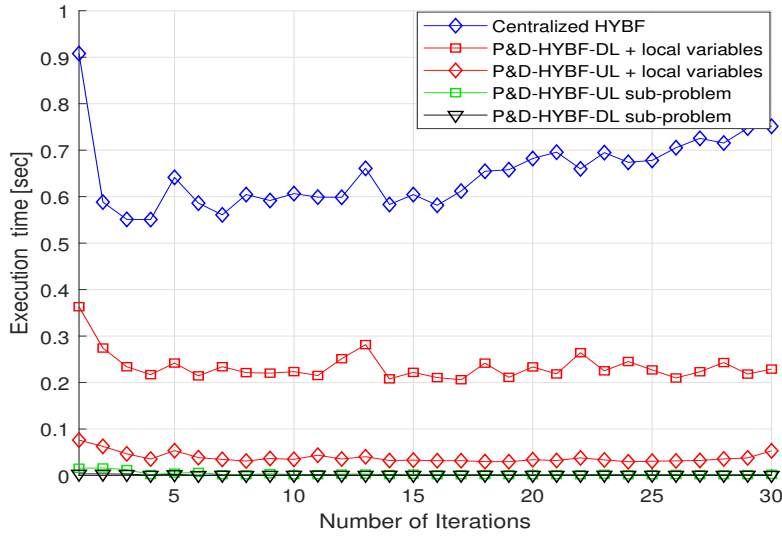


Figure 5.5: Execution time for the C-HYBF and the P&D HYBF schemes with 32-RF chains.

user. P&D-HYBF is evaluated on a computer consisting of 4 computational processors, equal to the number of users in the network, i.e., fully parallel implementation. BSs are assumed to have $M_b = 100$ transmit and $N_b = 60$ receive antennas. The RF chains in transmission and reception for FD BSs are chosen as $M_b^{RF} = N_b^{RF} = 32, 16, 10$ or 8 and the phase-shifters are assumed to be quantized with a uniform quantizer $\mathbb{Q}_b(\cdot)$ of 10 or 4 bits. The DL and UL users are assumed to have $N_{j_b} = N_{k_b} = 5$ antennas and are served with $d_{j_b} = d_{k_b} = 2$ data streams. The number of paths and number of clusters are chosen to be $N_{c,b} = N_{p,b} = 3$ and the AOA $\theta_{j_b}^{n_{p,b},n_{c,b}}$ and AOD $\phi_{k_b}^{n_{p,b},n_{c,b}}$ are assumed to be uniformly distributed in the interval $\mathcal{U} \sim [-30^\circ, 30^\circ], \forall j_b, k_b$. We assume uniform-linear-arrays (ULAs) for the FD BSs and users. For the FD BSs, the transmit and the receive array are assumed to be separated with distance $D_b = 20$ cm with a relative angle $\Theta_b = 90^\circ$ and $r_{m,n}$ in (6.3) is set given D_b and Θ_b as in (9) [54]. The Rician factor is chosen to be $\kappa_b = 1$ and the rate weights are set to be $w_{k_b} = w_{j_b} = 1$. Table 5.2 summarizes all the parameters' choice. Digital beamformers are initialized as the dominant eigenvectors of the channel covariance matrix of each user. The analog beamformers and combiners are initialized as the dominant eigenvectors of the sum of the channel covariance matrices across all the DL and UL users, respectively. The results reported herein are averaged over 100 channel realizations. Note that as we are assuming perfect CSI, the SI can be cancelled only up to the LDR noise floor, which reflects the residual SI power.

Fig. 5.5 shows a typical execution time to run the C-HYBF and the P&D-HYBF with 32 RF chains. We can see that the former requires significant computational time as it can only update different variables iteratively based on alternating optimization, one after the other. Transferring full CSI to the central node and communicating back the optimized variables to all cells will add significant additional time. For the latter, computation of the local variables takes place in parallel for each BS, which has to compute only the variables

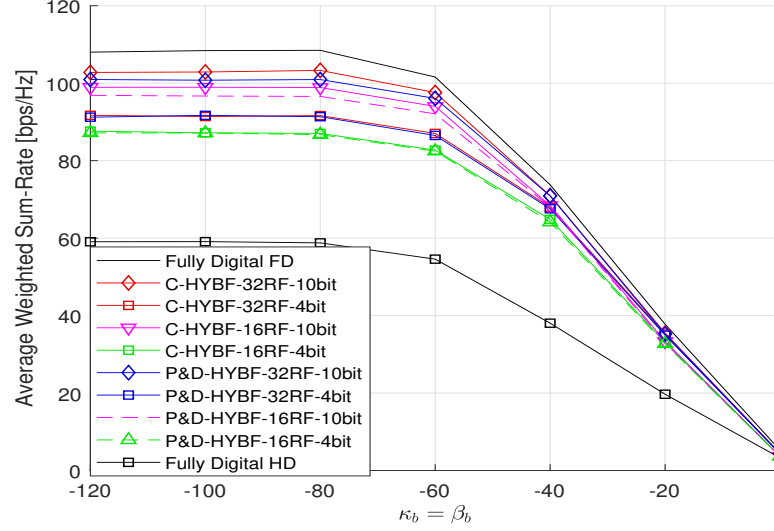


Figure 5.6: Average WSR as a function of the LDR noise at SNR= 20 dB with 32 or 16 RF chains and 10 or 4 bit phase-resolution.

associated with its users in parallel on different processors. We can see the P&D-HYBF requires $\sim 1/21$ and $\sim 1/2.3$ less time in UL and DL, respectively, than the average execution time of C-HYBF. The complexity of P&D-HYBF in DL is dominated by the computation of one large generalized dominant eigenvector to update the vectorized analog beamformer, which has complexity $\mathcal{O}(M_b^{RF^2} M_b^2)$. In UL, complexity of the analog combiner is only $\mathcal{O}(N_b^{RF} N_b^2)$. We also show the execution time to solve one sub-problem for the bottom layers in UL and DL, which is negligible compared to the average execution time of C-HYBF. Given the complexity analysis in Section-5.5.4, we can expect that the execution time of C-HYBF will increase quadratically as the number of users or cells increase. P&D-HYBF requires significantly less time and the execution time is expected to increase only linearly as the network size grows, which makes it very desirable.

Fig. 5.6 shows the average WSR achieved with both schemes as a function of the LDR noise with 32 or 16 RF chains and 10 or 4 bits phase-resolution. We can see that the P&D-HYBF performs very close to the C-HYBF scheme with the same number of RF chains and phase resolution. Fully digital FD achieves $\sim 83\%$ of additional gain than the fully digital HD for any LDR noise level. For a low LDR noise level $k_b < -80$ dB, C-HYBF and P&D-HYBF with 32 RF chains achieve $\sim 74\%$, 55% and $\sim 71\%$, 54% additional WSR with 10, 4 bits phase resolution, respectively. With 16 RF chains, the gain results to be $\sim 67\%$, 48% and $\sim 64\%$, 47% , with 10, 4 bits phase-resolution, respectively. As the LDR noise variance increases, the achievable WSR for both the FD and HD systems decreases considerably. For $k_b \geq -40$ dB, all of the beamforming schemes achieve a similar average WSR. Fig. 5.7 shows the average WSR as function of the LDR noise with only 12 or 10 RF chains and with 10 or 4 bits phase-resolution. We can see that both schemes achieve significant performance improvement in terms of WSR with a few RF chains and very low phase-resolution and significantly outperforms the fully digital HD system at any LDR noise level. In Fig. 5.7, for LDR noise $k_b \leq 80$ dB, C-HYBF and

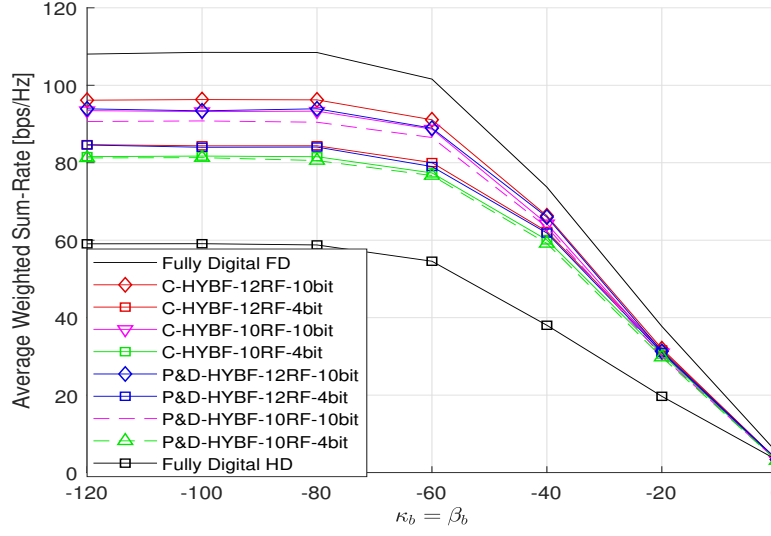


Figure 5.7: Average WSR as a function of the LDR noise at SNR= 20 dB with 12 or 10 RF chains and 10 or 4 bit phase-resolution.

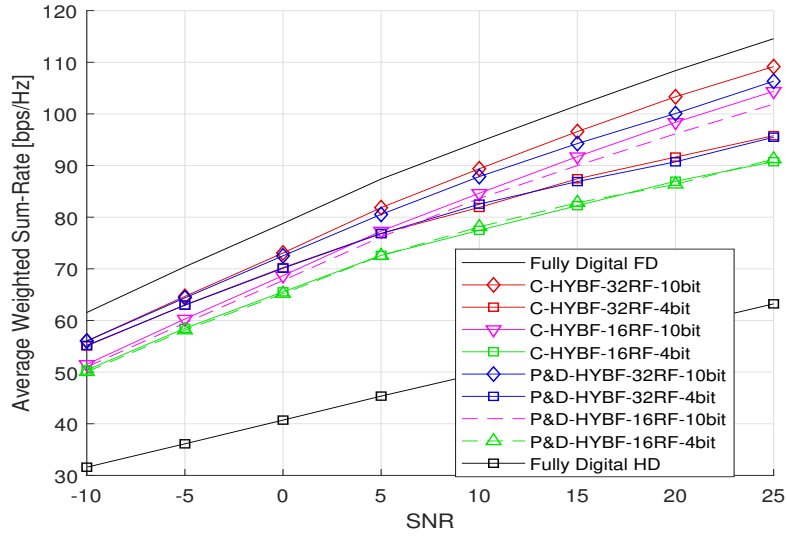


Figure 5.8: Average WSR as a function of the SNR with LDR noise $\kappa_{kb} = \kappa_b = \beta_b = \beta_{jb} = -80$ dB, with 32 or 16 RF chains and 10 or 4 bit phase-resolution.

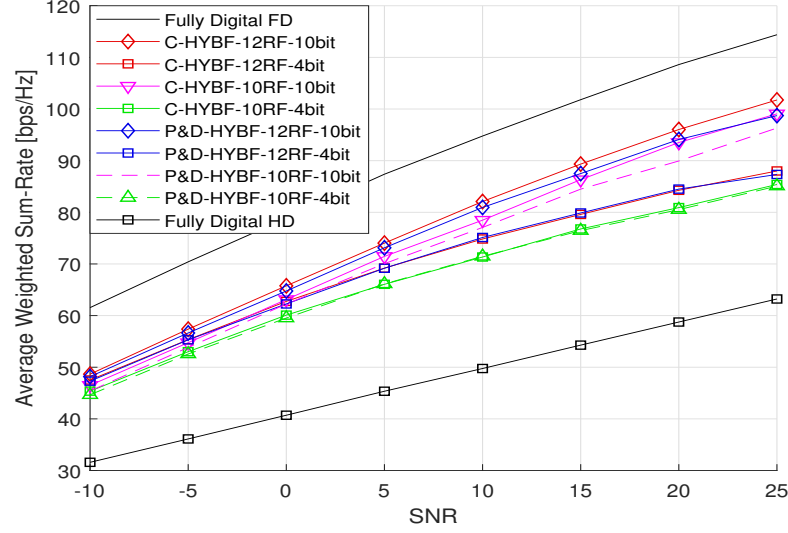


Figure 5.9: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -80$ dB, with 12 or 10 RF chains and 10 or 4 bit phase-resolution.

P&D-HYBF with 12 RF chains achieves $\sim 60, 43\%$ and $\sim 57, 43\%$ additional gain with 10, 4 bit phase-resolution, respectively. With 10 RF chains, the additional WSR results to be $\sim 58, 38\%$ and $\sim 53, 37\%$, with 10, 4 bit phase-resolution, respectively. As the LDR noise increases, the achievable average WSR decreases and results to be a fewer bps/Hz for any of the designs.

Fig. 5.8 shows the average WSR as a function of the SNR with 32 and 16 RF chains and with 10 or 4 bit phase-resolution affected with LDR noise $k_b = -80$ dB, in comparison with the benchmark schemes. A fully digital FD system achieves $\sim 94\%$ and $\sim 82\%$ additional gain at low and high SNR, respectively. With 32 RF chains and 10 bit phase-resolution, the C-HYBF scheme achieves $\sim 79\%$ gain at all the SNR levels and the P&D-HYBF achieves $\sim 77\%$ and $\sim 68\%$ gain at low and high SNR, respectively. As the phase-resolution decreases to 4-bits, we can see that the loss in WSR compared to the 10-bit phase-resolution case is much more evident at high SNR. Still, with 16 RF chains and 10 or 4 bit phase-resolution, both schemes significantly outperform the fully digital HD scheme for any SNR. Fig. 5.9 shows the average WSR as a function of the SNR with same LDR noise level as in Fig. 5.8, i.e., $k_b = -80$ dB, but with 10 or 12 RF chains and 10 or 4 bit phase-resolution. The achieved average WSR presents a similar behaviour as in the case of a high number of RF chains. We can see that both the proposed schemes significantly outperform the fully digital HD system with only 10 RF chains and with a very low phase-resolution of 4 bits.

Fig. 5.10 shows the achieved average WSR as a function of the SNR with LDR noise $k_b = -40$ dB, which reflects communication systems in which the signal suffers from a very high level of distortions. It is visible that at very high LDR noise level, the WSR does not increase as the SNR increases. When the LDR noise dominates, decreasing the thermal noise variance has negligible effect on the effective signal-to-LDR-plus-thermal-noise ratio (SLNR). Therefore, dominance of the LDR noise variance acts as a ceiling to the effective

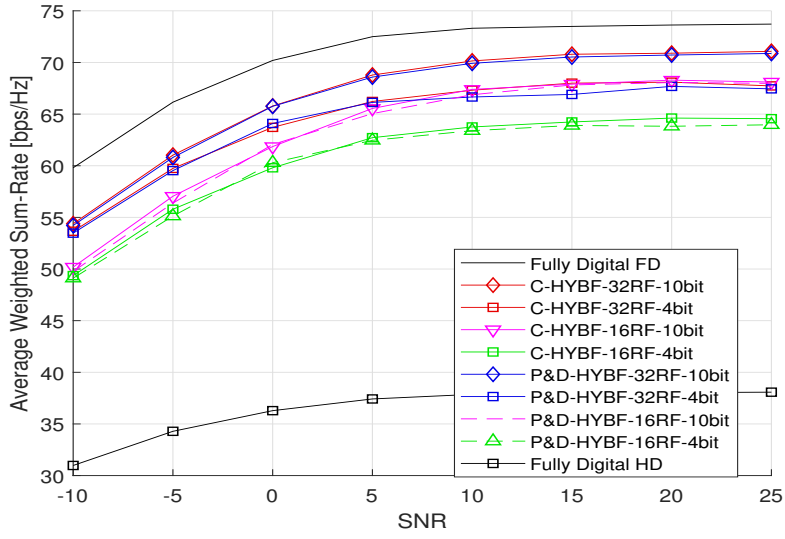


Figure 5.10: Average WSR as function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -40$ dB, with 32 or 16 RF chains and 10 or 4 bit phase-resolution.

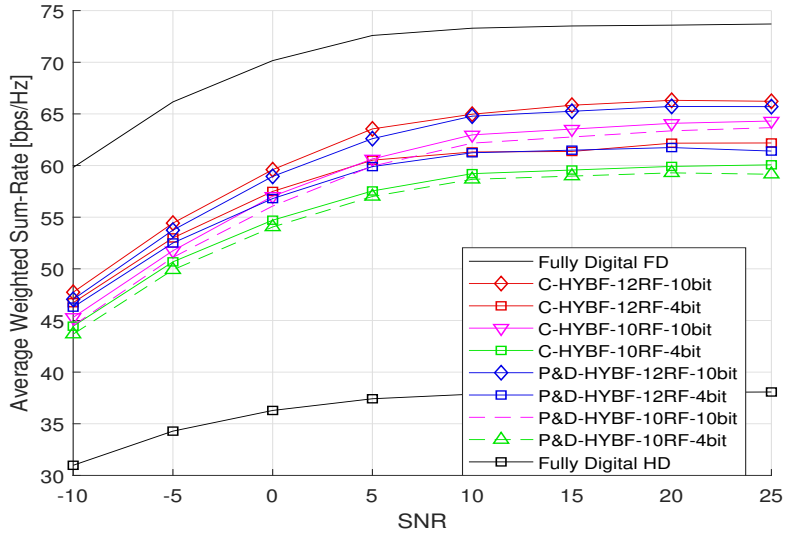


Figure 5.11: Average WSR as a function of the SNR with LDR noise $\kappa_{k_b} = \kappa_b = \beta_b = \beta_{j_b} = -40$ dB, with 12 or 10 RF chains and 10 or 4 bit phase-resolution.

SLNR ratio, which limits the achievable WSR. Consequently, increasing the SNR does not dictate higher WSR in a multi-cell mmWave FD system in the case of high LDR noise, which saturates at SNR=10 dB. We can also see that with a large LDR noise level, C-HYBF and P&D-HYBF perform similarly with the same phase-resolution and RF chains. At high SNR, both schemes achieve higher WSR at high LDR noise level with 16 RF chains and 10 bit phase-resolution than the case of 32 RF chains and 4 bit phase-resolution. Fig. 5.11 shows the average WSR as a function of the SNR with only 10 or 12 RF chains and with 10 or 4 bit phase-resolution. We can also see that, at very high LDR noise level, both schemes still perform similarly even with a very low number of RF chains and low phase-resolution. Fig. 5.11 also shows that both schemes with 10 RF chains and 10 bit phase resolution are more robust to the LDR noise than the case of 12 RF chains and 4 bit phase-resolution.

As the results reported above consider LDR for all the devices, the achieved WSR as a function of LDR noise variance can be expected in a practical multi-cell FD system with non-ideal hardware. From the results presented above, we can conclude that both the proposed HYBF schemes achieve significant additional gain and outperform the fully digital HD system with only a few RF chains. Both the schemes achieve similar average WSR with the same number of RF chains and phase-resolution. However, P&D HYBF is much more attractive because it can be distributed at each FD BS and allows a parallel update of all the variables on multiple computational processors. It eliminates the problem of transferring full CSI to the central node every channel coherence time. Moreover, it results to be also highly scalable as its complexity increases only linearly as the number of users or BSs increases. C-HYBF suffers from a quadratic increase in the computational complexity and requires a massive computational power to update all the variables jointly based on alternating optimization. P&D-HYBF achieves a similar average WSR as the C-HYBF but imposes a minimal computational burden on each processor and converges in a few iterations, thus requiring only a limited amount of information exchange among the BSs. We have also investigated the execution time for both designs and observed that parallel HYBF requires significantly less execution time. Therefore, it has the potential to be implemented in a real-time large and dense mmWave multi-cell massive mMIMO FD network and can deal with the optimization of numerous variables for the UL and DL users very efficiently.

5.7 Conclusions

This chapter has presented two joint HYBF schemes for WSR maximization in multi-cell mmWave mMIMO FD systems. Firstly, we have presented a C-HYBF scheme based on alternating optimization. It has several drawbacks and is not desirable for real-time implementation in a multi-cell mmWave FD system. To overcome all of its drawbacks, we have proposed a very low-complexity P&D-HYBF design. Its complexity scales only linearly as a function of the network size, making it desirable for the large and dense mmWave FD networks. Simulation results show that both the HYBF designs achieve similar average WSR and significantly outperform the centralized fully digital HD systems with only a few RF chains and very low phase-resolution at any LDR or SNR level.

Chapter 6

Near-Field Intelligent Reflecting Surfaces for Millimeter Wave MIMO Full Duplex

Wireless communication in mmWave or THz band is subject to severe propagation challenges, requiring the communication systems to be designed with a massive number of antennas. As discussed in the previous chapters, hybrid beamforming and combining is a cost-efficient solution to enable designing FD BSs with fewer radio frequency chains than the number of antennas. However, analog processing stage is not desirable as it has very high power consumption, high insertion loss and requires high resolution tuning of the phase shifters. Moreover, for FD, SI depends on the number of transmit antennas. Therefore, very high cost and complexity are also associated with the SI architectures to cancel the SI signal, which has the objective to mimic the transmit signal with an opposite sign from each transmit antenna and cancel it on the receive side.

The objective of this chapter is to combine the FD systems with the IRSs [95, 96, 97, 98, 99, 100]. IRS is a recently emerging hardware technology with the potential for significant energy consumption reductions. It is a meta-surface made of electronic circuits which can be flexibly programmed to shape the incoming electromagnetic field in a desired way. It is constructed with a single or fewer layer stack of planar structures which are fabricated with lithography and nano-printing methods. Each RIS unit is implemented by reflect-arrays that employ varactor diodes or other MicroElectrical-Mechanical Systems, and whose resonant frequency is electronically controlled [101, 102, 103, 104, 105, 106, 107]. The RIS units operating on the incoming field can be distributed over the meta-surface with continuity [108, 109, 110, 111] or in discrete positions [112, 113, 114, 115]. Literature on the FD systems with the IRSs is available in [116, 117, 118, 119, 120, 121], which assumes the IRSs to be placed in the far-field to improve the channel quality. This chapter aims to introduce the concept of the near-field intelligent reflecting surface (NF-IRS) for mmWave MIMO FD systems by removing the analog stage, which enables lower-dimensional active beamforming with digital beamformers and large dimensional passive beamforming with the NF-IRSs. As the FD system simultaneously transmits and

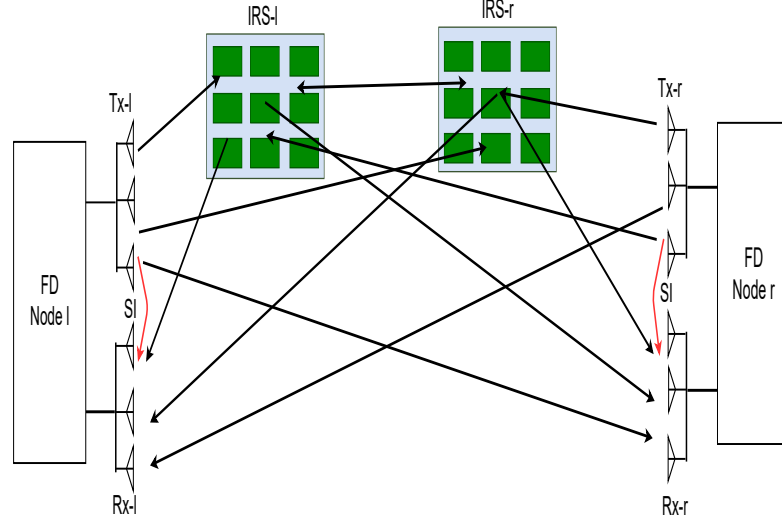


Figure 6.1: Near field IRSs assisted MIMO point-to-point FD communication in mmWave.

receives, each NF-IRS's objective will be to compensate simultaneously for the analog beamforming and combining stages of the FD systems. Moreover, the NF-IRS will also assist in shaping the SI channel, leading to a very advanced approach for SIC while keeping the cost and power consumption for the SIC very low. We remark that the proposed idea here is prominent for mmWave and THz MIMO FD systems. However, as the investigation of the FD operation in THz band is in its initial phase and the properties of the SI channel are unknown, we focus on applying the NF-IRSs for MIMO FD systems in the mmWave band.

6.1 System Model

We consider a mmWave point-to-point FD communication system consisting of two MIMO FD nodes assisted with one NF-IRS each, as shown in Figure 6.1. Let $\mathcal{F} = \{l, r\}$ denote the set containing the indices of such nodes. Let $\mathcal{I} = \{i_l, i_r\}$ denote the set containing the indices of the NF-IRSs, where the IRS $i_l \in \mathbb{C}^{L_r \times L_c}$ and IRS $i_r \in \mathbb{C}^{R_r \times R_c}$ are dedicated for node l and r , respectively. The MIMO FD node l is assumed to be equipped with M_l transmit and N_l receive antennas, respectively, and node r is assumed to be equipped with M_r transmit and N_r receive antennas, respectively. Let $\mathbf{V}_l \in \mathbb{C}^{M_l \times d_l}$ and $\mathbf{V}_r \in \mathbb{C}^{M_r \times d_r}$ denote the digital beamformers at the FD node l and r for the white and unitary variance data streams $\mathbf{s}_l \in \mathbb{C}^{d_l \times 1}$ and $\mathbf{s}_r \in \mathbb{C}^{d_r \times 1}$, respectively. Let ϕ_l and ϕ_r denote the vectors containing the collection of the phase shift response of the IRSs i_l and i_r , respectively. The elements of ϕ_l and ϕ_r at position i and j are phasors of the form $\phi_l(i) = e^{i\theta_l^i}$ and $\phi_r(j) = e^{i\theta_r^j}$, respectively. Let $\Phi_l = \text{diag}(\text{vec}(\phi_l)) \in \mathbb{C}^{L_r L_c \times L_r L_c}$ and $\Phi_r = \text{diag}(\text{vec}(\phi_r)) \in \mathbb{C}^{R_r R_c \times R_r R_c}$ denote the matrices containing the phase response of the IRSs i_l and i_r on its diagonal, with off-diagonal elements being zero.

Let $\mathbf{H}_{i_l, l} \in \mathbb{C}^{L_r L_c \times M_l}$ and $\mathbf{H}_{i_r, l} \in \mathbb{C}^{R_r R_c \times M_l}$ denote the direct channels from the

FD node l to the NF-IRS i_l and to the far-field (FF)-IRS i_r , respectively. The direct channels from the FD node r to the NF-IRS i_r and to the FF-IRS i_l are denoted with $\mathbf{H}_{i_r,r} \in \mathbb{C}^{R_r R_c \times M_r}$ and $\mathbf{H}_{i_l,r} \in \mathbb{C}^{L_r L_c \times M_r}$, respectively. Let $\mathbf{H}_{i_l,l} \in \mathbb{C}^{L_r L_c \times M_l}$ and $\mathbf{H}_{i_r,l} \in \mathbb{C}^{R_r R_c \times M_l}$ denote the direct channels from the FD node l to the NF-IRS i_l and to the FF-IRS i_r , respectively. The channels from IRSs i_l and i_r to the receive antenna array of node l are denoted with $\mathbf{H}_{l,i_l} \in \mathbb{C}^{N_l \times L_r L_c}$ and $\mathbf{H}_{l,i_r} \in \mathbb{C}^{N_l \times R_r R_c}$, respectively. The channels from the IRSs i_l and i_r to the receive antenna array of node r are denoted with $\mathbf{H}_{r,i_l} \in \mathbb{C}^{N_r \times L_r L_c}$ and $\mathbf{H}_{r,i_r} \in \mathbb{C}^{N_r \times R_r R_c}$, respectively. Let $\mathbf{H}_{i_r,i_l} \in \mathbb{C}^{R_r R_l \times L_r L_c}$ denote the channel from IRS i_l to the IRS i_r and the channel from IRS i_r to IRS i_l is denoted with $\mathbf{H}_{i_r,i_l} \in \mathbb{C}^{R_r R_c \times L_r L_c}$. Due to reciprocity, the channel from IRS r to IRS l is given as $\mathbf{H}_{i_l,i_r} = \mathbf{H}_{i_r,i_l}^T$. The direct channels from node l to node r and from node r to node l are denoted with $\mathbf{H}_{r,l} \in \mathbb{C}^{N_r \times M_l}$ and $\mathbf{H}_{l,r} \in \mathbb{C}^{N_l \times M_r}$, respectively. The SI channel for the MIMO nodes l and r are denoted with $\mathbf{H}_{l,l} \in \mathbb{C}^{N_l \times M_l}$ and $\mathbf{H}_{r,r} \in \mathbb{C}^{N_r \times M_r}$, respectively.

6.1.1 Channel Modelling

The direct channel $\mathbf{H}_{l,r}$ from the FD node r to node r can be modelled with the path-wise channel model as [54]

$$\mathbf{H}_{l,r} = \sqrt{\frac{M_r N_l}{N_c N_p}} \sum_{n_c=1}^{N_c} \sum_{n_p=1}^{N_p} \alpha_{l,r}^{(n_p, n_c)} \mathbf{a}_l(\phi_l^{n_p, n_c}) \mathbf{a}_r^T(\theta_r^{n_p, n_c}), \quad (6.1)$$

where N_c and N_p denote the number of clusters and number of rays (Figure 1 [54]), respectively, $\alpha_{l,r}^{(n_p, n_c)} \sim \mathcal{CN}(0, 1)$ is a complex Gaussian random variable with amplitudes and phases distributed according to the Rayleigh and uniform distribution, respectively, $\mathbf{a}_l(\phi_l^{n_p, n_c})$ and $\mathbf{a}_r^T(\theta_r^{n_p, n_c})$ denote the receive and transmit antenna array response at the MIMO node l and r , respectively, with angle of arrival (AoA) $\phi_l^{n_p, n_c}$ and angle of departure (AoD) $\theta_r^{n_p, n_c}$. Also, the far-field channels $\mathbf{H}_{r,l}$, $\mathbf{H}_{i_r,l}$, \mathbf{H}_{l,i_r} , \mathbf{H}_{i_r,i_l} , \mathbf{H}_{i_l,i_r} , $\mathbf{H}_{l,r}$, \mathbf{H}_{i_l,i_r} , \mathbf{H}_{i_r,i_l} can be modelled similarly as in (6.1). The SI channel for node l can be modelled as [54]

$$\mathbf{H}_{l,l} = \sqrt{\frac{\kappa_l}{\kappa_l + 1}} \mathbf{H}_l^{LoS} + \sqrt{\frac{1}{\kappa_l + 1}} \mathbf{H}_l^{ref}. \quad (6.2)$$

The scalar κ_l denotes the Rician factor, and \mathbf{H}_l^{LoS} and \mathbf{H}_l^{ref} denote the line-of-sight (LoS) and reflected contributions for the SI channel, respectively. The channel matrix \mathbf{H}_l^{ref} can be modelled as in (6.1) and the element at the m -th row and n -th columns of \mathbf{H}_l^{LoS} , using the NF channel model which will be discussed next, can be modelled as [54]

$$\mathbf{H}_l^{LoS}(m, n) = \frac{\rho}{r_{m,n}} e^{-i2\pi \frac{d_{m,n}}{\lambda}}. \quad (6.3)$$

where ρ denotes the power normalization constant to assure $\mathbb{E}(\|\mathbf{H}_l^{LoS}\|_F^2) = M_l N_l$ and λ denotes the wavelength. The scalar $d_{m,n}$ denotes the distance between m -th receive and n -th transmit antenna of node $l \in \mathcal{F}$, which depends on the transmit and receive array geometry (9) [54].

Due to the NF-IRSs, the channels $\mathbf{H}_{i_l,l}, \mathbf{H}_{l,i_l}, \mathbf{H}_{i_r,r}, \mathbf{H}_{r,i_r}$ cannot be modelled as in (6.1). To shed light on the NF concept, let D denote the size of the radiating element, which could be the transmit array of the FD node $i \in \mathcal{F}$ or the IRS i_l or i_r operating in the NF of nodes i_l and i_r , respectively. The space surrounding any transmitting device can be divided into 3 regions: 1) Reactive NF (RE-NF) 2) Radiating NF (RA-NF), and 3) Far field (FF), as shown in Figure 6.2. The radius R_1 and R_2 are the boundaries, which divides the three regions and depends on the size D and the wavelength λ . The regions RE-NF, RA-NF and FF are also known as inductive NF, Fresnel zone and Fraunhofer zone, respectively. In the RE-NF zone, the reactive fields dominate the radiation fields, and therefore its not of interest for wireless communications. In the RA-NF zone, the radiation fields dominate over the reactive fields, however, the wave received by any receiver in the RA-NF zone is characterized by a spherical wave-front. In the FF zone, the received wave-front is plane and the receiver sees the transmitter as a point-wise source. For the mmWave FD systems, the zone of interest to place NF-IRSs, such that it can compensate for the analog stage and assist with SIC, is the RA-NF zone, i.e., at distance $R_1 \leq r \leq R_2$.

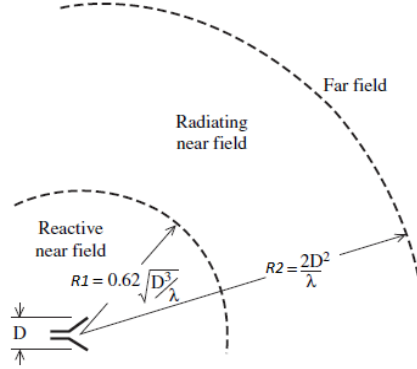


Figure 6.2: Field regions of the radiating element of size D .

Therefore, in contrast to the FF case, for the channels $\mathbf{H}_{i_l,l}, \mathbf{H}_{l,i_l}, \mathbf{H}_{i_r,r}, \mathbf{H}_{r,i_r}$, we have to consider a spherical wave-front, as considered also for the SI channel. We first consider the modelling of the NF channel $\mathbf{H}_{i_l,l}$, from the transmit antenna array of node l to the NF-IRS i_l . As we are in NF zone, for these channels we can assume that only LoS component dominates due to a very small path-loss. The impinging wave on the NF-IRS i_l transmitted from node $l \in \mathcal{F}$ is spherical, and therefore we must consider a phase shift from each antenna element to the each NF-IRS element, instead of treating the transmit antenna array or the NF-IRS as a point-wise element. Namely, the element $\mathbf{H}_{i_l,l}(m, n)$ of the channel response $\mathbf{H}_{i_l,l}$, can be modelled similarly as the near-field SI channel model (6.3), given as

$$\mathbf{H}_{i_l,l}(m, n) = \frac{\rho_{i_l,l}}{d_{m,n}} e^{-jkd_{m,n}}, \quad (6.4)$$

where $k = 2\pi/\lambda$ is the wavenumber, and $d_{n,m}$ denotes the distance between the n -th transmit antenna and m -th element of ϕ_l , with $1 \leq n \leq M_l$ and $1 \leq m \leq L_r L_c$ and $\rho_{i_l,l}$

is a normalization constant which assures $\mathbb{E}[\|\mathbf{H}_{l,i_l}\|_F^2] = L_r L_c M_l$. Similarly, for the NF channel \mathbf{H}_{l,i_l} , from the IRS i_l to the receive antenna array of l , its element $\mathbf{H}_{l,i_l}(i, j)$ can be modelled as

$$\mathbf{H}_{l,i_l}(i, j) = \frac{\rho_{l,i_l}}{d_{i,j}} e^{-jkd_{i,j}}, \quad (6.5)$$

where $d_{i,j}$ denotes the distance from the j -th element of ϕ_l to the transmit antenna i at the FD node l , with $1 \leq j \leq L_r L_c$ and $1 \leq i \leq N_l$, and ρ_{l,i_l} is the power normalization constant which assures $\mathbb{E}[\|\mathbf{H}_{l,i_l}\|_F^2] = L_r L_c N_l$. The elements of the channel response matrices $\mathbf{H}_{i_r,r}$ and \mathbf{H}_{r,i_r} can also be modelled similarly as (6.4) and (6.5), respectively.

For the mmWave band 30 – 300 GHz, the wavelengths corresponding to 30 GHz and 300 GHz result to be 10mm and 1mm, respectively. For the mmWave FD, the state-of-the-art reported results with ≤ 32 RF chains to achieve performance with HYBF comparable to the fully digital beamforming [54, 71]. Assuming uniform linear arrays (ULAs) with antenna separated by half-wavelength, for a typical antenna array with 32 antenna elements, $D = 16\lambda$. Typically, the IRS are made of hundreds of meta elements, and therefore, the NF region of each MIMO FD node is strictly dictated by the size of the NF-IRS. Assuming a NF-IRS i_l of size $100\lambda \times 100\lambda$, the boundary regions can be identified as $R_1 = 0.62\sqrt{L_{max}^2/\lambda}$ and $R_2 = 2L_{max}^2/\lambda$, where $L_{max} = \max\{L_r, L_c\}$, and for the values concerning the mmWave band are reported in Table 6.1.

Table 6.1: Field regions boundaries for an IRS of size $D = 100\lambda$.

λ	R_1	R_2
1mm	0.62m	20m
10mm	6.2m	200m

Scaling of the Channels Gains to Capture Distance Dependence

As we have defined above normalized channel gains (modelled with a statistical channel model), to capture the dependence of the channel gain on the distance, we consider scaling them with average channel gains which strictly depends on the distance. Each channel matrix is scaled with a scale factor $\sqrt{\beta}$ which dictates the average channel gain. It is chosen to be 1 for the direct links between the FD nodes l and r , respectively, which yields the minimum channel gain. As for the other links, the distance is shorter, we consider scaling them with scale factors greater than 1. For the other channels, e.g. for the SI channel $\mathbf{H}_{l,l}$, the scale factor is defined as $\sqrt{\beta_{l,l}} = \sqrt{\frac{D_{l,r}}{D_{l,l}}}$, where $D_{l,r}$ is the distance between the centre of the FD nodes l and r and $D_{l,l}$ is the distance between the centre of the transmit and receive antenna arrays of the FD node l . Also for the other channels, between any node m and n , the channel gains are scaled with a scale factor which is expressed as a square root of the ratio of $D_{l,r}$ and $D_{m,n}$, where $D_{l,r}$ is still the distance between the centre of the FD nodes l and r and $D_{m,n}$ is the distance between the centres of the nodes constructing the link m and n , respectively.

6.1.2 Problem Formulation

Let $\tilde{\mathbf{H}}_{l,r}$, $\tilde{\mathbf{H}}_{r,l}$ and $\tilde{\mathbf{H}}_{l,l}$, $\tilde{\mathbf{H}}_{l,l}$ denote the effective direct and the SI channels for the two nodes, defined as

$$\tilde{\mathbf{H}}_{l,r} = \mathbf{H}_{l,r} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{H}_{i_r,r} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{H}_{i_l,r} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{H}_{i_l,i_r} \Phi_r \mathbf{H}_{i_r,r} \quad (6.6a)$$

$$\tilde{\mathbf{H}}_{r,l} = \mathbf{H}_{r,l} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{H}_{i_l,l} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{H}_{i_r,l} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{H}_{i_r,i_l} \Phi_l \mathbf{H}_{i_l,l} \quad (6.6b)$$

$$\tilde{\mathbf{H}}_{l,l} = \mathbf{H}_{l,l} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{H}_{i_l,l} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{H}_{i_r,l} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{H}_{i_r,i_l} \Phi_l \mathbf{H}_{i_l,l} \quad (6.6c)$$

$$\tilde{\mathbf{H}}_{r,r} = \mathbf{H}_{r,r} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{H}_{i_r,r} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{H}_{i_l,r} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{H}_{i_l,i_r} \Phi_r \mathbf{H}_{i_r,r} \quad (6.6d)$$

Let \mathbf{y}_l and \mathbf{y}_r denote the signal received at the FD nodes l and r , respectively, which can be written as

$$\mathbf{y}_l = \tilde{\mathbf{H}}_{l,r} \mathbf{V}_r \mathbf{s}_r + \tilde{\mathbf{H}}_{l,l} \mathbf{V}_l \mathbf{s}_l + \mathbf{n}_l, \quad (6.7a)$$

$$\mathbf{y}_r = \tilde{\mathbf{H}}_{r,l} \mathbf{V}_l \mathbf{s}_l + \tilde{\mathbf{H}}_{r,r} \mathbf{V}_r \mathbf{s}_r + \mathbf{n}_r, \quad (6.7b)$$

where $\mathbf{n}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{n}_r \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ denote the noise at the MIMO FD nodes l and r , respectively. Let \bar{k} denote the indices in the set \mathcal{F} except the element k . Let (\mathbf{R}_l) $\mathbf{R}_{\bar{l}}$ and (\mathbf{R}_r) $\mathbf{R}_{\bar{r}}$ denote the (signal and) interference plus noise covariance matrices, which can be stated as

$$\mathbf{R}_l = \underbrace{\tilde{\mathbf{H}}_{l,r} \mathbf{V}_r \mathbf{V}_r^H \tilde{\mathbf{H}}_{l,r}^H}_{\triangleq \mathbf{S}_l} + \tilde{\mathbf{H}}_{l,l} \mathbf{V}_l \mathbf{V}_l^H \tilde{\mathbf{H}}_{l,l}^H + \mathbf{I}, \quad (6.8a)$$

$$\mathbf{R}_r = \underbrace{\tilde{\mathbf{H}}_{r,l} \mathbf{V}_l \mathbf{V}_l^H \tilde{\mathbf{H}}_{r,l}^H}_{\triangleq \mathbf{S}_r} + \tilde{\mathbf{H}}_{r,r} \mathbf{V}_r \mathbf{V}_r^H \tilde{\mathbf{H}}_{r,r}^H + \mathbf{I}, \quad (6.8b)$$

$$\mathbf{R}_{\bar{l}} = \mathbf{R}_l - \mathbf{S}_l, \quad \mathbf{R}_{\bar{r}} = \mathbf{R}_r - \mathbf{S}_r. \quad (6.8c)$$

The WSR maximization problem for the NF-IRSs assisted mmWave MIMO FD system, under the sum-power constraint for the FD nodes and the unit-modulus constraint for the NF-IRSs can be written as

$$\max_{\mathbf{V}_l, \mathbf{V}_r, \Phi_l, \Phi_r} w_l \ln \det(\mathbf{R}_{\bar{l}}^{-1} \mathbf{R}_l) + w_r \ln \det(\mathbf{R}_{\bar{r}}^{-1} \mathbf{R}_r) \quad (6.9a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \preceq p_k, \quad \forall k \in \mathcal{F} \quad (6.9b)$$

$$|\theta_i^b| = 1, \quad \forall i \text{ and } \forall b \in \mathcal{I}, \quad (6.9c)$$

where w_l and w_r denote the rate weights for node l and r , respectively, p_l and p_r denote their total sum power constraint, respectively.

6.2 Beamforming

The WSR maximization problem stated above is non-concave and finding its global optimum is challenging. To provide its suboptimal solution, we adopt the weighted minimum mean squared error (WMMSE) method. Such approach can be adopted as it is well-known that maximizing the WSR is equivalent to minimizing the weighted MSE [64].

We assume that the MIMO FD nodes l and r deploy the digital combiners \mathbf{F}_l and \mathbf{F}_r to estimate the received data streams. i.e.,

$$\hat{\mathbf{s}}_l = \mathbf{F}_l \mathbf{y}_l, \quad \hat{\mathbf{s}}_r = \mathbf{F}_r \mathbf{y}_r. \quad (6.10a)$$

By using (6.10), the MSE matrices \mathbf{E}_l and \mathbf{E}_r for node l and r can be written as

$$\begin{aligned} \mathbf{MSE}_l = & \mathbb{E}[(\mathbf{F}_l \mathbf{y}_l - \mathbf{s}_l)(\mathbf{F}_l \mathbf{y}_l - \mathbf{s}_l)^H] = \mathbf{F}_l \tilde{\mathbf{H}}_{l,r} \mathbf{V}_r \mathbf{V}_r^H \tilde{\mathbf{H}}_{l,r}^H \mathbf{F}_l^H \\ & - \mathbf{F}_l \tilde{\mathbf{H}}_{l,r} \mathbf{V}_r + \mathbf{F}_l \tilde{\mathbf{H}}_{l,l} \mathbf{V}_l \mathbf{V}_l^H \tilde{\mathbf{H}}_{l,l}^H \mathbf{F}_l^H + \mathbf{F}_l \mathbf{F}_l^H \\ & - \mathbf{V}_r^H \tilde{\mathbf{H}}_{l,r}^H \mathbf{F}_l^H + \mathbf{I} \end{aligned} \quad (6.11a)$$

$$\begin{aligned} \mathbf{MSE}_r = & \mathbb{E}[(\mathbf{F}_r \mathbf{y}_r - \mathbf{s}_r)(\mathbf{F}_r \mathbf{y}_r - \mathbf{s}_r)^H] = \mathbf{F}_r \tilde{\mathbf{H}}_{r,l} \mathbf{V}_l \mathbf{V}_l^H \tilde{\mathbf{H}}_{r,l}^H \mathbf{F}_r^H \\ & - \mathbf{F}_r \tilde{\mathbf{H}}_{r,l} \mathbf{V}_l + \mathbf{F}_r \tilde{\mathbf{H}}_{r,r} \mathbf{V}_r \mathbf{V}_r^H \tilde{\mathbf{H}}_{r,r}^H \mathbf{F}_r^H + \mathbf{F}_r \mathbf{F}_r^H \\ & - \mathbf{V}_l^H \tilde{\mathbf{H}}_{r,l}^H \mathbf{F}_r^H + \mathbf{I} \end{aligned} \quad (6.11b)$$

We assume that the combiners \mathbf{F}_l and \mathbf{F}_r are optimized based on the MMSE criteria, and therefore, they can be obtained by solving the following optimization problem

$$\min_{\mathbf{F}_l, \mathbf{F}_r} \text{Tr}(\mathbf{MSE}_l) + \text{Tr}(\mathbf{MSE}_r) \quad (6.12)$$

which leads to the following optimal MMSE combiners

$$\mathbf{F}_l = \mathbf{V}_r^H \tilde{\mathbf{H}}_{l,r}^H (\tilde{\mathbf{H}}_{l,r} \mathbf{V}_r \mathbf{V}_r^H \tilde{\mathbf{H}}_{l,r}^H + \tilde{\mathbf{H}}_{l,l} \mathbf{V}_l \mathbf{V}_l^H \tilde{\mathbf{H}}_{l,l}^H + \mathbf{I})^{-1} \quad (6.13a)$$

$$\mathbf{F}_r = \mathbf{V}_l^H \tilde{\mathbf{H}}_{r,l}^H (\tilde{\mathbf{H}}_{r,l} \mathbf{V}_l \mathbf{V}_l^H \tilde{\mathbf{H}}_{r,l}^H + \tilde{\mathbf{H}}_{r,r} \mathbf{V}_r \mathbf{V}_r^H \tilde{\mathbf{H}}_{r,r}^H + \mathbf{I})^{-1} \quad (6.13b)$$

Assuming the combiners to be fixed according to (6.13) and by plugging these expressions in the error covariance matrices, it is immediate to show the error covariance matrices can be written as

$$\mathbf{E}_l = (\mathbf{I} + \mathbf{V}_r^H \tilde{\mathbf{H}}_{l,r}^H \mathbf{R}_l \tilde{\mathbf{H}}_{l,r} \mathbf{V}_r)^{-1} \quad (6.14)$$

$$\mathbf{E}_r = (\mathbf{I} + \mathbf{V}_l^H \tilde{\mathbf{H}}_{r,l}^H \mathbf{R}_r \tilde{\mathbf{H}}_{r,l} \mathbf{V}_l)^{-1} \quad (6.15)$$

Therefore, maximizing the WSR is equivalent to minimizing the MSE error. The WSR maximization problem now can be stated with the minimization of the mean squared error criteria with respect to the digital beamformers and IRSs as

$$\min_{\substack{\mathbf{V}_l, \mathbf{V}_r, \\ \Phi_l, \Phi_r}} \text{Tr}(\mathbf{W}_l \mathbf{E}_l) + \text{Tr}(\mathbf{W}_r \mathbf{E}_r) \quad (6.16a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \preceq p_k, \quad \forall k \in \mathcal{F} \quad (6.16b)$$

$$|\theta_i^b| = 1, \quad \forall i \text{ and } \forall b \in \mathcal{I}, \quad (6.16c)$$

where \mathbf{W}_i and \mathbf{W}_l are the constant weight matrices associated with the MIMO FD nodes i_l and i_r , respectively. The gradient of the WSR and the WMMSE optimization problems results to be the same if the weight matrices are chosen as [41]

$$\mathbf{W}_l = \frac{w_l}{\ln 2} \mathbf{E}_l^{-1}, \quad \mathbf{W}_r = \frac{w_r}{\ln 2} \mathbf{E}_r^{-1}. \quad (6.17)$$

6.2.1 Active Digital Beamforming

We assume the combiners and the IRSs' response to be fixed and consider the optimization of the digital beamformers \mathbf{V}_l and \mathbf{V}_r for the MIMO FD nodes i_l and i_r , respectively. The active digital beamforming optimization problem can be formally stated as follows

$$\min_{\mathbf{V}_l, \mathbf{V}_r} \quad \text{Tr}(\mathbf{W}_l \mathbf{E}_l) + \text{Tr}(\mathbf{W}_r \mathbf{E}_r) \quad (6.18a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \preceq p_k, \quad \forall k \in \mathcal{F}. \quad (6.18b)$$

By taking the partial derivative of the Lagrangian function of (6.18) with respect to the digital beamformers \mathbf{V}_l and \mathbf{V}_r , we obtain the following optimal beamformers

$$\mathbf{V}_l = \underbrace{(\tilde{\mathbf{H}}_{r,l}^H \mathbf{F}_r^H \mathbf{W}_r \mathbf{F}_r \mathbf{H}_{r,l} + \mathbf{H}_{l,l}^H \mathbf{F}_l^H \mathbf{W}_r \mathbf{F}_l \mathbf{H}_{l,l} + \mu_l \mathbf{I})^{-1} \tilde{\mathbf{H}}_{r,l}^H \mathbf{F}_r^H \mathbf{W}_r}_{\triangleq \mathbf{X}_l} \quad (6.19a)$$

$$\mathbf{V}_r = \underbrace{(\tilde{\mathbf{H}}_{l,r}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \mathbf{H}_{l,r} + \mathbf{H}_{r,r}^H \mathbf{F}_r^H \mathbf{W}_l \mathbf{F}_r \mathbf{H}_{r,r} + \mu_r \mathbf{I})^{-1} \tilde{\mathbf{H}}_{l,r}^H \mathbf{F}_l^H \mathbf{W}_l}_{\triangleq \mathbf{X}_r} \quad (6.19b)$$

where μ_l and μ_r denote the Lagrange multipliers for the sum-power constraint for the MIMO FD nodes l and r , respectively. The Lagrange multipliers can be calculated while meeting the sum-power constraints. Namely, to find the Lagrange multiplier, we can consider the singular value decomposition of $\mathbf{X}_l = \mathbf{U}_l \mathbf{\Sigma}_l \mathbf{U}_l^H$ and $\mathbf{X}_r = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{U}_r^H$ and write the power constraints in (6.19), after simple steps, as

$$\text{Tr}(\mathbf{V}_l \mathbf{V}_l^H) = \sum_{i=1}^{N_l} \frac{\mathbf{G}_l(i, i)}{(\mu_l + \mathbf{\Sigma}_l(i, i))^2} = p_l, \quad (6.20a)$$

$$\text{Tr}(\mathbf{V}_r \mathbf{V}_r^H) = \sum_{i=1}^{N_r} \frac{\mathbf{G}_r(i, i)}{(\mu_r + \mathbf{\Sigma}_r(i, i))^2} = p_r, \quad (6.20b)$$

where the matrices \mathbf{G}_l and \mathbf{G}_r are defined as

$$\mathbf{G}_l = \mathbf{U}_l^H \tilde{\mathbf{H}}_{r,l}^H \mathbf{F}_r^H \mathbf{W}_r \mathbf{F}_r \tilde{\mathbf{H}}_{r,l} \mathbf{U}_l, \quad (6.21a)$$

$$\mathbf{G}_r = \mathbf{U}_r^H \tilde{\mathbf{H}}_{l,r}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \tilde{\mathbf{H}}_{l,r} \mathbf{U}_r. \quad (6.21b)$$

To find the optimal values of μ_l and μ_r , a linear search method can be adopted and we consider finding the multipliers with the Bisection method. If the values of the Lagrange multipliers results to be negative, then we assign them the value zero.

6.2.2 Passive Beamforming with the IRSs

We assume the digital combiners and beamformers to be fixed and consider the optimization of the Φ_r and Φ_l to optimize the passive beamforming response of the IRSs. To highlight their dependence on the WSR, we first write the effective channel responses either as

$$\tilde{\mathbf{H}}_{l,r} = \mathbf{A}_{1,r} + \mathbf{A}_{2,r} \Phi_r \mathbf{H}_{i_r,r}, \quad \tilde{\mathbf{H}}_{r,l} = \mathbf{B}_{1,r} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{B}_{2,r}, \quad (6.22a)$$

$$\tilde{\mathbf{H}}_{l,l} = \mathbf{C}_{1,r} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{C}_{2,r}, \quad \tilde{\mathbf{H}}_{r,r} = \mathbf{D}_{1,r} + \mathbf{D}_{2,r} \Phi_r \mathbf{H}_{i_r,r}. \quad (6.22b)$$

to optimize Φ_r or as

$$\tilde{\mathbf{H}}_{l,r} = \mathbf{A}_{1,l} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{A}_{2,l}, \quad \tilde{\mathbf{H}}_{r,l} = \mathbf{B}_{1,l} + \mathbf{B}_{2,l} \Phi_l \mathbf{H}_{i_l,l}, \quad (6.23a)$$

$$\tilde{\mathbf{H}}_{l,l} = \mathbf{C}_{1,l} + \mathbf{C}_{2,l} \Phi_l \mathbf{H}_{i_l,l}, \quad \tilde{\mathbf{H}}_{r,r} = \mathbf{D}_{1,l} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{D}_{2,l}. \quad (6.23b)$$

to highlight the effective channel responses as a function of Φ_l , where the auxiliary matrices appearing in (6.22)-(6.23a) are defined as

$$\mathbf{A}_{1,r} = \mathbf{H}_{l,r} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{H}_{i_l,r}, \quad \mathbf{A}_{2,r} = \mathbf{H}_{l,i_r} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{H}_{i_l,i_r} \quad (6.24a)$$

$$\mathbf{B}_{1,r} = \mathbf{H}_{r,l} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{H}_{i_l,l}, \quad \mathbf{B}_{2,r} = \mathbf{H}_{i_r,l} + \mathbf{H}_{i_r,i_l} \Phi_l \mathbf{H}_{i_l,l}, \quad (6.24b)$$

$$\mathbf{C}_{1,r} = \mathbf{H}_{l,l} + \mathbf{H}_{l,i_l} \Phi_l \mathbf{H}_{i_l,l}, \quad \mathbf{C}_{2,r} = \mathbf{H}_{i_r,l} + \mathbf{H}_{i_r,i_l} \Phi_l \mathbf{H}_{i_l,l}, \quad (6.24c)$$

$$\mathbf{D}_{1,r} = \mathbf{H}_{r,r} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{H}_{i_l,r}, \quad \mathbf{D}_{2,r} = \mathbf{H}_{r,i_r} + \mathbf{H}_{r,i_l} \Phi_l \mathbf{H}_{i_l,i_r}. \quad (6.24d)$$

$$\mathbf{A}_{1,l} = \mathbf{H}_{l,r} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{H}_{i_r,r}, \quad \mathbf{A}_{2,l} = \mathbf{H}_{i_l,r} + \mathbf{H}_{i_l,i_r} \Phi_r \mathbf{H}_{i_r,r} \quad (6.24e)$$

$$\mathbf{B}_{1,l} = \mathbf{H}_{r,l} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{H}_{i_r,l}, \quad \mathbf{B}_{2,l} = \mathbf{H}_{r,i_l} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{H}_{i_r,i_l}, \quad (6.24f)$$

$$\mathbf{C}_{1,l} = \mathbf{H}_{l,l} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{H}_{i_r,l}, \quad \mathbf{C}_{2,l} = \mathbf{H}_{l,i_l} + \mathbf{H}_{l,i_r} \Phi_r \mathbf{H}_{i_r,i_l}, \quad (6.24g)$$

$$\mathbf{D}_{1,l} = \mathbf{H}_{r,r} + \mathbf{H}_{r,i_r} \Phi_r \mathbf{H}_{i_r,r}, \quad \mathbf{D}_{2,l} = \mathbf{H}_{i_l,r} + \mathbf{H}_{i_l,i_r} \Phi_r \mathbf{H}_{i_r,r}. \quad (6.24h)$$

Let $\mathbf{S}_r, \mathbf{Z}_r, \mathbf{T}_r, \mathbf{S}_l, \mathbf{Z}_l, \mathbf{T}_l$ denote additional auxiliary matrices defined as

$$\begin{aligned} \mathbf{S}_r = & \mathbf{H}_{i_r,r} \mathbf{V}_r \mathbf{V}_r^H \mathbf{A}_{1,r}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \mathbf{A}_{2,r} - \mathbf{H}_{i_r,r} \mathbf{V}_r \mathbf{W}_l \mathbf{F}_l \mathbf{A}_{2,r} + \mathbf{C}_{2,r} \mathbf{V}_l \mathbf{V}_l^H \mathbf{C}_{1,r}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \mathbf{H}_{l,i_r} \\ & - \mathbf{B}_{2,r} \mathbf{V}_l \mathbf{W}_r \mathbf{F}_r \mathbf{H}_{r,i_r} + \mathbf{B}_{2,r} \mathbf{V}_l \mathbf{V}_l^H \mathbf{B}_{1,r}^H \mathbf{F}_r^H \mathbf{W}_r \mathbf{F}_r \mathbf{H}_{r,i_r} \\ & + \mathbf{H}_{i_r,r} \mathbf{V}_r \mathbf{V}_r^H \mathbf{D}_{1,r}^H \mathbf{F}_r^H \mathbf{W}_r \mathbf{F}_r \mathbf{D}_{2,r}, \end{aligned} \quad (6.25a)$$

$$\begin{aligned} \mathbf{Z}_r = & \mathbf{H}_{r,i_r}^H \mathbf{F}_r^H \mathbf{W}_r \mathbf{F}_r \mathbf{H}_{r,i_r} + \mathbf{D}_{2,r}^H \mathbf{A}_{2,r}^H \mathbf{W}_r \mathbf{F}_r \mathbf{D}_{2,r} + \mathbf{A}_{2,r}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \mathbf{A}_{2,r} \\ & + \mathbf{H}_{l,i_r}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \mathbf{H}_{l,i_r}, \end{aligned} \quad (6.25b)$$

$$\mathbf{T}_r = \mathbf{B}_{2,r} \mathbf{V}_l \mathbf{V}_l^H \mathbf{B}_{2,r}^H + \mathbf{H}_{i_r,r} \mathbf{V}_r \mathbf{V}_r^H \mathbf{H}_{i_r,r}^H + \mathbf{H}_{i_r} \mathbf{V}_r \mathbf{V}_r^H \mathbf{H}_{i_r,r}^H + \mathbf{C}_{2,r} \mathbf{V}_l \mathbf{V}_l^H \mathbf{C}_{2,r}^H. \quad (6.25c)$$

$$\begin{aligned} \mathbf{S}_l = & \mathbf{A}_{2,l} \mathbf{V}_r \mathbf{V}_r^H \mathbf{A}_{1,l}^H \mathbf{F}_l^H \mathbf{W}_l \mathbf{F}_l \mathbf{H}_{l,i_l} + \mathbf{H}_{i_l,l} \mathbf{V}_l \mathbf{V}_l^H \mathbf{C}_{1,l}^H \mathbf{A}_l^H \mathbf{W}_l \mathbf{A}_l \mathbf{C}_{2,l} - \mathbf{A}_{2,l} \mathbf{V}_r \mathbf{W}_l \mathbf{A}_l \mathbf{H}_{l,i_l} \\ & + \mathbf{H}_{i_l,l} \mathbf{V}_l \mathbf{V}_l^H \mathbf{B}_{1,l}^H \mathbf{A}_l^H \mathbf{W}_r \mathbf{A}_r \mathbf{B}_{2,l} + \mathbf{D}_{2,l} \mathbf{V}_r \mathbf{V}_r^H \mathbf{D}_{1,l}^H \mathbf{A}_r^H \mathbf{W}_r \mathbf{A}_r \mathbf{H}_{r,i_l} \\ & - \mathbf{H}_{i_l,l} \mathbf{V}_l \mathbf{W}_r \mathbf{A}_r \mathbf{B}_{2,l}, \end{aligned} \quad (6.26a)$$

$$\mathbf{Z}_l = \mathbf{H}_{l,i_l}^H \mathbf{A}_l^H \mathbf{W}_l \mathbf{A}_l \mathbf{H}_{l,i_l} + \mathbf{C}_{2,l}^H \mathbf{A}_l^H \mathbf{W}_l \mathbf{A}_l \mathbf{C}_{2,l} + \mathbf{A}_l^H \mathbf{W}_r \mathbf{A}_r \mathbf{B}_{2,l} + \mathbf{H}_{r,i_l}^H \mathbf{A}_r^H \mathbf{W}_r \mathbf{H}_{r,i_l}, \quad (6.26b)$$

$$\mathbf{T}_l = \mathbf{A}_{2,l} \mathbf{V}_r \mathbf{V}_r^H \mathbf{A}_{2,l}^H + \mathbf{H}_{i_l,l} \mathbf{V}_l \mathbf{V}_l^H \mathbf{H}_{i_l,l}^H + \mathbf{H}_{i_l,l} \mathbf{V}_l \mathbf{V}_l^H \mathbf{H}_{i_l,l}^H + \mathbf{D}_{2,l} \mathbf{V}_r \mathbf{V}_r^H \mathbf{D}_{2,l}^H. \quad (6.26c)$$

By substituting (6.25) and (6.26) in the expressions of the error covariance matrices, the minimization of the MSE optimization problem with respect to Φ_r and Φ_l , can be formally stated as

$$\min_{\Phi_r} \quad \text{Tr}(\Phi_r^H \mathbf{Z}_r \Phi_r \mathbf{T}_r) + \text{Tr}(\Phi_r^H \mathbf{S}_r^H) + \text{Tr}(\Phi_r \mathbf{S}_r) + c_r \quad (6.27a)$$

$$\text{s.t.} \quad |\theta_i^r| = 1, \quad \forall i, \quad (6.27b)$$

$$\min_{\Phi_l} \quad \text{Tr}(\Phi_l^H \mathbf{Z}_l \Phi_l \mathbf{T}_l) + \text{Tr}(\Phi_l^H \mathbf{S}_l^H) + \text{Tr}(\Phi_l \mathbf{S}_l) + c_l \quad (6.27c)$$

$$\text{s.t.} \quad |\theta_i^l| = 1, \quad \forall i, \quad (6.27d)$$

where c_l and c_r denote constant terms which are independent of Φ_l and Φ_r , respectively. We remark that when solving (6.27a)-(6.27b), Φ_l is assumed to fixed, and similary, when solving (6.27c)-(6.27d), Φ_r will be assumed to fixed.

Recall that ϕ_l and ϕ_r are vectors containing the diagonal elements of Φ_l and Φ_r , respectively. As the off-diagonal elements of the Φ_r are zero, we wish to maximize the WSR or minimize the MSE with respect to ϕ_l and ϕ_r . For such purpose, we use the identity 1.10.6 from [122], to restate the first trace term appearing in (6.27a) and (6.27c) as

$$\text{Tr}(\Phi_r^H \mathbf{Z}_r \Phi_r \mathbf{T}_r) = \phi_r^H \Sigma_r \phi_r, \quad \text{where } \Sigma_r = \mathbf{Z}_r \odot \mathbf{T}_r^T, \quad (6.28a)$$

$$\text{Tr}(\Phi_l^H \mathbf{Z}_l \Phi_l \mathbf{T}_l) = \phi_l^H \Sigma_l \phi_l, \quad \text{where } \Sigma_l = \mathbf{Z}_l \odot \mathbf{T}_l^T. \quad (6.28b)$$

Let \mathbf{s}_r and \mathbf{s}_l denote the vectors containing the diagonal elements of the matrices \mathbf{S}_r (6.25a) and \mathbf{S}_l (6.26a), respectively, given as $\mathbf{s}_r = [\mathbf{S}_r(1,1), \dots, \mathbf{S}_r(R_r, R_c)]^T$ and $\mathbf{s}_l = [\mathbf{S}_l(1,1), \dots, \mathbf{S}_l(R_r, R_c)]^T$. The second and third terms appearing in (6.27a) and (6.27c) can be restated as

$$\text{Tr}(\Phi_r^H \mathbf{S}_r^H) = \mathbf{s}_r^H \phi_r^*, \quad \text{Tr}(\Phi_r \mathbf{S}_r) = \mathbf{s}_r^T \phi_r. \quad (6.29a)$$

$$\text{Tr}(\Phi_l^H \mathbf{S}_l^H) = \mathbf{s}_l^H \phi_l^*, \quad \text{Tr}(\Phi_l \mathbf{S}_l) = \mathbf{s}_l^T \phi_l. \quad (6.29b)$$

By using the aforementioned properties, the optimization problem (6.27a) and (6.27c) to optimize ϕ_l and ϕ_r , respectively, can be written as

$$\min_{\Phi_r} \quad \phi_r^H \Sigma_r \phi_r + \mathbf{s}_r^H \phi_r^* + \mathbf{s}_r^T \phi_r, \quad (6.30a)$$

$$\text{s.t.} \quad |\phi_r(i)| = 1, \quad \forall i, \quad (6.30b)$$

$$\min_{\Phi_l} \quad \phi_l^H \Sigma_l \phi_l + \mathbf{s}_l^H \phi_l^* + \mathbf{s}_l^T \phi_l \quad (6.30c)$$

$$\text{s.t.} \quad |\phi_l(i)| = 1, \quad \forall i. \quad (6.30d)$$

which is non-convex due to the unit-modulus constraint. To render a feasible solution, we adopt the majorization-maximization optimization method [123]. Its objective is solve a difficult problem by constructing a series of more tractable problems. We carry out its explanation only for ϕ_r and a similar reasoning follows also for ϕ_l . Let $f(\phi_r^{(n)})$ denote the value of the objective function as a function of ϕ_r at the n -th iteration. We proceed by constructing an upper bound on the objective function (6.27a)-(6.27b), which is denoted with $g(\phi_r|\phi_r^{(n)})$. Instead of directly solving the optimization problem (6.27a)-(6.27b), we construct an approximate problem by using $g(\phi_r|\phi_r^{(n)})$ at the $n+1$ -th iteration. If $g(\phi_r|\phi_r^{(n)})$ satisfies the following conditions

1. $g(\phi_r^{(n)}|\phi_r^{(n)}) = f(\phi_r^{(n)})$,
2. $\nabla_{\phi_r} g(\phi_r|\phi_r^{(n)})|_{\phi_r=\phi_r^{(n)}} = \nabla_{\phi_r} f(\phi_r^{(n)})|_{\phi_r=\phi_r^{(n)}}$,
3. $g(\phi_r|\phi_r^{(n)}) \geq f(\phi_r^{(n)})$,

then the sequence of the solutions obtained at each iteration will result in a monotonically decreasing objective function and converge to the actual solution of $f(\cdot)$. The first two conditions dictate that $g(\phi_r|\phi_r^{(n)})$ and its first order gradient should be the same and the third condition state that it should construct an upper bound for the original function. It has been shown in [124] that a simple upper bound $g(\phi_r|\phi_r^{(n)})$ or $g(\phi_l|\phi_l^{(n)})$ for Φ_l for the problems of the form (6.27c)-(6.27d) is given by

$$g(\phi_r|\phi_r^{(n)}) = 2\text{Re}\{\mathbf{s}_r^H \mathbf{q}_r^{(n)}\} + o_r, \quad (6.31a)$$

$$g(\phi_l|\phi_l^{(n)}) = 2\text{Re}\{\mathbf{s}_l^H \mathbf{q}_l^{(n)}\} + o_l, \quad (6.31b)$$

where o_r and o_l denote constant terms, and $\mathbf{q}_r^{(n)}$ and $\mathbf{q}_l^{(n)}$ are given by

$$\mathbf{q}_r^{(n)} = (\lambda_r^{\max} \mathbf{I} - \Sigma_r) \phi_r^{(n)} - \mathbf{s}_r^*, \quad (6.32a)$$

$$\mathbf{q}_l^{(n)} = (\lambda_l^{\max} \mathbf{I} - \Sigma_l) \phi_l^{(n)} - \mathbf{s}_l^*, \quad (6.32b)$$

and λ_r^{\max} and λ_l^{\max} denote the maximum eigenvalues of Σ_r and Σ_l , respectively. The optimization problems (6.27a)-(6.27b) and (6.27c)-(6.27d), by ignoring all the terms which are constant, can be restated by using the upper bounds (6.31a) and (6.31b), respectively, as

$$\min_{\phi_r} 2\text{Re}\{\mathbf{s}_r^H \mathbf{q}_r^{(n)}\}, \quad (6.33a)$$

$$\text{s.t. } |\phi_r(i)| = 1, \quad \forall i, \quad (6.33b)$$

$$\min_{\phi_l} 2\text{Re}\{\mathbf{s}_l^H \mathbf{q}_l^{(n)}\}, \quad (6.33c)$$

$$\text{s.t. } |\phi_l(i)| = 1, \quad \forall i. \quad (6.33d)$$

Algorithm 11 Optimization of IRS $i \in \mathcal{I}$

Given: $\phi_i(0)$.

Initialize: iteration index $n = 1$, accuracy ϵ .

Evaluate: $f(\phi_i(0))$.

Repeat until convergence

 Calculate $\mathbf{q}_i^{(n)} = (\lambda_i^{\max} \mathbf{I} - \Sigma_i) \phi_i^{(n)} - \mathbf{s}_i^*$

 Update $\phi_i^{(n+1)}$ with $\phi_i^{(n+1)} = \phi_r^{(n+1)} = e^{i\angle \mathbf{q}_i^{(n)}}$.

if $|f(\phi_i^{(n+1)}) - f(\phi_i^{(n)})|/f(\phi_i^{(n+1)}) \leq \epsilon$

 Stop and return $\phi_i^{(n+1)}$.

else $n=n+1$ and repeat.

Algorithm 12 WSR maximization for MIMO FD with NF-IRSs

Given: The CSI and rate weights.

Initialize iteration index n , accuracy ϵ , beamformers and combiners.

Repeat until convergence

 for $\forall i \in \mathcal{F}$

 Update \mathbf{F}_i with (6.13).

 Update \mathbf{W}_i with (6.17).

 Update \mathbf{V}_i with (6.19).

 Update Φ_i with Algorithm 11.

if convergence condition is satisfied

 Stop and return the optimized variables.

else repeat.

By solving (6.33)-(6.33b) and (6.33c)-(6.33d) we get that the optimal ϕ_r and ϕ_l at $n + 1$ -th iteration is given as

$$\phi_r^{(n+1)} = e^{i\angle \mathbf{q}_r^{(n)}}. \quad (6.34a)$$

$$\phi_l^{(n+1)} = e^{i\angle \mathbf{q}_l^{(n)}}. \quad (6.34b)$$

At each iteration, when the digital beamformers are fixed, optimization of Φ_r and Φ_l consists in updating their response iteratively until convergence by solving a series of more tractable optimization problems. The procedure for the optimization of the phase response of IRS $i \in \mathcal{I}$ is stated in Algorithm 1, which needs to be iterated for each update of the digital beamformers and combiners. The overall procedure to jointly optimize the digital beamformers, combiners and the response of the NF-IRSs is formally stated in Algorithm 2.

6.2.3 Convergence

The proof of the convergence of the algorithm is based on the proof of a more general equivalent optimization problem, which includes the MSE weights and receive filters as

new optimization variables in addition to digital beamformers and the IRSs' response. The global optimization problem as function of the new variables can be formally stated as

$$\min_{\substack{\mathbf{F}_l, \mathbf{F}_r, \mathbf{V}_r \\ \mathbf{V}_l, \Phi_l, \Phi_r}} \sum_{f \in \mathcal{F}} \text{Tr}(\mathbf{W}_f \mathbf{MSE}_f) - w_f \log \det\left(\frac{\ln 2}{w_f} \mathbf{W}_f\right) + d_f \frac{w_f}{\ln 2} \quad (6.35)$$

By fixing the digital beamformers, IRSs' response and the weight matrices, the combiners are chosen to be the MMSE combiners as (6.13). By substituting the optimal MMSE receivers given in (6.35) in the \mathbf{MSE}_l and \mathbf{MSE}_r in (6.11a) and (6.11b), respectively, yields a new cost function

$$\min_{\substack{\mathbf{V}_r, \mathbf{V}_l, \\ \Phi_l, \Phi_r}} \sum_{f \in \mathcal{F}} \text{Tr}(\mathbf{W}_f \mathbf{E}_f) - w_f \log \det\left(\frac{\ln 2}{w_f} \mathbf{W}_f\right) + d_f \frac{w_f}{\ln 2}. \quad (6.36)$$

Optimization of (6.36) with respect to the weight matrix \mathbf{W}_f , yields the solution $\mathbf{W}_f = \frac{w_f}{\ln 2} \mathbf{E}_f^{-1}$. By plugging the optimized weight matrices in (6.36) leads to a new cost function, given as

$$\min_{\substack{\mathbf{V}_l, \mathbf{V}_r, \\ \Phi_l, \Phi_r}} \sum_{f \in \mathcal{F}} -w_f \ln \det((\mathbf{E}_f)^{-1}) \quad (6.37a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \preceq p_k, \quad \forall k \in \mathcal{F} \quad (6.37b)$$

$$|\theta_i^b| = 1, \quad \forall i \text{ and } b \in \mathcal{I}, \quad (6.37c)$$

which is exactly same as the original WSR cost function (6.9). Every update of the digital beamformers \mathbf{V}_l and \mathbf{V}_r and the IRSs' phase response by minimizing the MSE leads to a monotonic increase in the WSR, which assures convergence of the proposed joint active and passive beamforming design.

A typical convergence behaviour of the proposed beamforming design for the NF-IRSs assisted FD systems is highlighted in Fig. 6.3.

6.2.4 Simulation Results

In this section, we present simulation results to evaluate the performance of the proposed joint active and passive beamforming design to analyze its potential to replace HYBF and combining simultaneously for the FD systems while shaping the SI channel for SI management.

For comparison, we define the following benchmark schemes:

- mMIMO HD system with fully digital beamforming capability with 100 transmit and 50 receive antennas, denoted with HD-100 × 50.
- mMIMO FD system with fully digital beamforming capability with 100 transmit and 50 receive antennas, denoted with FD-100 × 50.

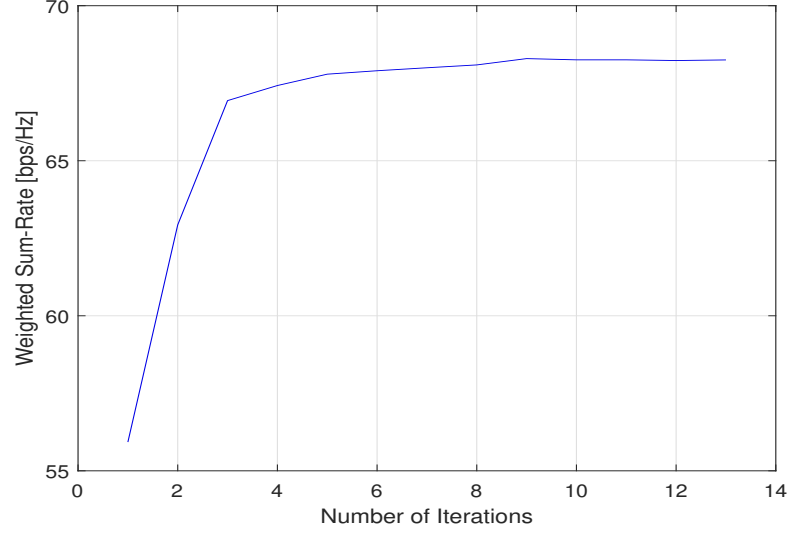


Figure 6.3: Typical convergence behaviour of the proposed joint active and passive beamforming design.

Note that digital beamforming designs represent an upper bound for the HYBF designs. Therefore we consider comparing the performance with the mMIMO fully digital systems only, which represent the maximum achievable gain for the HYBF designs. We define the SNR for the NF-IRSs assisted FD system as $\text{SNR} = p_k/\sigma_k^2 = p_l/\sigma_l^2$, where p_i and σ_i^2 denote the total transmit power and noise variance for node i , respectively, where $i \in \mathcal{F}$ or $i \in \mathcal{I}$. We assume that the systems operate at the frequency of 30 GHz, i.e., $\lambda = 10\text{mm}$. The noise σ_i^2 is set to 1, and the total transmit power is chosen to meet the desired SNR. For the NF-IRSs assisted FD system, the number of transmit antennas are chosen to be 5 times less, i.e. only $M_l = M_r = 20$ transmit and $N_l = N_r = 10$ receive antennas, respectively, and the number of data streams are set to be $d_l = d_r = 2$ for both the nodes. The FD nodes are assumed to be $D_{l,r} = 200\text{m}$ far, and the NF-IRSs are placed at the minimum distance of 3m from their FD nodes, which is varied up to 90 m. We assume the NF-IRSs for both nodes to be of the same size, i.e., $L_r = R_r$ and $L_c = R_c$, which is denoted with $\text{IRS} - X \times X$, and for performance evaluation, we choose $X = 10, 20$ or 30 . We assume ULAs for both the FD nodes and their transmit and the receive array are assumed to be separated with distance $D_b = 20\text{ cm}$ with a relative angle $\Theta_b = 90^\circ$ and $r_{m,n}$ in (6.3) is set given D_b and Θ_b as in (9) [54]. The Rician factor is chosen to be $\kappa_b = 1$ and the rate weights are set to be $w_l = w_r = 1$. The number of paths and number of clusters are chosen to be $N_{c,b} = N_{p,b} = 3$ and the AOA $\theta_r^{n_p, n_c}$ and AOD $\phi_l^{n_p, n_c}$ are assumed to be uniformly distributed in the interval $\mathcal{U} \sim [-30^\circ, 30^\circ]$. The digital beamformers are initialized as the dominant eigenvectors of effective channel covariance matrices, and the response of the NF-IRSs is initialized with random phases. The results reported herein are averaged over 200 channel realizations. Let \hat{x}, \hat{y} and \hat{z} denote the 3 versors on the three dimensional space. Any point on the 3 dimensional vector space can be written as $(x\hat{x}, y\hat{y}, z\hat{z})$. The FD nodes with ULAs are aligned with \hat{x} direction and the first transmit antenna of the FD nodes l and r are assumed to be

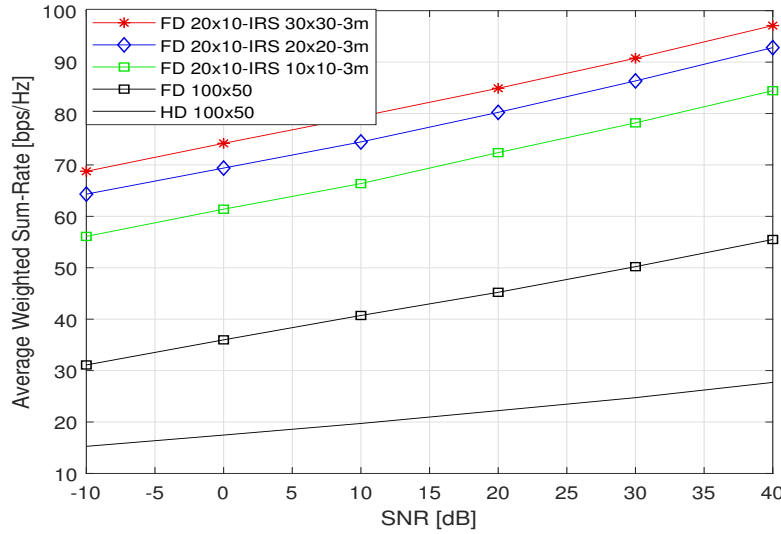


Figure 6.4: Average WSR as a function of SNR with NF-IRSs of size 30×30 , 20×20 and 10×10 , placed at distance of 3m.

placed in the positions $(0, 0, 0)$ and $(0, D_{l,r}\hat{y}, 0)$, respectively, with other antennas placed in positions $x+$. The NF-IRSs are placed on the plane identified with the versors $(0, \hat{y}, \hat{z})$, with reflecting elements pointing towards the direction $(\hat{x}, 0, 0)$.

Fig. 6.4 shows the achieved average WSR as a function of the SNR for the FD systems assisted with the NF-IRSs placed at 3m far from the transmit array. We can see that despite using 5 times fewer antennas than the mMIMO FD systems, the proposed idea has the potential to significantly outperform the HD system and the classical mMIMO fully digital FD system. The FD systems are well-known to double the spectral efficiency. However, as shown in Fig. 6.4, they can achieve significantly higher gains. For example, an FD system with NF-IRSs of size 30×30 can achieve 5 times higher gains than the traditional mMIMO HD system while operating with 5 times fewer antennas. Fig. 6.4 exhibits that the achievable gains of the FD systems are limited only by the size of the NF-IRSs, and having larger NF-IRSs will lead to higher gains. Fig. 6.5 shows the performance of the proposed design as a function of the SNR with NF-IRSs placed at 30m far from the transmit array. We can see that when the NF-IRSs are placed far from the FD node, they achieve significantly less gain. The reason lies in the fact that the IRSs placed far from the FD node cannot aid in shaping the SI channel, which is handled only with the digital beamformers, which leads to less achievable gain.

Fig. 6.6 shows the performance of the proposed joint active and passive beamforming design as a function of the distance between the FD node and their NF-IRSs, which is varied in the interval 3 – 90m, with the two FD nodes operating at the distance of 200m at SNR= 0 dB. It is visible from Fig. 6.6, that the FD systems benefit the most with NF-IRSs instead of the FF-IRSs by bringing significant improvement in terms of the achievable WSR compared to the traditional mMIMO FD systems. Placing the IRSs far from the FD systems leads to small performance gains as the FD systems improve

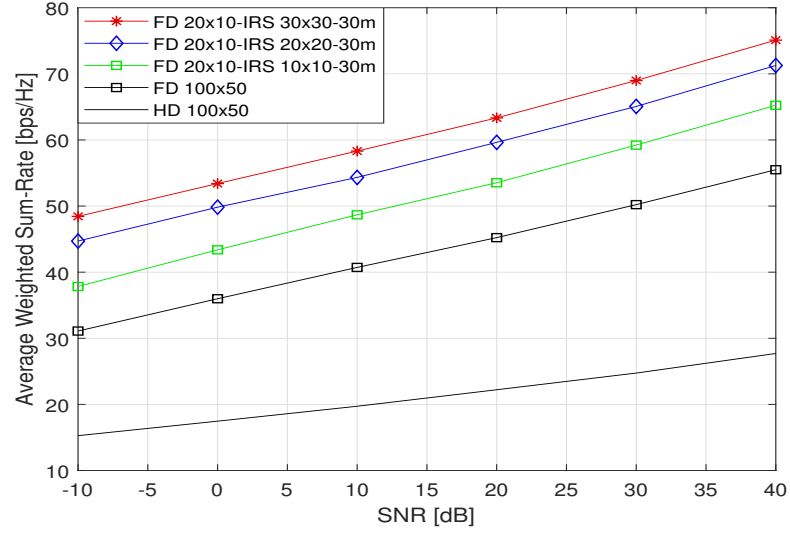


Figure 6.5: Average WSR as a function of SNR with NF-IRSs of size 30×30 , 20×20 and 10×10 , placed at distance of 30m.

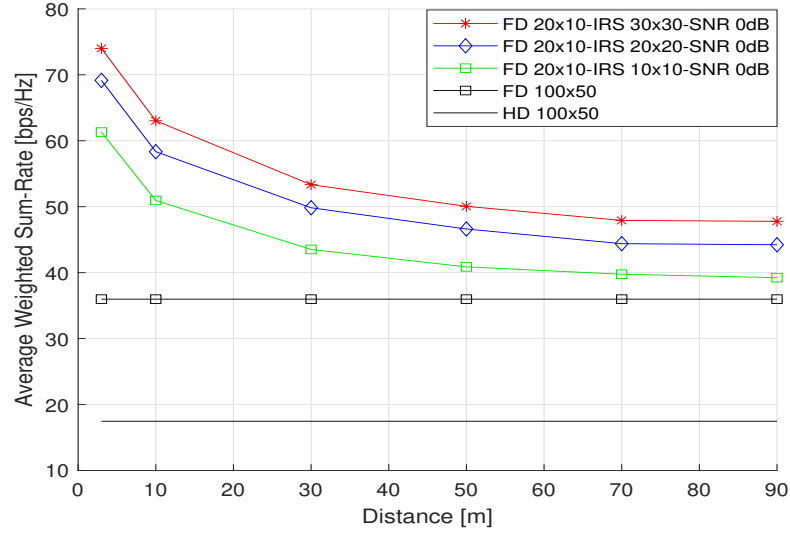


Figure 6.6: Average WSR as a function of the distance between the FD nodes and the NF-IRSs at SNR= 0 dB.

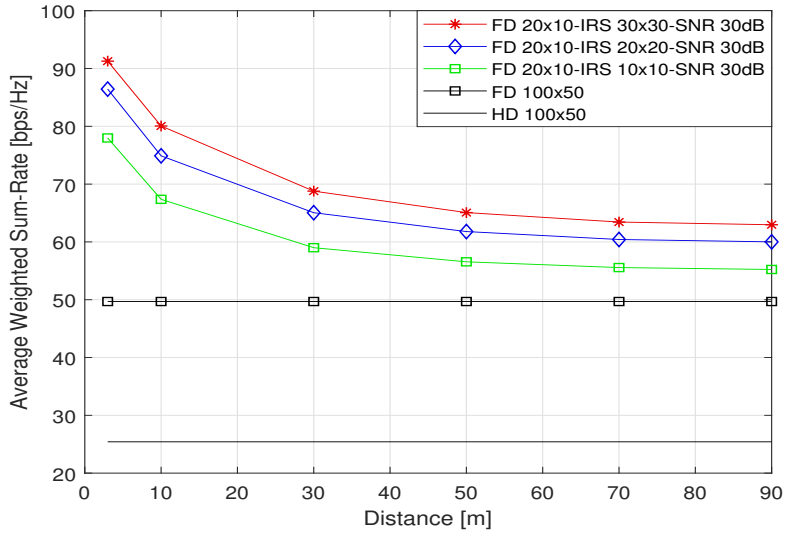


Figure 6.7: Average WSR as a function of the distance between the FD nodes and the NF-IRSs at SNR= 30 dB.

only the quality of the direct links and cannot assist in shaping the SI channel. Fig. 6.7 shows the performance as a function of the distance between the NF-IRSs from their FD node at SNR= 30dB. It is clearly visible that the regardless of the SNR at which the NF-IRSs assisted FD systems operate, the distance between the FD nodes and their NF-IRSs dictates the maximum achievable gain.

6.3 Conclusions

This chapter has introduced the idea of NF-IRSs assisted FD systems for enhancing the performance of the traditional FD systems while reducing the hardware cost for the SI cancellation and the number of active elements in the system. A novel joint active and passive beamforming design is proposed for WSR maximization based on the minimization of the MSE. Simulation results show that the proposed idea has the potential to leverage to the maximum the achievable gains of the FD system while operating with minimum cost. Results also show that the distance between the FD nodes and the IRSs can severely affect the FD systems' performance. Therefore, it is not desirable for the next generation of wireless FD systems to be aided with the FF-IRSs but only with the NF-IRSs.

Chapter 7

Dynamic TDD: A Special Case of Multi-Cell Full Duplex Systems

7.1 Introduction and Motivation

In HD cellular networks, UL and DL transmissions occur either at different times or over different frequency bands. In particular, with TDD, the BSs alternates time periods when it receives signals from its served users in the UL, or transmits to users in the DL. In advanced long-term evolution (A-LTE) networks, a fixed configuration of the TDD periods is used, to be chosen among a given set. In the new radio (NR) standard for fifth-generation (5G) networks, this rigid scheme is broken, and DTDD [125] enables cells to independently set the UL/DL configuration in each slot.

As well-discussed in Chapter 5, FD multi-cell FD systems are severely affected by the CI. DTDD represents a special case of the FD systems, which introduces UL-to-DL, i.e., from UL users towards DL users, and DL-to-UL, i.e., from BS-to-BS, CI among cells, as shown in Fig. 7.1(a)-7.1(b). As for FD systems, DTDD also requires CI management techniques to outperform the classical TDD systems. Fortunately, intelligent scheduling of the different cells in DTDD is possible, which if done correctly, can significantly limit the CI contributions.

In this chapter, we consider the joint scheduling and power allocation problem, aiming at a maximizing the average WSR of the network: this is mixed-integer and non-convex (MINC), which turns out to be NP-hard [126]. Thus, we decouple the problem: first, power is allocated with a low complexity, greedy, and decentralized algorithm, then, the mode is scheduled by playing a game among cells. We consider a non-cooperative game, wherein each base station plays an action (the UL or DL mode) with a given probability distribution, so as to maximize its utility function, given the choices made by its opponents. The neighboring cells' choices are crucial for the resulting utility, due to CI. The game is solved according to the mixed-strategy Nash equilibrium (MS-NE) concept, which always exists for finite games [127]. We consider two payoffs, using different models for the CI and requiring different overheads for the information exchange among cells. The proposed scheduling is performed at each slot independently, a solution that can

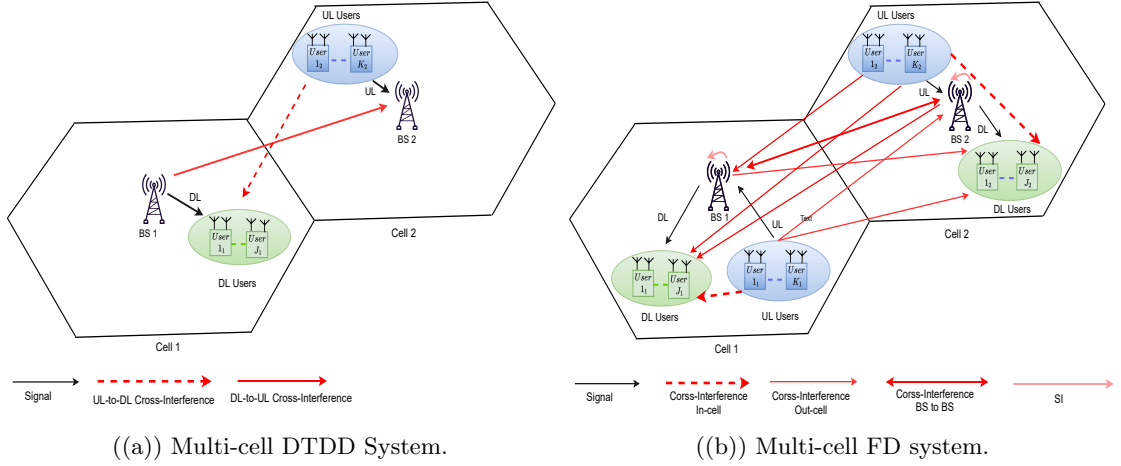


Figure 7.1: Multi-cell DTDD and FD systems, where both are affected by the CI.

be included in future releases of NR, as currently only a limited number of slot mode configurations can be selected.

Simulation results are presented for a dense, outdoor, and small-cells scenario, with line-of-sight (LoS) links. To our best knowledge, this is the first contribution to provide results for the most challenging DTDD scenario of [128], without resorting to clustering.

In summary, the contributions of our work are:

- the definition of the joint scheduling and power allocation problem for DTDD for average weighted sum-rate maximization;
- the decoupling of the power allocation and scheduling problem in DTDD; we obtain a much simpler system of (nonlinear) equations suited for a very efficient solution, without iterations or tuning of multiple parameters, as in [129];
- a novel scheduling procedure, implemented as a game among cells; the MS-NE solution is computed once over many slots, as long as the channel can be considered invariant, whereas DTDD modes are still selected for each slot. This leads to a significant saving of computational resources, while exploiting the full flexibility of DTDD.
- the definition of a simplified payoff model for the scheduling game that reduces the information exchange overhead among cells.

Lastly, note that we do not cluster cells for DTDD scheduling, which allows each cell to independently adapt the DTDD modes to its local traffic.

7.2 System Model

We consider N cells, where cell n , $n = 1, \dots, N$, serves users with indices in the set \mathcal{U}_n , and $|\mathcal{U}_n| = K, \forall n$. Thus, the total number of users is NK , which are identified by indices

$k = 1, \dots, NK$. Time is split into slots. Each slot comprises multiple *resource blocks* of 12 subcarriers each. For the sake of a simpler explanation, we assume that the same number of simultaneous resource blocks is allocated to each user for an entire slot. Moreover, the K users in a cell occupy all the available spectrum. We denote the resulting per-user portion of the time-frequency plane as *user block*.¹ Let $\phi_i \in \{1, \dots, K\}$ denote the user block assigned to user $i \in \mathcal{U}_n, \forall n$. As only one user is allocated to each user block in each cell, N users are served simultaneously in the network on the same user block.

In each slot, cell n can be in either DL or UL mode, indicated as $z_n = 1$ or $z_n = 0$, respectively. Moreover, at each slot, the UL and DL transmit powers are also set. Let $\mathbf{z} = [z_1, \dots, z_N]$ denote the chosen modes for all cells at each slot. Let $\mathbf{p}_k(\mathbf{z})$ be the power allocated either to or from user k , in DL or UL mode, respectively. Note that the dependence of power allocation arises due to CI, which must be taken into account. We consider the following power constraints for users and base stations:

$$\mathbf{p}_k \leq \alpha_0, \quad k \in \mathcal{U}_n, \text{ if } z_n = 0, \quad (7.1a)$$

$$\sum_{k \in \mathcal{U}_n} \mathbf{p}_k \leq \alpha_1, \text{ if } z_n = 1, \quad (7.1b)$$

with mode-dependent constants α_0 and α_1 , respectively.

We assume that the link between user k and its serving base station n is a narrowband channel with gain $H_{n,k}$ and we let \mathbf{H} be the $N \times NK$ matrix with entries $\{H_{n,k}\}$. Moreover, since users operating in UL mode interfere with the DL users of neighboring cells (UL-to-DL interference), we define the $NK \times NK$ symmetric matrix \mathbf{U} , whose entry U_{k_1, k_2} , $k_1, k_2 = 1, \dots, NK$, is the channel gain between users k_1 and k_2 . Similarly, base stations operating in DL interfere with base stations operating in UL (DL-to-UL interference), thus we define the $N \times N$ symmetric matrix \mathbf{B} , whose entry B_{n_1, n_2} , $n_1, n_2 = 1, \dots, N$, is the channel gain between base stations n_1 and n_2 . We assume also that channel reciprocity holds for \mathbf{H} , \mathbf{U} , and \mathbf{B} .

7.2.1 Scheduling and Power Allocation Problem

To compute the network WSR, used as optimization metric, let the power of interference suffered by user k served in cell n operating in DL mode ($z_n = 1$) be

$$I_{n,k}(\mathbf{z}) = \sum_{m=1, m \neq n}^N I'_{n,k,m}(\mathbf{z}), \quad (7.2)$$

$$I'_{n,k,m}(\mathbf{z}) = z_m H_{m,k} \sum_{j \in \mathcal{U}_m: \phi_j = \phi_k} \mathbf{p}_j + c_1 (1 - z_m) \sum_{j \in \mathcal{U}_m: \phi_j = \phi_k} U_{j,k} \mathbf{p}_j, \quad (7.3)$$

with $c_1 \in [0, 1]$ being the UL-to-DL CI management factor. In (7.3), we account for the interference from other base stations operating in DL (matrix \mathbf{H}) and other users

¹This can be easily extended to more general cases of different resources allocated to each user and different numbers of users per cell.

operating in UL (matrix \mathbf{U}). Similarly, the power of interference suffered by user k served in cell n operating in UL mode ($z_n = 0$) is given by (7.2), now with

$$I'_{n,k,m}(\mathbf{z}) = c_2 z_m B_{m,n} \sum_{j \in \mathcal{U}_m: \phi_j = \phi_k} \mathbf{p}_j + (1 - z_m) \sum_{j \in \mathcal{U}_m: \phi_j = \phi_k} H_{n,j} \mathbf{p}_j, \quad (7.4)$$

where $c_2 \in [0, 1]$ is the DL-to-UL CI management factor. In (7.4), we take into account the interference from other base stations operating in DL (matrix \mathbf{B}) and other users operating in UL (matrix \mathbf{H}). Note that c_1 and c_2 varying in the interval $[0, 1]$ assume value 0 or 1 in the case of perfect or no interference management, respectively. For a given set of modes \mathbf{z} , the WSR at base station n is then

$$W_n(\mathbf{z}) = \sum_{k \in \mathcal{U}_n} w_k(z_n) \log_2 \left(1 + \frac{H_{n,k} \mathbf{p}_k}{\sigma^2 + I_{n,k}(\mathbf{z})} \right), \quad (7.5)$$

where σ^2 is the variance of the noise, $w_k(z_n) \in [0, 1]$ is the weight of user $k \in \mathcal{U}_n$, when the serving base station n is operating in mode z_n . The network WSR is

$$W(\mathbf{z}) = \sum_{n=1}^N W_n(\mathbf{z}). \quad (7.6)$$

A possible target for scheduling optimization is the maximization of the instantaneous WSR, i.e.,

$$\max_{\{\mathbf{p}_k(\mathbf{z})\}, \mathbf{z}} W(\mathbf{z}), \text{ s.t. (7.1) and } z_n \in \{0, 1\}, \forall n. \quad (7.7)$$

However, the joint scheduling and power allocation problem (7.7) is MINC and turns out to be NP-hard [126]. In particular, its non-convexity comes from the interference terms. Note that, to fully exploit DTDD, the problem must be solved at each slot (millisecond scale), which is prohibitive.

7.3 Scheduling and Power Allocation Decoupling

To simplify and decentralize the problem, we assume that each base station randomly selects its own mode, thus z_n becomes a Bernoulli random variable. Let q_n be the probability that base station n selects $z_n = 1$, i.e.,

$$q_n = \mathbb{P}[z_n = 1], \quad 1 - q_n = \mathbb{P}[z_n = 0]. \quad (7.8)$$

The value of q_n is chosen to maximize the *average* WSR, i.e.,

$$\max_{q_n} \mathbb{E}_{\mathbf{z}} [W_n(\mathbf{z})], \quad (7.9)$$

with the expectation taken over the actions of all players. Note that problem (7.9) should be solved once over multiple slots, as long as channel conditions and rate weights do not change, thus the network WSR for a given mode configuration does not change. Still, the mode is chosen for each slot according to the obtained distribution q_n .

Problem (7.9) requires the computation of $W_n(\mathbf{z})$, which, from (7.5) requires the knowledge of all channel gains. To reduce the communication overhead needed to share this information among all base stations, we approximate the instantaneous interference power level with its average, as better detailed in the following.

7.3.1 Average Interference Power

We define the average interference power for cell n as

$$J_n(\mathbf{z}) = \sum_{m=1, m \neq n}^N \bar{I}_{n,m}(\mathbf{z}), \quad (7.10)$$

where $\bar{I}_{n,m}(\mathbf{z})$ is the average interference power suffered by users in cell n from cell m , i.e.,

$$\bar{I}_{n,m}(\mathbf{z}) = \mathbb{E}_{\mathbf{B}, \mathbf{H}, \mathbf{U}, \{\mathbf{p}_k\}} [I'_{n,k,m}(\mathbf{z})], \quad (7.11)$$

and the expectation is taken with respect to both the channel gains and the powers.

The computation of the average in (7.11) is rather complex, since the allocated power depends on scheduling and power assignments, which in turn depend on the channel gains. Thus, we consider an approximated value of the average interference power, where we keep the transmit power as a fixed term. In particular, we assume the DL transmit power as equally split (on average) among the K served users, i.e., $\mathbf{p}_k = \alpha_1/K$ for $k \in \mathcal{U}_m$ when $z_m = 1$, $\forall m$. In UL, we assume that all users transmit at their maximum power, i.e., $\mathbf{p}_k = \alpha_0$, for $k \in \mathcal{U}_m$ when $z_m = 0$, $\forall m$. Then, under this setting, the average interference power suffered by users in cell n from cell m becomes

$$\bar{I}_{n,m}(\mathbf{z}) = \mathbb{E}_{\mathbf{B}, \mathbf{H}, \mathbf{U}} [I'_{n,k,m}(\mathbf{z})], \quad (7.12)$$

while (7.10) still holds. When cell n is in DL ($z_n = 0$), since we have one user per user block from (7.3) we have

$$\bar{I}_{n,m}(\mathbf{z}) = \frac{\alpha_1}{K} z_m \mathbb{E}_{\mathbf{H}} [H_{m,k}] + c_1 (1 - z_m) \mathbb{E}_{\mathbf{U}} [U_{j,k}] \alpha_0, \quad (7.13)$$

while when cell n is in UL ($z_n = 1$), from (7.4) we have

$$\bar{I}_{n,m}(\mathbf{z}) = \frac{\alpha_1}{K} c_2 z_m \mathbb{E}_{\mathbf{B}} [B_{m,n}] + (1 - z_m) \mathbb{E}_{\mathbf{H}} [H_{n,j}] \alpha_0. \quad (7.14)$$

Note that the expectations here can be computed either using the channel statistics, if available, or by averaging multiple channel estimates over time. When using this approximation $J_n(\mathbf{z})$ in the following will be denoted as *reference interference power*.

7.3.2 Power Allocation Sub-problem

For power allocation, we assume that each cell allocates power to maximize only its utility function, regardless of interference and CI. This enables a greedy, decentralized, and low complexity power allocation.

In UL mode, we assume that each user transmits at its maximum power, to maximize its rate, i.e.,

$$\mathbf{p}_k^* = \alpha_0 \quad k \in \mathcal{U}_n, \text{ if } z_n = 0. \quad (7.15)$$

For the DL mode, we approximate $I_{n,k}(\mathbf{z}) \approx J_n(\mathbf{z})$ and we distribute the transmit power according to a modified version of the (weighted) water-filling algorithm [130], which takes into account the noise, interference, and CI and aims at maximizing $W_n^s(\mathbf{z})$ at each cell independently, i.e.,

$$\mathbf{p}_k^* = \left(\frac{w_k(1)}{\lambda_k} - \frac{\sigma^2 + J_n(\mathbf{z})}{H_{n,k}} \right)^+, \text{ if } z_n = 1, \quad (7.16)$$

where $\lambda_k > 0$ is the Lagrange multiplier that satisfies the power constraint (7.1b).

7.3.3 Approximated WSRs

Once powers have been allocated, let $W_n^s(\mathbf{z})$ be the WSR (7.5), with powers given either by (7.15) or (7.16), depending on the DTDD mode, to be used in (7.9).

A further simplification is obtained by replacing $I_{n,k}(\mathbf{z})$ with $J_n(\mathbf{z})$ in (7.5), providing the WSR $W_n^a(\mathbf{z})$, which is then used instead of $W_n^s(\mathbf{z})$ in (7.9).

7.4 Game-theoretic Scheduling

As the major contribution of this work, we tackle this issue by framing the scheduling problem (7.9) as a non-cooperative, decentralized, and instantaneous game, played by the base stations. The MS-NE solution of the game provides an efficient distributed scheduling, where no players desire to deviate from. Differently from a centralized scheduling scheme, each base station n acts as a player on its own, with action set $\mathcal{A}_n = \{0, 1\}$, corresponding to UL and DL modes. Let the overall set of actions be $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_N$. For each slot, each base station n chooses its mode over its action set \mathcal{A}_n , according to the distribution q_n . Define $\mathbf{q} = [q_1, \dots, q_N]^T$ as the vector of probabilities of choosing the DL mode for the N base stations: this is the vector of mixed strategies over the action set \mathcal{A} . Since we aim at maximizing the average WSR, the payoff of player n , denoted as $W_n(\mathbf{z})$, can be either the WSR $W_n^s(\mathbf{z})$ or its approximated version $W_n^a(\mathbf{z})$. The objective of each player is the maximization of his/her average payoff, through the choice of a mixed strategy.

Now, the scheduling problem translates into the following static game (in normal form) with players, actions, and payoffs

$$\mathcal{G} = (\{1, \dots, N\}, \mathcal{A}, \{W_n(\mathbf{z})\}). \quad (7.17)$$

7.4.1 Mixed Strategy Selection

We find an MS-NE, where each player aims at maximizing its average payoff (solution of (7.9)); if any player deviates, then its payoff degrades. In DTDD, this means that the equilibrium distribution locally maximizes the payoff at each cell. To solve the optimization problem in the MS-NE sense, we impose also the *principle of indifference* [127]. This principle imposes the further constraint that player n chooses q_n also to make his/her opponents indifferent over their actions. More precisely, the opponents would obtain the same average payoff, no matter which action they choose. Hence, the cross interference with optimal distribution yields that the same rate in UL and DL is achieved by neighboring cells.

This is achieved only in the asymptotic regime, with many opportunities to alternate between UL and DL, while in practice channel and rate weights can be considered constant only over a few slots, yielding only few opportunities.

For user n in UL and DL modes, the average payoff is denoted as $\mathbb{E}_{\mathbf{z}}[W_n(\mathbf{z})|z_n = a]$, with $a = 0$ or 1 , respectively, where the expectation is taken with respect to all the possible actions z_m , $m \neq n$ of the opponents of player n . By imposing the indifference principle, we obtain the following system of N nonlinear equations in N variables (q_n) [Sec. 4.6.2 [127]]

$$\mathbb{E}_{\mathbf{z}}[W_n(\mathbf{z})|z_n = 1] = \mathbb{E}_{\mathbf{z}}[W_n(\mathbf{z})|z_n = 0], \quad n=1, \dots, N. \quad (7.18)$$

The average payoff of player n , can be written as a function of the probability that the various modes are selected, i.e., [see (113.1) [127]]

$$\begin{aligned} \mathbb{E}_{\mathbf{z}}[W_n(\mathbf{z})|z_n = a] &= \sum_{\mathbf{z}' \in \mathcal{A}: z'_n = a} W_n(\mathbf{z}') \mathbb{P}[\mathbf{z} = \mathbf{z}'] \\ &= \sum_{\mathbf{z}' \in \mathcal{A}: z'_n = a} W_n(\mathbf{z}') \prod_{m=1, m \neq n}^N q_m^{z'_m} (1 - q_m)^{1-z'_m}, \end{aligned} \quad (7.19)$$

where the latter equality is obtained considering that modes are selected independently by each player n with probabilities q_n (see (7.8)).

Given the information shared from neighboring cells, (7.18) and (7.19) provide a system of polynomial equations in $q_m \in [0, 1]$, which can be solved by each cell using numerical methods (see [131] and references therein) to pick its mixed strategy. Moreover, each cell also knows the strategies of others as every player solve the same system (7.18), according to MS-NE.

Problem (7.7) is then shifted into a system of equations, which can be solved very efficiently. To further save computational complexity, (7.18) can be solved only at one base station and the optimal distribution communicated to the others, as all players solve the same system. The amount of information shared strictly depends on the type of payoffs, as discussed in Section 7.4.2. Lastly, note that the MS-NE system of equations has at least one solution as the actions set of DTDD players is finite [Prop. 116.1 [127]]. However, being nonlinear, it may exhibit multiple solutions. In that case, all the players must agree to choose the mixed strategy that maximizes the overall system performance, to avoid ambiguity.

7.4.2 Impact Of The Payoff Model

We now consider the different alternatives of payoff $W_n(\mathbf{z})$, and their consequences on the communication overhead.

Simplified-payoff (SIP) game with payoff $W_n^s(\mathbf{z})$ When $W_n(\mathbf{z}) = W_n^s(\mathbf{z})$, the interference power $I_{n,k}(\mathbf{z})$ has to be computed. In this case, although players independently solve (7.18), they need to share information on the mutual interference, in particular, matrices \mathbf{B} , \mathbf{H} , \mathbf{U} , weights $\{w_k(z_n)\}$, and the allocated powers.

Approximated-payoff (APP) game with payoff $W_n^a(\mathbf{z})$ When $W_n(\mathbf{z}) = W_n^a(\mathbf{z})$, the base stations only need to share information about weights $\{w_k(z_n)\}$, allocated powers and NK elements of matrix \mathbf{H} , corresponding to the direct signal channel gains. All the CI channel gains instead can be omitted. Still, we need the average interference power $J_n(\mathbf{z})$. From (7.10), we observe that this can be obtained from the $(N - 1)$ average interference terms $\bar{I}_{n,m}(z_m)$, $m = 1, \dots, N$, $m \neq n$. Therefore, a significant reduction of signaling overhead is obtained with this payoff. Moreover, as these average interference terms are slowly time-varying, they can be shared at a slower rate than \mathbf{H} , making the signaling overhead negligible.

7.5 Numerical Results

We now evaluate the performance of both SIP and APP games in a DTDD 5G network. We assume that channel varies every 10 ms (one frame) and each user block comprises one resource block of 12 subcarriers, continuously allocated to the same user for one DTDD frame. Therefore, only one computation of the MS-NE takes place to schedule 10 slots. For comparison purposes, we consider:

- a) the *instantaneous optimal approach* (OPT), which works as follows. At each slot, we first compute the WSR $W(\mathbf{z})$ of (7.6) with $W_n^s(\mathbf{z})$ for all possible scheduling modes \mathbf{z} , using the power allocation of (7.15)-(7.16), and then select the mode and powers yielding the maximum WSR. Note that we must evaluate 2^N scheduling modes (all possible \mathbf{z} vectors of size N with 0-1 entries). Moreover, the full channel state information must be transferred to a central node, where the WSRs are computed, and then the scheduling decision should be broadcast to all cells. These operations must be repeated at each slot;
- b) the static TDD (STDD) scheme with frame equally partitioned in both directions, i.e., 5 slots in UL and 5 in DL;
- c) the decentralized scheme [129], which always assigns the first 2 and the last 2 slots to DL and UL, respectively, whereas the rest of the slots are chosen by making the optimal allocation of a DL to UL switching point.

In all cases, power is allocated in the greedy fashion, as in (7.15) and (7.16). The expectations of (7.13) and (7.14) have been computed by averaging multiple channel estimates.

We consider an outdoor dense scenario, with seven hexagonal small cells, having side of size 333 m. LoS channels with path-loss exponent 2 are assumed. This results to

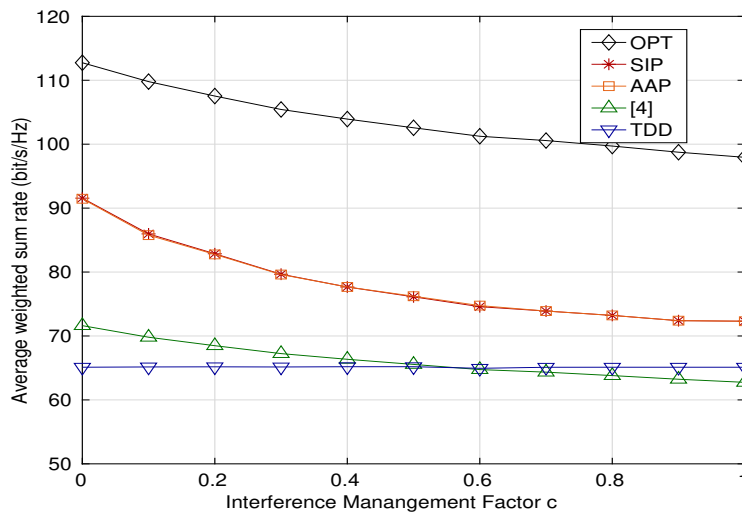


Figure 7.2: Average network payoff of a DTDD system for OPT, SIP, AAP, and STDD, as a function of c , with fixed 15 users per cell.

be the most challenging scenario for DTDD [128]. We wrap the scenario into a sphere, which leads to each cell having 6 interfering cells. We also assume $c_1 = c_2 = c$ and the maximum transmit powers at the base stations and users, $\alpha_1 = \alpha_0 = 23$ dBm, as in [132]. Each base station and user is equipped with one omni-directional antenna. The UL and DL rate weights are chosen from a uniform distribution. The noise power is -10 dB to the signal power.

We remark that, although under the APP model the players use an approximation of the payoff, we report here the resulting *true* payoffs obtained with their actions, according to (7.5), by considering the actual CI at each slot.

Fig. 7.2 shows the average WSR achieved by the various approaches, as a function of c . We note that both SIP and AAP significantly outperform TDD and [129], for all values of c . This is due to the higher flexibility of both SIP and AAP compared to TDD and [129]. Moreover, AAP performs strictly close to SIP, despite its use of the approximated payoff, which in turn makes AAP quite appealing for its reduced communication overhead. Still, SIP and AAP exhibit a performance gap with respect to OPT, as in OPT we optimize the cell mode, thus maximizing the *instantaneous* sum-rate, while AAP and SIP randomly select the cell modes, with densities that maximize the resulting *average* sum-rate. Moreover, with APP players do not have a perfect estimate of the interference when locally solving the MS-NE problem.

Fig. 7.3.a shows the cumulative distribution function (CDF) of the average WSR for the various methods and typical CI management factor $c = 0.3$. Both SIP and AAP shows a similar CDF, and provides higher average rates than static TDD and the method of [129]; this indicate the proposed solution operates fairly among cells. Fig. 7.3.b shows the average WSR as a function of the number of users per cell, again for $c = 0.3$. According to our assumptions in the System Model, the user block shrinks in the frequency domain

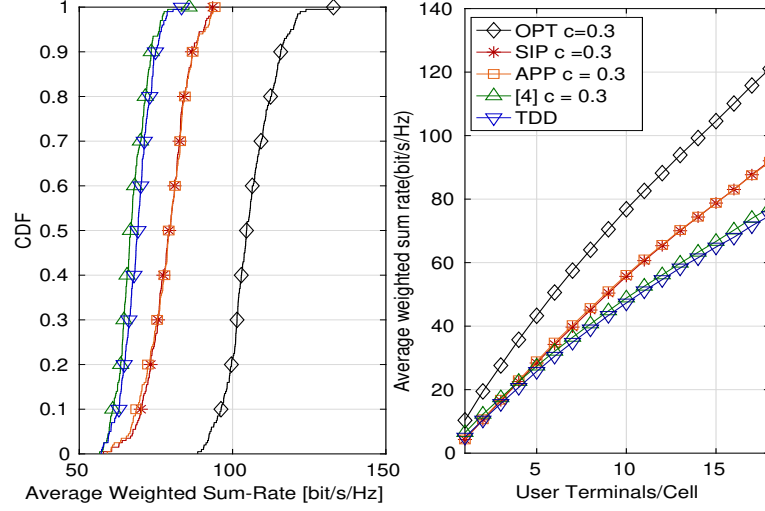


Figure 7.3: a) CDF for the WSR with $c = 0.3$ and 15 users/cell. b) Average network payoff of a DTDD system for OPT, SIP, APP, and TDD, as a function of the number of users per cell.

as the number of users increases. The figure shows that our approaches are particularly effective in dense networks, which are of interest for current and future generations of cellular systems. Moreover, both games result to be extremely scalable as the network size grows. With decentralized CI management, we provide a fully decentralized and low complexity DTDD scheduling for dense networks.

7.6 Conclusions

A new CI aware, decentralized, and low complexity power allocation is proposed. Scheduling of UL/DL transmissions in a DTDD scenario is cast into the MS-NE of a simultaneous game among the base stations. Our performance evaluation shows a significant advantage with respect to TDD and the state-of-the-art in terms of achievable WSR even in absence of further interference rejection techniques for the most challenging DTDD scenario. In particular, the APP solution also significantly reduces the communication overhead, while performing close to SIP.

Chapter 8

Conclusions and Future Work

This thesis presented several novel contributions for the SISO and MIMO FD systems in the sub-6GHz and for the mMIMO FD systems in the mmWave. This chapter provides the concluded remarks on the work items, lessons learned, and some future research directions to be explored.

- This thesis started with a novel idea to reconstruct the saturated signal in the FD systems with the LMMSE estimates of missing samples, presented in Chapter 2. Results showed an optimal tuning point dependent on the resolution of the ADCs, which exhibits the optimal compromise between the number of samples required to reconstruct the missing samples and the QN. Our work addressed an important issue of the FD systems, but it was presented in a simple scenario. Namely, it was assumed that the known samples were perfectly SI free and only the SISO FD case was investigated. Future work direction for this work may consider that the known samples are corrupted with the residual SI, and joint adaptation of the DSIC stage and signal reconstruction must be investigated, which can lead to a more realistic design. Extension of this work to MIMO or mMIMO FD systems is also desirable. In the second part of Chapter 2, a novel digital beamforming design for the MIMO FD system under the joint sum-power and the per-antenna power constraints was presented. Results show that the considered constraints are prominent when the transmit power increases. Namely, as the LDR noise is proportional to the total power transmitted from each antenna, imposing the power constraint also constraints the transmit LDR noise variance and thus leads to higher gain. Chapter 2 concluded with a novel SIC architecture, RF calibration algorithm and measurements results for the MIMO OFDM FD system. The proposed SIC has very low complexity, making it very desirable, and the proposed RF calibration is applicable to any MIMO FD system. The experimental results taken at EURECOM for an OFDM FD system showed that it is possible to achieve a low-rank SI channel when the FD systems transmit and receive antennas arrays are placed according to ULAs configurations. One interesting future research direction for this work is to show the validity of the presented idea for a mMIMO setting, possibly with multiple UL and DL users.

- In Chapter 3, we first studied a novel HYBF for the IAB mmWave FD system. Results showed that FD relay could lead to significant performance improvement. Apart from the WSR, which was the criteria assumed in the problem formulation, the FD relay also results in minimum latency for the communication from the mmWave mMIMO HD BS to the users via hybrid FD relay. Future work of this study item may consist in investigating the coexistence of multiple FD relays and mmWave HD BSs and the design of the beamformers which are global network interference aware. Then we studied a novel HYBF/multi-stage design for a point-to-point mmWave MIMO OFDM FD system. Future work for this study item may consist in extending the proposed design to multiple pairs of the OFDM FD links or to the case of a multi-user OFDM mMIMO FD system. Finally, Chapter 3 concluded with a novel HYBF design for a K-pair point-to-point FD system. Simulation results showed that the K-pair mMIMO FD system is able to achieve significant gain over an HD system. As this work assumed a single-carrier FD system, future work for this study item can consider the case of K-pair point-to-point mMIMO OFDM FD links and investigate the achievable performance.
- In Chapter 4, we studied a novel HYBF design for a mmWave mMIMO FD system with multi-antenna UL and DL users under several practical considerations. Simulation results showed that the proposed design can achieve significant higher gain over a single-cell mMIMO HD system by handling the SI and CI jointly with the beamformers. As this work considered a single-carrier FD system, future work for this contribution can consider the case of an OFDM single-cell FD system, possibly including the joint adaption of the SIC architecture. Another interesting research direction for this work could be the joint blind channel estimation and beamforming design, which could reduce the communication overhead significantly.
- In Chapter 5, we extended our HYBF design to the case of a multi-cell mmWave FD system by proposing a novel C-HYBF design. We also shed light on the fact that centralized HYBF designs are infeasible for a multi-cell FD system. Then we presented the first-ever P&D-HYBF design for mmWave, which enables per-link independent optimization of the beamformers at each multi-processor FD BS. Simulation results showed that both the proposed designs significantly outperformed the traditional multi-cell HD system and achieved similar performance. Even though the proposed P&D-HYBF design achieves similar performance as the C-HYBF, its drawback is that it requires sharing one analog beamformer and analog combiner of massive dimensions. Future work for this study item may consider the design of the analog stage based solely on statistical CSI, which can avoid sharing the analog beamformers and combiners of massive dimensions every channel coherence time and thus reducing the communication overhead significantly. Another interesting research direction could be studying non-cooperative P&D-HYBF schemes, which could be very prominent for the next generation of mmWave FD networks.
- In Chapter 6, we introduced the concept of NF-IRSs for the mmWave FD system, which leads to reduced power consumption and hardware cost. Results show that

the NF-IRSs not only can replace the analog stage of the mmWave FD systems but also lead to significantly higher gain. An interesting research direction for this work could be the analysis of the performance with multiple users. Distributed optimization of the beamformers and the NF-IRSs could also be an interesting research direction as it can lead to faster convergence of the algorithm, for which the joint optimization can require significant execution time, especially when large NF-IRSs are used.

- Chapter 7 studied the problem of joint scheduling and power optimization of the multi-cell DTDD system. The proposed solution is fully distributed and has low complexity. Simulation results showed that our design can achieve significant performance gain. This work was limited to the case of single-antenna users. Future work direction for this study item could consider the joint optimization of the beamformers, scheduling and power in a multi-cell MIMO DTDD network. Moreover, in our work, we considered a non-cooperative approach. Distributed solutions with cooperation by exchanging a small amount of information could also be investigated, which can surely lead to further performance improvement.

Appendices

Appendix A

Gradient Derivation for OFDM FD System

In this section we derive an expression for the gradient for the terms of the form,

$$\mathbf{Y} = \mathbf{A} \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D})\mathbf{B} + \mathbf{F}(\mathbf{X}), \quad \mathbf{R} = \mathbf{C}\mathbf{X}\mathbf{D}, \quad (\text{A.1})$$

where $\mathbf{F}(\mathbf{X})$ represents any matrix function in \mathbf{X} . Each element of \mathbf{Y} can be written as,

$$\begin{aligned} \mathbf{Y}_{i,j} &= \sum_{m,n} \mathbf{A}_{i,m} \mathbf{R}_{m,n} \mathbf{B}_{n,j} \delta_{m-n} + \mathbf{F}(\mathbf{X})_{i,j}, \\ \mathbf{R}_{m,n} &= \sum_{p,q} \mathbf{C}_{m,p} \mathbf{X}_{p,q} \mathbf{D}_{q,n}, \\ \mathbf{Y}_{i,j} &= \sum_{m,n} \mathbf{A}_{i,m} \left(\sum_{p,q} \mathbf{C}_{m,p} \mathbf{X}_{p,q} \mathbf{D}_{q,n} \right) \mathbf{B}_{n,j} \delta_{m-n} + \mathbf{F}(\mathbf{X})_{i,j}, \end{aligned} \quad (\text{A.2})$$

where δ_k represents the Kronecker delta function. We define $\mathbf{V}_{r,s}$ as zero-valued matrix except for a unity element at row r and column s and we obtain,

$$\begin{aligned} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} &= \sum_{r,s} \mathbf{V}_{r,s} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}_{r,s}} = \sum_{r,s} \mathbf{V}_{r,s} \sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \frac{\det(\mathbf{Y}_{i,j})}{\partial \mathbf{X}_{r,s}} \\ &= \sum_{r,s} \mathbf{V}_{r,s} \sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \left[\sum_{m,n} \mathbf{A}_{i,m} \mathbf{C}_{m,r} \mathbf{D}_{s,n} \mathbf{B}_{n,j} \delta_{m-n} + \frac{\det(\mathbf{F}(\mathbf{X})_{i,j})}{\partial \mathbf{X}_{r,s}} \right] \\ &= \sum_{r,s} \mathbf{V}_{r,s} \left(\sum_{m,n} \mathbf{C}_{m,r} \mathbf{D}_{s,n} \left(\sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \mathbf{A}_{i,m} \mathbf{B}_{n,j} \right) \delta_{m-n} + \right. \\ &\quad \left. \sum_{i,j} \frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}_{i,j}} \frac{\det(\mathbf{F}(\mathbf{X})_{i,j})}{\partial \mathbf{X}_{r,s}} \right) = [\mathbf{D} \text{diag}(\mathbf{B} \left(\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{Y}} \right)^T \mathbf{A}) \mathbf{C}]^T + \mathbf{F}' \end{aligned} \quad (\text{A.3})$$

For simplicity we call the second term in the summation \mathbf{F}' since that is not of interest here or the required gradients (needed forms of $\mathbf{F}(\mathbf{X})$) are derived in [133]. Further using

the result, $\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \det(\mathbf{Y})(\mathbf{Y}^{-1})^T$ we can simplify it as ,

$$\frac{\partial \det(\mathbf{Y})}{\partial \mathbf{X}} = \det(\mathbf{Y})[\mathbf{D} \text{diag}(\mathbf{B}\mathbf{Y}^{-1}\mathbf{A})\mathbf{C}]^T + \mathbf{F}'. \quad (\text{A.4})$$

Appendix B

Proofs for Chapter 5

The proof of Theorem 3 is based on the result derived in the following.

Lemma B.0.1. *Let $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B} + a \mathbf{A} \text{diag}(\mathbf{X} + \mathbf{Q})\mathbf{B} + b \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D} + \mathbf{E}) + \mathbf{F}$. The derivative of $\text{ln} \det(\mathbf{Y})$ with respect to \mathbf{X} is given by*

$$\begin{aligned} \frac{\partial \text{ln} \det \mathbf{Y}}{\partial \mathbf{X}} = & \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H + a \text{diag}(\mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H) \\ & + b \mathbf{C}^H \text{diag}(\mathbf{Y}^{-H}) \mathbf{D}^H. \end{aligned} \quad (\text{B.1})$$

Proof. By substituting $\phi = \text{ln} \det(\mathbf{Y})$, we can write

$$\partial \phi = \mathbf{Y}^{-H} : d\mathbf{Y} = \text{Tr}(\mathbf{Y}^{-1} d\mathbf{Y}), \quad (\text{B.2})$$

where operator $:$ denotes the Frobenius inner product, i.e. $\mathbf{G}_{RF} : \mathbf{H} = \text{Tr}(\mathbf{G}_{RF}^H \mathbf{H})$. Its derivative with respect to \mathbf{X} can be written as

$$\begin{aligned} \frac{\partial \phi}{\partial \mathbf{X}} = & \mathbf{Y}^{-H} : \left[\frac{d}{d\mathbf{X}} (\mathbf{A}\mathbf{X}\mathbf{B} + a \mathbf{A} \text{diag}(\mathbf{X})\mathbf{B} \right. \\ & \left. + b \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D} + \mathbf{E}) + \mathbf{F}) \right], \end{aligned} \quad (\text{B.3})$$

where the last term results to be zero as independent from \mathbf{X} . Substituting the Frobenius product with the trace operator, using its cyclic shift and separating terms, yields

$$\begin{aligned} \frac{\partial \phi}{\partial \mathbf{X}} = & \underbrace{\frac{\partial \text{Tr}(\mathbf{B}\mathbf{Y}^{-1}\mathbf{A}\mathbf{X})}{\partial \mathbf{X}}}_1 + a \underbrace{\frac{\partial \text{Tr}(\mathbf{B}\mathbf{Y}^{-1}\mathbf{A} \text{diag}(\mathbf{X}))}{\partial \mathbf{X}}}_2 \\ & + b \underbrace{\frac{\partial \text{Tr}(\mathbf{Y}^{-1} \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D}))}{\partial \mathbf{X}}}_3 + b \frac{\partial \text{Tr}(\mathbf{Y}^{-1} \text{diag}(\mathbf{E}))}{\partial \mathbf{X}}, \end{aligned} \quad (\text{B.4})$$

where the last term being independent of \mathbf{X} is also zero. To proof the aforementioned result, we proof the derivatives of 1, 2 and 3 separately. Firstly, for 1, by using $:$ and doing some simple algebraic manipulations leads to

$$\frac{\partial \text{Tr}(\mathbf{B}\mathbf{Y}^{-1}\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H : \partial \mathbf{X} = \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H. \quad (\text{B.5})$$

To obtain the derivative of 2, we first define $\text{diag}(\mathbf{X}) = \mathbf{Z}$. The diagonal of \mathbf{X} can be written as $\text{diag}(\mathbf{X}) = \mathbf{I} \circ \mathbf{X}$ where \circ denotes the Hadamard product. By writing 2 with : and expressing the diagonal term as a function of \circ , and using the commutative property of the Hadamard product leads to the following result

$$\begin{aligned} a \frac{\partial \text{Tr}(\mathbf{B}\mathbf{Y}^{-1}\mathbf{A}\mathbf{Z})}{\partial \mathbf{Z}} &= a \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H : \partial \mathbf{Z}, \\ &= a \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H : \mathbf{I} \circ \partial \mathbf{X}, \\ &= a \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H \circ \mathbf{I} : \partial \mathbf{X}, \\ &= a \text{diag}(\mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H). \end{aligned} \tag{B.6}$$

To compute the derivative of 3, we first define $\text{diag}(\mathbf{C}\mathbf{X}\mathbf{D}) = \mathbf{W}$. By using a similar approach as in (B.6), we get

$$\begin{aligned} b \frac{\partial \text{Tr}(\mathbf{Y}^{-1}\mathbf{W})}{\partial \mathbf{W}} &= b \mathbf{Y}^{-H} : \partial \mathbf{W}, \\ &= b \mathbf{Y}^{-H} : \mathbf{I} \circ \mathbf{C} \partial \mathbf{X} \mathbf{D}, \\ &= b \mathbf{Y}^{-H} \circ \mathbf{I} : \mathbf{C} \partial \mathbf{X} \mathbf{D}, \\ &= b \text{diag}(\mathbf{Y}^{-H}) : \mathbf{C} \partial \mathbf{X} \mathbf{D}, \\ &= b \mathbf{C}^H \text{diag}(\mathbf{Y}^{-1})^H \mathbf{D}^H. \end{aligned} \tag{B.7}$$

Combining the result from each term concludes the proof for Lemma (C.1). \square

To prove Theorem 3, note that the covariance matrices in 4.11 has a special (Hermitian) structure, i.e., $\mathbf{B} = \mathbf{A}^H$ and $\mathbf{D} = \mathbf{C}^H$. Therefore, the result of Lemma C.1.1 for this particular case is given in the following.

Lemma B.0.2. *Let $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B} + a \mathbf{A} \text{diag}(\mathbf{X} + \mathbf{Q})\mathbf{B} + b \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D} + \mathbf{E}) + \mathbf{F}$, where the size of matrices involved is such that the product is valid. Let $\mathbf{B} = \mathbf{A}^H$ and $\mathbf{D} = \mathbf{C}^H$ and the derivative of $\text{ln} \det(\mathbf{Y})$ is given by*

$$\begin{aligned} \frac{\partial \text{ln} \det \mathbf{Y}}{\partial \mathbf{X}} &= \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{A} + a \text{diag}(\mathbf{A}^H \mathbf{Y}^{-H} \mathbf{A}) \\ &\quad + b \mathbf{C}^H \text{diag}(\mathbf{Y}^{-H}) \mathbf{C}. \end{aligned} \tag{B.8}$$

Proof. The result follows directly by relying on the result given in Lemma C.1.1 by substituting $\mathbf{B} = \mathbf{A}^H$ and $\mathbf{D} = \mathbf{C}^H$ \square

Proof. Theorem 3 To prove the gradients to linearize the WSR with respect to \mathbf{T}_k and \mathbf{Q}_j , we proceed by simplifying the WSR as

$$\begin{aligned} \text{WSR} &= \sum_{k \in \mathcal{U}} w_k \text{ln} \det(\mathbf{R}_k) - w_k \text{ln} \det(\mathbf{R}_{\bar{k}}) \\ &\quad + \sum_{j \in \mathcal{D}} w_j \text{ln} \det(\mathbf{R}_j) - w_j \text{ln} \det(\mathbf{R}_{\bar{j}}). \end{aligned} \tag{B.9}$$

The $\text{WSR}_{\bar{k}}^{UL}$ and WSR^{DL} should be linearized for \mathbf{T}_k and $\text{WSR}_{\bar{j}}^{DL}$ and WSR^{UL} for \mathbf{Q}_j . Note from (4.11) that \mathbf{T}_k appears in $\text{WSR}_{\bar{k}}^{UL}$ and WSR^{DL} with the structure $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^H + a \mathbf{A} \text{diag}(\mathbf{X} + \mathbf{Q})\mathbf{A}^H + b \text{diag}(\mathbf{C}\mathbf{X}\mathbf{C}^H + \mathbf{E}) + \mathbf{F}$, where the scalars a and b are due to the LDR noise model, \mathbf{A} and \mathbf{C} are the interfering channels, \mathbf{F} and \mathbf{E} contain the noise contributions from other transmit covariance matrices but independent from \mathbf{T}_k . The same structure holds also for the DL covariance matrices $\mathbf{Q}_j, \forall j \in \mathcal{D}$. By applying the result from Lemma B.0.2 with $\mathbf{Y} = \mathbf{R}_k$ or $\mathbf{Y} = \mathbf{R}_{\bar{k}}$ repetitively $K - 1$ time for linearizing $\text{WSR}_{\bar{k}}$ with respect to \mathbf{T}_k yield the gradient \mathbf{A}_k . Similarly, by considering $\mathbf{Y} = \mathbf{R}_j$ or $\mathbf{Y} = \mathbf{R}_{\bar{j}}, \forall j \in \mathcal{D}$ and applying the result from Lemma B.0.2 yield the gradient \mathbf{B}_k .

The same reasoning holds also for \mathbf{Q}_j , which leads to the gradients $\hat{\mathbf{C}}_j$ and \mathbf{D}_j by applying the result provided in Lemma B.0.2 for $\text{WSR}_{\bar{j}}^{DL}$ $J - 1$ times and for WSR^{UL} K times, respectively, $\forall j \in \mathcal{D}$. \square

B.1 Proof of Theorem 5

The dominant generalized eigenvector solution maximizes the reformulated concave WSR maximization problem

$$\begin{aligned} \text{WSR} = & \sum_{k \in \mathcal{U}} \left[w_k \ln \det \left(\mathbf{I} + \mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_{\bar{k}}^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \right) \right. \\ & \left. - \text{Tr} \left(\mathbf{U}_k^H \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + l_k \mathbf{I} + \mathbf{\Psi}_k \right) \mathbf{U}_k \right) \right] \\ & + \sum_{j \in \mathcal{D}} \left[w_j \ln \det \left(\mathbf{I} + \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_{\bar{j}}^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j \right) \right. \\ & \left. - \text{Tr} \left(\mathbf{V}_j^H \mathbf{G}_{RF}^H \left(\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + l_0 \mathbf{I} + \mathbf{\Psi}_0 \right) \mathbf{G}_{RF} \mathbf{V}_j \right) \right]. \end{aligned} \quad (\text{B.10})$$

To prove Theorem 5 for solving (B.10), we first consider the UL digital beamforming solution by keeping the analog beamformer and the digital DL beamformers fixed. We proceed by considering user $k \in \mathcal{U}$ for which we wish to compute the WSR maximizing digital UL beamformer. The same proof will be valid $\forall k \in \mathcal{U}$. The proof relies on simplifying

$$\begin{aligned} \max_{\mathbf{U}_k} & w_k \ln \det \left(\mathbf{I} + \mathbf{U}_k^H \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_{\bar{k}}^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k \right) \\ & - \text{Tr} \left(\mathbf{U}_k^H \left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + l_k \mathbf{I} + \mathbf{\Psi}_k \right) \mathbf{U}_k \right) \end{aligned} \quad (\text{B.11})$$

until the Hadamard's inequality applies as in Proposition 1 [38] or Theorem 1 [134]. The Cholesky decomposition of the matrix $\left(\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + l_k \mathbf{I} + \mathbf{\Psi}_k \right)$ is given as $\mathbf{L}_k \mathbf{L}_k^H$ where \mathbf{L}_k is the lower triangular Cholesky factor. By defining $\tilde{\mathbf{U}}_k = \mathbf{L}_k^H \mathbf{U}_k$, (B.11) reduces to

$$\begin{aligned} \max_{\mathbf{U}_k} & w_k \ln \det \left(\mathbf{I} + \tilde{\mathbf{U}}_k^H \mathbf{L}_k^{-1} \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_{\bar{k}}^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \right. \\ & \left. \mathbf{L}_k^{-H} \tilde{\mathbf{U}}_k \right) - \text{Tr} \left(\tilde{\mathbf{U}}_k^H \tilde{\mathbf{U}}_k \right). \end{aligned} \quad (\text{B.12})$$

Let $\mathbf{E}_k \mathbf{D}_k \mathbf{E}_k^H$ be the eigen-decomposition of $\mathbf{L}_k^{-1} \mathbf{H}_k^H \mathbf{R}_{\bar{k}}^{-1} \mathbf{H}_k \mathbf{L}_k^{-H}$, where \mathbf{E}_k and \mathbf{D}_k are the unitary and diagonal matrices, respectively. Let $\mathbf{O}_k = \mathbf{E}_k^H \tilde{\mathbf{U}}_k \tilde{\mathbf{U}}_k^H \mathbf{E}_k$ and (B.12) can be expressed as

$$\max_{\mathbf{O}_k} w_k \ln \det(\mathbf{I} + \mathbf{O}_k \mathbf{D}_k) - \text{Tr}(\mathbf{O}_k). \quad (\text{B.13})$$

By Hadamard's inequality [Page 233 [135]], it can be seen that the optimal \mathbf{O}_k must be diagonal. Therefore, $\mathbf{U}_k = \mathbf{L}_k^{-H} \mathbf{E}_k \mathbf{O}_k^{\frac{1}{2}}$ and thereby

$$\begin{aligned} \mathbf{H}_k^H \mathbf{F}_{RF} \mathbf{R}_{\bar{k}}^{-1} \mathbf{F}_{RF}^H \mathbf{H}_k \mathbf{U}_k &= \mathbf{L}_k \mathbf{L}_k^H \mathbf{L}_k^{-H} \mathbf{E}_k \mathbf{O}_k^{\frac{1}{2}} \mathbf{D}_k \\ &= (\hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + l_k + \Psi_k) \mathbf{U}_k \mathbf{D}_k, \end{aligned} \quad (\text{B.14})$$

from which we select u_k dominant eigenvectors, which concludes the proof for the UL beamformer for user $k \in \mathcal{U}$. For the digital DL beamformers the proof follow similarly by considering the following optimization problem $\forall j$

$$\begin{aligned} \max_{\mathbf{V}_j} w_j \ln \det(\mathbf{I} + \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_{\bar{j}}^{-1} \mathbf{H}_j \mathbf{G}_{RF} \mathbf{V}_j) \\ - \text{Tr}(\mathbf{V}_j^H \mathbf{G}_{RF}^H (\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j + l_0 + \Psi_0) \mathbf{G}_{RF} \mathbf{V}_j). \end{aligned} \quad (\text{B.15})$$

and simplifying it until the Hadamard's inequality applies to yield a similar result as expressed in (C.8).

The proof for analog beamformer \mathbf{G}_{RF} does not apply directly as the KKT condition have the form $\mathbf{A}_1 \mathbf{G}_{RF} \mathbf{A}_2 = \mathbf{B}_1 \mathbf{G}_{RF} \mathbf{B}_2$, which are not resolvable. To solve it for the analog beamformer \mathbf{G}_{RF} , we apply the result $\text{vec}(\mathbf{A} \mathbf{X} \mathbf{B}) = \mathbf{B}^T \otimes \mathbf{A} \text{vec}(\mathbf{X})$ [136], which allows to rewrite (4.22) as

$$\begin{aligned} \sum_{j \in \mathcal{D}} w_j \left(\left(\mathbf{V}_j \mathbf{V}_j^H (\mathbf{I} + \mathbf{V}_j \mathbf{V}_j^H \mathbf{G}_{RF}^H \mathbf{H}_j^H \mathbf{R}_{\bar{j}}^{-1} \mathbf{H}_j \mathbf{G}_{RF}) \right)^{-1} \right)^T \otimes \\ \mathbf{H}_j^H \mathbf{R}_{\bar{j}}^{-1} \mathbf{H}_j \text{vec}(\mathbf{G}_{RF}) - \sum_{j \in \mathcal{D}} \left(\left(\mathbf{V}_j \mathbf{V}_j^H \right)^T \otimes (\hat{\mathbf{C}}_j \right. \\ \left. + \hat{\mathbf{D}}_j + \Psi_0 + l_0 \mathbf{I}) \right) \text{vec}(\mathbf{G}_{RF}) = 0. \end{aligned} \quad (\text{B.16})$$

The WSR maximizing analog beamformer can alternatively be derived as follows (which allows the proof for the digital beamformers to be applicable directly). First we apply a noise whitening procedure using the noise plus interference covariance matrix $\mathbf{R}_{\bar{j}}^{1/2}$ on the received signal. Further, we can rewrite the whitened signal as follows

$$\tilde{\mathbf{Y}}_j = \left((\mathbf{s}_{j_d}^T \mathbf{V}_j^T) \otimes \mathbf{R}_{\bar{j}}^{-1/2} \mathbf{H}_j \right) \text{vec}(\mathbf{G}_{RF}) + \tilde{\mathbf{n}}_j, \quad (\text{B.17})$$

where $\tilde{\mathbf{y}}_j = \mathbf{R}_{\bar{j}}^{-1/2} \mathbf{y}_j$ and $\tilde{\mathbf{n}}_j$ represents the whitened noise plus interference signal. We can write the resulting WSR optimization problem, after the approximation to concave form and some algebraic manipulations on the linearized term, as

$$\begin{aligned}
 \max_{\mathbf{G}_{RF}} \quad & \sum_{j \in \mathcal{D}} w_j \ln \det \left(\mathbf{I} + \text{vec}(\mathbf{G}_{RF})^H \left((\mathbf{V}_j \mathbf{V}_j^H)^T \otimes \mathbf{H}_j^H \mathbf{R}_j^{-1} \right. \right. \\
 & \left. \left. \mathbf{H}_j \right) \text{vec}(\mathbf{G}_{RF}) \right) - \text{Tr} \left(\text{vec}(\mathbf{G}_{RF})^H \left(\mathbf{V}_j \mathbf{V}_j^H \otimes \right. \right. \\
 & \left. \left. (\hat{\mathbf{C}}_j + \hat{\mathbf{D}}_j) + \boldsymbol{\Psi}_0 + l_0 \mathbf{I} \right) \text{vec}(\mathbf{G}_{RF}) \right). \tag{B.18}
 \end{aligned}$$

Taking the derivative of (C.12) for the conjugate of \mathbf{G}_{RF} leads to the same generalized eigenvector solution as in (4.23). Note that this alternative representation has the same form as (B.11), which is resolvable for the vectorized version of the analog beamformer \mathbf{G}_{RF} . Therefore, the proof for the UL and DL digital beamformers can now be applied directly on the vectorized analog beamformer $\text{vec}(\mathbf{G}_{RF})$, which is summed over all the DL users served by the mmWave FD BS.

Appendix C

Proofs for Chapter 5

C.1 Gradients

To derive the gradients (5.14) and (5.16) to construct the minorized WSR cost function, we apply the following result we derived in Lemma 3 [91].

Lemma C.1.1. *Let $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B} + a \mathbf{A} \text{diag}(\mathbf{X} + \mathbf{Q})\mathbf{B} + b \text{diag}(\mathbf{C}\mathbf{X}\mathbf{D} + \mathbf{E}) + \mathbf{F}$. The derivative of $\text{ln det}(\mathbf{Y})$ with respect to \mathbf{X} is given by*

$$\begin{aligned} \frac{\partial \text{ln det} \mathbf{Y}}{\partial \mathbf{X}} = & \mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H + a \text{diag}(\mathbf{A}^H \mathbf{Y}^{-H} \mathbf{B}^H) \\ & + b \mathbf{C}^H \text{diag}(\mathbf{Y}^{-H}) \mathbf{D}^H. \end{aligned} \quad (\text{C.1})$$

The result stated above can be used to construct the gradients for the multi-cell mMIMO mmWave FD system. To proceed, we write the WSR cost function (5.12) as

$$\begin{aligned} & \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} [w_{k_b} \text{ln det}(\mathbf{R}_{k_b}) - w_{k_b} \text{ln det}(\mathbf{R}_{k_b}^-)] \\ & + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} [w_{j_b} \text{ln det}(\mathbf{R}_{j_b}) - w_{j_b} \text{ln det}(\mathbf{R}_{j_b}^-)] \end{aligned} \quad (\text{C.2})$$

To compute the gradient $\hat{\mathbf{G}}_{k_b, b}^{UL}$ to optimize the transmit covariance matrix \mathbf{T}_{k_b} for UL user k_b , in the same cell for which \mathbf{T}_{k_b} acts as interference, we have to linearize with respect to the users $m_b \in \mathcal{U}_b$ with $m_b \neq k_b$. Applying the result in (C.1) for \mathbf{R}_{m_b} as \mathbf{X} and then $\mathbf{R}_{\overline{m_b}}$ as \mathbf{X} $U_b - 1$ times, and considering that \mathbf{X} is Hermitian, yields the gradient

$$\begin{aligned} \hat{\mathbf{G}}_{k_b, b}^{UL} = & \sum_{\substack{m_b \in \mathcal{U}_b \\ m_b \neq k_b}} w_{m_b} [\mathbf{H}_{m_b}^H \mathbf{F}_b^{RF} (\mathbf{R}_{\overline{m_b}}^{-1} - \mathbf{R}_{m_b}^{-1} + \beta_b \text{diag}(\mathbf{R}_{\overline{m_b}}^{-1} - \mathbf{R}_{m_b}^{-1} \\ &)) \mathbf{F}_b^{RFH} \mathbf{H}_{m_b} + k_{m_b} \text{diag}(\mathbf{H}_{m_b}^H \mathbf{F}_b^{RF} (\mathbf{R}_{\overline{m_b}}^{-1} - \mathbf{R}_{m_b}^{-1}) \mathbf{F}_b^{RFH} \mathbf{H}_{m_b})]. \end{aligned} \quad (\text{C.3a})$$

To linearize with respect to the DL users in the same cell for which the transmit covariance matrix \mathbf{T}_{k_b} acts as CI, we first replace \mathbf{R}_{j_b} as \mathbf{X} and then $\mathbf{R}_{j_b}^-$ as \mathbf{X} in (C.1), and repeating it for all the DL users in the same cell leads to the following gradient

$$\begin{aligned} \hat{\mathbf{G}}_{k_b, b}^{DL} = & \sum_{j_b \in \mathcal{D}_b} w_{j_b} [\mathbf{H}_{j_b, k_b}^H (\mathbf{R}_{j_b}^{-1} - \mathbf{R}_{j_b}^{-1} + \beta_{j_b} \text{diag}(\mathbf{R}_{j_b}^{-1} - \mathbf{R}_{j_b}^{-1} \\ &)) \mathbf{H}_{j_b, k_b} + k_{k_b} \text{diag}(\mathbf{H}_{j_b, k_b}^H (\mathbf{R}_{j_b}^{-1} - \mathbf{R}_{j_b}^{-1}) \mathbf{H}_{j_b, k_b})], \end{aligned} \quad (\text{C.3b})$$

By repeating a similar reasoning for all the remaining gradients and applying the results provided in (C.1) yields the gradients

$$\begin{aligned} \hat{\mathbf{G}}_{k_b, \bar{b}}^{UL} = & \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{k_c \in \mathcal{U}_c} w_{k_c} [\mathbf{H}_{c, k_b}^H \mathbf{F}_c^{RF} (\mathbf{R}_{k_c}^{-1} - \mathbf{R}_{k_c}^{-1} + \beta_c \text{diag}(\mathbf{R}_{k_c}^{-1} - \\ & \mathbf{R}_{k_c}^{-1})) \mathbf{F}_c^{RFH} \mathbf{H}_{c, k_b} + k_{k_b} \text{diag}(\mathbf{H}_{c, k_b}^H \mathbf{F}_c^{RF} (\mathbf{R}_{k_c}^{-1} - \mathbf{R}_{k_c}^{-1}) \\ & \mathbf{F}_c^{RFH} \mathbf{H}_{c, k_b})], \end{aligned} \quad (\text{C.3c})$$

$$\begin{aligned} \hat{\mathbf{G}}_{k_b, \bar{b}}^{DL} = & \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{j_c \in \mathcal{D}_c} w_{j_c} [\mathbf{H}_{j_c, k_b}^H (\mathbf{R}_{j_c}^{-1} - \mathbf{R}_{j_c}^{-1} + \beta_{j_c} \text{diag}(\mathbf{R}_{j_c}^{-1} - \\ & \mathbf{R}_{j_c}^{-1})) \mathbf{H}_{c, k_b} + k_{k_b} \text{diag}(\mathbf{H}_{c, k_b}^H (\mathbf{R}_{j_c}^{-1} - \mathbf{R}_{j_c}^{-1}) \mathbf{H}_{c, k_b})], \end{aligned} \quad (\text{C.3d})$$

$$\begin{aligned} \hat{\mathbf{G}}_{j_b, b}^{UL} = & \sum_{k_b \in \mathcal{U}_b} w_{k_b} [\mathbf{H}_{b, b}^H \mathbf{F}_b^{RF} (\mathbf{R}_{k_b}^{-1} - \mathbf{R}_{k_b}^{-1} + \beta_b \text{diag}(\mathbf{R}_{k_b}^{-1} - \mathbf{R}_{k_b}^{-1})) \\ & \mathbf{F}_b^{RFH} \mathbf{H}_{b, b} + k_{k_b} \text{diag}(\mathbf{H}_{b, b}^H \mathbf{F}_b^{RF} (\mathbf{R}_{k_b}^{-1} - \mathbf{R}_{k_b}^{-1}) \\ & \mathbf{F}_b^{RFH} \mathbf{H}_{b, b})], \end{aligned} \quad (\text{C.3e})$$

$$\begin{aligned} \hat{\mathbf{G}}_{j_b, b}^{DL} = & \sum_{\substack{l_b \in \mathcal{D}_b \\ l_b \neq j_b}} w_{l_b} [\mathbf{H}_{l_b}^H (\mathbf{R}_{l_b}^{-1} - \mathbf{R}_{l_b}^{-1} + \beta_{l_b} \text{diag}(\mathbf{R}_{l_b}^{-1} - \mathbf{R}_{l_b}^{-1})) \\ & \mathbf{H}_{l_b}^H + k_c \text{diag}(\mathbf{H}_{l_b}^H (\mathbf{R}_{l_b}^{-1} - \mathbf{R}_{l_b}^{-1}) \mathbf{H}_{l_b})], \end{aligned} \quad (\text{C.3f})$$

$$\begin{aligned} \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} = & \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{k_c \in \mathcal{U}_c} w_{k_c} [\mathbf{H}_{c, b}^H \mathbf{F}_c^{RF} (\mathbf{R}_{k_c}^{-1} - \mathbf{R}_{k_c}^{-1} + \beta_c \text{diag}(\mathbf{R}_{k_c}^{-1} - \\ & \mathbf{R}_{k_c}^{-1})) \mathbf{F}_c^{RFH} \mathbf{H}_{c, b} + k_b \text{diag}(\mathbf{H}_{c, b}^H \mathbf{F}_c^{RF} (\mathbf{R}_{k_c}^{-1} - \mathbf{R}_{k_c}^{-1}) \\ & \mathbf{F}_c^{RFH} \mathbf{H}_{c, b})], \end{aligned} \quad (\text{C.3g})$$

$$\begin{aligned} \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL} = & \sum_{\substack{c \in \mathcal{B} \\ c \neq b}} \sum_{j_c \in \mathcal{D}_c} w_{j_c} [\mathbf{H}_{j_c, b}^H (\mathbf{R}_{j_c}^{-1} - \mathbf{R}_{j_c}^{-1} + \beta_{j_c} \text{diag}(\mathbf{R}_{j_c}^{-1} - \\ & \mathbf{R}_{j_c}^{-1})) \mathbf{H}_{j_c, b} + k_b \text{diag}(\mathbf{H}_{j_c, b}^H (\mathbf{R}_{j_c}^{-1} - \mathbf{R}_{j_c}^{-1}) \mathbf{H}_{j_c, b})]. \end{aligned} \quad (\text{C.3h})$$

C.2 Proof of Theorem 6

The proof of Theorem 6 follows similarly as in Appendix B [91]. We proceed with the proof by considering the minorized version of the optimization problem

$$\begin{aligned}
 \max_{\substack{\mathbf{U}, \mathbf{V}, \\ \mathbf{G}^{RF}, \mathbf{F}^{RF}}} \sum_{b \in \mathcal{B}} \sum_{k_b \in \mathcal{U}_b} & \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RF} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RFH} \mathbf{H}_{k_b} \right. \\
 & \left. \mathbf{U}_{k_b}) - \text{Tr}(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{k_b, b}^{UL} + \hat{\mathbf{G}}_{k_b, b}^{DL} + \hat{\mathbf{G}}_{k_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b, \bar{b}}^{DL} \right. \\
 & \left. + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b}) \right] + \sum_{b \in \mathcal{B}} \sum_{j_b \in \mathcal{D}_b} \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \right. \\
 & \left. \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b, b}^{UL} + \hat{\mathbf{G}}_{j_b, b}^{DL} \right. \\
 & \left. + \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right]. \tag{C.4}
 \end{aligned}$$

We first consider the optimization only with respect to \mathbf{V}_{j_b} . The proof for the digital beamformer \mathbf{U}_{k_b} will follow similarly. When optimizing \mathbf{V}_{j_b} , all the remaining variables are fixed and their information from their previous update is saved in the gradients. Therefore, from (C.4), only the following optimization problem needs to be considered

$$\begin{aligned}
 \max_{\mathbf{V}_{j_b}} & \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right. \\
 & \left. - \text{Tr}(\mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b, b}^{UL} + \hat{\mathbf{G}}_{j_b, b}^{DL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL} \right. \\
 & \left. + \psi_b \mathbf{I}) \mathbf{G}_b^{RF} \mathbf{V}_{j_b}) \right]. \tag{C.5}
 \end{aligned}$$

The proof consists in simplifying (C.5), until the Hadamard's inequality applies as in Proposition 1 [38] or Theorem 1 [134]. The Cholesky decomposition of the matrix $(\mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b, b}^{UL} + \hat{\mathbf{G}}_{j_b, b}^{DL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b, \bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF})$ can be written as $\mathbf{L}_{j_b} \mathbf{L}_{j_b}^H$ where \mathbf{L}_{j_b} is the lower triangular Cholesky factor. We can define $\widetilde{\mathbf{V}}_{j_b} = \mathbf{L}_{j_b}^H \mathbf{V}_{j_b}$, which allows to write (C.5) as

$$\begin{aligned}
 \max_{\mathbf{V}_{j_b}} & \left[w_{j_b} \ln \det(\mathbf{I} + \widetilde{\mathbf{V}}_{j_b}^H \mathbf{L}_{j_b}^{-1} \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \right. \\
 & \left. \mathbf{L}_{j_b}^{-H} \widetilde{\mathbf{V}}_{j_b}) - \text{Tr}(\widetilde{\mathbf{V}}_{j_b}^H \widetilde{\mathbf{V}}_{j_b}) \right]. \tag{C.6}
 \end{aligned}$$

Let $\mathbf{E}_{j_b} \mathbf{D}_{j_b} \mathbf{E}_{j_b}^H$ be the eigen-decomposition of $\mathbf{L}_{j_b}^{-1} \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} \mathbf{L}_{j_b}^{-H}$, where \mathbf{E}_{j_b} is a unitary matrices and \mathbf{D}_{j_b} is diagonal. Let $\mathbf{O}_{j_b} = \mathbf{L}_{j_b}^H \widetilde{\mathbf{V}}_{j_b} \widetilde{\mathbf{V}}_{j_b}^H \mathbf{L}_{j_b}$, and we can express (C.6) as

$$\max_{\mathbf{O}_{j_b}} \left[w_{j_b} \ln \det(\mathbf{I} + \mathbf{O}_{j_b} \mathbf{D}_{j_b}) - \text{Tr}(\mathbf{O}_{j_b}) \right]. \tag{C.7}$$

By invoking the Hadamard's inequality [Page 233 [135]], we can see that the optimal \mathbf{O}_{j_b}

must be diagonal. Therefore, $\mathbf{U}_{j_b} = \mathbf{L}_{j_b}^{-H} \mathbf{E}_{j_b} \mathbf{O}_{j_b}^{\frac{1}{2}}$ and thus

$$\begin{aligned} \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF} &= \mathbf{L}_{j_b} \mathbf{L}_{j_b}^H \mathbf{L}_{j_b}^{-H} \mathbf{E}_{j_b} \mathbf{O}_{j_b}^{\frac{1}{2}} \mathbf{D}_k \\ &= \mathbf{G}_b^{RFH} (\hat{\mathbf{G}}_{j_b,b}^{UL} + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I}) \mathbf{G}_b^{RF}, \end{aligned} \quad (\text{C.8})$$

from which we select d_{j_b} dominant generalized eigenvectors, equal to the number of data streams to be transmitted, which concludes the proof for the digital beamformer \mathbf{V}_{j_b} . For the digital DL beamformers \mathbf{U}_{k_b} , the result mentioned above holds immediately by applying it to the optimization problem

$$\begin{aligned} \max_{\mathbf{U}_{k_b}} & \left[w_{k_b} \ln \det(\mathbf{I} + \mathbf{U}_{k_b}^H \mathbf{H}_{k_b}^H \mathbf{F}_b^{RFH} \mathbf{R}_{k_b}^{-1} \mathbf{F}_b^{RF} \mathbf{H}_{k_b} \mathbf{U}_{k_b}) \right. \\ & \left. - \text{Tr}(\mathbf{U}_{k_b}^H (\hat{\mathbf{G}}_{k_b,b}^{UL} + \hat{\mathbf{G}}_{k_b,b}^{DL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{k_b,\bar{b}}^{DL} + \lambda_{k_b} \mathbf{I}) \mathbf{U}_{k_b}) \right]. \end{aligned} \quad (\text{C.9})$$

and simplifying the terms in $\log \det(\cdot)$ and $\text{Tr}(\cdot)$ until the Hadamard's inequality applies to yield a result similar to (C.8).

C.3 Proof of Theorem 7

To provide the proof for the analog beamformer, the results stated above cannot be applied directly as the KKT condition (5.24) is not resolvable for \mathbf{G}_b^{RF} , having the form $\mathbf{A}_1 \mathbf{G}_b^{RF} \mathbf{A}_2 = \mathbf{B}_1 \mathbf{G}_b^{RF} \mathbf{B}_2$. To solve it for \mathbf{G}_b^{RF} , we apply the result $\text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{B}^T \otimes \mathbf{A} \text{vec}(\mathbf{X})$ [91], which allows to rewrite (5.24) as

$$\begin{aligned} \sum_{j_b \in \mathcal{D}_b} & ((\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H (\mathbf{I} + \mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \mathbf{G}_b^{RFH} \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b} \mathbf{G}_b^{RF})^{-1})^T \otimes \\ & \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \mathbf{H}_{j_b}) \text{vec}(\mathbf{G}_b^{RF}) - ((\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H)^T \otimes (\hat{\mathbf{G}}_{j_b,b}^{UL} \\ & + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I})) \text{vec}(\mathbf{G}_b^{RF}) = 0. \end{aligned} \quad (\text{C.10})$$

which now become resolvable for $\text{vec}(\mathbf{G}_b^{RF})$. To get to the KKT condition (C.10), we first consider rewriting the cost function such that, taking its derivative leads to (C.10). Firstly, we consider applying a noise whitening procedure using the noise plus interference covariance matrix $\mathbf{R}_{j_b}^{1/2}$ on the received signal. We can rewrite the whitened signal as

$$\tilde{\mathbf{y}}_{j_b} = \left((\mathbf{s}_{j_b}^T \mathbf{V}_{j_b}^T) \otimes \mathbf{R}_{j_b}^{-1/2} \mathbf{H}_{j_b} \right) \text{vec}(\mathbf{G}_b^{RF}) + \tilde{\mathbf{n}}_{j_b}, \quad (\text{C.11})$$

where $\tilde{\mathbf{y}}_{j_b} = \mathbf{R}_{j_b}^{-1/2} \mathbf{y}_{j_b}$ and $\tilde{\mathbf{n}}_{j_b}$ represents the whitened noise plus interference signal. The resulting WSR optimization problem by computing the minorizers for all the links with respect to the unconstrained analog beamformer \mathbf{G}_b^{RF} can be stated as

$$\begin{aligned} \max_{\mathbf{G}_{RF}} & \sum_{j_b \in \mathcal{D}} \left[w_{j_b} \ln \det(\mathbf{I} + \text{vec}(\mathbf{G}_{RF})^H ((\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H)^T \otimes \mathbf{H}_{j_b}^H \mathbf{R}_{j_b}^{-1} \right. \\ & \mathbf{H}_{j_b}) \text{vec}(\mathbf{G}_b^{RF})) - \text{Tr}(\text{vec}(\mathbf{G}_b^{RF})^H (\mathbf{V}_{j_b} \mathbf{V}_{j_b}^H \otimes (\hat{\mathbf{G}}_{j_b,b}^{UL} \\ & + \hat{\mathbf{G}}_{j_b,b}^{DL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{UL} + \hat{\mathbf{G}}_{j_b,\bar{b}}^{DL} + \psi_b \mathbf{I})) \text{vec}(\mathbf{G}_{RF})) \left. \right]. \end{aligned} \quad (\text{C.12})$$

By taking the derivative of (C.12) leads to the KKT condition (C.10). Note that the restatement of the whitened version stated with $\text{vec}(\mathbf{G}_b^{RF})$ has the same form as the digital beamformer \mathbf{V}_{j_b} or \mathbf{U}_{k_b} . By following a similar proof for \mathbf{V}_{j_b} , now it can be easily shown that we can optimize the analog beamformer $\text{vec}(\mathbf{G}_b^{RF})$ as one generalized dominant eigenvector, which is summed over all the users in DL in the same cell.

Bibliography

- [1] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive mimo networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [2] P. Xia, R. W. Heath, and N. Gonzalez-Prelcic, “Robust analog precoding designs for millimeter wave mimo transceivers with frequency and time division duplexing,” *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4622–4634, 2016.
- [3] P. W. Chan, E. S. Lo, R. R. Wang, E. K. Au, V. K. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. B. Letaief, “The evolution path of 4g networks: Fdd or tdd?” *IEEE Communications Magazine*, vol. 44, no. 12, pp. 42–50, 2006.
- [4] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, “Massive mimo performance—tdd versus fdd: What do measurements say?” *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2247–2261, 2018.
- [5] J. Choi, D. J. Love, and P. Bidigare, “Downlink training techniques for fdd massive mimo systems: Open-loop and closed-loop training with memory,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, 2014.
- [6] Y. Ramamoorthi and A. Kumar, “Dynamic time division duplexing for downlink/uplink decoupled millimeter wave-based cellular networks,” *IEEE Communications Letters*, vol. 23, no. 8, pp. 1441–1445, 2019.
- [7] J. M. B. da Silva, G. Wikström, R. K. Mungara, and C. Fischione, “Full duplex and dynamic TDD: Pushing the limits of spectrum reuse in multi-cell communications,” *IEEE Wireless Communications*, vol. 28, no. 1, pp. 44–50, Feb. 2021.
- [8] M. Heino, D. Korpi, T. Huusari, E. Antonio-Rodriguez, S. Venkatasubramanian, T. Riihonen, L. Anttila, C. Icheln, K. Haneda, R. Wichman *et al.*, “Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays,” *IEEE Communications magazine*, vol. 53, no. 5, pp. 91–101, 2015.
- [9] T. Riihonen, S. Werner, and R. Wichman, “Mitigation of loopback self-interference in full-duplex mimo relays,” *IEEE transactions on signal processing*, vol. 59, no. 12, pp. 5983–5993, 2011.

-
- [10] —, “Comparison of full-duplex and half-duplex modes with a fixed amplify-and-forward relay,” in *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009, pp. 1–5.
 - [11] D. Bharadia, E. McMillin, and S. Katti, “Full duplex radios,” in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, 2013.
 - [12] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, “In-band full-duplex wireless: Challenges and opportunities,” *IEEE Journal on selected areas in communications*, vol. 32, no. 9, pp. 1637–1652, 2014.
 - [13] C. B. Barneto, S. D. Liyanaarachchi, M. Heino, T. Riihonen, and M. Valkama, “Full duplex radio/radar technology: The enabler for advanced joint communication and sensing,” *IEEE Wireless Communications*, vol. 28, no. 1, pp. 82–88, Feb. 2021.
 - [14] R. Askar, T. Kaiser, B. Schubert, T. Haustein, and W. Keusgen, “Active self-interference cancellation mechanism for full-duplex wireless transceivers,” in *2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*. IEEE, 2014, pp. 539–544.
 - [15] D. Korpi, Y.-S. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, “Adaptive nonlinear digital self-interference cancellation for mobile inband full-duplex radio: Algorithms and rf measurements,” in *2015 IEEE global communications conference (GLOBECOM)*. IEEE, 2015, pp. 1–7.
 - [16] M. Duarte, C. Dick, and A. Sabharwal, “Experiment-driven characterization of full-duplex wireless systems,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4296–4307, 2012.
 - [17] A. Sahai, G. Patel, and A. Sabharwal, “Pushing the limits of full-duplex: Design and real-time implementation,” *arXiv preprint arXiv:1107.0607*, 2011.
 - [18] D. Bharadia and S. Katti, “Full duplex mimo radios,” in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2014.
 - [19] P. Rosson, C. K. Sheemar, N. Valecha, and D. Slock, “Towards massive mimo in-band full duplex radio,” in *IEEE 16th International Symposium on Wireless Communication Systems (ISWCS)*, 2019.
 - [20] G. C. Alexandropoulos and M. Duarte, “Joint design of multi-tap analog cancellation and digital beamforming for reduced complexity full duplex mimo systems,” in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–7.
 - [21] M. Duarte and G. C. Alexandropoulos, “Full duplex mimo digital beamforming with reduced complexity auxtx analog cancellation,” in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

- [22] M. A. Islam, G. C. Alexandropoulos, and B. Smida, "A unified beamforming and a/d self-interference cancellation design for full duplex mimo radios," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–7.
- [23] —, "Joint analog and digital transceiver design for wideband full duplex mimo systems," *arXiv preprint arXiv:2112.08970*, 2021.
- [24] —, "Integrated sensing and communication with millimeter wave full duplex hybrid beamforming," *arXiv preprint arXiv:2201.05240*, 2022.
- [25] I. P. Roberts, H. B. Jain, and S. Vishwanath, "Equipping millimeter-wave full-duplex with analog self-interference cancellation," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.
- [26] P. Stoica and Y. Selén, "Cyclic Minimizers, Majorization Techniques, and the Expectation-Maximization Algorithm: A Refresher," *IEEE Sig. Proc. Mag.*, Jan. 2004.
- [27] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Rate gain region and design tradeoffs for full-duplex wireless communications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, 2013.
- [28] M. A. Khojastepour and S. Rangarajan, "Wideband digital cancellation for full-duplex communications," in *IEEE Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012.
- [29] T. Riihonen and R. Wichman, "Analog and digital self-interference cancellation in full-duplex mimo-ofdm transceivers with limited resolution in a/d conversion," in *IEEE conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR)*, 2012.
- [30] M. Bernhardt, F. Gregorio, J. Cousseau, and T. Riihonen, "Self-interference cancellation through advanced sampling," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, 2018.
- [31] M. Zamanighomi, Z. Wang, K. Slavakis, and G. B. Giannakis, "Linear minimum mean-square error estimation based on high-dimensional data with missing values," in *IEEE 48th Annual Conference on Information Sciences and Systems (CISS)*, 2014.
- [32] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- [33] S. Wagner, R. Couillet, M. Debbah, and D. T. Slock, "Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback," *IEEE transactions on information theory*, vol. 58, no. 7, 2012.

-
- [34] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Transactions on signal processing*, vol. 55, no. 6, pp. 2646–2660, 2007.
 - [35] A. C. Cirik, L. Zhou, and T. Ratnarajah, "Linear transceiver design with per-antenna power constraints in full-duplex multi-user mimo systems," *IEEE Wireless Communications Letters*, vol. 5, no. 4, pp. 412–415, 2016.
 - [36] D. Korpi, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, "Full-duplex transceiver system calculations: Analysis of adc and linearity challenges," *IEEE Transactions on Wireless Communications*, 2014.
 - [37] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional mimo: Achievable rates under limited dynamic range," *IEEE Transactions on Signal Processing*, 2012.
 - [38] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for mimo ad hoc cognitive radio networks," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3117–3131, 2011.
 - [39] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2740–2746, 2007.
 - [40] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
 - [41] A. C. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-duplex mimo interference channels," *IEEE Transactions on communications*, 2015.
 - [42] A. C. Cirik, R. Wang, and Y. Hua, "Weighted-sum-rate maximization for bi-directional full-duplex mimo systems," in *ASILOMAR*. IEEE, 2013.
 - [43] K. Gopala and D. Slock, "Antenna array absolute self-calibration and application to separate tx/rx array full duplex mimo," in *2017 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–6.
 - [44] H. Abbas and K. Hamdi, "Full duplex relay in millimeter wave backhaul links," in *IEEE Wireless Communications and Networking Conference*, Apr. 2016, pp. 1–6.
 - [45] S. Han, Y. Zhang, W. Meng, C. Li, and Z. Zhang, "Full-duplex relay-assisted macrocell with millimeter wave backhauls: Framework and prospects," *IEEE Network*, vol. 33, no. 5, pp. 190–197, Oct. 2019.
 - [46] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 2, pp. 1199–1215, Jan. 2019.

- [47] L. Zhu, J. Zhang, Z. Xiao, X. Cao, X.-G. Xia, and R. Schober, "Millimeter-wave full-duplex uav relay: Joint positioning, beamforming, and power control," *IEEE JSAC*, 2020.
- [48] Y. Zhang, M. Xiao, S. Han, M. Skoglund, and W. Meng, "On precoding and energy efficiency of full-duplex millimeter-wave relays," *IEEE Transactions on Wireless Communications*, 2019.
- [49] Y. Cai, K. Xu, A. Liu, M. Zhao, B. Champagne, and L. Hanzo, "Two-timescale hybrid analog-digital beamforming for mmwave full-duplex mimo multiple-relay aided systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2086–2103, Jun. 2020.
- [50] Z. Luo, C. Zhan, L. Zhang, and R. Zhang, "Robust hybrid beamforming in millimeter wave relay networks with imperfect csi," *IEEE Access*, 2018.
- [51] R. López-Valcarce and N. González-Prelcic, "Beamformer design for full-duplex amplify-and-forward millimeter wave relays," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2019, pp. 86–90.
- [52] —, "Analog beamforming for full-duplex millimeter wave communication," in *IEEE 16th International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2019, pp. 687–691.
- [53] C. K. Sheemar and D. T. Slock, "Hybrid beamforming for bidirectional massive MIMO full duplex under practical considerations," in *IEEE 93rd Vehicular Technology Conference (VTC) Spring*, Apr. 2021.
- [54] K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, 2018.
- [55] E. Balti, N. Mensi, and S. Yan, "A modified zero-forcing max-power design for hybrid beamforming full-duplex systems," *arXiv preprint arXiv:2003.00147*, 2020.
- [56] S. Huang, Y. Ye, and M. Xiao, "Learning based hybrid beamforming design for full-duplex millimeter wave systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, pp. 120–132, Mar. 2020.
- [57] J. Palacios, J. Rodriguez-Fernandez, and N. González-Prelcic, "Hybrid precoding and combining for full-duplex millimeter wave communication," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [58] O. Taghizadeh, V. Radhakrishnan, A. C. Cirik, R. Mathar, and L. Lampe, "Hardware impairments aware transceiver design for bidirectional full-fuplex MIMO OFDM systems," *IEEE Trans. on Vehic. Tech.*, vol. 67, no. 8, Aug. 2018.

-
- [59] A. Alkhateeb and R. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. on Commun.*, vol. 64, no. 5, May 2016.
 - [60] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Commun.*, December 2008.
 - [61] A. C. Cirik, R. Wang, , Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-fuplex MIMO interference channels," *IEEE Trans. on Commun.*, vol. 63, no. 3, Mar. 2015.
 - [62] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," in *URL* <http://www2.imm.dtu.dk/pubdb/p.php?3274>., November 2011.
 - [63] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, Sep. 2011.
 - [64] 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*, 2008.
 - [65] C. K. Thomas, C. K. Sheemar, and D. Slock, "Multi-stage/hybrid bf under limited dynamic range for ofdm fd backhaul with mimo si nulling," in *IEEE 16th International Symposium on Wireless Communication Systems (ISWCS)*, 2019.
 - [66] C. K. Sheemar and D. Slock, "Massive MIMO mmwave full duplex relay for IAB with limited dynamic range," in *IEEE 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Apr. 2021, pp. 1–5.
 - [67] J. M. B. da Silva, A. Sabharwal, G. Fodor, and C. Fischione, "1-bit phase shifters for large-antenna full-duplex mmwave communications," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6916–6931, Oct. 2020.
 - [68] M.-M. Zhao, Y. Cai, M.-J. Zhao, Y. Xu, and L. Hanzo, "Robust joint hybrid analog-digital transceiver design for full-duplex mmwave multicell systems," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4788–4802, Aug. 2020.
 - [69] I. P. Roberts, J. G. Andrews, and S. Vishwanath, "Hybrid beamforming for millimeter wave full-duplex under limited receive dynamic range," *arXiv preprint arXiv:2012.11647*, 2020.
 - [70] I. P. Roberts and S. Vishwanath, "Beamforming cancellation design for millimeter-wave full-duplex," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2019, pp. 1–6.

- [71] C. K. Sheemar and D. Slock, “Hybrid beamforming and combining for millimeter wave full duplex massive MIMO interference channel,” *arXiv preprint arXiv:2108.00465*, 2021.
- [72] I. P. Roberts, H. B. Jain, and S. Vishwanath, “Frequency-selective beamforming cancellation design for millimeter-wave full-duplex,” in *IEEE International Conference on Communications (ICC)*, Jun. 2020, pp. 1–6.
- [73] C. K. Sheemar, L. Badia, and S. Tomasin, “Game-theoretic mode scheduling for dynamic TDD in 5G systems,” *IEEE Communications Letters*, vol. 25, no. 7, pp. 2425 – 2429, Apr. 2021.
- [74] H. Kim, J. Kim, and D. Hong, “Dynamic TDD systems for 5G and beyond: A survey of cross-link interference mitigation,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2315–2348, Jul. 2020.
- [75] E. de Olivindo Cavalcante, G. Fodor, Y. C. Silva, and W. C. Freitas, “Bidirectional sum-power minimization beamforming in dynamic TDD MIMO networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 9988–10 002, Aug. 2019.
- [76] S. Guo, X. Hou, and H. Wang, “Dynamic TDD and interference management towards 5G,” in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018, pp. 1–6.
- [77] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [78] A. C. Cirik, L. Zhou, and T. Ratnarajah, “Linear transceiver design with per-antenna power constraints in full-duplex multi-user mimo systems,” *IEEE Wireless Communications Letters*, 2016.
- [79] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato, “Hybrid multi-user precoding with amplitude and phase control,” in *IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.
- [80] M. Majidzadeh, J. Kaleva, N. Tervo, H. Pennanen, A. Tölli, and M. Latva-Aho, “Rate maximization for partially connected hybrid beamforming in single-user MIMO systems,” in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2018, pp. 1–5.
- [81] L. Chen, A. Liu, and X. Yuan, “Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4635–4639, Dec. 2017.
- [82] I. P. Roberts, J. G. Andrews, H. B. Jain, and S. Vishwanath, “Millimeter-wave full duplex radios: New challenges and techniques,” *IEEE Wireless Communications*, vol. 28, no. 1, pp. 36–43, Feb. 2021.

-
- [83] T. M. Kim, H. J. Yang, and A. J. Paulraj, "Distributed sum-rate optimization for full-duplex MIMO system under limited dynamic range," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 555–558, Feb. 2013.
- [84] R. Mudumbai, B. Wild, U. Madhow, and K. Ramchandran, "Distributed beamforming using 1 bit feedback: from concept to realization," in *IEEE 44th Annu. Allert. Conf. Commun. Control Comput.*, vol. 8, Sept. 2006, pp. 1020–1027.
- [85] R. Zakhour, Z. K. Ho, and D. Gesbert, "Distributed beamforming coordination in multicell MIMO channels," in *IEEE 69th Veh. Technol. Conf. (VTC)*, Apr. 2009, pp. 1–5.
- [86] R. A. Iltis, S.-J. Kim, and D. A. Hoang, "Noncooperative iterative MMSE beamforming algorithms for ad hoc networks," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 748–759, Apr. 2006.
- [87] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Info. Theory*, vol. 54, no. 12, pp. 5491–5499, Nov. 2008.
- [88] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distributed methods for constrained nonconvex optimization—part i: Theory," *IEEE Trans. Signal*, vol. 65, no. 8, pp. 1929–1944, Dec. 2016.
- [89] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, "A parallel decomposition method for nonconvex stochastic multi-agent optimization problems," *IEEE Trans. Signal*, vol. 64, no. 11, pp. 2949–2964, Feb. 2016.
- [90] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal*, vol. 62, no. 3, pp. 641–656, Nov. 2013.
- [91] C. K. Sheemar and C. K. T. D. Slock, "Practical hybrid beamforming for millimeter wave massive MIMO full duplex with limited dynamic range," *arXiv preprint arXiv:2104.11537*, 2021.
- [92] M. Rebato, L. Rose, and M. Zorzi, "Tilt angle optimization in dynamic TDD mmwave cellular scenarios," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2637–2641, Jul. 2020.
- [93] I. P. Roberts, J. G. Andrews, H. B. Jain, and S. Vishwanath, "Millimeter-wave full duplex radios: New challenges and techniques," *IEEE Wirel. Commun.*, vol. 28, no. 1, pp. 36–43, Feb. 2021.
- [94] L. T. H. An and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Oper. Res.*, vol. 133, no. 1, pp. 23–46, Jan. 2005.

- [95] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.
- [96] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable ai meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.
- [97] 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 *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–6.
- [98] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 9175–9179.
- [99] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, M. D. Renzo, and F. Chen, "Reconfigurable intelligent surfaces with reflection pattern modulation: Beamforming design and performance analysis," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 741–754, 2021.
- [100] G. C. Alexandropoulos, G. Lerosey, M. Debbah, and M. Fink, "Reconfigurable intelligent surfaces and metamaterials: The potential of wave propagation control for 6g wireless communications," *arXiv preprint arXiv:2006.11136*, 2020.
- [101] H.-T. Chen, A. J. Taylor, and N. Yu, "A review of metasurfaces: physics and applications," *Reports on progress in physics*, vol. 79, no. 7, p. 076401, 2016.
- [102] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018.
- [103] X. Tan, Z. Sun, J. M. Jornet, and D. Pados, "Increasing indoor spectrum sharing capacity using smart reflect-array," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.
- [104] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 270–278.
- [105] S. Foo, "Liquid-crystal reconfigurable metasurface reflectors," in *2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*. IEEE, 2017, pp. 2069–2070.

-
- [106] G. C. Alexandropoulos, V. I. Barousis, and C. B. Papadias, "Precoding for multiuser mimo systems with single-fed parasitic antenna arrays," in *2014 IEEE Global Communications Conference*. IEEE, 2014, pp. 3897–3902.
 - [107] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "Using any surface to realize a new paradigm for wireless communications," *Communications of the ACM*, vol. 61, no. 11, pp. 30–33, 2018.
 - [108] S. Hu, F. Rusek, and O. Edfors, "Beyond massive mimo: The potential of data transmission with large intelligent surfaces," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2746–2758, 2018.
 - [109] S. Hu, K. Chitti, F. Rusek, and O. Edfors, "User assignment with distributed large intelligent surface (lis) systems," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2018, pp. 1–6.
 - [110] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.
 - [111] S. Hu, F. Rusek, and O. Edfors, "Capacity degradation with modeling hardware impairment in large intelligent surface," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.
 - [112] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 1, pp. 183–198, 2013.
 - [113] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical csi," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8238–8242, 2019.
 - [114] L. Zhang, X. Q. Chen, S. Liu, Q. Zhang, J. Zhao, J. Y. Dai, G. D. Bai, X. Wan, Q. Cheng, G. Castaldi *et al.*, "Space-time-coding digital metasurfaces," *Nature communications*, vol. 9, no. 1, pp. 1–11, 2018.
 - [115] H. Yang, X. Cao, F. Yang, J. Gao, S. Xu, M. Li, X. Chen, Y. Zhao, Y. Zheng, and S. Li, "A programmable metasurface with dynamic polarization, scattering and focusing control," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.
 - [116] G. Pan, J. Ye, J. An, and M.-S. Alouini, "When full-duplex transmission meets intelligent reflecting surface: Opportunities and challenges," *arXiv preprint arXiv:2005.12561*, 2020.
 - [117] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Transactions on Signal Processing*, vol. 69, pp. 837–851, 2021.

- [118] G. Pan, J. Ye, J. An, and M.-S. Alouini, “Full-duplex enabled intelligent reflecting surface systems: Opportunities and challenges,” *IEEE Wireless Communications*, 2021.
- [119] Z. Abdullah, G. Chen, S. Lambbotharan, and J. A. Chambers, “Optimization of intelligent reflecting surface assisted full-duplex relay networks,” *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 363–367, 2020.
- [120] P. K. Sharma and P. Garg, “Intelligent reflecting surfaces to achieve the full-duplex wireless communication,” *IEEE Communications Letters*, vol. 25, no. 2, pp. 622–626, 2020.
- [121] Y. Ge and J. Fan, “Robust secure beamforming for intelligent reflecting surface assisted full-duplex miso systems,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 253–264, 2021.
- [122] X. Zhang, *Matrix analysis and applications*. Cambridge University Press, 2017.
- [123] B. R. Marks and G. P. Wright, “A general inner approximation algorithm for nonconvex mathematical programs,” *Operations research*, vol. 26, no. 4, pp. 681–683, 1978.
- [124] C. Pan, H. Ren, K. Wang, W. Xu, M. El-kashlan, A. Nallanathan, and L. Hanzo, “Multicell mimo communications relying on intelligent reflecting surfaces,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, 2020.
- [125] 3GPP Tech. Spec. Group RAN, “NR; physical layer procedures for control,” *TS 38.213 V15.5.0*, Mar. 2019.
- [126] S. Burer and A. N. Letchford, “Non-convex mixed-integer nonlinear programming: A survey,” *Surveys in Operations Research and Management Science*, vol. 17, no. 2, pp. 97–106, 2012.
- [127] M. Osborne, *An introduction to game theory*. Oxford Univ. Press, 2009.
- [128] Z. Shen, A. Khoryaev, E. Eriksson, and X. Pan, “Dynamic uplink-downlink configuration and interference management in TD-LTE,” *IEEE Communications Magazine*, vol. 50, no. 11, pp. 51–59, 2012.
- [129] A. A. Dowhuszko, O. Tirkkonen, J. Karjalainen, T. Henttonen, and J. Pirskanen, “A decentralized cooperative uplink/downlink adaptation scheme for TDD small cell networks,” in *Proc. Personal Indoor Mobile Radio Commun. Conf. (PIMRC)*, 2013.
- [130] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [131] S. Telen and M. Van Barel, “A stabilized normal form algorithm for generic systems of polynomial equations,” *Journal of Computational and Applied Mathematics*, vol. 342, pp. 119 – 132, 2018.

- [132] M. S. ElBamby, M. Bennis, W. Saad, and M. Latva-Aho, "Dynamic uplink-downlink optimization in TDD-based small cell networks," in *11th International Symposium on Wireless Communications Systems (ISWCS)*. IEEE, Aug. 2014, pp. 939–944.
- [133] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-Duplex bidirectional MIMO: achievable rates under limited dynamic range," *IEEE Trans. on Sig. Process.*, vol. 60, no. 7, July 2012.
- [134] D. Hoang and R. A. Iltis, "Noncooperative eigencoding for MIMO ad-hoc networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 865–869, Jan. 2008.
- [135] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [136] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2019.