MULTI-FUNCTIONAL WIRELESS NETWORKS: INTEGRATED SENSING AND COMMUNICATIONS, FULL-DUPLEX, AND UNLIMITED SENSING

ZIANG LIU

A dissertation submitted for the degree of Doctor of Philosophy of Imperial College London

Supervisor

Professor Bruno Clerckx

IMPERIAL

Communications and Signal Processing Group

Department of Electrical and Electronic Engineering

Imperial College London

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Declaration of Originality

I declare that this thesis is original and is the result of my own work. In the thesis, material, including information and ideas, derived from the published and unpublished work of others has been fully acknowledged. The content of the thesis has not been and will not be submitted to any other institution or university for fulfilling the requirements of another degree.

To my parents and all the people I love.

"Do not go gentle into that good night.
Rage, rage against the dying of the light."

— Dylan Thomas

Abstract

Next generation wireless communication networks are expected to be multi-functional to support emerging requirements, *e.g.* high data transmission rates, high-quality wireless connectivity supporting a large number of devices, and highly accurate and robust sensing capability. To realize these demands, cutting-edge technologies such as integrated sensing and communication (ISAC), massive multiple-input multiple-output (MIMO), satellite communications, and full-duplex (FD) communications system requires further exploration and development. Specifically, challenges persist in optimizing these systems. In mono-static ISAC system, radar returns are drowned by the strong residual self interference (SI) stemming from the transmitter. ISAC satellite communications systems encounter challenges on interference control and power limitations. Massive MIMO systems confront the high power consumption and receiver analog-to-digital converter (ADC) saturation. Additionally, FD communication systems face the difficulties of suppressing SI. To overcome the mentioned challenges, this thesis introduces novel approaches, which are based on convex optimization techniques and unlimited sensing theorem. Subsequently, this thesis provides introductions to the background context, followed by a thorough review of the techniques utilized in the proposed solutions. In the following chapters, explanations of these solutions are presented. Finally, this thesis wraps up with a comprehensive conclusion.

Acknowledgements

It has been quite a long journey since the first day I began my PhD. During this time, every day has been busy. When I look back, I realize how far I have come. During my journey, there are many people I would like to thank for their help.

First of all, I would like to express my gratitude to my supervisor, Bruno Clerckx, for his guidance and support throughout my PhD. He has helped me develop as an independent researcher by providing me with research freedom from the beginning and encouraging me to explore new areas rather than following his established research path. This has greatly influenced this thesis.

During my time at Imperial, there is one person who has been a true friend, collaborator, and mentor. I would like to deeply thank Ayush Bhandari, who has supported me both in research and in navigating research life. I still remember the pizza nights, especially during the pandemic. We discussed research and life, which implicitly enhanced my research skills and career planning. I am grateful for his encouragement when I faced negative thoughts. The most important lesson I learned from him is, "Do not set boundaries on yourself.

I am also grateful to Wonjae Shin, who has helped me extend my ideas to satellite communications. It has been a wonderful collaboration. I still remember the coffee breaks when we talked about research and career development, which benefited me greatly. Thank you also for the many meetings where we discussed research and shared research resources.

I would like to thank Anna Scaglione for hosting me at Cornell University. It was an amazing time in New York City. She introduced me to control theory

and signal processing.

During my PhD, Xinze Lyu has always been supportive. I appreciate the coffee breaks, travels, and our countless discussions. We interacted almost daily during my years in London, and our friendship helped me navigate the stressful times. I also want to thank the members of Bruno Clerckx's lab, which made my PhD journey enjoyable.

There are a few individuals who have also helped me. I would like to thank Antonio Lozano Ortega for our friendship and his encouragement during tough times.

I would also like to express my deep gratitude to Wanxin Zhong, who accompanied me throughout my PhD. Thank you for always being there to help me stay positive about both research and life.

Finally, I want to express my wholehearted gratitude to my family. Words cannot convey how thankful I am to my parents. You gave me the freedom to choose my path and helped me pursue my goals. Your love and support have shaped who I am. The encouragement from my family has provided me with infinite energy

London
17 May, 2024

Contents

1 Introduction **22**

1.1 Overview 23

1.1.1 Integrated Sensing and Communications (ISAC) 23

1.1.2 Low earth orbit (LEO)-ISAC system 25

1.1.3 Massive MIMO (M-MIMO) systems 25

1.1.4 Full-duplex systems 26

1.2 Challenges, Scope and Objectives 28

1.2.1 Challenges and Objectives of ISAC Systems 28

1.2.2 Challenges and Objectives of ISAC LEO Satellite Systems 28

1.2.3 Challenges and Objectives of M-MIMO Systems 29

1.2.4 Challenges and Objectives of Full-duplex Systems 30

1.3 Outline and Contributions 32

1.4 Publications 39

2 Background **40**

2.1 Unlimited Sampling 41

2.2 Rate-splitting Multiple Access (RSMA) 47

2.3 Optimization Techniques 50

2.3.1 Block Coordination Descent (BCD) 50

2.3.2 Penalty-dual Decomposition (PDD) 51

2.3.3 Weighted Minimum Mean Square Error (WMMSE) 53

2.3.4 Fractional Programming/Dinkelbach's Transformation . . 54

2.3.5 Successive Convex Approximation (SCA) 56

2.3.6 Karush-Kuhn-Tucker (KKT) Conditions 58
2.3.7 Computational Complexity Comparison 59

3 Joint Transmit and Receive Beamforming Design in Full-Duplex Integrated Sensing and Communications 61
3.1 Introduction 62
3.2 System Model 67
3.2.1 Signal Model 67
3.2.2 Channel Model 69
3.2.3 Problem Formulation 71
3.3 Joint TX-RX Beamformer Design 74
3.3.1 Problem Transformation 74
3.3.2 Penalty-based Block Update Algorithm 76
3.3.3 Algorithm Implementation 82
3.4 Convergence and Complexity Analysis 83
3.4.1 Convergence Analysis 83
3.4.2 Complexity Analysis 85
3.5 Numerical evaluation 86
3.5.1 SI Power After Cancellation 86
3.5.2 Beampattern Power Performance 88
3.5.3 Sum-rate Performance 89
3.5.4 Sensing Performance 90
3.6 Summary 94

4 Rate-Splitting Multiple Access for Quantized ISAC LEO Satellite Systems: A Max-Min Fair Energy-Efficient Beam Design 95
4.1 Introduction 96
4.2 System model 101
4.2.1 Channel Model 102
4.2.2 Signal Model 104
4.3 Minimal Energy Efficiency Maximization and Problem Formulation 107
4.3.1 Communication Metric 107

4.3.2 Sensing Metric: Cramér–Rao Bound . . . 109
4.3.3 Problem Formulation . . . 110
4.4 The Proposed Algorithm for Minimal Energy Efficiency Maximization . . . 112
4.4.1 SDP Transformation . . . 113
4.4.2 CRB Transformation by Schur Complement . . . 118
4.4.3 Rank-one Constraint Transformation by Penalty Function 118
4.4.4 Computational Complexity and Convergence Analysis . . 120
4.4.5 Total EE maximization . . . 121
4.5 Numerical evaluation . . . 124
4.5.1 Simulation Environment . . . 124
4.5.2 Communication Performances . . . 126
4.5.3 Sensing performances . . . 128
4.6 Summary . . . 133

5 λ–MIMO: Massive MIMO via Modulo Sampling 134
5.1 Introduction . . . 135
5.2 System model . . . 142
5.2.1 Transmitter . . . 142
5.2.2 Channel Model . . . 144
5.2.3 Receiver . . . 146
5.3 Receiver Design with $\mathscr{M}_\lambda$-ADC . . . 150
5.3.1 Discrete-Time to Digital Modulo Samples . . . 151
5.3.2 Recovery from Modulo Samples: The Unlimited Sampling Algorithm . . . 153
5.4 Achievable sum-rate analysis . . . 156
5.4.1 SQNR of Uniform Distribution Signal . . . 157
5.4.2 SQNR of Gaussian Distribution Signal . . . 158
5.4.3 Achievable sum-rate Analysis for $\mathscr{M}_\lambda$–ADC in MU λ–MIMO. 159
5.5 Numerical evaluation . . . 162
5.5.1 Recovery and Detection Performance . . . 162
5.5.2 Hardware Experiment . . . 164

5.5.3 Achievable Uplink sum-rates and Analytical Approximations . . . 164
5.6 Summary . . . 172

6 Full-Duplex Beyond Self-Interference: The Unlimited Sensing Way 173
6.1 Introduction . . . 174
6.2 System model . . . 176
6.2.1 Receiver . . . 176
6.2.2 Recovery from folded samples . . . 178
6.2.3 Modulo-domain channel estimation and digital SIC . . . 179
6.2.4 Estimation Procedure . . . 180
6.2.5 Recovery Condition . . . 182
6.3 Numerical evaluation . . . 184
6.3.1 Recovery of received signal and SoI after SIC . . . 184
6.4 Summary . . . 186

7 Conclusions 187
7.1 Summary of Thesis Achievements . . . 188
7.2 Future Research . . . 191
7.2.1 Future Work for Chapter 3 . . . 191
7.2.2 Future Work for Chapter 4 . . . 192
7.2.3 Future Work for Chapter 5 and 6 . . . 193

Appendix 195

Bibliography 197

List of Figures

1.1 The receiver architecture of conventional M-MIMO (M–MIMO) system with uplink user equipment (UE). Each RF chain consists of the low noise amplifier (LNA), mixer, variable gain amplifier (VGA), automatic gain controller (AGC), and ADC. 29

1.2 Illustration of output samples by conventional ADC. f_{Input}: continuous input signal before ADC. f_{Output}: continuous output signal shaped by limited ADC dynamic range. Samples: Sampled f_{Output} signals. In conventional ADC, saturation and clipping errors happen in f_{Output} signal and samples, because some values in f_{Input} exceed the dynamic threshold λ. 30

2.1 Illustration of modulo-folded signal via USF. 41

2.2 Hardware of $\mathscr{M}_{\lambda}$−ADC. 42

2.3 Flow Chart of Reconstruction Algorithm. 44

2.4 Overview of the main idea behind recovering a bandlimited function from modulo samples [1]. 45

2.5 RSMA structure. 47

3.1 Full-duplex integrated sensing and communication system, and illustration of the boundaries and contents of the SI suppression in the propagation, analog, and digital domains for multi-antenna configurations. 64

3.2 Convergence of Algorithm 5 to a stable objective function value with $\epsilon = 10^{-5}$. (a) Objective function versus the number of iterations; (b) Relative difference versus the number of iterations. . . . 85

3.3 Performance comparison of SI cancellation with varying priority parameter ϱ and level of SI. The residual SI level is computed with respect to the noise floor. 87
3.4 Performance comparison of SoI (e.g., radar echo and uplink data) power over SI power in dB. 87
3.5 **Left panel**: Transmit beampattern, **Right panel**: Receive beampattern of the ISAC FD system with a effective cancellation of 60 dB SI, where $N_{\mathrm{t}} = 16$, target at $\theta_{\mathrm{r}} = 45^{\circ}$, and downlink user at $\theta_{\mathrm{d}} = -30^{\circ}$: (a, d) $\varrho = 1$; (b, e) $\varrho = 100$; (c, f) $\varrho = 0.01$. 88
3.6 Frame structures of (a) half-duplex and (b) FD systems. 89
3.7 Average sum-rate in one unit time-frequency resource versus varying level of SI power. 90
3.8 Examples of range and velocity estimation based on the FD ISAC system with 60 dB SI with $N_{\mathrm{t}} = 16, N_{\mathrm{r}} = 16$, $M = 512$ blocks, and $N = 1024$ symbols. 92
3.9 An example of angle estimation for a simulated scenario with a target at $\theta_{\mathrm{r}} = 45^{\circ}$, a uplink communication user at $\theta_{\mathrm{u}} = -50^{\circ}$ with antennas $N_{\mathrm{t}} = 16$ and 60 dB SI power. 93

4.1 Integrated sensing and communications enabled quantized rate-splitting multiple access satellite system. 101
4.2 Transmitter Architecture of RSMA-assisted ISAC-LEO system. . . 102
4.3 Convergence illustration for 2-bit cases (a) Algorithm 6, and (b) Algorithm 7 with 20 dB transmit SNR and 40 dB radar SNR. . . . 125
4.4 Max-min fairness EE versus transmit SNR with the quantization bit $b \in \{2, 4, 8\}$ and 40 dB radar SNR. 127
4.5 Power allocation of each precoder with 2-bit quantization and 20 dB transmit SNR and 40 dB radar SNR for (a) RSMA with $\mathbf{p}_r$ and SIC, (b) RSMA with $\mathbf{p}_r$ and no SIC, (c) RSMA without $\mathbf{p}_r$, (d) RSMA with $\mathbf{p}_r$ by maximizing total EE, (e) RSMA with $\mathbf{p}_r$ by maximizing sum-rate, (f) SDMA by maximizing total EE. 128

4.6 Under the case that 20 dB transmit SNR and 40 dB radar SNR, (a) Max-min fairness EE versus RCRB constraint ρ with $b = 2$, (b) Max-min fairness EE versus number of bits b. 129
4.7 Detection results of Doppler and range with 2-bit quantization, 20 dB transmit SNR and 40 dB radar SNR: Range-Doppler Maps for (a) RSMA with $\mathbf{p}_{\mathrm{r}}$ and SIC, (b) RSMA with $\mathbf{p}_{\mathrm{r}}$ and no SIC, (c) RSMA without $\mathbf{p}_{\mathrm{r}}$, (d) SDMA by maximizing total EE. 130
4.8 Under 2-bit quantization, 20 dB transmit SNR, and 40 dB radar SNR. Case 1: users and targets are distant, transmit beampatterns of (a) common stream precoder $\mathbf{p}_{\mathrm{c}}$, (c) radar sequence precoder $\mathbf{p}_{\mathrm{r}}$. Case 2: users and targets are near, transmit beampatterns of (b) common stream precoder $\mathbf{p}_{\mathrm{c}}$, (d) radar sequence precoder $\mathbf{p}_{\mathrm{r}}$. 131

5.1 The receiver architecture of conventional massive MIMO (M–MIMO) system with uplink user equipment (UE). Each RF chain consists of the LNA, mixer, VGA, AGC, and ADC. 136
5.2 Illustration of output samples by conventional and $\mathscr{M}_{\lambda}$–ADCs. f_{Input}: continuous input signal before ADC. f_{Output}: continuous output signal shaped by limited ADC dynamic range. Samples: Sampled f_{Output} signals. a) In conventional ADC, saturation and clipping errors happen in f_{Output} signal and samples, because some values in f_{Input} exceeds dynamic threshold λ, b) In $\mathscr{M}_{\lambda}$–ADC, the input continuous signal f_{Input} is modulo-folded to $f_{\mathsf{Output},(\mathsf{m})}$. No saturation occurs in the $\mathscr{M}_{\lambda}$–ADC. 137
5.3 Hardware output of $\mathscr{M}_{\lambda}$–ADC (cf. Fig. 5.15) used in the Unlimited Sensing Framework [1–4]. An oscilloscope screen shot plotting the real component of a quadrature phase shift keying (QPSK) signal (ground truth, red) and its folded version (blue). The input signal with dynamic range 13.6V peak-to-peak ($\approx 7\lambda$) is folded into a 4.02V peak-to-peak signal. 139
5.4 Proposed M–MIMO uplink system architecture with $\mathscr{M}_{\lambda}$–ADC in each RF chain. 140

5.5 Channel model between m^{th} user and n^{th} antenna in the BS. . . . 142
5.6 Receiver architectures with $\mathscr{M}_\lambda$–ADCs for single-carrier system. . . 146
5.7 Receiver architectures with $\mathscr{M}_\lambda$–ADCs for multi-carrier (OFDM) system. 147
5.8 The architecture of the $\mathscr{M}_\lambda$–ADC and quantizer. 150
5.9 An illustration of quantization process of normal samples and modulo-folded samples with bit resolution $b = 3$. 152
5.10 Fourier spectrum of r, and 4-bit quantized samples for conventional ADC and $\mathscr{M}_\lambda$–ADC with $\zeta = 0.1$. 159
5.11 Constellation diagrams. (a) $M = 10$ users with $N = 50$ BS antennas and MMSE detector; (b) $M = 10$ users with $N = 50$ BS antennas and ZF detector; (c) $M = 10$ users with $N = 50$ BS antennas and MRC detector; (d) $M = 10$ users with $N = 1000$ BS antennas and MRC detector. 167
5.12 Modulo samples and recovery of QPSK (a) single-carrier signal, (b) OFDM signal, 1024 QAM (c) single-carrier signal, (d) OFDM signal; 2-order difference on samples and modulo-folded samples of QPSK (e) single-carrier signal, (f) OFDM signal, 1024 QAM (g) single-carrier signal, (h) OFDM signal. 168
5.13 SQNR of $\mathscr{M}_\lambda$ and conventional ADCs with $\lambda = 0.1\|r\|_\infty$. The curves with theory are calculated by the theoretical SQNR equations, *i.e.* (5.29), (5.30), (5.33), and (5.34). The curves with simulation are obtained from our numerical simulations, wherein we calculated the power of the transmitted signal, σ_r^2, and the power of the quantization noise, σ_{q}^2, from the simulated signal. Subsequently, the SQNR is calculated by (5.25). 168
5.14 Eye diagrams with roll-off factor $\alpha = 0.45$ for signals (a) after reconstruction, (b) before reconstruction (modulo signal). 169

5.15 Hardware experiments and actual reconstruction results. (a) $\mathscr{M}_\lambda$–ADC circuit. (b) Live screen shot of the oscilloscope plotting the real part of QPSK signal (red) and its modulo-folded signal (blue). (c) Reconstruction results with actual modulo-folded signals from the circuit as the input, and 2-bit and 12-bit resolution quantization. 169
5.16 Uplink sum-rate per cell vs the number of BS antennas, when $M = 10$ users with Tx power $p_{\mathsf{u}} = 10$ dB and BS using (a), (c) MRC detector; (b), (d) ZF detector. 170
5.17 Uplink sum-rate per cell vs the number of BS antennas, when $M = 10$ users with scaled-down Tx power $p_{\mathsf{u}} = \frac{E_{\mathsf{u}}}{M}$ dB and BS using MRC detector. 170
5.18 (a) Uplink sum-rate versus ADC bit resolution b, (b) Energy efficiency versus ADC bit resolution b. 171

6.1 FD model with modulo-domain channel estimation method and digital cancellation based on $\mathscr{M}_\lambda$–ADC. 177
6.2 Simulation results with 4-bit quantization, 1/SIR = 40 dB, and SNR = 40 dB of (a) modulo samples and recovery; (b) SoI reconstruction by our proposed SIC approach. 184

List of Tables

2.1 Computational complexity of various algorithms. 59

3.1 Novelty Comparison with Existing FD ISAC Literature 65

4.1 Novelty Comparison with Existing RSMA-ISAC/Satellite Literature 98

Abbreviations

ADC	Analog-to-digital converter
AGC	Automatic gain controller
AoA	Angle-of-Arrival
AoD	Angle-of-Departure
AQNM	Additive quantization noise model
BCD	Block coordination descent
BER	Bit error rate
BPSK	Binary phase-shift keying
CRB	Cramér–Rao Bound
CP	Cyclic prefix
CSI	Channel state information
CSIT	Channel state information at the transmitter
CPI	Coherent processing interval
DAC	Digital-to-analog converter
DC	Difference of concave
DFT	Discrete Fourier transformation
DR	Dynamic range
EE	Energy efficiency
EVD	Eigenvalue decomposition
FD	Full-duplex
FIM	Fisher information matrix
FFT	Fast Fourier transform
FP	Fractional programming
GEO	Geostationary earth orbit
HDR	High dynamic range
IDFT	Inverse discrete Fourier transformation
IoT	Internet of things
IRS	Intelligent reflecting surfaces
ISAC	Integrated sensing and communication
KKT	Karush–Kuhn–Tucker condition
LEO	Low earth orbit
LDR	Low dynamic range

LoS	Line-of-sight
LMI	Linear matrix inequality
LNA	Low noise amplifier
LS	Least square
MEO	Medium earth orbit
MLE	Maximum likelihood estimation
M-MIMO	Massive multiple-input multiple-output
MMSE	Minimum mean-squared error
MRC	Maximum ratio combining
MSE	Mean squared error
MU	Multi-user
NOMA	Non-orthogonal multiple access
NSBC	Narrowband Single-Carrier
NSP	Null-space projection
OFDM	Orthogonal frequency- division multiplexing
OMA	Orthogonal multiple access
PA	Power amplifer
PDD	Penalty-dual decomposition
QAM	Quadrature amplitude modulation
QoS	Quality-of-service
QPSK	Quadrature phase shift keying
RCS	Radar cross-section
RF	Radio frequency
RSMA	Rate-splitting multiple access
RTT	Round-trip times
SCA	Successive convex approximation
SDMA	Space-division multiple access
SDP	Semi-definite programming
SER	Symbol error rate
SI	Self-interference
SIC	Self-interference cancellation
SNR	Signal to noise ratio
SOC	Second-order cone
SoI	Signal of interest

SQNR	Signal-to-quantization noise ratio
ToA	Time-of-arrival
Tx/Rx	Transmit/Receive antenna
ULA	Uniform linear arrays
UPA	Uniform planar array
USF	Unlimited sensing framework
VGA	Variable gain amplifier
WMMSE	Weighted minimum mean-squared error
WSE	Weighted sum rate
ZF	Zero forcing

Notations

$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of reals
$\mathbb{Z}$	The set of integers
$\mathbb{C}$	The set of integer complex numbers
$\lfloor \mathbb{K} \rceil$	Non-negative integers until K
	$\{k : k \in \mathbb{Z} \cap [1, K+1)\}$ for $K \in \mathbb{R}$
$x(t)$	Continuous signals $t \in \mathbb{R}$
$x[k]$	Discrete sequences $k \in \mathbb{Z}$
$\mathrm{Re}(x)$	Real component of a complex number
$\mathrm{Im}(x)$	Imaginary component of a complex number
$\mathbf{X}$	Matrices are written in capital boldface
$\mathbf{x}$	Vectors are written in small boldface
x	Scalars are written in normal fonts
$[\mathbf{X}]_{i,j}$	The entry of the matrix at index (i, j)
$[\mathbf{x}]_i$	The entry of the vector $\mathbf{x}$
$\mathbf{X}^H$	Conjugate-transpose operation
$\mathbf{X}^\top$	Transpose operation
$\lfloor x \rfloor$	Floor operation that is $\sup\{k \in \mathbb{Z} : k \leqslant x\}$ for $x \in \mathbb{R}$
$\lceil x \rceil$	Ceiling operation that is $\inf\{k \in \mathbb{Z} : k \geqslant x\}$ for $x \in \mathbb{R}$
$\langle , \rangle$	Inner-product defined by $\langle a, b \rangle = \int a(x) b^*(x) dx$
Δ	First-order difference $(\Delta x)[k]$=$x[k+1] - x[k]$
Δ^L	L^{th} order finite difference of the sequence $x[k]$
	$\left(\Delta^L x\right)[k] = \Delta^{L-1}(\Delta x)[k]$
$\underline{*}_T$	Semi-discrete convolution operator with sampling interval T
	$\underline{*}_T : c \in \ell_2 \longmapsto \left(c \underline{*}_T \phi\right)(t) = \sum_{n \in \mathbb{Z}} c[n]\, \phi(t - nT)$
$\underline{*}$	Semi-discrete convolution operator with sampling interval $T = 1$
$\mathscr{E}(\cdot)$	Statistical expectation
$\lvert \cdot \rvert$	Absolute value
$\lVert \cdot \rVert_{\ell_p}$	Euclidean norm
$\circ$	Hadamard product
$f \in C^N(\mathbb{R})$	Ω-bandlimited or, $f \in \mathcal{B}_\Omega \Leftrightarrow \widehat{f}(\omega) = \mathbb{1}_{[-\Omega,\Omega]}(\omega) \widehat{f}(\omega)$
$\mathbb{1}_{\mathcal{D}}(t)$	Indicator function on domain $\mathcal{D}$

Chapter 1

Introduction

1.1 Overview

Next generation wireless communication networks are expected to be *multi-functional* and ultra-flexible, accommodating high data transmission rates, high-quality wireless connectivity with massive devices, while highly accurate and robust sensing are required to be supported [5]. Fulfilling such diverse demands necessitates the development of key enabling technologies. While some technologies such as massive multiple-input multiple-output(MIMO), satellite communications, and full-duplex communications have been actively explored in previous generations, there is a continuous quest for better solutions to existing challenges. For example, in massive MIMO (M-MIMO) systems, issues such as high power consumption and receiver analog-to-digital converter (ADC) saturation have been addressed individually. Is there a method that can simultaneously tackle both of these challenges? Additionally, is there potential to simplify the self-interference (SI) cancellation process, which depends on both the analog and digital domains? Meanwhile, emerging technologies *i.e.* integrated sensing and communications (ISAC), Low earth orbit (LEO)-ISAC, demand thorough investigation to harness their potential for achieving sensing capabilities. These questions raised above have become meaningful unresolved issues deserving further attention and investigation. In the following parts, the details, challenges, and objectives of the studied key enabling technologies are elaborated.

1.1.1 Integrated Sensing and Communications (ISAC)

Sensing capabilities are recognized as a key technique in the next generation of wireless communication networks (*i.e.* 6G) for achieving emerging applications, such as assisted navigation, activity detection and movement tracking, and environmental monitoring [5]. Furthermore, given the significant development in the wireless communications and radar sensing industries, spectrum resources are becoming increasingly scarce and valuable. As a promising solution, ISAC has been proposed to meet these requirements. In ISAC, the dual-functional signal is jointly designed to achieve both communication and sens-

ing functionalities in a single hardware. By adopting the ISAC system, spectral, hardware, and energy resources are efficiently utilized. This efficiency is achieved through unified signal processing procedures and a shared hardware framework between sensing and communication systems.

The original aims of sensing and communication capabilities are different. Sensing, or detection, utilizes known signals to extract target information from noisy observations. In comparison, communication focuses on transmitting random information between devices via modulated signals. Based on the objectives, ISAC systems can be categorized into radar-centric, communication-centric, and joint dual-functional designs [6, 7]. In radar-centric design, radar functionality is prioritized, ensuring that the beampattern performance and target detection capabilities are comparable to conventional radar system. However, since either pulsed or continuous radar waveforms are used, embedding communication data to the radar signal is necessary. Typical methods of information embedding include time-frequency embedding, code-domain embedding, and spatial embedding [8]. This design limits the communication rates as communication is treated as an additional function. Conversely, in communication-centric design, priority is given to communication function with sensing as an adjunct. Communication waveforms, *e.g.* single carrier, and orthogonal frequency division multiplexing (OFDM), are employed. To facilitate the sensing function, enhancement on the hardware and algorithms, *e.g.* beamforming design, are utilized. While communication rates remain unaffected, radar detection performance is dependent on the scenario. The joint dual-functional design is the most flexible and radical approach, which requires fundamental redesign of the hardware, signal, and algorithm. It is thus not limited by the existing communication and radar standards. This thesis primarily delves into communication-centric ISAC studies as the objective is to retain communication capabilities while integrating sensing functions to enable new applications.

1.1.2 Low earth orbit (LEO)-ISAC system

Much of the past ISAC research has been primarily devoted to terrestrial networks, resulting in limitations in providing global services [9]. Actually, satellite communications have been envisioned as a solution to satisfy the growing demand for ubiquitous and high-capacity global connectivity in unserved (*e.g.* marine, and desert) and underserved areas (*e.g.* rural areas). The deployment of satellites and terrestrial stations offers global coverage; thus, it enhances the service availability in such areas. If one satellite in the non-terrestrial network is broken, other satellites can replace it and provide the communication service, thereby ensuring service reliability. In comparison to geostationary and medium earth orbit (GEO/MEO) satellites, which suffer from high propagation delay and require large investments, low earth orbit (LEO) satellites offer a lower latency, higher data rate, and near-complete visibility due to their relatively lower orbit altitudes (*i.e.* $500 - 2000$ km). Furthermore, the proximity of LEO satellites to Earth has prompted the exploration of remote sensing functionality in LEO satellite systems. In recent years, the necessity for sensing capabilities and enhanced spectrum utilization efficiency in LEO satellites has naturally driven the development of ISAC-LEO systems.

1.1.3 Massive MIMO (M-MIMO) systems

In early communication systems, only single antenna was used in both transmitter and receiver, which is known as the single-input single-output (SISO) system. The maximum rate of error-free data transmission under additive Gaussian white noise condition is given in Shannon's capacity equation [10]. However, the multipath propagation causes severe interference to the real-world communications systems due to the different time delays of the received signals. To address this challenge, the concept of antenna diversity or spatial diversity emerged, which is realized by deployment of multiple antennas in both transmitter and receiver. This technique, known as MIMO, has been demonstrated to increase spectral efficiency and finds widespread application in 4G wireless communication networks [11].

M-MIMO architecture was proposed by Marzetta [12], and relies on a large number of antennas at the base station (BS). This approach brings the advantage that channels among users become orthogonal, and thereby enabling effective mitigation of interference among users [13]. This characteristic simplifies interference mitigation, allowing for the efficacy of straightforward linear precoding methods such as zero forcing (ZF) and maximal ratio transmission (MRT). Additionally, M-MIMO exhibits advantages on anti-fading, suppressing noise, and compensating for hardware deficiencies. Due to the above reasons, M-MIMO can further improve the spectral efficiency and thus has attracted much interest in the past decade [14, 15].

1.1.4 Full-duplex systems

A promising solution to the *spectrum scarcity* problem has been the design of In-Band Full-Duplex (FD) wireless communications systems [16, 17]. In such systems, data can be transmitted bi-directionally between two devices simultaneously. Consequently, FD systems can potentially double the spectral efficiency of communication [16, 17]. Thus, there has been a surge of research interest towards this solution in past decades. However, when considering the FD solution as it is, there is a price to pay. That is, the coupling between the co-located transmitter and receiver causes severe self-interference (SI) (*e.g.* up to 100 dB [16, 18]).

The SI cancellation problem has been actively studied in FD wireless communication systems [17, 19, 20]. In general, SI cancellation techniques can be implemented in the propagation [21], analog [22–24], and digital domains [25, 26]. As shown in Fig. 3.1, the propagation-domain cancellation aims to minimize the coupling between the transmit and receive direct paths. This kind of cancellation is achieved by techniques based on path loss, cross-polarization, and antenna directionality [27]. Beyond this, the analog-domain cancellation aims to suppress the SI before the ADC, where a negative copy of the transmit waveform generated by the canceller circuit is subtracted from the received signal [20, 22, 23, 28]. Finally, as the last defense against SI, the digital-domain

cancellation block utilizes either linear or non-linear adaptive filters to generate the negative of the residual SI [25, 26], and add it to the digital signal after the ADC.

1.2 Challenges, Scope and Objectives

1.2.1 Challenges and Objectives of ISAC Systems

Mono-static ISAC setup, where the transmitter and receiver are colocated, is assumed in many ISAC studies [29–33] since there is no need to share reference signals. However, typically communication frames are much longer than the radar echo round-trip times (RTTs). Hence, the radar echo is masked by the severe SI from the transmitter. We term this phenomenon as the *echo-miss* problem. Thus, it is important to sufficiently suppress the residual SI to manageable levels. In addition, to preserve the fast-coming radar echo, full-duplex (FD) ISAC is required.

Another challenge is the *correlation between radar echo and SI*, which makes it challenging to effectively suppress the SI without distorting the radar echoes. This is the point different from the SI cancellation techniques, which have been actively studied in FD wireless communication systems [17, 19, 20]. Therefore, the aim of this research is to address the *echo-miss* problem in mono-static ISAC system without impairing the SoI (*e.g.* radar echo and uplink data).

1.2.2 Challenges and Objectives of ISAC LEO Satellite Systems

The first challenge of a ISAC LEO satellite system is the interference control, which incorporates inter-beam interference among the users and the control of the trade-off between communication and sensing capabilities. Another challenge is the power limitations typically faced by LEO satellites due to the lack of suitable power sources. This issue becomes notably more severe, especially whenever the number of transmit antennas increases for the improved spectral efficiency and spectrum reuse factor. Therefore, a high energy efficiency (EE) is required in ISAC LEO satellite systems.

Additionally, given the expected increase in connected devices and data rates in forthcoming satellite communication systems [34], it is crucial to adopt EE as a

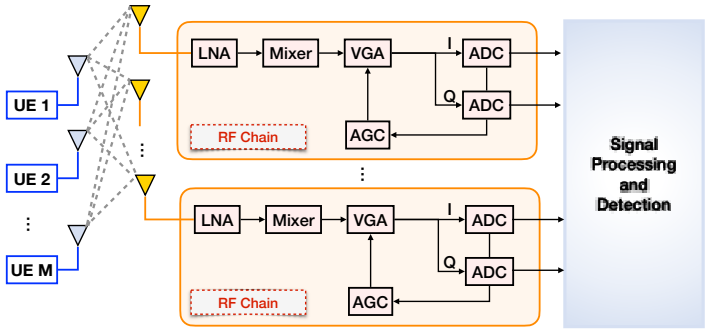


Figure 1.1 The receiver architecture of conventional M-MIMO (M–MIMO) system with uplink user equipment (UE). Each RF chain consists of the low noise amplifier (LNA), mixer, variable gain amplifier (VGA), automatic gain controller (AGC), and ADC.

significant performance indicator. From a system-wide viewpoint, prioritizing total EE appears favorable. Nonetheless, this approach tends to drop links with poor channel conditions, resulting in unfairness across all links. This can be severe in scenarios with numerous users. Consequently, it is more prudent to seek to maximize the minimum EE across all links, thereby ensuring fairness.

Here, the pressing issue lies in effectively managing inter-beam interference between users, while balancing the trade-off between communication and sensing functionalities within the ISAC-LEO system. Simultaneously, ensuring efficient power utilization and guaranteeing minimal EE among all users.

1.2.3 Challenges and Objectives of M-MIMO Systems

The technological promise of M–MIMO is limited by two fundamental challenges [35], namely, 1) *power consumption*, and 2) *receiver saturation*. High power consumption is due to that the M-MIMO has many antennas, where RF chains consume high power. In the RF chain, ADC dominates the total power. In addition, the power consumption of the ADC grows linearly with the sampling rate but exponentially with the quantization bit b.

Another impediment to the M–MIMO systems is due to receiver-ADC saturation, which is prevalent in wireless communication systems. This impediment

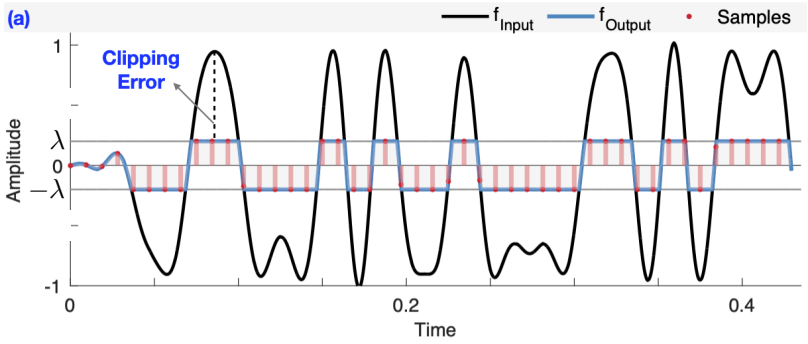


Figure 1.2 Illustration of output samples by conventional ADC. f_{Input}: continuous input signal before ADC. f_{Output}: continuous output signal shaped by limited ADC dynamic range. Samples: Sampled f_{Output} signals. In conventional ADC, saturation and clipping errors happen in f_{Output} signal and samples, because some values in f_{Input} exceed the dynamic threshold λ.

is caused by the fact that ADCs are limited by their dynamic range (DR) as illustrated in Fig. 1.2. When the received signal amplitude is larger than λ, the ADC output is distorted. This results in a clipped or saturated signal. Such undesirable, yet unavoidable, effects lead to high-frequency components in the digital signal, thus creating aliasing in the measurements. Consequently, the performance of the whole system deteriorates.

Hence, it is imperative to devise an approach that tackles both the challenges of high power consumption and receiver saturation simultaneously.

1.2.4 Challenges and Objectives of Full-duplex Systems

The severe SI in the full-duplex systems drowns the signal of interest (SoI) which contains uplink data. The composite of SI and SoI naturally necessitates a high dynamic range (HDR) ADC to prevent clipping. Consequently, subtraction of an accurately estimated SI can unmask the SoI. On the other hand, for a fixed bit budget, HDR ADCs entail higher quantization noise which can compromise the capability of existing algorithms.

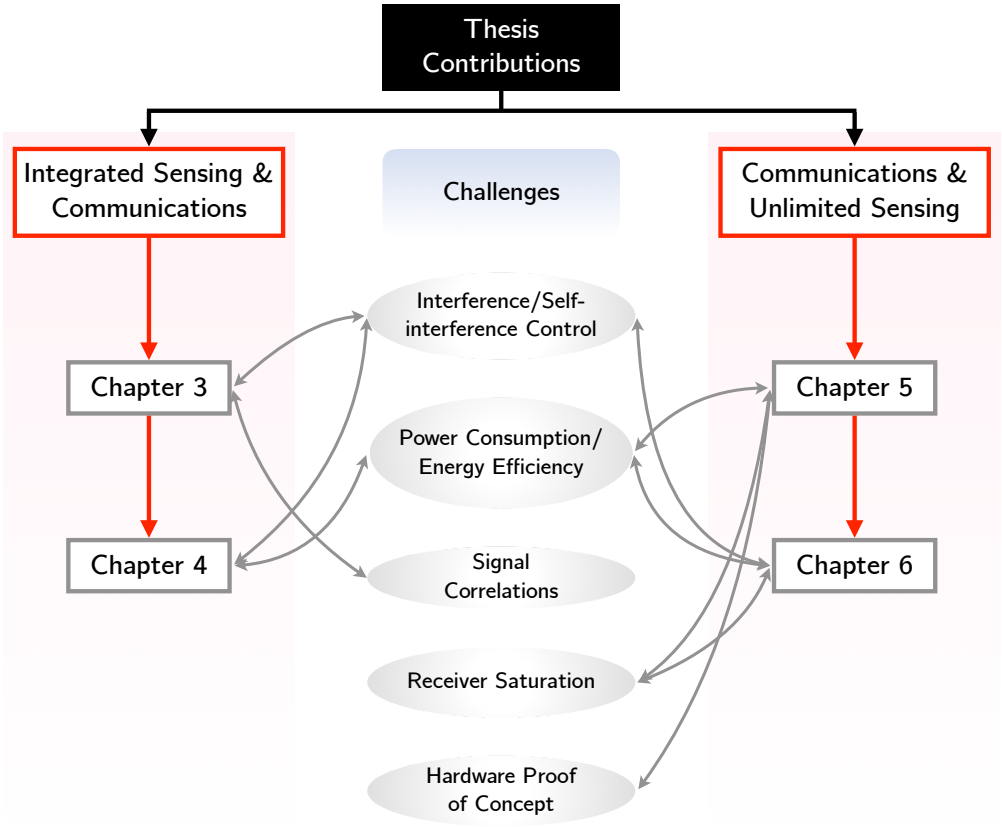
Therefore, the objective of this research is to simultaneously tackle the ADC saturation and quantization noise issues through a fully digital-domain SI can-

cellation method in FD wireless communications systems.

1.3 Outline and Contributions

As the first part of the thesis' title "Multi-functional Wireless Networks" suggests, this thesis addresses the inherent challenges and/or emerging challenges caused by integration of functionalities, which are:

- Interference/Self-interference Control;
- Power Consumption/Energy Efficiency;
- Signal Correlations;
- Receiver Saturation;
- Hardware Proof of Concept.



The interdependencies between the research in each chapter and focused challenges are shown in the roadmap. Specifically, this thesis builds upon two main bodies: integration of communication and sensing, and communication with unlimited sensing. The first body consists of Chapters 3 and 4, while the second body comprises Chapters 5 and 6.

In Chapter 3, we address two challenges: severe SI from the transmitter in FD ISAC and the difficulty of SI cancellation due to the correlation between the radar echo and SI. Although FD in wireless communications without sensing capabilities has been extensively studied for decades, it relies on the uncorrelated nature between the SI and SoI, such as the uplink data stream. Thus, the SoI will not be impaired. However, the correlation between SI and radar echo makes SI cancellation very challenging. To solve this, we utilize the angle-of-arrival difference of incoming signals to suppress the SI in ISAC. In Chapter 4, we address the interference control problem and EE issues. In previous research, inter-beam interference and the trade-off between communication and sensing capabilities have been difficult to cope with. Additionally, LEO satellites have power limitations, thereby emphasizing in the consideration of EE. However, the total EE metric lacks fairness among all links; it may drop users in bad channel conditions. To address these issues, we propose an approach based on rate-splitting multiple access (RSMA), minimal EE maximization, and low-resolution digital to analogue converters (DACs).

In Chapter 5, we focus on the high power consumption and receiver ADC saturation problems. The high power consumption is due to the considerable power consumed by a large number of RF chains in the antennas of M-MIMO systems. Receiver ADC saturation occurs when a strong interference signal is received (*e.g.* in FD systems) or when the received signals have significant differences (*e.g.* near-far users in the mobile cell). These two challenges are addressed through a new receiver-side design and the introduction of the unlimited sensing theorem. Subsequently, in Chapter 6, we consider a more severe case of receiver saturation, namely, in full-duplex systems. To cancel the strong SI, we further explore the unlimited sensing theorem, through which we propose a fully digital domain SI cancellation method and a modulo domain

channel estimation approach.

The outline and the detailed contributions for each chapter are listed as follows:

- **Chapter 2: Background**
  In Chapter 2, the related concepts and techniques for solving the problems in the following chapters are reviewed. Specifically, we briefly review the details of ISAC, unlimited sampling, and RSMA. Subsequently, to better understand the approaches for addressing the mentioned challenges, we explain the optimization techniques, *i.e.* block coordination descent (BCD), penalty-dual decomposition (PDD), fractional programming/Dinkelbach's algorithm, and successive convex approximation (SCA).

- **Chapter 3: Joint Transmit and Receive Beamforming Design in Full-Duplex Integrated Sensing and Communications**
  In this chapter, we first model a FD ISAC mono-static system to capture the *echo-miss* problem. To suppress the residual SI and preserve the two types of SoI (e.g., radar echo and uplink data), we formulate the joint TX-RX beamformer design problem for FD ISAC, where the objective function incorporates (a) uplink and downlink rates as the communications metric, (b) the transmit and receive radar beampattern power at a target as the sensing metric, and (c) the post-beamforming SI residual as a penalty term. Based on the equivalence between the rate maximization and the mean square error (MSE) minimization [36, 37], and inspired by the PDD method, an iterative algorithm with guaranteed convergence and acceptable complexity is developed using block coordination descent (BCD) methods. In contrast to the existing literature, our optimization framework concentrates on joint TX-RX beamformer design and directly cancel residual SI. The performance of the designed beamformers is validated via simulations, which show that up to 60dB residual SI can be effectively suppressed with better sum-rates for the uplink and downlink users and better radar parameter estimation performance (i.e., range-velocity and AoA detection) compared with the null-space projection (NSP) method [38–40].

- **Chapter 4: Max-Min Fair Energy-Efficient Beam Design for Quantized ISAC LEO Satellite Systems: A Rate-Splitting Approach**
  Besides the design of FD ISAC system in Chapter 3, we further consider the ISAC in the power limited LEO satellite scenario. Specifically, in this Chapter, we investigate the maximization of the minimum EE problem within the RSMA framework for an ISAC-LEO satellite system. Here, low-resolution DACs are employed in the satellite to enhance the EE. Specifically, we propose an RSMA-assisted ISAC satellite system design to facilitate multi-user communications and target sensing, simultaneously. Considering the non-linearity of the quantization effect in low-resolution DACs, we define a max-min fairness EE to ensure equitable EE performance across all users. We adopt the RSMA transmission scheme to enable effective interference control, and achieve an optimized communication sensing trade-off. Our proposed design addresses two challenges: interference control and fairness EE guarantee among all users in the context of ISAC-LEO systems. The latter is normally neglected as most previous studies adopt the total EE metric. We show that our proposed design has a higher max-min fairness EE compared to design by total EE metric. It is observed that adopting low-resolution DACs effectively improves the max-min fairness EE. Specifically, it is enhanced with satisfactory detection performances (*i.e.* Doppler, and range), when the quantization bit decreases. Compared to previous RSMA ISAC studies [41–46], the consistent observation is that radar sequence is not utilized when maximizing sum-rate metric. Consequently, a common stream is used for both communication and sensing functions. However, the new observation is that, when adopting the max-min fairness EE as the metric, the radar sequence is used for sensing. This is due to the utilization of a separated radar sequence facilitating a higher EE; the power of the radar sequence is used for sensing and the power consumed by common and private streams, which determines EE, is minimized mainly for communication use. In addition, we formulate the maximization of the minimum EE as a non-convex optimization problem to obtain the optimal precoders.

The formulated problem is subject to total transmit power, communication and sensing constraints. We ensure communication performance through quality-of-service (QoS) requirements while maintaining sensing estimation performance by considering the Cramér–Rao bound (CRB) on angle estimates. We transform the problem into a semi-definite programming (SDP) problem. Optimal precoders can be then obtained by solving the SDP problem through an iterative algorithm based on SCA and Dinkelbach's method. The numerical results are provided to illustrate the performance of the proposed RSMA-assisted ISAC-LEO satellite architecture. Key observations include: (a) The proposed design outperforms the strategies that maximize the total EE of the system and conventional space-division multiple access (SDMA) in terms of max-min fairness EE; (b) RSMA yields an improved trade-off between communication (*i.e.* max-min fairness EE) and sensing functionalities; (c) RSMA, combined with an additional precoded radar sequence with successive interference cancellation (SIC) operation, can lead to the best max-min fairness EE; (d) A lower DAC quantization bit results in a higher max-min fairness EE with satisfactory detection performances.

- **Chapter 5: λ–MIMO: Massive MIMO via Modulo Sampling**
  In this chapter, we formulate the unlimited sensing framework (USF) acquisition and recovery problem in the context of two typical communication systems, *i.e.* Narrowband single-carrier or the NBSC case, and wideband, multi-carrier orthogonal frequency-division multiplexing (OFDM) M–MIMO or MIMO–OFDM. For the two setups, we consider both *reconstruction* and *detection* performances. Here, our key contribution is the receiver design for USF enabled M–MIMO. Since $\mathscr{M}_\lambda$ maps a bandlimited function into a non-bandlimited one, existing receiver architectures do not apply to our case and this is the non-trivial aspect of our work. Our receiver design enables 1024 QAM for both (a) narrowband single-carrier and (b) wideband OFDM signal and detection is possible with as little as 2 bit resolution; this has not been reported in previous literature. Addtionally, we present the *Achievable sum-rate Analysis* for our new ap-

proach considering different bit resolutions. Surprisingly, with few bits *e.g.* 1- or 2-bits, the λ–MIMO architecture achieves a similar sum-rate performance when compared with ∞-bit resolution, conventional ADC. We attribute this finding to the fact that for a given bit budget, the $\mathscr{M}_\lambda$–ADC offers a higher quantization resolution, thus resulting in a lower quantization noise. We also provide *approximates* for the achievable sum-rate in the context of maximum ratio combiner and zero forcing combiner. We analyze the EE of the λ–MIMO approach and contrast it with low-resolution, conventional ADC as a baseline. Numerical results illustrate that the EE of the $\mathscr{M}_\lambda$–ADC is higher than the conventional ADC while offering competitive detection performance (*e.g.* 1024 QAM), as listed in C_1). Finally, we validate our approach via computer simulations as well as experiments based on the $\mathscr{M}_\lambda$–ADC hardware [4], thus corroborating our theoretical findings.

- **Chapter 6: Full-Duplex Beyond Self-Interference: The Unlimited Sensing Way**
  Similar to the system architecture for M-MIMO systems that is based on the USF in Chapter 5, we also do not employ AGC and VGA in the system design. However, different from the system in Chapter 5, our design considers the FD case, thus it encompasses new aspects of modulo-domain sparse channel estimation and digital SIC without any analog-domain SIC circuitry. Specifically, we propose a *full-stack* FD receiver architecture design, which can handle HDR signals and achieve higher quantization resolution with a given budget. The *full-stack* design in this paper enables understanding of the trade-off between bit-budget and dynamic range, and the best achievable digital SI cancellation capability. In the algorithmic front, we present recovery algorithms for the FD case. Additionally, we also propose a novel modulo-domain sparse channel estimation method, resulting in a more accurate digital SI cancellation compared with previous methods [47]. Finally, from the hardware perspective, since our approach is devoid of any analog-domain SIC circuitry, *e.g.* AGC and VGA, it lends itself to a low complexity implementation.

- **Chapter 7: Conclusions**
  We conclude this thesis with a summary of its major contributions and a discussion of possible directions for future research.

1.4 Publications

The work in the thesis has resulted in several papers, which have been published or submitted for publication.

1. Integrated sensing and communications:

   - **Z. Liu**, S. Aditya, H. Li and B. Clerckx, "Joint Transmit and Receive Beamforming Design in Full-Duplex Integrated Sensing and Communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 9, pp. 2907-2919, Sept. 2023, doi: 10.1109/JSAC.2023.3287542

   - **Z. Liu**, L. Yin, W. Shin, and B. Clerckx, "Rate-Splitting Multiple Access for Quantized ISAC LEO Satellite Systems: A Max-Min Fair Energy- Efficient Beam Design," Accept to *IEEE Transactions on Wireless Communications*, January, 2024, available online at arXiv:2402.09253

   - L. Yin, **Z. Liu**, M. R. Bhavani Shankar, M. Alaee-Kerahroodi and B. Clerckx, "Integrated Sensing and Communications Enabled Low Earth Orbit Satellite Systems," *IEEE Network*, 2024, doi: 10.1109/MNET.2024.3371172

2. Modulo sampling in wireless communications:

   - **Z. Liu**, A. Bhandari, and B. Clerckx, "λ–MIMO: Massive MIMO via modulo sampling," *IEEE Transactions on Communications*, vol. 71, no. 11, pp. 6301-6315, Nov. 2023, doi: 10.1109/TCOMM.2023.3305528

   - **Z. Liu**, A. Bhandari, and B. Clerckx, "Full-Duplex Beyond Self-Interference: The Unlimited Sensing Way," Submitted to *IEEE Communications Letters*, June, 2024, available online at arXiv:2404.08610

Chapter 2

Background

2.1 Unlimited Sampling

We replace conventional ADCs by $\mathscr{M}_\lambda$–ADCs, where it maps a continuous-time signal into folded, digital samples by injecting a modulo non-linearity in the analog domain. This prevents distortion due to clipping/saturation that is prevalent in conventional ADCs. This is shown in Fig. 2.1. Additionally, the hardware of $\mathscr{M}_\lambda$–ADCs is shown in Fig. 2.2 [48]. However, $\mathscr{M}_\lambda$–ADC lead to a different form of information loss as one still needs to recover the original HDR signal from folded samples.

This recovery is mathematically guaranteed provided that a Shannon-Nyquist like density criterion is satisfied. Specifically, reconstruction of the HDR input signal, $r\,(t)$, from its modulo samples entails the "inversion" of the modulo non-linearity $\mathscr{M}_\lambda(\cdot)$. This is an ill-posed problem due to the many-to-one nature of $\mathscr{M}_\lambda(\cdot)$. The key to unfolding modulo samples is the observation that the modulo and the finite-difference operators commute in a certain sense [1]. This allows us to extract the higher-order differences of r from its modulo samples. Then, reconstruction boils down to the problem of stably inverting the difference-operator. Next, we focus on these two aspects separately and for the case of non-quantized measurements (5.22). This is solely for the ease of expo-

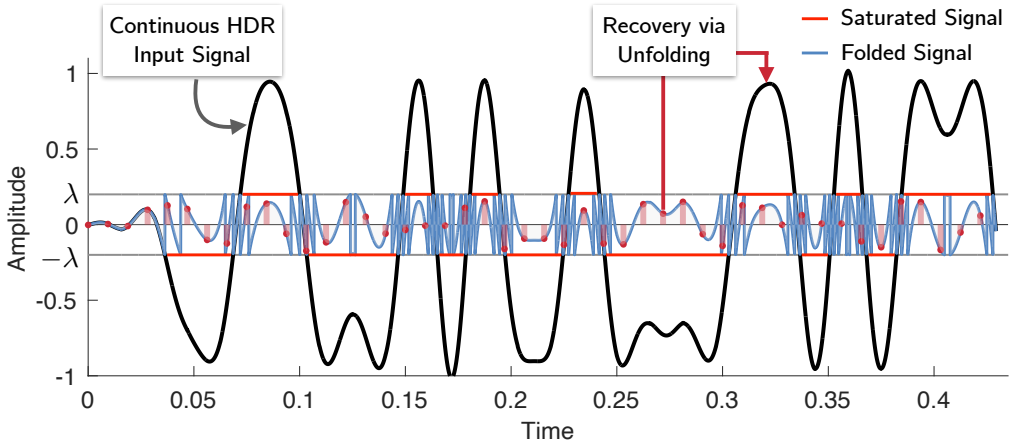


Figure 2.1 Illustration of modulo-folded signal via USF.

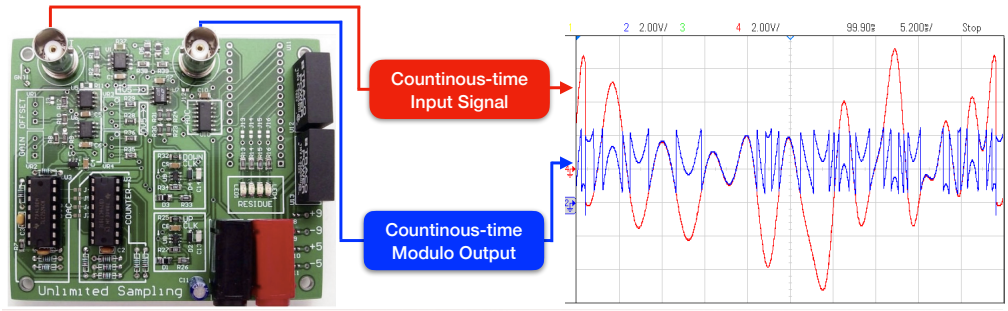


Figure 2.2 Hardware of $\mathscr{M}_\lambda$–ADC.

sition as the same recovery procedure applies to the case of quantized modulo samples in (5.23).

Accessing Higher-Order Differences. Let $(\Delta r)[k] \overset{\text{def}}{=} r[k+1] - r[k]$ denote the first-order difference operator; higher-order differences recursively defined by $\Delta^L r \overset{\text{def}}{=} \Delta^{L-1}(\Delta r), L > 1$. According to the modulo decomposition property (see Fig. 4 in [1]), we have, $r(t) = \mathscr{M}_\lambda(r(t)) + \varepsilon_r(t)$, where $\varepsilon_r(t) \in 2\lambda\mathbb{Z}$ is a piecewise-constant function. The same property also applies to the point-wise samples. On one hand, since $\varepsilon_r(kT) \in 2\lambda\mathbb{Z}$, we have that $\Delta^L \varepsilon_r(kT) \in 2\lambda\mathbb{Z}$ and hence, $\mathscr{M}_\lambda(\Delta^L \varepsilon_r(kT)) = 0$. On the other hand, for Ω-bandlimited functions, *i.e.* $r \in \mathcal{B}_\Omega$, we note that Δ^L results in amplitude shrinkage in the oversampled regime. This is made precise via the following Lemma.

Lemma 1 (Difference-Derivative Inequality [1, 2]). *For any $r(t) \in \mathcal{B}_\Omega \cap L_\infty(\mathbb{C})$, its samples $r[n] \overset{\text{def}}{=} r(nT)$ satisfy, $\|\Delta^L r\|_{\ell_\infty(\mathbb{R})} \leqslant (T\Omega e)^L \|r\|_{L_\infty(\mathbb{R})}$ where e is Euler's number.*

Clearly, choosing $T < 1/(\Omega e)$ amounts to controlling the factor $(T\Omega e)^L$. More precisely, choosing $L^\star \geqslant \left\lceil \frac{\log \lambda - \beta_r}{\log(T\Omega e)} \right\rceil$ with $\beta_r \geqslant 2\lambda\mathbb{Z}$ and $\beta_r \geqslant \|r\|_{L_\infty(\mathbb{R})}$ also ensures that $\|\Delta^{L^\star} r\|_{\ell_\infty(\mathbb{R})} \leqslant \lambda$. According to Proposition 2 in [1], we have, $\mathscr{M}_\lambda(\Delta^L(r)) = \mathscr{M}_\lambda(\Delta^L(\mathscr{M}_\lambda(r)))$ and hence, whenever $L \geqslant L^\star$, oversampling or choosing $T\Omega e < 1$ implies that $\Delta^L r = \mathscr{M}_\lambda(\Delta^L r_\lambda)$. In summary, oversampling $r(t) \in \mathcal{B}_\Omega$ allows us to access $\Delta^L r$ from its modulo samples $r_\lambda[k]$.

Stable Inversion of Δ^L. Instead of inverting $(\Delta^L r)[k] \in \mathbb{R}$, the recovery algo-

rithm in [1,2] capitalizes the piece-wise constant structure of the residue $\varepsilon_r(t)$, that is, $\varepsilon_r(kT) \in 2\lambda\mathbb{Z}$ and hence, $\Delta^L\varepsilon_r(kT) \in 2\lambda\mathbb{Z}$. This narrows down the range of $\varepsilon_r(kT)$ and hence has a stabilizing effect during inversion. We initialize the sequence $s^{\{0\}}[k] = \left(\mathscr{M}_\lambda(\Delta^L r_\lambda) - \Delta^L r_\lambda\right)[k] \equiv \Delta^L\varepsilon_r(kT) \in 2\lambda\mathbb{Z}$. Next, $\Delta^{L-1}\varepsilon_r(kT)$ is obtained from $\Delta^L\varepsilon_r(kT)$ by invoking the anti-difference operator $\mathsf{S} = \Delta^{-1}$ and rounded-off to $2\lambda\mathbb{Z}$. Note that the unknown constant-of-integration in the kernel of Δ must be estimated for unique reconstruction. It has been shown in [1,2] that for $r \in \mathcal{B}_\Omega$, this unknown constant can be estimated via,

$$\kappa^{\{l\}} = \left\lfloor \frac{\mathsf{S}\left(s^{\{l+1\}}\right)[k] - \mathsf{S}\left(s^{\{l+1\}}\right)[J_\lambda + 1]}{12\beta_r} + \frac{1}{2} \right\rfloor,$$

where $J_\lambda = 6\beta_r/\lambda, l = 0, \ldots, L-2$. This completes the estimation of $s^{\{1\}}[k]$ and more generally,

$$s^{\{l+1\}}[k] = 2\lambda \left\lceil \frac{1}{2} \left\lfloor \frac{\mathsf{S}\left(s^{\{l+1\}}\right)[k]}{\lambda} \right\rfloor \right\rceil + 2\lambda\kappa^{\{l\}}, \quad l = 0, \ldots, L-2.$$

Finally, we obtain the recovered samples $\widetilde{r}[k]$ up to an unknown constant,

$$\widetilde{r}[k] = \mathsf{S}\left(s^{\{L-1\}}\right)[k] + r_\lambda[k] + 2\lambda\mathbb{Z} = \widetilde{\varepsilon}_r + r_\lambda[k]. \tag{2.1}$$

The reconstruction approach is summarized in Algorithm 1 which also works with quantized measurements [1]. A flow chart as given in Fig. 2.3 is also provided for illustration. Compared with conventional ADC, adopting $\mathscr{M}_\lambda$–ADC has this additional reconstruction operation. It has a complexity of the order of $\mathcal{O}(LN_\mathsf{s})$, where N_s is the length of the samples. The recovery guarantee is summarized below.

A figure for explaining the Algorithm is provided below. Given the sequence of modulo samples, our basic strategy will be to apply a higher order finite difference operators. We will be exploiting that such operators commute with the modulo operation. So after applying the amplitude folding to the resulting sequence, one obtains the same output as if one had started with $\gamma[k] = g(kT)$ instead of $y[k]$. That in turn will allow for recovery if the higher order finite

Algorithm 1: Recovery from Noisy Modulo Communication Samples

Input: $q_\lambda[k]$ and $2\lambda\mathbb{Z} \ni \beta_r \geqslant \|r\|_\infty$.

Output: $\widetilde{r}[k] \approx r(kT)$.

1) Compute $L = \left\lceil \frac{\log\lambda - \log\beta_r}{\log(\mathrm{T}\Omega e)} \right\rceil$.

2) Compute $\left(\Delta^L q_\lambda\right)[k]$.

3) Compute $\left(\Delta^L \varepsilon_r\right)[k] = \left(\mathscr{M}_\lambda\left(\Delta^L q_\lambda\right) - \Delta^L q_\lambda\right)[k]$.

Set $s^{\{0\}}[k] = \left(\Delta^L \varepsilon_r\right)[k]$

for $l = 0 : L-2$ **do**

i) $s^{\{l+1\}}[k] = \left(\mathsf{S} s^{\{l\}}\right)[k]$

ii) $s^{\{l+1\}} = 2\lambda \left\lceil \frac{\lfloor s^{\{l+1\}}/\lambda \rfloor}{2} \right\rceil$ (rounding to $2\lambda\mathbb{Z}$)

iii) Compute $\kappa^{\{l\}} = \left\lfloor \frac{\left(\mathsf{S}^2\Delta^l\varepsilon_r\right)[1] - \left(\mathsf{S}^2\Delta^l\varepsilon_r\right)[J_\lambda+1]}{12\beta_r} + \frac{1}{2} \right\rfloor$

iv) $s^{\{l+1\}}[k] = s^{\{l+1\}}[k] + 2\lambda\kappa^{\{l\}}$

end

5) $\widetilde{r}[k] = \mathsf{S}\left(s_{\{L-1\}}\right)[k] + q_\lambda[k] + 2v\lambda, v \in \mathbb{Z}$

differences of the $\gamma[k]$'s are so small that the amplitude folding has no effect. Sub-figure (1) shows modulo decomposition, $\gamma[k] = y[k] + \varepsilon_\gamma[k]$. Sub-figure (2) shows pointwise difference of (1)-a, the bandlimited function, pointwise difference of (1)-b, that is, folded version of bandlimited function in (1)-a and pointwise difference of the residual in (1)-c. Sub-figure (3) shows the modulo of plots in (2)-a, (2)-b and (2)-c, respectively. When the sampling density meets certain criterion, the differences are always bounded by the threshold λ. More precisely, $T < 1/\Omega e \Rightarrow \left\|\Delta^N \gamma\right\|_\infty < \lambda$. Hence modulo operation has no effect

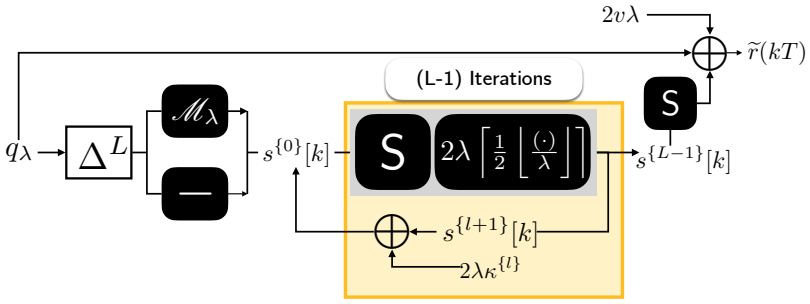


Figure 2.3 Flow Chart of Reconstruction Algorithm.

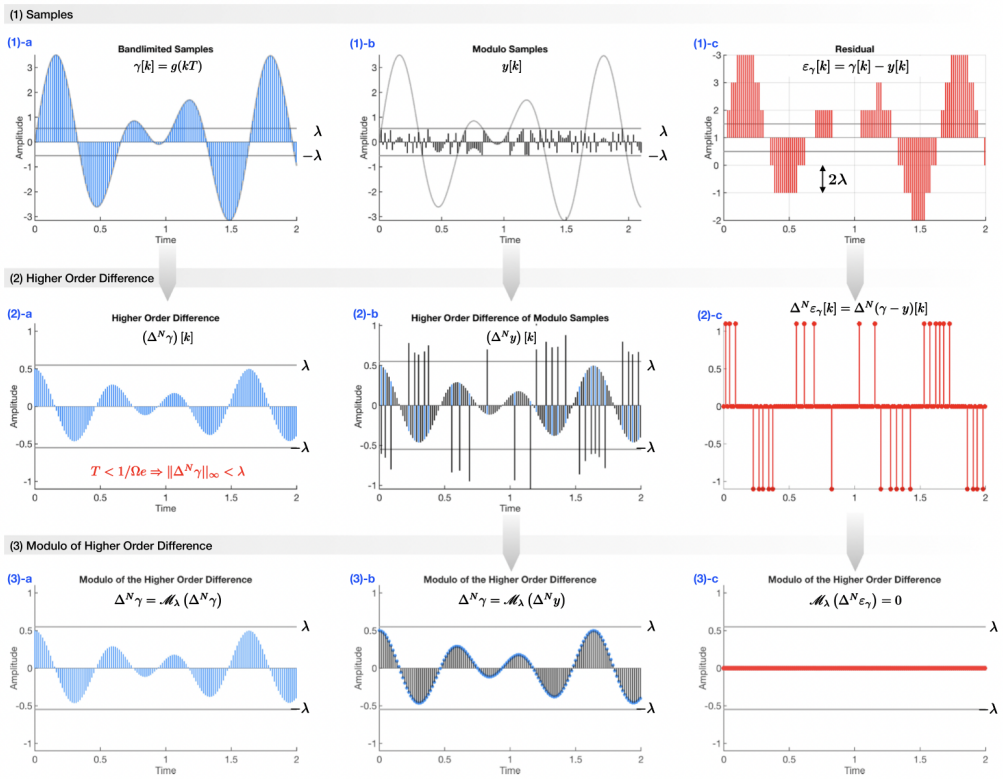


Figure 2.4 Overview of the main idea behind recovering a bandlimited function from modulo samples [1].

implying that plots (2)-a and (3)-a exactly coincide. As shown in Sub-figure (3)-b, the plots in (3)-a and (3)-b exactly coincide. Hence, starting with the modulo measurements $y[k]$ in (1)-b, it is possible to recover the higher differences of 1)-a, that is, $\Delta^N \gamma$ which is shown in (2)-a. To recover $\gamma[k]$ from $\Delta^N \gamma[k]$, we first estimate $\Delta^N \varepsilon_\gamma = \mathscr{M}_\lambda(\Delta^N y) - \Delta^N y$. Since ε_γ only takes amplitudes on the grid of $2\lambda\mathbb{Z}$, we exploit this restriction on amplitudes to recover ε_γ in (1)-c. There on, adding ε_γ in (1)-c and the modulo samples $y[k]$ in (1)-b results in the bandlimited samples in (1)-a.

Theorem 1 (Unlimited Sampling Theorem with Bounded Noise [1]). *Let $r(t)$ be a finite energy, bandlimited signal with maximum frequency Ω (rads/s) and assume that $\beta_r \in 2\lambda\mathbb{Z}$ is known with $\|r\|_\infty \leqslant \beta_r$. For the dynamic range we work with the*

normalization $\rho_\lambda = \beta_r/\lambda$. Let the noise ϵ_λ be the additive form on the modulo samples with a noise bound given in terms of the dynamic range as

$$\|\epsilon_\lambda\|_\infty \leqslant \frac{\lambda}{4}(2\rho_\lambda)^{-\frac{1}{\alpha}}, \quad \alpha \in \mathbb{N}. \tag{2.2}$$

Then, a sufficient condition for Algorithm 1 with

$$L = -\left\lceil \frac{\log\lambda - \log\beta_r - 1}{\log(\mathrm{T}\Omega e)} \right\rceil, \tag{2.3}$$

to approximately recover the bandlimited samples $r[k]$ in the sense that it returns $\widetilde{r}[k] = r[k] + \epsilon_\lambda + 2v\lambda$, $v \in \mathbb{Z}$ is that

$$T \leqslant \frac{1}{2^\alpha \Omega e}. \tag{2.4}$$

When α increases, the sampling rate increases, as shown in (2.4). This will affect the noise bound as given in (2.2), thereby enlarging the range of noise signals that Algorithm 1 can handle. In our simulations, α is set to 1.

Remark 1. *Compared to [1], the novelty of our work on λ–MIMO in chapter 3 is as follows:*

- *We focus on the communication design with the USF and conduct proof-of-concept hardware experiments.*
- *We derive the analysis of quantization noise and achievable uplink sum-rate for the $\mathscr{M}_\lambda$–ADC.*
- *We analyze the energy efficiency of λ–MIMO and conventional massive MIMO systems.*

2.2 Rate-splitting Multiple Access (RSMA)

RSMA, recognized as a generalized and powerful non-orthogonal transmission framework and robust interference management strategy for multi-antenna multi-user networks, has been demonstrated to be effective in handling various interference challenges in satellite communication [9, 49–52]. In RSMA, the transmitter splits each message into common and private messages, which are then linearly precoded at the satellite gateway. At the receiver, the common stream is first decoded by treating all private streams as noise. Subsequently, each user eliminates the decoded common stream and decodes their respective private stream, treating the other private streams as noise. Due to the ability to partially decode the interference and partially treat it as noise, RSMA outperforms the existing multiple access schemes, *e.g.* SDMA, orthogonal multiple access (OMA), and non-orthogonal multiple access (NOMA), in terms of spectral efficiency and user fairness [53]. Furthermore, RSMA has been shown to facilitate a better trade-off between the communication and sensing functionalities, when compared to conventional beamforming techniques [41–44, 54].

We explain the structure of RSMA as shown in Fig. 2.5. We firstly assume that this system serves K single-antenna users, indexed by $\mathcal{K} = \{1, \cdots, K\}$. The message W_k for the k^{th} user is categorized into a common message $W_{c,k}$ and a private message $W_{p,k}$. Subsequently, all common messages, $\{W_{c,1}, \cdots, W_{c,K}\}$,

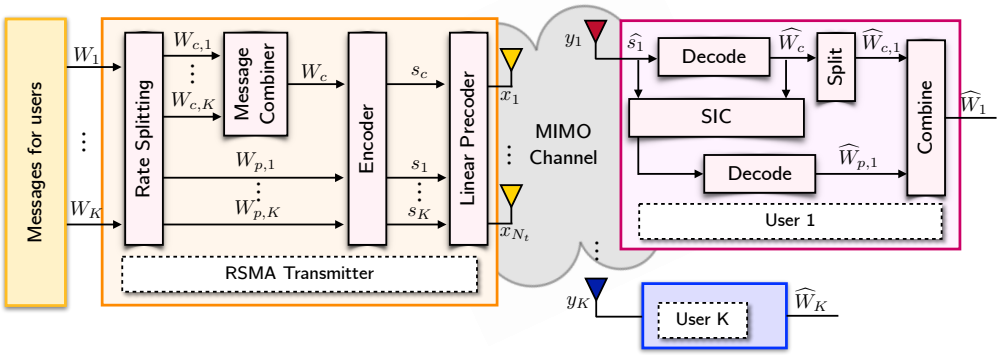


Figure 2.5 RSMA structure.

are jointly encoded into a common stream s_c. All the private parts, which are denoted by $\{W_{p,1}, \cdots, W_{p,K}\}$, are encoded into private streams, $\{s_1, \cdots, s_K\}$. The transmit streams comprise common stream, private streams and radar sequence are expressed as $\mathbf{s}[\ell] = [s_c[\ell], s_1[\ell], \cdots, s_K[\ell], s_r[\ell]]^\top \in \mathbb{C}^{(K+2)\times 1}$, where $\ell \in \mathcal{L} = \{1, \cdots, L\}$ is the discrete time index. Subsequently, the data streams are linearly precoded by the precoder $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots, \mathbf{p}_K] \in \mathbb{C}^{N_t \times (K+1)}$, where $\mathbf{p}_c \in \mathbb{C}^{N_t}$ is the precoder of the common stream, and $\{\mathbf{p}_k\}_{k\in\mathcal{K}}$ is the precoder of the private stream. Thus, the resulting signal is $\mathbf{x} = \mathbf{P}\mathbf{s} = \mathbf{p}_c s_c + \sum_{k\in\mathcal{K}} \mathbf{p}_k s_k + \mathbf{p}_r s_r$. The total transmit power is constrained by $\mathrm{tr}(\mathbf{P}\mathbf{P}^H) \leqslant P_t$. At user sides, all users firstly decode the common stream s_c by treating all private streams, $\{s_1, \cdots, s_K\}$, as interference. At this step, each user decodes a portion of the message transmitted by the other interfering users, which is encoded in the common stream s_c. Thus, we can say that the interference is partially decoded at each user. The signal-to-interference-plus-noise (SINR) for the common stream at the k^{th} user is given by:

$$\gamma_{c,k} = \frac{\left|\mathbf{h}_k^H \mathbf{p}_c\right|^2}{\sum_{i\in\mathcal{K}} \left|\mathbf{h}_k^H \mathbf{p}_i\right|^2 + \sigma_k^2}. \tag{2.5}$$

The successfully decoded common message s_c is subtracted from the received signal y_k by SIC. Subsequently, the k^{th} user decodes its private stream s_K by treating the private streams from other $K-1$ users as noise. The SINR of private streams at the k^{th} user is written as

$$\gamma_k = \frac{\left|\mathbf{h}_k^H \mathbf{p}_k\right|^2}{\sum_{i\in\mathcal{K}, i\neq k} \left|\mathbf{h}_k^H \mathbf{p}_i\right|^2 + \sigma_k^2}. \tag{2.6}$$

The normalized rates of the common and private streams with regard to SINR are written as

$$R_{c,k} = B \log_2 (1 + \gamma_{c,k}) = \log_2 (1 + \gamma_{c,k}), \tag{2.7a}$$

$$R_k = B \log_2 (1 + \gamma_k) = \log_2 (1 + \gamma_k), \tag{2.7b}$$

where B is the bandwidth, and we normalize it as 1. In the RSMA framework, the common stream needs to be decodable by all receivers. To guarantee this,

the common rate, denoted by R_c, should be set to the minimum of the $R_{c,k}, \forall k \in \mathcal{K}$ [53], *i.e.*

$$R_c = \min_{k \in \mathcal{K}} \{R_{c,k}\} = \sum_{k \in \mathcal{K}} C_k, \tag{2.8}$$

where C_k is the rate of the common part of the k^{th} user. Thus, the total achievable rate of the k^{th} user is $R_{k,\text{total}} = C_k + R_k$.

2.3 Optimization Techniques

2.3.1 Block Coordination Descent (BCD)

BCD is an optimization framework to iteratively update one block of variables to find the minimum of the function [55, 56]. In the application, we can divide the original objective problems into several sub-problems via optimization of the blocks of variables. In the descent process, it searches the direction of one dimension. When the minimum in this dimension is achieved, it searches other dimensions iteratively, and finally the optimal is obtained. The update order of blocks can be certain or random. Different from gradient descent method, BCD requires no computation of gradient of the objective function. Due to this reason, BCD only suits for smooth functions, otherwise the BCD may stuck in non-stationary points.

Consider the following problem

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \tag{2.9a}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \tag{2.9b}$$

- $\mathcal{X} : \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \subseteq \mathbb{R}$ is closed, nonempty and convex sets,
- f: $\mathcal{X} \to \mathbb{R}$ is continuous function.

The general BCD algorithm can be then summarized as below. The BCD method

Algorithm 2: BCD framework

Find a feasible point $x^0 \in \mathcal{X}$ and set $r = 0$
repeat
  $r = r + 1, i = (r - 1 \bmod m) + 1$
  Let $\boldsymbol{x}_i^\star \in \arg\min_{\boldsymbol{x} \in \mathcal{X}_i} f\left(\boldsymbol{x}_1^{r-1}, \ldots, \boldsymbol{x}_{i-1}^{r-1}, \boldsymbol{x}, \boldsymbol{x}_{i+1}^{r-1}, \ldots, \boldsymbol{x}_m^{r-1}\right)$
  Set $\boldsymbol{x}_i^r = \boldsymbol{x}_i^\star$ and $\boldsymbol{x}_k^r = \boldsymbol{x}_k^{r-1}, \forall k \neq i$
until *convergence criterion is met*;

has the following advantages: 1) the sub-problems are easier to solve and may have closed-form solutions. For example, in our research, we transform the

sub-problem to a convex function with convex constraints, thus closed-form solutions are given; 2) BCD updates are non-decreasing; 3) BCD framework is adaptive to parallel or distributed implementations. Meanwhile, BCD method also has some drawbacks such that the convergence rate is slower than that of the gradient descent method.

2.3.2 Penalty-dual Decomposition (PDD)

PDD is a widely-used technique to solve optimization problems involving non-smooth, non-convex functions [57, 58]. When the constraints involve the coupling among variables, the BCD method in 2.3.1 is no longer useful. In contrast, PDD method can be used. Not only the coupling constraints, but also the non-convex and non-differentiable constraints can be tackled by the PDD method. The core to solve such problems is to introduce the penalty term and adopt the Lagrangian transformation to appropriately scale the coupling constraints. Specifically, PDD is a double-loop optimization framework. In the inner loop, the Augmented Lagrangian problem is approximately solved by BCD. In the outer loop, the dual variable and/or penalty parameters are updated.

Consider a general multi-variate non-convex non-smooth problem with coupling constraints and non-convex/non-differentiable constraints

$$\min_{\boldsymbol{x},\,\boldsymbol{y}} \quad f(\boldsymbol{x},\boldsymbol{y}) + \sum_{j=1}^{N} \phi_j(s_j(\boldsymbol{y}_j)), \tag{2.10a}$$

$$\text{s.t.} \quad \mathbf{h}(\boldsymbol{x},\boldsymbol{y}) = 0, \tag{2.10b}$$

$$\mathbf{g}(\boldsymbol{x}_i) \leqslant 0, \forall i, \tag{2.10c}$$

- $\boldsymbol{x} \in \mathcal{X}$;
- $f(\boldsymbol{x},\boldsymbol{y})$ is a scalar continuously differentiable function;
- $\phi_j(s_j(\boldsymbol{y}_j))$ is a composite function;
- $\mathbf{h}(\boldsymbol{x},\boldsymbol{y})$ is the coupling constraint;
- $\mathbf{g}(\boldsymbol{x}_i) \leqslant 0, \forall i$ is the possible non-convex constraint.

Subsequently, we can transform the original objective problem (2.10) to an augmented Lagrange function as given by

$$\min_{\boldsymbol{x},\boldsymbol{y}} \quad f(\boldsymbol{x},\boldsymbol{y}) + \sum_{j=1}^{N} \phi_j(s_j(\boldsymbol{y}_j)) + \lambda^\top \mathbf{h}(\boldsymbol{x},\boldsymbol{y}) + \frac{1}{\rho}\|\mathbf{h}(\boldsymbol{x},\boldsymbol{y})\|^2 \tag{2.11a}$$

$$\text{s.t.} \quad \mathbf{g}(\boldsymbol{x}_i) \leqslant 0, \forall i, \tag{2.11b}$$

where λ and ρ are the introduced dual variable and penalty parameter corresponding to the coupling constraint $\mathbf{h}(\boldsymbol{x},\boldsymbol{y})$. Note that since PDD method is a double-loop framework, the convergence speed of it is highly depends on the methods utilized to solve each block in the inner loop and approaches used to update dual variable/penalty parameter in the outer loop. Moreover, the PDD method is sensitive to the initialization. Due to the above reasons, fine tuning is required for specific problems. A general PDD framework is summarized below.

Algorithm 3: PDD framework

initialize $\boldsymbol{z}^0 = (\boldsymbol{x}^0, \boldsymbol{y}^0), \rho_0 > 0, \boldsymbol{\lambda}_0$, set $k = 1$, and pick two sequences $\{\eta_k > 0\}, \{\epsilon_k > 0\}$
repeat
  $\left(\boldsymbol{z}^k, \boldsymbol{\nu}^k\right) = \text{optimize}\left(P_{\rho_k,\boldsymbol{\lambda}_k}, \boldsymbol{z}^{k-1}, \epsilon_k\right)$
  if $\left\|\boldsymbol{h}\left(\boldsymbol{z}^k\right)\right\|_\infty \leqslant \eta_k$ **then**
    $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \frac{1}{\rho_k}\boldsymbol{h}\left(\boldsymbol{z}^k\right)$
    $\rho_{k+1} = \rho_k$
  else
    $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k$
    update ρ_{k+1} by decreasing ρ_k
  end
  $k = k + 1$
until *convergence criterion is met*;

2.3.3 Weighted Minimum Mean Square Error (WMMSE)

Weighted minimum mean square error (WMMSE) is to circumvent the non-convexity of the communication rate term based on the equivalence between the rate and MSE matrix [36, 37]. The basic idea in this method is to transform the $\log(\cdot)$ function in the rate term to a more tractable forms based on MSE function.

Consider a sum-rate maximization problem in a downlink multi-user MIMO system, where a BS serves K multi-antenna users. We assume that the precoding matrix $\mathbf{P}$ is designed in the BS, thus the received signal at the k^{th} user is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{P}_k \mathbf{s}_k + \sum_{j \neq k} \mathbf{H}_k \mathbf{P}_j \mathbf{s}_j + \mathbf{n}_k, \forall k, \tag{2.12}$$

where $\mathbf{H}_k$ is the channel matrix between the BS and the k^{th} user, $\mathbf{n}_k$ is the thermal noise at the k^{th} user, and $\mathbf{s}_k$ is the stream for the k^{th} user.

Therefore, the sum-rate maximization problem by designing transmit precoder is written as

$$\max_{\{\mathbf{P}_k\}_{k=1}^K} \quad \sum_{k=1}^{K} \log_2 \left| \mathbf{I} + \mathbf{P}_k^H \mathbf{H}_k^H \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{P}_k \right| \tag{2.13a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \|\mathbf{P_k}\|^2 \leqslant P_t, \tag{2.13b}$$

where P_t is the transmit power budget, and $\mathbf{R}_k$ is the equivalent noise covariance matrix, given by

$$\mathbf{R}_k = \mathbf{I} + \sum_{j \neq k} \mathbf{H}_k \mathbf{P}_j \mathbf{P}_j^H \mathbf{H}_k^H. \tag{2.14}$$

If all users adopt MMSE combiner, the receive matrix at k^{th} user is expressed as

$$\mathbf{W}_k^{\mathsf{MMSE}} = \mathbf{P}_k^H \mathbf{H}_k^H \left(\mathbf{I} + \sum_{j=1}^{K} \mathbf{H}_k \mathbf{P}_j \mathbf{P}_j^H \mathbf{H}_k^H \right)^{-1}. \tag{2.15}$$

Note that if MMSE combiner is not adopted at the users' side, the equivalence

between the rate and MSE matrix no longer exists. Subsequently, the MSE matrix is obtained as

$$\mathbf{E}_k = \mathbb{E}\left\{\left(\mathbf{W}_k^{\mathsf{MMSE}}\mathbf{y}_k - \mathbf{s}_k\right)\left(\mathbf{W}_k^{\mathsf{MMSE}}\mathbf{y}_k - \mathbf{s}_k\right)^H\right\} = \left(\mathbf{I} + \mathbf{P}_k^H\mathbf{H}_k^H\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{P}_k\right)^{-1} \tag{2.16}$$

Comparing (2.13a) and (2.16), we can clearly observe the equivalence relationship between communication rate term and MSE. Based on [36, 37], the sum-rate maximization problem can be transformed to a weighted MSE minimization problem by introducing the weight matrix $\mathbf{V}_k$, given by

$$\min_{\{\mathbf{P}_k, \mathbf{W}_k, \mathbf{V}_k\}_{k=1}^K} \quad \sum_{k=1}^K \left(\mathrm{Tr}(\mathbf{V}_k\mathbf{E}_k) - \log_2|\mathbf{V}_k|\right) \tag{2.17a}$$

$$\text{s.t.} \quad \sum_{k=1}^K \|\mathbf{P}_k\|^2 \leqslant P_t. \tag{2.17b}$$

Finally, this reformulated multivariate optimization problem can be solved by the BCD method.

2.3.4 Fractional Programming/Dinkelbach's Transformation

Fractional programming (FP) is a class of optimization problems where the objective function is a ratio of two functions. The fractional structure is prevalent in wireless communications such as:

- Energy efficiency maximization problem

$$\max_{\mathbf{p}} \quad \frac{\log(1 + \mathsf{SINR}(\mathbf{p}))}{P} \tag{2.18a}$$

$$\text{s.t.} \quad \|\mathbf{p}\|^2 \leqslant P_t, \tag{2.18b}$$

  where P is the total power consumption, $\mathbf{p}$ is the precoding vector, and P_t is transmit power budget.

- SINR related problem

$$\frac{|h_{k,k}|^2 p_k}{\sum_{j\neq k} |h_{k,j}|^2 p_j + \sigma^2}. \tag{2.19}$$

The above problems can be expressed by

$$\max_{\boldsymbol{x}} \quad \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} \tag{2.20a}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \tag{2.20b}$$

- $f(\boldsymbol{x})$ is a non-negative scalar function;
- $g(\boldsymbol{x})$ is a positive scalar function;
- $\mathcal{X}$ is a non-empty constraint set.

To solve problems with fractional structure, the core solution idea is to separate the numerator and denominator of the fractional terms, and change the form of division into the easier-to-solve $\pm$ form to facilitate the solution. In [59], the Dinkelbach's method is proposed to reformulate the original optimization problem (2.20) to

$$\max_{\boldsymbol{x}} \quad f(\boldsymbol{x}) - \lambda g(\boldsymbol{x}) \tag{2.21a}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \tag{2.21b}$$

- λ is an auxiliary variable, which is updated at each iteration as $\lambda^t = f(\boldsymbol{x}^t)/g(\boldsymbol{x}^t)$;
- When λ is fixed and the FP is concave-convex, solving $\boldsymbol{x}$ is a convex problem, which is easier to solve.

When multi-ratio fractional programming is considered, the influence on the value of a single fraction changes. Therefore, the optimal solution will differ from that of the original problem, making Dinkelbach's algorithm infeasible.

The multi-ratio FP is required and given by

$$\max_{\boldsymbol{x}} \quad \sum_{i=1}^{m} h_i\left(\frac{f_i(\boldsymbol{x})}{g_i(\boldsymbol{x})}\right), \tag{2.22a}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}. \tag{2.22b}$$

- $h_i(\cdot), i = 1, \cdots, m$ are non-decreasing functions.

According to [60, 61], to solve multi-ratio FP problems, the reformulated function should satisfy the following four properties: 1) (*Decoupling*) the transformed new function $F(\boldsymbol{x}, \lambda)$ should decouple the numerator and denominator in the original objective function into difference or summation form; 2) (Equivalent Solution) The optimal solution $\boldsymbol{x}^\star$ of $f(\boldsymbol{x})/g(\boldsymbol{x})$ and the designed function $F(\boldsymbol{x}, \lambda)$ are equal when λ is equal to a certain value; 3) (Equivalent Objective) The objective values of $f(\boldsymbol{x})/g(\boldsymbol{x})$ and the designed function $F(\boldsymbol{x}, \lambda)$ are equal when λ is equal to a certain value; 4) (Concavity) Function $F(\boldsymbol{x}, \lambda)$ is concave over λ for fixed $\boldsymbol{x}$, *i.e.* $\frac{\partial^2 F}{\partial \lambda^2} \leqslant 0$. Thus, a quadratic transform for FP problem which meet all above conditions is proposed in [60], given by

$$\max_{\boldsymbol{x}} \quad \sum_{i=1}^{m} h_i\left(2\lambda_i\sqrt{f_i(\boldsymbol{x})} - \lambda_i^2 g_i(\boldsymbol{x})\right), \tag{2.23a}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \lambda_i \in \mathbb{R}. \tag{2.23b}$$

The original optimization problem is hence transformed to a multi-block optimization. It can be subsequently solved by the BCD method.

2.3.5 Successive Convex Approximation (SCA)

SCA is an iterative optimization framework to solve non-convex optimization problems [62–64]. The basic idea of SCA is to solve a sequence of simpler problems and thus solve the originally difficult problem. Specifically, SCA replaces the non-convex components via a convex surrogate function in each iteration. The surrogate function for SCA should satisfy the following conditions: 1) Continuity; 2) Convexity; 3) Function value consistency; 4) Gradient consistency.

Algorithm 4: SCA framework

initialize $\boldsymbol{x}^0 \in \mathcal{X}$, set $t = 0$, step-size $\{\gamma^t\} \in (0, 1]$
repeat
  Solve $\hat{\boldsymbol{x}}\left(\boldsymbol{x}^t\right) = \arg\min_{\boldsymbol{x}\in\mathcal{X}} g\left(\boldsymbol{x} \mid \boldsymbol{x}^t\right)$
  Smooth the surrogate problem $\boldsymbol{x}^{t+1} = \boldsymbol{x}^t + \gamma^t\left(\hat{\boldsymbol{x}}\left(\boldsymbol{x}^t\right) - \boldsymbol{x}^t\right)$
  $t \leftarrow t + 1$
until *Some convergence criterion is met*;
return $\boldsymbol{x}^t$

Consider the following optimization problem

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \tag{2.24a}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}, \tag{2.24b}$$

- $\mathcal{X}$: $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \subseteq \mathbb{R}$ is closed, nonempty and convex sets,
- f: $\mathcal{X} \to \mathbb{R}$ is continuous function.

Subsequently, we can construct the surrogate function $g(\boldsymbol{x})$ by the widely adopted gradient descent method

$$g\left(\boldsymbol{x} \mid \boldsymbol{x}^t\right) = g\left(\boldsymbol{x}^t\right) + \nabla g\left(\boldsymbol{x}^t\right)^\top\left(\boldsymbol{x} - \boldsymbol{x}^t\right) + \frac{1}{2\alpha^t}\left\|\boldsymbol{x} - \boldsymbol{x}^t\right\|^2. \tag{2.25}$$

The update rule in each iteration is obtained by setting the gradient of surrogate function $g(\boldsymbol{x})$ to zero, yielding

$$\nabla g\left(\boldsymbol{x}^t\right) + \frac{1}{\alpha^t}\left(\boldsymbol{x} - \boldsymbol{x}^t\right) = 0, \tag{2.26}$$

$$\boldsymbol{x}^{t+1} = \boldsymbol{x}^t - \alpha^t \nabla g(\boldsymbol{x}^t), \tag{2.27}$$

where α^t is a constant at iteration t.

The surrogate functions can be classified as 1) Taylor expansion-based method, *e.g.* gradient descent method, and Newton's method; surrogates based on 2) block-wise convexity; 3) proximal gradient-like; 4) sum-utility function; 5) product of function; 6) composition of functions [64]. A general SCA framework is

summarized in Algorithm 4.

2.3.6 Karush-Kuhn-Tucker (KKT) Conditions

Karush-Kuhn-Tucker (KKT) conditions are the necessary conditions for optimality in optimization problem with constraints. Consider the following general optimization problem (primal problem)

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & f(\boldsymbol{x}), && \text{(2.28a)} \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) \leqslant 0, i = 1, \cdots, m, && \text{(2.28b)} \\ & \ell_j(\boldsymbol{x}) = 0, j = 1, \cdots, r, && \text{(2.28c)} \end{aligned}$$

where (2.28b) is the inequality constraints, and (2.28c) is the equality constraints. Subsequently, the Lagrangian function is defined by introducing the dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, given by

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i h_i(\boldsymbol{x}) + \sum_{j=1}^{r} \boldsymbol{\mu}_j \ell_j(\boldsymbol{x}). \tag{2.29}$$

The Lagrange dual function associated with (2.28) is then given by

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{2.30}$$

and the subsequent dual problem is

$$\begin{aligned} \max_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^m, \\ \boldsymbol{\mu} \in \mathbb{R}^r}} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\mu}), && \text{(2.31a)} \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geqslant \mathbf{0} && \text{(2.31b)} \end{aligned}$$

A necessary condition for $\boldsymbol{x}^\star$ to be a local optimal solution of the original primal problem is the KKT conditions [65, 66] such that

Table 2.1 Computational complexity of various algorithms.

Algorithm	Computational Complexity
BCD	$\mathcal{O}(B_{\mathsf{num}} \cdot C_{\text{block}})$ [55]
PDD	$\mathcal{O}(C_{\text{dual}} + C_{\text{primal}})$ [57]
WMMSE	$\mathcal{O}((K^2 N_t N_r^2 + K^2 N_r N_t^2 + K^2 N_t^3 + K N_r^3))$ [37]
FP	$\mathcal{O}(N_t^4 + N_t K + (N_t^3 + N_t)K + (K + N_t)^3)$ [61]
SCA	$\mathcal{O}(K^2(N_t N_r^2 + N_t^2 N_r) + K(N_t^3 + N_r^3))$ [63]

1. Stationarity Condition

$$\partial_{\boldsymbol{x}} L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = 0. \tag{2.32}$$

2. Complementary Slackness

$$\lambda_i h_i(\boldsymbol{x}) = 0, i = 1, \cdots, m. \tag{2.33}$$

3. Primal feasibility

$$h_i(\boldsymbol{x}) \leqslant 0, i = 1, \cdots, m, \tag{2.34}$$

$$\ell_j(\boldsymbol{x}) = 0, j = 1, \cdots, r. \tag{2.35}$$

4. Dual Feasibility

$$\lambda_i \geqslant 0, i = 1, \cdots, m. \tag{2.36}$$

The KKT conditions are necessary but not sufficient for global optimalty. However, if the optimization problem is convex, the conditions are necessary and sufficient.

2.3.7 Computational Complexity Comparison

We compare the per-round computational complexity of each optimization algorithm as in the Table 2.1. One round refers to one update for all users of the algorithm. In the computational complexity expressions, K, N_t, and N_r, are the number of users, the number of transmit antenna elements, and the number

of receive antenna elements. The users are equipped with a single antenna. In the computational complexity of BCD, B_{num} denotes the block number, and C_{block} is the complexity of each block. In addition, C_{dual} and C_{primal} are the complexities of updating the dual variables, and solving the primal subproblems, respectively.

Chapter 3

Joint Transmit and Receive Beamforming Design in Full-Duplex Integrated Sensing and Communications

3.1 Introduction

Next generation wireless communication networks are expected to support high data transmission rates, high-quality wireless connectivity with massive devices, and highly accurate and robust sensing capability [7, 67]. To realize these requirements, ISAC, which unifies the signal processing procedures and hardware framework between sensing and communication systems, is widely viewed as a promising solution to efficiently utilize the available spectral, hardware and energy resources.

Challenge: *Echo miss*. In the existing ISAC literature, many studies assume that the sensing takes place in mono-static mode [29–33] due to its relative simplicity compared to other sensing configurations (e.g., multi-static). In a mono-static ISAC system, the transmit (TX)/receive (RX) antennas are co-located, resulting in the transmit (dual-function) waveform being known at the receiver. Hence, the receiver can use the transmit waveform as a reference waveform to extract target information from the radar echo, thereby saving on the overhead associated with reference sharing. However, the transmit waveform is also expected to serve communications users in parallel, and typically communication frames are much longer than the radar echo round-trip times (RTTs). For example, in the 5G NR specifications [68], a standard radio frame has 10ms duration. For a target located at 100-1000m from the radar, its echo RTT is of the order of 1-10μs - orders of magnitude smaller than even the minimum unit of data scheduling (i.e., 1 slot = 0.5ms). Hence, the radar echo is drowned by the severe SI from the transmitter, which causes receiver saturation, where the power of the received signal exceeds the ADC dynamic range. Even if there is no ADC saturation due to the use of sufficient SI cancellation techniques before quantization, the radar echo may still be difficult to detect because it is masked by strong residual SI. We term this phenomenon the *echo-miss* problem. Thus, it is important to sufficiently suppress the residual SI to manageable levels.

To address the *echo-miss* problem and suppress the SI, one straightforward method is to physically separate the TX and RX antennas. The measurement-based study [27] shows that limited isolation capability can be achieved by a

combination of directional isolation, absorber, and cross-polarization. Specifically, in the experiments, a 35cm separation between the TX and RX antennas, along with an absorber, was shown to realize 45dB passive suppression. However, the power level of the SI can be large (i.e., up to 100dB larger than the receiver thermal noise floor [16, 18]). Thus, physical separation of TX and RX antennas may not entirely solve the *echo-miss* problem. Consequently, to integrate communication and sensing functions, the transceiver should work in the full-duplex (FD) mode to simultaneously transmit a dual-functional signal, receive the echo signal, and suppress the SI, caused by the leakage of the transmit signal to the receiver. Some attempts have been made for ISAC in the FD context, with [69] concentrating on waveform design by utilizing the waiting time of conventional pulsed radars to transmit communication signal in a single-antenna setup. For the multi-antenna case, [70] jointly optimizes relay beamformer, receive filter, and transmit power of the radar for a FD ISAC relay system, wherein the residual SI is assumed to be cancelled in advance by SI cancellation techniques. However, for the most part, the SI cancellation problem is underestimated by many ISAC studies [30, 31, 33, 71–73], which assume either ideal isolation between the TX and RX or rely on the radar-function-focused SI cancellation method in [74].

Previous Approaches for SI Cancellation. The SI cancellation problem has been actively studied in FD wireless communication systems [17, 19, 20] (i.e., without the additional sensing functionality). In general, SI cancellation techniques can be implemented in the propagation [21], analog [22–24], and digital domains [25, 26]. As shown in Fig. 3.1, the propagation-domain cancellation aims to minimize the coupling between the transmit and receive direct paths. This kind of cancellation is achieved by techniques based on path loss, cross-polarization, and antenna directionality [27]. Beyond this, the analog-domain cancellation aims to suppress the SI before the ADC, where a negative copy of the transmit waveform generated by the canceller circuit is subtracted from the received signal [20, 22, 23, 28]. Finally, as the last defense against SI, the digital-domain cancellation block utilizes either linear or non-linear adaptive filters to generate the negative of the residual SI [25, 26], and add it to the digital signal

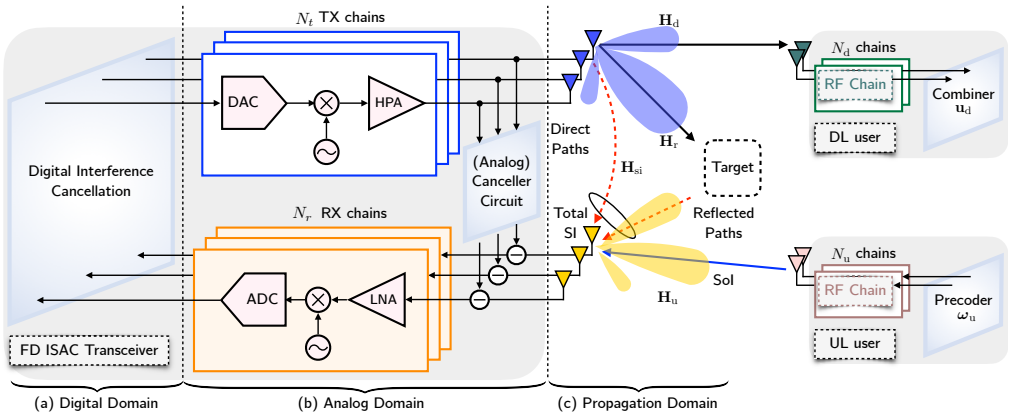


Figure 3.1 Full-duplex integrated sensing and communication system, and illustration of the boundaries and contents of the SI suppression in the propagation, analog, and digital domains for multi-antenna configurations.

after the ADC.

The above SI cancellation techniques for communications rely on the uncorrelated nature between the SI (e.g., downlink data stream) and the SoI (e.g., the uplink data stream); thus, the SI can be suppressed by adding its negative to the received signal without impairing the SoI. However, since the SoI in ISAC consists of uplink communications data and radar echoes that are correlated with the SI, it is challenging to effectively suppress the SI without distorting the radar echoes. To tackle this problem, we utilize the time-of-arrival (ToA) difference or the spatial angle-of-arrival (AoA) difference between the SI and echo are two promising approaches. With respect to the first approach, the early study [74] utilizes the temporal difference to generate a negative counterpart of the SI signal before the ADC (cf. Fig. 3.1 (b)), based on a gradient-learning method. Apart from adding this negative counterpart, an adaptive filter is also employed to generate a negative signal in digital domain to cancel the residual SI (cf. Fig. 3.1 (a)). In practice, many factors (e.g., RF taps, adaptive filter taps, and update algorithms) affect the accuracy of the generated negatives in both domains, which in turn, have a huge impact on the SI cancellation performance. If the SI signal is fast-changing, this approach may fail in tracking and can have high computational complexity.

Table 3.1 Novelty Comparison with Existing FD ISAC Literature

	Our work	[69]	[70]	[74]	[38–40]
Relay			✓		
Analog hardware design				✓	
Waveform design		✓			
Tx/Rx beamformers NSP design					✓
Tx/Rx beamformers joint design	✓		✓		
Uplink user	✓		✓		
Downlink user	✓	✓	✓		✓
Multi-antenna at users	✓				
Radar performance	✓	✓	✓	✓	✓
Communication performance	✓	✓			

In multi-antenna systems, an alternative way to suppress the SI in ISAC is by employing the spatial AoA difference between the SI and SoI. In [38], the SI cancellation based on TX and RX beamforming design is adopted in the digital domain (cf. Fig. 3.1a). Specifically, the TX beamformer is the weighted sum of two separate beams probing at a downlink communication user and a radar target, whose power allocation is controlled by a parameter. In terms of the RX beamformer, the NSP based on pseudoinverse operation is used to generate nulls in the angles of the downlink communication beam and SI. In [39] and [40], the NSP method is utilized to design hybrid TX and RX beamformers for sensing the target and communicating to a downlink user (cf. Figs. 3.1a and 3.1b). However, in these studies, the TX and RX beamformers are separately designed and only the RX beamformer is used for SI cancellation. Recently, intelligent reflecting surfaces (IRS) have emerged as a means to boost the SoI, and thus reduce the effect of SI [75,76], but this approach cannot actively cancel SI, and induces additional hardware and beamforming design complexity.

Thus, the potential of *joint ISAC TX-RX beamformer design* at transceiver to further suppress the SI has not been explored. In addition, the uplink communication performance in the FD ISAC system has not been analyzed. Given that the research on FD ISAC is still in its infancy (cf. Table 3.1), we consider a mono-static FD ISAC multiple input multiple output (MIMO) system. In this system, (a) the TX-RX beamformers at the transceiver, (b) transmit precoder at the uplink user, and (c) receive combiner at the downlink user are jointly

optimized. The aim of the optimization is to simultaneously (a) maximize the uplink and downlink rate, (b) maximize the transmit and receive radar beampattern power at the target, and (c) suppress the residual SI. Inspired by research adopting the penalty-dual decomposition (PDD) method (e.g., for FD mmWave hybrid beamforming [77], and IRS [78]), a penalty-based iterative algorithm is proposed to solve the optimization problem. Our proposed scheme can work when the received signal exceeds ADC dynamic range by adopting effective SI cancellation techniques before quantization.

To conclude, the main contribution is that we propose a promising approach to tackle SI cancellation and correlation between the radar echo and SI in FD ISAC transceiver. In this work we jointly design the transmit and receive beamformers at the transceiver, transmit precoder at the uplink user, and receive combiner at the downlink user to simultaneously (a) maximize the uplink and downlink communication rate, (b) maximize the transmit and receive radar beampattern power at the target, and (c) suppress the residual SI. To solve this optimization problem, we proposed a penalty-based iterative algorithm. Numerical results illustrate that the proposed design can effectively achieve up to 60 dB digital-domain SI cancellation, a higher average sum-rate, and more accurate radar parameter estimation compared with previous ISAC FD studies.

3.2 System Model

3.2.1 Signal Model

As shown in Fig. 3.1, we consider a single-cell narrowband FD MIMO ISAC transceiver equipped with N_{t} transmit antennas and N_{r} receive antennas. All antenna arrays are assumed to be uniform linear arrays (ULA) with half-wavelength spacing between adjacent antenna elements. The transceiver simultaneously serves one uplink user with N_{u} antennas, one downlink user with N_{d} antennas, and probes a target direction. We assume perfect channel state information (CSI) in the system model and optimization.

Let $s_{\mathrm{d}} \in \mathbb{C}$ denote the ISAC downlink transmit symbol, and $s_{\mathrm{u}} \in \mathbb{C}$ the uplink symbol. We assume both s_{d} and s_{u} have unit power. The received signal $\mathbf{y}_{\mathrm{d}} \in \mathbb{C}^{N_{\mathrm{d}} \times 1}$ at the downlink user is given by

$$\mathbf{y}_{\mathrm{d}} = \mathbf{H}_{\mathrm{d}} \mathbf{p} s_{\mathrm{d}} + \mathbf{n}_{\mathrm{d}}, \tag{3.1}$$

where $\mathbf{p} \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$ denotes the transmit precoder at the transceiver, $\mathbf{H}_{\mathrm{d}} \in \mathbb{C}^{N_{\mathrm{d}} \times N_{\mathrm{t}}}$ the downlink communication channel, $\mathbf{n}_{\mathrm{d}} \in \mathbb{C}^{N_{\mathrm{d}} \times 1}$ the independent and identically distributed (i.i.d) additive complex Gaussian noise (i.e., $\mathbf{n}_{\mathrm{d}} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{d}}^2 \mathbf{I}_{N_{\mathrm{d}}})$). At the downlink user, an estimate of s_{d}, denoted by $\widehat{s}_{\mathrm{d}}$, is obtained by filtering $\mathbf{y}_{\mathrm{d}}$ by a combiner $\mathbf{u}_{\mathrm{d}} \in \mathbb{C}^{N_{\mathrm{d}} \times 1}$, as follows:

$$\widehat{s}_{\mathrm{d}} = \mathbf{u}_{\mathrm{d}}^H \mathbf{y}_{\mathrm{d}} = \mathbf{u}_{\mathrm{d}}^H \mathbf{H}_{\mathrm{d}} \mathbf{p} s_{\mathrm{d}} + \mathbf{u}_{\mathrm{d}}^H \mathbf{n}_{\mathrm{d}}. \tag{3.2}$$

At the FD ISAC transceiver, the received signal incorporates the SoI (i.e., the radar echo and the uplink symbol, s_{u}), along with residual SI. We assume that the received signal has no clipping error due to sufficient SI cancellation techniques in propagation and analog domains, and a high ADC dynamic range[1],

[1]The dynamic range is defined by the ratio between the largest and smallest possible values of the input signal, which are respectively the residual SI and radar echo in our system model. From our link budget simulation, the ratio between the residual SI and radar echo is 134.15dB, thus we assume that the SI cancellation techniques before digital domain can achieve larger than 54.15dB SI cancellation.

of the order of 80dB, which can be achieved by a variety of ways, e.g., logarithmic ADC [79], modulo ADC [80, 81]). Hence, the received signal at the transceiver is given by

$$\mathbf{y}_{\mathrm{u}} = \mathbf{H}_{\mathrm{u}}\boldsymbol{\omega}_{\mathrm{u}}s_{\mathrm{u}} + \mathbf{H}\mathbf{p}s_{\mathrm{d}} + \mathbf{n}_{\mathrm{u}}, \tag{3.3}$$

where $\boldsymbol{\omega}_{\mathrm{u}} \in \mathbb{C}^{N_{\mathrm{u}}\times 1}$ denotes the precoder vector of the uplink user, $\mathbf{H}_{\mathrm{u}} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{u}}}$ the uplink communication channel, $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$ the aggregate channel comprising the radar channel, $\mathbf{H}_{\mathrm{r}} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$, and the self-interference channel, $\mathbf{H}_{\mathrm{si}} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$ (i.e., $\mathbf{H} = \mathbf{H}_{\mathrm{r}} + \mathbf{H}_{\mathrm{si}}$), and finally $\mathbf{n}_{\mathrm{u}} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{u}}^2\mathbf{I}_{N_{\mathrm{r}}}) \in \mathbb{C}^{N_{\mathrm{r}}\times 1}$ the i.i.d additive complex Gaussian noise. The above-mentioned channels will be elaborated in Section 3.2.2.

At the FD ISAC transceiver, an estimate of s_{u}, denoted by $\widehat{s}_{\mathrm{u}}$, is obtained by filtering $\mathbf{y}_{\mathrm{u}}$ by a receive beamformer $\mathbf{w} \in \mathbb{C}^{N_{\mathrm{r}}\times 1}$ as follows:

$$\begin{aligned}\widehat{s}_{\mathrm{u}} &= \mathbf{w}^H\mathbf{y}_{\mathbf{u}} \\ &= \underbrace{\mathbf{w}^H\mathbf{H}_{\mathrm{u}}\boldsymbol{\omega}_{\mathbf{u}}s_{\mathrm{u}}}_{\text{Uplink stream}} + \underbrace{\mathbf{w}^H\mathbf{H}_{\mathrm{r}}\mathbf{p}s_{\mathrm{d}}}_{\text{Radar return}} + \underbrace{\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p}s_{\mathrm{d}}}_{\text{Residual SI}} + \mathbf{w}^H\mathbf{n}_{\mathrm{u}},\end{aligned} \tag{3.4}$$

In Section 3.3, we design beamformers $\mathbf{w}$ and $\mathbf{p}$ such that $\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p} \approx 0$ (i.e., SI suppression) while the uplink symbol and the radar returns are amplified. To decode s_{u}, we treat the interfering radar return as noise, while we use the favorable correlation properties of communications signals [82] to isolate the uplink data stream from the radar returns to estimate the target parameters. We provide more details regarding this in Section 3.5.

Remark 2. *It is intuitive to choose a separate radar signal, and design two separate beamformers for radar and communications. However, for a fixed transmit power budget, this leads to the important question of power allocation across the communication and sensing symbols. Hence, it is not obvious that the radar performance may be better with separate signals/precoders. Furthermore, for full-duplex operation, beamformer design with separate radar and communication signals is more complex as residual SI needs to be suppressed twice. For these reasons, we adopt the dual-functional transmit*

signal and a common beamformer.

3.2.2 Channel Model

Communication Channel

The communication channels, $\mathbf{H}_{\mathrm{u}}$ and $\mathbf{H}_{\mathrm{d}}$, are assumed to experience both small-scale and large-scale fading, which is modelled as follows, taking $\mathbf{H}_{\mathrm{u}}$ as an example

$$\mathbf{H}_{\mathrm{u}} = \sqrt{\eta_{\mathrm{u}}}\,\mathbf{G}_{\mathrm{u}}, \tag{3.5}$$

where $\eta_{\mathrm{u}} \in \mathbb{C}$ is the large-scale fading coefficient (including geometric attenuation and shadow fading), and $\mathbf{G}_{\mathrm{u}} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{u}}}$ is the small-scale fading matrix modelled by the classic Rician fading model [83], as follows

$$\mathbf{G}_{\mathrm{u}} = \sqrt{\frac{\kappa}{\kappa+1}}\bar{\mathbf{G}}_{\mathrm{u}} + \sqrt{\frac{1}{\kappa+1}}\tilde{\mathbf{G}}_{\mathrm{u}}, \tag{3.6}$$

where $\kappa \quad (\geqslant 0)$ captures the proportion of the energy in the direct path (LoS), relative to the energy of the scattered paths (NLoS). We assume that the size of the antenna arrays is negligible compared with the distance between transceiver and user. Thus, the channel matrix corresponding to the LoS path (denoted by $\bar{\mathbf{G}}_{\mathrm{u}}$ in (3.6)) is given by

$$\bar{\mathbf{G}}_{\mathrm{u}} = \mathbf{a}(\theta_{\mathrm{u}})\mathbf{b}^{\top}(\theta'_{\mathrm{u}}), \tag{3.7}$$

where

— θ_{u} and θ'_{u} denote the AoA at the transceiver and angle-of-departure (AoD) at the uplink user, respectively.

— $\mathbf{a}(\theta_{\mathrm{u}}) = \left[1, e^{j2\pi\delta\sin(\theta_{\mathrm{u}})}, \ldots, e^{j2\pi(N_{\mathrm{r}}-1)\delta\sin(\theta_{\mathrm{u}})}\right]^{\top} \in \mathbb{C}^{N_{\mathrm{r}}\times 1}$ is the steering vector in the direction of the AoA at the transceiver, with $\delta = 0.5$.

— $\mathbf{b}(\theta'_{\mathrm{u}}) = \left[1, e^{j2\pi\delta\sin(\theta'_{\mathrm{u}})}, \ldots, e^{j2\pi(N_{\mathrm{u}}-1)\delta\sin(\theta'_{\mathrm{u}})}\right]^{\top} \in \mathbb{C}^{N_{\mathrm{u}}\times 1}$ is the steering vector in the direction of the AoD at the uplink user.

Additionally, $\tilde{\mathbf{G}}_{\mathrm{u}}$ is the i.i.d. complex Gaussian matrix, where each element follows $[\tilde{\mathbf{G}}_{\mathrm{u}}]_{i,j} \sim \mathcal{CN}(0,1)$.

Radar Channel

We assume that the radar operates in track model, where an initial estimate of the target's AoA is available from previous scanning. As in (3.3), the FD ISAC transceiver will receive a reflected echo with the target's AoA, range and velocity information (i.e., θ_{r}, r, v, respectively). Since we assume the transceiver to operate as a mono-static radar, the AoA and AoD are both equal to θ_{r}. We model the radar channel as varying slowly with time, with its expression given by

$$\mathbf{H}_{\mathrm{r}}(t) = \eta_{\mathrm{r}} e^{j2\pi f_{\mathrm{d}} t} \mathbf{a}(\theta_{\mathrm{r}}) \mathbf{b}^{\top}(\theta_{\mathrm{r}}), \tag{3.8}$$

where $f_{\mathrm{d}} = 2v f_{\mathrm{c}}/c$ is the Doppler frequency, f_{c} and c are the signal carrier frequency and the speed of light. Moreover, η_{r} is the attenuation factor, including round-trip path-loss and radar cross-section (RCS) σ_{r}, given in (3.9) derived from the radar range equation [84]:

$$\eta_{\mathrm{r}} = \sqrt{\frac{\lambda^2 \sigma_{\mathrm{r}}}{(4\pi)^3 r^4}}. \tag{3.9}$$

Self-Interference Channel

In this work, we focus on the residual digital-domain SI cancellation assuming that the analog SI cancellation method is effective, i.e., there is no saturation at the ADC of the receive RF chains at the transceiver.

To obtain $\mathbf{H}_{\mathrm{si}}$, channel estimation methods based on classic estimation algorithms (e.g., least square (LS) [25], maximum-likelihood (ML) [85]) or machine learning-based methods [86] can be used. For illustration and without loss of generality, we consider a channel model derived from experiments [19, 87], which is similar to (3.6) and given by:

$$\mathbf{H}_{\mathrm{si}} = \sqrt{\eta_{\mathrm{si}}}\, \mathbf{G}_{\mathrm{si}}, \tag{3.10}$$

where η_{si} and $\mathbf{G}_{\text{si}}$ denote the large-scale fading coefficient and small-scale fading matrix, respectively. Specifically, $\mathbf{G}_{\text{si}}$ has the same structure as $\mathbf{G}_{\text{u}}$ in (3.6), containing LoS and NLoS components for generalization.

3.2.3 Problem Formulation

Communication Rate

With the uplink channel $\mathbf{H}_{\text{u}}$, radar channel $\mathbf{H}_{\text{r}}$[2], and SI channel $\mathbf{H}_{\text{si}}$ in (3.5), (3.8), and (3.10), the normalized uplink communication rate is given by

$$R_{\text{u}} = \log_2(1 + \gamma_{\text{u}}), \tag{3.11}$$

where γ_{u} is the Signal-to-Interference-plus-Noise Ratio (SINR) of the received uplink signal. The expression for γ_{u} is given by

$$\gamma_{\text{u}} = \frac{|\mathbf{w}^H \mathbf{H}_{\text{u}} \boldsymbol{\omega}_{\text{u}}|^2}{|\mathbf{w}^H \mathbf{H} \mathbf{p}|^2 + \|\mathbf{w}\|^2 \sigma_{\text{u}}^2}. \tag{3.12}$$

Similarly, the normalized downlink communication rate is given by

$$R_{\text{d}} = \log_2(1 + \gamma_{\text{d}}), \tag{3.13}$$

where γ_{d} is the SINR of the downlink signal given by

$$\gamma_{\text{d}} = \frac{|\mathbf{u}_{\text{d}}^H \mathbf{H}_{\text{d}} \mathbf{p}|^2}{\|\mathbf{u}_{\text{d}}\|^2 \sigma_{\text{d}}^2}. \tag{3.14}$$

In FD systems, in general, the downlink user would experience interference from uplink transmissions, as well. However, we assume that this interference is effectively managed by scheduling [88–90], and thus ignore its effect in (3.13).

[2]For notation simplicity, we drop the time instant, (t), in $\mathbf{H}_{\text{r}}(t)$.

Radar Beampattern Power

The transmit and receive beampattern power in the target direction, denoted by G_{t} and G_{r}, respectively, as a function of $\mathbf{p}$ and $\mathbf{w}$, is given by

$$G_{\mathrm{t}} = |\mathbf{b}^H(\theta_{\mathrm{r}})\mathbf{p}|^2, \quad G_{\mathrm{r}} = |\mathbf{w}^H\mathbf{a}(\theta_{\mathrm{r}})|^2. \tag{3.15}$$

Remark 3. *We use beampattern power as a proxy for radar SINR. While the latter may be a more intuitive sensing metric, its fractional form makes it non-convex – and thus, intractable – as well as difficult to obtain good bounds. On the other hand, the tractability of the beampattern power metric enables us to obtain closed-form solutions for key sub-problems (see Section 3.3.2) that help Algorithm 5 converge faster.*

Objective Function and Constraints

We aim to design the transmit and receive beamformer $\mathbf{p}, \mathbf{w}$ at the transceiver, precoder $\boldsymbol{\omega}_{\mathrm{u}}$ at the uplink user, and combiner $\mathbf{u}_{\mathrm{d}}$ at the downlink user to simultaneously maximize the uplink and downlink rate, the transmit and receive radar beampattern power at the target, and suppress the residual SI. In addition, the following constraints should be satisfied: 1) transmit power constraint at the transceiver and uplink user, i.e. $\|\mathbf{p}\|^2 \leqslant P_{\mathrm{d}}$, $\|\boldsymbol{\omega}_{\mathrm{u}}\|^2 \leqslant P_{\mathrm{u}}$, where P_{d} and P_{u} are the maximum transmit power permitted at transceiver and uplink user, respectively; 2) receive power constraint at the transceiver, i.e. $\|\mathbf{w}\|^2 = 1$; 3) SI suppression constraint, i.e., $\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p} = 0$. With the above constraints, we can formulate the joint ISAC TX-RX beamformers design as

$$\begin{aligned}
\max_{\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}} \quad & \alpha_1 R_{\mathrm{u}} + \alpha_2 R_{\mathrm{d}} + \alpha_3 G_{\mathrm{t}} + \alpha_4 G_{\mathrm{r}} && \text{(3.16a)}\\
\text{s.t.} \quad & \|\mathbf{p}\|_2^2 \leqslant P_{\mathrm{d}}, && \text{(3.16b)}\\
& \|\boldsymbol{\omega}_{\mathrm{u}}\|_2^2 \leqslant P_{\mathrm{u}}, && \text{(3.16c)}\\
& \|\mathbf{w}\|_2^2 = 1, && \text{(3.16d)}\\
& \mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p} = 0. && \text{(3.16e)}
\end{aligned}$$

The weight of the four components in (3.16a) is controlled by the coefficient $\alpha_n, \forall n \in \{1, 2, 3, 4\}$. The problem (3.16) is difficult to solve due to the following reasons: 1) the coupling between the transmit and receive beamformer in constraint (3.16e); 2) the non-convex communication metrics (i.e., $R_{\mathrm{u/d}}$) in objective (3.16a); 3) the quadratic radar metrics (i.e. $G_{\mathrm{t/r}}$) in objective (3.16a) in the maximization problem. To solve (3.16), we transform it into an alternative form that permits an iterative solution with guaranteed convergence to a feasible solution. Note that maximizing rates alone might lead to suboptimal use of power resources, resulting in inefficient beampatterns. Therefore, a regularization approach by adding the transmit and receive beampattern power, i.e., $G_{\mathrm{t/r}}$, can be used to balance the rates and power efficiently.

3.3 Joint TX-RX Beamformer Design

3.3.1 Problem Transformation

To address the difficulties in the previous section, we transform problem (3.16) based on the PDD method [57], the weighted Minimum Mean Square Error (MMSE) approach [37], and the reformulation of the quadratic beampattern power function [91]. Each of these techniques addresses a specific difficulty, as explained below:

Coupling Constraint Transformation by Penalty Term

To tackle the coupling between $\mathbf{w}$ and $\mathbf{p}$ in constraint (3.16e), we transform the constraint $\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p} = 0$ into a penalty term $\frac{1}{2\beta}\|\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p}\|_2^2$, and add it to the objective function with the penalty parameter $\beta > 0$. In particular, β determines the penalty intensity of the digital-domain SI suppression, with lower values of β prioritizing SI suppression over communication and sensing performance. Thus, problem (3.16) can be reformulated as (3.17):

$$\begin{aligned} \max_{\{\mathbf{p},\mathbf{w}\},\{\boldsymbol{\omega}_{\mathrm{u}},\mathbf{u}_{\mathrm{d}}\}} \quad & \alpha_1 R_{\mathrm{u}} + \alpha_2 R_{\mathrm{d}} + \alpha_3 G_{\mathrm{t}} + \alpha_4 G_{\mathrm{r}} \\ & - \frac{1}{2\beta}\|\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p}\|_2^2 \qquad (3.17a) \\ \text{s.t.} \quad & (3.16b) - (3.16d). \end{aligned}$$

Communication Rate Transformation by WMMSE Method

To circumvent the non-convexity of the rate terms in (3.17a), we adopt the WMMSE method to reformulate these terms based on the equivalence between the rate and MSE matrix [36,37]. With the received signal (3.4) at the transceiver, the MSE matrix at the transceiver is defined as

$$\begin{aligned} \mathsf{E}_{\mathrm{BS}} &\triangleq \mathscr{E}\left\{(\widehat{\mathbf{s}}_{\mathrm{u}} - \mathbf{s}_{\mathrm{u}})\,(\widehat{\mathbf{s}}_{\mathrm{u}} - \mathbf{s}_{\mathrm{u}})^H\right\} \\ &= \mathbf{w}^H\mathbf{H}_{\mathrm{u}}\boldsymbol{\omega}_{\mathrm{u}}\boldsymbol{\omega}_{\mathrm{u}}^H\mathbf{H}_{\mathrm{u}}^H\mathbf{w} - 2\,\mathrm{Re}\{\mathbf{w}^H\mathbf{H}_{\mathrm{u}}\boldsymbol{\omega}_{\mathrm{u}}\} \\ &\quad + \mathbf{w}^H(\mathbf{H}\mathbf{p}\mathbf{p}^H\mathbf{H}^H + \sigma_{\mathrm{u}}^2\mathbf{I})\mathbf{w} + 1. \end{aligned} \qquad (3.18)$$

Similarly, the MSE matrix function of downlink user can be also defined as

$$\begin{aligned}\mathsf{E}_{\mathrm{d}} &\triangleq \mathscr{E}\left\{(\widehat{\mathbf{s}}_{\mathrm{d}}-\mathbf{s}_{\mathrm{d}})\,(\widehat{\mathbf{s}}_{\mathrm{d}}-\mathbf{s}_{\mathrm{d}})^{H}\right\}\\ &= \mathbf{u}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}}\mathbf{p}\mathbf{p}^{H}\mathbf{H}_{\mathrm{d}}^{H}\mathbf{u}_{\mathrm{d}} - 2\,\mathrm{Re}\{\mathbf{u}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}}\mathbf{p}\}\\ &\quad + \sigma_{\mathrm{d}}^{2}\mathbf{u}_{\mathrm{d}}^{H}\mathbf{u}_{\mathrm{d}} + 1.\end{aligned} \tag{3.19}$$

Then, we introduce ρ_{u} and ρ_{d} as the auxiliary variables, and obtain the transformed objective function shown in (3.20) on the basis of rate-WMMSE relationship.

$$\begin{aligned}\max_{\substack{\{\mathbf{p},\mathbf{w}\},\{\boldsymbol{\omega}_{\mathrm{u}},\mathbf{u}_{\mathrm{d}}\},\\ \{\rho_{\mathrm{u}},\rho_{\mathrm{d}}\}}} \quad & \alpha_1(\log_2\rho_{\mathrm{u}} - \rho_{\mathrm{u}}\mathsf{E}_{\mathrm{BS}}) + \alpha_2(\log_2\rho_{\mathrm{d}} - \rho_{\mathrm{d}}\mathsf{E}_{\mathrm{d}})\\ & + \alpha_3 G_{\mathrm{t}} + \alpha_4 G_{\mathrm{r}} - \frac{1}{2\beta}\|\mathbf{w}^{H}\mathbf{H}_{\mathrm{si}}\mathbf{p}\|_2^2 \qquad (3.20a)\\ \text{s.t.} \quad & (3.16\mathrm{b}) - (3.16\mathrm{d}).\end{aligned}$$

Beampattern Power Transformation

To make G_{t} and G_{r} concave, we reformulate them with $V_{\mathrm{t}} = \|\mathbf{b}(\theta_{\mathrm{r}})\|^2$ as

$$\begin{aligned}G_{\mathrm{t}} &= |\mathbf{b}^{H}(\theta_{\mathrm{r}})\mathbf{p}|^2 = \mathbf{p}^{H}\mathbf{b}(\theta_{\mathrm{r}})\mathbf{b}^{H}(\theta_{\mathrm{r}})\mathbf{p}\\ &= \mathbf{p}^{H}\mathbf{b}(\theta_{\mathrm{r}})\mathbf{b}^{H}(\theta_{\mathrm{r}})\mathbf{p} - V_{\mathrm{t}}P_{\mathrm{d}} + V_{\mathrm{t}}P_{\mathrm{d}}\\ &\leqslant \mathbf{p}^{H}\mathbf{b}(\theta_{\mathrm{r}})\mathbf{b}^{H}(\theta_{\mathrm{r}})\mathbf{p} - \mathbf{p}^{H}V_{\mathrm{t}}\mathbf{I}\mathbf{p} + V_{\mathrm{t}}P_{\mathrm{d}}\ (\text{Utilizing } \mathbf{p}^{H}\mathbf{p}\leqslant P_{\mathrm{d}})\\ &= \mathbf{p}^{H}(\mathbf{b}(\theta_{\mathrm{r}})\mathbf{b}^{H}(\theta_{\mathrm{r}}) - V_{\mathrm{t}}\mathbf{I})\mathbf{p} + V_{\mathrm{t}}P_{\mathrm{d}}\\ &= \mathbf{p}^{H}\mathbf{Z}_{\mathrm{t}}(\theta_{\mathrm{r}})\mathbf{p} + V_{\mathrm{t}}P_{\mathrm{d}},\end{aligned} \tag{3.21}$$

where we denote $\mathbf{Z}_{\mathrm{t}}(\theta_{\mathrm{r}}) = \mathbf{b}(\theta_{\mathrm{r}})\mathbf{b}^{H}(\theta_{\mathrm{r}}) - V_{\mathrm{t}}\mathbf{I}$. The equality holds when the power of $\mathbf{p}$ reaches P_{d}. With the expression of the steering vector, it is straightforward to know that $\mathbf{b}(\theta_{\mathrm{r}})\mathbf{b}^{H}(\theta_{\mathrm{r}})$ is a rank-1 matrix with the eigenvalue $\|\mathbf{b}(\theta_{\mathrm{r}})\|^2$. Therefore, $\mathbf{Z}_{\mathrm{t}}(\theta_{\mathrm{r}})$ is negative semi-definite, and $\mathbf{p}^{H}Z(\theta_{\mathrm{r}})\mathbf{p}$ is concave. Subsequently, maximizing G_{t} can be approximated as the maximization of $\mathbf{p}^{H}\mathbf{Z}_{\mathrm{t}}(\theta_{\mathrm{r}})\mathbf{p}$ by ignoring the constant term $V_{\mathrm{t}}P_{\mathrm{d}}$.

Analogous to the reformulation of G_{t}, the expression of G_{r} is given by

$$G_{\mathrm{r}} \leqslant \mathbf{w}^H \mathbf{Z}_{\mathrm{r}}(\theta_{\mathrm{r}})\mathbf{w} + V_{\mathrm{r}}, \tag{3.22}$$

where $\mathbf{Z}_{\mathrm{r}}(\theta_{\mathrm{r}}) = \mathbf{a}(\theta_{\mathrm{r}})\mathbf{a}_{\mathrm{r}}^H(\theta_{\mathrm{r}}) - V_{\mathrm{r}}\mathbf{I}$ and $V_{\mathrm{r}} = \|\mathbf{a}(\theta_{\mathrm{r}})\|^2$. The equality holds when $\|\mathbf{w}\|_2^2 = 1$. Consequently, problem (3.16) is reformulated as problem (3.23) shown below.

$$\begin{aligned}
\max_{\substack{\{\mathbf{p},\mathbf{w}\},\{\boldsymbol{\omega}_{\mathrm{u}},\mathbf{u}_{\mathrm{d}}\},\\ \{\rho_{\mathrm{u}},\rho_{\mathrm{d}}\}}} \quad & \alpha_1(\log_2 \rho_{\mathrm{u}} - \rho_{\mathrm{u}}\mathsf{E}_{\mathrm{BS}}) + \alpha_2(\log_2 \rho_{\mathrm{d}} - \rho_{\mathrm{d}}\mathsf{E}_{\mathrm{d}}) + \alpha_3 \mathbf{p}^H \mathbf{Z}_{\mathrm{t}}(\theta_{\mathrm{t}})\mathbf{p} && \text{(3.23a)}\\
& + \alpha_4 \mathbf{w}^H \mathbf{Z}_{\mathrm{r}}(\theta_{\mathrm{r}})\mathbf{w} - \frac{1}{2\beta}\|\mathbf{w}^H \mathbf{H}_{\mathrm{si}}\mathbf{p}\|_2^2 \\
\text{s.t.} \quad & \text{(3.16b)} - \text{(3.16d)}.
\end{aligned}$$

3.3.2 Penalty-based Block Update Algorithm

With the BCD method, an iterative approach is utilized where each variable is optimized while keeping the others fixed, till convergence.

Block $\{\rho_{\mathrm{u}}, \rho_{\mathrm{d}}\}$

When $\{\mathbf{p}, \mathbf{w}\}$, $\{\boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}\}$ are fixed, the two separate sub-problems with regards to ρ_{u} and ρ_{d} are both convex and unconstrained given by

$$\max_{\rho_{\mathrm{u}}} \quad \log_2 \rho_{\mathrm{u}} - \rho_{\mathrm{u}}\mathsf{E}_{\mathrm{BS}}, \tag{3.24}$$

and

$$\max_{\rho_{\mathrm{d}}} \quad \log_2 \rho_{\mathrm{d}} - \rho_{\mathrm{d}}\mathsf{E}_{\mathrm{d}}. \tag{3.25}$$

Thus, the optimal ρ_{u}^* and ρ_{d}^* can be obtained by setting the partial derivative with respect to the two optimized variables to zero. Subsequently, the expres-

sions for ρ_{u}^* and ρ_{d}^* are:

$$\begin{aligned}\rho_{\mathrm{u}}^* &= \mathsf{E}_{\mathrm{BS}}^{-1} \\ &= (\mathbf{w}^H \mathbf{H}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}}^H \mathbf{H}_{\mathrm{u}}^H \mathbf{w} - 2\,\mathrm{Re}\{\mathbf{w}^H \mathbf{H}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}}\} \\ &\quad + \mathbf{w}^H (\mathbf{H}\mathbf{p}\mathbf{p}^H \mathbf{H}^H + \sigma_{\mathrm{u}}^2 \mathbf{I})\mathbf{w} + 1)^{-1},\end{aligned} \tag{3.26}$$

$$\begin{aligned}\rho_{\mathrm{d}}^* &= \mathsf{E}_{\mathrm{d}}^{-1} \\ &= (\mathbf{u}_{\mathrm{d}}^H \mathbf{H}_{\mathrm{d}} \mathbf{p}\mathbf{p}^H \mathbf{H}_{\mathrm{d}}^H \mathbf{u}_{\mathrm{d}} - 2\,\mathrm{Re}\{\mathbf{u}_{\mathrm{d}}^H \mathbf{H}_{\mathrm{d}} \mathbf{p}\} \\ &\quad + \sigma_{\mathrm{d}}^2 \mathbf{u}_{\mathrm{d}}^H \mathbf{u}_{\mathrm{d}} + 1)^{-1}.\end{aligned} \tag{3.27}$$

The complexity of (3.26) and (3.27) are $\mathcal{O}(N_{\mathrm{r}}^2)$, and $\mathcal{O}(N_{\mathrm{t}} N_{\mathrm{d}})$, respectively, mainly due to matrix multiplication.

Block $\{\boldsymbol{\omega}_{\mathrm{u}}\}$

With fixed $\{\mathbf{p}, \mathbf{w}\}$, $\{\rho_{\mathrm{u}}, \rho_{\mathrm{d}}\}$, the sub-problem with regard to $\boldsymbol{\omega}_{\mathrm{u}}$ is given by

$$\begin{aligned}\min_{\boldsymbol{\omega}_{\mathrm{u}}} \quad & \alpha_1 \mathsf{E}_{\mathrm{BS}} \\ \text{s.t.} \quad & \|\boldsymbol{\omega}_{\mathrm{u}}\|_2^2 \leqslant P_{\mathrm{u}}.\end{aligned} \tag{3.28}$$

Since the objective function is convex with the convex constraint, the Lagrange multiplier method based on the KKT condition is used with expression:

$$\min_{\boldsymbol{\omega}_{\mathrm{u}}} \quad \alpha_1 \mathsf{E}_{\mathrm{BS}} + \mu(\|\boldsymbol{\omega}_{\mathrm{u}}\|_2^2 - P_{\mathrm{u}}). \tag{3.29}$$

To solve this problem, we first substitute (3.29) with (3.18). Subsequently, we take the partial derivative of the Lagrangian function with regard to $\boldsymbol{\omega}_{\mathrm{u}}$ and μ, respectively, and set both to zero. As a result, we obtain

$$\boldsymbol{\omega}_{\mathrm{u}} = (\alpha_1 \mathbf{w}^H \mathbf{H}_{\mathrm{u}} \mathbf{H}_{\mathrm{u}}^H \mathbf{w} + \mu)^{-1} \alpha_1 \mathbf{H}_{\mathrm{u}}^H \mathbf{w}, \tag{3.30}$$

and

$$\|\boldsymbol{\omega}_{\mathrm{u}}\|_2^2 = P_{\mathrm{u}}. \tag{3.31}$$

By substituting (3.31) with (3.30), we can then utilize the bisection search to obtain μ^*. Thereon, the optimal solution $\boldsymbol{\omega}_{\mathrm{u}}^*$ is obtained by replacing μ with μ^* in (3.30):

$$\boldsymbol{\omega}_{\mathrm{u}}^* = (\alpha_1 \mathbf{w}^H \mathbf{H}_{\mathrm{u}} \mathbf{H}_{\mathrm{u}}^H \mathbf{w} + \mu^*)^{-1} \alpha_1 \mathbf{H}_{\mathrm{u}}^H \mathbf{w}. \tag{3.32}$$

Updating $\boldsymbol{\omega}_{\mathrm{u}}$ requires $\mathcal{O}(I_1 N_{\mathrm{r}} N_{\mathrm{u}})$ complexity due to the bisection search for μ with I_1 the number of iterations and the matrix multiplication.

Block $\{\mathbf{w}\}$

Under the condition that $\mathbf{p}$, $\boldsymbol{\omega}_{\mathrm{u}}$, $\mathbf{u}_{\mathrm{d}}$, ρ_{u}, and ρ_{d} are all fixed, $\mathbf{w}$ is only related to the first and the last two terms in (3.23a), yielding the following convex sub-problem

$$\min_{\mathbf{w}} \quad \alpha_1 \rho_{\mathrm{u}} \mathsf{E}_{\mathrm{BS}} + \alpha_4 \mathbf{w}^H \mathbf{Z}_{\mathrm{r}}(\theta_{\mathrm{r}}) \alpha_4 \mathbf{w} + \frac{1}{2\beta} \|\mathbf{w}^H \mathbf{H}_{\mathrm{si}} \mathbf{p}\|_2^2. \tag{3.33}$$

Although the constraint (3.16d) that $\|\mathbf{w}\|_2^2 = 1$ is a non-convex constraint, we firstly address this sub-problem with respect to $\mathbf{w}$ as an unconstrained complex optimization problem, and normalize the obtained solution after the convergence of the algorithm is satisfied. Equating the partial derivative of (3.33) with respect to $\mathbf{w}$ to zero, the optimal $\mathbf{w}^*$ is given by

$$\mathbf{w}^* = (2\mathbf{X}_{\mathrm{u}} + \frac{1}{\beta} \mathbf{H}_{\mathrm{si}} \mathbf{p} \mathbf{p}^H \mathbf{H}_{\mathrm{si}}^H)^{-1} 2\alpha_1 \rho_{\mathrm{u}} \mathbf{H}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}}, \tag{3.34}$$

where $\mathbf{X}_{\mathrm{u}}$ is defined as

$$\begin{aligned} \mathbf{X}_{\mathrm{u}} = {} & \alpha_1 \rho_{\mathrm{u}} \mathbf{H}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}}^H \mathbf{H}_{\mathrm{u}}^H + \alpha_1 \rho_{\mathrm{u}} \mathbf{H} \mathbf{p} \mathbf{p}^H \mathbf{H}^H \\ & + \alpha_1 \rho_{\mathrm{u}} \sigma_{\mathrm{u}}^2 \mathbf{I} - \alpha_4 \mathbf{Z}_{\mathrm{r}}(\theta_{\mathrm{r}}). \end{aligned} \tag{3.35}$$

The complexity of this step is $\mathcal{O}(N_{\mathrm{r}}^3)$ mainly due to the matrix inversion. Note that the high computational complexity of this step makes it infeasible for massive MIMO. Future work should focus on developing a low-complexity algorithm.

Block $\{\mathbf{u}_\mathrm{d}\}$

The uplink precoder is only related to the second term in (3.23a) with other blocks fixed. Therefore, the unconstrained convex sub-problem is given by

$$\begin{aligned}\min_{\mathbf{u}_\mathrm{d}} \quad & \mathbf{u}_\mathrm{d}^H \mathbf{H}_\mathrm{d} \mathbf{p}\mathbf{p}^H \mathbf{H}_\mathrm{d}^H \mathbf{u}_\mathrm{d} - 2\,\mathrm{Re}\{\mathbf{u}_\mathrm{d}^H \mathbf{H}_\mathrm{d} \mathbf{p}\} \\ & + \sigma_\mathrm{d}^2 \mathbf{u}_\mathrm{d}^H \mathbf{u}_\mathrm{d} + 1.\end{aligned} \tag{3.36}$$

The optimal solution is given by

$$\mathbf{u}_\mathrm{d}^* = (\mathbf{H}_\mathrm{d} \mathbf{p}\mathbf{p}^H \mathbf{H}_\mathrm{d}^H + \sigma_\mathrm{d}^2 \mathbf{I})^{-1} \mathbf{H}_\mathrm{d} \mathbf{p}. \tag{3.37}$$

Updating $\mathbf{u}_\mathrm{d}$ in (3.37) requires $\mathcal{O}(N_\mathrm{t} N_\mathrm{d})$ complexity mainly due to matrix multiplication.

Block $\{\mathbf{p}\}$

Under fixed ρ_u, ρ_d, $\mathbf{w}$, $\boldsymbol{\omega}_\mathrm{u}$ and $\mathbf{u}_\mathrm{d}$, the sub-problem with respect to $\mathbf{p}$ is a convex optimization problem with a convex constraint on the transceiver transmit power, P_d, as follows

$$\begin{aligned}\min_{\mathbf{p}} \quad & \alpha_1 \rho_\mathrm{u} \mathsf{E}_\mathrm{BS} + \alpha_2 \rho_\mathrm{d} \mathsf{E}_\mathrm{d} - \alpha_3 \mathbf{p}^H \mathbf{Z}_\mathrm{t}(\theta_\mathrm{t}) \mathbf{p} + \frac{1}{2\beta} \|\mathbf{w}^H \mathbf{H}_\mathrm{si} \mathbf{p}\|^2 \\ \text{s.t.} \quad & \|\mathbf{p}\|_2^2 \leqslant P_\mathrm{d}.\end{aligned} \tag{3.38}$$

Similar to the design of block $\{\boldsymbol{\omega}_\mathrm{u}\}$ in 3.3.2, this sub-problem with respect to $\mathbf{p}$ can be reformulated by the Lagrange multiplier method based on the KKT condition written below:

$$\begin{aligned}\min_{\mathbf{p}} \quad & \alpha_1 \rho_\mathrm{u} \mathsf{E}_\mathrm{BS} + \alpha_2 \rho_\mathrm{d} \mathsf{E}_\mathrm{d} - \alpha_3 \mathbf{p}^H \mathbf{Z}_\mathrm{t}(\theta_\mathrm{t}) \mathbf{p} + \frac{1}{2\beta} \|\mathbf{w}^H \mathbf{H}_\mathrm{si} \mathbf{p}\|^2 \\ & + \Gamma(\|\mathbf{p}\|_2^2 - P_\mathrm{d}).\end{aligned} \tag{3.39}$$

By taking the partial derivative with respect to Γ and $\mathbf{p}$, respectively, and set-

ting them to zero, we obtain

$$\|\mathbf{p}\|_2^2 = P_{\mathrm{d}}, \tag{3.40}$$

and

$$\begin{aligned}\mathbf{p} = (2\mathbf{X}_{\mathrm{d}} + \frac{1}{\beta}(\mathbf{H}_{\mathrm{si}}^H \mathbf{w}\mathbf{w}^H \mathbf{H}_{\mathrm{si}}) + 2\Gamma\mathbf{I})^{-1} \\ \cdot (2\alpha_2 \rho_{\mathrm{d}} \mathbf{H}_{\mathrm{d}}^H \mathbf{u}_{\mathrm{d}}),\end{aligned} \tag{3.41}$$

where $\mathbf{X}_{\mathrm{d}}$ is defined in (3.42) for expression simplicity,

$$\mathbf{X}_{\mathrm{d}} = \alpha_1 \rho_{\mathrm{u}} \mathbf{H}^H \mathbf{w}\mathbf{w}^H \mathbf{H} + \alpha_2 \rho_{\mathrm{d}} \mathbf{H}_{\mathrm{d}} \mathbf{u}_{\mathrm{d}} \mathbf{u}_{\mathrm{d}}^H \mathbf{H}_{\mathrm{d}}^H - \alpha_3 \mathbf{Z}_{\mathrm{t}}(\theta_{\mathrm{t}})\mathbf{I}. \tag{3.42}$$

Subsequently, we substitute (3.41) into (3.40), and utilize the bisection search to obtain the optimal Γ^*. Finally, the optimal $\mathbf{p}^*$ is given by

$$\begin{aligned}\mathbf{p}^* = (2\mathbf{X}_{\mathrm{d}} + \frac{1}{\beta}(\mathbf{H}_{\mathrm{si}}^H \mathbf{w}\mathbf{w}^H \mathbf{H}_{\mathrm{si}}) + 2\Gamma^*\mathbf{I})^{-1} \\ \cdot (2\alpha_2 \rho_{\mathrm{d}} \mathbf{H}_{\mathrm{d}}^H \mathbf{u}_{\mathrm{d}}).\end{aligned} \tag{3.43}$$

The complexity of updating $\mathbf{p}$ in (3.43) is $\mathcal{O}(I_2 N_{\mathrm{t}}^3)$ caused by the matrix inversion, where I_2 denotes the number of iterations in the bisection search for Γ.

Summary

Following the framework of penalty-based approach, iterative BCD update process, and setting appropriate initial value of the optimized variables, ρ_{u}, ρ_{d}, $\mathbf{w}$, $\boldsymbol{\omega}_{\mathrm{u}}$, $\mathbf{u}_{\mathrm{d}}$, and $\mathbf{p}$, the optimal solution in each block can be found after reaching the convergence condition. To satisfy the receive power constraint of $\mathbf{w}$, we normalize $\mathbf{w}$ after the whole iteration as below:

$$\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|}. \tag{3.44}$$

We summarize our proposed joint ISAC TX-RX beamformers design algorithm in Algorithm 5.

Remark 4. *An alternative to* (3.16e) *is to restrict the residual SI power to below a*

Algorithm 5: Proposed Penalty-based Joint Transmit and Receive Beamformer Design

Input: $\mathbf{H}_{\mathrm{d}}$, $\mathbf{H}_{\mathrm{u}}$, $\mathbf{H}_{\mathrm{r}}$, $\mathbf{H}_{\mathrm{si}}$, P_{d}, P_{u}.
Output: $\rho_{\mathrm{u}}^*, \rho_{\mathrm{d}}^*, \mathbf{p}^*, \mathbf{w}^*, \boldsymbol{\omega}_{\mathrm{u}}^*, \mathbf{u}_{\mathrm{d}}^*$.

Initialize $\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}$ randomly, $\beta = 10^{-25}$;
while no convergence of objective function (3.23a) **do**
    Update ρ_{u}^* by (3.26). Complexity = $\mathcal{O}(N_{\mathrm{r}}^2)$.
    Update ρ_{d}^* by (3.27). Complexity = $\mathcal{O}(N_{\mathrm{t}}N_{\mathrm{d}})$.
    Update $\boldsymbol{\omega}_{\mathrm{u}}^*$ by (3.32). Complexity = $\mathcal{O}(I_1 N_{\mathrm{r}} N_{\mathrm{u}})$.
    Update $\mathbf{w}^*$ by (3.34). Complexity = $\mathcal{O}(N_{\mathrm{r}}^3)$.
    Update $\mathbf{u}_{\mathrm{d}}^*$ by (3.37). Complexity = $\mathcal{O}(N_{\mathrm{t}}N_{\mathrm{d}})$.
    Update $\mathbf{p}^*$ by (3.43). Complexity = $\mathcal{O}(I_2 N_{\mathrm{t}}^3)$.
end
Normalize $\mathbf{w}^*$ by (3.44);
Return $\rho_{\mathrm{u}}^*, \rho_{\mathrm{d}}^*, \mathbf{p}^*, \mathbf{w}^*, \boldsymbol{\omega}_{\mathrm{u}}^*, \mathbf{u}_{\mathrm{d}}^*$.

threshold (i.s., $\frac{1}{2\beta}\|\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p}\|_2^2 < \varepsilon$), which has been shown to improve the communications performance in [92]. However, we have adopted a stronger constraint in (3.16e) *because with the inclusion of sensing, where the radar returns are both very weak and strongly correlated with the residual SI, the residual SI needs to be suppressed to a far greater extent.*

Nevertheless, enforcing such a strong constraint deteriorates both the communications and sensing performance. Our compromise to this trade-off is detailed in the following comment.

Remark 5. ***Comparison between PDD and Algorithm 5*** *PDD is a widely-used technique to solve optimization problems involving non-smooth, non-convex functions [57]. Algorithm 5 is similar to PDD due to the same problem formulation as shown in* (3.17a). *However, the PDD is a two-loop iterative algorithm, wherein the inner loop solves an augmented Lagrangian problem by the BCD method (i.e., the steps within the while loop in Algorithm 5), and the outer loop updates the dual variable and penalty term until convergence (i.e., $\lambda\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p} + \frac{1}{2\beta}\|\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p}\|_2^2 < \varepsilon$). In Algorithm 5, we omit the outer loop iteration of the PDD framework because for sufficiently small β, the residual SI power can be suppressed below a suitable threshold (e.g. noise floor).*

In contrast, having the outer loop would needlessly suppress the residual SI further (i.e., enforcing (3.16e)*), which not only deteriorates both communications and sensing performance, but also leads to slower convergence.*

Remark 6. *In general, the ISAC TX-RX beamformer design problem can be formulated in several ways; for instance, an alternative to* (3.16a)-(3.16e) *is to maximize the communications performance by setting constraints on radar sensing performance.*

However, an advantage of our formulation is that it permits a closed-form solution for each step of the BCD (i.e., each of the subproblems in Section 3.3.2). On the other hand, incorporating the radar sensing performance as a constraint would have made it difficult to realize this, and would therefore slow down the convergence of our Algorithm. On the other hand, a softer constraint on sensing performance can be imposed by increasing the weights α_3 and α_4, which would have the effect of prioritizing the sensing performance. Note that for simplicity, we set α_3 and α_4 to be the same in order to simultaneously control the transmit and receive beampattern power.

Remark 7. *The joint TX-RX beamformer design can be extended to underloaded multi-user scenarios as well. For example, consider K downlink users and let $\mathbf{p}_k$ denote the downlink precoder for the k^{th} downlink user and $\mathbf{P} := [\mathbf{p}_1, \cdots \mathbf{p}_K] \in \mathbb{C}^{N_t \times K}$. For $K < \min(N_t, N_r)$, the rows of $\mathbf{H}_{\text{si}}\mathbf{P}$ have a non-trivial nullspace and hence, there exists non-zero $\mathbf{w}$ for which $\mathbf{w}^H\mathbf{H}_{\text{si}}\mathbf{P} = 0$.*

3.3.3 Algorithm Implementation

In this thesis, we consider a centralized implementation of Algorithm 5, which involves the overhead of communicating all the channel state information ($\mathbf{H}_{\text{u}}$, $\mathbf{H}_{\text{si}}$, $\mathbf{H}_d$) and the target direction to a controller as well as transferring the optimized $\mathbf{p}$ and $\mathbf{w}$ to the ISAC transceiver, $\boldsymbol{\omega}_{\text{u}}$ to the uplink user and $\boldsymbol{\omega}_{\text{d}}$ to the downlink user. Thus, a decentralized implementation of Algorithm 5 is key for minimizing the overhead. Furthermore, the order in which the variables are optimized can be different, and thus, a different ordering along with possibly different update rates may speed up convergence. These questions are left for future work. In the following section, we discuss the convergence behavior and the complexity of Algorithm 5.

3.4 Convergence and Complexity Analysis

3.4.1 Convergence Analysis

In this section, the convergence of Algorithm 5 to at least a local optimum is proved. Let $f(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}})$ denote the objective function in (3.17a), and $f_{\rho}(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}, \rho_{\mathrm{u}}, \rho_{\mathrm{d}})$ the reformulated objective function in (3.20a), based on the rate-WMMSE relationship explained in Section 3.3.1. Thus, we have

$$f(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}) \geqslant f_{\rho}(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}, \rho_{\mathrm{u}}, \rho_{\mathrm{d}}), \tag{3.45}$$

where ρ_{u}^{*}, and ρ_{d}^{*} are the optimal solutions as shown in (3.26) and (3.27). The inequality condition in (3.45) holds when ρ_{u} satisfies (3.26) and ρ_{d} satisfies (3.27) are both the global optimum with fixed other optimized variables $\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}$. Let $f_{\mathrm{g}}(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}, \rho_{\mathrm{u}}, \rho_{\mathrm{d}})$ denote the objective in (3.23a). Then, we have

$$\begin{aligned} f_{\rho}(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}, \rho_{\mathrm{u}}, \rho_{\mathrm{d}}) \leq\ & f_{\mathrm{g}}(\mathbf{p}, \mathbf{w}, \boldsymbol{\omega}_{\mathrm{u}}, \mathbf{u}_{\mathrm{d}}, \rho_{\mathrm{u}}, \rho_{\mathrm{d}}) \\ & + \alpha_3 V_{\mathrm{t}} P_{\mathrm{d}} + \alpha_4 V_{\mathrm{r}}. \end{aligned} \tag{3.46}$$

Let $\mathbf{p}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n})}, \rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}$ be the n-th iteration of the variables whose expressions are given by (3.43), (3.34), (3.32), (3.37), (3.26) and (3.27), respectively. $\rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}$ are obtained by the n-th iteration beamformers $\{\mathbf{p}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}\}$ and $\{\mathbf{p}^{(\mathrm{n})}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n})}\}$, respectively. Since the power of $\mathbf{p}^{(\mathrm{n})}$ reaches P_{d}, we obtain (3.47) from (3.45) and (3.46), given by

$$\begin{aligned} & f(\mathbf{p}^{(\mathrm{n}+1)}, \mathbf{w}^{(\mathrm{n}+1)}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n}+1)}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n}+1)}) \\ & \geqslant f_{\mathrm{g}}(\mathbf{p}^{(\mathrm{n}+1)}, \mathbf{w}^{(\mathrm{n}+1)}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n}+1)}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n}+1)}, \rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}) \\ & \quad + \alpha_3 V_{\mathrm{t}} P_{\mathrm{d}} + \alpha_4 V_{\mathrm{r}}. \end{aligned} \tag{3.47}$$

Following Algorithm 5, the update sequences are $\boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n}+1)} \overset{\text{(a)}}{\leftarrow} \{\mathbf{w}^{(\mathrm{n})}\}$, $\mathbf{w}^{(\mathrm{n}+1)} \overset{\text{(b)}}{\leftarrow} \{\mathbf{p}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{u}}^{(\mathrm{n})}\}$, $\mathbf{u}_{\mathrm{d}}^{(\mathrm{n}+1)} \overset{\text{(c)}}{\leftarrow} \{\mathbf{p}^{(\mathrm{n})}\}$, and $\mathbf{p}^{(\mathrm{n}+1)} \overset{\text{(d)}}{\leftarrow} \{\rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n}+1)}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n}+1)}\}$. There-

fore, we have

$$\begin{aligned} &f_{\mathrm{g}}(\mathbf{p}^{(\mathrm{n}+1)}, \mathbf{w}^{(\mathrm{n}+1)}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n}+1)}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n}+1)}, \rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}) \\ &\geqslant f_{\mathrm{g}}(\mathbf{p}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n})}, \rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}). \end{aligned} \tag{3.48}$$

Subsequently, by using (3.45) and (3.46) and due to the power of $\mathbf{p}^{(\mathrm{n})}$ reaching P_{d}, we have

$$\begin{aligned} &f_{\mathrm{g}}(\mathbf{p}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n})}, \rho_{\mathrm{u}}^{(\mathrm{n})}, \rho_{\mathrm{d}}^{(\mathrm{n})}) + \alpha_3 V_{\mathrm{t}} P_{\mathrm{d}} + \alpha_4 V_{\mathrm{r}} \\ &= f(\mathbf{p}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n})}). \end{aligned} \tag{3.49}$$

From (3.45) - (3.49), we obtain

$$f(\mathbf{p}^{(\mathrm{n}+1)}, \mathbf{w}^{(\mathrm{n}+1)}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n}+1)}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n}+1)}) \geqslant f(\mathbf{p}^{(\mathrm{n})}, \mathbf{w}^{(\mathrm{n})}, \boldsymbol{\omega}_{\mathrm{u}}^{(\mathrm{n})}, \mathbf{u}_{\mathrm{d}}^{(\mathrm{n})}), \tag{3.50}$$

which shows that the objective function in (3.17a) is non-decreasing after each BCD update iteration. Since the uplink and downlink rates are upper bounded [11], as are the beampattern gains when beamformers $\mathbf{p}$ and $\mathbf{w}$ both probe at target direction, θ_{r}, the value of objective in (3.17a) is bounded above. Consequently, the proposed algorithm will converge to at least a local optimum, which guarantees the convergence of the proposed Algorithm 5. In the following, we provide numerical results to demonstrate the convergence of the proposed design.

Fig. 3.2 shows the behavior of the objective function (3.17a) and the relative difference ζ, defined in (3.51), with respect to the number of iterations, for different N_{t} and SI power levels:

$$\zeta \triangleq \frac{f_{\mathrm{g}}^{(\mathrm{n})} - f_{\mathrm{g}}^{(\mathrm{n}-1)}}{f_{\mathrm{g}}^{(\mathrm{n}-1)}}, \tag{3.51}$$

where $\zeta \leqslant \epsilon$ is the iteration end criterion. As shown in Fig. 3.2, objective (3.23a) converges within a small number of iterations.

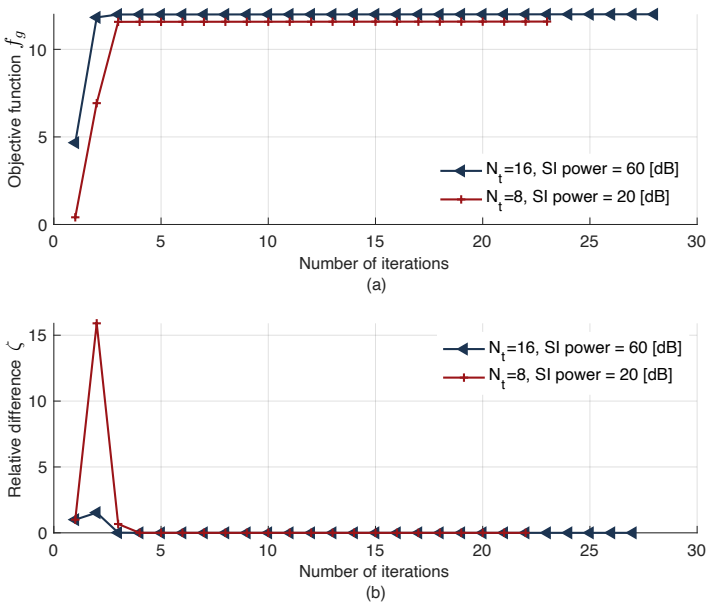


Figure 3.2 Convergence of Algorithm 5 to a stable objective function value with $\epsilon = 10^{-5}$. (a) Objective function versus the number of iterations; (b) Relative difference versus the number of iterations.

3.4.2 Complexity Analysis

From Algorithm 5, the bottleneck step is the step 8, which updates $\mathbf{p}$, with assumption $N_{\mathrm{d}} < N_{\mathrm{t}} \leqslant N_{\mathrm{r}}$, $N_{\mathrm{u}} < N_{\mathrm{t}} \leqslant N_{\mathrm{r}}$. Thus, the overall complexity of Algorithm 5 is given by

$$\mathcal{O}(I_3(N_{\mathrm{r}}^2 + N_{\mathrm{t}}N_{\mathrm{d}} + I_1 N_{\mathrm{r}}N_{\mathrm{u}} + N_{\mathrm{r}}^3 + I_2 N_{\mathrm{t}}^3)) \approx \mathcal{O}(I_3(\max(N_{\mathrm{r}}^3, I_2 N_{\mathrm{t}}^3))), \tag{3.52}$$

where I_3 is the iteration number. In summary, Algorithm 5 has polynomial complexity.

3.5 Numerical evaluation

In this section, we provide numerical results to validate the performance of the proposed joint TX-RX beamformer design for the FD ISAC system. The number of transmit and receive antennas at the FD ISAC transceiver are set the same, which are $N_{\mathrm{t}} = N_{\mathrm{r}} = 16$. The number of antennas at uplink and downlink user are set to $N_{\mathrm{u}} = 2$ and $N_{\mathrm{d}} = 2$, respectively. The carrier frequency of the FD ISAC system is $f_{\mathrm{c}} = 2.4$GHz. The sampling rate is $T_{\mathrm{s}} = 1/\Omega$, where $\Omega = 20$MHz is the bandwidth. We use the tuple (θ, r, v) to represent a target/user's coordinates and bearing w.r.t the transceiver, where θ denotes the direction, r the range and v the velocity. The radar target is assumed to be at $(45°, 7.5\mathrm{m}, 20\mathrm{m/s})$, the uplink user at $(-50°, 10\mathrm{m}, 0\mathrm{m/s})$, and the downlink user at $(\theta_{\mathrm{d}} = -30°, 100\mathrm{m}, 0\mathrm{m/s})$. The transmit powers of the ISAC transceiver and uplink user are $P_{\mathrm{d,[dBm]}} = 20$dBm, and $P_{\mathrm{u,[dBm]}} = 10$dBm, respectively. The thermal noise floor at the ISAC receiver is $P_{\mathrm{noise,[dBm]}} = -94$dBm. The path loss (in dB) is modelled as follows $\eta(d) = -20\log_{10}(\lambda/(4\pi d_0)) + 10n\log_{10}(d/d_0)$, where $d_0 = 1$m is the reference distance, and $n = 2.2$ is the path loss exponent [93], respectively. The SI power, $P_{\mathrm{si,[dBm]}} - P_{\mathrm{noise,[dBm]}}$, ranges from 10 - 60dB at the ISAC receiver. We run 1000 times Monte Carlo simulations with perfect CSI for observing the performances of the SI cancellation, SoI over SI, and sum-rates. The signals, s_{u}, and s_{d}, comprise uncoded i.i.d zero mean, unit-energy QPSK symbols.

3.5.1 SI Power After Cancellation

Firstly, we measure the SI cancellation performance of Algorithm 5 in terms of the residual digital-domain SI power level $P_{\mathrm{res,[dB]}}$, which is defined as $P_{\mathrm{res,[dB]}} := 10\log_{10}(|\mathbf{w}^H\mathbf{H}_{\mathrm{si}}\mathbf{p}|^2/1\mathrm{mW}) - P_{\mathrm{noise,[dBm]}}$. Thus, a negative value of P_{res} implies that the residual SI power level is lower than the noise floor. Additionally, we assume $\alpha_{\mathrm{com}} = \alpha_1 = \alpha_2$ (i.e., equal weight for uplink and downlink communications performance) and $\alpha_{\mathrm{radar}} = \alpha_3 = \alpha_4$. Furthermore, the priority given to communications performance is captured by the parameter, $\varrho = \alpha_{\mathrm{com}}/\alpha_{\mathrm{radar}}$. As shown in Fig. 3.3, when ϱ varies from 1 to 1000, the SI power up to 60 dB

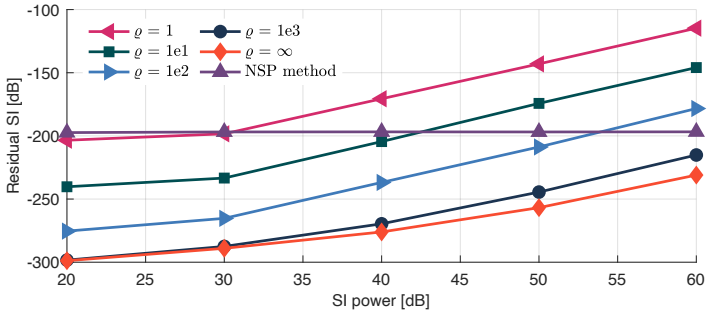


Figure 3.3 Performance comparison of SI cancellation with varying priority parameter ϱ and level of SI. The residual SI level is computed with respect to the noise floor.

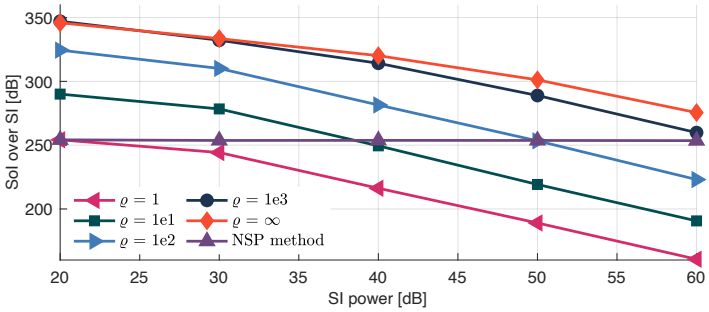


Figure 3.4 Performance comparison of SoI (e.g., radar echo and uplink data) power over SI power in dB.

with regards to noise floor can be effectively suppressed due to the residual SI power, $P_{\text{res,[dB]}} < 0$. To observe whether the SoI is preserved in the process of SI suppression, we consider the ratio between the SoI and SI powers, defined as $P_{\text{SoI/SI,[dB]}} := 10\log_{10}((|\mathbf{w}^H\mathbf{H}_\text{r}\mathbf{p}|^2 + |\mathbf{w}^H\mathbf{H}_\text{u}\boldsymbol{\omega}_\text{u}|^2)/|\mathbf{w}^H\mathbf{H}_\text{si}\mathbf{p}|^2)$. The performance of $P_{\text{SoI/SI,[dB]}}$ with varying ϱ is shown in Fig. 3.4, where a positive value implies that the SoI (e.g., radar echo and uplink data) is not drowned by the strong SI. As expected, the performance of SI cancellation degrades and $P_{\text{SoI/SI,[dB]}}$ decreases with increasing of the SI power. Additionally, when the priority parameter ϱ increases, the performance of SI cancellation improves and SoI over SI power increases. This is because a higher ϱ leads to more priority given to communications, resulting in reduced correlation between SI and SoI. It should be noted that this correlation is one of the challenges in suppressing SI when

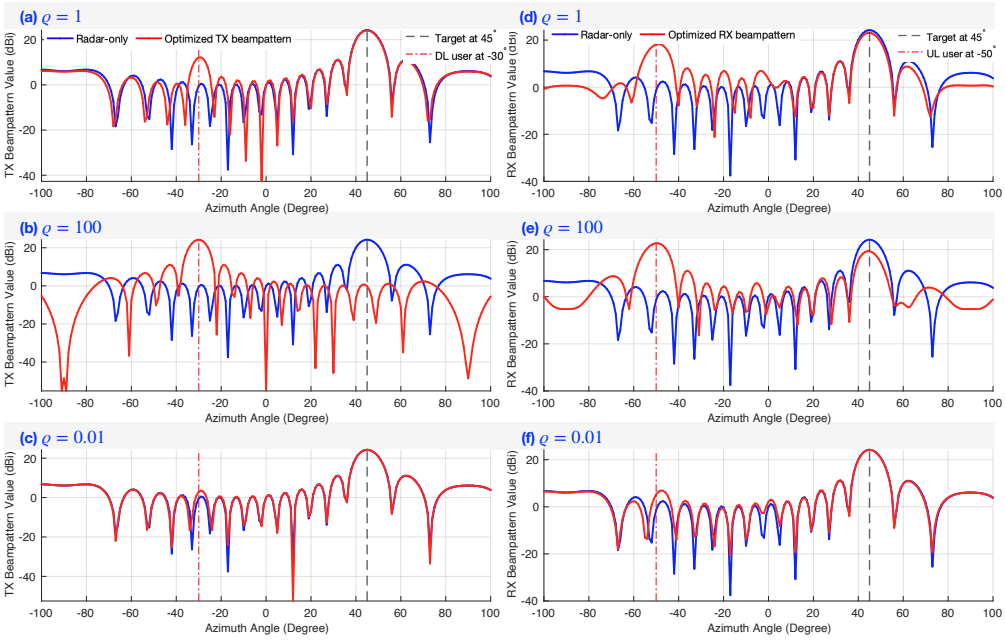


Figure 3.5 Left panel: Transmit beampattern, **Right panel**: Receive beampattern of the ISAC FD system with a effective cancellation of 60 dB SI, where $N_{\mathrm{t}} = 16$, target at $\theta_{\mathrm{r}} = 45°$, and downlink user at $\theta_{\mathrm{d}} = -30°$: (a, d) $\varrho = 1$; (b, e) $\varrho = 100$; (c, f) $\varrho = 0.01$.

the sensing function is present. Thus, the joint ISAC TX-RX beamformers design can effectively preserve the SoI and suppress SI in FD ISAC. While NSP method has better performances than the proposed method with some certain parameter settings, other system capabilities of the proposed method (i.e., average sum-rate of downlink and uplink users, radar parameter estimation performance) are enhanced, which are explained in Section 3.5.3 and Section 3.5.4, respectively.

3.5.2 Beampattern Power Performance

Fig 3.5 illustrate examples of the transmit and receive beampatterns with varying priority parameters ϱ with effective cancellation of 60dB residual SI. As shown in Figs. 3.5a and 3.5d, when $\varrho = 1$, the transmit beamformer focuses

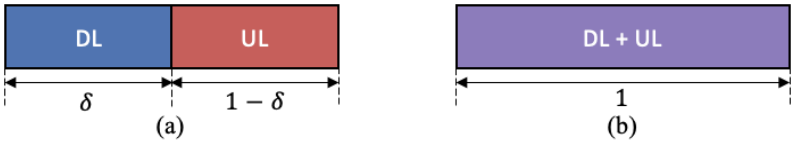


Figure 3.6 Frame structures of (a) half-duplex and (b) FD systems.

the transmit power towards the radar target and the downlink user directions. Meanwhile, the receive beamformer concentrates on the uplink user and radar target. According to Figs. 3.5b and 3.5e, the transmit and receive beams focus more on the downlink user and uplink user compared with the transmit and receive beampatterns of $\varrho = 1$, respectively, due to a higher priority on the communication that is $\varrho = 100$. When $\varrho = 0.01$, the radar function has higher priority, thus the transmit and receive beams concentrate more on the target in comparison with the transmit and receive beampatterns of $\varrho = 1$. Additionally, the residual SI power level $P_{\text{res,[dB]}}$ for the cases that $\varrho = 1, 100, 0.01$ are -130.97dB, -197.37dB, and -108.48dB, respectively. Hence, the SI is effectively suppressed.

3.5.3 Sum-rate Performance

In half-duplex systems, a fraction, δ, of the (time-frequency) resources are allocated for uplink and the remaining for downlink, while for FD, all the resources are allocated simultaneously for both uplink and downlink, as shown in Fig. 3.6. The sum-rate in one unit time-frequency resource for the half-duplex system is defined as $R_{\text{hd,[unit]}} = \delta R_{\text{dl}} + (1-\delta)R_{\text{ul}}$, and sum-rate for the FD system is defined as $R_{\text{fd,[unit]}} = R_{\text{dl}} + R_{\text{ul}}$. In our simulation, δ is set as 0.5, which means equal time-frequency resource allocation for both uplink and downlink.

The sum-rate performance with respect to SI power level with varying priority parameter ϱ is shown in Fig. 3.7. When ϱ increases, the sum-rate increases correspondingly. Specifically, the upper bound is given by the sum-rate of the FD communication functions (i.e., $\varrho = \infty$). The lower bound is given by the sum-rate in one unit time-frequency resource of the half-duplex system. With in-

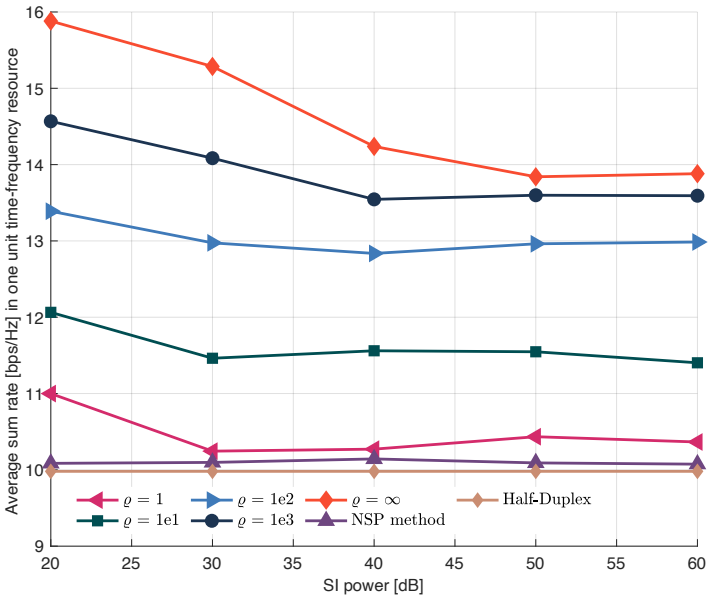


Figure 3.7 Average sum-rate in one unit time-frequency resource versus varying level of SI power.

creasing SI power, the sum-rate in unit time-frequency resource first decreases and then tends to be flat thanks to the effective suppression of the residual SI.

3.5.4 Sensing Performance

For range-Doppler sensing, we explicitly consider the signal stream in (3.4) in discrete time, ignoring the residual SI. The resulting signal can then be represented as

$$\begin{aligned}\hat{s}_u[n,m] &= \mathbf{w}^H \mathbf{H}_{\mathrm{u}} \boldsymbol{\omega}_{\mathrm{u}} s_{\mathrm{u}}[n,m] \\ &\quad + \eta_{\mathrm{r}} e^{j2\pi(f_{\mathrm{d}} T_{\mathrm{s}})m} \mathbf{w}^H \mathbf{a}(\theta_{\mathrm{r}}) \mathbf{b}^\top(\theta_{\mathrm{r}}) \mathbf{p} s_{\mathrm{d}}[n - i_\tau, m] \\ &\quad + \mathbf{w}^H \mathbf{n}_{\mathrm{u}}[n,m], \\ &\quad (n = 0, \cdots, N; m = 0, \cdots, M-1),\end{aligned} \tag{3.53}$$

where the indices m and n respectively capture the slow and fast time-scales commonly assumed in range-Doppler processing[3], and $i_\tau \in \mathbb{Z}$ is the round-

[3]This framework results in a decoupling between the effects of the delay and Doppler shifts in (3.53), which is a reasonable assumption when the Doppler frequency (320Hz in our case) is

trip delay of the radar echo (also known as the target range bin). The signal stream, $s_d[n,m] := s_d[n+mM]$ (and likewise, with $s_u[n,m]$) can be viewed as a concatenation of M blocks, with each block comprising N symbols.

The range bin, i_τ, is estimated from (3.53) by applying the matched filter w.r.t $s_d[n,m]$ along the fast-time axis. In [82], it was shown that the cross-correlation function between $s_u[\cdot,m]$ and $s_d[\cdot,m]$, and the autocorrelation function of $s_{\mathrm{d}}[\cdot,m]$ asymptotically converged to the all-zero function and $\delta[\cdot]$, respectively, for large N. Hence, the uplink data stream has negligible impact on the range-Doppler sensing performance. Subsequently, we perform an M-point DFT operation, resulting in the range-Doppler map.

With respect to angular domain sensing, since we implicitly assume that the target is in the vicinity of the direction in which $\mathbf{w}$ is "pointing" towards (i.e., the radar operates in *track mode*), we are more interested in the interference suppression capability of the latter from an unwanted direction, θ, captured by the following metric:

$$P(\theta) = |\mathbf{w}^H(\mathbf{a}(\theta) + \mathbf{H}_{\mathrm{si}}\mathbf{p})|^2. \tag{3.54}$$

Consequently, the AoA estimate is given by $\arg\max_\theta P(\theta)$.

For comparison, we consider the NSP method, radar-only method, and communication only method. Specifically, the TX and RX beamformers for the radar-only method are given by $\mathbf{p}_{\mathrm{r}} = \mathbf{b}(\theta_{\mathrm{r}})$, and $\mathbf{w}_{\mathrm{r}} = \mathbf{a}(\theta_{\mathrm{r}})$, respectively. Similarly, the TX and RX beamformers for the communication-only method are given by $\mathbf{p}_{\mathrm{u}} = \mathbf{b}(\theta_{\mathrm{u}})$, and $\mathbf{w}_{\mathrm{u}} = \mathbf{a}(\theta_{\mathrm{u}})$, respectively. The priority parameter ϱ is set to 1. The range-velocity map is shown in Fig. 3.8. We observe that Algorithm 5 has effective residual SI suppression compared with methods that do not consider SI suppression (i.e., the radar-only and communications-only methods), where the desired radar echo drowns in the SI, which can be seen as a strong signal with $v = 0\mathrm{m/s}, r = 0\mathrm{m}$. Additionally, the radar-only and communication-only methods both have interference around the original point

much smaller than the signal bandwidth (20MHz in our case).

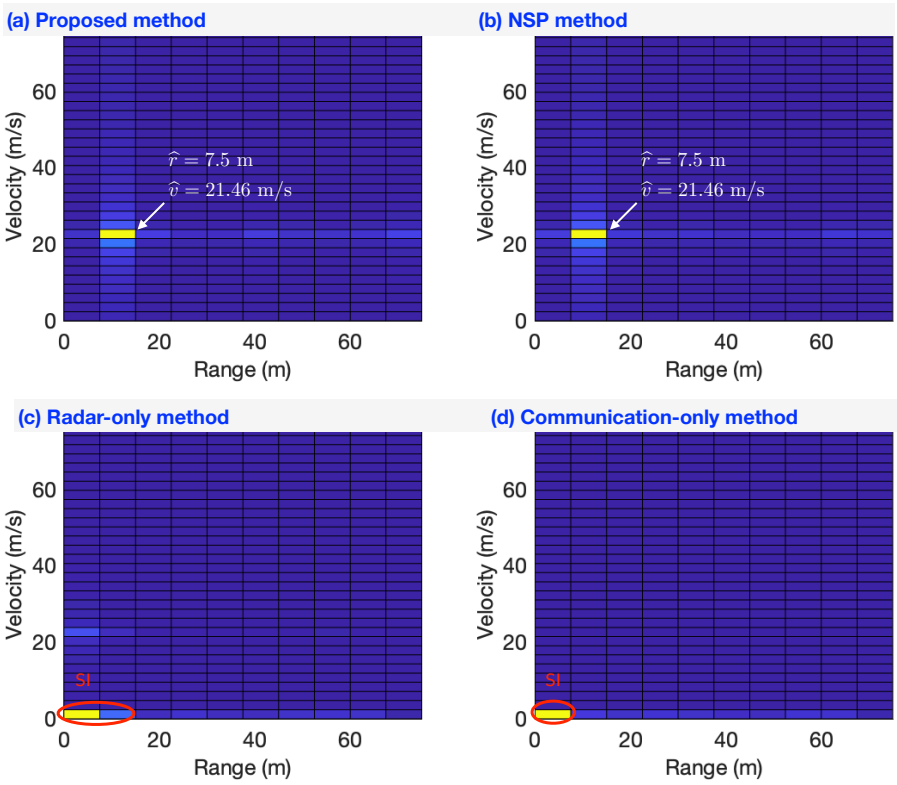


Figure 3.8 Examples of range and velocity estimation based on the FD ISAC system with 60 dB SI with $N_{\mathrm{t}} = 16, N_{\mathrm{r}} = 16$, $M = 512$ blocks, and $N = 1024$ symbols.

due to the unsuppressed residual SI.

Fig. 3.9 contains plots of the output power, $P(\theta)$, of different methods. In the direction of target ($45°$), our method and the radar-only method have similar output power. This indicates similar angle detection results of the two methods. The proposed method also has a similar output power compared with the communication-only method. Additionally, the proposed method has smaller interference compared with the NSP method, especially around the angle $0°$, where the SI is located.

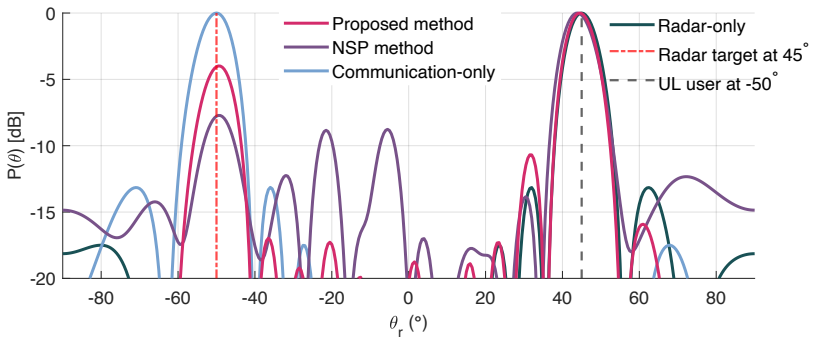


Figure 3.9 An example of angle estimation for a simulated scenario with a target at $\theta_{\mathrm{r}} = 45°$, a uplink communication user at $\theta_{\mathrm{u}} = -50°$ with antennas $N_{\mathrm{t}} = 16$ and 60 dB SI power.

3.6 Summary

In this thesis, we designed the transmit and receive beamformer $\mathbf{p}, \mathbf{w}$ at the transceiver, precoder $\boldsymbol{\omega}_{\mathrm{u}}$ at the uplink user, combiner $\mathbf{u}_{\mathrm{d}}$ at the downlink user to simultaneously maximize the uplink and downlink rate, the transmit and receive radar beampattern power at the target, and suppress the residual SI. In the objective function, the TX and RX beampattern gains are used as the radar metric, and the uplink and downlink rates are used as the communication metric. With the aid of the equivalence of the rate maximization and the MSE minimization, and penalty-based transformation, we used the BCD method to solve the optimization problem. Subsequently, we gave a convergence analysis. Numerical results showed that up to 60 dB residual SI in digital domain can be efficiently suppressed. Additionally, the optimized TX and RX beampatterns can probe at the desired target, uplink and downlink user directions with a satisfactory average sum-rate, a more accurate radar parameter estimation with regards to range, velocity, and angle, and outperforms the NSP beamformer design method, which validates the effectiveness of the proposed algorithm. For future studies, a more effective algorithm on suppressing SI is worthy investigating. In addition, an extended transceiver design for a more general multi-data transmission can be studied. Moreover, it would be interesting to extend the SI cancellation technique to both the digital and analog domains in view of more accurate residual SI model and saturation caused by analog SI.

Chapter 4

Rate-Splitting Multiple Access for Quantized ISAC LEO Satellite Systems: A Max-Min Fair Energy-Efficient Beam Design

4.1 Introduction

Sensing capabilities are recognized as a key technique in the next generation of wireless communication networks (*i.e.* 6G) for achieving emerging applications, such as assisted navigation, activity detection and movement tracking, and environmental monitoring [5, 94]. Furthermore, given the significant development in the wireless communications and radar sensing industries, spectrum resources are becoming increasingly scarce and valuable. As a promising solution, ISAC has been proposed to meet these requirements. By adopting the ISAC system, spectral, hardware, and energy resources are efficiently utilized. This efficiency is achieved through unified signal processing procedures and a shared hardware framework between sensing and communication systems. Nonetheless, much of the past ISAC research has been primarily devoted to terrestrial networks, resulting in limitations in providing global services [9].

On the other hand, satellite communications have been envisioned as a solution to satisfy the growing demand for ubiquitous and high-capacity global connectivity in unserved (*e.g.* marine, and desert) and underserved areas (*e.g.* rural areas). The deployment of satellites and terrestrial stations offers global coverage; thus, it enhances the service availability in such areas and enables service reliability through the redundancy in the satellite-terrestrial networks. In comparison to GEO and MEO satellites, which suffer from high propagation delay and require large investments, LEO satellites offer a lower latency, higher data rate, and near-complete visibility due to their relatively lower orbit altitudes (*i.e.* $500-2000$ km). Furthermore, the proximity of LEO satellites to Earth has prompted the exploration of remote sensing functionality in LEO satellite systems. In recent years, the necessity for sensing capabilities and enhanced spectrum utilization efficiency in LEO satellites has naturally driven the development of ISAC-LEO systems. However, the implementation of ISAC in LEO satellite systems poses certain challenges, potentially limiting the technological promise of ISAC-LEO systems, as follows.

Challenge #1: Interference Control. The first challenge is the inter-beam interference among the users and the control of the trade-off between communica-

tion and sensing capabilities, as these factors determine the overall system performance. RSMA, recognized as a generalized and powerful non-orthogonal transmission framework and robust interference management strategy for multi-antenna multi-user networks, has been demonstrated to be effective in handling various interference challenges in satellite communication [9, 49–52, 54, 95]. In RSMA, the transmitter splits each message into common and private messages, which are then linearly precoded at the satellite gateway. At the receiver, the common stream is first decoded by treating all private streams as noise. Subsequently, each user eliminates the decoded common stream and decodes their respective private stream, treating the other private streams as noise. Due to the ability to partially decode the interference and partially treat it as noise, RSMA outperforms the existing multiple access schemes, *e.g.* SDMA, OMA, and non-orthogonal multiple access NOMA, in terms of spectral efficiency and user fairness [53]. Furthermore, RSMA has been shown to facilitate a better trade-off between the communication and sensing functionalities, when compared to conventional beamforming techniques [41–44].

Challenge #2: Energy Efficiency. The second challenge pertains to the power limitations typically faced by LEO satellites due to the lack of its suitable power sources. This issue becomes notably more severe, especially as the number of transmit antennas increases for the improved spectral efficiency and spectrum reuse factor. Consequently, the satellite power needs to be used with high efficiency. When considering the downlink RF chain, it has been shown that the DAC dominates the total power consumption of the RF chain [96]. Specifically, the power consumption of DAC grows exponentially with the number of bits b that determine the resolution of the DAC. By leveraging this property, we can employ low-resolution DACs, thereby reducing the power consumption in each RF chain and the total power consumed at the satellite.

Furthermore, considering the expected increase in the number of connected devices and data rates in future satellite communication systems [34], coupled with the limitation of satellite power, it is crucial to adopt energy EE as a significant performance indicator in transmission scheme design. The EE is defined

Table 4.1 Novelty Comparison with Existing RSMA-ISAC/Satellite Literature

	Our work	[41,45]	[42]	[43]	[44]	[46]	[100–102]	[103]	[104]	[105]	[106]
ISAC	✓	✓	✓	✓	✓	✓					
Satellite	✓		✓				✓	✓	✓	✓	✓
Rate metric		✓		✓		✓	✓	✓	✓	✓	✓
EE metric					✓						
Max-min EE metric	✓										
Beampattern matching		✓			✓						
CRB	✓		✓	✓		✓					
Beamforming design	✓	✓	✓		✓	✓	✓				✓
Power allocation								✓	✓	✓	
Low-resolution DAC	✓				✓						
Radar sequence	✓	✓									

as the ratio between the transmission rate and the consumed energy, measured in bits/Joule. From a system-level perspective, it is preferable to consider a total EE, which is defined as the sum-rate of the communication system over the total power consumption [97]. However, this metric tends to prioritize the EE in a few links with favorable channel conditions. This prioritization potentially neglects the needs of others. In essence, the total EE metric lacks fairness among all links, especially when dealing with a substantial number of users. Therefore, it is more reasonable to aim for maximizing the minimum EE among all links to ensure fairness.

Related Works. Existing ISAC studies primarily concentrate on the terrestrial networks [9,98], exploring applications such as autonomous vehicles, environment monitoring, smart cities, smart manufacturing, and industrial IoT, etc [7]. Nevertheless, terrestrial ISAC systems fail to provide global service, and are constrained by limited data reception and processing resources. To address these challenges, the concept of an ISAC-LEO satellite system has naturally emerged, driven by the development of on-board processing capacities on the satellite. A hybrid beamforming design for ISAC-LEO systems are proposed in [99], taking into account the effects of beam squint. This issue is formulated as an EE optimization, subject to the constraint of matching the beampattern. It has been demonstrated that a trade-off can be achieved between communication EE and the sensing beampattern.

RSMA is an effective interference management strategy, resulting in enhanced spectral efficiency and energy efficiency. Therefore, RSMA has been utilized in satellite systems [100–106] to improve communication performances. In detail, RSMA for multigroup multicast beamforming is investigated in (multigateway) multibeam satellite system [100–102] to improve the rate performance. A deep reinforcement learning method is proposed in [103] to obtain the optimal power allocation to maximize the sum rate in satellite system. In [104], the power allocation over different beams in satellite system is optimized by maximizing the sum rate. A comprehensive comparison of sum rate performance among RSMA, SDMA, and OMA in satellite systems is provided in [105]. Additionally, the secure delivery issue in investigated in [106].

RSMA is also proposed in the ISAC design to enhance the management of inter-beam interference among communication users, and achieve a satisfactory trade-off between communication and sensing functionalities. Starting from the RSMA-assisted ISAC system without the satellite set-up, the concept of integrating RSMA with ISAC systems is initially introduced in [41]. Building upon this, [45] delves into the study of RSMA-assisted ISAC systems, considering practical scenarios that involve partial channel state information at the transmitter (CSIT) and the mobility of communication users. This investigation revealed that RSMA exhibited superior interference management capabilities, thereby improving the trade-off between the weighted sum rate (WSR) and mean square error (MSE) in beamforming approximation, when compared to conventional strategies like SDMA and NOMA. In [44], a design for RSMA-assisted ISAC utilizing low-resolution DACs is investigated. This design shows the potential of RSMA to enhance EE with a reduced number of RF chains.

The sensing performance in the above research is optimized by the beampattern matching constraint, which solely captures the transmitter design. In contrast, Cramér–Rao Bound (CRB) is a theoretical lower bound on the variance of unbiased estimators, resulting in a focus on the optimization of estimation performance [107]. Using the CRB metric, RSMA-assisted ISAC-LEO system has been proposed recently [42, 43]. RSMA has been proven to effectively manage different types of interference and enable a better trade-off between communi-

cation and sensing in ISAC-LEO systems compared to SDMA-assisted methods. Though the extension of the multi-target case is recently investigated in [46], it does not delve into aspects such as the EE fairness among users, low-resolution DACs, and the unique features of the satellite scenario. A novelty comparison table (cf. Table 4.1) on existing RSMA-ISAC/Satellite is provided. In summary, previous research has not rigorously investigated the max-min fairness EE problem and interference control in the quantized ISAC-LEO satellite system, which motivates this research.

To conclude, the main contribution is that we propose an approach to effectively control the interference through RSMA. On the other hand, to address the limited power supply at the LEO satellites, we consider low-resolution quantization DACs at the transmitter and the max-min fairness EE metric. Specifically, we optimize the precoders for maximizing the minimal EE among all users, considering the power consumption of each RF chain under communication and sensing constraints. To tackle this optimization problem, we proposed an iterative algorithm based on SCA and Dinkelbach's method. Numerical results illustrate that the proposed design outperforms the strategies that aim to maximize the total EE of the system and SDMA in terms of max-min fairness EE and the communication-sensing trade-off.

4.2 System model

As depicted in Fig. 4.1, we consider a downlink ISAC-LEO satellite system equipped with co-located N_t transmit antennas and N_r receive antennas (*i.e.* mono-static setup). This system operates in full-duplex mode to prevent the echo-miss problem, which arises from signal transmission durations being considerably longer than the radar echo round-trip time [108]. This system serves K single-antenna users, indexed by $\mathcal{K} = \{1, \cdots, K\}$, and detects a target. We assume that all antenna arrays are uniform linear arrays (ULA) [109], featuring half-wavelength spacing between adjacent antenna elements. As shown in Fig. 4.2, since RSMA is adopted, the message W_k for the k^{th} user is categorized into a common message $W_{c,k}$ and a private message $W_{p,k}$. Subsequently, all common messages, $\{W_{c,1}, \cdots, W_{c,K}\}$, are jointly encoded into a common stream s_c. All the private parts, $\{W_{p,1}, \cdots, W_{p,K}\}$, are encoded into private streams, $\{s_1, \cdots, s_K\}$. The transmit streams comprise common stream, private streams, and radar sequence are expressed as $\mathbf{s}[\ell] = [s_c[\ell], s_1[\ell], \cdots, s_K[\ell], s_r[\ell]]^\top \in \mathbb{C}^{(K+2)\times 1}$, where $\ell \in \mathcal{L} = \{1, \cdots, L\}$ is the discrete time index. Specifically, the radar sequence is deterministic and known at the receiver for target information extraction.

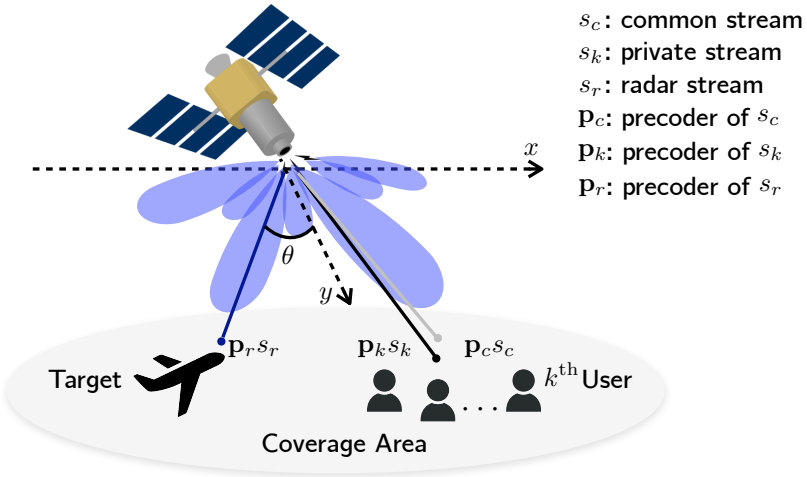


Figure 4.1 Integrated sensing and communications enabled quantized rate-splitting multiple access satellite system.

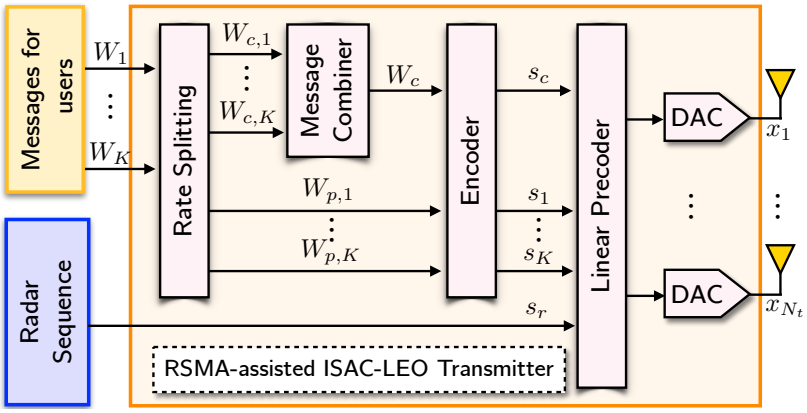


Figure 4.2 Transmitter Architecture of RSMA-assisted ISAC-LEO system.

4.2.1 Channel Model

Satellite Communication Channel Model. The satellite channel comprises free space loss, radiation pattern and rain attenuation, thus the downlink channel between the LEO ISAC satellite and the k^{th} user is modelled as [49, 110]

$$\mathbf{h}_k = \mathbf{b}_k \circ \mathbf{q}_k \circ \exp\{\jmath \boldsymbol{\phi}_k\}, \tag{4.1}$$

where $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$ is the downlink channel between the k^{th} user and the LEO satellite based on the satellite setup [111], $\mathbf{b}_k = [b_{k,1}, b_{k,2}, \cdots, b_{k,N_t}] \in \mathbb{C}^{N_t}$ consists of satellite beam radiation pattern and free space loss. The n_t^{th} element of the $\mathbf{b}_k$ is approximated as

$$b_{k,n_t} \simeq \frac{\sqrt{G_{\mathsf{u}} G_{k,n_t}}}{4\pi \frac{d_k}{\lambda_{\mathsf{c}}} \sqrt{\iota T_{\mathsf{sys}} B}}, \tag{4.2}$$

where G_{u} is the antenna gain at the user, d_k is the distance between the satellite and the k^{th} user, λ_{c} is the carrier wavelength, ι is the Boltzman's constant, T_{sys} is the receive noise temperature, B is the bandwidth. G_{k,n_t} is the beam gain from

the n_t^{th} feed to k^{th} user, and is approximated by

$$G_{k,n_t} = G_{\max}\left[\frac{J_1\left(u_{k,n_t}\right)}{2u_{k,n_t}} + 36\frac{J_3\left(u_{k,n_t}\right)}{u_{k,n_t}^3}\right]^2, \tag{4.3}$$

where $G_{\max}$ denotes the maximum beam gain for each beam and $u_{k,n_t} = 2.07123\sin\left(\theta_{k,n_t}\right)/\sin\left(\theta_{3\text{ dB}}\right)$. θ_{k,n_t} is the angle between the k^{th} user and the center of n_t^{th} beam, and $\theta_{3\text{ dB}}$ is the 3 dB angle loss compared with the gain of the beam center. J_1 and J_3 are the first and third order of the first-kind Bessel function, respectively. Moreover, $\mathbf{q}_k = [q_{k,1}, q_{k,2}, \cdots, q_{k,N_t}] \in \mathbb{C}^{N_t}$ is the rain attenuation coefficient vector, where each element $q_{k,n_t} = \zeta_{k,n_t}^{1/2}$. The power gain ζ_{k,n_t} in dB, $\zeta_{k,n_t}(\text{dB}) = 20\log_{10}(\zeta_{k,n_t})$, commonly follows the lognormal distribution, *i.e.* $\ln\left(\zeta_{k,n_t}(\text{dB})\right) \sim \mathcal{N}(\mu,\sigma)$. Additionally, $\boldsymbol{\phi}_k = [\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,N_t}] \in \mathbb{C}^{N_t}$ is the uniformly distributed phase vector, where $\phi_{k,n_t} \sim \mathcal{U}(0, 2\pi)$. We denote the satellite channels between the satellite and all the users by $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K] \in \mathbb{C}^{N_t \times K}$.

Radar Channel Model. We consider a radar operating in a track model, with an initial AoA estimate for the target obtained from prior scanning. By processing the received reflected echo, we can extract the information of the target, *i.e.* target's AoA, θ, range, r, and the relative velocity, v, between the LEO satellite and target in the line-of-sight (LOS) direction. Since the AoA and AoD are the same for the monostatic radar, they are both denoted as θ. We assume that the radar channel varies slowly with time, thus it is modeled by [108]

$$\mathbf{H}_r(t) = \alpha e^{\jmath 2\pi f_d \ell T}\mathbf{b}(\theta)\mathbf{a}^\top(\theta), \tag{4.4}$$

– α is the attenuation factor, comprising the radar cross-section (RCS) and round-trip path-loss; ν is derived from the radar range equation [84], given in (4.5); r is the distance between the receiver and the target:

$$\alpha = \sqrt{\frac{\lambda_c^2\nu}{(4\pi)^3 r^4}}. \tag{4.5}$$

– ℓ is the discrete time index; T is the sampling interval; $t = \ell T$ is the time instant; $f_d = 2vf_c/c$ is the Doppler shift; the signal carrier frequency and the speed of light are denoted by f_c and c, respectively.

– Transmit and receive steering vectors of the ULA, $\mathbf{a}(\theta) \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{b}(\theta) \in \mathbb{C}^{N_r \times 1}$, are respectively written as

$$\mathbf{a}(\theta) = \left[1, e^{\jmath 2\pi\omega \sin(\theta)}, \ldots, e^{\jmath 2\pi(N_t-1)\omega \sin(\theta)}\right]^\top, \tag{4.6a}$$

$$\mathbf{b}(\theta) = \left[1, e^{\jmath 2\pi\omega \sin(\theta)}, \ldots, e^{\jmath 2\pi(N_r-1)\omega \sin(\theta)}\right]^\top, \tag{4.6b}$$

where ω represents the normalized spacing between adjacent array elements in relation to the wavelength.

4.2.2 Signal Model

The discrete baseband dual-functional signal at time index $\ell \in \mathcal{L}$ is as follows

$$\mathbf{x}[\ell] = \mathbf{P}\mathbf{s}[\ell] = \underbrace{\overbrace{\mathbf{p}_c s_c[\ell]}^{\text{Common stream}} + \overbrace{\sum_{k\in\mathcal{K}} \mathbf{p}_k s_k[\ell]}^{\text{Private streams}}}_{\text{Communication streams}} + \underbrace{\mathbf{p}_r s_r[\ell]}_{\text{Radar sequence}}, \tag{4.7}$$

where $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots, \mathbf{p}_K, \mathbf{p}_r] \in \mathbb{C}^{N_t \times (K+2)}$ is the dual-functional beamforming matrix to be optimized. Specifically, $\mathbf{p}_c \in \mathbb{C}^{N_t}$ is the precoder of the common stream, $\{\mathbf{p}_k\}_{k\in\mathcal{K}}$ is the precoder of the private stream, and $\mathbf{p}_r \in \mathbb{C}^{N_t}$ is the precoder of the radar sensing sequence. Note that the radar sequence, s_r, is deterministic for extracting the target's information, and it has no useful data for the users.

Before the transmission, the baseband signal, $\mathbf{x}[\ell]$, is quantized by the DAC. The quantization is a non-linear mapping, thus the additive quantization noise model (AQNM) [44, 112, 113] is adopted to approximate the quantization to a linear form. Therefore, the analog transmit signal at time ℓ after quantization is given by

$$\mathscr{Q}(\mathbf{x}) \approx \mathbf{x_q}[\ell] = \mathbf{\Delta}\mathbf{x}[\ell] + \boldsymbol{\epsilon}[\ell], \tag{4.8}$$

where $\mathbf{x}_{\mathsf{q}}[\ell] = \mathbf{x}_{\mathsf{q},c}[\ell] + \sum_{k\in\mathcal{K}} \mathbf{x}_{\mathsf{q},k}[\ell] + \mathbf{x}_{\mathsf{q},r}[\ell] \in \mathbb{R}^{N_t\times 1}$. The matrix $\mathbf{\Delta} \triangleq \text{diag}(\delta_1, \cdots, \delta_{N_t}) \in \mathbb{R}^{N_t\times N_t}$, where $\delta_m \in (0,1)$ is the linear gain of the quantization at the m^{th} transmit RF chain when adopting b_m bit DAC quantization, written as

$$\delta_m = 1 - \varrho = 1 - \frac{\pi\sqrt{3}}{2} 2^{-2b_m}, \tag{4.9}$$

where ϱ is the quantization loss. Additionally, $\boldsymbol{\epsilon}$ is the additive quantization noise vector, $\boldsymbol{\epsilon} \sim \mathcal{CN}(\mathbf{0}_{N_t}, \mathbf{\Sigma})$, where

$$\mathbf{\Sigma} = \mathbf{\Delta}(\mathbf{I}_{N_t} - \mathbf{\Delta})\mathscr{E}\big[\mathbf{x}[\ell]\mathbf{x}^H[\ell]\big]. \tag{4.10}$$

The coherent matrix of transmitted dual-functional waveform after quantization is as follows

$$\mathbf{R}_{\mathsf{q}} = \mathscr{E}\big[\mathbf{x}_{\mathsf{q}}[\ell]\mathbf{x}_{\mathsf{q}}^H[\ell]\big] = \mathbf{\Delta}\mathbf{P}\mathbf{P}^H\mathbf{\Delta}^H. \tag{4.11}$$

The received signal at all K users is given by

$$\mathbf{y}[\ell] = \mathbf{H}^H\mathbf{x}_{\mathsf{q}}[\ell] + \mathbf{n}[\ell], \tag{4.12}$$

where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_K, \sigma_k^2\mathbf{I}_K)$ is the additive complex Gaussian noise at each user. We assume that the CSI is perfectly known at the satellite. At the k^{th} user, the received signal is written as

$$\begin{aligned} y_k[\ell] &= \mathbf{h}_k^H\mathbf{x}_{\mathsf{q}}[\ell] + n_k[\ell] \\ &= \mathbf{h}_k^H\mathbf{\Delta}\mathbf{p}_c s_c[\ell] + \mathbf{h}_k^H\mathbf{\Delta}\sum_{k\in\mathcal{K}}\mathbf{p}_k s_k[\ell] \\ &\quad + \mathbf{h}_k^H\mathbf{\Delta}\mathbf{p}_r s_r[\ell] + \mathbf{h}_k^H\boldsymbol{\epsilon}[\ell] + n_k[\ell]. \end{aligned} \tag{4.13}$$

At ISAC-LEO satellite receiver, the radar echo, $\mathbf{z}[\ell] \in \mathbb{C}^{N_r\times 1}$, for the sensing functionality is given by

$$\begin{aligned} \mathbf{z}[\ell] &= \mathbf{H}_r\mathbf{x}_{\mathsf{q}}[\ell] + \mathbf{n}_r[\ell] \\ &= \alpha e^{\jmath 2\pi f_d \ell T}\mathbf{b}(\theta)\mathbf{a}^\top(\theta)\mathbf{\Delta}\mathbf{x}[\ell] \\ &\quad + \alpha e^{\jmath 2\pi f_d \ell T}\mathbf{b}(\theta)\mathbf{a}^\top(\theta)\boldsymbol{\epsilon}[\ell] + \mathbf{n}_r[\ell], \end{aligned} \tag{4.14}$$

where $\mathbf{n}_r[\ell]$ is the independent, zero-mean complex Gaussian noise, and $\mathbf{n}_r[\ell] \sim \mathcal{CN}(\mathbf{0}_{N_r}, \sigma_r^2 \mathbf{I}_{N_r})$.

4.3 Minimal Energy Efficiency Maximization and Problem Formulation

In this section, the metrics in the objective function and constraints for an optimization problem of interest are explained. Additionally, we formulate the optimization problem.

4.3.1 Communication Metric

Communication Rates. Considering the EE of each user, we adopt the max-min fairness EE. Since RSMA is adopted, the communication metric needs to be modified due to the common stream. Recalling the decoding process of RSMA [53], each user first decodes the common stream, s_c, by treating all private streams as noise. After the common stream is successfully decoded, it is subtracted from the received signal, $\mathbf{y}$, by SIC. In RSMA, the interference is partially subtracted and the remaining interference is treated as noise, resulting in significant improvement of sum-rate [114] or minimum rate among all users [115]. This is facilitated by the common and private message split and SIC in the decoding.

The signal-to-interference-plus-noise (SINRs) for the common and private streams at k^{th} user with the assumption that channels vary slowly with time are respectively written as

$$\gamma_{c,k} = \frac{\left|\mathbf{h}_k^H \boldsymbol{\Delta}\mathbf{p}_c\right|^2}{\sum_{i\in\mathcal{K}} \left|\mathbf{h}_k^H \boldsymbol{\Delta}\mathbf{p}_i\right|^2 + \psi \left|\mathbf{h}_k^H \boldsymbol{\Delta}\mathbf{p}_r\right|^2 + \mathbf{h}_k^H \boldsymbol{\Sigma}\mathbf{h}_k + \sigma_k^2}, \tag{4.15a}$$

$$\gamma_k = \frac{\left|\mathbf{h}_k^H \boldsymbol{\Delta}\mathbf{p}_k\right|^2}{\sum_{i\in\mathcal{K}, i\neq k} \left|\mathbf{h}_k^H \boldsymbol{\Delta}\mathbf{p}_i\right|^2 + \psi \left|\mathbf{h}_k^H \boldsymbol{\Delta}\mathbf{p}_r\right|^2 + \mathbf{h}_k^H \boldsymbol{\Sigma}\mathbf{h}_k + \sigma_k^2}, \tag{4.15b}$$

where ψ is an index variable to unify the different processing methods for radar sequence, s_r. Specifically, we consider:

– $\psi = 0$: Since the radar sequence, s_r, is commonly designed and determined before transmission, the user side can access to the radar sequence and im-

plement SIC to remove this interference. This case is hence denoted by $\psi = 0$.

– $\psi = 1$: SIC at the user receiver can be computationally intensive due to the iterative interference cancellation, thus to facilitate low computational and hardware complexity at the user side, we can adopt no SIC at the expense of treating radar sequence, s_r, as the interference.

The normalized rates of the common and private streams with regard to SINR are written as

$$R_{c,k} = \log_2 (1 + \gamma_{c,k}), \tag{4.16a}$$

$$R_k = \log_2 (1 + \gamma_k). \tag{4.16b}$$

We recall that the common stream needs to be decodable by all receivers in the RSMA framework. To guarantee this, the common rate, denoted by R_c, should set to the minimum of the $R_{c,k}, \forall k \in \mathcal{K}$ [53], *i.e.*

$$R_c = \min_{k \in \mathcal{K}} \{R_{c,k}\} = \sum_{k \in \mathcal{K}} C_k, \tag{4.17}$$

where C_k is the rate of the common part of the k^{th} user. Thus, the total achievable rate of the k^{th} user is $R_{k,\mathsf{total}} = C_k + R_k$.

Max-min Energy Efficiency. Communication rates are determined on the K user end, making them related to the transmission design. Thus, we consider the power consumption of the transmitter as the denominator in the max-min EE metric. In each RF chain, the power consumption comes from the low-pass filter, mixer, local oscillator, 90° hybrid with buffer, DAC, whose power consumption are denoted by $P_{\mathsf{LP}}, P_{\mathsf{M}}, P_{\mathsf{LO}}, P_{\mathsf{H}}, P_{\mathsf{DAC}}$, respectively. Specifically, the power consumption of one RF chain is $P_{\mathsf{RF}} = 2P_{\mathsf{DAC}} + 2P_{\mathsf{LP}} + 2P_{\mathsf{M}} + P_{\mathsf{H}} + P_{\mathsf{LO}}/N_t$, where all transmit RF chains share one local oscillator, thus P_{LO} is divided by N_t for simplicity while we assume that all RF chains adopt the same quantization (*i.e.* $b_m = b, \forall m \in \mathcal{N}_t = \{1, \cdots, N_t\}$). The DAC power consumption P_{DAC} in Watt is as follows [116, 117]

$$P_{\mathsf{DAC}}(b, f_s) = 1.5 \times 10^{-5} \cdot 2^b + 9 \times 10^{-12} \cdot f_s \cdot b, \tag{4.18}$$

where f_s is the sampling rate. Subsequently, the total power consumption for each user is written as [116]

$$P_{\mathsf{chain},k} = P_{\mathsf{RF}} + \kappa^{-1} P_{\mathsf{TX},k}, \tag{4.19}$$

where κ is the power amplifier (PA) efficiency (*i.e.* $\kappa = P_{\mathsf{TX},k}/P_{\mathsf{PA}}$), $P_{\mathsf{TX},k}$ is the actual transmit power for the k^{th} user, given by

$$P_{\mathsf{TX},k} = \mathrm{tr}(\mathscr{E}[\mathbf{x}_{\mathsf{q},k}\mathbf{x}_{\mathsf{q},k}^H]) = \|\boldsymbol{\Delta}\mathbf{p}_k\|^2 + 1/K\|\boldsymbol{\Delta}\mathbf{p}_c\|^2, \tag{4.20}$$

where we consider the power of the common stream for each user by dividing user number K. Therefore, the EE of the k^{th} user in the considered system is as follows:

$$\eta_k = \frac{R_{k,\mathsf{total}}}{P_{\mathsf{chain},k}} = \frac{C_k + R_k}{P_{\mathsf{RF}} + \kappa^{-1}(\|\boldsymbol{\Delta}\mathbf{p}_k\|^2 + 1/K\|\boldsymbol{\Delta}\mathbf{p}_c\|^2)}. \tag{4.21}$$

4.3.2 Sensing Metric: Cramér–Rao Bound

For a target sensing, the target information extraction can be formed as a classic maximum likelihood estimation (MLE) problem [107]. After the estimated parameters are obtained, the mean square error (MSE) is used as the sensing performance metric. However, the MSE has no closed-form expression, thus intractable. Instead, the Cramér–Rao Bound is employed as the performance metric, which provides a lower bound for the MSE and has a closed-form expression. The CRB is calculated as $\mathsf{CRB} = \mathbf{J}^{-1}$, where $\mathbf{J}$ is the Fisher information matrix (FIM). The FIM [107] for estimating the unknown parameter vector $\boldsymbol{\xi} = [\theta, \alpha]^\top$ from (4.14), is written as

$$[\mathbf{J}]_{i,j} = \frac{2}{\sigma_r^2}\,\mathrm{Re}\left\{\sum_{\ell=1}^{L} \frac{\partial \boldsymbol{\mu}[\ell]^H}{\partial \xi_i}\frac{\partial \boldsymbol{\mu}[\ell]}{\partial \xi_j}\right\}, i,j \in \{1,2\}, \tag{4.22}$$

where $\boldsymbol{\mu} \triangleq \alpha e^{\jmath 2\pi f_d \ell T}\mathbf{b}(\theta)\mathbf{a}^\top(\theta)\boldsymbol{\Delta}\mathbf{x}[\ell]$. Therefore, the FIM for estimation is partitioned as

$$\mathbf{J}_{\boldsymbol{\xi}} = \begin{bmatrix} \mathbf{J}_{\theta\theta} & \mathbf{J}_{\theta\alpha} \\ \mathbf{J}_{\theta\alpha}^\top & \mathbf{J}_{\alpha\alpha} \end{bmatrix}. \tag{4.23}$$

$$\begin{aligned}\mathsf{CRB}(\theta) &= \left[J_{\theta\theta} - J_{\theta\alpha}J_{\alpha\alpha}^{-1}J_{\theta\alpha}^{\top}\right]^{-1} \\ &= \frac{\mathrm{tr}\left(\mathbf{A}(\theta)\mathbf{R}_{\mathsf{q}}\mathbf{A}^{H}(\theta)\right)}{2\,\mathsf{SNR}_{\mathsf{radar}}\left(\mathrm{tr}\left(\dot{\mathbf{A}}(\theta)\mathbf{R}_{\mathsf{q}}\dot{\mathbf{A}}^{H}(\theta)\right)\mathrm{tr}\left(\mathbf{A}(\theta)\mathbf{R}_{\mathsf{q}}\mathbf{A}^{H}(\theta)\right) - \left|\mathrm{tr}\left(\mathbf{A}(\theta)\mathbf{R}_{\mathsf{q}}\dot{\mathbf{A}}^{H}(\theta)\right)\right|^{2}\right)},\end{aligned} \tag{4.24}$$

The CRB matrix for estimating AoA, θ, is written in (4.24) [43,118] where $\mathbf{A}(\theta) \triangleq \mathbf{b}(\theta)\mathbf{a}^{\top}(\theta)$; $\dot{\mathbf{A}} \triangleq \partial\mathbf{A}(\theta)/\partial\theta$; the SNR of received radar echo $\mathsf{SNR}_{\mathsf{radar}} = P_t L|\alpha|^2/\sigma_r^2$; P_t is the transmitted power; L represents the number of transmit symbols within one coherent processing interval (CPI), which facilities the multi-pulse processing and improves the received radar SNR [119]. The expression of the elements in FIM (*i.e.* $[\mathbf{J}]_{\theta\theta}, [\mathbf{J}]_{\theta\alpha}, [\mathbf{J}]_{\alpha\alpha}$) are derived in Appendix 7.2.3. Note that constraining the CRB on angle effectively limits the minimum achievable error in estimating the AoA. Thus, it indirectly necessitates the transmit beamforming pattern to concentrate its power on the target. This, in turn, leads to an overall enhancement of detection performance, such as Doppler detection. In practice, the AoA of the radar echo, θ, changes slowly compared with the adjacent coherent transmission time blocks [120]. Thus, we assume that the AoA, θ, is fixed in the optimization. For the same reason, we also assume the Doppler shift, f_d, is a constant in the optimization [121].

Remark 8. *We employ CRB as the sensing metric instead of radar SINR or beampattern matching. This choice is made due to the non-convex and thus intractable nature of the fractional form in radar SINR. Additionally, both of these metrics concentrate on transmit-side design, and neglect real target detection performance. In contrast, CRB prioritizes the optimization of the estimation performance and implicitly constrains the beampattern direction.*

4.3.3 Problem Formulation

We aim to maximize the minimum EE among all users to guarantee fairness among users in this section. We formulate the max-min EE fairness problem and optimize the precoders $\mathbf{p}_c, \mathbf{p}_r, \{\mathbf{p}_k\}_{k\in\mathcal{K}}$, and the vector of common rate portions, $\mathbf{c} = [C_1, \cdots, C_K]^{\top}$. Therefore, the optimization problem is formulated as

$$\max_{\substack{\mathbf{p}_c,\mathbf{p}_r,\mathbf{c}\\ \{\mathbf{p}_k\}_{k\in\mathcal{K}}}} \quad \min_{k\in\mathcal{K}} \quad \frac{(C_k+R_k)}{P_{\mathsf{RF}}+1/\kappa\left(\|\mathbf{\Delta p}_k\|^2+1/K\|\mathbf{\Delta p}_c\|^2\right)} \tag{4.25a}$$

$$\text{s.t.} \quad \|\mathbf{\Delta p}_c\|^2+\sum_{k\in\mathcal{K}}\|\mathbf{\Delta p}_k\|^2+\psi\|\mathbf{\Delta p}_r\|^2\leqslant P_{\mathrm{t}}, \tag{4.25b}$$

$$R_{\mathrm{c},k}\geqslant\sum_{i\in\mathcal{K}}C_i,\quad \forall k\in\mathcal{K}, \tag{4.25c}$$

$$C_k\geqslant 0,\quad \forall k\in\mathcal{K}, \tag{4.25d}$$

$$R_k+C_k\geqslant R_{\mathsf{th}},\quad \forall k\in\mathcal{K}, \tag{4.25e}$$

$$\mathsf{CRB}(\theta)\leqslant\rho, \tag{4.25f}$$

where R_{th} is a parameter for the required rate for each user, ρ is a parameter as the upper limit of CRB. Constraint (4.25b) limits the total transmit power of the ISAC satellite. (4.25c) guarantees the successful decoding of the common stream by all users. (4.25d) ensures all common rate portions are non-negative. (4.25e) is imposed to guarantee the QoS requirement of each user. To guarantee the target estimation performance, the CRB of the angle is upper limited as in (4.25f).

4.4 The Proposed Algorithm for Minimal Energy Efficiency Maximization

In this section, we tackle the non-convexity of problem (4.25). To have computational efficiency in the problem solving, we transform the original problem to a SDP form. Since the non-convexity of the objective function (4.25a) attributable to the fractional structure, we rewrite the objective function as a form of the difference of concave functions (DC). Subsequently, SCA is employed to transform the communication rate expression to a convex form. To handle the non-convexity in the fractional form of the objective function, we adopt Dinkelbach's algorithm [59] to transform the problem to an equivalent linear objective function. Subsequently, the objective function is transformed to a constraint. The constraints (4.25c) and (4.25e) are intractable due to the non-convex rate expressions, thus we introduce auxiliary variables and adopt SCA to have the linear approximation. The CRB constraint (4.25f) is transformed to a linear matrix inequality (LMI) by Schur Complement. To satisfy the rank-one constraint, a penalty function is added to the objective function. The details of the transformation and algorithm is given below.

4.4.1 SDP Transformation

To make our algorithm computationally efficient, the original problem (4.25) is first transformed into a SDP problem, given by

$$\max_{\substack{\mathbf{P}_c,\mathbf{P}_r,\mathbf{c},\\ \{\mathbf{P}_k\}_{k\in\mathcal{K}}}} \quad \min_{k\in\mathcal{K}} \quad \frac{\left(C_k + R_k(\mathbf{P}_c,\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi)\right)}{P_{\mathsf{chain},k}(\mathbf{P}_k,\boldsymbol{\Delta})} \tag{4.26a}$$

$$\text{s.t.} \quad \mathrm{tr}(\boldsymbol{\Delta}\mathbf{P}\mathbf{P}^H\boldsymbol{\Delta}) \leqslant P_{\mathrm{t}}, \tag{4.26b}$$

$$\mathbf{P}_c \succeq 0, \mathbf{P}_r \succeq 0, \mathbf{P}_k \succeq 0, \forall k \in \mathcal{K} \tag{4.26c}$$

$$\mathrm{rank}(\mathbf{P}_c) = 1, \mathrm{rank}(\mathbf{P}_r) = 1, \mathrm{rank}(\mathbf{P}_k) = 1, \quad \forall k \in \mathcal{K}, \tag{4.26d}$$

$$\log_2\left(1+\Gamma_{c,k}\right) \geqslant \sum_{i\in\mathcal{K}} C_i, \quad \forall k \in \mathcal{K}, \tag{4.26e}$$

$$C_k \geqslant 0, \quad \forall k \in \mathcal{K}, \tag{4.26f}$$

$$\log_2\left(1+\Gamma_k\right) \geqslant r_k, \quad \forall k \in \mathcal{K}, \tag{4.26g}$$

$$C_k + r_k \geqslant R_{\mathsf{th}} \quad \forall k \in \mathcal{K}, \tag{4.26h}$$

$$\mathsf{CRB}(\theta) \leqslant \rho, \tag{4.26i}$$

where we define $\mathbf{P}_c = \mathbf{p}_c\mathbf{p}_c^H, \mathbf{P}_r = \mathbf{p}_r\mathbf{p}_r^H, \mathbf{P}_k = \mathbf{p}_k\mathbf{p}_k^H, \mathbf{H}_k = \mathbf{h}_k\mathbf{h}_k^H$ for the SDP transformation; $\mathbf{r} = [r_1, \cdots, r_K]^\top$ is an auxiliary variable; $P_{\mathsf{chain},k}(\mathbf{P}_k,\boldsymbol{\Delta}) = P_{\mathsf{RF}} + 1/\kappa\left(\mathrm{tr}(\boldsymbol{\Delta}\mathbf{P}_k\boldsymbol{\Delta}) + 1/K\left(\boldsymbol{\Delta}\mathbf{P}_c\boldsymbol{\Delta}\right)\right)$,

$$\Gamma_{c,k} = \frac{\mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_c\boldsymbol{\Delta}\right)}{\sum\limits_{j\in\mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_j\boldsymbol{\Delta}\right) + \psi\,\mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_r\boldsymbol{\Delta}\right) + \mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Sigma}\right) + \sigma_k^2}, \tag{4.27a}$$

$$\Gamma_k = \frac{\mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_k\boldsymbol{\Delta}\right)}{\sum\limits_{\substack{j\in\mathcal{K},\\ j\neq k}} \mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_j\boldsymbol{\Delta}\right) + \psi\,\mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_r\boldsymbol{\Delta}\right) + \mathrm{tr}\left(\mathbf{H}_k\boldsymbol{\Sigma}\right) + \sigma_k^2}, \tag{4.27b}$$

and $R_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi)$ is given in (4.28).

Objective Function Transformation by SCA and Dinkelbach's Method. Since the rate expression in the numerator, $R_{\mathsf{total},k}(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},C_k,\psi)$, of the objective function (4.26) is intractable, and it is a classic DC programming problem, we

$$R_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi)=\log_2\big(\sum_{j\in\mathcal{K}}\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_j\boldsymbol{\Delta})+\psi\,\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_r\boldsymbol{\Delta})+\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Sigma})+\sigma_k^2\big)$$
$$-\log_2\big(\sum_{j\in\mathcal{K},j\neq k}\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_j\boldsymbol{\Delta})+\psi\,\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_r\boldsymbol{\Delta})+\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Sigma})+\sigma_k^2\big) \tag{4.28}$$

---

can tackle this problem by SCA method [122]. Thus, the numerator in (4.26a) is rewritten as

$$\begin{aligned}R_{\mathsf{total},k}(\mathbf{P}_k,\boldsymbol{\Delta},C_k,\psi)&=C_k+\log_2(1+\Gamma_k)\\&=f_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},C_k,\psi)-g_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi),\end{aligned} \tag{4.29}$$

where

$$\begin{aligned}f_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},C_k,\psi)\triangleq C_k+\log_2\big(&\sum_{j\in\mathcal{K}}\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_j\boldsymbol{\Delta})\\&+\psi\,\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_r\boldsymbol{\Delta})+\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Sigma})+\sigma_k^2\big),\end{aligned} \tag{4.30a}$$

$$\begin{aligned}g_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi)\triangleq\log_2\big(&\sum_{j\in\mathcal{K},j\neq k}\mathrm{tr}(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_j\boldsymbol{\Delta})\\&+\psi\,\mathrm{tr}(\mathbf{H}_k\boldsymbol{\Delta}\mathbf{P}_r\boldsymbol{\Delta})+\mathrm{tr}\,(\mathbf{H}_k\boldsymbol{\Sigma})+\sigma_k^2\big).\end{aligned} \tag{4.30b}$$

By using the SCA, the concave function $g_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi)$ is upper bounded by its linear approximation at any feasible point with respect to $\mathbf{P}_r$ and $\mathbf{P}_k$, rewritten as

$$\begin{aligned}&g_k(\mathbf{P}_r,\mathbf{P}_k,\boldsymbol{\Delta},\psi)\leqslant g_k(\mathbf{P}_r^{(t)},\mathbf{P}_k^{(t)},\boldsymbol{\Delta},\psi)\\&+\sum_{j\in\mathcal{K},j\neq k}\mathrm{tr}(\boldsymbol{\nabla}^H_{\mathbf{P}_j}g_k(\mathbf{P}_r^{(t)},\mathbf{P}_k^{(t)},\boldsymbol{\Delta},\psi)(\mathbf{P}_j-\mathbf{P}_j^{(t)}))\\&+\mathrm{tr}(\boldsymbol{\nabla}^H_{\mathbf{P}_r}g_k(\mathbf{P}_r^{(t)},\mathbf{P}_k^{(t)},\boldsymbol{\Delta},\psi)(\mathbf{P}_r-\mathbf{P}_r^{(t)})\\&\triangleq\widetilde{g}(\mathbf{P}_r^{(t)},\mathbf{P}_k^{(t)},\boldsymbol{\Delta},\psi),\end{aligned} \tag{4.31}$$

where t represents the t^{th} SCA iteration, and

$$\begin{aligned}\boldsymbol{\nabla}^H_{\mathbf{P}_j}g_k(\mathbf{P}_r^{(t)},\mathbf{P}_k^{(t)},\boldsymbol{\Delta},\psi)&=\boldsymbol{\nabla}^H_{\mathbf{P}_r}g_k(\mathbf{P}_r^{(t)},\mathbf{P}_k^{(t)},\boldsymbol{\Delta},\psi)\\&=\frac{\boldsymbol{\Delta}\mathbf{H}_k\boldsymbol{\Delta}}{\ln 2},\quad\forall j.\end{aligned} \tag{4.32}$$

Since $\widetilde{g}(\mathbf{P}_r^{(t)}, \mathbf{P}_k^{(t)}, \mathbf{\Delta}, \psi)$ is now an affine function, $\widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r^{(t)}, \mathbf{P}_k^{(t)}, \mathbf{\Delta}, C_k^{(t)}, \psi) \triangleq f_k(\mathbf{P}_r, \mathbf{P}_k, \mathbf{\Delta}, C_k, \psi) - \widetilde{g}(\mathbf{P}_r^{(t)}, \mathbf{P}_k^{(t)}, \mathbf{\Delta}, \psi)$ is a concave function. Therefore, the transformed problem is as follows

$$\max_{\substack{\mathbf{P}_c, \mathbf{P}_r, \mathbf{c}, \\ \{\mathbf{P}_k\}_{k\in\mathcal{K}}}} \quad \min_{k\in\mathcal{K}} \quad \frac{\widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r, \mathbf{P}_k, \mathbf{\Delta}, C_k, \psi)}{P_{\mathsf{chain},k}(\mathbf{P}_k, \mathbf{\Delta})} \tag{4.33a}$$

$$\text{s.t.} \qquad \text{(4.26b)} - \text{(4.26i)}.$$

Subsequently, this fractional programming (FP) problem (4.33) can be iteratively solved by the Dinkelbach's algorithm [59], the transformed problem with integral expression in the objective function (4.33a) is rewritten as

$$\max_{\substack{\mathbf{P}_c, \mathbf{P}_r, \mathbf{c}, \\ \{\mathbf{P}_k\}_{k\in\mathcal{K}}}} \quad \min_{k\in\mathcal{K}} \widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r, \mathbf{P}_k, \mathbf{\Delta}, C_k, \psi) - \lambda \left(P_{\mathsf{chain},k}(\mathbf{P}_k, \mathbf{\Delta})\right) \tag{4.34a}$$

$$\text{s.t.} \qquad \text{(4.26b)} - \text{(4.26i)}.$$

Next, the variable λ is updated as follows:

$$\lambda^{(t+1)} = \min_{k\in\mathcal{K}} \quad \frac{\widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r^{(t)}, \mathbf{P}_k^{(t)}, \mathbf{\Delta}, C_k^{(t)}, \psi)}{P_{\mathsf{chain},k}(\mathbf{P}_k^{(t)}, \mathbf{\Delta})}, \tag{4.35}$$

where t denotes the t^{th} iteration in the optimization process.

To make the max-min objective function (4.34a) smooth, we introduce an auxiliary variable μ [123], and the equivalent problem is as follows

$$\max_{\substack{\mathbf{P}_c, \mathbf{P}_r, \mathbf{c}, \\ \{\mathbf{P}_k\}_{k\in\mathcal{K}}, \mu}} \quad \mu \tag{4.36a}$$

$$\text{s.t.} \qquad \widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r, \mathbf{P}_k, \mathbf{\Delta}, C_k, \psi) - \lambda \left(P_{\mathsf{chain},k}(\mathbf{P}_k, \mathbf{\Delta})\right) \geqslant \mu, \forall k \in \mathcal{K} \tag{4.36b}$$

$$\text{(4.26b)} - \text{(4.26i)}.$$

Rate Constraints Transformation by SCA. To tackle the non-convexity in (4.26e) and (4.26g), we utilize the SCA technique to transform these two constraints.

Inspired by [42], we first introduce slack variables $\eta_{c,k}, \eta_k, \beta_{c,k}, \beta_k, \forall k \in \mathcal{K}$. Thus, (4.26e) is transformed to

$$\eta_{c,k} - \beta_{c,k} \geqslant \sum_{i \in \mathcal{K}} C_i \log 2, \tag{4.37}$$

$$\begin{aligned} e^{\eta_{c,k}} \leqslant & \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_c \boldsymbol{\Delta}\right) + \psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right) + \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) \\ & + \sum_{j \in \mathcal{K}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right) + \sigma_k^2, \end{aligned} \tag{4.38}$$

$$\begin{aligned} e^{\beta_{c,k}} \geqslant & \sum_{j \in \mathcal{K}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right) + \psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right) \\ & + \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) + \sigma_k^2, \end{aligned} \tag{4.39}$$

where $\forall k \in \mathcal{K}$. Next, (4.26g) is transformed to

$$\eta_k - \beta_k \geqslant r_k \log 2, \tag{4.40}$$

$$e^{\eta_k} \leqslant \sum_{j \in \mathcal{K}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right) + \psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right) + \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) + \sigma_k^2, \tag{4.41}$$

$$e^{\beta_k} \geqslant \sum_{\substack{j \in \mathcal{K}, \\ j \neq k}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right) + \psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right) + \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) + \sigma_k^2, \tag{4.42}$$

where $\forall k \in \mathcal{K}$. However, (4.39) and (4.42) are still non-convex, thus the first-order Taylor approximation is utilized to tackle these two constraints, given by

$$\begin{aligned} & \sum_{j \in \mathcal{K}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right) + \psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right) + \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) \\ & \quad + \sigma_k^2 \leqslant e^{\beta_{c,k}^{(t)}} \left(\beta_{c,k} - \beta_{c,k}^{(t)} + 1\right), \end{aligned} \tag{4.43}$$

$$\begin{aligned} & \sum_{j \in \mathcal{K}, j \neq k} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right) + \psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right) \\ & \quad + \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) + \sigma_k^2 \leqslant e^{\beta_k^{(t)}} \left(\beta_k - \beta_k^{(t)} + 1\right). \end{aligned} \tag{4.44}$$

(4.38) and (4.41) belong to generalized nonlinear convex programming, and result in high computational complexity. To efficiently cope with (4.38) and (4.41), we introduce the auxiliary variables $\tau_{c,k}, \tau_k$, and then compute a first-

order Taylor approximation. Therefore, (4.38) and (4.41) are rewritten as

$$\begin{aligned}\tau_{c,k} \leqslant & \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_c \boldsymbol{\Delta}\right)+\psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right)+\operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right) \\ & +\sum_{j \in \mathcal{K}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right)+\sigma_k^2,\end{aligned} \tag{4.45}$$

$$\tau_{c,k}^{(t)} \log\left(\tau_{c,k}^{(t)}\right)+\left(\tau_{c,k}-\tau_{c,k}^{(t)}\right)\left[\log\left(\tau_{c,k}^{(t)}\right)+1\right] \geqslant \tau_{c,k} \eta_{c,k}, \tag{4.46}$$

$$\tau_k \leqslant \sum_{j \in \mathcal{K}} \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_j \boldsymbol{\Delta}\right)+\psi \operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Delta} \mathbf{P}_r \boldsymbol{\Delta}\right)+\operatorname{tr}\left(\mathbf{H}_k \boldsymbol{\Sigma}\right)+\sigma_k^2, \tag{4.47}$$

$$\tau_k^{(t)} \log\left(\tau_k^{(t)}\right)+\left(\tau_k-\tau_k^{(t)}\right)\left[\log\left(\tau_k^{(t)}\right)+1\right] \geqslant \tau_k \eta_k. \tag{4.48}$$

To increase efficiency and numerical robustness of the solution, we transform (4.46) and (4.48) to second-order cone (SOC) forms. Taking the transformation of (4.46) as an example, (4.46) can be rewritten by multiplying 4 in both sides as

$$-4\tau_{c,k}(\eta_{c,k}-(\log\left(\tau_{c,k}^{(t)}\right)+1)) \geqslant 4\tau_{c,k}^{(t)}. \tag{4.49}$$

Through completing the square in both sides of the inequality (4.49), it is further transformed as

$$\begin{aligned}&\left(\tau_{c,k}+\left(\eta_{c,k}-(\log\left(\tau_{c,k}^{(t)}\right)+1)\right)\right)^2+4\tau_{c,k}^{(t)} \\ &\leqslant\left(\tau_{c,k}-\left(\eta_{c,k}-(\log\left(\tau_{c,k}^{(t)}\right)+1)\right)\right)^2.\end{aligned} \tag{4.50}$$

Therefore, the equivalent second-order cone (SOC) forms of (4.46) and (4.48) are as follows

$$\begin{aligned}&\left\|\left[\tau_{c,k}+\eta_{c,k}-\left(\log\left(\tau_{c,k}^{(t)}\right)+1\right), 2\sqrt{\tau_{c,k}^{(t)}}\right]\right\|_2 \\ &\leqslant \tau_{c,k}-\eta_{c,k}+\log\left(\tau_{c,k}^{(t)}\right)+1,\end{aligned} \tag{4.51}$$

$$\begin{aligned}&\left\|\left[\tau_k+\eta_k-\left(\log\left(\tau_k^{(t)}\right)+1\right), 2\sqrt{\tau_k^{(t)}}\right]\right\|_2 \\ &\leqslant \tau_k-\eta_k+\log\left(\tau_k^{(t)}\right)+1.\end{aligned} \tag{4.52}$$

Thus, the transformed problem is rewritten as

$$\begin{aligned} \max_{\substack{\mathbf{P}_c,\mathbf{P}_r,\mathbf{c},\\ \{\mathbf{P}_k\}_{k\in\mathcal{K}},\mu}} \quad & \mu && \text{(4.53a)}\\ \text{s.t.} \quad & \text{(4.26b)} - \text{(4.26d)}, \text{(4.26f)}, \text{(4.26h)}, \text{(4.26i)}, \text{(4.36b)}, \\ & \text{(4.37)}, \text{(4.40)}, \text{(4.43)}, \text{(4.44)}, \text{(4.45)}, \text{(4.47)}, \text{(4.51)}, \text{(4.52)}. \end{aligned}$$

4.4.2 CRB Transformation by Schur Complement

To handle the non-convexity in the $\mathsf{CRB}(\theta)$ (4.24) due to the fractional structure, we first reformulate it as

$$\mathrm{tr}\left(\dot{\mathbf{A}}(\theta)\mathbf{R}_{\mathsf{q}}\dot{\mathbf{A}}^H(\theta)\right) - \frac{\left|\mathrm{tr}\left(\mathbf{A}(\theta)\mathbf{R}_{\mathsf{q}}\dot{\mathbf{A}}^H(\theta)\right)\right|^2}{\mathrm{tr}\left(\mathbf{A}(\theta)\mathbf{R}_{\mathsf{q}}\mathbf{A}^H(\theta)\right)} - \frac{1}{2\rho\mathsf{SNR}} \geqslant 0. \tag{4.54}$$

By leveraging Schur complement [120,124], we have the following linear matrix inequality, given by

$$\begin{bmatrix} \mathrm{tr}\left(\dot{\mathbf{A}}\mathbf{R}_{\mathsf{q}}\dot{\mathbf{A}}^H\right) - \frac{1}{2\rho\mathsf{SNR}} & \mathrm{tr}\left(\mathbf{A}\mathbf{R}_{\mathsf{q}}\dot{\mathbf{A}}^H\right) \\ \mathrm{tr}\left(\dot{\mathbf{A}}\mathbf{R}_{\mathsf{q}}\mathbf{A}^H\right) & \mathrm{tr}\left(\mathbf{A}\mathbf{R}_{\mathsf{q}}\mathbf{A}^H\right) \end{bmatrix} \succeq \mathbf{0}. \tag{4.55}$$

Therefore, the reformulated problem is written as

$$\begin{aligned} \max_{\substack{\mathbf{P}_c,\mathbf{P}_r,\mathbf{c},\\ \{\mathbf{P}_k\}_{k\in\mathcal{K}},\mu}} \quad & \mu && \text{(4.56a)}\\ \text{s.t.} \quad & \text{(4.26b)} - \text{(4.26d)}, \text{(4.26f)}, \text{(4.26h)}, \text{(4.36b)}, \text{(4.37)}, \\ & \text{(4.40)}, \text{(4.43)}, \text{(4.44)}, \text{(4.45)}, \text{(4.47)}, \text{(4.51)}, \text{(4.52)}, \text{(4.55)}. \end{aligned}$$

4.4.3 Rank-one Constraint Transformation by Penalty Function

The rank-one constraints (4.26d) are non-convex, we adopt the penalty function and a corresponding iterative method to tackle this issue [49, 65]. Given that a rank-one matrix implies the presence of only a single nonzero eigenvalue, we

can rewrite the non-convex constraints (4.26d) in an alternative form as

$$\mathrm{tr}(\mathbf{P}_c) - \chi_{\mathsf{max}}(\mathbf{P}_c) = 0, \tag{4.57a}$$

$$\mathrm{tr}(\mathbf{P}_r) - \chi_{\mathsf{max}}(\mathbf{P}_r) = 0, \tag{4.57b}$$

$$\mathrm{tr}(\mathbf{P}_k) - \chi_{\mathsf{max}}(\mathbf{P}_k) = 0, \quad \forall k \in \mathcal{K}, \tag{4.57c}$$

where $\chi_{\mathsf{max}}()$ denotes the maximum eigenvalue of a matrix. Subsequently, we formulate a penalty function to incorporate these constraints into the objective function (4.56a), as follows

$$\begin{aligned} \max_{\substack{\mathbf{P}_c,\mathbf{P}_r,\mathbf{c},\\ \{\mathbf{P}_k\}_{k\in\mathcal{K}},\mu}} \quad & \mu - \xi\Big([\mathrm{tr}(\mathbf{P}_c) - \chi_{\mathsf{max}}(\mathbf{P}_c)] + \psi[\mathrm{tr}(\mathbf{P}_r) - \chi_{\mathsf{max}}(\mathbf{P}_r)] \\ & \qquad + \sum_{k\in\mathcal{K}} [\mathrm{tr}(\mathbf{P}_k) - \chi_{\mathsf{max}}(\mathbf{P}_k)]\Big) \qquad (4.58a) \\ \text{s.t.} \quad & (4.26b) - (4.26d), (4.26f), (4.26h), (4.36b), (4.37), \\ & (4.40), (4.43), (4.44), (4.45), (4.47), (4.51), (4.52), (4.55), \end{aligned}$$

where ξ serves as a pre-defined penalty factor to minimize the penalty function to the greatest extent possible. However, since the penalty function exists, the objective function (4.58a) is non-concave. Next, an iterative approach is employed to handle this challenge, leveraging the inequality given below by taking $\mathbf{P}_c$ as an example

$$\mathrm{tr}\left(\mathbf{P}_c\right) - \left(\mathbf{m}_{c,\mathsf{max}}^{(t)}\right)^H \mathbf{P}_c \mathbf{m}_{c,\mathsf{max}}^{(t)} \geqslant \mathrm{tr}\left(\mathbf{P}_c\right) - \xi_{\mathsf{max}}\left(\mathbf{P}_c\right) \geqslant 0, \tag{4.59}$$

where $\mathbf{m}_{c,\mathsf{max}}^{(t)}$ is the normalized eigenvector at the t^{th} optimization iteration with respect to the maximum eigenvalue $\xi_{\mathsf{max}}\left(\mathbf{P}_c\right)$. Same to this definition, we define $\mathbf{m}_{r,\mathsf{max}}^{(t)}$, $\mathbf{m}_{k,\mathsf{max}}^{(t)}$ as the normalized eigenvectors with regard to $\mathbf{P}_c$ and $\mathbf{P}_k$, respectively. Finally, the approximate problem at the t^{th} iteration is formulated in (4.60), where PF is defined by

$$\begin{aligned} \max_{\mathbf{P}_c,\mathbf{P}_r,\mathbf{c},\{\mathbf{P}_k\}_{k\in\mathcal{K}},\mu} \quad & \mu - \xi\mathsf{PF} && \text{(4.60a)} \\ \text{s.t.} \quad & \text{(4.26b)} - \text{(4.26d)}, \text{(4.26f)}, \text{(4.26h)}, \text{(4.36b)}, \text{(4.37)}, \text{(4.40)}, \text{(4.43)}, \\ & \text{(4.44)}, \text{(4.45)}, \text{(4.47)}, \text{(4.51)}, \text{(4.52)}, \text{(4.55)}. \end{aligned}$$

$$\begin{aligned} \mathsf{PF} \triangleq & [\mathrm{tr}\left(\mathbf{P}_c\right) - \left(\mathbf{m}_{c,\mathsf{max}}^{(t)}\right)^H \mathbf{P}_c \mathbf{m}_{c,\mathsf{max}}^{(t)}] \\ & + \psi[\mathrm{tr}\left(\mathbf{P}_r\right) - \left(\mathbf{m}_{r,\mathsf{max}}^{(t)}\right)^H \mathbf{P}_r \mathbf{m}_{r,\mathsf{max}}^{(t)}] \\ & + \sum_{k\in\mathcal{K}} [\mathrm{tr}\left(\mathbf{P}_k\right) - \left(\mathbf{m}_{k,\mathsf{max}}^{(t)}\right)^H \mathbf{P}_k \mathbf{m}_{k,\mathsf{max}}^{(t)}]. \end{aligned} \tag{4.61}$$

The transformed problem is convex and only involves SOC constraints, and LMI, thus it can be efficiently solved via numerical tools, *e.g.* CVX [125]. In each iteration, a feasible solution is given around the previous solution. We summarize our proposed algorithm for addressing (4.60) in Algorithm 6, where ε is the iteration tolerance coefficient, t_{max} is the maximum iteration number. After iteration ends, an optimal solution is given, and eigenvalue decomposition (EVD) is used to obtain the optimized beamforming vectors, $\mathbf{p}_c, \mathbf{p}_r, \{\mathbf{p}_k\}_{k\in\mathcal{K}}$, from the SDP expressions.

4.4.4 Computational Complexity and Convergence Analysis

The formulated problem in (4.60) involves SDP, SOC and LMI constraints, which can be efficiently solved by standard interior-point method (IPM) [126]. The computational complexity is given by $\mathcal{O}\big(\sqrt{n}\log(\varepsilon^{-1})(mn^3+m^2n^2+m^3)\big)$, where ε is the convergence tolerance, n is the size of the positive semi-definite matrix, m is the number of the SDP constraints. In our case (4.26), $m = 4K+3, n = N_t$, therefore the computational complexity of the proposed Algorithm 6 is $\mathcal{O}\big(\log(\varepsilon^{-1})(N_t^{3.5} + 4KN_t^{2.5})\big)$.

The convergence of the proposed Algorithm 6 is also analyzed here. In each iteration, a feasible solution is given around the previous solution. In addition, the optimal solution $\{\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}, \lambda^{(t)}, \eta^{(t)}, \beta^{(t)}, \tau^{(t)}\}$ at the t^{th} iteration

Algorithm 6: Proposed Algorithm for Max-min Fairness Beamforming Design

Input: $\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}, \mathbf{H}, \mathbf{H}_r, P_t$.
Output: $\mathbf{P}_c^*, \mathbf{P}_r^*, \{\mathbf{P}_k^*\}_{k\in\mathcal{K}}, \lambda^*, \eta^*, \beta^*, \tau^*$.

Initialize $t \leftarrow 0, \lambda^{(t)} = 0, \xi = 0.5 \times 10^{-2}, \varepsilon = 10^{-3}, t_{\mathsf{max}} = 100$, $\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}$ by SVD;
while no convergence of objective function (4.60a) **&** $t < t_{\mathsf{max}}$ **do**
  $t \leftarrow t + 1$.
  Solve problem (4.60) to obtain $\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}$.
  Update $\lambda^{(t)} = \min\limits_{k\in\mathcal{K}} \quad \frac{\widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r^{(t-1)}, \mathbf{P}_k^{(t-1)}, \mathbf{\Delta}, C_k^{(t-1)}, \psi)}{P_{\mathsf{chain},k}(\mathbf{P}_k^{(t-1)}, \mathbf{\Delta})}$.
end
Set $\lambda^* = \lambda^{(t)}$.
Return $\mathbf{P}_c^*, \mathbf{P}_r^*, \{\mathbf{P}_k^*\}_{k\in\mathcal{K}}, \lambda^*, \eta^*, \beta^*, \tau^*$.

is also a feasible solution of the $(t+1)^{\text{th}}$ iteration. Hence, the objective function (4.60a) is non-decreasing after each iteration. Additionally, the objective function is bounded by lower bounds to satisfy the rank-one constraint [65] as given in (4.59), and thus guarantee a convergence. Consequently, the obtained feasible solution of (4.60) satisfies the Karush–Kuhn–Tucker (KKT) optimality conditions, and (4.58) and (4.60) have the same KKT conditions at convergence [65, 127]. Due to the convexification of the original problem, the feasible solution is a local optimum. This algorithm cannot guarantee that the solution is a global optimum.

Note that the global optimum is not guaranteed at the given feasible solution point.

4.4.5 Total EE maximization

The total EE maximization problem in the considered system model with low-resolution DACs is used as a benchmark for comparison. The total EE is defined by the sum-rate over the total power consumption as given in (4.62a). This problem can be solved by a similar algorithm as Algorithm 6 due to the

Algorithm 7: Algorithm for Total EE Beamforming Design

Input: $\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}, \mathbf{H}, \mathbf{H}_r, P_t$.

Output: $\mathbf{P}_c^*, \mathbf{P}_r^*, \{\mathbf{P}_k^*\}_{k\in\mathcal{K}}, \lambda_{\mathsf{total}}^*, \eta^*, \beta^*, \tau^*$.

Initialize $t \leftarrow 0, \lambda_{\mathsf{total}}^{(t)} = 0, \xi = 0.5 \times 10^{-2}, \varepsilon = 10^{-3}, t_{\mathsf{max}} = 100$, $\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}$ by SVD;

while no convergence of objective function (4.63a) **&** $t < t_{\mathsf{max}}$ **do**

  $t \leftarrow t + 1$.

  Solve problem (4.63) to obtain $\mathbf{P}_c^{(t)}, \mathbf{P}_r^{(t)}, \left\{\mathbf{P}_k^{(t)}\right\}_{k\in\mathcal{K}}$.

  Update $\lambda^{(t)} = \frac{\sum_{k\in\mathcal{K}} \widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r^{(t-1)}, \mathbf{P}_k^{(t-1)}, \mathbf{\Delta}, C_k^{(t-1)}, \psi)}{\sum_{k\in\mathcal{K}} P_{\mathsf{chain},k}(\mathbf{P}_k^{(t-1)}, \mathbf{\Delta})}$.

end

Set $\lambda_{\mathsf{total}}^* = \lambda^{(t)}$.

Return $\mathbf{P}_c^*, \mathbf{P}_r^*, \{\mathbf{P}_k^*\}_{k\in\mathcal{K}}, \lambda_{\mathsf{total}}^*, \eta^*, \beta^*, \tau^*$.

summation is an operation preserving convexity [65].

$$\begin{aligned} \max_{\substack{\mathbf{p}_c, \mathbf{p}_r, \mathbf{c} \\ \{\mathbf{p}_k\}_{k\in\mathcal{K}}}} \quad & \frac{\sum_{k\in\mathcal{K}}(C_k + R_k)}{\sum_{k\in\mathcal{K}} \left(P_{\mathsf{RF}} + 1/\kappa \left(\|\mathbf{\Delta}\mathbf{p}_k\|^2 + 1/K\|\mathbf{\Delta}\mathbf{p}_c\|^2\right)\right)} \\ \text{s.t.} \quad & \text{(4.25b)} - \text{(4.25f)}. \end{aligned} \tag{4.62a}$$

Without introducing the auxiliary variable μ to handle max-min form as in (4.36), the transformed problem is rewritten as

$$\begin{aligned} \max_{\substack{\mathbf{P}_c, \mathbf{P}_r, \mathbf{c}, \\ \{\mathbf{P}_k\}_{k\in\mathcal{K}}}} \quad & \sum_{k\in\mathcal{K}} \widetilde{R}_{\mathsf{total},k}(\mathbf{P}_r, \mathbf{P}_k, \mathbf{\Delta}, C_k, \psi) \\ & -\lambda\big(\sum_{k\in\mathcal{K}} P_{\mathsf{chain},k}(\mathbf{P}_k, \mathbf{\Delta})\big) - \xi\mathsf{PF} \\ \text{s.t.} \quad & \text{(4.26b)} - \text{(4.26d)}, \text{(4.26f)}, \text{(4.26h)}, \text{(4.37)}, \text{(4.40)}, \\ & \text{(4.43)}, \text{(4.44)}, \text{(4.45)}, \text{(4.47)}, \text{(4.51)}, \text{(4.52)}, \text{(4.55)}, \end{aligned} \tag{4.63a}$$

Therefore, we have the following Algorithm 7. Subsequently, we select the minimal EE among all the users to be the baseline.

Remark 9. *The ULA in the LEO-ISAC satellite system can be easily extended to uniform planar array (UPA) setup. If we assume that the transmit and receive UPAs on the LEO satellite are equipped M_t and M_r antenna elements, respectively. Azimuth and elevation angles of the target regarding the satellite UPAs are denoted by θ_h and θ_v, respectively. Thus, the radar echo model in* (4.14) *can be rewritten as* $\mathbf{Z}[\ell] = \alpha e^{\jmath 2\pi f_d \ell T}\mathbf{b}(\theta_h,\theta_v)\mathbf{a}^\top(\theta_h,\theta_v)\mathbf{x}_\mathbf{q}[\ell] + \mathbf{N}_r[\ell]$. *Here,* $\mathbf{a}(\theta_h,\theta_v) \in \mathbb{C}^{M_t\times 1}$ *and* $\mathbf{b}(\theta_h,\theta_v) \in \mathbb{C}^{M_r\times 1}$ *denote the transmit and receive steering vectors, respectively, given by* $\mathbf{a}(\theta_h,\theta_v) = \exp(-\jmath[\mathbf{r}_X,\mathbf{r}_Y,\mathbf{r}_Z]\mathbf{k}(\theta_h,\theta_v))$, *and* $\mathbf{b}(\theta_h,\theta_v) = \exp(-\jmath[\bar{\mathbf{r}}_X,\bar{\mathbf{r}}_Y,\bar{\mathbf{r}}_Z]\mathbf{k}(\theta_h,\theta_v))$, *where* $\mathbf{r}_Z = \bar{\mathbf{r}}_Z = \mathbf{0}$ *if the UPAs are deployed within* (X,Y) *plane. Additionally,* $\mathbf{k}(\theta_h,\theta_v) = \frac{2\pi}{\lambda_c}[\cos\theta_h\cos\theta_v, \sin\theta_h\cos\theta_v, \sin\theta_v]^\top$. *Subsequently, we can obtain the FIM and CRB on the azimuth and elevation angles, and utilize a similar optimization algorithm to design the system.*

4.5 Numerical evaluation

The objective of this section is to illustrate the superiority of the RSMA-assisted ISAC-LEO scheme using the proposed algorithm. We start from the investigation of the convergence behaviours. Next, we solve the optimization problem as in (4.60) to obtain the optimized precoders. We compare the communications and sensing performances of the following schemes: max-min fairness EE optimization: (a) RSMA with radar sequence with SIC, (b) RSMA with radar sequence without SIC, (c) RSMA without radar sequence, the schemes of total EE optimization: (d) RSMA with radar sequence with SIC, (e) RSMA with radar sequence without SIC, (f) SDMA. Note that the SDMA's optimization problem can be solved via Algorithm 7 by turning off the common stream.

4.5.1 Simulation Environment

We consider an overloaded ($N_t < K$) RSMA-assisted ISAC-LEO satellite system, where the LEO satellite is equipped $N_t = 4$ transmit antennas and $N_r = 4$ receive antennas. The provided method can be extended to a massive MIMO system, e.g., $N_r = 16$, with a higher computational complexity. We assume that $K = 5$ users are in the coverage area; the target is located at $\theta = 0^\circ$ with distance $r = 2000$ m; the relative velocity between the LEO satellite and target in LOS direction is $v = 10$ m/s. Note that no user grouping is considered in the simulation. The transmit power budget is $P_t = 20$ dBm, and we normalize the noise power by $\iota T_{\mathsf{sys}} B$, *i.e.* $\sigma_k^2 = 1, \forall k \in \mathcal{K}$ [49]. For the DAC power consumption model, we set $P_{\mathsf{LP}} = 14\,\text{mW}, P_{\mathsf{M}} = 0.3\,\text{mW}, P_{\mathsf{LO}} = 22.5\,\text{mW}, P_{\mathsf{H}} = 3\,\text{mW}$ and $\kappa = 0.27$ [116]. The simulation parameters of the LEO satellite channel are given by carrier frequency, $f_c = 20$ GHz, satellite height $d_{\mathsf{sat}} = 600$ km, bandwidth $B = 25$ MHz, 3 dB angle equaling to 0.4°, satellite antenna gain, $G_u = 17$ dBi, Boltzman's constant, $\iota = 1.38 \times 10^{-23}$ J/m, system noise temperature $T_{\mathsf{sys}} = 517$K, and rain fading parameters, $(\mu_{\mathsf{rain}}, \sigma^2_{\mathsf{rain}}) = (-2.6, 1.63)$. In the constraint, QoS requirement for each user, and CRB threshold are set as, $R_{\mathsf{th}} = 1$ and $\rho = 10^{-2}$, respectively. As for the quantized DACs, we assume that all RF chains are equipped with the same quantization bit DACs for simplic-

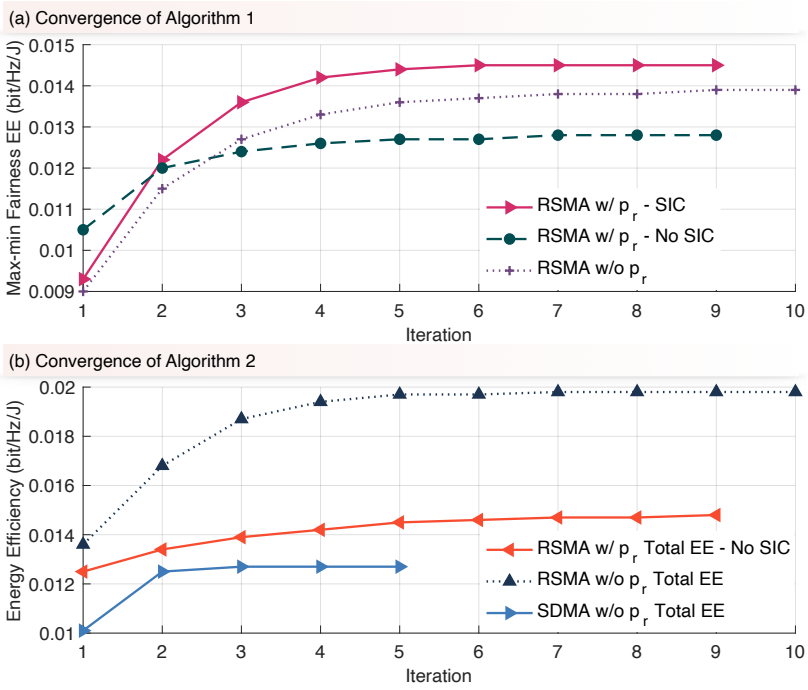


Figure 4.3 Convergence illustration for 2-bit cases (a) Algorithm 6, and (b) Algorithm 7 with 20 dB transmit SNR and 40 dB radar SNR.

ity, *i.e.* $b_m = b, \forall m \in \mathbb{Z}$. We adopt the values for the quantization loss ϱ when $b \leqslant 5$ as in [128]. When $b > 5$, we calculate ϱ as given in (4.9). The number CPI is $L = 1024$, and the signals, s, consist of uncoded i.i.d zero mean, unit-energy quadrature phase shift keying (QPSK) symbols. Recall that the SNR of received radar echo $\mathsf{SNR}_{\mathsf{radar}} = P_t L|\alpha|^2/\sigma_r^2$. We conduct 100 Monte Carlo simulations with perfect CSI to observe the performances.

Starting from the convergence analysis, Fig. 4.3 (a) and (b) respectively demonstrate the changes of the max-min fairness EE values and total EE values with respect to the iterations. The proposed algorithm can converge fast in 10 iterations for all cases. Specifically, RSMA with $\mathbf{p}_r$ and SIC at the receiver with metric max-min fairness EE has the best performances in the optimization process via Algorithm 6. In contrast, RSMA with $\mathbf{p}_r$ and no SIC has the lowest max-min fairness EE due to the interference from the radar sequence.

4.5.2 Communication Performances

We investigate the max-min fairness EE of the system as shown in Fig. 4.4 under the case that quantization bit, $b \in \{2, 4, 8\}$, and QoS requirements for each user, $R_{\mathsf{th}} = 1$. The results are obtained by solving (4.60) for the max-min fairness EE and (4.63) for the maximization of total EE. According to Fig. 4.4, the observations are that the max-min fairness EE increases with the decrease of the quantization bits, especially in the high transmit SNR range (*i.e.* $20-30$ dB). The reason behind is due to a lower power consumption caused by fewer quantization bit. Additionally, RSMA with $\mathbf{p}_r$ and SIC at the receiver has the largest max-min fairness EE. In contrast, SDMA has the lowest one. This is due to a better inter-beam interference management of the RSMA compared to SDMA. In the five RSMA scenarios, RSMA with $\mathbf{p}_r$ and SIC has the best max-min EE performance, followed by RSMA without $\mathbf{p}_r$ and RSMA with $\mathbf{p}_r$ and no SIC. This is because that RSMA with $\mathbf{p}_r$ and SIC has no radar sequence interference (*i.e.* $\psi = 0$) compared with RSMA with $\mathbf{p}_r$ and no SIC (*i.e.* $\psi = 1$). In addition, RSMA with $\mathbf{p}_r$ and SIC has a better communication-sensing trade-off, enabled by the precoder, $\mathbf{p}_r$, for sensing, when it is compared with other cases. The max-min fairness EE of RSMA without $\mathbf{p}_r$ is higher than that of RSMA with $\mathbf{p}_r$ without SIC due to no radar sequence interference (*i.e.* $\psi = 0$). The max-min fairness EE performances of the two cases, which are the maximization of the total EE with $\mathbf{p}_r$ and without $\mathbf{p}_r$, are worse than the three cases of directly maximizing minimum EE. In detail, the performance of total EE without $\mathbf{p}_r$ is better than total EE with $\mathbf{p}_r$ in terms of the max-min fairness EE. This is also because the latter case has no radar sequence interference.

We analyze the precoder power allocation here. The results are obtained by solving (4.60) and (4.63). As shown in Fig. 4.5 (a), the power allocation of each precoder in our proposed method is plotted. It is obvious that precoders $\mathbf{p}_c$ and $\mathbf{p}_r$ are well-utilized. Since our proposed design maximizes the max-min fairness EE under constraints, which are QoS for each user and sensing guarantee, a balance between communication (*i.e.* max-min fairness EE) and sensing performances are achieved. However, as shown in Fig. 4.5 (e), if the sum-rate maximization problem is considered, the power allocation of precoder $\mathbf{p}_r$ is

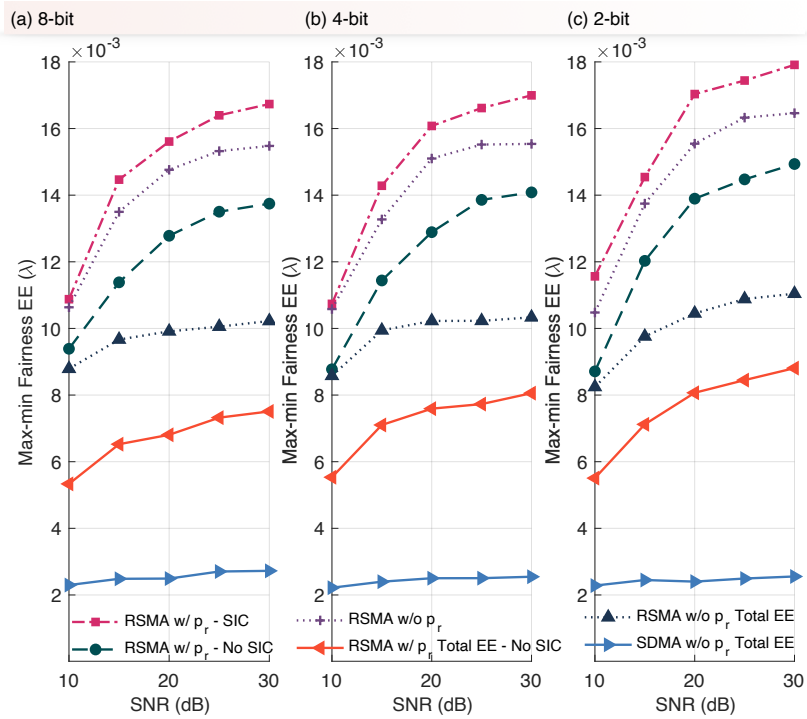


Figure 4.4 Max-min fairness EE versus transmit SNR with the quantization bit $b \in \{2, 4, 8\}$ and 40 dB radar SNR.

almost zero. This indicates that $\mathbf{p}_r$ is not well utilized when adopting the sum-rate metric. This conclusion is also verified by beampattern performances of the max-min fairness EE and sum-rate metrics. To further explain the reason why $\mathbf{p}_r$ is effective under the metric of max-min fairness EE, we compare the power allocations of this metric and sum-rate metric. Based on the optimization problem (4.25), we change the objective function to sum-rate, which is the numerator in (4.25a) without minimization. Subsequently, we follow the same problem and constraints transformation, and obtain the power allocation of sum-rate metric. It is observed that the sum-rate metric tends to turn off the radar sequence, thus $\mathbf{p}_c$ will work for both communication and sensing functionalities, and be probed at the target direction. In contrast, the metric that is max-min fairness EE can utilize both $\mathbf{p}_c$ and $\mathbf{p}_r$, resulting in a better trade-off between communication and sensing capabilities.

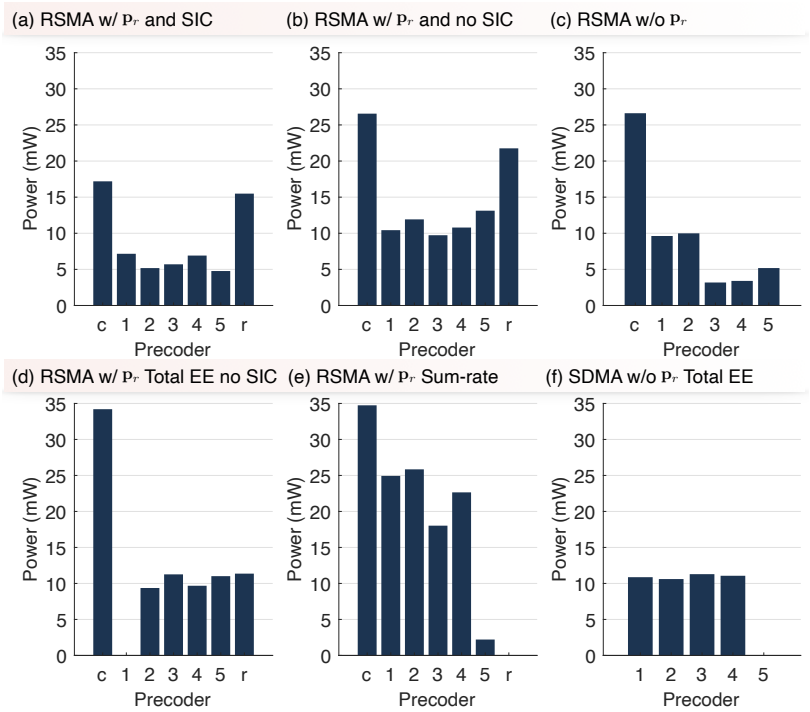


Figure 4.5 Power allocation of each precoder with 2-bit quantization and 20 dB transmit SNR and 40 dB radar SNR for (a) RSMA with $\mathbf{p}_r$ and SIC, (b) RSMA with $\mathbf{p}_r$ and no SIC, (c) RSMA without $\mathbf{p}_r$, (d) RSMA with $\mathbf{p}_r$ by maximizing total EE, (e) RSMA with $\mathbf{p}_r$ by maximizing sum-rate, (f) SDMA by maximizing total EE.

In addition, as shown in Fig. 4.6 (a), the max-min fairness EE decreases with b owing to the increased power consumption.

4.5.3 Sensing performances

The sensing performance is constrained by $\mathsf{CRB}(\theta) \leqslant \rho$, where a smaller ρ indicates a stronger constraint and results in a better sensing performance. As shown in Fig. 4.6 (b), RSMA with $\mathbf{p}_r$ and SIC can achieve the best trade-off between the communication and sensing functionalities; the sensing constraint is satisfied while the highest max-min fairness EE can be achieved. In addition, the max-min fairness EE of each curve gradually increases with a looser ρ

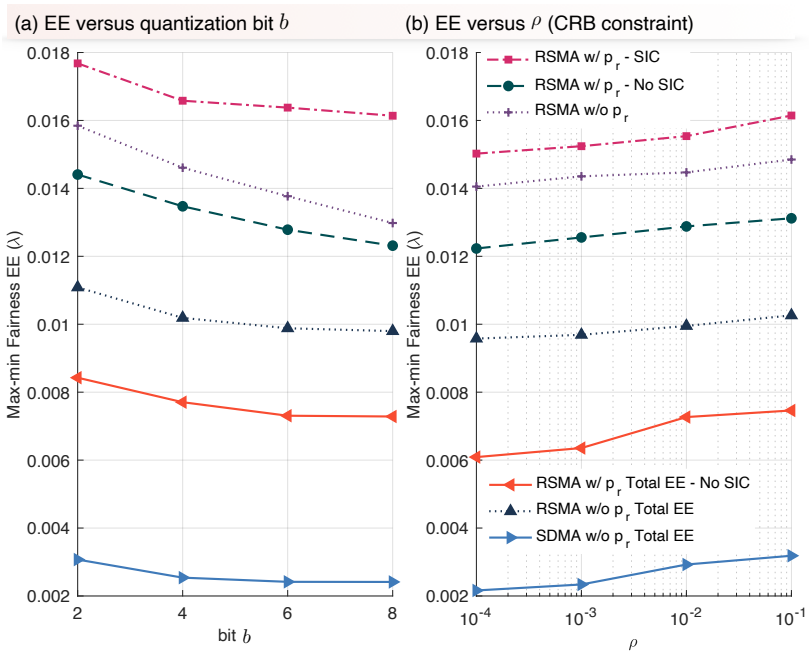


Figure 4.6 Under the case that 20 dB transmit SNR and 40 dB radar SNR, (a) Max-min fairness EE versus RCRB constraint ρ with $b = 2$, (b) Max-min fairness EE versus number of bits b.

constraint.

The real performances of the sensing functionality are captured by the information extraction. Here, we investigate the range-Doppler detection and beampattern performances. We use the same sensing detection pipeline as in [42, 108]. Next, we perform a discrete Fourier transform (DFT) operation to extract the range-Doppler map based on the discrete-time signal stream in (4.14). According to Fig. 4.7, observations are that target's Doppler-range information can be extracted by solving the precoder optimization problems with CRB constraint. Specifically, RSMA with $\mathbf{p}_r$ and SIC has the highest value followed by RSMA with $\mathbf{p}_r$ and no SIC, and RSMA without $\mathbf{p}_r$. In contrast, SDMA has the smallest detection value. These results indicate the superiority of our proposed approach in terms of the Doppler-range detection. To investigate how the precoders $\mathbf{p}_\mathrm{c}$ and $\mathbf{p}_\mathrm{r}$ are utilized in the two metrics (*i.e.* max-min fairness EE and

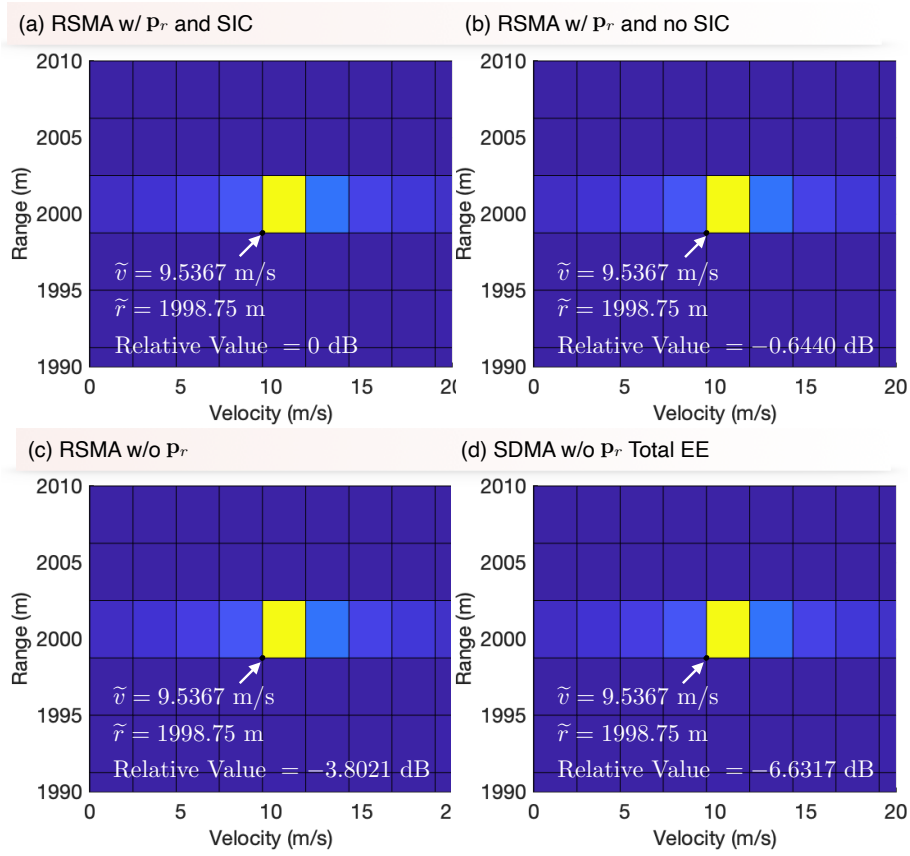


Figure 4.7 Detection results of Doppler and range with 2-bit quantization, 20 dB transmit SNR and 40 dB radar SNR: Range-Doppler Maps for (a) RSMA with $\mathbf{p}_\mathrm{r}$ and SIC, (b) RSMA with $\mathbf{p}_\mathrm{r}$ and no SIC, (c) RSMA without $\mathbf{p}_\mathrm{r}$, (d) SDMA by maximizing total EE.

sum-rate), we plot the beampatterns under two scenarios. The two scenarios are (1) users and target are distant (*i.e.* all users are located around -45° and target is at 30°); (2) users and target are near (*i.e.* all users are located around -45° and target is at -10°). As shown in Fig. 4.8, in both cases, $\mathbf{p}_\mathrm{r}$ is probed at the target direction when maximizing max-min fairness EE. In contrast, $\mathbf{p}_\mathrm{r}$ is not used when maximizing sum-rate. To achieve the sensing function, $\mathbf{p}_\mathrm{c}$ of the sum-rate metric is probed at the target direction. This observation is consistent with previous RSMA ISAC studies [41]. Moreover, $\mathbf{p}_\mathrm{c}$ of the max-min fairness

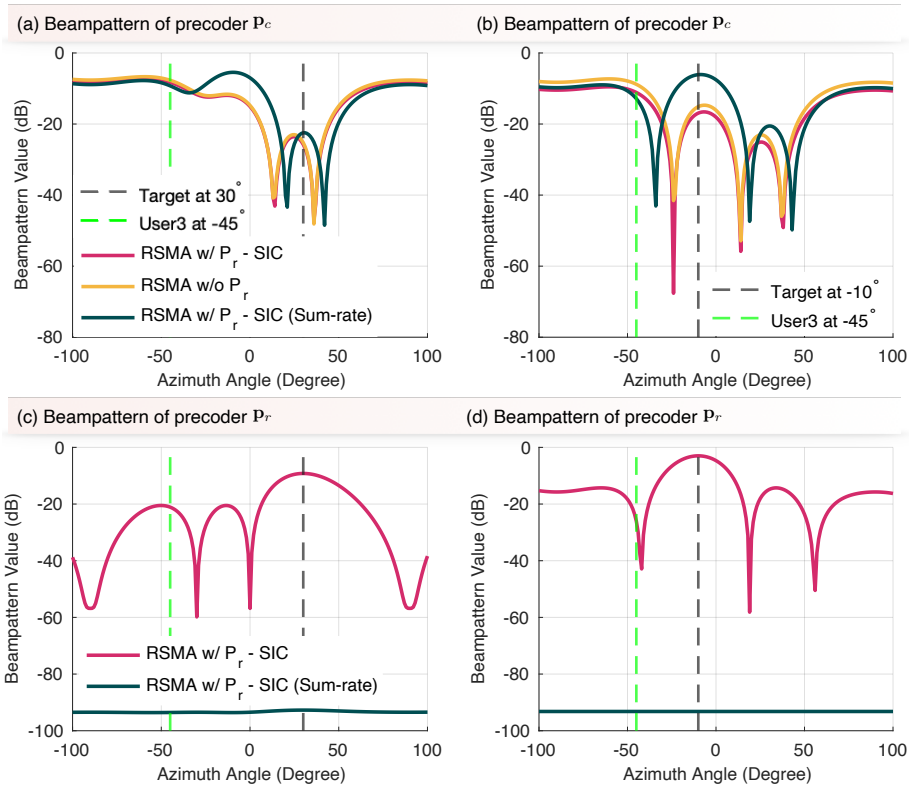


Figure 4.8 Under 2-bit quantization, 20 dB transmit SNR, and 40 dB radar SNR. Case 1: users and targets are distant, transmit beampatterns of (a) common stream precoder $\mathbf{p}_c$, (c) radar sequence precoder $\mathbf{p}_r$. Case 2: users and targets are near, transmit beampatterns of (b) common stream precoder $\mathbf{p}_c$, (d) radar sequence precoder $\mathbf{p}_r$.

EE metric is not completely probed at the target direction, and thus enables flexibility and a higher max-min fairness EE performance.

Remark 10. *Compared to prior RSMA ISAC studies [41–46], the consistent observation is that when maximizing the sum-rate, radar sequence will not be utilized, and the common stream will be used for both communication and sensing functions. Specifically, it will be probed at the target direction for sensing guarantee. The new observation is that when adopting max-min fairness EE as the metric, radar sequence can be utilized, resulting in the improvement of the minimal EE among users. The reason behind this is that the utilization of a separated radar sequence is more energy-efficient*

in terms of communication use. In detail, the power of radar sequence is used for sensing, and the power consumed by common and private streams, which determines EE, is minimized mainly for communication use.

4.6 Summary

In this chapter, we have explored an RSMA-assisted ISAC-LEO satellite system that incorporates the low-resolution DACs at the transmitter. This system serves multiple communication users and detects a moving target, simultaneously. We have investigated the max-min fairness EE of the proposed strategy under communication and sensing constraints. This problem has been formulated as a non-convex optimization problem by designing the precoders. An iterative algorithm based on SCA and Dinkelbach's method has been proposed to solve the problem. Numerical results have demonstrated that the proposed design outperforms the strategies that maximize the total EE of the system and SDMA in terms of max-min fairness EE and communication-sensing trade-off.

Chapter 5

λ–MIMO: Massive MIMO via Modulo Sampling

5.1 Introduction

As the name suggests, *massive* MIMO or M–MIMO architecture relies on a large number of antennas at the BS. This allows for serving a large number of users. M–MIMO has attracted much interest in the past decade because of its significant potential for *spectral efficiency gain* [14, 15]. However, the technological promise of M–MIMO is limited by two fundamental challenges [35], namely, 1. *power consumption*, and 2. *receiver saturation*. It is established wisdom that these two bottlenecks inhibit the practical implementation of M–MIMO in future wireless communication system. Specifically, each antenna in M–MIMO is connected to a corresponding RF chain consisting of i) a LNA, ii) a mixer, iii) a VGA with an AGC, and iv) an ADC. The RF-chain is illustrated in Fig. 5.1. Next, we elaborate on the two challenges that limit M–MIMO systems.

Challenge #1: Power Consumption. Among all the components in the RF chain, the ADC maps the received analog signal into the digital baseband samples and dominates the total power consumption of the RF-chain [129]. Note that the power consumption of the ADC grows linearly with the sampling rate but exponentially with the number of bits b that decide the ADC's resolution. Thus, the total power consumption of M–MIMO is a function of the ADC's bit budget considering current commercial M-MIMO BS with 64 antennas. For example, the power consumption of the BS equipped with 256 antenna elements, where each RF chain is equipped with two ADCs (quadrature pair), will be as high as $256\,\mathrm{W}$ [130], attributed to high bit resolution (8-12 bits) and high sampling rate ($\geqslant 20$ GSample/s). Such a power consumption is *unsustainable* in realistic deployments [130].

Challenge #2: Receiver Saturation. Another impediment to the M–MIMO systems is due to receiver-ADC saturation; typically this is the case with wireless communication systems. This impediment is caused by the fact that ADCs are limited by their dynamic range (DR), say $[-\lambda, \lambda]$ where the *lower* threshold, $-\lambda$ and *upper* threshold λ of the ADC are assumed to be symmetric (as illustrated in Fig. 5.2 (a)). When the received signal amplitude is larger than λ, the ADC output is distorted. This results in a clipped or saturated signal. Such undesir-

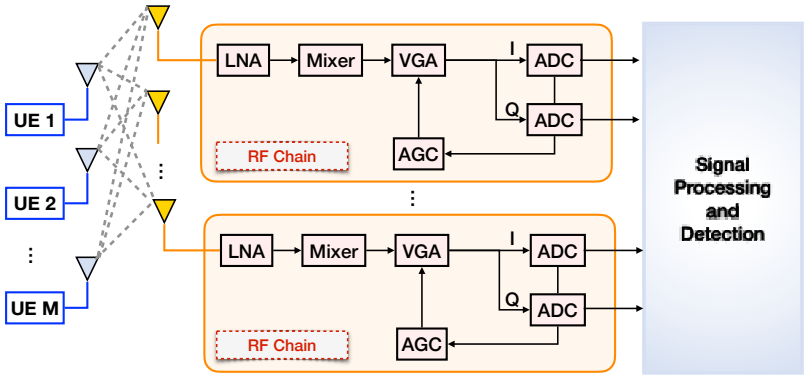


Figure 5.1 The receiver architecture of conventional massive MIMO (M–MIMO) system with uplink user equipment (UE). Each RF chain consists of the LNA, mixer, VGA, AGC, and ADC.

able yet unavoidable effects lead to high-frequency components in the digital signal, thus creating aliasing in the measurements. Consequently, the performance of the whole system deteriorates.

Previous Approaches for M–MIMO. To overcome the above stated problems—namely, *high power consumption* and *ADC saturation*—researchers typically resort to a solution strategy that disentangles the two challenges.

- High power consumption problem can be tacked by reducing the sampling rate and/or the bit resolution. Since Shannon's sampling theorem places a fundamental lower bound on the sampling rate, the only degree-of-freedom we can leverage is the bit resolution. This has been the driving force behind the conceptualization of low-resolution ADC based solutions [129–134]. That said, low-resolution ADCs in M–MIMO face a challenging signal detection problem due to the quantization noise and the non-linearity caused by the low-resolution quantization, resulting in sub-optimal system performance. One popular solution is to adopt low-order modulation schemes (*e.g.* BPSK) to increase the inter-symbol distance in the constellation diagram, thus simplifying the signal detection problem. However, if a higher symbol rate is required, higher-order modulations will have to be adopted. As a result, signal detection becomes difficult and the system performance deteriorates.

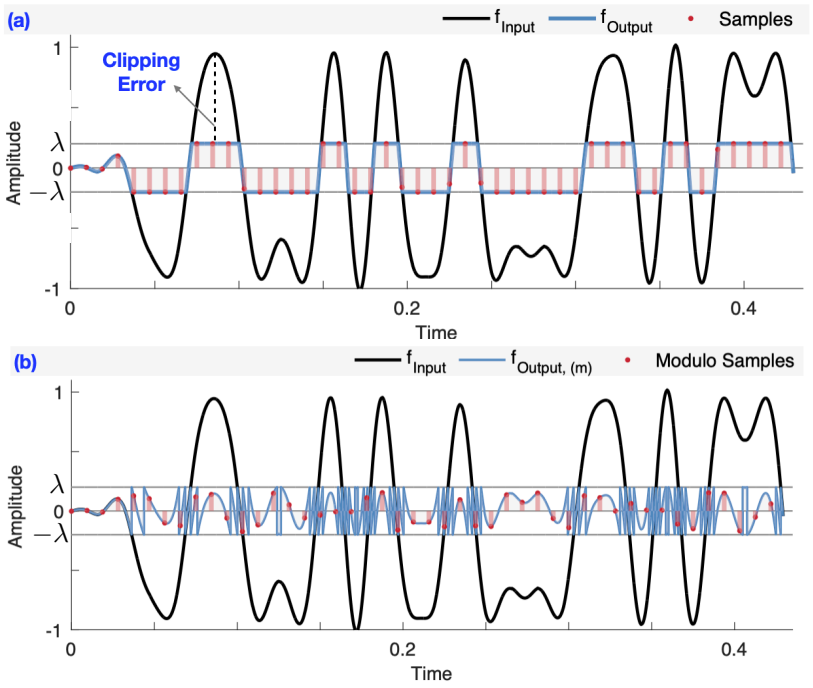


Figure 5.2 Illustration of output samples by conventional and $\mathscr{M}_\lambda$−ADCs. f_{Input}: continuous input signal before ADC. f_{Output}: continuous output signal shaped by limited ADC dynamic range. Samples: Sampled f_{Output} signals. a) In conventional ADC, saturation and clipping errors happen in f_{Output} signal and samples, because some values in f_{Input} exceeds dynamic threshold λ, b) In $\mathscr{M}_\lambda$−ADC, the input continuous signal f_{Input} is modulo-folded to $f_{\text{Output,(m)}}$. No saturation occurs in the $\mathscr{M}_\lambda$−ADC.

To improve the detection accuracy of higher-order modulations with low-resolution ADCs, two techniques are adopted in general,

—• *Spatial Oversampling* [132, 135] leverages multiple receive antennas receiving the same signal to improve the detection performance of high-order modulation (*e.g.* 16 QAM).

—• *Temporal Oversampling* [136] leverages a higher sampling rate to increase the resolution of the ADC.

It is clear that both of these strategies entail an increased power consumption thus leading to an inherent trade-off between *detection performance* and the

power consumption.

- To cope with the saturation problem, *automatic gain control* (AGC) and *variable gain amplifiers* (VGA) are adopted along with the ADC (cf. Fig. 5.1) to ensure that the amplitude of the input signal is within the dynamic range. The AGC tracks the input signal and returns the feedback information of the signal amplitude range to the VGA. Then, the VGA adjusts the amplification of the input signal based on the feedback. Unfortunately, the received communication signal is fast-changing and not deterministic; thus the AGC may sometimes fail to track the range of the signal. Consequently, clipping errors still occur. Alternatively, the logarithmic ADC [79, 137] can be used to extend the dynamic range of the ADC; however, the quantization resolution in the high amplitude region is coarse due to the logarithm operation, and information in this region may be lost, leading to system performance deterioration.

It is intuitive to note that strategies that take a decoupled approach, *i.e.* address the high power consumption and receiver saturation problems independently, run into balancing trade-offs that induce limitations on the overall system performance. To the best of our knowledge, there is no combined method with satisfactory performance that solves these two problems in a holistic fashion. This motivates investigation of new approaches and we take a step in this direction by introducing modulo ADCs in the pipeline.

Our Approach: Unlimited Sampling for M–MIMO. In this thesis, we take a radically different approach to M–MIMO and this done by replacing a conventional ADC (cf. Fig. 5.1) by a modulo ADC or $\mathscr{M}_\lambda$–ADC. The resulting approach is referred to as λ–MIMO. The $\mathscr{M}_\lambda$–ADCs are at the core of the Unlimited Sensing Framework (USF) [1–4, 138–141] which enables recovery of *high dynamic range* (HDR) signals by acquiring *low dynamic range* (LDR), modulo samples. The $\mathscr{M}_\lambda$–ADC maps a continuous-time signal into folded, digital samples by injecting a modulo non-linearity in the analog domain. This prevents distortion due to clipping/saturation that is prevalent in conventional ADCs [142–145]. However, as shown in the hardware example in Fig. 5.3, $\mathscr{M}_\lambda$–ADCs lead to a different

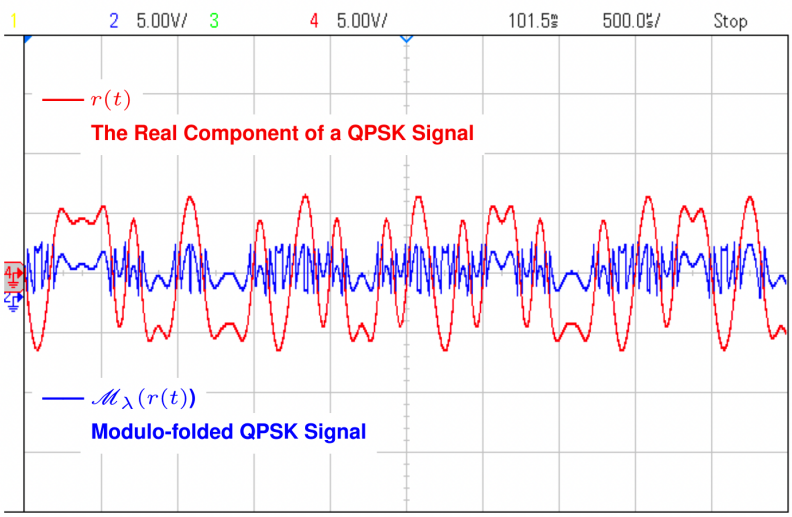


Figure 5.3 Hardware output of $\mathscr{M}_\lambda$–ADC (cf. Fig. 5.15) used in the Unlimited Sensing Framework [1–4]. An oscilloscope screen shot plotting the real component of a quadrature phase shift keying (QPSK) signal (ground truth, red) and its folded version (blue). The input signal with dynamic range 13.6V peak-to-peak ($\approx 7\lambda$) is folded into a 4.02V peak-to-peak signal.

form of information loss as one still needs to recover the original HDR signal from folded samples. This recovery is mathematically guaranteed provided that a *Shannon-Nyquist like* density criterion is satisfied. We recall the modulo sampling theorem here.

Theorem 2 (Unlimited Sampling Theorem [1, 2])**.** *Let $f(t)$ be a continuous-time function with maximum frequency Ω (rads/s). Then, a sufficient condition for recovery of $f(t)$ from its modulo samples (up to an additive constant) taken every T seconds apart is $T \leqslant 1/(2\Omega e)$ where e is Euler's number.*

If this sufficient condition is satisfied, the recovery algorithms can be then used to recover the unfolded signal $f(t)$. Subsequently, the samples of the original HDR input are recovered by the unlimited sampling algorithm (US-Alg) accompanying the theorem. A hardware example of modulo acquisition of a QPSK signal is shown in Fig. 5.3. This is further used in Section 5.5.1 to provide a proof-of-concept demonstration. We refer to Fig. 5.12 and Fig. 5.15 for

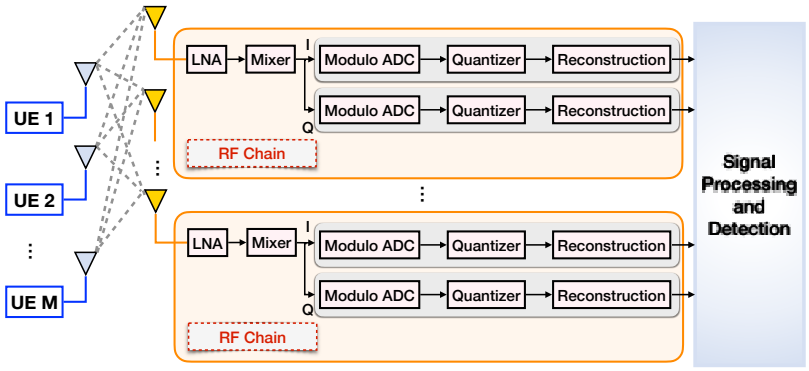


Figure 5.4 Proposed M–MIMO uplink system architecture with $\mathscr{M}_\lambda$–ADC in each RF chain.

a first look at our hardware experiments. Apart from US-Alg, a "Fourier-Prony" method operating at much lower sampling rates was introduced in [4,141]. Extensive experiments via a prototype $\mathscr{M}_\lambda$–ADC validated the empirical robustness of this approach while serving as the first hardware demonstration of USF. It was shown that signals as large as $\approx 25\lambda$ can be recovered via USF. With some side-information, *e.g.* the knowledge of a certain number of unfolded samples, $f_s > f_{\text{Nyquist}}$ also suffices recovery [146], but uses a different recovery mechanism. Similarly, using multiple modulo thresholds [147] *e.g.* Gaussian integers, allows for a different reconstruction approach. For other signal models and recovery methods, we refer to [141,148–153].

To conclude, the main contribution is that we take a radically different approach to address the high power consumption and receiver ADC saturation *simultaneously*. Specifically, we propose λ–MIMO architecture which uses modulo ADCs ($\mathscr{M}_\lambda$–ADC) instead of a conventional ADC. Our work is inspired by the Unlimited Sampling Framework. $\mathscr{M}_\lambda$–ADC in the RF chain folds high dynamic range signals into low dynamic range modulo samples, thus alleviating the ADC saturation problem. At the same time, digitization of modulo signals results in high resolution quantization. In the novel λ–MIMO context, we discuss baseband signal reconstruction, detection and uplink achievable sum-rate performance. The key takeaways of our work include, (a) leveraging higher

signal-to-quantization noise ratio (SQNR), (b) detection and average uplink sum-rate performances comparable to a conventional, infinite-resolution ADC when using a 1-2 bit $\mathscr{M}_\lambda$–ADC. This enables higher order modulation schemes *e.g.* 1024 QAM that was previously seemingly impossible, (c) superior trade-off between energy efficiency and bit budget thus resulting in higher power efficiency. Numerical simulations and modulo ADC based hardware experiments corroborate our theory and reinforce the clear benefits of λ–MIMO approach.

To conclude, the main contribution is that we consider the USF enabled full-duplex system to SI cancellation in fully digital domain. We show that by injecting folding non-linearities in the sensing pipeline, one can not only suppress self-interference but also recover the SoI. This approach leads to novel design of the receiver architecture that is complemented by a modulo-domain channel estimation method. Numerical experiments show that the USF enabled receiver structure can achieve up to 40 dB digital SIC by using as few as 4-bits per sample. Our method outperforms the previous approach based on adaptive filters when it comes to SoI reconstruction, detection, and digital SIC performance.

5.2 System model

In this section, we revisit the transmitter, channel, and receiver models, which will help us build the receiver architecture of the λ–MIMO system. As shown in Fig. 5.5, we assume that there are M single-antenna users and one BS equipped with N receive antenna elements (with $N \gg M$) in one cell. Note that the users transmit the information to the BS in the same time-frequency band. We denote $\lfloor \mathbb{K} \rceil$, $\lfloor \mathbb{M} \rceil$, $\lfloor \mathbb{N} \rceil$, and $\lfloor \mathbb{D} \rceil$ as the set of indices of subcarriers, users, receive antennas, and channel taps, respectively.

5.2.1 Transmitter

At each user location, we revisit the transmitter structure. To demonstrate the applicability of the $\mathscr{M}_\lambda$–ADC in communication systems, we consider two popular scenarios, namely, NBSC and wideband MIMO–OFDM systems.

Narrowband Single-Carrier Signal Model

Let us begin our discussion with the definition of the baseband signal denoted by $x_{m,\mathsf{b}}(t)$. Starting with an upsampled sequence $x^{\uparrow}_{m,\mathsf{b}}[k]$, filtering the same

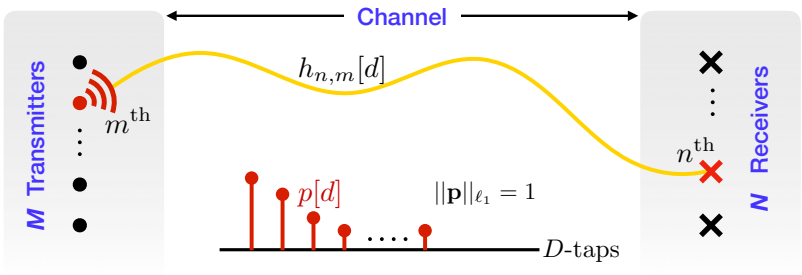


Figure 5.5 Channel model between m^{th} user and n^{th} antenna in the BS.

with a pulse-shaping filter[1] results in the continuous-time signal,

$$\underbrace{x^{\uparrow}_{m,\mathsf{b}}[k] \longrightarrow \boxed{\psi}}_{\text{Pulse-Shaping}} \longrightarrow \underbrace{x_{m,\mathsf{b}}(t) \stackrel{\text{def}}{=} \left(x^{\uparrow}_{m,\mathsf{b}}[k] \underline{*}_T \psi\right)(t)}_{\text{Baseband Signal}}. \tag{5.2}$$

There on, given the same for I/Q channels, denoted by $x^{\mathsf{I}}_{m,\mathsf{b}}(t)$ and $x^{\mathsf{Q}}_{m,\mathsf{b}}(t)$, respectively, we generate the passband signal by modulating the baseband sources,

$$\begin{aligned} x_{m,\mathsf{p}}(t) &= \begin{bmatrix} \cos(\omega_c t) & \\ & -\sin(\omega_c t) \end{bmatrix} \begin{bmatrix} x^{\mathsf{Q}}_{m,\mathsf{b}}(t) \\ x^{\mathsf{I}}_{m,\mathsf{b}}(t) \end{bmatrix} \equiv \mathbf{M}_{\omega_c} \mathbf{x}_m \\ &= \sqrt{2}\,\mathrm{Re}\left(x_{m,\mathsf{b}}(t)\, e^{\jmath\omega_c t}\right), \end{aligned} \tag{5.3}$$

where $\mathbf{M}_{\omega_c}$ is the modulation matrix parameterized by the carrier frequency $\omega_{\mathrm{c}} = 2\pi f_{\mathrm{c}}$ (rads/sec).

Wideband OFDM Signal Model (MIMO–OFDM)

The baseband OFDM signal is denoted by $x'_{m,\mathsf{b}}(t)$. Starting with a modulated complex sequence of symbols $\widehat{x}_{m,\mathsf{b}}[k]$ to be transmitted using OFDM. Its vector representation is $\widehat{\mathbf{x}}_{m,\mathsf{b}} = [\widehat{x}_{m,\mathsf{b}}[1], \ldots, \widehat{x}_{m,\mathsf{b}}[K]]^{\top}$. By using the K-point inverse discrete Fourier transformation (IDFT), the sequence of the baseband OFDM symbol $x'_{m,\mathsf{b}}[k]$ with vector expression $\mathbf{x}'_{m,\mathsf{b}} = [x'_{m,\mathsf{b}}[1], \ldots, x'_{m,\mathsf{b}}[K]]^{\top}$ is given by

$$\mathbf{x}'_{m,\mathsf{b}} = \mathbf{F}^{H}_{\mathsf{DFT}} \widehat{\mathbf{x}}_{m,\mathsf{b}}, \tag{5.4}$$

where $\mathbf{F}_{\mathsf{DFT}} \in \mathbb{C}^{K\times K}$ is the normalized discrete Fourier transform (DFT) matrix with element,

$$[\mathbf{F}_{\mathsf{DFT}}]_{i,j} = \frac{1}{\sqrt{K}} e^{\frac{-\jmath 2\pi(i-1)(j-1)}{K}}. \tag{5.5}$$

[1]The pulse-shaping filter can be the raised-cosine Nyquist pulse [154], with excess bandwidth $\alpha, 0 \leqslant \alpha \leqslant 1$, given by

$$f_{\mathsf{RC}}(t) = \mathrm{sinc}\left(\frac{t}{T_{\mathsf{rep}}}\right) \frac{\cos(\pi\alpha t/T_{\mathsf{rep}})}{1 - 4\alpha^2 t^2/T^2_{\mathsf{rep}}}, \tag{5.1}$$

where $T_{\mathsf{rep}} = 1/(2\Omega)$ is the symbol repetition rate.

There on, a cyclic prefix with length N_{cp} is prepended to the sequence $x'_{m,\mathsf{b}}[k]$ to compensate for the multipath effect, where $x'_{m,\mathsf{b}}[k] = x'_{m,\mathsf{b}}[K+k]$ for $k = -(N_{\mathsf{cp}}-1), ..., 0$. We then upsample $x'_{m,\mathsf{b}}[k]$ resulting in the upsampled sequence $x'^{\uparrow}_{m,\mathsf{b}}[k]$. Subsequently, it is filtered with a pulse-shaping filter leading to the continuous-time OFDM signal,

$$\underbrace{x'^{\uparrow}_{m,\mathsf{b}}[k] \longrightarrow \boxed{\psi}}_{\text{Pulse-Shaping}} \longrightarrow \underbrace{x'_{m,\mathsf{b}}(t) \stackrel{\text{def}}{=} \left(x'^{\uparrow}_{m,\mathsf{b}}[k] \underline{*}_T \psi\right)(t)}_{\text{Baseband OFDM Signal}}. \tag{5.6}$$

Finally, we generate the passband signal by modulating the baseband I/Q channels, denoted by $x'^{\mathsf{I}}_{m,\mathsf{b}}(t)$ and $x'^{\mathsf{Q}}_{m,\mathsf{b}}(t)$, respectively

$$\begin{aligned} x'_{m,\mathsf{p}}(t) &= \begin{bmatrix} \cos(\omega_c t) & \\ & -\sin(\omega_c t) \end{bmatrix} \begin{bmatrix} x'^{\mathsf{Q}}_{m,\mathsf{b}}(t) \\ x'^{\mathsf{I}}_{m,\mathsf{b}}(t) \end{bmatrix} \equiv \mathbf{M}_{\omega_c} \mathbf{x}'_m \\ &= \sqrt{2}\,\mathrm{Re}\left(x'_{m,\mathsf{b}}(t) e^{\jmath\omega_c t}\right). \end{aligned} \tag{5.7}$$

Therefore, we can generalize the passband signals for the two baseband signals

$$x_m(t) = \begin{cases} \sqrt{2}\,\mathrm{Re}\left(x_{m,\mathsf{b}}(t) e^{\jmath\omega_c t}\right), & \text{if single-carrier} \\ \sqrt{2}\,\mathrm{Re}\left(x'_{m,\mathsf{b}}(t) e^{\jmath\omega_c t}\right), & \text{if OFDM} \end{cases}, \tag{5.8}$$

which is transmitted at each user location into the wireless channel.

5.2.2 Channel Model

We revisit the wideband MIMO–OFDM model. It can be simplified to the narrowband channel model with single path. The channel of the wideband MIMO–OFDM system encapsulates multipath propagation and hence, it is modeled by D taps ($D \leqslant N_{\mathsf{cp}}$). As shown in Fig. 5.5, the d^{th} path impulse response $h_{n,m}[d]$ between the m^{th} user and the n^{th} antenna element at the BS is the product of a large-scale fading coefficient $\sqrt{\eta_m}$ and a small-scale fading parameter $g_{n,m}[d]$ and is given by,

$$h_{n,m}[d] = \sqrt{\eta_m}\, g_{n,m}[d]. \tag{5.9}$$

Here, we assume that the large fading coefficients are the same across the antenna array at the BS, therefore we only have one coefficient η_m between the m^{th} user and all antenna elements.

Following the conventional propagation model[2] [83], we assume that the channel fading follows an i.i.d. Rayleigh distribution, or $g_{n,m}[d] \sim \mathcal{CN}(0, p[d])$. Hence, the variance of the small-scale fading in each path is its power delay profile, or $\mathscr{E}[|g_{n,m}[d]|^2] = p[d]$ that satisfies the relation,

$$\|\mathbf{p}\|_{\ell_1} = 1, \quad \mathbf{p} = \begin{bmatrix} p\,[1] & \cdots & p\,[D] \end{bmatrix}^\top, \quad p\,[d] > 0. \tag{5.10}$$

In the NBSC M–MIMO system, since the multiple paths are not resolvable in narrow bandwidth, a single path (frequency-flat) model is used, thus D reduces to 1. The frequency-flat channel is expressed as $\mathbf{H} \in \mathbb{C}^{N \times M}$,

$$\mathbf{H} = \underbrace{\begin{bmatrix} g_{1,1} & \cdots & g_{1,N} \\ \vdots & \ddots & \vdots \\ g_{M,1} & \cdots & g_{M,N} \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \sqrt{\eta_1} & & \\ & \ddots & \\ & & \sqrt{\eta_M} \end{bmatrix}}_{\mathbf{D}_{\boldsymbol{\eta}}} \equiv \mathbf{G}\mathbf{D}_{\boldsymbol{\eta}},$$

where $g_{n,m} \sim \mathcal{CN}\,(0, 1)$. For both channels, the channel state information (CSI) is assumed to be known at the receiver end. Channel estimation of the λ–MIMO occurs after reconstructing the unfolded signals, allowing classic channel estimation methods to be implemented on the recovered unfolded digital signals without modification. Additionally, if the thermal noise is small enough, the modulo domain channel estimation method [157] can be used to reduce the computational complexity of the whole system.

[2] Adopting large antenna arrays can result in different channel characteristics (*e.g.* spherical wavefronts [155] and spatial non-stationarity [156]). For simplicity, we utilize the conventional channel model in this work.

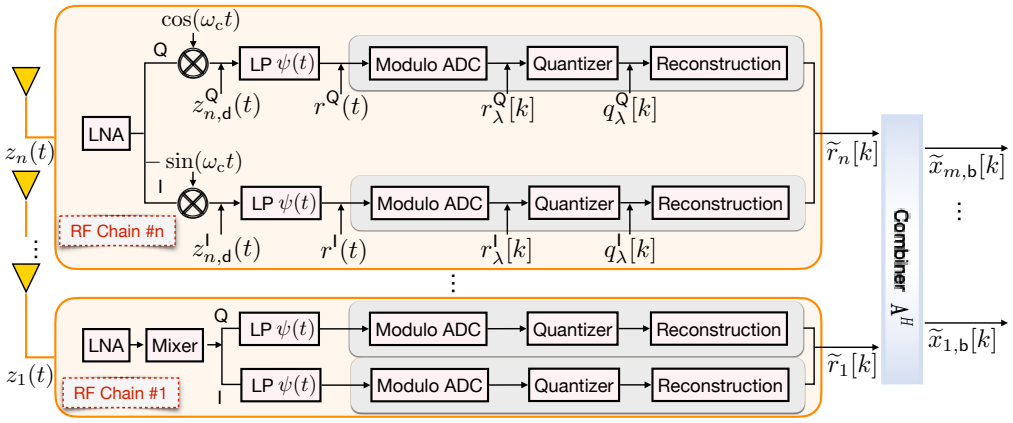


Figure 5.6 Receiver architectures with $\mathscr{M}_\lambda$–ADCs for single-carrier system.

5.2.3 Receiver

The received signal at the n^{th} antenna is given by

$$z_n(t) \overset{\text{def}}{=} \sum_{m \in \lfloor \mathbb{M} \rceil} \sqrt{p_{\text{u}}} (h_{n,m} \underset{T}{*} x_m)(t) + \epsilon_n(t), \tag{5.11}$$

where p_{u} is the average transmit power at the user, and $\epsilon_n(t) \sim \mathcal{CN}(0,1), \forall n \in \lfloor \mathbb{N} \rceil$ is the additive Gaussian noise at the receiver. Subsequently, the received signal at each antenna enters the RF chain and is processed using the $\mathscr{M}_\lambda$–ADC receiver. The novel RF chains and signal processing blocks for both single-carrier and multi-carrier (OFDM) signals are shown in Fig. 5.6 and 5.7, respectively.

At the receiver side, the continuous-time signal $z_n(t)$ is firstly amplified by LNA and demodulated to the baseband by means of coherent detection. This is facilitated by,

$$\begin{bmatrix} z^{\mathsf{Q}}_{n,\mathsf{d}}(t) \\ z^{\mathsf{I}}_{n,\mathsf{d}}(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_c t) & \\ & -\sin(\omega_c t) \end{bmatrix} \begin{bmatrix} z_n(t) \\ z_n(t) \end{bmatrix} \equiv \mathbf{M}_{\omega_c} \mathbf{z}_n. \tag{5.12}$$

Subsequently, the quadrature and in-phase signals are low-pass filtered with $\psi(t)$; this eliminates the images located outside from the baseband. Quantita-

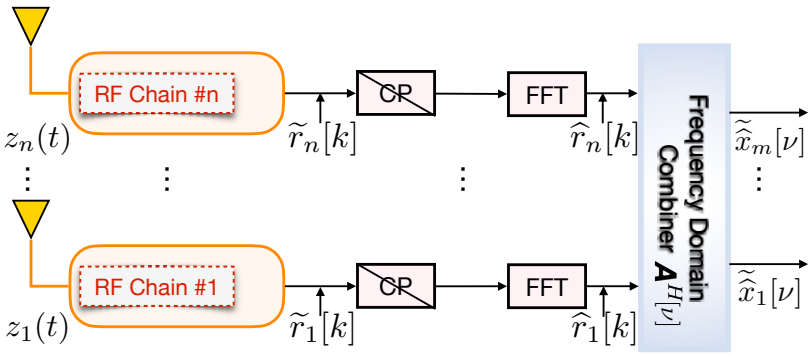


Figure 5.7 Receiver architectures with $\mathscr{M}_\lambda$–ADCs for multi-carrier (OFDM) system.

tively, we have,

$$r_n^{\mathsf{Q}}(t) = \langle z_{n,\mathsf{d}}^{\mathsf{Q}}, \overline{\psi} \rangle \quad \text{and} \quad r_n^{\mathsf{I}}(t) = \langle z_{n,\mathsf{d}}^{\mathsf{I}}, \overline{\psi} \rangle, \tag{5.13}$$

where $\overline{\psi}(t) = \psi(-t)$ (time-reversal). Then, the baseband signal is acquired by the $\mathscr{M}_\lambda$–ADC to reduce power consumption and prevent ADC saturation, as detailed in Section 5.3.

Single-Carrier Signal Detection

It is well known that maximum likelihood detector achieves the optimal performance in theory [158]. However, the computational complexity increases exponentially with user number M. Thus, a commonly adopted approach is to use the maximum ratio combining (MRC), zero forcing (ZF), or minimum mean-squared error (MMSE) linear combiners [11]. This is used to separate the streams from each uplink user. The matrix form of the combiner is given by

$$\widetilde{\mathbf{x}}_{\mathsf{b}} = \mathbf{A}^H \mathbf{r}, \tag{5.14}$$

where $\widetilde{\mathbf{x}}_{\mathsf{b}} \in \mathbb{C}^{M\times 1}$ is the estimated baseband signal for all M users, $\mathbf{r} \in \mathbb{C}^{N\times 1}$ is the recovered signal for all N antennas, and $\mathbf{A}^H$ is the combiner, *i.e.*

$$\mathbf{A} = \begin{cases} \mathbf{H}, & \text{for MRC} \\ \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}, & \text{for ZF} \\ \mathbf{H}(\mathbf{H}^H\mathbf{H} + \frac{\sigma_r^2}{\sigma_n^2}\mathbf{I})^{-1}, & \text{for MMSE}, \end{cases} \tag{5.15}$$

where σ_r^2 and σ_n^2 are the signal power and noise power, respectively.

OFDM Signal Detection

As shown in Fig. 5.7, we proceed by undoing the cyclic prefix (CP). Then, the samples at n^{th} antenna are transformed to frequency domain by fast Fourier transform (FFT), which yields the frequency-domain signal $\widehat{\mathbf{r}}_n$

$$\widehat{\mathbf{r}}_n = \mathbf{F}_{\mathsf{DFT}}\mathbf{r}_n. \tag{5.16}$$

The received signal at the ν^{th} subcarrier can also be written as (5.17)

$$[\widehat{\mathbf{r}}_n]_\nu \equiv \widehat{r}_n[\nu] = \sum_{m\in\lfloor\mathbb{M}\rceil} \sqrt{p_{\text{u}}\eta_m}\,\widehat{g}_{n,m}[\nu]\widehat{x}_m[\nu] + \widehat{\epsilon}_n[\nu], \quad \forall n, \tag{5.17}$$

where $\widehat{\mathbf{g}}_{n,m} \in \mathbb{C}^K$, $[\widehat{\mathbf{g}}_{n,m}]_\nu = \widehat{g}_{n,m}[\nu]$ is the frequency response of the small-scale fading between the n^{th} antenna and m^{th} user,

$$\widehat{\mathbf{g}}_{n,m} = \mathbf{F}_{\mathsf{DFT}}\mathbf{g}_{n,m}^{\mathsf{pad}}. \tag{5.18}$$

In (5.18), $\mathbf{g}_{n,m}^{\mathsf{pad}} = [\mathbf{g}_{n,m}^\top \mid \mathbf{0}_{1\times(K-D)}]^\top$, and $\widehat{\epsilon}_n[\nu] \sim \mathcal{CN}(0,1)$ is the unitary Fourier transform of the noise vector. Therefore, the transmit-receive relation at the ν^{th} subcarrier can be written in matrix form as,

$$\widehat{\mathbf{r}}[\nu] = \sqrt{p_{\text{u}}}\widehat{\mathbf{H}}[\nu]\widehat{\mathbf{x}}[\nu] + \widehat{\boldsymbol{\epsilon}}[\nu], \tag{5.19}$$

where $[\widehat{\mathbf{H}}[\nu]]_{n,m} = \eta_m\widehat{g}_{n,m}[\nu]$ is the frequency response of the channel at the ν^{th} subcarrier. Subsequently, frequency domain combiner given in (5.20) is

adopted on each subcarrier, and the estimated data $\widetilde{\widehat{\mathbf{x}}}[\nu] = \widehat{\mathbf{A}}^H[\nu]\widehat{\mathbf{r}}[\nu]$ from uplink users at the ν^{th} subcarrier is obtained [134], where

$$\widehat{\mathbf{A}}[\nu] = \begin{cases} \widehat{\mathbf{H}}[\nu], & \text{for MRC} \\ \widehat{\mathbf{H}}[\nu](\widehat{\mathbf{H}}^H[\nu]\widehat{\mathbf{H}}[\nu])^{-1}, & \text{for ZF} \\ \widehat{\mathbf{H}}[\nu](\widehat{\mathbf{H}}^H[\nu]\widehat{\mathbf{H}}[\nu] + \frac{\sigma_r^2}{\sigma_n^2}\mathbf{I})^{-1}, & \text{for MMSE .} \end{cases} \tag{5.20}$$

5.3 Receiver Design with $\mathscr{M}_\lambda$-ADC

Compared to the conventional setup, our newly designed receiver deploys a $\mathscr{M}_\lambda$–ADC to simultaneously reduce the power consumption and prevent potential ADC saturation caused by limited dynamic range. As shown in Fig. 5.8, an arbitrary baseband signal $r(t)$ (continuous-time) is acquired by the $\mathscr{M}_\lambda$–ADC for both the I/Q channels. This is done by implementing the modulo mapping in hardware [4]. Formally, the centered modulo operator is defined by,

$$\mathscr{M}_\lambda : r \mapsto 2\lambda \left(\left[\!\!\left[\frac{r}{2\lambda} + \frac{1}{2} \right]\!\!\right] - \frac{1}{2} \right), \tag{5.21}$$

where $\lambda > 0$ is the folding threshold and $[\![r]\!] \stackrel{\text{def}}{=} r - \lfloor r \rfloor$ is the fractional part of r. If $|r(t)| < \lambda$, the $\mathscr{M}_\lambda$–ADC works the same as the conventional ADC.

Clearly, the action of $\mathscr{M}_\lambda$–ADC on $r(t)$ results in folding $r\,(t)$ into low dynamic range signal that is upper bounded (in absolute sense), by λ; this also avoids the need for any VGA or AGC in the pipeline, thus facilitating lower complexity hardware. Finally, the continuous-time modulo-folded signal is sampled using impulse-train

$$\otimes_{kT} \stackrel{\text{def}}{=} \sum\nolimits_{k\in\mathbb{Z}} \delta(t - kT), \quad T > 0$$

where T is the sampling interval. This results in modulo samples $r_\lambda\,[k]$,

$$r_\lambda\,[k] \stackrel{\text{def}}{=} r_\lambda(kT) = \mathscr{M}_\lambda(r(kT)), \quad k \in \mathbb{Z}. \tag{5.22}$$

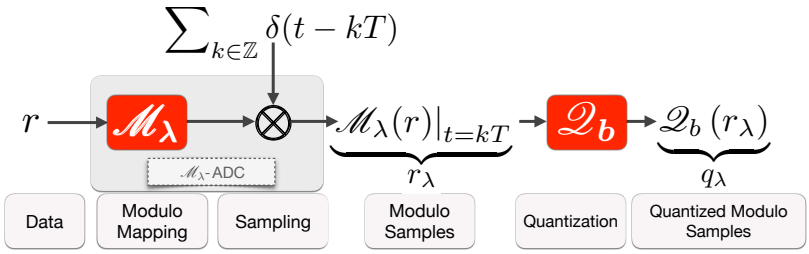


Figure 5.8 The architecture of the $\mathscr{M}_\lambda$–ADC and quantizer.

5.3.1 Discrete-Time to Digital Modulo Samples

Accounting for finite precision operations in practice, the modulo samples in (5.22) are quantized. Given a bit budget of b bits, each modulo sample is rounded to the closest element in the set

$$\Upsilon_{b,\lambda} = \left\{ \pm \frac{(2n+1)}{2^b} \lambda \mid n \in \left\{0, \ldots, 2^{b-1} - 1\right\} \right\}.$$

Hence, the "digital version" of modulo samples is written as

$$q_\lambda[k] = \mathscr{Q}_b(r_\lambda[k]), \quad q_\lambda[k] \in \Upsilon_{b,\lambda}, \tag{5.23}$$

where $\mathscr{Q}_b(\cdot)$ is the quantization mapping. Specifically, the number of the mid-rise quantization levels is $B = 2^b$ with the step size $q_0 = 2^{-b}\rho_\lambda$ where $\rho_\lambda = 2\lambda$ is the dynamic range of the $\mathscr{M}_\lambda$–ADC.

To measure the distortion caused by the quantization, we define the quantization noise as $e_{\mathsf{q}}[k] = r_\lambda[k] - \mathscr{Q}_b(r_\lambda[k])$. As shown in Fig. 5.9, the synergistic interplay between the bit budget b and dynamic range ρ_λ decides the quantization noise e_{q}. Given the same bit budget, it is clear that the quantization of modulo samples has *smaller* quantization noise due to smaller dynamic range, determined by λ. The implication is that when compared to a conventional ADC, the $\mathscr{M}_\lambda$–ADC results in higher resolution of quantized samples. This advantage of the $\mathscr{M}_\lambda$–ADC over quantization on the normal samples *i.e.* $r(kT)$ is leveraged in the subsequent reconstruction.

Remark 11. *In practice, given the power consumption budget P_{b} and antenna number N, we can calculate the exact needed bit resolution b as*

$$P_{\mathsf{b}} = c_0 N 2^b + c_1 \Rightarrow b = \log_2\left(\frac{P_{\mathsf{b}} - c_1}{c_0 N}\right)$$

with $c_0 = 10^{-4}$ Watt and $c_1 = 0.02$ Watt [159].
To determine the value of λ, we can obtain the suitable value of λ from the algorithmic and hardware perspectives. From the algorithmic viewpoint, with Lemma 2, we have

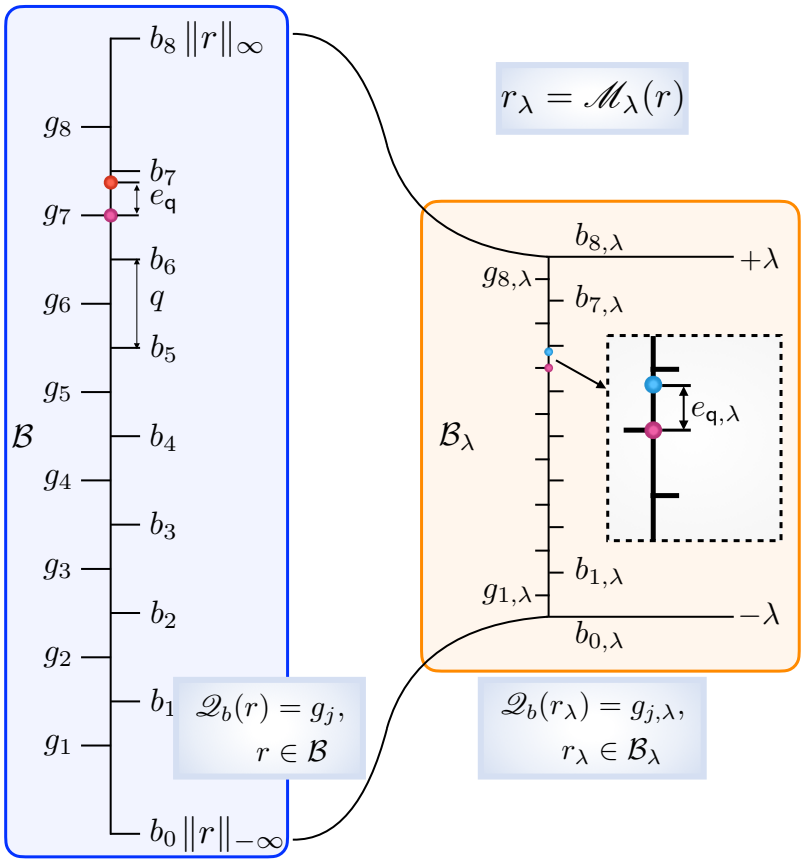


Figure 5.9 An illustration of quantization process of normal samples and modulo-folded samples with bit resolution $b = 3$.

the following corollary:

$$(T\Omega e)^L \|r\|_{\ell_\infty(\mathbb{R})} < \lambda,$$

where $(T\Omega e)^L \|r\|_{\ell_\infty(\mathbb{R})}$ is a function of sampling interval, T, and finite difference order, L. In practice, a lower order of finite difference order, L is desirable, because it makes the reconstruction numerically stable. Consequently, for a fixed L, and T, we can obtain the value of λ for a given β_r. From the viewpoint of hardware, since the equivalent bit of $\mathscr{M}_\lambda$–ADC, b_λ, in terms of the bit of the conventional ADC, b, is $b_\lambda = b + \log_2 \zeta$, thus we can set the λ to $\lambda = \zeta\rho/2$ with normal setting of DR, ρ, in conventional ADC.

5.3.2 Recovery from Modulo Samples: The Unlimited Sampling Algorithm

Reconstruction of the HDR input signal, $r(t)$, from its modulo samples entails the "inversion" of the modulo non-linearity $\mathscr{M}_\lambda(\cdot)$. This is an ill-posed problem due to the many-to-one nature of the $\mathscr{M}_\lambda(\cdot)$. The key to unfolding modulo samples is the observation that the modulo and the finite-difference operators commute in a certain sense [1]. This allows us to extract the higher-order differences of r from its modulo samples. Then, reconstruction boils down to the problem of stably inverting the difference-operator. Next, we focus on these two aspects separately and for the case of non-quantized measurements (5.22). This is solely for the ease of exposition as the same recovery procedure applies to the case of quantized modulo samples in (5.23).

Accessing Higher-Order Differences. Let $(\Delta r)[k] \stackrel{\text{def}}{=} r[k+1] - r[k]$ denote the first-order difference operator; higher-order differences recursively defined by $\Delta^L r \stackrel{\text{def}}{=} \Delta^{L-1}(\Delta r), L > 1$. According to the modulo decomposition property (see Fig. 4 in [1]), we have, $r(t) = \mathscr{M}_\lambda(r(t)) + \varepsilon_r(t)$ where $\varepsilon_r(t) \in 2\lambda\mathbb{Z}$ is a piecewise-constant function. The same property also applies to the point-wise samples. On one hand, since $\varepsilon_r(kT) \in 2\lambda\mathbb{Z}$, we have that $\Delta^L \varepsilon_r(kT) \in 2\lambda\mathbb{Z}$ and hence, $\mathscr{M}_\lambda(\Delta^L \varepsilon_r(kT)) = 0$. On the other hand, for Ω-bandlimited functions, *i.e.* $r \in \mathcal{B}_\Omega$, we note that Δ^L results in amplitude shrinkage in the oversampled regime. This is made precise via the following Lemma.

Lemma 2 (Difference-Derivative Inequality [1, 2])**.** *For any $r(t) \in \mathcal{B}_\Omega \cap L_\infty(\mathbb{C})$, its samples $r[n] \stackrel{\text{def}}{=} r(nT)$ satisfy, $\|\Delta^L r\|_{\ell_\infty(\mathbb{R})} \leqslant (T\Omega e)^L \|r\|_{L_\infty(\mathbb{R})}$ where e is the Euler's number.*

Clearly, choosing $T < 1/(\Omega e)$ amounts to controlling the factor $(T\Omega e)^L$. More precisely, choosing $L^\star \geqslant \left\lceil \frac{\log \lambda - \beta_r}{\log(T\Omega e)} \right\rceil$ with $\beta_r \geqslant 2\lambda\mathbb{Z}$ and $\beta_r \geqslant \|r\|_{L_\infty(\mathbb{R})}$ also ensures that $\|\Delta^{L^\star} r\|_{\ell_\infty(\mathbb{R})} \leqslant \lambda$. According to Proposition 2 in [1], we have, $\mathscr{M}_\lambda(\Delta^L(r)) = \mathscr{M}_\lambda(\Delta^L(\mathscr{M}_\lambda(r)))$ and hence, whenever $L \geqslant L^\star$, oversampling or choosing $T\Omega e < 1$ implies that $\Delta^L r = \mathscr{M}_\lambda(\Delta^L r_\lambda)$. In summary, oversampling $r(t) \in \mathcal{B}_\Omega$ allows us to access $\Delta^L r$ from its modulo samples $r_\lambda[k]$.

Stable Inversion of Δ^L. Instead of inverting $\left(\Delta^L r\right)[k] \in \mathbb{R}$, the recovery algorithm in [1,2] capitalizes the piece-wise constant structure of the residue $\varepsilon_r(t)$, that is, $\varepsilon_r(kT) \in 2\lambda\mathbb{Z}$ and hence, $\Delta^L \varepsilon_r(kT) \in 2\lambda\mathbb{Z}$. This narrows down the range of $\varepsilon_r(kT)$ and hence has a stabilizing effect during inversion. We initialize the sequence $s^{\{0\}}[k] = \left(\mathscr{M}_\lambda(\Delta^L r_\lambda) - \Delta^L r_\lambda\right)[k] \equiv \Delta^L \varepsilon_r(kT) \in 2\lambda\mathbb{Z}$. Next, $\Delta^{L-1}\varepsilon_r(kT)$ is obtained from $\Delta^L \varepsilon_r(kT)$ by invoking the anti-difference operator $\mathsf{S} = \Delta^{-1}$ and rounded-off to $2\lambda\mathbb{Z}$. Note that the unknown constant-of-integration in the kernel of Δ must be estimated for unique reconstruction. It has been shown in [1,2] that for $r \in \mathcal{B}_\Omega$, this unknown constant can be estimated via,

$$\kappa^{\{l\}} = \left\lfloor \frac{\mathsf{S}\left(s^{\{l+1\}}\right)[k] - \mathsf{S}\left(s^{\{l+1\}}\right)[J_\lambda + 1]}{12\beta_r} + \frac{1}{2} \right\rfloor,$$

where $J_\lambda = 6\beta_r/\lambda, l = 0, \ldots, L-2$. This completes the estimation of $s^{\{1\}}[k]$ and more generally,

$$s^{\{l+1\}}[k] = 2\lambda \left\lceil \frac{1}{2} \left\lfloor \frac{\mathsf{S}\left(s^{\{l+1\}}\right)[k]}{\lambda} \right\rfloor \right\rceil + 2\lambda\kappa^{\{l\}}, \quad l = 0, \ldots, L-2.$$

Finally, we obtain the recovered samples $\widetilde{r}[k]$ up to an unknown constant,

$$\widetilde{r}[k] = \mathsf{S}\left(s^{\{L-1\}}\right)[k] + r_\lambda[k] + 2\lambda\mathbb{Z} = \widetilde{\varepsilon}_r + r_\lambda[k]. \tag{5.24}$$

The reconstruction approach is summarized in Algorithm 8 which also works with quantized measurements [1]. Compared with conventional ADC, adopting $\mathscr{M}_\lambda$–ADC has this additional reconstruction operation. It has a complexity of the order of $\mathcal{O}(LN_\mathsf{s})$, where N_s is the length of the samples. We have explained the reconstruction in detail in chapter 2.1.

Algorithm 8: Recovery from Noisy Modulo Communication Samples

Input: $q_\lambda[k]$ and $2\lambda\mathbb{Z} \ni \beta_r \geqslant \|r\|_\infty$.

Output: $\widetilde{r}[k] \approx r(kT)$.

1) Compute $L = \left\lceil \frac{\log\lambda - \log\beta_r}{\log(\mathrm{T}\Omega e)} \right\rceil$.

2) Compute $\left(\Delta^L q_\lambda\right)[k]$.

3) Compute $\left(\Delta^L \varepsilon_r\right)[k] = \left(\mathscr{M}_\lambda\left(\Delta^L q_\lambda\right) - \Delta^L q_\lambda\right)[k]$.

Set $s^{\{0\}}[k] = \left(\Delta^L \varepsilon_r\right)[k]$

for $l = 0 : L-2$ **do**

i) $s^{\{l+1\}}[k] = \left(\mathsf{S} s^{\{l\}}\right)[k]$

ii) $s^{\{l+1\}} = 2\lambda \left\lceil \frac{\left\lfloor s^{\{l+1\}}/\lambda \right\rfloor}{2} \right\rceil$ (rounding to $2\lambda\mathbb{Z}$)

iii) Compute $\kappa^{\{l\}} = \left\lfloor \frac{\left(\mathsf{S}^2\Delta^l\varepsilon_r\right)[1] - \left(\mathsf{S}^2\Delta^l\varepsilon_r\right)[J_\lambda+1]}{12\beta_r} + \frac{1}{2} \right\rfloor$

iv) $s^{\{l+1\}}[k] = s^{\{l+1\}}[k] + 2\lambda\kappa^{\{l\}}$

end

5) $\widetilde{r}[k] = \mathsf{S}\left(s_{\{L-1\}}\right)[k] + q_\lambda[k] + 2v\lambda, v \in \mathbb{Z}$

5.4 Achievable sum-rate analysis

To show the potential impact of the lower quantization noise achieved by $\mathscr{M}_\lambda$–ADC on the performance of communication, we analyze the SQNR and achievable sum-rate for both MRC and ZF combiners. Our analysis shows that with as few as 1- or 2-bits, the λ–MIMO architecture achieves very similar sum-rate performance when compared with ∞-bit resolution, conventional ADC.

To be able to compare with the benchmark (*i.e.* conventional ADC), and clearly show the gain of $\mathscr{M}_\lambda$–ADC on the quality of quantization, we assume that the dynamic range ρ_λ of the conventional ADC matches the signal r. Hence, the clipping error does not exist. Additionally, the quality of the quantization is measured by the signal to quantization noise ratio (SQNR) defined by [160]

$$\mathsf{SQNR} = 10 \log_{10} \frac{\sigma_r^2}{\sigma_{\mathsf{q}}^2}, \tag{5.25}$$

where σ_r^2 is the signal power with the assumption that signal r has zero mean and σ_{q}^2 is the quantization error power.

As shown in Fig. 5.9, to calculate σ_{q}^2 at a given time instant during quantization, we first partition ρ_λ into $J = B + 1$ subsets expressed as $\rho_{\lambda,j}$, $j \in \lfloor \mathbb{J} \rceil$. Each partition region is then bounded by the boundary values b_j, *i.e.* $\rho_{\lambda,j} = [b_{j-1}, b_j)$. Specifically, the samples inside the same region $\rho_{\lambda,j}$ are mapped to the same quantization value g_j based on the mapping (5.23). Therefore, the variance σ_{q}^2 of the quantization error is calculated as [160], [161]

$$\begin{aligned} \sigma_{\mathsf{q}}^2 &= \mathscr{E}\left\{ |r - \mathscr{Q}_b(r)|^2 \right\} \\ &= \sum_{j \in \lfloor \mathbb{J} \rceil} P\left(r \in \rho_{\lambda,j}\right) \int_{b_{j-1}}^{b_j} (r - g_j)^2 \, p\left(r \mid r \in \rho_{\lambda,j}\right) dy, \end{aligned} \tag{5.26}$$

where

— $p(r)$ is the probability density function (pdf) of r.

— $P\left(r \in \rho_{\lambda,j}\right) = \int_{\rho_{\lambda,j}} p(r) dr$ is the probability when signal r is inside the $\rho_{\lambda,j}$

region *i.e.* $r \in \rho_{\lambda,j}$.

— $p\left(r \mid r \in \rho_{\lambda,j}\right) = p(r)/P(r \in \rho_{\lambda,j})$ is the conditional pdf of r in region $\rho_{\lambda,j}$.

In comparison, the quantization error of $\mathscr{M}_\lambda$–ADC is decreased to $\sigma^2_{\lambda,\mathsf{q}} = \zeta^2\sigma^2_{\mathsf{q}}$ due to modulo operation. The ratio ζ is defined by

$$\zeta = \frac{\lambda}{\|r\|_\infty}, \quad \|r\|_\infty = \max|r\left(t\right)|. \tag{5.27}$$

With the above, we derive the SQNRs for baseband sampled single carrier and OFDM signals as follows.

5.4.1 SQNR of Uniform Distribution Signal

The baseband sampled single-carrier signal is subject to uniform distribution [162] that $r[k] \sim \mathcal{U}(0,1)$, hence the variance of the sample $r[k]$ is $\sigma^2_{r,\mathsf{unif}} = \rho^2_\lambda/12$, and variance of quantization noise $\sigma^2_{\mathsf{q,unif}}$ is given by

$$\sigma^2_{\mathsf{q,unif}} = \frac{q_0^2}{12} = \sigma^2_{r,\mathsf{unif}}\, 2^{-2b}. \tag{5.28}$$

Subsequently, the SQNR of conventional ADC is

$$\mathsf{SQNR}_{\mathsf{unif}} = 10\log_{10}\frac{\sigma^2_{r,\mathsf{unif}}}{\sigma^2_{\mathsf{q,unif}}} = 6.02b\ (\mathrm{dB}). \tag{5.29}$$

Since the modulo-folded samples follow uniform distribution that $r_\lambda[k] \sim \mathcal{U}(0,\zeta^2)$, the quantization error of $\mathscr{M}_\lambda$–ADC is decreased to $\sigma^2_{\mathsf{q},\lambda} = \zeta^2\sigma^2_{\mathsf{q,unif}}$ with the SQNR given by

$$\mathsf{SQNR}_{\mathsf{unif},\lambda} = 10\log_{10}\frac{\sigma^2_{r,\mathsf{unif}}}{\sigma^2_{\mathsf{q},\lambda}} = 6.02b + 20\log_{10}(1/\zeta)\ (\mathrm{dB}). \tag{5.30}$$

Therefore, there is a $20\log_{10}(1/\zeta)$ dB enhancement on SQNR by adopting $\mathscr{M}_\lambda$–ADC. Note that the oversampling is required to guarantee the successful recon-

struction. In addition, ζ should satisfy the constraint

$$(T\Omega e)^L < \frac{\lambda}{\|r\|_{L_\infty}} = \zeta,$$

thus ζ cannot be 0, and the gain in SQNR is bounded by

$$\mathsf{SQNR}_{\mathsf{gain}} < 20\log_{10}\left(\frac{1}{(\pi\Omega e)^L}\right).$$

5.4.2 SQNR of Gaussian Distribution Signal

The OFDM signal is commonly assumed to be a centered, wide-sense stationary, ergodic Gaussian process [163], [164] that is $r[k] \sim \mathcal{N}(0,1)$, thus the variance of quantization noise $\sigma^2_{\mathsf{q,Gau}}$ is given by [160, 161, 165]

$$\sigma^2_{\mathsf{q,Gau}} = \varepsilon^2 \sigma^2_{r,\mathsf{Gau}} 2^{-2b}, \tag{5.31}$$

where

$$\varepsilon^2 = \frac{1}{12\sigma^2_{r,\mathsf{Gau}}}\left(\int_{-\infty}^{\infty} \sqrt[3]{p(r)}dr\right)^3. \tag{5.32}$$

Specifically, for the case when $r[k] \sim \mathcal{N}(0,1)$, we have $\varepsilon^2 = \sqrt{3}\pi/2$. Thus, the SQNR in dB of the conventional ADC is given by

$$\mathsf{SQNR}_{\mathsf{Gau}} = 10\log_{10}\left(\frac{\sigma^2_{r,\mathsf{Gau}}}{\sigma^2_{\mathsf{q,Gau,(m)}}}\right) \approx 6.02b - 4.35. \tag{5.33}$$

Modulo folding on the input distribution leads to a family of so-called "wrapped distributions" [166]. In particular, when the input signal is drawn from $\mathcal{N}(0,\sigma^2)$ and the modulo threshold λ is chosen to be a fraction of σ^2, empirically, we have found that the resulting density function for $r_\lambda[k]$ follows a uniform distribution. Assuming that $r_\lambda[k] \sim \mathcal{U}(0,\zeta^2)$, the quantization error of $\mathscr{M}_\lambda$–ADC shrinks to $\sigma^2_{\mathsf{q},\lambda} \approx \zeta^2\sigma^2_{\mathsf{q,unif}}$, and the SQNR in dB is given by

$$\mathsf{SQNR}_{\mathsf{Gau},\lambda} = 10\log_{10}\left(\frac{\sigma^2_{r,\mathsf{Gau}}}{\sigma^2_{\mathsf{q},\lambda}}\right) \approx 6.02b + 20\log_{10}(1/\zeta). \tag{5.34}$$

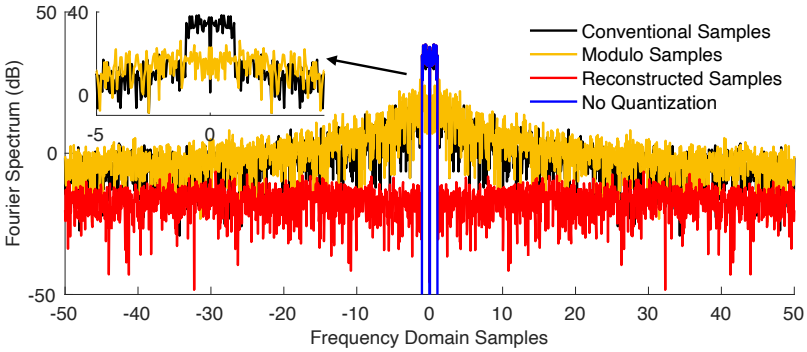


Figure 5.10 Fourier spectrum of r, and 4-bit quantized samples for conventional ADC and $\mathscr{M}_\lambda$–ADC with $\zeta = 0.1$.

In summary, when capitalizing on the $\mathscr{M}_\lambda$–ADC the resulting samples entail lower quantization noise as shown in Fig. 5.10. This in turn leads to a factor $20\log_{10}(1/\zeta)+10\log_{10}(\varepsilon^2)$ dB enhancement in the SQNR. For bandlimited functions $r(t)$ with maximum frequency Ω, from [1, 2] we know that $\left\|\Delta^L r\right\|_\infty \leqslant (T\Omega e)^L \|r\|_{L_\infty}$. Choosing $T < 1/(\Omega e)$ for an appropriate value of L entails that,

$$(T\Omega e)^L < \frac{\lambda}{\|r\|_{L_\infty}} = \zeta.$$

For example, choosing $\zeta \in \mathcal{U}(1/100, 1/10)$ we can see that SQNR (5.34) can be increased from 24.5 dB to 44.5 dB (in noiseless conditions).

5.4.3 Achievable sum-rate Analysis for $\mathscr{M}_\lambda$–ADC in MU λ–MIMO.

To be able to show the benefits of using $\mathscr{M}_\lambda$–ADC in communication systems, we analyze the achievable sum-rate of MU λ–MIMO architecture. The nonlinear nature of the quantization operator makes the analysis challenging. To overcome this aspect, we adopt the AQNM model, previously described in [112, 131, 167]. Consequently, the quantized signal $\mathbf{r}_\mathsf{q}$ is modeled by,

$$\mathbf{r}_\mathsf{q} = \gamma_\lambda \mathbf{r} + \boldsymbol{\epsilon}_\mathsf{q} = \gamma_\lambda \sqrt{p_\mathrm{u}}\mathbf{H}\mathbf{x} + \gamma_\lambda \boldsymbol{\epsilon} + \boldsymbol{\epsilon}_\mathsf{q}, \tag{5.35}$$

where p_{u} is the average transmitted power of each user, $\gamma_\lambda = 1 - \sigma^2_{\mathsf{q},\lambda}/\sigma^2_r$, $\boldsymbol{\epsilon}_{\mathsf{q}} \sim \mathcal{N}(0, (1-\gamma_\lambda)\gamma_\lambda\sigma^2_r)$ is the additive Gaussian quantization noise independent with $\mathbf{r}$. In this setting, the signal vector after the linear detection is given by

$$\widetilde{\mathbf{x}}_{\mathsf{b}} = \mathbf{A}^H \mathbf{r}_{\mathsf{q}}, \tag{5.36}$$

where $\mathbf{A}$ is given in (5.15). By substituting (5.36) into (5.35), we derive the following expression

$$\widetilde{\mathbf{x}}_{\mathsf{b}} = \gamma_\lambda \sqrt{p_{\mathrm{u}}} \mathbf{A}^H \mathbf{H}\mathbf{x} + \gamma_\lambda \mathbf{A}^H \boldsymbol{\epsilon} + \mathbf{A}^H \boldsymbol{\epsilon}_{\mathsf{q}}. \tag{5.37}$$

Then, the m^{th} user data is $[\widetilde{\mathbf{x}}_{\mathsf{b}}]_m = \widetilde{x}_{m,\mathsf{b}}$ is given by

$$\begin{aligned} \widetilde{x}_{m,\mathsf{b}} = \sqrt{p_{\mathrm{u}}}\gamma_\lambda x_{m,\mathsf{b}} \left\langle \mathbf{a}_m, \mathbf{h}_m \right\rangle + \sqrt{p_{\mathrm{u}}}\gamma_\lambda \sum_{\substack{i=1 \\ i\neq m}}^{M} x_{i,\mathsf{b}} \left\langle \mathbf{a}_m, \mathbf{h}_i \right\rangle \\ + \gamma_\lambda \left\langle \mathbf{a}_m, \boldsymbol{\epsilon} \right\rangle + \left\langle \mathbf{a}_m, \boldsymbol{\epsilon}_{\mathbf{q}} \right\rangle, \end{aligned} \tag{5.38}$$

where $\mathbf{a}_m$ is the m^{th} column of $\mathbf{A}$. From (5.38), we can also obtain the interference and noise term by excluding the m^{th} user's data, given by

$$\mathrm{I}_{\mathbf{H}} = p_{\mathrm{u}}\gamma^2_\lambda \sum_{\substack{i=1 \\ i\neq m}}^{N} \left|\left\langle \mathbf{a}_m, \mathbf{h}_i \right\rangle\right|^2 + \gamma^2_\lambda \left\|\mathbf{a}_m\right\|^2 + \left\langle \mathbf{a}_m, \mathbf{R}_{\boldsymbol{\epsilon}_{\mathsf{q}},\boldsymbol{\epsilon}_{\mathsf{q}}} \mathbf{a}_m \right\rangle \tag{5.39}$$

where $\mathbf{R}_{\boldsymbol{\epsilon}_{\mathsf{q}},\boldsymbol{\epsilon}_{\mathsf{q}}}$ is the covariance of the additive quantization noise $\boldsymbol{\epsilon}_{\mathsf{q}}$ in a given channel realization with

$$\mathbf{R}_{\boldsymbol{\epsilon}_{\mathsf{q}},\boldsymbol{\epsilon}_{\mathsf{q}}} = \gamma_\lambda(1-\gamma_\lambda)\,\mathrm{diag}\left(p_{\mathrm{u}}\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^H + \mathbf{I}\right). \tag{5.40}$$

Specifically, $\mathbf{R}_{\mathbf{x}}$ is the input signal covariance, and $\mathbf{R}_{\mathbf{x}} = \mathbf{I}$ when $r[k] \sim \mathcal{N}$. The above derivations allow us to obtain the ergodic achievable uplink rate of the m^{th} user [131] that is,

$$R_m = \mathscr{E}\left\{\log_2\left(1 + \frac{p_{\mathrm{u}}\gamma^2_\lambda \left\|\mathbf{a}_m\right\|^4}{\mathrm{I}_{\mathbf{H}}}\right)\right\}. \tag{5.41}$$

Since the direct computation of the achievable uplink rate in (5.41) is challenging, in the sequel, we will utilize an analytical approximation for the λ–MIMO case.

MRC Analytical Approximation of Achievable sum-rate

The uplink rate of the m^{th} user with MRC is given by [131]

$$\widetilde{R}_{m,(\mathsf{MRC})} = \log_2\left(1 + \frac{p_{\text{u}}\gamma_\lambda \eta_m (N+1)}{\mathrm{I}_{\mathsf{MRC}}}\right), \tag{5.42}$$

where $\mathrm{I}_{\mathsf{MRC}}$ is the expectation of the noise-plus-interference term with

$$\mathrm{I}_{\mathsf{MRC}} = p_{\text{u}}\gamma_\lambda \sum_{\substack{i=1 \\ i\neq n}}^{M} \eta_i + p_{\text{u}}(1-\gamma_\lambda)\left(\sum_{i=1}^{M}\eta_i + \eta_m\right) + 1, \tag{5.43}$$

and the sum-rate for the system is $R = \sum_{m=1}^{M} \widetilde{R}_m$.

ZF Analytical Approximation of Achievable sum-rate

Similarly, the uplink rate of the m^{th} user with ZF is given by [168]

$$\tilde{R}_{m,(\mathsf{ZF})} = \log_2\left(1 + \frac{\gamma_\lambda p_{\text{u}}}{\frac{\gamma_\lambda}{(N-M)\eta_m} + (1-\gamma_\lambda)\mathrm{I}_{\mathsf{ZF}}}\right), \tag{5.44}$$

where I_{ZF} is the expectation of the noise-plus-interference term given in (5.45), $\delta < 1$ is the attenuation coefficient.

$$\begin{aligned}
\mathrm{I}_{\mathsf{ZF}} = N\left(\frac{\delta}{N+M}\right)^2 \Bigg\{ & 4\eta_m - \frac{4\delta}{N+M}\eta_m\left(\sum_{j=1}^{M}\eta_j + \eta_m\right) + N\eta_m\left(\frac{\delta}{N+M}\right)^2\left(\sum_{j=1}^{M}\eta_j^2 + \eta_m^2\right) \\
& + p_{\text{u}}\sum_{i=1, i\neq k}^{M}\left[4\eta_m\eta_i - \eta_m\eta_i\frac{4\delta}{N+M}\left(\sum_{j=1}^{M}\eta_j + \eta_m + \eta_i\right) + N\eta_m\eta_i\left(\frac{8}{N+M}\right)^2\left(\sum_{j=1}^{M}\eta_j^2 + \eta_m^2 + \eta_i^2\right)\right] \\
& + 8p_{\text{u}}{\eta_m}^2 - p_{\text{u}}{\eta_m}^2\frac{8\delta}{N+M}\left(\sum_{j=1}^{M}\eta_j + 2\eta_m\right) + p_{\text{u}}\eta_m^2\left(\frac{\delta}{N+M}\right)^2\left((N+1)\sum_{j=1}^{M}\eta_j^2 + (3N+1)\eta_m^2\right)\Bigg\}.
\end{aligned} \tag{5.45}$$

5.5 Numerical evaluation

The focus of this section is to demonstrate the advantages of our proposed λ–MIMO approach. Our experiments are based on both computer simulations as well as modulo ADC hardware.

Starting with computer experiments, we show the recovery and detection performance of the high-order modulation schemes (*e.g.* 1024 QAM OFDM signal) when using the $\mathscr{M}_\lambda$–ADC. The SQNRs of the uniform and Gaussian distributed baseband samples in $\mathscr{M}_\lambda$–ADC and conventional ADC are compared with theory in Section 5.5.1. We show proof-of-concept hardware experiments in Section 5.5.2. Additionally, the ergodic achievable sum-rate of $\mathscr{M}_\lambda$–ADC with MRC and ZF detectors, and the analytic approximations are discussed in Section 5.5.3.

5.5.1 Recovery and Detection Performance

We consider two cases that NBSC and MIMO–OFDM in the simulation. Specifically, NBSC is with channel tap $D = 1$, small-scale fading coefficients $g_{n,m} \sim \mathcal{CN}(0, 1)$. The MIMO–OFDM considers the frequency selective channel with channel taps $D = 15$ with small-scale fading coefficient[3] $g_{n,m}[d] \sim \mathcal{CN}(0, 1/D)$. We assume that perfect CSI, maximum value of baseband signal $\|r\|_\infty$, $\lambda = \frac{1}{10}\|r\|_\infty$ are known at the receiver. To satisfy the sufficient condition for recovery, we use $f_s \geqslant 2\pi e f_{\text{Nyquist}}$ in the simulation.

At the receiver with $\mathscr{M}_\lambda$–ADC, the received signals are modulo folded, sampled, quantized, recovered, and combined by MMSE, ZF and MRC detectors. In Fig. 5.11, the MMSE, ZF, and MRC receiver with $\mathscr{M}_\lambda$–ADCs outputs are plotted for the case that single-antenna multiple users transmit 64QAM modulation signals. The observations are that the recovered signals from modulo-folded samples can be detected by normal linear combiners. The ZF and MMSE combiners have close detection performances, and are better than the detection performance of MRC combiner. From Fig. 5.11 (c) and (d), we can observe that a

[3]Uniform power delay profile is assumed for simplicity of illustration. More realistic channel models can be used.

better mapping in the constellation can be achieved by increasing the antenna number in the BS.

In Fig. 5.12 (a), (b), the recovery results of one user's data for the cases that users transmit QPSK single-carrier and QPSK OFDM signals are demonstrated. In addition, in Fig. 5.12 (c), (d), the 1024 QAM single-carrier and OFDM signals after the same processes are shown. The recovery performance is also analyzed by the reconstruction mean squared error (MSE) between the input signal r and its estimate $\tilde{r}$ recovered by Algorithm 8, symbol error rate (SER), and bit error rate (BER). Specifically, the result of the recovery single-carrier signal with 1024 QAM is that MSE $= 3.8 \times 10^{-4}$. When adopting ZF combiner, the detection result is that BER $= 7.7 \times 10^{-4}$ and SER $= 7.9 \times 10^{-3}$. Using the same simulation parameters, OFDM signal with 1024 QAM can also be fully recovered with MSE $= 6.4 \times 10^{-4}$. Additionally, when adopting ZF combiner, the detection result is that BER $= 3.0 \times 10^{-3}$ and SER $= 1.7 \times 10^{-2}$.

From Fig. 5.11 and Fig. 5.12, it is known that the once incapable recovery and detection of very high order modulation signal can be realized with only 2 bit resolution $\mathscr{M}_\lambda$–ADC with $f_\mathrm{s} = 50 f_{\mathsf{Nyquist}}$. According to the Fig. 5.12 (e), (f) and Fig. 5.12 (g), (h), we can know $\Delta^L r = \mathscr{M}_\lambda \left(\Delta^L q_\lambda\right)$, which means the modulo samples of higher order of bandlimited samples r and modulo samples q_λ are the same. This fact provides the evidence of the unlimited sampling theory's effectiveness on communication signals.

When the bit resolution of $\mathscr{M}_\lambda$–ADC increases from $b = 2$ to 12, the MSE of 1024 QAM single-carrier signal decreases from 3.8×10^{-4} to 2.0×10^{-10}. Additionally, the MSE of 1024 QAM OFDM signal decreases from 6.4×10^{-4} to 2.0×10^{-10}. These results show that the recovery results of $\widetilde{r}$ based on Algorithm 8 become better. Considering a noisy case, the detection performances of 2-bit quantization resolution 1024 QAM single-carrier signal are BER $= 3.1434 \times 10^{-4}$ and SER $= 3.1 \times 10^{-3}$ with SNR $= 55$ dB and ZF combiner.

Currently, our inversion/recovery algorithm is not geared to handle low-SNR conditions and hence, this requires some future work before the MMSE combiner can be tightly integrated with our framework.

In addition, the SQNRs of single-carrier and OFDM signals with different ADC resolutions b are shown in Fig. 5.13. Specifically, the curves with theory are calculated by the theoretical SQNR equations, *i.e.* (5.29), (5.30), (5.33), and (5.34). For example, if the OFDM signal is transmitted and $\mathscr{M}_\lambda$–ADC is used in the receiver, the OFDM signal theory curve is calculated by (5.34), and then plotted. The curves with simulation are obtained from our numerical simulations, wherein we calculated the power of the transmitted signal, σ_r^2, and the power of the quantization noise, σ_{q}^2, from the simulated signal. Subsequently, the SQNR is calculated by (5.25). It is clear that a higher SQNR is achieved by using $\mathscr{M}_\lambda$–ADC than the conventional ADC. We also show the eye diagrams of signal after and before reconstruction in Fig. 5.14. The distortion of the reconstructed signal is low, which validates the effectiveness of the reconstruction algorithm.

5.5.2 Hardware Experiment

We show proof-of-concept hardware experiments of $\mathscr{M}_\lambda$–ADC in this section. Here, we use modulo samples of the real component of the QPSK signal. Specifically, the settings are $f_s = 50 f_{\mathsf{Nyquist}}$, dynamic range 13.6V peak-to-peak ($\approx 7\lambda$), and folded dynamic range 4.02V peak-to-peak. The experimental data and recovery results are shown in Fig. 5.15. We can know that the original real component of the QPSK signal can be recovered with MSE $= 4.6 \times 10^{-2}$ and 7.5×10^{-1} when we use 12-bit and 2-bit quantization resolution, respectively.

5.5.3 Achievable Uplink sum-rates and Analytical Approximations

In this simulation, a hexagonal cell with a radius of 1000m and a central disk radius $d_{\mathsf{min}} = 100$m is considered. The large scale attenuation η_{n} is modeled as $\eta_{\mathsf{n}} = z_{\mathsf{n}} \left(d_m / d_{\mathsf{min}}\right)^{-v}$, where z_{n} is a log-norm random variable with standard deviation $\sigma_{\mathsf{shadow}} = 8$dB, $v = 3.8$ is the path loss exponent, d_m is the distance between m^{th} user and BS. We assume a narrowband system and $D = 1$, therefore the small-scale fading coefficient $g_{n,m} \sim \mathcal{CN}(0, 1)$. Additionally, $M = 10$ users are randomly distributed with a normalized transmit power $p_{\mathsf{u}} = 10$dB.

Then, the achievable sum-rates of $\mathscr{M}_\lambda$–ADC and conventional ADC are compared in Fig. 5.16. In Fig. 5.16 (a), (c), analytical approximations and simulation results for $\mathscr{M}_\lambda$–ADC and conventional ADCs with MRC detector and bit resolution $b = 1, 2, \infty$ are shown. Moreover, results for $\mathscr{M}_\lambda$–ADC and conventional ADCs with ZF detector and $b = 2, 3, \infty$ are presented in Fig. 5.16 (b), (d). In all cases, a clear observation can be found that by adopting $\mathscr{M}_\lambda$–ADC the achievable uplink sum-rate can approach the sum-rate achieved by $b = \infty$. Also, the analytical approximations and simulations of achievable sum-rate for $\mathscr{M}_\lambda$–ADC are in agreement.

The power scaling law is also examined by setting the user transmit power to $p_{\mathsf{u}} = E_{\mathsf{u}}/M$. We consider a scenario that $M = 10$ users in the cell, $E_{\mathsf{u}} = 10$ dB, $b = 1, 2, \infty$, and MRC detector in the BS with different number of antennas. According to Fig. 5.17, the results show that $\mathscr{M}_\lambda$–ADC has close achievable uplink sum-rate even with $b = 1$ to the sum-rate of conventional ADC with perfect quantization. In addition, with the increase of the number of BS antennas, there is no obvious rise on the achievable sum-rate. Thus, there is no need to equip the BS with very large quantity of antennas regarding achievable sum-rate. If the $\mathscr{M}_\lambda$–ADC is used, there is no need to have very high-resolution (8-12 bits) ADCs.

The relationship between achievable uplink sum-rate and ADC bit resolution b and energy efficiency are investigated. Simulation results of uplink sum-rate versus bit resolution are presented for $\mathscr{M}_\lambda$ and conventional ADC with MRC and ZF detector in Fig. 5.18 (a). The simulation settings are users, $M = 10$, $p_{\mathsf{u}} = 10$ dB, and BS antennas, $N = 50$. The observation is that $\mathscr{M}_\lambda$–ADC outperforms conventional ADC with respect to achievable sum-rate with low resolution b. In the case of MRC detector, the achievable uplink sum-rate of $\mathscr{M}_\lambda$–ADC with only 2 bit resolution nearly approaches the limit with infinite resolution. The similar result is also obtained in the ZF detector case, with $b = 5$ bit resolution, the $\mathscr{M}_\lambda$–ADC can achieve close result to the conventional ADC with perfect quantization.

Finally, the energy efficiency $\xi \triangleq \Omega R/P$ [159] is measured, where $\Omega = 1$ MHz

is the bandwidth, R is the sum-rate, $P = c_0 N 2^b + c_1$ with $c_0 = 10^{-4}$ Watt and $c_1 = 0.02$ Watt. We can know that the energy efficiency of $\mathscr{M}_\lambda$–ADC is higher than conventional ADC when $b \leqslant 6$, and lower resolutions b have higher energy efficiency. Considering the results in Fig. 5.18 (b) that the achievable sum-rate saturates after some resolution b in MRC and ZF, low resolution b (e.g., $b = 1$-3) of $\mathscr{M}_\lambda$–ADC should be selected to reach high energy efficiency and reduce power consumption.

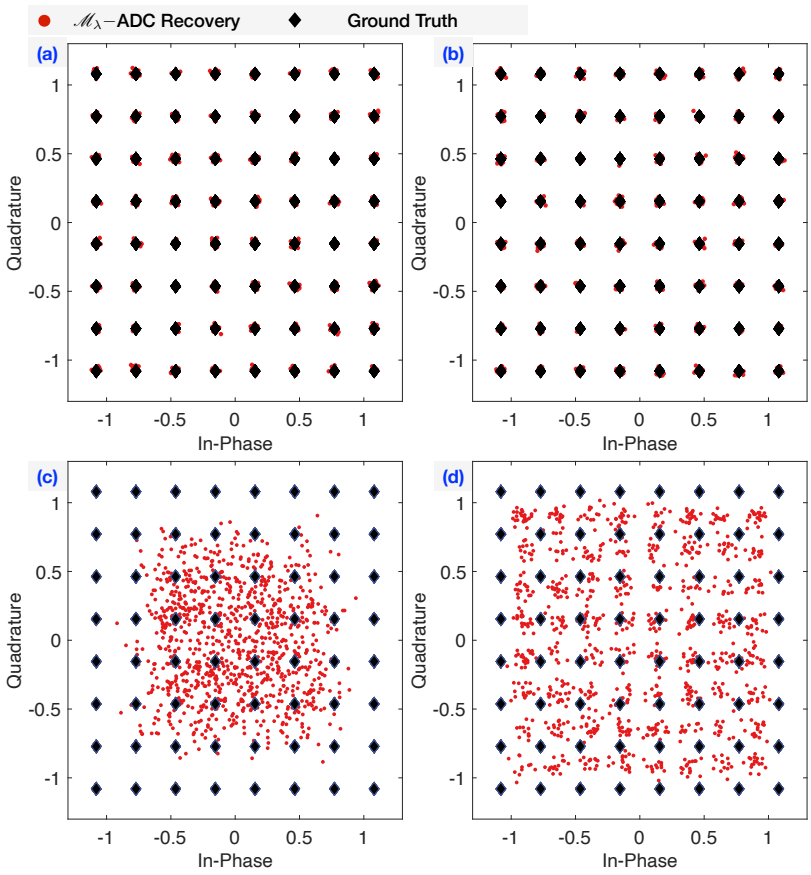


Figure 5.11 Constellation diagrams. (a) $M = 10$ users with $N = 50$ BS antennas and MMSE detector; (b) $M = 10$ users with $N = 50$ BS antennas and ZF detector; (c) $M = 10$ users with $N = 50$ BS antennas and MRC detector; (d) $M = 10$ users with $N = 1000$ BS antennas and MRC detector.

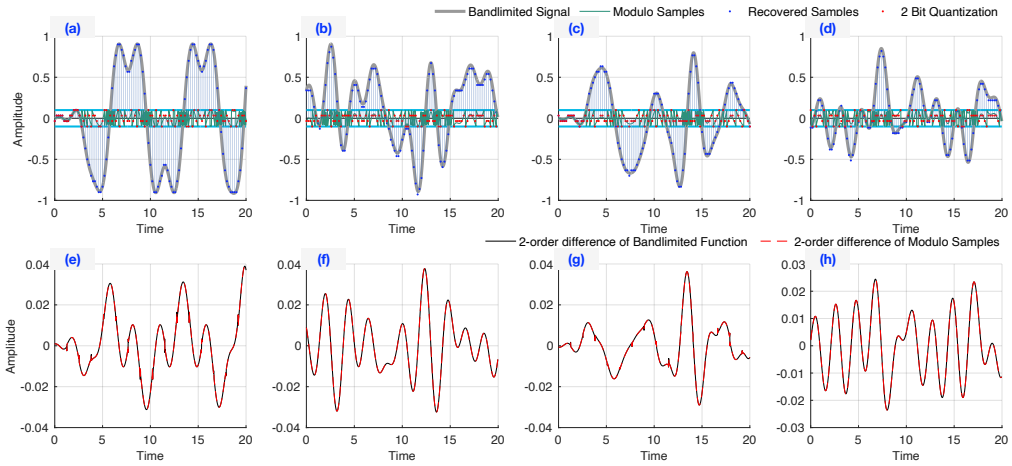


Figure 5.12 Modulo samples and recovery of QPSK (a) single-carrier signal, (b) OFDM signal, 1024 QAM (c) single-carrier signal, (d) OFDM signal; 2-order difference on samples and modulo-folded samples of QPSK (e) single-carrier signal, (f) OFDM signal, 1024 QAM (g) single-carrier signal, (h) OFDM signal.

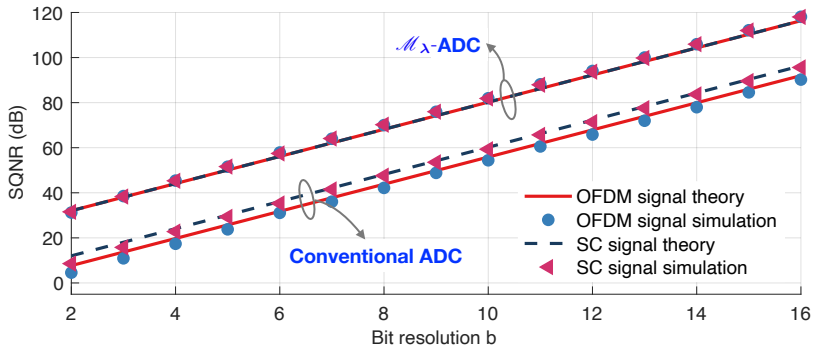


Figure 5.13 SQNR of $\mathscr{M}_\lambda$ and conventional ADCs with $\lambda = 0.1\|r\|_\infty$. The curves with theory are calculated by the theoretical SQNR equations, *i.e.* (5.29), (5.30), (5.33), and (5.34). The curves with simulation are obtained from our numerical simulations, wherein we calculated the power of the transmitted signal, σ_r^2, and the power of the quantization noise, σ_{q}^2, from the simulated signal. Subsequently, the SQNR is calculated by (5.25).

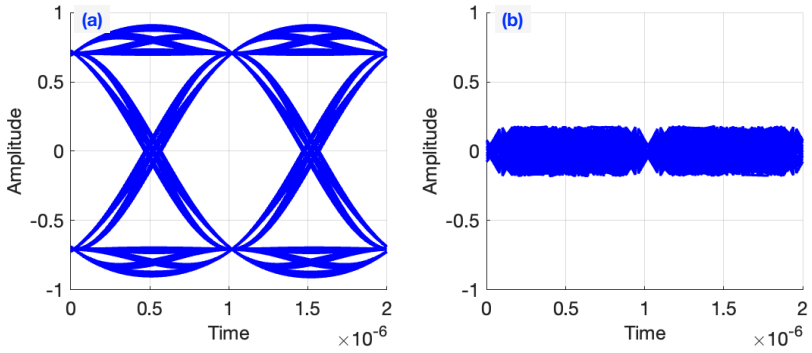


Figure 5.14 Eye diagrams with roll-off factor $\alpha = 0.45$ for signals (a) after reconstruction, (b) before reconstruction (modulo signal).

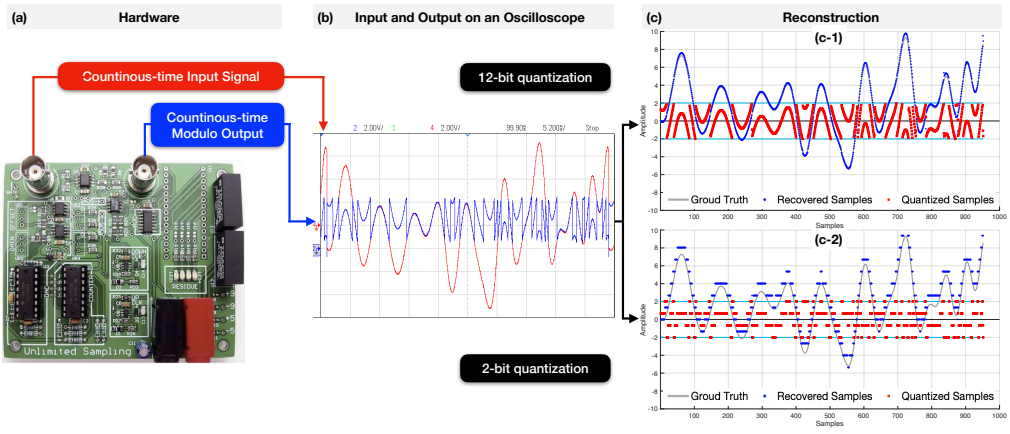


Figure 5.15 Hardware experiments and actual reconstruction results. (a) $\mathscr{M}_\lambda$–ADC circuit. (b) Live screen shot of the oscilloscope plotting the real part of QPSK signal (red) and its modulo-folded signal (blue). (c) Reconstruction results with actual modulo-folded signals from the circuit as the input, and 2-bit and 12-bit resolution quantization.

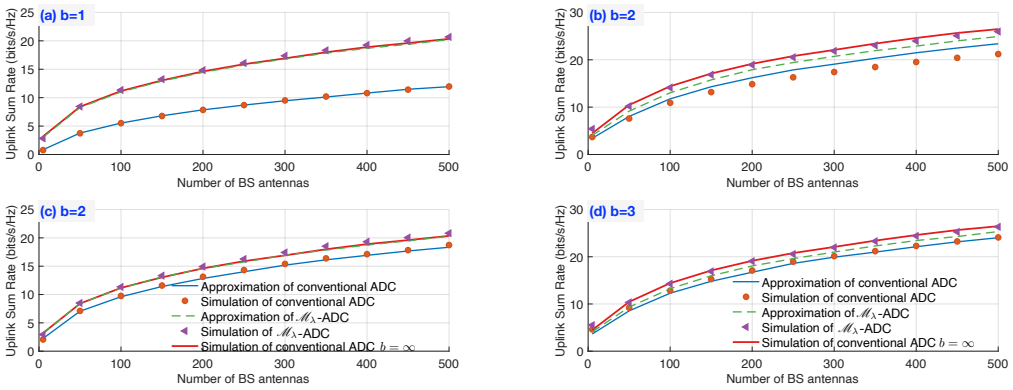


Figure 5.16 Uplink sum-rate per cell vs the number of BS antennas, when $M = 10$ users with Tx power $p_{\mathsf{u}} = 10$ dB and BS using (a), (c) MRC detector; (b), (d) ZF detector.

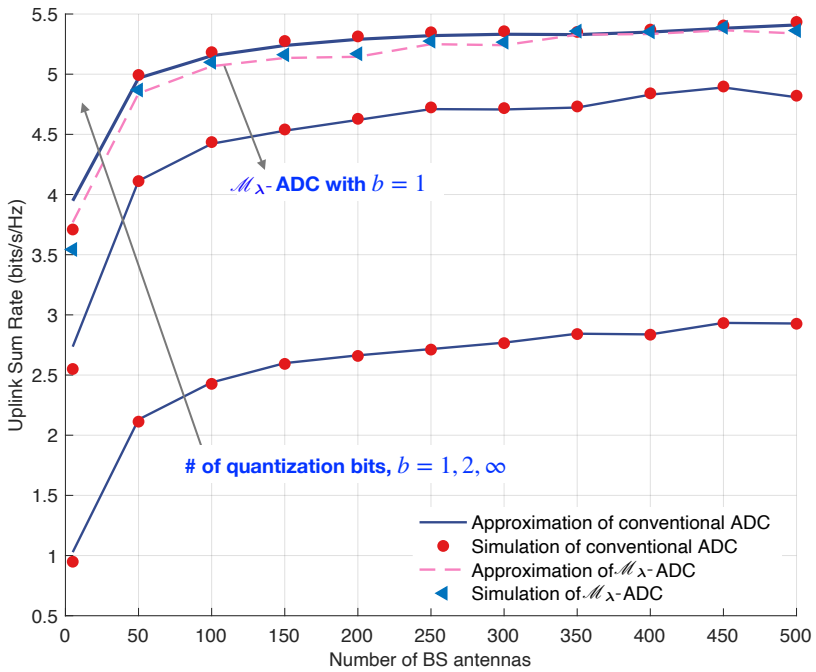


Figure 5.17 Uplink sum-rate per cell vs the number of BS antennas, when $M = 10$ users with scaled-down Tx power $p_{\mathsf{u}} = \frac{E_{\mathsf{u}}}{M}$ dB and BS using MRC detector.

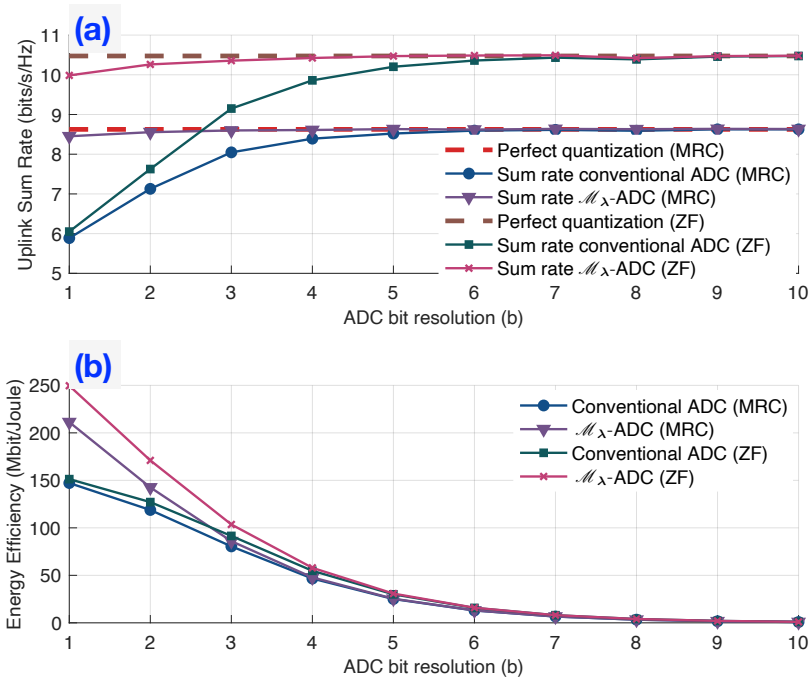


Figure 5.18 (a) Uplink sum-rate versus ADC bit resolution b, (b) Energy efficiency versus ADC bit resolution b.

5.6 Summary

Modern communication systems are pivoted on the M–MIMO technology. The success of this technology is tied to the power consumption and receiver/ADC saturation problems. In this work, we have described a new receiver design that can tackle the two challenges simultaneously. Our approach is based on the use of $\mathscr{M}_\lambda$–ADCs that are at the heart of the Unlimited Sensing Framework. By introducing modulo non-linearity in the RF-chain, we are able to fold arbitrary high dynamic range signals into low dynamic range modulo samples. This folding effect offers higher quantization resolution, given a fixed bit budget. The distinguishing features of our work include, (a) leveraging higher SQNR, (b) approaching the detection and average uplink sum-rate performances comparable to an ∞-bit resolution, conventional ADC, with a very few bit $\mathscr{M}_\lambda$–ADC. This is the key to enabling higher order modulation schemes *e.g.* 1024 QAM, (c) higher power efficiency attributing to a superior trade-off between energy efficiency and bit budget. We validate our approach using computer simulations and $\mathscr{M}_\lambda$–ADC hardware experiments.

Chapter 6

Full-Duplex Beyond Self-Interference: The Unlimited Sensing Way

6.1 Introduction

Rapidly emerging frontiers of wireless communication systems have enabled new applications. Future growth of this area in terms of applications and transmission bandwidth poses strong restrictions on the frequency spectrum [169]. A promising solution to the *spectrum scarcity* problem has been the design of In-Band Full-Duplex (FD) systems [16, 17]. Recently, there has been a surge of research interest towards this solution. This is mainly because FD systems can potentially double the spectral efficiency of communication [16, 17]. However, when considering the FD solution as it is, there is a price to pay. That is, the coupling between the co-located transmitter and receiver causes severe self-interference (SI) (*e.g.* up to 100 dB [16, 18]). This results in an undesirable scenario where the signal of interest (SoI) containing the uplink data irretrievably drowns in the interference. On one hand, simultaneous occurrence of SI and SoI naturally necessitates a high dynamic range (HDR) analog-to-digital converters or ADCs, or else, there will be no hope of recovering the uplink data when the SoI is drawn in the SI. On the other hand, for a fixed bit budget, HDR ADCs entail higher quantization noise which can compromise the capability of existing algorithms. In this case, the only available remedy is to boost the number of bits used for digitization. This strategy, proves to be sub-optimal because higher bit-budget entails higher power consumption and cost. In crux, combating the effect of SI creates an inevitable trade-off between the bit-budget and the dynamic range.

Given the significance of this problem, several promising techniques for SI cancellation have been proposed in the literature, including [16, 22, 23, 170, 171]. Broadly, these methods can be classified as follows: 1) **Analog Domain:** The key concept here is to generate a negative copy by the canceller circuit, and this replica is subtracted from the received signal [22, 23]. However, the analog-domain signal processing is especially difficult and entails high hardware complexity (compared to digital implementation). And even then, this does not enable complete cancellation of the SI, resulting in unwanted residuals [16] and can even sometimes increase the SI based on the experiment [172]. This has

motivated the development of digital methods. 2) **Digital Domain:** In contrast, the SI channel estimation and signal processing for cancelling SI becomes relatively easy in the digital domain [170, 171]. This is due to the flexibility of the digital hardware and algorithms. However, the digital-domain cancellation is limited by the ADC's dynamic range leading to clipping errors and large quantization noise, as introduced in the first paragraph.

Our work is based on the latter, and more prevalent approach of digital domain SI cancellation. In this context, so that we can simultaneously tackle the ADC saturation and quantization noise issues, we propose to use USF. In summary, instead of using conventional ADCs, we use modulo ADCs or $\mathscr{M}_\lambda$–ADCs at the core of the USF [1, 2, 80]. We develop a fully digital-domain SI cancellation and propose a full-stack, FD receiver architecture design stemming from the use of modulo ADCs. It considers the digital signal processing, *i.e.* "inversion" of the modulo non-linearity, channel estimation, digital SIC, and signal detection. In the previous chapter, we have proposed a receiver architecture design for Massive MIMO systems [80] that is based on the USF. Similar to that case, our approach here does not employ any analog-domain SIC circuitry, AGC and VGA, resulting in a low hardware complexity. Distinguished from [80], our design considers FD case, thus it encompasses new aspects of modulo-domain sparse channel estimation and digital SIC. This is elaborated in the contributions. We validate our approach through simulations and show its superiority over recent approach. Specifically, our method can achieve 40 dB digital SIC with as little as 4-bit quantization resolution.

6.2 System model

The FD communication system we consider is shown in Fig. 6.1. The baseband signals, $\gamma^{\mathsf{Q}}[k]$ and $\gamma^{\mathsf{I}}[k]$, are pulse-shaped and up-converted to form the pass-band signal, $x_{\mathsf{d}}(t)$. The FD transceiver simultaneously serves an uplink and a downlink user. The received antenna signal is given by

$$z(t) \overset{\text{def}}{=} \underbrace{\sqrt{p_{\mathsf{u}}}\left(h_{\mathsf{u}} \underline{*}_T x_{\mathsf{u}}\right)(t)}_{\mathsf{SoI}} + \underbrace{\sqrt{p_{\mathsf{d}}}\left(h_{\mathsf{SI}} \underline{*}_T x_{\mathsf{d}}\right)(t)}_{\mathsf{SI}} + n(t), \tag{6.1}$$

- $\underline{*}_T$ is the semi-discrete convolution operator with sampling interval T and some sequence $c \in \ell^2$, defined by [173]

$$\underline{*}_T : c \longmapsto \left(c \underline{*}_T \phi\right)(t) = \sum\nolimits_{n \in \mathbb{Z}} c[n]\, \phi\left(t - kT\right).$$

- p_{u} and p_{d} are the transmit power of the user and BS, respectively.
- h_{u} and h_{SI} are the discrete uplink and SI channels in the FD system, respectively.
- $x_{\mathsf{u}}(t)$ and $x_{\mathsf{d}}(t)$ are the pulse-shaped continuous-time uplink signal and downlink signal, respectively.
- $n(t) \sim \mathcal{CN}(0, 1)$ is the additive complex-Gaussian noise at the receiver.

Subsequently, the received signal enters the RF chain and is processed by the $\mathscr{M}_\lambda$–ADC at the quadrature and in-phase channels, respectively. As shown in Fig. 6.1, we design a novel RF chain, signal processing blocks for SoI reconstruction, digital SIC, and detection.

6.2.1 Receiver

Compared with conventional receiver architecture for FD system, our novel receiver deploys a $\mathscr{M}_\lambda$–ADC to simultaneously prevent drowning of SoI caused by the saturation due to the limited dynamic range, and reduce quantization noise. Specifically, the continuous-time signal $z_n(t)$ is firstly amplified by LNA

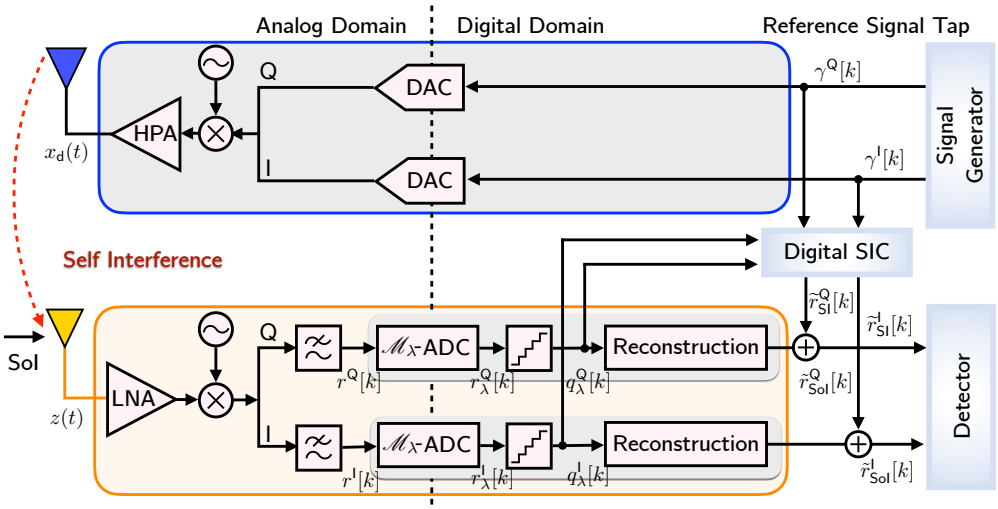


Figure 6.1 FD model with modulo-domain channel estimation method and digital cancellation based on $\mathscr{M}_\lambda$–ADC.

and demodulated to the baseband by means of coherent detection. This is facilitated by,

$$\begin{bmatrix} z_d^Q(t) \\ z_d^I(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_c t) & \\ & -\sin(\omega_c t) \end{bmatrix} \begin{bmatrix} z(t) \\ z(t). \end{bmatrix} \tag{6.2}$$

Subsequently, the quadrature and in-phase signals are low-pass filtered with $\psi(t)$; this eliminates the images located outside from the baseband. Quantitatively, we have,

$$r^Q(t) = \langle z_d^Q, \overline{\psi} \rangle \quad \text{and} \quad r^I(t) = \langle z_d^I, \overline{\psi} \rangle, \tag{6.3}$$

where $\overline{\psi}(t) = \psi(-t)$ (time-reversal). Then, the continuous-time baseband signal for I/Q channels, $r^Q(t)$ and $r^I(t)$, are folded with modulo threshold $\lambda > 0$ by the $\mathscr{M}_\lambda$–ADC hardware [4].

Since,

i) the modulo signal $r(t)$ is bounded by λ (in absolute sense), no VGA and AGC [174] is needed

ii) there is no requirement for analog-domain SIC circuitry,

our approach results in a lower hardware complexity. The continuous-time modulo-folded signal is sampled and quantized, resulting in,

$$r(t) \xrightarrow{\text{Folding}} \boxed{\mathscr{M}_\lambda} \to r_\lambda[k] \xrightarrow{\text{b Bits}} \boxed{\mathscr{Q}_b} \to q_\lambda[k] = \mathscr{Q}_b(r_\lambda[k]) \tag{6.4}$$

where $\mathscr{M}_\lambda(\cdot)$ [80,175] is the centered-modulo operator producing modulo samples, $r_\lambda[k] \overset{\text{def}}{=} \mathscr{M}_\lambda(r(t))|_{t=kT}$ and $\mathscr{Q}_b$ is the quantization operator with bit budget of b bits. These quantized samples $q_\lambda[k]$ enable the reconstruction of the SI r_{SI}, and thus we can achieve SIC by subtracting it from the received signal r.

6.2.2 Recovery from folded samples

To reconstruct the HDR input signal, $r(t)$, from its modulo samples, an "inversion" of modulo non-linearity $\mathscr{M}_\lambda(\cdot)$ is needed. The key to unfolding the modulo samples is based on the observation that the finite difference operator commutes with the modulo in some sense [1]. This allows us to obtain the higher-order differences of r from its modulo samples, r_λ. Thus, the reconstruction amounts to the problem of stably inverting the difference-operator.

Let us define the L^{th} order finite difference of the sequence $x[k]$ by the recursion, $\left(\Delta^L x\right)[k] = \Delta^{L-1}(\Delta x)[k], L > 1$. When $L = 1$, we define $\underline{x}[k] = (\Delta x)[k]{=}x[k+1] - x[k] \iff \underline{\mathbf{x}} = \mathbf{\Delta x}$. Specifically, according to the unlimited sampling theorem with bounded noise [1], oversampling $r(t) \in \mathcal{B}_\Omega$ with $\mathrm{T} \leqslant 1/2^\alpha \Omega e$, where $\alpha \in \mathbb{N}$, Ω is the bandwidth, and e is the Euler's number, allows us to access the L^{th} order difference, $\Delta^L r$, from its modulo samples $r_\lambda[k]$. A suitable L can be $L^\star \geqslant \left\lceil \frac{\log \lambda - \beta_r}{\log(T\Omega e)} \right\rceil$ with $\beta_r \in 2\lambda\mathbb{Z}$ and $\beta_r \geqslant \|r\|_\infty$. To stably invert the difference operator, we use the anti-difference operation S, defined by $\mathsf{S} : \{s[k]\}_{k\in\mathbb{Z}^+} \to \sum_{m=1}^{k} s[m]$. Instead of recovering r via anti-difference operation S, we capitalize the piece-wise constant structure of the residue $\varepsilon_r(t) = r(t) - \mathscr{M}_\lambda(r(t))$, that is, $\Delta^L \varepsilon_r(kT) \in 2\lambda\mathbb{Z}$, to reconstruct the unfolded signal. We conclude the reconstruction approach is summarized in Algorithm 9 which also works with quantized measurements [1].

Algorithm 9: Recovery from Noisy Modulo Samples

Input: $q_\lambda[k]$ and $2\lambda\mathbb{Z} \ni \beta_r \geqslant \|r\|_\infty$.

Output: $\widetilde{r}[k] \approx r(kT)$.

1) Compute $L = \left\lceil \frac{\log \lambda - \log \beta_r}{\log(\mathrm{T}\Omega e)} \right\rceil$.

2) Compute $\left(\Delta^L q_\lambda\right)[k]$.

3) Compute $\left(\Delta^L \varepsilon_r\right)[k] = \left(\mathscr{M}_\lambda\left(\Delta^L q_\lambda\right) - \Delta^L q_\lambda\right)[k]$.

Set $s^{\{0\}}[k] = \left(\Delta^L \varepsilon_r\right)[k]$

for $l = 0 : L-2$ **do**

i) $s^{\{l+1\}}[k] = \left(\mathsf{S} s^{\{l\}}\right)[k]$

ii) $s^{\{l+1\}} = 2\lambda \left\lceil \frac{\lfloor s^{\{l+1\}}/\lambda \rfloor}{2} \right\rceil$ (rounding to $2\lambda\mathbb{Z}$)

iii) Compute $\kappa^{\{l\}} = \left\lfloor \frac{\left(\mathsf{S}^2\Delta^l \varepsilon_r\right)[1] - \left(\mathsf{S}^2\Delta^l \varepsilon_r\right)[J_\lambda+1]}{12\beta_r} + \frac{1}{2} \right\rfloor$

iv) $s^{\{l+1\}}[k] = s^{\{l+1\}}[k] + 2\lambda\kappa^{\{l\}}$

end

5) $\widetilde{r}[k] = \mathsf{S}\left(s_{\{L-1\}}\right)[k] + q_\lambda[k] + 2v\lambda, v \in \mathbb{Z}$

6.2.3 Modulo-domain channel estimation and digital SIC

To achieve a full-stack fully digital-domain SIC FD system, we propose a modulo-domain channel estimation method to estimate SI channel, h_{SI}, as in Algorithm 10. In addition, we design a corresponding digital SIC approach based on $\mathscr{M}_\lambda$–ADC, inspired by [48, 157]. Since the received signal can be acquired without saturation, and it incorporates SI and SoI as given in (6.1), our aim is to reconstruct SI via our proposed approach and subtract it from the received signal.

Specifically, in the proposed modulo-domain channel estimation method, a pilot is firstly transmitted, and we assume no SoI is transmitted during the SI channel estimation period. Since the channel is a single-tap channel and the estimation requires good auto-correlation to estimate the delay, for simplicity, we use the basic random quadrature phase shift keying (QPSK), i.e., $x_{\mathsf{pilot}} = e^{\jmath(2m+1)\frac{\pi}{4}}, m \in \{0,1,2,3\}$, signal as the pilot. For a more complex multipath channel, other pilot sequences with good auto-correlation, low cross-correlation, and robustness to multipath and Doppler effects should be adopted. Considering a single path SI channel, h_{SI}, with delay, τ, and amplitude, A, the

SI channel is hence modelled as $h_{\mathsf{SI}}[k] = A\delta[k-\tau]$. After down-conversion, the pointwise received SI samples are given by

$$r_{\mathsf{SI}}[k] = A\gamma[k-\tau] = \sum\nolimits_{|p|\leqslant P} \left(\widehat{\gamma}_p \widehat{h}_{\mathsf{SI}}\right) e^{\jmath p\omega_0 k}, \tag{6.5}$$

where γ is a T_p-periodic baseband transmit signal (reference signal, cf. transmitter in Fig. 6.1), $\widehat{h}_{\mathsf{SI}} = Ae^{-\jmath p\omega_0\tau}$, $P = \lceil \Omega/\omega_0 \rceil$, $\omega_0 = 2\pi/T_p$, and $K \geqslant 2P+1$. Additionally, $\widehat{\gamma}_p$ is the Fourier series (FS) coefficient and $\widehat{\gamma}_{-p} = \widehat{\gamma}_p^*$. We then have the bandlimitedness that is $\widehat{\gamma}_p = 0, |p| > P = [\Omega/\omega_0]$ denoted by $f \in \mathcal{B}_\Omega$. In vector-matrix format, (6.5) is given by

$$\mathbf{r}_{\mathsf{SI}} = \mathbf{U}^* \boldsymbol{\mathcal{D}}_{\widehat{\gamma}} \widehat{\mathbf{h}}_{\mathsf{SI}}, \tag{6.6}$$

- $\mathbf{r}_{\mathsf{SI}} \in \mathbb{R}^{K\times 1}$ is the samples with $[\mathbf{r}_{\mathsf{SI}}]_k = r_{\mathsf{SI}}(kT)$.
- $\mathbf{U} \in \mathbb{C}^{K\times(2P+1)}$ is the Discrete Fourier Transform (DFT) matrix with element $[\mathbf{U}]_{k,p} = e^{-\jmath p\omega_0 kT}$.
- $\boldsymbol{\mathcal{D}}_{\widehat{\gamma}}$ is the matrix of FS coefficients. It is a diagonal matrix and the FS coefficients of the kernel (transmit signal), γ, are on the diagonal, *i.e.* $[\boldsymbol{\mathcal{D}}_{\widehat{\gamma}}]_{p,p} = \widehat{\gamma}_p$.
- $\widehat{\mathbf{h}}_{\mathsf{SI}} \in \mathbb{C}^{(2P+1)\times 1}$ is a vector of exponentials parameterized by the unknown channel, *i.e.* $\widehat{\mathbf{h}}_{\mathsf{SI},p} = \widehat{h}_{\mathsf{SI},p}$.

In our case, the forward model takes the form of,

$$\mathbf{r}_{\mathsf{SI},\lambda} = \mathscr{M}_\lambda(\mathbf{U}^* \boldsymbol{\mathcal{D}}_{\widehat{\gamma}} \widehat{\mathbf{h}}_{\mathsf{SI}}) \in \mathbb{R}^{K\times 1}, \tag{6.7}$$

and the goal is to estimate $\mathbf{r}_{\mathsf{SI}}$ in (6.6).

6.2.4 Estimation Procedure

Starting from the modulo decomposition property, we have

$$\mathbf{r}_{\mathsf{SI},\lambda} = \mathbf{r}_{\mathsf{SI}} - \boldsymbol{\varepsilon}_{\mathsf{SI}} \xrightarrow{\Delta} \underline{\mathbf{r}}_{\mathsf{SI},\lambda} = \underline{\mathbf{r}}_{\mathsf{SI}} - \underline{\boldsymbol{\varepsilon}}_{\mathsf{SI}}, \quad \underline{\mathbf{r}}_{\mathsf{SI}} = \Delta\mathbf{r}_{\mathsf{SI}} \tag{6.8}$$

Algorithm 10: Modulo-domain Sparse Channel Estimation

Input: $\{r_{\mathsf{SI}}[k], r_{\mathsf{SI},\lambda}[k]\}_{k=0}^{K-1}, T_p$, and P.

Output: $\mathbf{h}_{\mathsf{SI}}$.

1) Compute $\widehat{\underline{\mathbf{r}}}_{\mathsf{SI},\lambda} = \mathbf{V}\underline{\mathbf{r}}_{\mathsf{SI},\lambda}$ (DFT).
2) Define $[\mathbf{z}]_n = -[\widehat{\mathbf{r}}_{\mathsf{SI},\lambda}]_n, n \in \mathbb{I}_{K-1}\backslash\mathbb{E}_{P,K-1}$.
3) Estimate number of folds, M, by forming a Toeplitz or Hankel matrix from $[\mathbf{z}]_n$ [48] and then thresholding using *second order statistic* of eigenvalues [176].
4) Estimate folds using $\texttt{Prony}(\mathbf{z}) \mapsto \left(\{\mu_m, \nu_m\}_{m=0}^{M-1}, \mathbf{f}\right)$.
5) Construct $\widehat{\underline{\boldsymbol{\varepsilon}}}_r$ for $n \in [0, K-2]$.
6) Estimate $[\widehat{\underline{\mathbf{r}}}_{\mathsf{SI}}]_n = [\widehat{\underline{\mathbf{r}}}_{\mathsf{SI},\lambda}]_n + [\widehat{\underline{\boldsymbol{\varepsilon}}}_r]_n, n \in \mathbb{E}_{P,K-1}$.
7) Calculate $\widehat{\underline{\boldsymbol{\gamma}}}$ with given reference signal γ. Compute element-wise division that is $\mathbf{d} = \widehat{\underline{\mathbf{r}}}_{\mathsf{SI}}/\widehat{\underline{\boldsymbol{\gamma}}}$, where $[\mathbf{d}]_n = e^{-\jmath\omega_0\tau}[\mathbf{d}]_{n-1}$.
8) Calculate $q \triangleq ([\mathbf{d}]_n - [\mathbf{d}]_{n+1})/([\mathbf{d}]_{n-1} - [\mathbf{d}]_n)$.
9) Compute $\widetilde{\tau} = -K\arctan(q)/(2\pi)$, $\widetilde{A} = \mathscr{E}(|\mathbf{d}|)$, and $\mathbf{h}_{\mathsf{SI}} = \widetilde{A}\delta(k - \widetilde{\tau})$.
10) Reconstruct $r_{\mathsf{SI}} = h_{\mathsf{SI}} * \gamma$ and subsequently subtract it from the received signal (r)

where Δ is the finite difference operator yielding, $[\Delta\mathbf{r}_{\mathsf{SI}}]_k = [\mathbf{r}_{\mathsf{SI}}]_{k+1} - [\mathbf{r}_{\mathsf{SI}}]_k$. We use $\mathbb{I}_K = \{0, \cdots, K-1\}, K \in \mathbb{Z}^+$ to denote the set of K contiguous integers, and $\mathbb{E}_{P,K} = [0, P] \cup [K-P, K-1], |\mathbb{E}_{P,K}| = 2P+1$ to denote the effective spectrum. Applying the DFT on (6.8) results in,

$$\mathbf{V}\underline{\mathbf{r}}_{\mathsf{SI},\lambda} = \mathbf{V}\underline{\mathbf{r}}_{\mathsf{SI}} - \mathbf{V}\underline{\boldsymbol{\varepsilon}}_{\mathsf{SI}}, \quad [\mathbf{V}]_{n,k} = e^{-\jmath\left(\frac{2\pi}{K-1}\right)nk}, k \in \mathbb{I}_{K-1} \tag{6.9}$$

where $\mathbf{V} \in \mathbb{C}^{(K-1)\times(K-1)}$ is the DFT matrix. Noting that $\mathbf{V}\underline{\mathbf{r}}_{\mathsf{SI}}$ is the contribution due to the bandlimited part, choosing $n \in (\mathbb{I}_{K-1}\backslash\mathbb{E}_{P,K-1}) + (K-1)\mathbb{Z}$ provides us with access to the modulo folds, which in the DFT domain assume a parametric form. In particular, $\forall n \in (\mathbb{I}_{K-1}\backslash\mathbb{E}_{P,K-1}) + (K-1)\mathbb{Z}$, $\mathbf{V}\underline{\mathbf{r}}_{\mathsf{SI},\lambda} = -\mathbf{V}\underline{\varepsilon}_{\mathsf{SI}} = \sum_{m=0}^{M-1} \mu_m e^{-\jmath\left(\frac{2\pi}{K-1}\right)\frac{n}{T}\nu_m}$ is sum-of-complex-exponentials [48,157] which can be estimated via Prony's method [177]. We denote this procedure by, $\texttt{Prony}(\mathbf{z}) \mapsto \left(\{\mu_m, \nu_m\}_{m=0}^{M-1}, \mathbf{f}\right)$ where $\mathbf{z} = -\mathbf{V}\underline{\varepsilon}_{\mathsf{SI}}, \forall n \in (\mathbb{I}_{K-1}\backslash\mathbb{E}_{P,K-1})$.

With $\widehat{\varepsilon}_{\mathsf{SI}}$ known, we can remove its contribution and obtain $\widehat{\underline{\mathbf{r}}}_{\mathsf{SI}} = \widehat{\underline{\mathbf{r}}}_{\mathsf{SI},\lambda} + \widehat{\underline{\varepsilon}}_{\mathsf{SI}}$. Instead of recovering the signal $\mathbf{r}_{\mathsf{SI}}$, we follow a direct Fourier domain deconvolution

approach to estimate $\mathbf{h}_{\mathsf{SI}}$. To this end, note that,

$$\forall n \in (\mathbb{E}_{P,K-1}) + (K-1)\mathbb{Z}, \quad \widehat{\underline{r}}_{\mathsf{SI}}[n] = \widehat{\underline{\gamma}}[n] A e^{-\jmath\omega_0 n\tau} \tag{6.10}$$

The process of deconvolution now entails point-wise division,

$$d[n] \triangleq \frac{\widehat{\underline{r}}_{\mathsf{SI}}[n]}{\widehat{\underline{\gamma}}[n]} = Ae^{-\jmath n\omega_0\tau} \equiv \widehat{h}_{\mathsf{SI}}[n]. \tag{6.11}$$

With $[\mathbf{d}]_n = d[n]$, we can now estimate $A = |\mathbf{d}|$ and more generally, via $\mathscr{E}(|\mathbf{d}|)$, where $\mathscr{E}$ denotes expectation operator. The unknown delay is estimated via,

$$[\mathbf{d}]_n/[\mathbf{d}]_{n-1} = e^{-\jmath\omega_0\tau} \Rightarrow \tau = \tfrac{1}{-\jmath\omega_0} \angle \left([\mathbf{d}]_n/[\mathbf{d}]_{n-1}\right). \tag{6.12}$$

Furthermore, one may leverage measurement redundancy to obtain a robust estimate via least-squares solution. The procedural steps are outlined in Algorithm 10.

Having estimated h_{SI} via the proposed modulo-domain channel estimation method, we can reconstruct r_{SI} and subsequently subtract it from the received signal (r), to obtain the SoI, r_{SoI}.

6.2.5 Recovery Condition

For Algorithm 10 to succeed, the exponential fitting part (*i.e.* Prony's method) in step 2 requires

$$K \geqslant 2(P + M + 1)$$

while step 8 (see (6.12)) requires that $P \geqslant 1$. With the period $\tau = KT$, the recovery condition thus is, $T \leqslant \frac{\tau}{2M+4}$ with $P = \lceil \frac{\Omega}{\omega_0} \rceil \geqslant 1$, resulting the following theorem.

Theorem 3 (Modulo-domain Channel Estimation)**.** *Let $g = h_{\mathsf{SI}} * \gamma$ where $h_{\mathsf{SI}}(t) = A\delta(t - t_\tau)$ is the unknown 1-sparse channel, and $\gamma \in \mathcal{B}_\Omega$ is a known, T_p-periodic, transmitted signal. Given K modulo samples $g_\lambda[k] = \mathscr{M}_\lambda(g(kT)), T > 0$ folded at most M times, the sufficient condition for estimation of h_{SI} from $\{g_\lambda[k]\}_{k=0}^{K-1}$ is that*

$T \leqslant \frac{T_p}{K}$ *and* $K \geqslant 2(M+2)$.

6.3 Numerical evaluation

This section focuses on the unique advantages of adopting the USF for full-stack FD architecture. According to the simulations, we show the saturation-free recovered (received) signal, reconstructed SoI and detection performance. We consider single-carrier QPSK based transmission, and the SI channel is modelled as a single path channel with delay. Additionally, the SI power is measured by $\mathsf{SIR} = 10\log_{10}(P_{\mathsf{SoI}}/P_{\mathsf{SI}})$ dB. We assume that no uplink data is transmitted during the SI channel estimation period, and perfect CSI of the uplink channel is known at the receiver. With respect to the recovery algorithm, we follow $f_s \geqslant 2\pi e f_{\mathsf{Nyquist}}$ to reconstruct the signal in the simulation. We remind the reader that this is a loose bound, a near Nyquist-rate recovery is also possible [178], specially with some side-information [179], which is not assumed in this paper.

6.3.1 Recovery of received signal and SoI after SIC

Since we consider the FD system, it is natural to work with signals such that $\|r\|_\infty \triangleq \max|r(t)| \gg \lambda$. To prevent saturation, the received signal is folded by the $\mathscr{M}_\lambda$–ADCs. Subsequently, modulo-domain channel estimation as in Algorithm 10 is adopted to estimate SI channel, $\mathbf{h}_{\mathsf{SI}}$. Then, the unfolded received signal $\widetilde{\mathbf{r}}$ is recovered from the modulo samples $\mathbf{r}_\lambda$. With the estimated SI chan-

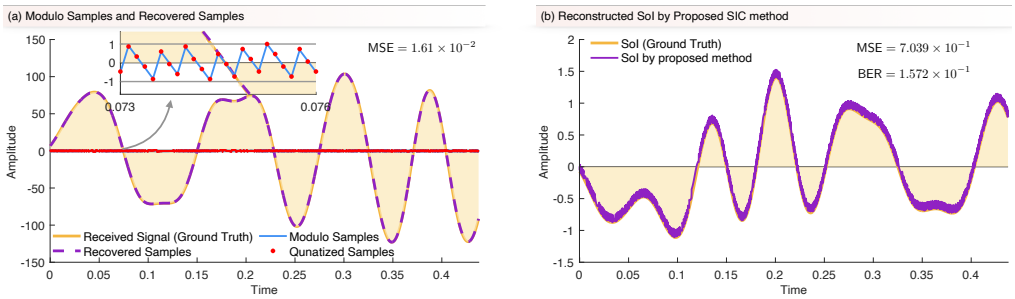


Figure 6.2 Simulation results with 4-bit quantization, 1/SIR = 40 dB, and SNR = 40 dB of (a) modulo samples and recovery; (b) SoI reconstruction by our proposed SIC approach.

nel and baseband transmit signal γ (reference signal), the digital-domain SIC is applied. Specifically, we reconstruct the SI signal $\mathbf{r}_{\mathsf{SI}}$, and subtract it from the reconstruction of the received signal $\widetilde{\mathbf{r}}$ to obtain $\mathbf{r}_{\mathsf{SoI}}$. Finally, the QPSK signal detection is executed on the $\mathbf{r}_{\mathsf{SoI}}$.

As shown in Fig. 6.2, considering low-resolution quantization (*i.e.* $b = 4$) with 1/SIR = 40 dB (*e.g.* $\lambda = 1, \|r\|_\infty = 141.1621$), and SNR = 40 dB, the reconstruction error of the received signal is $\mathsf{MSE} = 1.61 \times 10^{-2}$. The reconstruction error of the SoI is $\mathsf{MSE} = 7.039 \times 10^{-1}$. The bit error rate (BER) is 1.572×10^{-1}. If we consider a smaller SI power that is 1/SIR = 20 dB, and keep other simulation parameters fixed, the following results are obtained. The reconstruction error of the received signal is $\mathsf{MSE} = 1.17 \times 10^{-2}$. The reconstruction error of the SoI is $\mathsf{MSE} = 3.47 \times 10^{-2}$. The bit error rate (BER) is 7.47×10^{-2}. We attribute the reconstruction with few bits to the fact that for a given bit budget, the $\mathscr{M}_\lambda$–ADC offers a higher signal to quantization noise ratio (SQNR). With the definition of ratio $\zeta \triangleq \lambda/\|r\|_\infty$, it is shown that a $20\log_{10}(1/\zeta)$ dB enhancement in the SQNR [80]. In contrast, adaptive filter based strategies [47] are not feasible as they can not handle modulo samples, and hence, the SoI reconstruction and detection fail in the simulations.

Remark 12. *Compared to chapter 5, the main mathematical difference of this chapter is that a modulo-domain channel estimation method is provided.*

6.4 Summary

In this work, we presented a method for full-stack FD receiver architecture, utilizing the unlimited sensing framework. We demonstrated the advantages of substituting the traditional ADC with a modulo ADC ($\mathscr{M}_\lambda$−ADC). This replacement effectively avoids ADC saturation, enabling the capture of high dynamic range signals. Additionally, the low quantization noise feature of the $\mathscr{M}_\lambda$−ADC allows for the use of few-bits quantization. The us of $\mathscr{M}_\lambda$−ADC necessitated the creation of algorithms for sparse channel estimation. To address this, we introduced a novel approach for sparse channel estimation in the modulo domain, accompanied by a digital SI cancellation method. Our numerical results demonstrate that the proposed full-stack FD receiver structure can tackle with very strong SI (*e.g.* 40 dB) with few-bits quantization, and outperforms the previous approach in terms of SoI reconstruction, digital SIC, and detection performance. To summarize, our work enables: 1) fully digital-domain SI cancellation in FD system; 2) modulo-domain sparse channel estimation; 3) a low complexity implementation of FD system due to no VGA, AGC and analog-domain SIC circuitry.

Chapter 7

Conclusions

7.1 Summary of Thesis Achievements

Multi-functional wireless communication networks facilitate new capabilities and applications in future networks. However, the realization of the multi-functional wireless networks requires exploration and development of key technologies. In this thesis, we have addressed a set of designs and optimizations in M-MIMO, satellite communications, full-duplex communications, and ISAC systems to enable the multi-functional wireless networks.

The main achievements of our thesis are summarised as follows. We have proposed the λ–MIMO architecture simultaneously tackling high power consumption and receiver ADC saturation problems, rather than employing decoupled methods. We have designed a fully digital-domain SI cancellation structure in FD communication systems. Moreover, to preserve the radar echo in the mono-static ISAC system, we have proposed a FD ISAC structure, which suppresses the residual SI and preserve the two types of SoI (e.g., radar echo and uplink data). In addition, we have designed an ISAC-LEO satellite to enable the sensing functionality on LEO satellite. In the designed system, two challenges namely interference control and fairness EE guarantee among all users have been addressed. In more detail, we expand upon each of these achievements in the following discussion.

In Chapter 3, we have firstly modelled a FD ISAC mono-static system to capture the echo-miss problem in Section 3.2. As in Section 3.3, we have jointly designed the transmit and receive beamformers at the transceiver, transmit precoder at the uplink user, and receive combiner at the downlink user to simultaneously 1) maximize the uplink and downlink communication rate; 2) maximize the transmit and receive radar beampattern power at the target; and 3) suppress the residual SI. To solve this optimization problem, we have proposed a penalty-based iterative algorithm. Using the designed algorithm, we have illustrated that the proposed design can effectively achieve up to 60 dB digital-domain SI cancellation, a higher average sum-rate, and more accurate radar parameter estimation compared with previous ISAC FD studies in Section 3.5.

We have extended the exploration on ISAC to LEO satellite system in Chapter 4. In Section 4.2, we have designed an RSMA-assisted ISAC satellite system design to facilitate multi-user communications and target sensing, simultaneously. We have illustrated the inference control is effective via RSMA. In the design, low-resolution DACs are adopted at the transmitter to reduce power consumption. Considering the above requirements and also the max-min fairness EE among all users, we have formulated a optimization problem and proposed an iterative algorithm based on successive convex approximation (SCA) and Dinkelbach's method as explained in Section 4.2. In Section 4.5, adopting the designed algorithm, we have showed that the interference among users can be effectively tackled, and the max-min fairness EE can be guaranteed. Moreover, the performance of the proposed method outperforms the strategies that aim to maximize the total EE of the system and SDMA.

In Chapter 5, λ–MIMO architecture which uses modulo ADCs instead of a conventional ADC has been proposed to address the high power consumption and receiver ADC saturation problems. The massive MIMO with single carrier and OFDM signals, and MRC, ZF, MMSE combiners have reviewed in Section 5.2. Specifically, the radically changed receiver design and the unlimited sampling reconstruction guarantee have been elaborated in Section 5.3. With the newly-designed λ–MIMO, we have showed in Section 5.4, and Section 5.5 that detection and average uplink sum-rate performances comparable to a conventional, infinite-resolution ADC when using a $1-2$ bit $\mathscr{M}_{\lambda}$–ADC, this is achieved by leveraging higher SQNR. This enables higher order modulation schemes e.g. 1024 QAM that seemed previously impossible, superior trade-off between energy efficiency and bit budget, thus resulting in higher power efficiency. Furthermore, in Section 5.5, we have demonstrated that numerical simulations and modulo ADC based hardware experiments corroborate our theory and reinforce the clear benefits of λ–MIMO approach.

Finally, to tackle the SI in FD communication system, we have designed a full-stack fully digital domain SIC FD system in Chapter 6. We have showed that by injecting folding non-linearities in the sensing pipeline, one can not only suppress SI but also recover the SoI. We have elaborated the novel design of

the receiver architecture that is complemented by a modulo-domain channel estimation method in Section 6.2. In Section 6.3, we have also showed that Our method outperforms the previous approach based on adaptive filters when it comes to SoI reconstruction, detection, and digital SIC performance.

7.2 Future Research

To conclude, we discuss some open questions and possible directions for future research.

7.2.1 Future Work for Chapter 3

In Chapter 3, we utilized the incoming angle difference to suppress the SI in FD ISAC system by jointly design the transmit and receive beamformers.

- *SI Suppression via Time-of-Arrival Difference*: The time-of-arrival difference between the SI and SoI can also be investigated to suppress SI in ISAC system. Specifically, the temporal difference can be utilized to generate a negative counterpart of the SI signal before the ADC, based on some update algorithms, such as gradient-learning method. Since there is still some residual SI left, apart from adding this negative counterpart, an adaptive filter is also required to generate a negative in digital domain to cancel this residual SI. In practice, many factors (e.g., RF taps, adaptive filter taps, and update algorithms) affect the accuracy of the generated negatives in both domains, which in turn, have a huge impact on the SI cancellation performance. Thus, the key to implement this future SI suppression method on time-of-arrival is accurate generation of SI. Moreover, this requires tracking ability on fast-changing signals.
- *Waveform Design*: Waveform design can be used as a promising technique to effectively cancel SI. It can reduce the SI in the received signal.
- *SoI Boosting*: Emerging techniques, e.g., intelligent reflecting surfaces (IRS) in wireless communications can be used as a means to boost the SoI. As a result, the effect of SI is reduced.
- *Channel Estimation*: The SI channel estimation affects the performance of the beamforming design, making it necessary to further explore it in the presence of SI. Additionally, the estimation of the SI channel is also an important topic for future study.

7.2.2 Future Work for Chapter 4

In Chapter 4, we solved the interference and power efficiency problem in satellite ISAC system through RSMA, minimal EE, and low resolution ADCs. Some directions in the satellite ISAC are meaningful to explore:

- *Energy-efficient Transceiver Design*: LEO satellites typically operate under power constraints, which necessitates efficient power utilization. Hybrid beamforming offers a potential solution by optimizing RF chain usage, consequently lowering power consumption. Meanwhile, low-resolution electronic component such as phase shifter can be adopted to increase the energy efficiency. This is because the power consumption of electronic components in the RF chain is proportional to their bit budget.

- *Multi-Static LEO-ISAC System*: The incorporation of numerous LEO satellites within the constellation network can act as a multi-static radar. Such deployment enhances detection and tracking abilities, increases survivability, and enhances flexibility in the multistatic LEO-ISAC system. However, the coordination and data exchange among LEOs, such as sharing radar reference waveforms, require further exploration. Utilizing Inter-Satellite Links (ISL) is a potential solution for enabling data sharing among LEOs.

- *Security*: Due to the spectrum sharing mechanism as well as the broadcasting nature of satellite communications, the ISAC system is susceptible to unique security challenges. Since the radar tar- get has a high reception SINR on the embedded confidential signal, it greatly amplifies the vulnerability of information eavesdropping by the target. Possible solutions include cryptographic techniques, e.g., advanced encryption algorithms, and physical layer (PHY) security techniques, e.g., secure beamforming, artificial noise approaches jamming and cooperative security.

7.2.3 Future Work for Chapter 5 and 6

Given the end-to-end nature of the work of λ–MIMO and FD via unlimited sensing, it opens up new research questions in the areas of theory, algorithms and hardware design.

Theoretical Front

Using $\mathscr{M}_\lambda$–ADC in the RF-chain necessitates the development of new theoretical models. This is important for precise understanding of the fundamental limits of our work.

- *Trade-off between quantization noise and modulo threshold*: Currently we have assumed that modulo threshold is low enough to map a Gaussian source into its uniform counterpart. Mathematical analysis based on the theory of wrapped distributions will guide towards a clearer quantitative trade-off between quantization noise and modulo threshold. This is the key to leveraging higher order modulation schemes which was previously highly challenging, specially with low-resolution ADCs.
- *Tailored MMSE combiner*: Due to the non-linear, many-to-one nature of modulo mapping, its inversion is non-trivial and Algorithm 8 is one such example. Developing robust recovery methods is the key to a tight integration of the conventional MMSE combiner which are designed to work with pointwise samples. More generally, current MMSE combiners can not work with encoded (modulo) measurements, which is the case in our work. This requires a fundamental re-thinking of the MMSE combiner for λ-MIMO which remains to be investigated.

Algorithmic Front

- *Noise Robustness*: Practical deployment of modulo ADCs [4] may result in non-ideal measurements. Further to this, influence of noise in the RF-chain remains to be investigated.
- *Reconstruction assisted by correlation*: Since correlation among signals ex-

ists in received antenna array, the correlation can be exploited to enhance the reconstruction [180, 181]. Additionally, multi-channel modulo ADCs where the modulo threshold can be adjusted variably to improve reconstruction is also an promising direction.

These aspects necessitate the design of advanced algorithms for reconstruction and detection, specially the ones that can operate with moderate levels of oversampling [4].

Communication Theory

- *Full Duplex Integrated Sensing and Communications via Unlimited Sampling*: As we have discussed in chapter 3, the radar echo is drown in the strong SI when we consider mono-static ISAC setup. To preserve the radar echo, it naturally requires the FD functionality of the system to enable transmit and receive signal simultaneously. Unlimited sensing, which is a promising solution to the SI cancellation problem in the FD system, can also be adopted to address the radar echo miss problem in mono-static ISAC system. Meanwhile, unlimited sensing facilitates high energy efficiency and high quantization resolution, which can enhance both the communication and sensing performances.

Hardware Front

There are steps to be taken before our proof-of-concept hardware experiments can be implemented in practice. The first step in this direction would be a tight integration of $\mathscr{M}_\lambda$–ADC in the RF-chain. This requires custom designed hardware with technical specifications matched to the existing communication systems.

Appendix

Derivation of The Fisher Information Matrices

The elements at the i^{th} column and j^{th} row in the FIM, $\mathbf{J}$, is given by [107, 118]

$$[\mathbf{J}]_{i,j} = \frac{2}{\sigma_r^2} \operatorname{Re}\left\{ \sum_{\ell=1}^{L} \frac{\partial \boldsymbol{\mu}[\ell]^H}{\partial \xi_i} \frac{\partial \boldsymbol{\mu}[\ell]}{\partial \xi_j} \right\}, i, j \in \{1, 2\}, \tag{7.1}$$

where $\boldsymbol{\mu} \triangleq \alpha e^{\jmath 2\pi f_d \ell T} \mathbf{A}(\theta) \boldsymbol{\Delta}\mathbf{x}[\ell]$. The partial derivatives of $\boldsymbol{\mu}$ with respect to the unknown parameters are given by

$$\frac{\partial \boldsymbol{\mu}}{\partial \theta} = \alpha e^{\jmath 2\pi f_d \ell T} \dot{\mathbf{A}}(\theta) \boldsymbol{\Delta}\mathbf{x}[\ell], \tag{7.2}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial \alpha} = (1, \jmath) \otimes e^{\jmath 2\pi f_d \ell T} \mathbf{A}(\theta) \boldsymbol{\Delta}\mathbf{x}[\ell]. \tag{7.3}$$

Therefore, the expression of each element is given by

$$J_{\theta\theta} = \frac{2L|\alpha|^2}{\sigma_r^2} \operatorname{tr}\left(\dot{\mathbf{A}}(\theta) \mathbf{R}_{\mathsf{q}} \dot{\mathbf{A}}^H(\theta) \right), \tag{7.4}$$

$$J_{\theta\alpha} = \frac{2L|\alpha|^2}{\sigma_r^2} \operatorname{Re}\left\{ \alpha^* \operatorname{tr}\left(\mathbf{A}\left(\theta_p\right) \mathbf{R}_{\mathsf{q}} \dot{\mathbf{A}}^H\left(\theta_l\right) \right) (1, \jmath) \right\}, \tag{7.5}$$

$$J_{\alpha\alpha} = \frac{2L|\alpha|^2}{\sigma_r^2} \operatorname{Re}\left\{ (1, \jmath)^H (1, \jmath) \operatorname{tr}\left(\mathbf{A}(\theta) \mathbf{R}_{\mathsf{q}} \mathbf{A}^H(\theta) \right) \right\}, \tag{7.6}$$

where the derivative of $\mathbf{A}(\theta)$ is given by

$$\begin{aligned} \dot{\mathbf{A}}(\theta) = \frac{\partial \mathbf{A}(\theta)}{\partial \theta} &= \frac{\partial \mathbf{b}(\theta)}{\partial \theta} \mathbf{a}^\top(\theta) + \mathbf{b}(\theta) \frac{\partial \mathbf{a}^\top(\theta)}{\partial \theta} \\ &= \jmath 2\pi\delta \cos\theta \big(\operatorname{diag}(\bar{\mathbf{r}}_X) \mathbf{b}(\theta) \mathbf{a}^\top(\theta) \\ &\quad + \mathbf{b}(\theta) \mathbf{a}^\top(\theta) \operatorname{diag}(\mathbf{r}_X) \big). \end{aligned} \tag{7.7}$$

In (7.7), $\mathbf{r}_X = [0, \cdots, N_r - 1]^\top$ and $\bar{\mathbf{r}}_X = [0, \cdots, N_t - 1]^\top$ represent the Cartesian coordinates of the transmit and receive antenna arrays.

Bibliography

[1] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar. On unlimited sampling and reconstruction. *IEEE Trans. Signal Process.*, 69:3827–3839, December 2020.

[2] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar. On unlimited sampling. In *Intl. Conf. on Sampling Theory and Applications (SampTA)*, July 2017.

[3] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar. Methods and apparatus for modulo sampling and recovery, May 2020.

[4] Ayush Bhandari, Felix Krahmer, and Thomas Poskitt. Unlimited sampling from theory to practice: Fourier-Prony recovery and prototype ADC. *IEEE Trans. Signal Process.*, 70:1131–1141, September 2021.

[5] Ahmet Yazar, S Dogan-Tusha, and Huseyin Arslan. 6G vision: An ultra-flexible perspective. *ITU J. Future Evol. Technol.*, 1(1):121–140, 2020.

[6] J Andrew Zhang, Fan Liu, Christos Masouros, Robert W Heath, Zhiyong Feng, Le Zheng, and Athina Petropulu. An overview of signal processing techniques for joint communication and radar sensing. *IEEE J. Sel. Topics Signal Process.*, 15(6):1295–1315, 2021.

[7] Fan Liu, Yuanhao Cui, Christos Masouros, Jie Xu, Tony Xiao Han, Yonina C. Eldar, and Stefano Buzzi. Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. *IEEE J. Sel. Areas Commun.*, 40(6):1728–1767, 2022.

[8] Fan Liu and Christos Masouros. A tutorial on joint radar and communication transmission for vehicular networks—part i: Background and fundamentals. *IEEE Commun. Lett.*, 25(2):322–326, 2020.

[9] Longfei Yin, Ziang Liu, MR Bhavani Shankar, Mohammad Alaee-Kerahroodi, and Bruno Clerckx. Integrated sensing and communications enabled low earth orbit satellite systems. *IEEE Network*, 2024.

[10] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[11] Bruno Clerckx and Claude Oestges. *MIMO wireless networks: channels, techniques and standards for multi-antenna, multi-user and multi-cell systems*. Academic Press, 2013.

[12] Thomas L Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wireless Commun.*, 9(11):3590–3600, 2010.

[13] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G Larsson, Thomas L Marzetta, Ove Edfors, and Fredrik Tufvesson. Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Process. Mag.*, 30(1):40–60, 2012.

[14] Erik G Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L Marzetta. Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.*, 52(2):186–195, 2014.

[15] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5G be? *IEEE J. Sel. Areas Commun.*, 32(6):1065–1082, 2014.

[16] Ashutosh Sabharwal, Philip Schniter, Dongning Guo, Daniel W Bliss, Sampath Rangarajan, and Risto Wichman. In-band full-duplex wireless: Challenges and opportunities. *IEEE J. Sel. Areas Commun.*, 32(9):1637–1652, 2014.

[17] Kenneth E Kolodziej, Bradley T Perry, and Jeffrey S Herd. In-band full-duplex technology: Techniques and systems survey. *IEEE Trans. Microw. Theory Techn.*, 67(7):3025–3041, 2019.

[18] Zhongshan Zhang, Keping Long, Athanasios V Vasilakos, and Lajos Hanzo. Full-duplex wireless communications: Challenges, solutions, and future research directions. *Proc. IEEE*, 104(7):1369–1409, 2016.

[19] Melissa Duarte, Chris Dick, and Ashutosh Sabharwal. Experiment-driven characterization of full-duplex wireless systems. *IEEE Trans. Wireless Commun.*, 11(12):4296–4307, 2012.

[20] Kenneth E Kolodziej, Joseph G McMichael, and Bradley T Perry. Multitap RF canceller for in-band full-duplex wireless communications. *IEEE Trans. Wireless Commun.*, 15(6):4321–4334, 2016.

[21] Evan Everett, Clayton Shepard, Lin Zhong, and Ashutosh Sabharwal. SoftNull: Many-antenna full-duplex wireless via digital beamforming. *IEEE Trans. Wireless Commun.*, 15(12):8077–8092, 2016.

[22] Björn Debaillie, Dirk-Jan van den Broek, Cristina Lavin, Barend van Liempd, Eric AM Klumperink, Carmen Palacios, Jan Craninckx, Bram Nauta, and Aarno Pärssinen. Analog/RF solutions enabling compact full-duplex radios. *IEEE J. Sel. Areas Commun.*, 32(9):1662–1673, 2014.

[23] Dani Korpi, Joose Tamminen, Matias Turunen, Timo Huusari, Yang-Seok Choi, Lauri Anttila, Shilpa Talwar, and Mikko Valkama. Full-duplex mobile device: Pushing the limits. *IEEE Commun. Mag.*, 54(9):80–87, 2016.

[24] Steven Hong, Joel Brand, Jung Il Choi, Mayank Jain, Jeff Mehlman, Sachin Katti, and Philip Levis. Applications of self-interference cancellation in 5G and beyond. *IEEE Commun. Mag.*, 52(2):114–121, 2014.

[25] Elsayed Ahmed and Ahmed M Eltawil. All-digital self-interference cancellation technique for full-duplex systems. *IEEE Trans. Wireless Commun.*, 14(7):3519–3532, 2015.

[26] Kazuki Komatsu, Yuichi Miyaji, and Hideyuki Uehara. Basis function selection of frequency-domain hammerstein self-interference canceller for in-band full-duplex wireless communications. *IEEE Trans. Wireless Commun.*, 17(6):3768–3780, 2018.

[27] Evan Everett, Achaleshwar Sahai, and Ashutosh Sabharwal. Passive self-interference suppression for full-duplex infrastructure nodes. *IEEE Trans. Wireless Commun.*, 13(2):680–694, 2014.

[28] Insoo Hwang, Bongyong Song, Cong Nguyen, and Samir S Soliman. Digitally controlled analog wideband interference cancellation for in-device spectrum sharing and aggregation. *IEEE J. Sel. Areas Commun.*, 34(11):2838–2850, 2016.

[29] Nanchi Su, Fan Liu, and Christos Masouros. Secure radar-communication systems with malicious targets: Integrating radar, communications and jamming functionalities. *IEEE Trans. Wireless Commun.*, 20(1):83–95, 2020.

[30] Musa Furkan Keskin, Visa Koivunen, and Henk Wymeersch. Limited feedforward waveform design for OFDM dual-functional radar-communications. *IEEE Trans. Signal Process.*, 69:2955–2970, 2021.

[31] Kai Wu, J Andrew Zhang, Xiaojing Huang, and Y Jay Guo. Integrating low-complexity and flexible sensing into communication systems. *IEEE J. Sel. Areas Commun.*, 40(6):1873–1889, 2022.

[32] Xiangnan Liu, Haijun Zhang, Keping Long, Mingyu Zhou, Yonghui Li, and H. Vincent Poor. Proximal policy optimization-based transmit beamforming and phase-shift design in an IRS-aided ISAC system for the THz band. *IEEE J. Sel. Areas Commun.*, 40(7):2056–2069, 2022.

[33] Qin Shi, Liang Liu, Shuowen Zhang, and Shuguang Cui. Device-free sensing in OFDM cellular network. *IEEE J. Sel. Areas Commun.*, 40(6):1838–1853, 2022.

[34] Shanzhi Chen, Ying-Chang Liang, Shaohui Sun, Shaoli Kang, Wenchi Cheng, and Mugen Peng. Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed. *IEEE Wireless Commun.*, 27(2):218–228, 2020.

[35] Xinyu Gao, Linglong Dai, Shuangfeng Han, I Chih-Lin, and Robert W Heath. Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. *IEEE J. Sel. Areas Commun.*, 34(4):998–1009, 2016.

[36] Søren Skovgaard Christensen, Rajiv Agarwal, Elisabeth De Carvalho, and John M Cioffi. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. *IEEE Trans. Wireless Commun.*, 7(12):4792–4799, 2008.

[37] Qingjiang Shi, Meisam Razaviyayn, Zhi-Quan Luo, and Chen He. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Trans. Signal Process.*, 59(9):4331–4340, 2011.

[38] Carlos Baquero Barneto, Sahan Damith Liyanaarachchi, Taneli Riihonen, Mikko Heino, Lauri Anttila, and Mikko Valkama. Beamforming and waveform optimization for OFDM-based joint communications and sensing at mm-Waves. In *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 895–899. IEEE, 2020.

[39] Sahan Damith Liyanaarachchi, Carlos Baquero Barneto, Taneli Riihonen, Mikko Heino, and Mikko Valkama. Joint multi-user communication and MIMO radar through full-duplex hybrid beamforming. In *2021 1st IEEE International Online Symposium on Joint Communications & Sensing (JC&S)*, pages 1–5. IEEE, 2021.

[40] Carlos Baquero Barneto, Taneli Riihonen, Sahan Damith Liyanaarachchi, Mikko Heino, Nuria González-Prelcic, and Mikko Valkama. Beamformer design and optimization for full-duplex joint communication and sensing at mm-Waves. *arXiv preprint arXiv:2109.05932*, 2021.

[41] Chengcheng Xu, Bruno Clerckx, Shiwa Chen, Yijie Mao, and Jianyun Zhang. Rate-splitting multiple access for multi-antenna joint radar and communications. *IEEE J. Sel. Topics Signal Process.*, 15(6):1332–1347, 2021.

[42] Longfei Yin and Bruno Clerckx. Rate-splitting multiple access for dual-functional radar-communication satellite systems. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2022.

[43] Longfei Yin, Yijie Mao, Onur Dizdar, and Bruno Clerckx. Rate-splitting multiple access for 6G—part II: Interplay with integrated sensing and communications. *IEEE Commun. Lett.*, 26(10):2237–2241, 2022.

[44] Onur Dizdar, Aryan Kaushik, Bruno Clerckx, and Christos Masouros. Energy efficient dual-functional radar-communication: Rate-splitting multiple access, low-resolution DACs, and RF chain selection. *IEEE Open J. of the Comm. Society.*, 3:986–1006, 2022.

[45] Rafael Cerna Loli, Onur Dizdar, and Bruno Clerckx. Rate-splitting multiple access for multi-antenna joint radar and communications with partial CSIT: Precoder optimization and link-level simulations. *arXiv preprint arXiv:2201.10621*, 2022.

[46] Kexin Chen, Yijie Mao, Longfei Yin, Chengcheng Xu, and Yang Huang. Rate-splitting multiple access for simultaneous multi-user communication and multi-target sensing. *arXiv preprint arXiv:2306.06458*, 2023.

[47] Luis G Ordoñez, Paul Ferrand, Melissa Duarte, Maxime Guillaud, and Ganghua Yang. On Full-Duplex Radios with Modulo-ADCs. *IEEE Open J. of the Comm. Society*, 2021.

[48] Ayush Bhandari, Felix Krahmer, and Thomas Poskitt. Unlimited sampling from theory to practice: Fourier-Prony recovery and prototype ADC. *IEEE Trans. Signal Process.*, 70:1131–1141, 2021.

[49] Longfei Yin and Bruno Clerckx. Rate-Splitting Multiple Access for Dual-Functional Radar-Communication Satellite Systems. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2022.

[50] Bruno Clerckx, Yijie Mao, Eduard A. Jorswieck, Jinhong Yuan, David J. Love, Elza Erkip, and Dusit Niyato. A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions. *IEEE J. Sel. Areas Commun.*, 41(5):1265–1308, 2023.

[51] Jeonghun Park, Byungju Lee, Jinseok Choi, Hoon Lee, Namyoon Lee, Seok-Hwan Park, Kyoung-Jae Lee, Junil Choi, Sung Ho Chae, Sang-Woon Jeon, Kyung Sup Kwak, Bruno Clerckx, and Wonjae Shin. Rate-splitting multiple access for 6G networks: Ten promising scenarios and applications. *IEEE Netw.*, pages 1–1, 2023.

[52] Juhwan Lee, Jungwoo Lee, Longfei Yin, Wonjae Shin, and Bruno Clerckx. Coordinated rate-splitting multiple access for integrated satellite-terrestrial networks with super-common message. *IEEE Trans. Veh. Technol.*, pages 1–6, 2023.

[53] Yijie Mao, Onur Dizdar, Bruno Clerckx, Robert Schober, Petar Popovski, and H. Vincent Poor. Rate-splitting multiple access: Fundamentals, survey, and future research trends. *IEEE Commun. Surveys Tuts.*, 24(4):2073–2126, 2022.

[54] Xinze Lyu, Sundar Aditya, Junghoon Kim, and Bruno Clerckx. Rate-splitting multiple access: The first prototype and experimental validation of its superiority over SDMA and NOMA. *IEEE Trans. Wireless Commun.*, 2024.

[55] Julie Nutini, Issam Laradji, and Mark Schmidt. Let's make block coordinate descent converge faster: faster greedy rules, message-passing, active-set complexity, and superlinear convergence. *Journal of Machine Learning Research*, 23(131):1–74, 2022.

[56] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.

[57] Qingjiang Shi and Mingyi Hong. Penalty dual decomposition method for nonsmooth nonconvex optimization—Part I: Algorithms and convergence analysis. *IEEE Trans. Signal Process.*, 68:4108–4122, 2020.

[58] Qingjiang Shi, Mingyi Hong, Xiao Fu, and Tsung-Hui Chang. Penalty dual decomposition method for nonsmooth nonconvex optimization—part ii: Applications. *IEEE Trans. Signal Process.*, 68:4242–4257, 2020.

[59] Werner Dinkelbach. On nonlinear fractional programming. *Management science*, 13(7):492–498, 1967.

[60] Kaiming Shen and Wei Yu. Fractional programming for communication systems—part i: Power control and beamforming. *IEEE Trans. Signal Process.*, 66(10):2616–2630, 2018.

[61] Kaiming Shen and Wei Yu. Fractional programming for communication systems—part ii: Uplink scheduling via matching. *IEEE Trans. Signal Process.*, 66(10):2631–2644, 2018.

[62] Meisam Razaviyayn. *Successive convex approximation: Analysis and applications*. PhD thesis, University of Minnesota, 2014.

[63] Gesualdo Scutari, Francisco Facchinei, Peiran Song, Daniel P Palomar, and Jong-Shi Pang. Decomposition by partial linearization: Parallel optimization of multiagent systems. *IEEE Trans. Signal Process.*, 62(3):641–656, 2013.

[64] Angelia Nedić, Jong-Shi Pang, Gesualdo Scutari, Ying Sun, Gesualdo Scutari, and Ying Sun. Parallel and distributed successive convex approximation methods for big-data optimization. *Multi-Agent Optimization: Cetraro, Italy 2014*, pages 141–308, 2018.

[65] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[66] Zhi-Quan Luo and Wei Yu. An introduction to convex optimization for communications and signal processing. *IEEE J. Sel. Areas Commun.*, 24(8):1426–1438, 2006.

[67] Walid Saad, Mehdi Bennis, and Mingzhe Chen. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Netw.*, 34(3):134–142, 2019.

[68] Xingqin Lin, Jingya Li, Robert Baldemair, Jung-Fu Thomas Cheng, Stefan Parkvall, Daniel Chen Larsson, Havish Koorapaty, Mattias Frenne, Sorour Falahati, Asbjorn Grovlen, et al. 5G new radio: Unveiling the essentials of the next generation wireless access technology. *IEEE Commun. Stand. Mag.*, 3(3):30–37, 2019.

[69] Zhiqiang Xiao and Yong Zeng. Waveform design and performance analysis for full-duplex integrated sensing and communication. *IEEE J. Sel. Areas Commun.*, 40(6):1823–1837, 2022.

[70] Yinghui He, Yunlong Cai, Guanding Yu, and Kai-Kit Wong. Joint transceiver design for dual-functional full-duplex relay aided radar-communication systems. *IEEE Trans. Commun.*, 70(12):8355–8369, 2022.

[71] Christian Sturm and Werner Wiesbeck. Waveform design and signal processing aspects for fusion of wireless communications and radar sensing. *Proc. IEEE*, 99(7):1236–1259, 2011.

[72] Preeti Kumari, Sergiy A Vorobyov, and Robert W Heath. Adaptive virtual waveform design for millimeter-wave joint communication–radar. *IEEE Trans. Signal Process.*, 68:715–730, 2019.

[73] Preeti Kumari, Junil Choi, Nuria González-Prelcic, and Robert W Heath. IEEE 802.11 ad-based radar: An approach to joint vehicular communication-radar system. *IEEE Trans. Veh. Technol.*, 67(4):3012–3027, 2017.

[74] Carlos Baquero Barneto, Taneli Riihonen, Matias Turunen, Lauri Anttila, Marko Fleischer, Kari Stadius, Jussi Ryynänen, and Mikko Valkama. Full-duplex OFDM radar with LTE and 5G NR waveforms: Challenges, solutions, and measurements. *IEEE Trans. Microw. Theory Techn.*, 67(10):4042–4054, 2019.

[75] Prabhat Kumar Sharma and Parul Garg. Intelligent reflecting surfaces to achieve the full-duplex wireless communication. *IEEE Commun. Lett.*, 25(2):622–626, 2020.

[76] Ahmad Massud Tota Khel and Khairi Ashour Hamdi. Performance analysis of irs-assisted full-duplex wireless communication systems with interference. *IEEE Commun. Lett.*, 26(9):2027–2031, 2022.

[77] José Mairton Barros da Silva, Ashutosh Sabharwal, Gábor Fodor, and Carlo Fischione. 1-bit phase shifters for large-antenna full-duplex mmWave communications. *IEEE Trans. Wireless Commun.*, 19(10):6916–6931, 2020.

[78] Ming-Min Zhao, Qingqing Wu, Min-Jian Zhao, and Rui Zhang. Exploiting amplitude control in intelligent reflecting surface aided wireless communication with imperfect CSI. *IEEE Trans. Commun.*, 69(6):4216–4231, 2021.

[79] Jongwoo Lee, Joshua Kang, Sunghyun Park, Jae Sun Seo, Jens Anders, Jorge Guilherme, and Michael P. Flynn. A 2.5 mW 80 dB DR 36 dB SNDR 22 MS/s logarithmic pipeline ADC. *IEEE J. Solid-State Circuits*, 44(10):2755–2765, 2009.

[80] Ziang Liu, Ayush Bhandari, and Bruno Clerckx. λ–MIMO: Massive MIMO via modulo sampling. *IEEE Trans. Commun.*, 71(11):6301–6315, 2023.

[81] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar. On unlimited sampling and reconstruction. *IEEE Trans. Signal Process.*, 2020.

[82] Sundar Aditya, Onur Dizdar, Bruno Clerckx, and Xueru Li. Sensing using coded communications signals. *IEEE Open J. of the Comm. Society*, 4:134–152, 2023.

[83] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[84] Merrill Ivan Skolnik. Introduction to radar systems. *New York*, 1980.

[85] Ahmed Masmoudi and Tho Le-Ngoc. A maximum-likelihood channel estimator for self-interference cancellation in full-duplex systems. *IEEE Trans. Veh. Technol.*, 65(7):5122–5132, 2015.

[86] Konstantin Muranov, Md Atiqul Islam, Besma Smida, and Natasha Devroye. On deep learning assisted self-interference estimation in a full-duplex relay link. *IEEE Wireless Commun. Lett.*, 10(12):2762–2766, 2021.

[87] Dan Nguyen, Le-Nam Tran, Pekka Pirinen, and Matti Latva-aho. On the spectral efficiency of full-duplex small cell wireless systems. *IEEE Trans. Wireless Commun.*, 13(9):4896–4910, 2014.

[88] Jingwen Bai and Ashutosh Sabharwal. Distributed full-duplex via wireless side-channels: Bounds and protocols. *IEEE Trans. Wireless Commun.*, 12(8):4162–4173, 2013.

[89] Achaleshwar Sahai, Suhas Diggavi, and Ashutosh Sabharwal. On degrees-of-freedom of full-duplex uplink/downlink channel. In *2013 IEEE Information Theory Workshop (ITW)*, pages 1–5. IEEE, 2013.

[90] Can Karakus and Suhas Diggavi. Opportunistic scheduling for full-duplex uplink-downlink networks. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1019–1023. IEEE, 2015.

[91] Chengcheng Xu, Bruno Clerckx, and Jianyun Zhang. Multi-antenna joint radar and communications: Precoder optimization and weighted sum-rate vs probing power tradeoff. *IEEE Access*, 8:173974–173982, 2020.

[92] Omid Taghizadeh, Jianshu Zhang, and Martin Haardt. Transmit beamforming aided amplify-and-forward MIMO full-duplex relaying with limited dynamic range. *Signal Processing*, 127:266–281, 2016.

[93] Hao Xu, Vikas Kukshya, and Theodore S Rappaport. Spatial and temporal characteristics of 60-GHz indoor channels. *IEEE J. Sel. Areas Commun.*, 20(3):620–630, 2002.

[94] Jiacheng Wang, Hongyang Du, Dusit Niyato, Mu Zhou, Jiawen Kang, and H Vincent Poor. Acceleration estimation of signal propagation path length changes for wireless sensing. *arXiv preprint arXiv:2401.00160*, 2023.

[95] Yanyu Cheng, Dusit Niyato, Hongyang Du, Jiawen Kang, Zehui Xiong, Chunyan Miao, and Dong In Kim. Resource allocation and common message selection for task-oriented semantic information transmission with RSMA. *IEEE Trans. Wireless Commun.*, 2023.

[96] Robert H Walden. Analog-to-digital converter survey and analysis. *IEEE J. Sel. Areas Commun.*, 17(4):539–550, 1999.

[97] Alessio Zappone, Eduard Jorswieck, et al. Energy efficiency in wireless networks via fractional programming theory. *Foundations and Trends® in Communications and Information Theory*, 11(3-4):185–396, 2015.

[98] Fan Liu, Christos Masouros, Athina P Petropulu, Hugh Griffiths, and Lajos Hanzo. Joint radar and communication design: Applications, state-of-the-art, and the road ahead. *IEEE Trans. Commun.*, 68(6):3834–3862, 2020.

[99] Li You, Xiaoyu Qiang, Christos G Tsinos, Fan Liu, Wenjin Wang, Xiqi Gao, and Björn Ottersten. Beam squint-aware integrated sensing and communications for hybrid massive MIMO LEO satellite systems. *IEEE J. Sel. Areas Commun.*, 40(10):2994–3009, 2022.

[100] Longfei Yin and Bruno Clerckx. Rate-splitting multiple access for multigroup multicast and multibeam satellite systems. *IEEE Trans. Commun.*, 69(2):976–990, 2020.

[101] Zhi Wen Si, Longfei Yin, and Bruno Clerckx. Rate-splitting multiple access for multigateway multibeam satellite systems with feeder link interference. *IEEE Trans. Commun.*, 70(3):2147–2162, 2022.

[102] Longfei Yin, Onur Dizdar, and Bruno Clerckx. Rate-splitting multiple access for multigroup multicast cellular and satellite communications: PHY layer design and link-level simulations. In *Proc. IEEE Int. Conf. Commun. Workshop*, pages 1–6, 2021.

[103] Jingfei Huang, Yang Yang, Longfei Yin, Dazhong He, and Qing Yan. Deep reinforcement learning-based power allocation for rate-splitting multiple access in 6G LEO satellite communication system. *IEEE Wireless Commun. Lett.*, 11(10):2185–2189, 2022.

[104] Wali Ullah Khan, Zain Ali, Eva Lagunas, Asad Mahmood, Muhammad Asif, Asim Ihsan, Symeon Chatzinotas, Björn Ottersten, and Octavia A Dobre. Rate splitting multiple access for next generation cognitive radio enabled LEO satellite networks. *IEEE Trans. Wireless Commun.*, 2023.

[105] Alea Schröder, Maik Röper, Dirk Wübben, Bho Matthiesen, Petar Popovski, and Armin Dekorsy. A comparison between RSMA, SDMA, and OMA in multibeam LEO satellite systems. In *WSA & SCC 2023; 26th International ITG Workshop on Smart Antennas and 13th Conference on Systems, Communications, and Coding*, pages 1–6. VDE, 2023.

[106] Minjue He, Hui Zhao, Xiaqing Miao, Shuai Wang, and Gaofeng Pan. Secure rate-splitting multiple access transmissions in LMS systems. *IEEE Commun. Lett.*, 2023.

[107] Steven M. Kay. *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, Inc., 1993.

[108] Ziang Liu, Sundar Aditya, Hongyu Li, and Bruno Clerckx. Joint transmit and receive beamforming design in full-duplex integrated sensing and communications. *IEEE J. Sel. Areas Commun.*, 41(9):2907–2919, 2023.

[109] Maik Röper, Bho Matthiesen, Dirk Wübben, Petar Popovski, and Armin Dekorsy. Beamspace MIMO for satellite swarms. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1307–1312. IEEE, 2022.

[110] Gan Zheng, Symeon Chatzinotas, and Bjorn Ottersten. Generic optimization of linear precoding in multibeam satellite systems. *IEEE Trans. Wireless Commun.*, 11(6):2308–2320, June 2012.

[111] Ahmad Gharanjik, MR Bhavani Shankar, Pantelis-Daniel Arapoglou, Mats Bengtsson, and Björn Ottersten. Robust precoding design for multibeam downlink satellite channel with phase uncertainty. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3083–3087. IEEE, 2015.

[112] Oner Orhan, Elza Erkip, and Sundeep Rangan. Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance. In *2015 Information Theory and Applications Workshop (ITA)*, pages 191–198. IEEE, 2015.

[113] Seokjun Park, Jinseok Choi, Jeonghun Park, Wonjae Shin, and Bruno Clerckx. Rate-splitting multiple access for quantized multiuser MIMO communications. *IEEE Trans. Wireless Commun.*, 22(11):7696–7711, November 2023.

[114] Yijie Mao, Bruno Clerckx, and Victor OK Li. Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA. *EURASIP J. Wireless Commun. Netw.*, 2018:1–54, 2018.

[115] Byungju Lee and Wonjae Shin. Max-min fairness precoder design for rate-splitting multiple access: Impact of imperfect channel knowledge. *IEEE Trans. Veh. Technol.*, 72(1):1355–1359, 2022.

[116] Lucas N Ribeiro, Stefan Schwarz, Markus Rupp, and André LF de Almeida. Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs. *IEEE J. Sel. Topics Signal Process.*, 12(2):298–312, 2018.

[117] Jinseok Choi, Jeonghun Park, and Namyoon Lee. Energy efficiency maximization precoding for quantized massive MIMO systems. *IEEE Trans. Wireless Commun.*, 21(9):6803–6817, 2022.

[118] Ilya Bekkerman and Joseph Tabrikian. Target detection and localization using MIMO radars and sonars. *IEEE Trans. Signal Process.*, 54(10):3873–3883, 2006.

[119] Mark A Richards, Jim Scheer, William A Holm, and William L Melvin. *Principles of modern radar*, volume 1. Citeseer, 2010.

[120] Fan Liu, Ya-Feng Liu, Ang Li, Christos Masouros, and Yonina C. Eldar. Cramér-Rao bound optimization for joint radar-communication beamforming. *IEEE Trans. Signal Process.*, 70:240–253, 2022.

[121] Christian R. Berger, Bruno Demissie, JÖrg Heckenbach, Peter Willett, and Shengli Zhou. Signal processing for passive radar using OFDM waveforms. *IEEE J. Sel. Topics Signal Process.*, 4(1):226–238, 2010.

[122] Yang Yang and Marius Pesavento. Energy efficiency in MIMO interference channels: Social optimality and max-min fairness. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3689–3693. IEEE, 2018.

[123] Alessio Zappone, Emil Björnson, Luca Sanguinetti, and Eduard Jorswieck. Globally optimal energy-efficient power control and receiver design in wireless networks. *IEEE Trans. Signal Process.*, 65(11):2844–2859, 2017.

[124] Fuzhen Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.

[125] Michael Grant, Stephen Boyd, and Yinyu Ye. CVX: Matlab software for disciplined convex programming, 2009.

[126] Yinyu Ye. *Interior point algorithms: Theory and analysis*. John Wiley & Sons, 2011.

[127] Barry R Marks and Gordon P Wright. A general inner approximation algorithm for nonconvex mathematical programs. *Operations research*, 26(4):681–683, 1978.

[128] Li Fan, Shi Jin, Chao-Kai Wen, and Haixia Zhang. Uplink achievable rate for massive MIMO systems with low-resolution ADC. *IEEE Commun. Lett.*, 19(12):2186–2189, December 2015.

[129] Jiayi Zhang, Linglong Dai, Shengyang Sun, and Zhaocheng Wang. On the spectral efficiency of massive MIMO systems with low-resolution ADCs. *IEEE Commun. Lett.*, 20(5):842–845, 2016.

[130] Jiayi Zhang, Linglong Dai, Xu Li, Ying Liu, and Lajos Hanzo. On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems. *IEEE Commun. Mag.*, 56(7):205–211, 2018.

[131] Li Fan, Shi Jin, Chao-Kai Wen, and Haixia Zhang. Uplink achievable rate for massive MIMO systems with low-resolution ADC. *IEEE Commun. Lett.*, 19(12):2186–2189, 2015.

[132] Sven Jacobsson, Giuseppe Durisi, Mikael Coldrey, Ulf Gustavsson, and Christoph Studer. Throughput analysis of massive MIMO uplink with low-resolution ADCs. *IEEE Trans. Wireless Commun.*, 16(6):4038–4051, 2017.

[133] Jinseok Choi, Gilwon Lee, Ahmed Alkhateeb, Alan Gatherer, Naofal Al-Dhahir, and Brian L Evans. Advanced receiver architectures for millimeter-wave communications with low-resolution ADCs. *IEEE Commun. Mag.*, 58(8):42–48, 2020.

[134] Christopher Mollen, Junil Choi, Erik G Larsson, and Robert W Heath. Uplink performance of wideband massive MIMO with one-bit ADCs. *IEEE Trans. Wireless Commun.*, 16(1):87–100, 2016.

[135] Sven Jacobsson, Giuseppe Durisi, Mikael Coldrey, Ulf Gustavsson, and Christoph Studer. One-bit massive MIMO: Channel estimation and high-order modulations. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1304–1309, 2015.

[136] Tim Halsig, Lukas Landau, and Gerhard Fettweis. Information rates for faster-than-nyquist signaling with 1-bit quantization and oversampling at the receiver. In *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2014.

[137] Ji-Jon Sit and Rahul Sarpeshkar. A micropower logarithmic A/D with offset and temperature compensation. *IEEE J. Solid-State Circuits*, 39(2):308–319, 2004.

[138] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar. Unlimited sampling of sparse signals. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018.

[139] Ayush Bhandari, Felix Krahmer, and Ramesh Raskar. Unlimited sampling of sparse sinusoidal mixtures. In *IEEE Intl. Sym. on Information Theory (ISIT)*, June 2018.

[140] Ayush Bhandari and Felix Krahmer. On identifiability in unlimited sampling. In *Intl. Conf. on Sampling Theory and Applications (SampTA)*, July 2019.

[141] Ayush Bhandari. Back in the US-SR: Unlimited sampling and sparse super-resolution with its hardware validation. *IEEE Signal Process. Lett.*, 29:1047–1051, March 2022.

[142] B. F. Logan. Signals designed for recovery after clipping–II. Fourier transform theory of recovery. *AT&T Bell Laboratories Technical Journal*, 63(2):287–306, February 1984.

[143] J.S. Abel and J.O. Smith. Restoring a clipped signal. In *IEEE Intl. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, May 1991.

[144] Amir Adler, Valentin Emiya, Maria G. Jafari, Michael Elad, Rémi Gribonval, and Mark D. Plumbley. Audio inpainting. *IEEE Trans. Acoust., Speech, Signal Process.*, 20(3):922–932, March 2012.

[145] Fabian Esqueda, Stefan Bilbao, and Vesa Valimaki. Aliasing reduction in clipped signals. *IEEE Trans. Signal Process.*, 64(20):5255–5267, October 2016.

[146] Elad Romanov and Or Ordentlich. Above the Nyquist rate, modulo folding does not hurt. *IEEE Signal Process. Lett.*, 26(8):1167–1171, August 2019.

[147] Yicheng Gong, Lu Gan, and Hongqing Liu. Multi-channel modulo samplers constructed from gaussian integers. *IEEE Signal Process. Lett.*, 28:1828–1832, 2021.

[148] Sunil Rudresh, Aniruddha Adiga, Basty Ajay Shenoy, and Chandra Sekhar Seelamantula. Wavelet-based reconstruction for unlimited sampling. In *IEEE Intl. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, pages 4584–4588, April 2018.

[149] Osman Musa, Peter Jung, and Norbert Goertz. Generalized approximate message passing for unlimited sampling of sparse signals. In *IEEE Global Conf. on Sig. and Info. Proc. (GlobalSIP)*, November 2018.

[150] Or Ordentlich, Gizem Tabak, Pavan Kumar Hanumolu, Andrew C. Singer, and Gregory W. Wornell. A modulo-based architecture for analog-to-digital conversion. *IEEE J. Sel. Topics Signal Process.*, pages 1–1, 2018.

[151] Ayush Bhandari and Felix Krahmer. HDR imaging from quantization noise. In *IEEE Intl. Conf. on Image Processing (ICIP)*, pages 101–105, October 2020.

[152] Samuel Fernandez-Menduina, Felix Krahmer, Geert Leus, and Ayush Bhandari. Computational array signal processing via modulo non-linearities. *IEEE Trans. Signal Process.*, 70:2168–2179, July 2021.

[153] Dorian Florescu, Felix Krahmer, and Ayush Bhandari. The surprising benefits of hysteresis in unlimited sampling: Theory, algorithms and experiments. *IEEE Trans. Signal Process.*, 70:616–630, 2022.

[154] Norman C Beaulieu and Mohamed Oussama Damen. Parametric construction of Nyquist-I pulses. *IEEE Trans. Commun.*, 52(12):2134–2142, 2004.

[155] Zhou Zhou, Xiang Gao, Jun Fang, and Zhi Chen. Spherical wave channel and analysis for large linear array in los conditions. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2015.

[156] Elisabeth De Carvalho, Anum Ali, Abolfazl Amiri, Marko Angjelichinoski, and Robert W. Heath. Non-stationarities in extra-large-scale massive MIMO. *IEEE Wireless Communications*, 27(4):74–80, August 2020.

[157] Ayush Bhandari. Back in the US-SR: Unlimited sampling and sparse super-resolution with its hardware validation. *IEEE Signal Process. Lett.*, 29:1047–1051, 2022.

[158] Wim van Etten. Maximum likelihood receiver for multiple channel transmission systems. *IEEE Trans. Commun.*, 24(2):276–283, 1976.

[159] Qing Bai and Josef A Nossek. Energy efficiency maximization for 5G multi-antenna receivers. *Transactions on Emerging Telecommunications Technologies*, 26(1):3–14, 2015.

[160] Allen Gersho and Robert M Gray. *Vector quantization and signal compression*, volume 159. Springer Science & Business Media, 2012.

[161] Yao Wang, Jörn Ostermann, and Ya-Qin Zhang. *Video processing and communications*, volume 1. Prentice hall Upper Saddle River, NJ, 2002.

[162] Makoto Tanahashi and Hideki Ochiai. On the distribution of instantaneous power in single-carrier signals. *IEEE Trans. Wireless Commun.*, 9(3):1207–1215, 2010.

[163] Davide Dardari. Joint clip and quantization effects characterization in OFDM receivers. *IEEE Trans. Circuits Syst. I*, 53(8):1741–1748, 2006.

[164] Michael Bernhard, David Rörich, Thomas Handte, and Joachim Speidel. Analytical and numerical studies of quantization effects in coherent optical OFDM transmission with 100 Gbit/s and beyond. *ITG-Fachtagung Photonische Netze*, pages 34–40, 2012.

[165] PF Panter and Wu Dite. Quantization distortion in pulse-count modulation with nonuniform spacing of levels. *Proceedings of the IRE*, 39(1):44–48, 1951.

[166] Kanti V Mardia, Peter E Jupp, and KV Mardia. *Directional statistics*, volume 2. Wiley Online Library, 2000.

[167] Jianhua Mo and Robert W Heath. Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information. *IEEE Trans. Signal Process.*, 63(20):5498–5512, 2015.

[168] Dan Qiao, Weiqiang Tan, Yuyan Zhao, Chao-Kai Wen, and Shi Jin. Spectral efficiency for massive MIMO zero-forcing receiver with low-resolution ADC. In *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*, pages 1–6. IEEE, 2016.

[169] Ping Yang, Yue Xiao, Ming Xiao, and Shaoqian Li. 6G wireless communications: Vision and potential techniques. *IEEE Netw.*, 33(4):70–75, 2019.

[170] DW Bliss, PA Parker, and AR Margetts. Simultaneous transmission and reception for improved wireless network performance. In *IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 478–482. IEEE, 2007.

[171] Panagiota Lioliou, Mats Viberg, Mikael Coldrey, and Fredrik Athley. Self-interference suppression in full-duplex MIMO relays. In *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, pages 658–662. IEEE, 2010.

[172] Melissa Duarte, Chris Dick, and Ashutosh Sabharwal. Experiment-driven characterization of full-duplex wireless systems. *IEEE Trans. Wireless Commun.*, 11(12):4296–4307, December 2012.

[173] Ayush Bhandari and Ahmed I Zayed. Shift-invariant and sampling spaces associated with the fractional fourier transform domain. *IEEE Trans. Signal Process.*, 60(4):1627–1637, 2011.

[174] Chandan Kumar Sheemar and Dirk Slock. Receiver design and AGC optimization with self interference induced saturation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5595–5599. IEEE, 2020.

[175] Ayush Bhandari, Miguel Heredia Conde, and Otmar Loffeld. One-bit time-resolved imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(7):1630–1641, July 2020.

[176] Zhaoshui He, Andrzej Cichocki, Shengli Xie, and Kyuwan Choi. Detecting the number of clusters in n-way probabilistic clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(11):2006–2021, 2010.

[177] Petre Stoica. List of references on spectral line analysis. *Signal Processing*, 31(3):329–340, 1993.

[178] Ayush Bhandari and Felix Krahmer. On identifiability in unlimited sampling. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–4. IEEE, 2019.

[179] Or Ordentlich, Gizem Tabak, Pavan Kumar Hanumolu, Andrew C Singer, and Gregory W Wornell. A modulo-based architecture for analog-to-digital conversion. *IEEE J. Sel. Topics Signal Process.*, 12(5):825–840, 2018.

[180] Amir Weiss, Everest Huang, Or Ordentlich, and Gregory W Wornell. Blind modulo analog-to-digital conversion. *IEEE Trans. Signal Process.*, 70:4586–4601, 2022.

[181] Amir Weiss, Everest Huang, Or Ordentlich, and Gregory W Wornell. Blind modulo analog-to-digital conversion of vector processes. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.