SENSING, BEAMFORMING AND SELF-INTERFERENCE MITIGATION METHODS FOR FULL-DUPLEX COMMUNICATIONS

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

BY

ANIL KURT

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

AUGUST 2025

Approval of the thesis:

SENSING, BEAMFORMING AND SELF-INTERFERENCE MITIGATION METHODS FOR FULL-DUPLEX COMMUNICATIONS

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

Prof. Dr. Naci Emre Altun
Dean, **Graduate School of Natural and Applied Sciences** _____

Prof. Dr. Nevzat Güneri Gençer
Head of Department, **Electrical and Electronics Engineering** _____

Assist. Prof. Dr. Gökhan Muzaffer Güvensen
Supervisor, **Electrical and Electronics Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Buyurman Baykal
Electrical and Electronics Engineering, METU _____

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

Prof. Dr. Tolga Mete Duman
Electrical and Electronics Engineering, Bilkent University _____

Prof. Dr. Berkan Dülek
Electrical and Electronics Engineering, Hacettepe University _____

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

Date:26.08.2025

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

Name, Surname: Anıl Kurt

Signature :

ABSTRACT

SENSING, BEAMFORMING AND SELF-INTERFERENCE MITIGATION METHODS FOR FULL-DUPLEX COMMUNICATIONS

Kurt, Anıl
Ph.D., Department of Electrical and Electronics Engineering
Supervisor: Assist. Prof. Dr. Gökhan Muzaffer Güvensen

August 2025, 186 pages

Full-Duplex (FD) communication systems enable simultaneous transmission and reception using the same frequency and time resources, potentially doubling the data rate. However, the self-interference (SI) is a bottleneck for FD operation. This thesis investigates SI mitigation and beamforming techniques for FD communication systems, with a focus on time-varying channels and millimeter-wave (mmWave) multiple-input multiple-output (MIMO) architectures. The first part addresses SI cancellation in single-antenna FD systems, examining the effects of nonlinear distortions caused by power amplifier impairments and rapid variations in the SI channel. A novel adaptive-iterative cancellation algorithm is proposed to combat signal-of-interest contamination and reduce complexity by separating nonlinear learning from real-time operation. The second part focuses on far-field SI in mmWave MIMO systems, where environmental echoes from scatterers create SI with spatial and temporal characteristics. A hybrid scheme combining angle-based spatial suppression and delay-based digital SI cancellation is developed. Using sensing methods and a scatterer map, the system adaptively selects the optimal mitigation action for each scatterer to maintain robust communication performance in dynamic environments.

Finally, the third part complements the FD operation by enabling robust angular estimation and beam tracking of user equipments. A low-complexity two-stage beamforming framework is proposed for multi-cluster, time-varying massive MIMO channels, leveraging reduced-dimensional statistical methods and distinguishing fast- and slow-time channel parameters. Overall, the contributions in this thesis offer a holistic SI mitigation and beamforming strategy for emerging FD mmWave systems, addressing both near- and far-field interference while ensuring robustness to environmental dynamics and hardware limitations.

Keywords: Millimeter-wave, MIMO, Full-Duplex, Self-Interference Cancellation, Spatial Suppression.

ÖZ

TAM ÇİFT YÖNLÜ HABERLEŞMELER İÇİN ALGILAMA, HÜZME ŞEKİLLENDİRME VE ÖZ GİRİŞİM AZALTMA YÖNTEMLERİ

Kurt, Anıl
Doktora, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi. Gökhan Muzaffer Güvensen

Ağustos 2025 , 186 sayfa

Tam Çift Yönlü (Full-Duplex, FD) haberleşme sistemleri, aynı frekans ve zaman kaynaklarını kullanarak eşzamanlı gönderim ve alım yapma imkanı sunar. Bu sayede veri iletim oranı teorik olarak iki katına çıkabilir. Ancak, öz girişim (self-interference, SI), FD çalışmanın önündeki en büyük engeldir. Bu tez, zamanla değişen kanallar ve milimetre dalga çoklu giriş-çoklu çıkış (MIMO) mimarileri odağında, FD haberleşme sistemleri için SI bastırma ve hüzme şekillendirme tekniklerini incelemektedir. Tezin ilk bölümü, tek antenli FD sistemlerde SI iptali için güç yükseltici kaynaklı doğrusal olmayan bozulmalar ve hızlı SI kanal değişimlerinin etkisini ele almaktadır. İlgilenilen sinyalin kirleticiliğini azaltmak ve gerçek zamanlı çalışmada doğrusal olmayan öğrenmeyi ayrıştırarak karmaşıklığı düşürmek amacıyla, yeni bir uyarlamalı-iteratif bastırma algoritması önerilmektedir. İkinci bölüm, milimetre dalga MIMO sistemlerinde çevredeki saçıcı nesnelerden gelen yansımaların oluşturduğu uzak-alan SI'ya odaklanmaktadır. Açıya dayalı uzamsal bastırma ile gecikmeye dayalı sayısal SI iptalini birleştiren melez bir yöntem geliştirilmiştir. Sistem, algılama yöntemleri ve bir saçıcı haritası kullanarak her saçıcı için en uygun bastırma eylemini uyarlamalı olarak

seçer ve dinamik çevre koşullarında sağlam haberleşme performansı sağlar. Son olarak, FD çalışmayı destekleyici olarak kullanıcı cihazlarının açısal kestirimi ve hüzme takibi işlenmiştir. Çok kümeli ve zamanla değişen büyük ölçekli MIMO kanalları için, boyut indirgemeli istatistiksel yöntemler ve kanal parametrelerinin hızlı ve yavaş zamanlı ayrımı kullanılarak, düşük karmaşıklıklı iki adımlı bir hüzme şekillendirme çerçevesi önerilmiştir. Bu tezdeki katkılar; gelişmekte olan FD milimetre dalga sistemler için yakın ve uzak alan girişimlerini kapsayan, çevresel dinamiklere ve donanım sınırlamalarına karşı dayanıklı, bütüncül bir SI bastırma ve hüzme şekillendirme stratejisi sunmaktadır.

Anahtar Kelimeler: Milimetre-dalga, MIMO, Tam Çift Yönlü, Öz Girişim İptali, Uzamsal Bastırma.

To my family and friends

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Assistant Professor Gökhan Muzaffer Güvensen, for his constant support and guidance throughout my graduate studies. His wisdom, availability, and encouragement have been invaluable in shaping both my academic and professional development. I have learned a lot through our discussions on research and industrial projects. He is a perfect mentor and the nicest person, to whom I owe a great deal.

I am also thankful to the members of my thesis monitoring committee, Assoc. Prof. Ayşe Melda Yüksel Turgut and Prof. Berkan Dülek, for their insightful comments and constructive feedback, which helped me to improve my work and perspective.

I would like to thank Aselsan Inc. and its engineers for providing the opportunity to participate in several projects that significantly enriched my development. In particular, the full-duplex communication project formed the basis of this thesis. I especially thank Soner Yeşil, Altuğ Kaya, and Hasan Aykut Şatana for their valuable contributions. Our discussions greatly enhanced my knowledge, and their experimental input was instrumental for this research.

I acknowledge the support of TÜBİTAK (2224-A programme, term 2024/3, application no. 1919B022404862) for funding my participation in PIMRC 2024, held in Valencia, Spain, where I presented a paper related to this thesis.

I am also grateful to my friends from our office and research group, Safa Çelik, Uğur Berkay Saraç, Zhichao Zhang, Murat Babek Salman, Hakan Özen, Musa Burak Baytok, Furkan Odunkıran, Cansu Ecemiş, and Melih Can Zerin. I also thank my friends from the department, Nurullah Gülmüş, Lütfullah Tomak, and Haktan Yalçın. I have shared a whole graduate-level education and participated in industrial projects with Uğur and Babek, and spent a lot of time outside of work with Safa, Nurullah, Lütfullah, and Zhichao. Their friendship, academic and non-academic discussions, and countless shared moments over tea, lunches, and dinners made this journey enjoyable.

Finally, I am deeply grateful to my family, who have always supported me, believed in me, and stood by me through the most difficult times. Their love, patience, and encouragement have been my greatest source of strength.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xii

LIST OF TABLES xix

LIST OF FIGURES xxi

LIST OF ABBREVIATIONS xxvi

NOMENCLATURE xxx

CHAPTERS

1 INTRODUCTION 1

1.1 Full-Duplex Communications 1

1.1.1 Transceiver Architectures and Self-Interference 2

1.1.2 SI Effects and Mitigation Strategies 3

1.1.3 Literature Review 5

1.2 Beamforming Towards Users 9

1.2.1 Literature Review 11

1.3 Contributions 13

1.3.1 Contributions of Chapter 3 . . . 13
1.3.2 Contributions of Chapter 4 . . . 15
1.3.3 Contributions of Chapter 5 . . . 18
1.4 The Outline of the Thesis . . . 20
1.5 Notation . . . 20
2 PRELIMINARIES . . . 23
2.1 Channel Model . . . 23
2.1.1 Single-Ray Model . . . 23
2.1.2 Carrier Phase . . . 24
2.1.2.1 Doppler Effect . . . 25
2.1.2.2 Channel Coherence Time . . . 26
2.1.2.3 Correlated Time Variation Model . . . 26
2.1.3 Information Signal . . . 27
2.1.3.1 Modulation Types . . . 27
2.1.3.2 Sensitivity to Distance and Information Delay . . . 28
2.1.4 Spatial Dimension . . . 29
2.1.4.1 Angular Coherence . . . 30
2.1.5 Power . . . 32
2.1.5.1 Element and Array Gain . . . 33
2.1.6 Fast-Time and Slow-Time Parameters . . . 33
2.1.7 Multipath Channel . . . 34
2.1.7.1 Millimeter-Wave Effects . . . 36
2.2 Processing . . . 36

2.2.1 Parameter-Based Processing With Slow-Time Parameters . . . 37
2.2.2 Beamforming . . . 38
2.2.2.1 Instantaneous and Statistical Beamforming Approaches 40
2.2.3 Nonlinear Modeling . . . 41
2.2.4 Adaptive Filtering and Aging . . . 45
2.3 Power Analysis for Self-Interference . . . 47
2.3.1 Power Equations . . . 48
2.3.2 Radiation Regions and Spherical-wave Model . . . 49
2.3.3 Approximated NFSI Power Expression with Far-Field Model . 50
2.3.4 Analog-to-Digital Conversion and Dynamic Range . . . 55
3 AN ADAPTIVE-ITERATIVE NONLINEAR INTERFERENCE CANCELLATION IN TIME-VARYING FULL-DUPLEX CHANNELS . . . 57
3.1 Signal Model . . . 57
3.1.1 SC Transmission Model . . . 58
3.1.2 Time-Varying SI Channel Model . . . 59
3.1.3 Operation Modes . . . 61
3.1.4 Digitally Controlled ASIC for Full-Duplex Radios . . . 61
3.2 Nonlinear Adaptive DSIC with Iterative SoI Detection: Turbo-DSIC . 62
3.2.1 Two-Stage Signal Decomposition for NL SI Prediction . . . 63
3.2.1.1 First Stage - Slow-Time NPL . . . 65
3.2.1.2 Second Stage - Fast-Time Adaptive Linear SI Estimation 67
3.2.2 Receiver Processing for SoI Detection and Regeneration . . . 69
3.2.3 Online Turbo-DSIC Summary . . . 69

3.3 Analysis on SIC Performance under Time-Varying SI Channel and SoI Contamination 72

3.3.1 Residual SI Power after DSIC 74

3.3.2 Offline/Online Learning Schemes 74

3.4 Complexity Analysis 75

3.5 Performance Evaluation 77

3.5.1 Discussion on the feasibility of FD communications under TV SI channels 77

3.5.2 Parameter Selection for Proposed System 82

3.5.3 Synthetic Case with Saleh PA Model 85

3.5.4 Hardware Implementation 88

4 SENSING AND MITIGATION OF MULTI-SCATTERER SELF-INTERFERENCE FOR FULL-DUPLEX MIMO COMMUNICATIONS 93

4.1 System Model 94

4.1.1 Transmitted Signals 94

4.1.2 Channels 95

4.1.3 Received Signals 96

4.1.4 Multi-User Equalization and Precoding 97

4.1.5 Power Expressions 98

4.2 Two-stage SI Mitigation Scheme 99

4.2.1 NFSI Cancellation 100

4.2.2 FFSI Mitigation Methods 102

4.2.2.1 Spatial SI Suppression 102

4.2.2.2 Digital SI Cancellation (DSIC) 103

4.3 Selection Policy for FFSI Mitigation 104
4.3.1 No-SI and No-Action Cases 104
4.3.2 Spatial Suppression via Rx or Tx Beamforming 105
4.3.3 Digital SI Cancellation (DSIC) 106
4.3.4 Determining Actions for Scatterers Based on the Metrics . . . 107
4.4 Slow-Time and Fast-Time Operations 107
4.4.1 Sensing for Scatterer Map 109
4.4.2 Detection-Triggered Fast-Time Sensing for Emerging Scatterers 110
4.4.2.1 Detection and Delay Estimation 110
4.4.2.2 Repetition of Channel Estimation and DSIC 112
4.4.2.3 Initial Angle Estimation 112
4.5 Complexity Analysis 112
4.6 Performance Evaluation 114
4.6.1 Initial Insights for the Single-Scatterer Case 115
4.6.2 Performance Evaluation for the Multi-Scatterer Case 117
4.6.2.1 Illustration of Selected Actions for Scatterers 117
4.6.2.2 Overall Communication Performance on Randomized Scenarios 118
4.6.2.3 Full-Duplex Communication Rate 120
4.6.3 Sensing Performance 120
5 ROBUST STATISTICAL BEAMFORMING WITH MULTI-CLUSTER TRACKING FOR TIME-VARYING MASSIVE MIMO 123
5.1 Utilization in Full-Duplex Systems 124
5.2 System Model 124

5.2.1 Signal Model for Single Carrier Uplink Transmission 125
5.2.2 Multi-Cluster Massive MIMO Channel Model 126
5.2.3 Time Variation Model 128
5.2.4 Practical Two-Stage Training Modes for Time-Varying Massive MIMO 129
5.3 General System Structure 130
5.3.1 DFT-Based Analog Beamformer for Full Cluster Coverage . . 131
5.3.2 Statistical Digital Beamformer for Inter-Cluster Separation . . 131
5.3.3 Intra-Cluster Spatial Channel Matched Filtering (ICS-CMF) . 133
5.4 Two-Stage Parameter Estimation Procedure 133
5.4.1 FT Estimation: Instantaneous Effective Channel (IEC) Estimation 134
5.4.1.1 Beam-Aware Least Squares (BA-LS) Estimation 135
5.4.1.2 Conventional Joint Estimation Techniques 136
5.4.2 ST Estimation: Beam Tracking 136
5.4.2.1 Assumed Slow-Time Model 136
5.4.2.2 Beam-Aware Maximum-Likelihood (BA-ML) Estimator 138
5.4.2.3 Statistical Extended Kalman Filter (SEKF) 139
5.5 Statistical Beamformer Construction 140
5.5.1 Full Dimensional GEB (FD-GEB) 140
5.5.2 Reduced Dimensional GEB (RD-GEB) 141
5.6 Computational Complexity 143
5.7 Numerical Evaluations 144
5.7.1 Auxiliary Definitions 144

5.7.2 Performance Measures . . . 146
5.7.2.1 Average NMSE for BA-LS . . . 146
5.7.2.2 Angular RMSE for Beam Tracking . . . 146
5.7.2.3 Average SINR after ICS-CMF . . . 146
5.7.3 Simulation Settings . . . 147
5.7.4 Numerical Results . . . 149
6 CONCLUSIONS . . . 161
REFERENCES . . . 163
APPENDICES . . . 177
A Information Decorrelation . . . 177
B Calculation of RSIP After Linear SIC . . . 178
B.1 Channel Aging . . . 178
B.2 Derivation of RSIP Expressions for $k_{obs} < k$ . . . 178
B.3 Derivation of RSIP Expressions for $k_{obs} = k$ . . . 181
C Derivations Related to NFSI Cancellation . . . 182
D Expressions for Robust Statistical Beamforming . . . 183
D.1 Angular Estimate for BA-ML . . . 183
D.2 Covariance Matrix in SEKF . . . 183
D.3 Analytical NMSE Expression for BA-LS . . . 184
CURRICULUM VITAE . . . 185

LIST OF TABLES

TABLES

Table 1.1 Literature Comparison for Chapter 4 16

Table 2.1 Boundaries for Radiation Regions 48

Table 2.2 INR Levels 55

Table 3.1 Time Variation and Normalized Doppler Spread 60

Table 3.2 Training Modes 75

Table 3.3 Complexity per CPI 76

Table 3.4 Various SIC Schemes 81

Table 3.5 Simulation Parameters 83

Table 3.6 Hardware Settings and Measurements 89

Table 3.7 BER and SINR for Hardware Data 90

Table 4.1 Prior Performance Metrics 107

Table 4.2 Complexities of Operations per OFDM Symbol 113

Table 5.1 Nomenclature for Chapter 5 125

Table 5.2 FT-CPI and ST-CPI 129

Table 5.3 Average Computational Complexity per FT-CPI 144

Table 5.4 Simulation Parameters 147

LIST OF FIGURES

FIGURES

Figure 1.1 Time-frequency domain resource usage for uplink (UL) and downlink (DL) communications in Time-Division Duplex (TDD), Frequency-Division Duplex (FDD), and Full-Duplex (FD) communications. 1

Figure 1.2 Transceiver types and self-interference types. 3

Figure 1.3 Illustration of a full-duplex communication scenario. 4

Figure 2.1 Illustration of incident waves on a ULA, incidence angle, and relative differences in the propagation distance. 30

Figure 2.2 The spectrum of the nonlinear (NL) information signal. (a) Severe nonlinearity, (b) Mild nonlinearity, (c,d) The mildly nonlinear signal when received at (c) high SNR and (d) low SNR. The vertical axis is normalized to the linear component in (a) and (b), and to the noise level in (c) and (d). 44

Figure 2.3 Analytic MSE versus RLS forgetting factor plots. 47

Figure 2.4 Considered setting of multipanel FD MIMO transceiver. 47

Figure 2.5 DFT beampattern. ($N_a = 32$, $\theta_{BF} = 60^\circ$.) 52

Figure 2.6 Calculated powers against the beamformed angle θ_{BF} with (a) $N_a = 8$, (b) $N_a = 32$. 54

Figure 3.1 Digitally controlled ASIC architecture. 59

Figure 3.2 System diagram for Turbo-DSIC. 62

Figure 3.3 SI estimation topologies. (LTI: linear time-invariant system, NLTI: nonlinear time-invariant system.) 64

Figure 3.4 Progress of Turbo-DSIC algorithm in time. 71

Figure 3.5 ASE versus N. (SI SNR=70dB, SoI SNR=20dB, f_D^{norm}=1 ppm.) 78

Figure 3.6 ASE versus K. (SI SNR=70dB, SoI SNR=20dB, N=100 Symbols.) 79

Figure 3.7 BER Performances of different SIC schemes against time variation in the SI channel. ($N = 100$, GMP with (5,16,0) for online and offline schemes.) 81

Figure 3.8 Performance against forgetting factor λ. (Saleh model, 30 dB SoI SNR, GMP for AONL, $L_{NL} = 16$, $L_{dsic} = 3$, f_D^{norm}=1 ppm.) 82

Figure 3.9 BER performance against time variation in SoI Channel. (Saleh Model, Turbo-DSIC with 5 iterations, GMP for AONL.) 84

Figure 3.10 Spectra of signals before and after the ASIC stage for the Saleh model. 85

Figure 3.11 BER performance against SoI SNR. (Saleh model, GMP for AONL.) 86

Figure 3.12 AIR performance against SoI SNR. (Saleh model, GMP for AONL.) 88

Figure 3.13 Experimental RF setup for in-band full-duplex transceiver. . . . 89

Figure 3.14 BER performance against SoI SNR. (HW data, $L_{NL} = 4$.) 90

Figure 3.15 BER performance against assumed CPI length. (HW Data, GMP for AONL, $L_{NL} = 16$.) 91

Figure 4.1 General system diagram. (Ch.: Channel, Est.: Estimator, MU: Multi-user, Eq.: Equalizer, FFT: Fast Fourier transform, IFFT: Inverse FFT). 94

Figure 4.2 Exemplary scenario illustration and angular convention. 95

Figure 4.3 Regular FD communication and slow-time sensing intervals, complemented by detection-triggered fast-time sensing for emerging scatterers using UL communication signals within fast-time Rx processing flow. (Detection-triggered operations are dashed. ST: Slow-Time, FT: Fast-Time, Op.: Operations, Sc.: Scatterer.) 109

Figure 4.4 Beampatterns and exposed far-field SI power. ($K_d = 4$, $K_u = 4$, $K_s = 1$. User positions are as shown with markers.) 115

Figure 4.5 Calculated action metrics and color-coded rankings of actions according to the angle of the considered scatterer. ($K_d = 4$, $K_u = 4$, $K_s = 1$. User positions are as shown with markers.) 115

Figure 4.6 Worst-case UL SINR and DL SNR against the angle of a single scatterer. ($K_d = 4$, $K_u = 4$, $K_s = 1$. Markers show the user positions.) . 116

Figure 4.7 Action choices and performances on an exemplary scenario. ($K_d = 4$, $K_u = 4$, $K_s = 13$. User and scatterer positions are as shown with markers.) 117

Figure 4.8 Worst-case UL SINR against the number of scatterers. ($K_d = 4$, $K_u = 4$, with the scenario described in Section 4.6.2.2.) 118

Figure 4.9 Worst-case UL SINR against RCS σ. ($K_d = 4$, $K_u = 4$, $K_s = 13$, with the scenario described in Section 4.6.2.2. 10 dBsm RCS is used for all the other simulations. Adding 36 dB to RCS gives the maximum INR $P_k^{(s)}/N_0$.) 119

Figure 4.10 CDF of UL/DL AIR where user and scatterer positions are random. ($K_d = 4$, $K_u = 4$, $K_s = 13$, with the scenario described in Section 4.6.2.2.) 121

Figure 4.11 Performance of detection and mitigation mechanisms and initial angular estimation in case of an emerging single scatterer. ($K_d = 4$, $K_u = 4$, and $K_s = 13$ known scatterers. Markers show user and scatterer positions.) 122

Figure 5.1 Signaling schemes and coherent processing intervals. 126

Figure 5.2 Illustration of users and signal clusters. ($U = 2$, $M = 3$, $L = 3$; $\mathcal{U}(m) = u$ is defined by $\mathcal{U}(1) = 1, \mathcal{U}(2) = 2, \mathcal{U}(3) = 2$). 127

Figure 5.3 Beamforming structure. 131

Figure 5.4 Notation of variables according to the processing stage for M=2. 132

Figure 5.5 Estimation structure for the $\bar{m}^{\text{th}}$ signal cluster. 134

Figure 5.6 One realization of AoAs from the movement model. 148

Figure 5.7 Effect of D_m on the beam pattern after DBF for the first cluster ($m = 1$). (a) All clusters, (b) interference area, (c) intended area. 149

Figure 5.8 Comparison of instantaneous and statistical beamforming. ($m = 1$, $p_{\text{max}} = 1000$, Genie-Aided mode.) 151

Figure 5.9 Comparison of instantaneous channel estimation techniques. ($m = 1$, $p_{\text{max}} = 1000$, Genie-Aided mode.) 152

Figure 5.10 RMS error of angular estimates at the end of an ST-CPI depending on a given error upon which beamformers are constructed at the beginning of the ST-CPI. ($m = 1$, $P = 1000$.) 153

Figure 5.11 Tracking behaviors of angular estimation techniques: (a) BA-ML with RD-GEB, (b) SEKF with RD-GEB, (c) OMP with ABF, (d) Periodogram with ABF (Blue lines: true AoA, red lines: estimates, $p_{\text{max}} = 100000$, $P = 1000$, Self-Driven mode.) 154

Figure 5.12 CDF of error in the angular estimates. ($m = 1$, $p_{\text{max}} = 100000$, $P = 1000$, Self-Driven mode.) 155

Figure 5.13 CDF of NMSE for BA-LS with different beamformers. ($m = 1$, $p_{\max} = 100000$, $P = 1000$.) 156

Figure 5.14 Performances against the length of an ST-CPI: (a) RMS error for angular estimators, (b) Average SINR. ($m = 1$, $p_{\max} = 100000$.) 157

Figure 5.15 Performances against the length of an ST-CPI: (a) RMS error for angular estimators, (b) Average SINR. ($m = 1$, $p_{\max} = 100000$, $T_{\mathrm{F}} = 100\mu s$, RD-GEB, Self-Driven BF.) 158

Figure 5.16 Performances against the speed of the sources: (a) RMS error for angular estimators, (b) Average SINR. ($m = 1$, $p_{\max} = 100000$, RD-GEB, Self-Driven BF.) 159

LIST OF ABBREVIATIONS

ABBREVIATIONS	
5G	5th Generation
6G	6th Generation
ABF	Analog Beamformer
ADC	Analog-to-Digital Converter
AIR	Achievable Information Rate
AoA	Angle of Arrival
AONL	ASIC-Output Nonlinear
AS	Angular Spread
ASE	Average Spectral Efficiency
ASIC	Analog Self-Interference Cancellation
AWGN	Additive White Gaussian Noise
BA	Beam-Aware
BA-LS	Beam-Aware Least Squares
BA-ML	Beam-Aware Maximum Likelihood
BER	Bit Error Rate
BF	Beamforming/Beamformer
BS	Base Station
CCM	Channel Covariance Matrix
CDF	Cumulative Distribution Function
CFAR	Constant False Alarm Rate
CMF	Channel Matched Filter
CPI	Coherent Processing Interval
DAC	Digital-to-Analog Converter

dB	Decibel
DBF	Digital Beamformer
DFT	Discrete Fourier Transform
DL	Downlink
DSIC	Digital Self-Interference Cancellation
EKF	Extended Kalman Filter
FD	Full-Dimensional (Only in Chapter 5)
FD	Full-Duplex
FDD	Frequency-Division Duplex
FFSI	Far-Field Self-Interference
FFT	Fast Fourier Transform
FT	Fast-time
GEB	Generalized Eigenbeamformer
GMI	Generalized Mutual Information
GMP	Generalized Memory Polynomial
GPR	Gaussian Process Regression
HBF	Hybrid Beamformer
HD	Half-Duplex
ICI	Inter-Carrier Interference
IEC	Instantaneous Effective Channel
IFFT	Inverse Fast Fourier Transform
iid	Independent and Identically Distributed
INR	Interference-to-Noise Ratio
ISAC	Integrated Sensing and Communications
ISI	Inter-Symbol Interference
LMS	Least Mean Squares
LNA	Low-Noise Amplifier

LOS	Line-of-Sight
LS	Least Squares
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
mmWave	Millimeter-wave
MSE	Mean Squared Error
MU	Multi-User
MUI	Multi-User Interference
NDS	Normalized Doppler Spread
NFSI	Near-Field Self-Interference
NL	Nonlinear
NLOS	Non-Line-of-Sight
NMSE	Normalized Mean Squared Error
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
OMP	Orthogonal Matching Pursuit
PA	Power Amplifier
PDF	Probability Density Function
PDP	Power-Delay Profile
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
RCS	Radar Cross Section
RD	Reduced-Dimensional
RF	Radio Frequency
RFC	Radio Frequency Chain
RLS	Recursive Least Squares

RMS	Root Mean Square
RMSE	Root Mean Squared Error
RRC	Root Raised Cosine
RSI	Residual Self-Interference
RSIP	Residual Self-Interference Power
Rx	Receive
SEKF	Statistical Extended Kalman Filter
SI	Self-Interference
SIC	Self-Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SoI	Signal of Interest
ST	Slow-time
TDD	Time-Division Duplex
TV	Time-varying
Tx	Transmit
UE	User Equipment
UL	Uplink
ULA	Uniform Linear Array
ZF	Zero-Forcing

NOMENCLATURE

Common Variables

$\boldsymbol{a}(\theta)$	ULA steering vector for angle θ
c	Speed of light
f_c	Carrier frequency
λ	Wavelength OR RLS forgetting factor
n	Time/Delay index
P	Power OR number of FT-CPIs/OFDM symbols in an interval
p	FT-CPI/OFDM symbol index
σ^2	Variance
θ	Angle
v	Speed/velocity
W	Bandwidth

Chapter 3

α	AR(1) correlation coefficient across CPIs for the FFSI channel
$a_n[k]$	Transmitted M-QAM symbols
$d_n[k]$	SoI symbols
$\eta_n[k]$	Noise
$\eta_n^{\downarrow}[k]$	Downsampled noise
f_D	Doppler spread
f_D^{norm}	Normalized Doppler spread
f_n^{Rice}	Rician factor for SI channel
$\bar{h}_n$	Mean (NFSI) component of the SI channel
$\hat{h}_n^{SoI\downarrow}$	Estimated symbol-rate SoI channel

$\tilde{h}_n[k]$	Time-varying Rayleigh (FFSI) component of the SI channel
h_n^{aux}	Auxiliary line channel
$h_n^{SI}[k]$	Self-interference (SI) channel (CPI k, delay n)
$h_n^{SoI}[k]$	Signal-of-interest (SoI) channel (CPI k, delay n)
I	Number of iterations
i	Iteration index
K	CPIs in data mode
k	CPI index
K_1	CPIs in mode-1
K_2	CPIs in mode-2
K_{aux}	GMP nonlinearity parameter for auxiliary NL model
K_{GMP}	GMP nonlinearity parameter for AONL model
K_{NN}	Number of neurons in NN for AONL model
L	SI channel size
L_{asic}	Linear ASIC filter memory
L_{aux}	GMP memory parameter for auxiliary NL model
L_{dsic}	Linear SI predictor memory at symbol-rate
L_{NL}	Common memory for AONL model
μ	Upsampling ratio
M	QAM Size
M_{aux}	GMP cross memory parameter for auxiliary NL model
M_{GMP}	GMP cross memory parameter for AONL model
N	Symbols in a CPI
n	Delay index (represented time depends on the downsampling situation)
N_{M0}	Symbols in mode-0
$\mathbf{\Phi}[k]$	Joint convolution matrix (from $\phi^{old\downarrow}, \hat{\psi}^{\downarrow}$)
$\hat{\boldsymbol{\phi}}^{\downarrow}[k,i]$	Estimated residual SI (iteration i)

$\hat{\Phi}(\cdot)$	AONL Model
$\hat{\Psi}(\cdot)$	Auxiliary NL model
$\Phi(\cdot)$	True NL mapping from transmitted signal $\{x_n\}$ to residual SI
$\phi_n[k]$	Residual SI after ASIC
$\phi_n^{\downarrow}[k]$	Downsampled residual SI after ASIC
ϕ_n^{M0}	Residual SI during Mode-0
$\phi_n^{old}[k]$	Residual-SI prediction using actual transmit signal and aged AONL model
$\Psi(\cdot)$	PA nonlinearity function
$\psi_n[k]$	PA output sample at time n
p_n	Unit-energy pulse shaping filter
$\boldsymbol{R}[k]$	RLS correlation matrix
$\hat{r}_n^{\downarrow}[k,i]$	Regenerated SoI at symbol rate
ρ_n	Power-delay profile for the SI channel (before normalization)
ρ_s^2	SoI power
ρ_x^2	SI power
$\tilde{r}_n^{\downarrow}[k,i]$	Signal after DSIC in iteration i
$r_n[k]$	SoI signal after SoI channel (received)
$r_n^{\downarrow}[k]$	Downsampled SoI
R_{avg}	Average spectral efficiency (ASE)
σ_n^2	Variance of the FFSI
σ_η^2	Noise variance
$s_n[k]$	SoI signal before SoI channel (at the transmitting end)
v	Scatterer velocity
$v_n[k]$	Innovation for the FFSI channel variation model
$v_n^{SI}[k,i]$	Residual interference after DSIC
$v_n^{SoI}[k,i]$	Residual SoI contamination after SoI cancellation
$\boldsymbol{w}^{\downarrow}[k,i]$	Linear SI predictor (two-input) at iteration i

$\boldsymbol{w}_0^{\downarrow}$	Initial SI predictor from Mode-2
$w_n^1[k]$	Linear SI predictor component mapping $\{\phi_n^{old}\}$ to ϕ_n
$w_n^2[k]$	Linear SI predictor component mapping $\{\hat{\psi}_n\}$ to ϕ_n
w_n^{asic}	Linear ASIC filter
$w_n^{dsic,1\downarrow}[k]$	Symbol-rate counterpart of w_n^1
$w_n^{dsic,2\downarrow}[k]$	Symbol-rate counterpart of w_n^2
$x_n[k]$	Transmit signal in the kth CPI
x_n^{M0}	Training sequence used in Mode-0
$\boldsymbol{y}[k]$	RLS recursion vector obtained at the end of CPI k
$\tilde{\boldsymbol{y}}[k,i]$	RLS recursion vector at iteration i in CPI k
$y_n[k]$	Total received signal after ASIC
$y_n^{\downarrow}[k]$	Downsampled total received signal
$y_n^{MF}[k]$	Matched-filtered received signal
y_n^{SI}	Received signal (SI and noise) without SoI in training modes
$y_n^{total}[k]$	Total received signal before processing
$z_n^{\downarrow}[k,i]$	Residual SI observation after SoI cancellation

Chapter 4

$\alpha_k^{(d)}$	Complex gain of DL channel for DL user k
$\alpha_k^{(s)}$	Complex gain of SI channel for scatterer k
$\alpha_k^{(u)}$	Complex gain of UL channel for UL user k
$\mathcal{A}$	Action set
$a^{(d)}$	Superscript d indicates relation to DL
$a^{(s)}$	Superscript s indicates relation to scatterer/SI
$a^{(u)}$	Superscript u indicates relation to UL
$\boldsymbol{C}$	Rx beamforming matrix
$\bar{\boldsymbol{D}}$	Training data stack used in joint estimation
$\boldsymbol{D}^{(d)}$	Frequency-domain DL data matrix

$\boldsymbol{D}^{(u)}$	Frequency-domain UL data matrix
$\hat{\boldsymbol{D}}^{(d)}$	Soft estimates of DL symbols
$\hat{\boldsymbol{D}}^{(u)}$	Soft estimates of UL symbols after MU equalization
d_s	Distance from scatterer to BS
d_u	Distance from UL user to BS
$d_{k,m}^{(d)}$	Information symbol of DL user k on subcarrier m
$d_{k,m}^{(u)}$	Information symbol of UL user k on subcarrier m
$\boldsymbol{E}_0^{(s)}$	NFSI channel estimation error
$\boldsymbol{E}_l$	Diagonal delay matrix
$\boldsymbol{F}$	DFT (FFT) matrix with
$\bar{\gamma}_{k_s}^{(a)}$	Mean normalized performance metric for scatterer k_s under action a
$\boldsymbol{G}^{(d)}$	Effective DL channel matrix
$\boldsymbol{G}^{(u)}$	Effective UL channel matrix
$\boldsymbol{g}_k^{(d)}$	Effective DL channel for user k
$\boldsymbol{G}_k^{(s)}$	Effective FFSI channel for scatterer k
$\boldsymbol{g}_k^{(u)}$	Effective UL channel for user k
$\Gamma^{(d,w)}$	Worst-case DL SNR across DL users
$\Gamma^{(u,w)}$	Worst-case UL SINR across UL users
$\Gamma_k^{(d)}$	Measured post-processing SINR/SNR for DL user k
$\gamma_k^{(d)}$	SNR metric for DL user k (case-dependent)
$\Gamma_k^{(u)}$	Measured post-processing SINR/SNR for UL user k
$\gamma_k^{(u)}$	SINR metric for UL user k (case-dependent)
$\hat{\boldsymbol{G}}^{(u)}$	Estimated effective UL channel matrix
$\boldsymbol{H}^{(d)}$	DL channel matrix
$\boldsymbol{H}^{(u)}$	UL channel matrix
$\boldsymbol{H}_0^{(s)}$	NFSI
$\boldsymbol{h}_k^{(d)}$	Spatial channel vector for DL user k

$\boldsymbol{H}_k^{(s)}$	FFSI channel of scatterer k
$\boldsymbol{h}_k^{(u)}$	Spatial channel vector for UL user k
$\hat{\boldsymbol{H}}_0^{(s)}$	Estimate of the NFSI channel
$I_k^{(d)}$	AIR for DL user
$I_k^{(u)}$	AIR for UL user
k	UL user, DL user, or scatterer index, determined by superscripts u, d, s
K_d	Number of DL users
K_s	Number of scatterers
K_u	Number of UL users
$l_k^{(s)}$	Delay (in samples) associated with scatterer k
$\boldsymbol{M}^{(d)}$	DL effective channel on intended DL symbols
$\boldsymbol{M}^{(s)}$	New SI channel seen after MU equalization
$\boldsymbol{M}^{(u)}$	UL effective channel seen after MU equalization
$\hat{\boldsymbol{M}}_l^{(s)}$	Estimate for SI channel at delay l seen after MU equalization
m	Subcarrier index
$\boldsymbol{N}$	Thermal noise in $\boldsymbol{R}$
$\boldsymbol{N}^{(d)}$	DL noise term at users
$\boldsymbol{N}^{(z)}$	Post-beamforming noise in frequency-domain
n	Time index
N_0	Noise variance
N_a	Number of antenna elements in each ULA at the BS
N_c	Number of OFDM subcarriers
$N_{c,tr}$	Number of OFDM subcarriers for training (pilots)
p	OFDM symbol index
$P_k^{(d)}$	Channel variance for DL user k
$P_k^{(s)}$	FFSI channel variance for scatterer k
$P_k^{(u)}$	Channel variance for UL user k

P_{BS}	Total BS transmit power
P_{FA}	False-alarm probability in CFAR detection
P_{UE}	UL user transmit power
$\boldsymbol{q}_k^{(d)}$	Steering vector for DL user k at angle $\theta_k^{(d)}$
$\boldsymbol{q}_k^{(s)}$	Steering vector for scatterer k at angle $\theta_k^{(s)}$
$\boldsymbol{q}_k^{(u)}$	Steering vector for UL user k at angle $\theta_k^{(u)}$
$Q_k^{(ds)}$	The inner product of steering vectors for intended scatterer and DL user k
$Q_k^{(us)}$	The inner product of steering vectors for intended scatterer and UL user k
$\boldsymbol{R}$	BS received signal
$\boldsymbol{R}_0$	Residual after NFSI cancellation in $\boldsymbol{R}$
ρ_l	Detection metric for delay l
$\boldsymbol{S}^{(d)}$	Time-domain DL data matrix
$\boldsymbol{S}^{(u)}$	Time-domain UL data matrix
$\boldsymbol{S}_l^{(d)}$	Delayed DL time-domain signal
$\mathcal{S}$	Set of indices of scatterers
$\mathcal{S}_{DSIC}$	Set of scatterers mitigated by digital SI cancellation
$\mathcal{S}_{NA}$	Set of scatterers requiring no action
$\mathcal{S}_{Rx}$	Set of scatterers mitigated by Rx beamforming
$\mathcal{S}_{Tx}$	Set of scatterers mitigated by Tx beamforming
σ	Radar cross section (RCS) of a scatterer
$s_{k,n}^{(d)}$	DL time-domain sample for DL user k at time index n
$s_{k,n}^{(u)}$	UL time-domain sample of UL user k at time index n
$\boldsymbol{T}$	Multi-user zero-forcing precoder
$\theta_k^{(d)}$	Angle of DL user k
$\theta_k^{(s)}$	Angle of scatterer k
$\theta_k^{(u)}$	Angle of UL user k

$\mathcal{U}_d$	Set of indices of DL users
$\mathcal{U}_u$	Set of indices of UL users
$\boldsymbol{W}$	Statistical Tx beamforming matrix
$\boldsymbol{\Xi}^{(u)}$	Aggregate residual term (residual SI and noise) in UL equalization
$\boldsymbol{X}$	Beamformed DL signal at the BS
$\boldsymbol{Z}$	Rx-beamformed, frequency-domain UL received signal
$\boldsymbol{Z}_0^{(s)}$	Residual NFSI term in $\boldsymbol{Z}$
$\hat{\boldsymbol{Z}}_{si}$	Regenerated SI to be subtracted in DSIC

Chapter 5

$\alpha_{l,p}^{(m)}$	Channel complex gain
$\boldsymbol{A}$	Transition matrix for movement model
$b_{n,p}^{(u)}$	Symbols from users
$\Delta_p^{(m)}$	Angular spread
D_m	Number of DBF outputs
$E^{(m)}$	Cluster power
$\bar{\boldsymbol{h}}_p^{(m)}$	Effective channel after ABF
$\boldsymbol{h}_p^{(m)}$	Channel at antennas
$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}$	IEC estimate
$\tilde{\boldsymbol{h}}_p^{(m,m')}$	Effective channel of cluster m' after DBF for cluster m
$\tilde{\boldsymbol{h}}_p^{(m,m)}$	IEC for cluster m
k	ST-CPI index
M	Number of clusters
m	Cluster index
$\boldsymbol{\nu}_p^{(m)}$	Innovation for movement model
N	Number of antennas
n	Symbol index
N_0	Noise variance

N_{F}	Number of training symbols in an FT-CPI
N_{S}	Number of data symbols in an FT-CPI
$\bar{\boldsymbol{\Psi}}_k^{(\mathrm{ST})}$	Total covariance matrix after ABF assumed for ST processing design
$\bar{\boldsymbol{\Psi}}_p$	Total covariance matrix after ABF
$\boldsymbol{\Psi}_k^{(\mathrm{ST})}$	Total covariance matrix at antennas assumed for ST processing design
$\boldsymbol{\Psi}_p$	Total covariance matrix at antennas
$\tilde{\boldsymbol{\Psi}}_p^{(m)}$	Total covariance matrix after DBF for cluster m
P	Number of FT-CPIs in an ST-CPI
p	FT-CPI index
$\bar{\boldsymbol{R}}_k^{(\mathrm{ST},m)}$	Channel covariance matrix after ABF assumed for ST processing design
$\bar{\boldsymbol{R}}_p^{(m)}$	Channel covariance matrix after ABF
$\boldsymbol{R}_k^{(\mathrm{ST},m)}$	Channel covariance matrix at antennas assumed for ST processing design
$\boldsymbol{R}_p^{(m)}$	Channel covariance matrix at antennas
$\boldsymbol{r}_{n,p}$	ABF output signal
$\tilde{\boldsymbol{R}}_p^{(m,m')}$	Channel covariance matrix of cluster m' after DBF for cluster m
R	Number of ABF outputs (RFCs)
$\boldsymbol{S}$	ABF matrix
$s_{n,p}^{(m)}$	Symbols from clusters
$\boldsymbol{T}^{(m)}$	Total BF matrix
$\theta_k^{(\mathrm{ST},m)}$	AoA assumed for ST processing design
$\theta_p^{(m)}$	Mean AoA
U	Number of users
u	User index
$\boldsymbol{W}^{(m)}$	DBF matrix
$\boldsymbol{\xi}_p^{(m)}$	IEC estimation error
$\boldsymbol{x}_k^{(\mathrm{ST},m)}$	Movement state vector assumed for ST processing design

$\boldsymbol{x}_p^{(m)}$	State vector for movement model
$\boldsymbol{y}_{n,p}$	Received signal at antennas
$\boldsymbol{z}_{n,p}^{(m)}$	DBF output signal

CHAPTER 1

INTRODUCTION

1.1 Full-Duplex Communications

Wireless communication systems have traditionally relied on *half-duplex* techniques to avoid interference between transmission and reception. Two common approaches are *time-division duplexing (TDD)*, where uplink and downlink occur in different time slots, and *frequency-division duplexing (FDD)*, where separate frequency bands are used. These methods ensure isolation between transmission (downlink) and reception (uplink) but divide the available spectral resources, limiting efficiency.

In contrast, *full-duplex (FD)* communication enables simultaneous transmission and reception over the same frequency and time resources, as shown in Fig. 1.1. Theoretically, this can double the spectral efficiency. For a base station, downlink corresponds to transmission and uplink to reception; for user equipments (UEs), the roles are reversed. Eliminating resource division, FD holds the promise of substantial capacity enhancement in next-generation wireless systems.

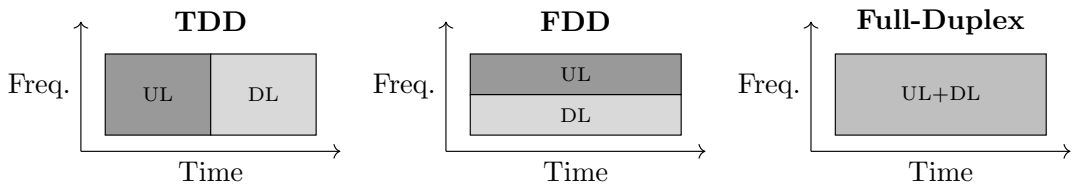


Figure 1.1: Time-frequency domain resource usage for uplink (UL) and downlink (DL) communications in Time-Division Duplex (TDD), Frequency-Division Duplex (FDD), and Full-Duplex (FD) communications.

The potential of doubling spectral efficiency makes in-band full-duplex (FD) systems attractive for future communication standards [1,2]. For instance, FD communication is an important candidate to replace half-duplex (HD) ones in short-reach communications, where high throughput demand is present [3]. Beyond traditional use cases, FD communication supports a wide range of applications, including:

- Wireless local area networks [1],
- Decode-and-forward relay architectures [4, 5],
- Integrated sensing and communication (ISAC) [6],
- Full-duplex wireless backhaul and fronthaul [7, 8].

1.1.1 Transceiver Architectures and Self-Interference

FD systems can be implemented using either *shared-antenna* or *separate-antenna* architectures, as shown in Fig. 1.2. Shared-antenna systems rely on a *circulator* to route signals through a single-antenna, whereas separate-antenna systems isolate Tx and Rx paths physically. Both architectures may include antenna arrays for beamforming. Figure 1.3 demonstrates a separate-antenna array form.

FD transmission introduces several types of interference [9]. Among these, *self-interference (SI)*, the leakage from transmitting antennas to receiving antennas within a transceiver, is a bottleneck for FD communications [1]. SI manifests in two primary forms [10, 11] as shown in Fig. 1.3:

- **Near-field SI (NFSI):**
  - In separate-antenna systems, it is caused by direct electromagnetic coupling between Tx and Rx antennas.
  - In shared-antenna systems, it originates from circuit-level leakage due to non-ideal circulators.

  NFSI is typically stronger, nearly delay-free, and time-invariant due to the fixed transceiver structure.

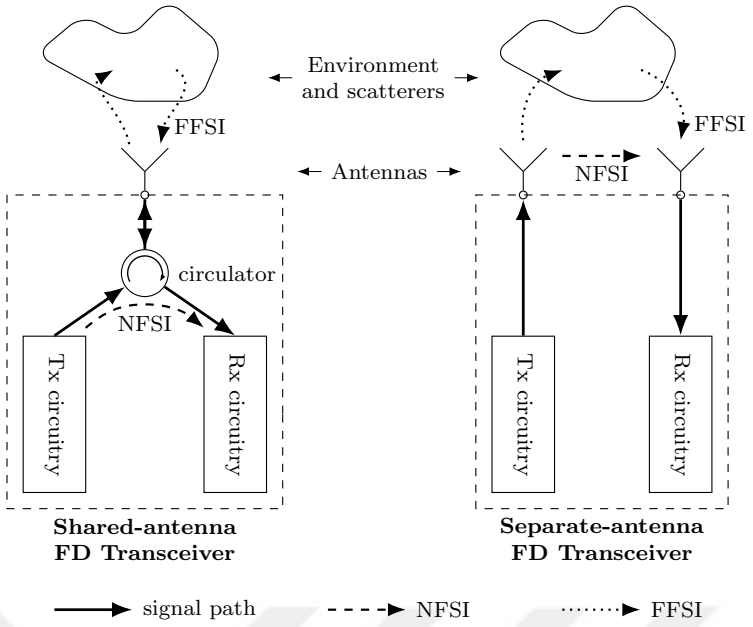


Figure 1.2: Transceiver types and self-interference types.

- **Far-field SI (FFSI):**
  - Arises from environmental reflections and echoes of the transmitted signal.
  - Delayed, weaker due to propagation loss, and time-varying due to mobility and environmental dynamics.

While NFSI dominates in magnitude, FFSI becomes significant in rich-scattering or mobile environments, particularly at millimeter-wave (mmWave) frequencies.

1.1.2 SI Effects and Mitigation Strategies

SI impacts the FD receiver at two critical levels: analog-to-digital converter (ADC) saturation and contamination. At the first level, SI can exceed the dynamic range of the ADC when it is too strong. This causes a complete loss of the received signal, necessitating strong *isolation* and *analog SI cancellation*. This method uses an auxiliary transmit chain to regenerate the SI waveform and RF cancellers to subtract it before digitization. This is mandatory for shared-antenna systems and sometimes optional for separate-antenna ones with sufficient isolation.

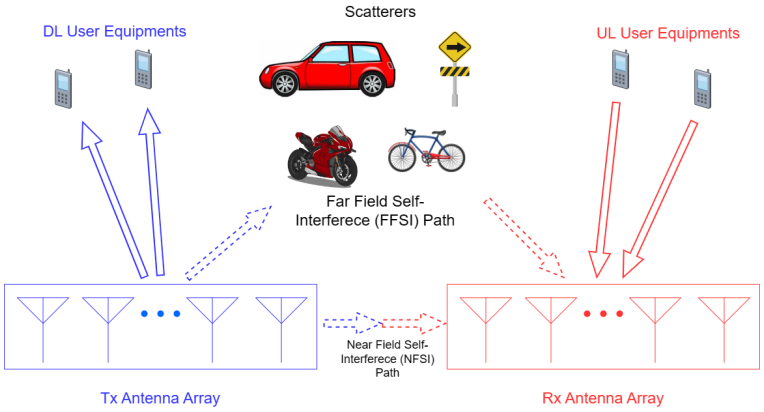


Figure 1.3: Illustration of a full-duplex communication scenario.

Second, the residual SI can corrupt the digitized UL signal (i.e., the signal of interest (SoI)) even if ADC saturation is prevented. Two mitigation techniques are used:

- *Digital SI Cancellation (DSIC):* It reconstructs and subtracts SI in the digital domain.
- *Spatial Suppression:* Applicable when antenna arrays are available, it uses beamforming to place nulls in the direction of SI.

For DSIC, NFSI-related parameters are typically static and derivable from the transceiver configuration. In contrast, mitigating FFSI requires knowledge of the environment, especially the delay and angle characteristics of scatterers. Also, the FFSI channel becomes time-varying due to the free motion in the environment, which should be taken into account in the system design.

Furthermore, nonlinear effects introduced by the power amplifier (PA) and I/Q imbalance complicate SI regeneration. Nonlinear modeling becomes crucial for strong SI components, especially NFSI. However, as shown in Chapter 2, the effect of nonlinearity is minimal for weaker FFSI and can often be neglected.

Finally, SoI reception is the main goal of the system. Since SI contaminates the SoI, SI mitigation is necessary. However, the presence of SoI has adverse effects on SI mitigation too. In *DSIC*, the SoI leaks into the SI channel estimate, deteriorating the mitigation performance. On the other hand, allocation of extra resources for SI channel estimation (e.g., a time interval without communication) compromises the

motivation towards FD communication. Therefore, the coexistence of SoI and SI is a challenge for DSIC, studied in Chapter 3. In *spatial suppression*, a similar challenge is the trade-off between beamforming for SoI reception and null steering for SI suppression. A transceiver attempting to receive UL signals from UEs must beamform toward them, but SI suppression requires placing spatial nulls toward the SI directions. In addition, multi-user interference should be handled in the beamformer design. These objectives can conflict, particularly in the presence of time-varying FFSI. Also, required information, angles of users and scatterers, should be estimated and tracked. Related to these challenges, Chapter 4 provides a comparison of spatial suppression and DSIC, and Chapter 5 studies angular estimation and tracking.

1.1.3 Literature Review

Elimination of the SI signal is of vital importance to benefit from FD systems effectively since the intended signal is much weaker ($\sim 100\ dB$ [10]) compared to SI [1]. Unfortunately, SI cancellation (SIC) is not a straightforward task. On the contrary, it involves several challenges, which require significant attention [12]. Hardware impairments, such as power amplifier (PA) nonlinearities, in-phase/quadrature (I/Q) imbalance, low-resolution digital-to-analog converter (DAC), and ADC nonlinearities induce significant nonlinear SI distortion [12, 13]. Precise modeling is crucial for achieving sufficient signal of interest (SoI) detection performance. In addition to hardware impairments, the time-varying (TV) environment, yielding a TV SI channel, further degrades the performance of the FD system [4, 14–16]. Hence, tracking channel variation is necessary to achieve reliable SoI detection.

Since SI is a strong interference, passive and active methods are applied in sequence. Regarding passive measures, shared-antenna systems exploit suppression via circulator circuitry [7] to facilitate the Tx and Rx tasks at the same antenna. Circulator prevents SI leakage to the receiver chain by electromagnetic isolation, and the active radio frequency (RF) SIC stage [17, 18]. The systems with separate Tx and Rx antennas, such as multi-panel multiple-input multiple-output (MIMO) systems, employ techniques such as antenna separation, directional isolation, cross-polarization, absorptive shielding [19, 20], and even movable antennas [21].

Following passive suppression, different techniques in both analog and digital domains are incorporated to reduce the SI power to an acceptable level [19, 22, 23]. Especially for shared antennas, limited passive SI suppression causes a huge difference in SI and SoI powers, extending the problems to be solved. For instance, in [24], design requirements and some fundamental analog and digital SIC techniques are presented. The objective of the analog SIC (ASIC) stage is to prevent the low-noise amplifier (LNA) and ADC from operating near saturation [25–27]. In active ASIC, a replica of the SI signal is generated in the digital domain and subtracted from the received signal in the analog domain by using an auxiliary transmit chain [17]. Due to hardware impairments, a linear analog canceller is not sufficient to reduce the SI power below the saturation region of the LNA. Hence, the implementation of nonlinear ASIC algorithms is proposed in the literature [17, 18, 28]. Furthermore, the effects of the auxiliary transmit chain are studied in the literature. In [29], the filtered least mean squares (LMS) (FxLMS) algorithm is adopted to mitigate the effects of the auxiliary transmit channel.

However, the ASIC stage is not capable of eliminating SI completely, and the residual SI (RSI) signal remains to be canceled in the digital domain [30]. Hence, DSIC techniques have been proposed to suppress the RSI signal as close as possible to the noise floor. In the literature, RSI signal estimation is studied for SI channels impaired by transceiver nonlinearities. In [31], a widely linear adaptive algorithm is proposed to model the nonlinearity caused by I/Q imbalance. However, this model is not capable of capturing the characteristics of a nonlinear PA. Therefore, algorithms, which use higher-order nonlinear terms, are required to improve the SI suppression performance [32–35]. In [32], memory polynomial is used to model nonlinearity together with the SI channel. In [33], different sources of nonlinearity (such as PA, I/Q imbalance, and LNA nonlinearities) and SI channels are iteratively estimated. In [34, 36, 37], cascaded spline-based Hammerstein model is adopted to reduce the computational complexity. In addition, machine-learning algorithms are proposed in the context of SI channel estimation [35, 38, 39]. For instance, a neural network (NN) architecture is proposed in [38] to estimate the nonlinear SI channel, which includes linear multipath components, and an NN architecture based on the Hammerstein-Wiener model is proposed in [39].

The aforementioned DSIC algorithms assume a relatively static SI propagation environment. However, due to the mobility of scatterers and users, time variations in SI and SoI channel responses may have significant effects on FD system performance [15, 24, 40–43]. In [40], the authors investigated the impact of SoI channel variations on the FD system performance in detail. Specifically, in outdoor environments, the SI channel exhibits significant time-varying characteristics due to mobile vehicles, which alter scatterer distribution [43–45]. For instance, in [46], it is reported that moving vehicles may yield Doppler shifts up to 40 Hz. Also, it is shown in [16] that even for the relatively slowly varying SI channels, aging of the channel estimate has a tremendous impact on the performance of the FD systems, whereas SoI channel response variation with similar Doppler frequencies has negligible effects on HD systems. Furthermore, it is stated in [44] that the time-varying SI channel reduces the SI suppression capability by 20 dB. Consequently, SIC algorithms developed for static channel conditions fail to provide sufficient SI suppression for mobile scenarios. Thus, SIC algorithms that take the dynamic characteristics of the SI channel into account should be deployed to benefit from FD communications fully.

Estimation of the SI signal during FD operation is particularly challenging in the presence of TV SI channels since SoI contamination on the received signal decreases the SI estimation accuracy in mobile environments [45, 47]. Therefore, recent literature on TV SI channel estimation mainly focuses on SoI-free SI channel estimation, where the channel estimation is performed in HD mode [45]. In [48], an adaptive frequency-domain DSIC method is proposed to track the TV SI channel, where phase noise is present. Tracking TV SI channels under PA nonlinearity is also studied in the literature. For instance, in [47], SI modeling based on generalized memory polynomial (GMP) is obtained via the recursive least squares (RLS) method. This method adaptively tracks the SI channel; however, it does not consider the effect of the SoI presence while deriving the SI estimator. Moreover, the effect of using SoI-free training for SI channel estimation on the spectral efficiency is not considered in these works.

On the other hand, the effect of SoI contamination is not extensively studied in the literature. There are a few studies that consider the SI channel estimation during FD communication without interrupting FD mode and instead operate the channel

estimation in HD mode [47, 49–51]. An expectation-maximization-based algorithm is proposed in [49], where carrier frequency offset is an issue. In [47], an analysis is carried out to demonstrate the effects of SoI contamination on the convergence of the proposed algorithm, and [50] proposes a single-stage, symbol-rate Hammerstein model-based nonlinear SI channel estimation scheme. In [50], coefficients of nonlinear basis functions and linear finite impulse response filter are obtained in the frequency domain by using the RLS algorithm. This approach estimates SoI by using an aged SI channel prediction and subtracts the SoI estimate from the received signal to increase the SI channel estimation accuracy in the subsequent blocks. The nonlinear SI channel is estimated under the assumption of a first-order Markov model. However, in case of a model mismatch, model-based SI estimate might yield an RSI that surpasses the SoI, and undermine SoI detection accuracy since SI is generally much stronger than SoI. Moreover, the Hammerstein model reduces the complexity of the nonlinear SI prediction; however, it is not fully capable of capturing hardware nonlinearities with memory, as explained in [34].

For MIMO systems, NFSI spatial characteristics are challenging to model because of the near-field effects [52,53], whereas conventional radar signal models are applicable for FFSI [6,54]. Compared to the single-antenna transceivers, the analog SIC requires RF cancellers for each RF chain [25], which multiplies the impracticality due to the required intrusion into the transceiver circuitry. While separate-antenna systems are known to decrease the need for analog SIC, multi-panel MIMO systems (separate arrays for Tx and Rx) are further resilient to NFSI as the panels remain outside each other's beams. Millimeter-wave measurements show a 40 dB interference-to-noise ratio (INR) even without active mitigation efforts [53], suggesting that analog SIC can be omitted. In the digital stage, spatial suppression is generally favored over DSIC, where beamformer weights are optimized based on the spherical wave model or an estimate of the spatial NFSI channel [54–57]. However, as spatial modeling efforts, including the spherical wave model, remain inadequate [53], model-based spatial suppression seems invalid.

The literature on FFSI is weaker. For single-antenna systems, a single SI channel is considered where larger delays represent time-varying FFSI components with Rayleigh or Rician fading [16, 26, 58]. However, this joint SI channel is both nonli-

near and time-varying. For MIMO systems, NFSI and FFSI channels are modeled separately due to spatial near-field effects on NFSI [57, 59]. While [59] considers angles of scatterers for FFSI, [57] introduces a time-varying model, but scatterers are not discussed. Generally, the delay difference due to the distinct propagation distances is overlooked. Although more rigorous models are seen in monostatic ISAC studies [6], where FFSI is actually the intended signal, effects on communication are generally not the primary focus. In terms of mitigation, FFSI is generally handled with spatial suppression via beamformer optimization [56, 57, 59, 60]. Comparing spatial suppression and DSIC for FFSI mitigation, [56] discusses the effects of the number of antennas and spatial correlation, while the published work [60] discusses the effects of user/scatterer positioning.

1.2 Beamforming Towards Users

Massive MIMO is a widely studied technology that is the fundamental basis for new-generation communication systems [61]. Using large antenna arrays, massive MIMO increases the data rate, brings angular selectivity and angular focusing of the radiated power, and compensates for huge path loss in millimeter-wave communications [62]. It also enables spatial multiplexing of multiple users, removing the need for temporal or spectral division [63, 64]. Studies show that it increases the performance of full-duplex communications [65] and radar applications [66]. In addition, it is the basis for various new applications such as dual-functional radar-communication (DFRC) [67], reconfigurable intelligent surfaces (RIS) [68], and cell-free massive MIMO [69], and a myriad of 6G technologies such as high-speed train (HST) and vehicle-to-vehicle (V2V) communications [70].

As technology evolves, the number of actively communicating devices is expected to grow significantly [70]. In a time-varying scenario where multiple users are mobile, besides serving, channel estimation is a significant problem that results in the explore/exploit dilemma and a huge computational load.

The hybrid beamformer (HBF) structure also complicates the estimation procedure. HBF is a widely offered structure for large antenna arrays instead of fully digital or

fully analog beamformers for its efficiency regarding performance and physical viability [71]. It includes both analog beamformer (ABF) and digital beamformer (DBF) stages, with a reduced number of radio-frequency chains (RFCs). Since the ABF always needs to be directed to some angle(s), and angle of arrival (AoA) estimation is held in the digital stage after the ABF, AoA estimation is limited by the prior ABF beam coverage. In addition, user mobility might result in beam loss or unintended suppression due to ABF. These factors obligate either a scanning operation or a beam tracking procedure.

Computational complexity is a significant factor affecting the feasibility of a system design. In time-varying systems, varying channel parameters should be estimated frequently, and systems using these parameters in their constructions should be updated accordingly. These estimation and update operations should be repeated within the interval in which the related parameters can be assumed to remain almost the same, namely the coherence time. On the other hand, the main task that brings the most complexity is interference mitigation, which requires the inversion of matrices in the majority of techniques. Interference might be caused by the presence of multiple users and multipath components, or simply multiple signal clusters in a MIMO system.

For time-varying channels, the rate of variation imposed on channel parameters might be different [72,73]. Among the channel parameters, slowly varying and rapidly varying ones are often called slow-time (ST) and fast-time (FT) parameters, respectively. In the massive MIMO case, AoA and angular spread (AS) could be classified as ST parameters, while gains of individual micro-components inside a signal cluster could be FT parameters. Based on this classification, instantaneous channel estimation should be repeated after shorter intervals (in fast-time), while beam tracking can be repeated more rarely (in slow-time). Upon the estimates from these procedures, beamformer weights can be updated either in fast-time or slow-time depending on the method that is used, namely, instantaneous or statistical beamforming.

1.2.1 Literature Review

Angular estimation for massive MIMO systems is a widely studied topic. While exhaustive or hierarchical search is offered for initial estimation [74–78], lower-complexity beam tracking operations are offered after the initial phase to adaptively modify the estimate under variations through time.

In the vast majority of the studies, beam tracking is held after ABF consisting of phase shifters, which create discrete Fourier transform (DFT) beams, to probe angles in space. In [74, 79–82], beam tracking means the selection of the optimum set of AoAs to be probed under time or physical (number of RFCs) limitations, after which the most powerful DFT beams are selected. The use of two perturbed beams around the main AoA is also common [83–85], from which the final AoA estimate is calculated accordingly. In the aforementioned studies, beam tracking is a phase where the beamformer is varied for a search. On the contrary, AoA variation is deduced from the variations on the current ABF output in [75, 86, 87] with the help of the extended Kalman filter (EKF), and in [76] via the maximum a posteriori (MAP) estimator. [88] and [76] employ fully digital beamformers, and [89] assumes erroneous AoA estimates and tracks the beams using recursive filtering.

Beam tracking in multi-user channels is investigated in only a few studies. Further, the near-far effect, to the best of the author's knowledge, is not studied in the literature in the beam tracking context. That is the reason why the ABF output, whose side-lobe suppression is limited, is generally chosen for beam tracking. Against multi-user interference, [90] proposes a user scheduling algorithm, [88] employs orthogonal pilots, and [79] and [85] offer time or code domain multiplexing. In [81], multi-user estimation is performed either jointly by extending the MAP estimator or sequentially by successive interference cancellation. [83] also applies joint estimation after ABF with the help of a maximum likelihood (ML) estimator. Contrary to the algorithms that operate after ABF, [74] proposes measurement after a digital zero-forcing (ZF) beamformer following ABF, to suppress the multi-user interference further. However, since ZF is constructed with the channel estimates, it is an *instantaneous beamformer*, which necessitates frequent training phases.

The estimation of rapidly changing channel gain coefficients is important for angular estimation since the observed signal also depends on these complex gains, besides the angular parameters. While [91] assumes them constant, [75, 81, 83, 86, 87, 92] track the gain coefficients together with the AoA. However, the studies in [72, 73] report that the variation in the angular parameters is much slower than the variation in the complex gains. [72] defines gain coherence time and angle coherence time such that the angle coherence time is much longer. In addition, [73] analyses the coherence times of channel parameters assuming a constant velocity movement, defining channel coherence time and beam coherence time, where the beam coherence time is about 1000 times longer. Therefore, beam-tracking can be held at much slower rates than the conventional channel estimation for the gain coefficients. Indeed, the optimum period for beam tracking is studied in [84], while the estimation for the fast-varying parameters is repeated more frequently.

Training sequence transmission is needed for beam tracking in most studies. Although uplink training is proposed generally, [80–82, 85] propose downlink training at the user equipment (UE), which requires feedback to the base station (BS). In [74], different DFT beams are tried inside the data mode, accepting data rate losses during the process.

To sum up, the beam tracking literature lacks a complete design for practical situations for millimeter-wave massive MIMO communications. These practical situations are multi-user channels, near-far effect, mobility of users, and the widely promoted HBF structure. The literature includes well-established methodologies and solutions for these practical situations, but the majority of the studies leave the acquisition of varying angular parameters out of scope. In this regard, beam-tracking studies can fill this gap in the sophisticated millimeter-wave massive MIMO literature. However, beam-tracking studies generally consider either a fully analog beamformer, which is primitive for millimeter-wave massive MIMO communications, or a fully digital beamformer, for which beam-tracking is less needed because it acquires the full-dimensional channel response. In addition, high-performance beamformer structures and multidimensional effective channels are not seen in beam-tracking literature, although they are widely seen in millimeter-wave massive MIMO literature. Furthermore, the angular parameters are tracked at the same rate as the fast-varying complex

gains, although there are studies that report a huge difference between their coherence times. Finally, a separate training sequence transmission is generally considered for beam tracking in the literature, although the system actively communicates with the user and has information about the channel through fast-time training phases.

1.3 Contributions

This thesis investigates SI mitigation strategies and beamforming techniques for FD mmWave systems under practical constraints, including time variation, nonlinearities, and environmental dynamics. Briefly, the following chapters are organized as follows.

- **Chapter 3** addresses SI in shared single-antenna systems. A nonlinear, adaptive, and low-complexity digital cancellation method is proposed to mitigate NFSI and FFSI jointly. The proposed method significantly improves robustness to SoI contamination and SI channel aging [51, 58].
- **Chapter 4** investigates FFSI in mmWave MIMO systems. A hybrid mitigation approach combining DSIC and spatial suppression is introduced, supported by a scatterer-aware sensing mechanism to enable real-time environmental adaptation [60].
- **Chapter 5** focuses on angular estimation and beam tracking for mobile UEs in massive MIMO systems. A reduced-dimensional statistical beamforming framework is developed to handle multi-cluster propagation and mobility efficiently [93].

Together, these studies offer a holistic FD transceiver design that adapts to both near-field and far-field interference in dynamically changing environments. In the remainder of this section, the contributions of each chapter will be detailed.

1.3.1 Contributions of Chapter 3

This chapter focuses on nonlinear and time-varying channel effects in SI cancellation for a shared-antenna type FD transceiver. The studies toward Chapter 3 started with

the work in [51], where a turbo-DSIC scheme is introduced by using an iterative SoI detection and cancellation approach to enhance the nonlinear SI channel learning and SIC quality for a single data block. That work is improved in [58] in terms of complexity and training overhead by decoupling linear and nonlinear parts of the SI channel. In the resulting nonlinear SIC scheme in Chapter 3, only the reduced-dimensional TV linear channel is continuously tracked at each data block, while nonlinear learning is repeated very infrequently.

To the best of the author's knowledge, the effects of time variation in the SI channel on the FD system performance under SoI contamination have not been investigated comprehensively. Whereas in this study, all these effects are taken into account while proposing an efficient SIC operation. The main contributions are elaborated below.

The sensitivity of FD systems to the time variation in the SI channel is demonstrated via performance analysis in Section 3.3. Under time variation in the SI channel, it is shown that:

- Frequent model updates are needed, which results in higher training overhead and computational complexity.
- SoI contamination is detrimental for SIC performance in the online learning scheme.
- Fundamental training schemes (online & offline) fail to provide a practical operation considering the concerns such as spectral efficiency, channel aging and learning accuracy, SoI contamination, and computational complexity.

In response to the aforementioned practical concerns, a novel DSIC algorithm, namely *Turbo-DSIC*, is proposed:

- The nonlinear SI channel modeling is performed in a 2-stage structure where computationally complex nonlinear modeling is performed very rarely, and the SI channel is continuously tracked via adaptive linear modeling.
- Contrary to conventional online and offline learning schemes, *Turbo-DSIC* reduces SoI contamination through iterations using tentative SoI symbol decisions, which improves SI prediction accuracy without an additional training

overhead. Thus, a remarkable performance improvement is shown compared to state-of-the-art methods.

- Computational complexity is reduced further by realizing SI and SoI predictions at symbol rate.

- Different nonlinearity modeling techniques, based on NN, Gaussian process regression (GPR), and GMP, are adapted and compared within the proposed 2-stage DSIC algorithm.

1.3.2 Contributions of Chapter 4

The state-of-the-art for FD communications has provided analyses and solutions for some problems, especially hardware nonlinearity effects have been widely studied for single-antenna systems. However, antenna arrays and MIMO technology bring some new aspects that remain unexplored. These aspects are grouped into three main categories, which are NFSI/FFSI differences, FFSI characteristics in millimeter-wave (mmWave) channels, and the multiple available methods for SI mitigation. A literature comparison is tabulated on these categories in Table 1.1 for MIMO studies.

The components of SI, namely NFSI and FFSI, possess distinct characteristics. The NFSI is time-invariant, strong, but can be weakened by passive suppression. It stands around the zero delay, and its spatial models are shown to be inadequate. The FFSI is time-varying, weaker, and possesses a delay according to distance, with solid spatial models. These differences require distinct processing approaches. For example, the power strength directly reflects the nonlinear modeling need, as residuals due to modeling errors are proportional to the power. In addition, time variance limits the durations for both training and usage of the SI models. The joint mitigation of NFSI and FFSI components should satisfy the requirements of both within the same design, which decreases the performance when resources are constrained [58]. In this work, the proposed time-variance-based separation enables separate and simplified mitigation stages for NFSI and FFSI. For NFSI, provided solutions in the wide literature, such as nonlinear modeling, can be implemented with continuous training and time-integration for SIC, leveraging time-invariance. Consequently, the FFSI mitigation

stage can be designed accounting for special features of FFSI, without nonlinearity concerns. This approach is partially adopted in almost all monostatic ISAC [94, 95], and some FD communication studies [59], relying solely on passive suppression of NFSI.

MIMO systems enable spatial suppression for FFSI mitigation besides SIC, bringing the geometry of transceivers and scatterers into the discussion. Angle-based correlated spatial models are available for FFSI in mmWave channels [11, 59], known for their angle-delay sparsity [96–98]. Together with the temporal effect of the SI delay, DSIC and beamforming can be designed parametrically for mmWave channels. In FD literature, although complete and partial ($\sim$ in Table 1.1) parametric channel models are seen, parameter-based suppression is not common, as seen in Table 1.1. Such a design connects with the idea of construction and utilization of a scatterer map with

Table 1.1: Literature Comparison for Chapter 4

	Article	[11]	[25]	[55]	[56]	[57]	[59]	[60] (Published Work)	Chapter 4
NFSI / FFSI	Inadequate Modeling, Joint Processing		✓	✓	✓				
	Separate Models, Joint Processing	✓				✓			
	Separate Models, Separate Processing						✓	✓	✓
mmWave FFSI	Delay/Angle-Sparse Parametric Model	$\sim$	$\times$	$\times$	$\times$	$\sim$	$\sim$	✓	✓
	Parametric Suppression (Scatterer Map)						✓	✓	✓
	Parameter Sensing (Map Construction)							✓	✓
Method	Only DSIC	✓	✓						
	Only BF			✓		✓	✓		
	Comparison (DSIC/BF)				✓			✓	✓
	Joint usage (DSIC+BF)								✓

angle and delay parameters.

The published work [60] has shown that widely overlooked FFSI can also deteriorate communication performance, even when NFSI is assumed to be completely canceled. Besides designing and comparing various channel estimators for SI and signal of interest (SoI), [60] compares DSIC and spatial suppression for a single-scatterer case, providing insights into their distinct characteristics. Extending the comparison of spatial suppression and DSIC for single-scatterer FFSI in [60], Chapter 4 presents an optimized policy for the joint utilization of both methods for the case of multiple scatterers. To the best of the author's knowledge, the joint utilization of these methods is proposed for the first time in this work, as summarized in Table 1.1.

The proposed FFSI mitigation method provides a significant performance increase within a meticulously designed system-level practicality, considering NFSI mitigation, slow-time/fast-time operations, and scatterer sensing. The contributions of this chapter are as follows:

- The NFSI and FFSI mitigation procedures are separated by the novel *time-variance-based two-stage SI mitigation* scheme, preventing trade-offs and providing flexibility to employ specific designs for each,
- *DSIC* and *spatial suppression* are investigated and compared for FFSI, demonstrating their overlooked advantages and disadvantages,
- A novel technique is offered for using DSIC and spatial suppression *together* for the *multi-scatterer* case, avoiding the disadvantages of each. Each scatterer is intelligently associated with the best SI mitigation method. For this, *simple metrics* are derived for the expected communication performance in the case of each SI mitigation action, namely, Tx beamforming, Rx beamforming, DSIC, and no-action. Based on these metrics, the optimal action for *each scatterer* is selected to maximize the communication performance.
- The requirements of parameter-based processing, namely the *SI delay* for DSIC and the *scatterer angle* for spatial suppression, are also considered. The proposed detector for *emerging scatterers* provides awareness of the changes in the *scatterer map* during communication, exhibiting a *sensing for communication*.

Detection triggers immediate estimation and mitigation, reducing the overall sensing workload from exhaustive search to verification.

1.3.3 Contributions of Chapter 5

In this chapter, a novel per-cluster estimation and beam tracking scheme is proposed for millimeter-wave massive MIMO systems employing an HBF structure. The proposed system considers angular sparsity, variation rates of the channel parameters under mobility, the angular spread phenomenon, near-far effect due to the multi-user channel, computational complexity, and time scheduling of operations. The main contribution of Chapter 5 is the design of novel beam tracking methods, namely beam-aware ML (BA-ML) and statistical EKF (SEKF). While the BA-ML method is a variant of the nonlinear least squares method [99], SEKF is an EKF application whose observation is the second-order statistics, namely the effective channel covariance matrix (CCM). The proposed novel beam tracking methods

- are suitable for HBF structures thanks to their design that allows processing in reduced-dimensional subspaces, which is a natural requirement forced by HBF structures,
- can operate in multi-user/multi-cluster channels under notable near-far effect, thanks to the per-cluster estimation scheme that is enabled by the statistical beamformer design,
- operate with longer repetition intervals than the conventional channel estimation,
- leverage a batch of readily available instantaneous effective channel (IEC) estimates collected from previous conventional channel estimation phases, without assuming given or highly correlated complex gains of the channel
- remove the need for an additional training sequence transmission for beam tracking purposes.

The proposed system requires the collaborative operation of three main tasks: statistical beamforming, IEC estimation, and beam tracking. Among these, each task

determines the performance of another task in a circular fashion. That is, the performance of beam tracking is determined by the accuracy of IEC estimates, which is determined by the performance of the statistical beamformer. Further, the performance of the statistical beamformer is dependent on the beam tracker. Therefore, the aforementioned contributions regarding the proposed beam tracking methods cannot be considered separately from the design of the statistical beamformer and IEC estimator. The existing methods from the literature are well adapted for these tasks, namely beam-aware least squares (BA-LS) for IEC estimation, and reduced-dimensional generalized eigenbeamformer (RD-GEB) for statistical beamformer. With the adaptation of these stages, a novel overall system design is obtained that can operate self-driven in mobile multi-user millimeter-wave massive MIMO communications.

The statistical beamformer RD-GEB is the adaptation of the generalized eigenbeamformer (GEB) from the prior work [100] for HBFs. Statistical beamformers do not depend on instantaneous channel coefficients, and they provide a wider selection and null areas in the angular domain, based on the channel model that includes AS. Therefore, the validity of the beamformer lasts longer, which enables the definition of a separate coherent processing interval (CPI) for beam tracking, namely the slow-time CPI (ST-CPI), which is longer than the CPI for the instantaneous channel coefficients, namely the fast-time CPI (FT-CPI). The indifference of the statistical beamformer to instantaneous variations of the channel is handled by a simple intra-cluster spatial channel matched filter (ICS-CMF), which is matched to the IEC. The proposed statistical beamforming method RD-GEB mitigates the multipath and multi-user interference better than widely proposed DFT beams, and creates interference-free subspaces for all clusters even under strong near-far effects. Consequently, multi-user channel estimation and beam tracking are performed *per-cluster* via angular division in these subspaces, without spending resources via user scheduling algorithms, orthogonal pilot transmissions, or time division/synchronization between users.

The *per-cluster* estimation scheme is a result of the proposed overall system of the aforementioned three tasks, statistical beamforming, IEC estimation, and beam tracking. It yields additional benefits for IEC estimation besides those listed above for beam tracking. The first benefit is a high multi-user estimation performance thanks to the reduced-dimensional interference-free subspaces, which also enables the us-

age of simpler estimators designed for single-user cases. In addition, the per-cluster estimation is suitable for frequency-selective channels, caused by multiple clusters that have different delays. Also, it removes the need for synchronization between the users. Finally, it provides a reduced delay spread for IEC estimation compared with joint estimation.

The computational complexities of per-cluster estimation operations are very low because interference mitigation is handled previously by the statistical beamformer, and the dimension is further reduced. Also, beamformer construction and beam tracking are performed rarely in slow-time, but the IEC estimator operates frequently with RD inputs in fast-time. Since the dominant complexity of the interference mitigation is undertaken by the ST beamformer, and the FT estimator is of low complexity, the time-averaged computational complexity also reduces.

1.4 The Outline of the Thesis

The remainder of the thesis is as follows. Chapter 2 introduces technical preliminaries. Chapter 3 examines SI cancellation in single-antenna systems, focusing on the effects of nonlinearities due to hardware impairments and time variation in SI channels. Chapter 4 studies SI mitigation in mmWave MIMO systems, where the introduced spatial domain characteristics, in addition to the previously studied time domain ones, enable a scatterer-based modeling and multiple mitigation methods. Chapter 5 addresses the acquisition of required angular information for beamforming in FD mmWave MIMO systems. Finally, Chapter 6 concludes the thesis.

1.5 Notation

Throughout the thesis, small or capital plain characters (x, X) denote scalar numbers, small bold characters ($\boldsymbol{x}$) denote (generally column) vectors, and capital bold characters ($\boldsymbol{X}$) denote matrices. For an arbitrary matrix $\boldsymbol{X}$, $(\boldsymbol{X})_{r,c}$ gives the entry in r^{th} row and c^{th} column, $(\boldsymbol{X})_{r,:}$ is the r^{th} row, $(\boldsymbol{X})_{:,c}$ is the c^{th} column. In addition, $\boldsymbol{X}^T$ and $\boldsymbol{X}^H$ are the transpose and the conjugate transpose of the matrix $\boldsymbol{X}$. Finally, $\boldsymbol{I}_N$ is the

identity matrix of size N, and $\mathrm{tr}(\boldsymbol{X})$ denotes the trace operation.

Note that each chapter has its own definitions of variables. Therefore, the same entity might be denoted differently across the chapters. The notation is clearly defined in the introductory parts of the chapters.

CHAPTER 2

PRELIMINARIES

2.1 Channel Model

2.1.1 Single-Ray Model

In a simple communication scenario with one transmitter and one receiver, consider that the transmitter sends information signals $a(t)$ and $b(t)$. Using quadrature amplitude modulation, which has a generic expression that covers more advanced modulations too, the transmitted signal $s(t)$ is written as

$$s(t) = a(t)\cos(2\pi f_c t) - b(t)\sin(2\pi f_c t), \tag{2.1}$$

where f_c is the carrier frequency. The signal arrives at the receiver with a propagation delay t_p and a scaling factor $\gamma \in \mathbb{R}$, which models path loss. The received signal $r(t) \triangleq \gamma s(t - t_p)$ is written as

$$r(t) = \gamma\left[a(t - t_p)\cos(2\pi f_c(t - t_p)) - b(t - t_p)\sin(2\pi f_c(t - t_p))\right]. \tag{2.2}$$

Then, a coherent demodulator implements operations in two parallel branches, producing two outputs. These operations start with multiplication by $2\cos(2\pi f_c t)$ and $-2\sin(2\pi f_c t)$, and then continue with low-pass filtering, denoted by $LP\{\cdot\}$. Let us express these two outputs as real and imaginary parts of a single complex signal $y(t)$. Then, the output is expressed as

$$Re\{y(t)\} \triangleq LP\{+2\cos(2\pi f_c t)r(t)\}, \tag{2.3}$$

$$Im\{y(t)\} \triangleq LP\{-2\sin(2\pi f_c t)r(t)\}. \tag{2.4}$$

Using trigonometric identities, they are computed as

$$Re\{y(t)\} = \gamma[+a(t-t_p)\cos(2\pi f_c t_p) + b(t-t_p)\sin(2\pi f_c t_p)], \tag{2.5}$$

$$Im\{y(t)\} = \gamma[-a(t-t_p)\sin(2\pi f_c t_p) + b(t-t_p)\cos(2\pi f_c t_p)], \tag{2.6}$$

where the components at frequency $2f_c$ are assumed to be removed completely by the low-pass filters. Then, it is seen that the signal $y(t)$ can also be expressed as

$$y(t) = \gamma[\cos(2\pi f_c t_p) - j\sin(2\pi f_c t_p)][a(t-t_p) + jb(t-t_p)]. \tag{2.7}$$

As seen, the received signal is factorized into information-related and carrier-related parts. According to this, the received signal is expressed more compactly as

$$y(t) = hx(t-t_p), \tag{2.8}$$

$$x(t) \triangleq a(t) + jb(t), \tag{2.9}$$

$$h \triangleq \gamma e^{-j2\pi f_c t_p}, \tag{2.10}$$

where $x(t)$ is the complex information signal and h is the channel.

The electromagnetic waves propagate at the speed of light c in free space. Therefore, the propagation delay t_p is calculated by $t_p = d/c$, where d is the distance between the transmitter and receiver. In addition, the wavelength of electromagnetic waves is $\lambda = c/f_c$. Then, the channel can also be written as

$$h = \gamma e^{-j2\pi d/\lambda}, \tag{2.11}$$

which means that the phase advances 2π over each wavelength λ. The power-wise effect $|h|^2 = \gamma^2$ is called path loss, and the nature of the scalar γ will be discussed in Section 2.1.5.

2.1.2 Carrier Phase

The channel model in (2.11) shows a dependency between the distance and the channel in terms of carrier phase rotation. However, this dependency is too sensitive to be utilized in a communication design when a reference is absent, especially at higher carrier frequencies.

For example, a 30 GHz carrier frequency in millimeter-wave communication yields a wavelength of 1 cm. If the distance is 100 m, the system should be aware of the distance at the level of millimeters. That is, erroneous information of 100.005 m instead of 100 m gives a phase error of π, which is the worst possible error.

Therefore, the phase rotation is often assumed to be uniformly random in the $[0, 2\pi]$ interval when a reference is absent. However, when a reference is present, it can be leveraged. This reference might be in time, frequency, or space. For example, in beamforming, relative differences between multiple rays give useful information on the incidence angle. In addition, while the phase is initially unpredictable, it exhibits temporal correlation. Thanks to this correlation, the channel estimates can be utilized for an interval, namely the coherence time.

2.1.2.1 Doppler Effect

Variations in the distance d create the Doppler effect. If a time-dependent distance $d(t)$ is assumed, the channel in (2.11) becomes

$$h(t) = \gamma e^{-j2\pi d(t)/\lambda} \tag{2.12}$$

If the time-dependency is due to a constant movement, the effect on the channel is the fundamental form of the Doppler effect. For example, if the transmitter or receiver is moving towards the other at speed v, that is $d(t) = d_0 - vt$, the Doppler effect can be shown as

$$\begin{aligned} h(t) &= \gamma e^{-j2\pi(d_0 - vt)/\lambda} && (2.13)\\ &= \gamma e^{-j2\pi d_0/\lambda} e^{j2\pi vt/\lambda} && (2.14)\\ &= h_0\, e^{j2\pi f_D t} && (2.15) \end{aligned}$$

where h_0 is the initial channel at $t = 0$, and $f_D \triangleq v/\lambda$ is called the Doppler frequency. As seen, when the distance gets shorter, the observed frequency increases by f_D. Apart from this simple example, the movement may exhibit more complex behavior, making the distance $d(t)$ and the channel $h(t)$ practically random.

2.1.2.2 Channel Coherence Time

The movement model in the previous discussion was exhibiting a single speed v and a single Doppler frequency $f_D = v/\lambda$. If the movement is more complex, such that the speed is random between v_1 and v_2, a Doppler spread is created between v_1/λ and v_2/λ. As a quantity, the Doppler spread is said to be $|v_1 - v_2|/\lambda$. Then, the coherence time is assumed to be the reciprocal of the Doppler spread [101].

The discussions in the next chapters will assume a movement with a random speed between $+v$ and $-v$. Accordingly, the channel coherence time is defined as

$$T_c = \frac{\lambda}{2v}. \tag{2.16}$$

2.1.2.3 Correlated Time Variation Model

In the next chapters, the first-order autoregressive model (AR(1)) will often be used to describe the time variation of channels. A discretized approach assumes that the channel remains constant for durations Δ_t and slightly changes between each interval. Considering $h_k = h(k\Delta_t)$, the model can be written as

$$h_{k+1} = ch_k + \sqrt{1-c^2}\,\nu_{k+1}, \tag{2.17}$$

where $h_k, \nu_k \sim \mathcal{CN}(0, \sigma_h^2)$ and c is the correlation, which yields the relations

$$h_{k+x} = c^x h_k + \sqrt{1-c^2}\sum_{l=0}^{x-1} c^l \nu_{k+x-l}, \tag{2.18}$$

$$\mathbb{E}\{h_{k_1} h_{k_2}^*\} = c^{|k_1-k_2|}\sigma_h^2. \tag{2.19}$$

As time progresses and $|k_1 - k_2|$ gets larger, the correlation between the two channels fades. The coherence time can be defined as the time required for the correlation coefficient $\mathbb{E}\{h_{k_1} h_{k_2}^*\}/\sigma_h^2$ to drop to a small value c_0. Accordingly, the coherence time T_c is calculated as

$$T_c = \frac{\log c_0}{\log c}\Delta_t \tag{2.20}$$

This expression can be equated to the previous expression in (2.16) to find a practical value for the correlation c. Such a consideration yields

$$c = c_0^{\frac{2v\Delta_t}{\lambda}} \tag{2.21}$$

2.1.3 Information Signal

The information signal $x(t)$ is defined in (2.9). In the received signal, it is affected by the propagation delay t_p, as shown in (2.11). In this section, the information signal will be described, and the effects of the channel will be discussed.

2.1.3.1 Modulation Types

The form of the information signal $x(t)$ is dependent on the modulation type. Two main categories of modulation types are single-carrier (SC) and orthogonal frequency-division multiplexing (OFDM). For these, the information signal $x(t)$ can be written as

$$x_{SC}(t) = \sum_{n=0}^{N-1} s_n p_{SC}(t - nT) \tag{2.22}$$

$$x_{OFDM}(t) = \sum_{n=0}^{N-1} s_n e^{j2\pi n\Delta_f t} p_{OFDM}(t) \tag{2.23}$$

where $p_{SC}(t)$ and $p_{OFDM}(t)$ are the pulses, and s_n are information symbols. These pulses should obey the Nyquist-1 criterion, satisfying the orthogonality in time or frequency. This criterion prevents interference between information s_n, which is called inter-symbol interference (ISI) in SC and inter-carrier interference (ICI) in OFDM. The pulses are desired to be time and band-limited. These limits are inversely related. That is, if the pulse duration is T and the bandwidth is W, the bandwidth-time product tends to be unity. Within the family of suitable pulses, the rectangular pulse will be adopted for simplicity, although the result will be valid for all pulses. In this case, the pulses are written as

$$p_{SC}(t) = \frac{1}{\sqrt{T_s}}\text{rect}(\frac{t}{T_s}) \tag{2.24}$$

$$p_{OFDM}(t) = \frac{1}{\sqrt{NT_s}}\text{rect}(\frac{t}{NT_s}) \tag{2.25}$$

which have unit energy and equalize the overall time for $x_{SC}(t)$ and $x_{OFDM}(t)$ at NT_s. Their frequency domain correspondents are

$$P_{SC}(f) = \sqrt{T_s}\,\text{sinc}(T_s f) \tag{2.26}$$

$$P_{OFDM}(f) = \sqrt{NT_s}\text{sinc}(NT_s f) \tag{2.27}$$

Note that $P_{OFDM}(f) = 0$ for $f = \pm\frac{1}{NT_s}, \pm\frac{2}{NT_s}, \ldots$, forcing the minimum subcarrier spacing to be $\Delta_f = \frac{1}{NT_s}$ to prevent ICI. Since there are N subcarriers, and the half-power bandwidth is approximately $\frac{1}{NT_s}$ for the OFDM pulse, the overall bandwidth for $x_{OFDM}(t)$ is $N \times \frac{1}{NT_s} = \frac{1}{T_s}$. On the other hand, the bandwidth for the $x_{SC}(t)$ is equal to that for the pulse, which is $\frac{1}{T_s}$. Therefore, the overall bandwidth for both modulations is $W = \frac{1}{T_s}$.

2.1.3.2 Sensitivity to Distance and Information Delay

The distance d, hence the propagation delay t_p, affects the information signal represented by $x(t - t_p)$ in (2.11). This effect can create a sensitivity for some operations. For example, the system uses a known information signal $x(t)$ to estimate the channel from the observations in the form of 2.8. However, the propagation delay is unknown, and the system should try a possible set of delays. Instead of this, the system can utilize the information from the previous estimations. However, this might bring a sensitivity when the delay is changed, say, from t_p to $t_p + \Delta t$.

The decorrelation will be useful to assess the sensitivity to information delay for the discussions in the next chapters. They are expressed as

$$c_{SC}(\Delta_t) \triangleq \mathbb{E}\{\int_{-\infty}^{+\infty} x_{SC}(t - t_p)x_{SC}^*(t - (t_p + \Delta t))dt\} \tag{2.28}$$

$$= \mathbb{E}\{\int_{-\infty}^{+\infty} x_{SC}(t)x_{SC}^*(t - \Delta t)dt\} \tag{2.29}$$

$$= N\int_{-\infty}^{+\infty} p_{SC}(t)p_{SC}^*(t - \Delta t)dt \tag{2.30}$$

$$c_{OFDM}(\Delta_t) \triangleq \mathbb{E}\{\int_{-\infty}^{+\infty} x_{OFDM}(t - t_p)x_{OFDM}^*(t - (t_p + \Delta t))dt\} \tag{2.31}$$

$$= \mathbb{E}\{\int_{-\infty}^{+\infty} x_{OFDM}(t)x_{OFDM}^*(t - \Delta t)dt\} \tag{2.32}$$

$$= e^{j\pi(N-1)\Delta_f\Delta_t}\frac{\sin(\pi N\Delta_f\Delta_t)}{\sin(\pi\Delta_f\Delta_t)}\int_{-\infty}^{+\infty} p_{OFDM}(t)p_{OFDM}^*(t - \Delta t)dt \tag{2.33}$$

where the last expressions are derived in Appendix A. According to the definitions of pulses in the previous section, both correlations are N when $\Delta_t = 0$, and they

both drop to zero[1] first time when $\Delta t = \pm\frac{1}{W}$. Considering this continuous descent, a useful correlation between $\Delta t = +\frac{1}{2W}$ and $\Delta t = -\frac{1}{2W}$ can be assumed. Based on the relation $t_p = d/c$, this corresponds to a useful correlation with a maximum of

$$\Delta d = \pm\frac{c}{2W} \tag{2.34}$$

change in the distance. That is, the correlation holds within a disk centered at the transceiver, between distances (radii) $d = d_0 + \frac{c}{2W}$ and $d = d_0 - \frac{c}{2W}$, where d_0 is the initial distance. With 10 and 100 MHz bandwidth, $\frac{c}{2W}$ is 15 and 1.5 meters, respectively.

2.1.4 Spatial Dimension

The spatial characteristics of the channel can be observed when antenna arrays are used. Antenna arrays include multiple antennas, and these antennas can be located in various ways. The array in which antenna elements are uniformly placed on a line is called a uniform linear array (ULA). The spacing d_a between the antennas can vary, but half-wavelength spacing ($d_a = \lambda/2$) is adopted in this thesis. In Fig. 2.1, the relative differences Δ_d in the propagation distances are shown. It is related to the incidence angle θ as $\Delta_d = d_a \sin(\theta) = (\lambda/2)\sin(\theta)$. Denoting the received signals as $y_k(t)$ for the k^{th} antenna ($k \in \{0, \ldots, N_a - 1\}$), they can be written as

$$y_k(t) = \gamma e^{-j2\pi d_k/\lambda} x(t - t_{p,k}) \tag{2.35}$$

where $t_{p,k} = d_k/c$ and d_k are

$$d_k = d_0 - k\Delta_d \tag{2.36}$$

$$= d_0 - k\frac{\lambda}{2}\sin(\theta) \tag{2.37}$$

$$t_{p,k} = t_{p,0} - k\frac{1}{2f_c}\sin(\theta) \tag{2.38}$$

Substituting these to the expression of $y_k(t)$ yields

$$y_k(t) = \gamma e^{-j2\pi d_0/\lambda} e^{-j2\pi(-k\frac{\lambda}{2}\sin(\theta))/\lambda}\, x(t - t_{p,0} + k\frac{1}{2f_c}\sin(\theta)) \tag{2.39}$$

$$y_k(t) = \gamma e^{-j2\pi d_0/\lambda} e^{jk\pi\sin(\theta)} x(t - t_{p,0} + k\frac{1}{2f_c}\sin(\theta)) \tag{2.40}$$

[1] The zero crossing is due to the integral of the pulse for SC, and it is due to the $\sin(\pi N\Delta_f\Delta_t)$ term for OFDM. Note that $N\Delta_f = W$ for OFDM.

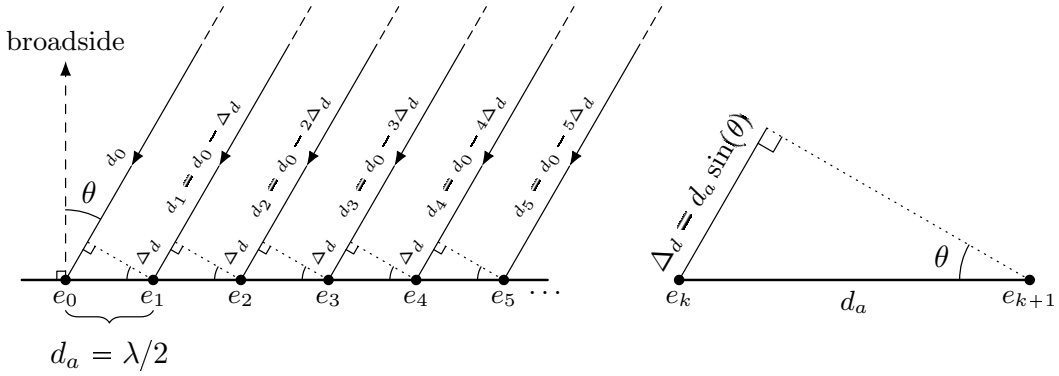


Figure 2.1: Illustration of incident waves on a ULA, incidence angle, and relative differences in the propagation distance.

In Section 2.1.3.2, it is established that the information signal is affected by the delay variations on the order of $1/W$. However, the delay difference between the received signals is on the order of $1/f_c$. Assuming the carrier frequency is at least a hundred times larger than the frequency, the delay differences in the information signals are neglected. This consideration is widely known as the narrowband assumption. Then, the expression is reduced to

$$y_k(t) = h_0 e^{jk\pi \sin(\theta)} x(t - t_{p,0}) \tag{2.41}$$

where h_0 is the channel of antenna $k = 0$. An equivalent expression would be in the vector form. It can be written as

$$\boldsymbol{y}(t) = h_0 \boldsymbol{a}(\theta) x(t - t_{p,0}) \tag{2.42}$$

$$\boldsymbol{y}(t) \triangleq \begin{bmatrix} y_0(t) & y_1(t) & y_2(t) & \dots & y_{N_a-1}(t) \end{bmatrix}^T \tag{2.43}$$

$$\boldsymbol{a}(\theta) \triangleq \begin{bmatrix} 1 & e^{j\pi \sin(\theta)} & e^{j2\pi \sin(\theta)} & \dots & e^{j(N_a-1)\pi \sin(\theta)} \end{bmatrix}^T \tag{2.44}$$

2.1.4.1 Angular Coherence

To estimate the channel effect and reach the information signal $x(t)$ from the observation $\boldsymbol{y}(t)$, the system estimates the channel effect. During this estimation, a known information signal $x(t)$ is used. The parameters of the channel are h_0, $\boldsymbol{a}(\theta)$, and t_p according to (2.42). The previous sections established that the h_0 is unpredictable, but a coarse knowledge of t_p can be utilized, considering its effect on the information signal. Then, assuming known $x(t - t_p)$, the system should estimate the vector

$h_0\boldsymbol{a}(\theta)$. That is, N_a unknowns should be estimated. Let us assume that the system also knows the angle θ. This knowledge reduces the number of unknowns from N_a to one. However, as a disadvantage, a possible error in the knowledge of θ can cause performance degradation. This error might be due to an erroneous estimation or a variation in the actual angle.

To assess the sensitivity, consider that there is a Δ_θ difference between the known and actual channels. In this case, the correlation of the observation and knowledge should be high to estimate the unknowns. The correlation can be written as

$$c(\Delta_\theta) \triangleq \mathbb{E}\left\{\int_{-\infty}^{+\infty} \left[\boldsymbol{a}(\theta+\Delta_\theta)x(t-t_{p,0})\right]^H \boldsymbol{y}(t)dt\right\} \tag{2.45}$$

$$= \mathbb{E}\left\{h_0\boldsymbol{a}^H(\theta+\Delta_\theta)\boldsymbol{a}(\theta)\int_{-\infty}^{+\infty} x(t-t_{p,0})x^*(t-t_{p,0})dt\right\} \tag{2.46}$$

$$= h_0\boldsymbol{a}^H(\theta+\Delta_\theta)\boldsymbol{a}(\theta)\mathbb{E}\left\{\int_{-\infty}^{+\infty} x(t-t_{p,0})x^*(t-t_{p,0})dt\right\} \tag{2.47}$$

$$\boldsymbol{a}^H(\theta+\Delta_\theta)\boldsymbol{a}(\theta) = \sum_{k=0}^{N_a-1} e^{-jk\pi\sin(\theta+\Delta_\theta)}e^{jk\pi\sin(\theta)} \tag{2.48}$$

$$= \sum_{k=0}^{N_a-1} e^{-jk\pi(\sin(\theta+\Delta_\theta)-\sin(\theta))} \tag{2.49}$$

$$= \sum_{k=0}^{N_a-1} e^{-jk\pi\phi}, \tag{2.50}$$

$$= \frac{1-e^{-j\pi N_a\phi}}{1-e^{-j\pi\phi}} \tag{2.51}$$

$$= e^{-j\frac{N_a}{2}\pi\phi}\frac{\sin(\pi\frac{N_a}{2}\phi)}{\sin(\pi\frac{1}{2}\phi)} \tag{2.52}$$

where $\phi \triangleq \sin(\theta+\Delta_\theta)-\sin(\theta)$. This expression yields N_a for $\phi = 0$, or equivalently $\Delta_\theta = 0$. The correlation drops to zero for the first time at $\phi = \pm\frac{2}{N_a}$, due to $\sin(\pi\frac{N_a}{2}\phi)$ term. To reach the corresponding angle for $\phi = \pm\frac{2}{N_a}$, ϕ can be written approximately as

$$\phi = \sin(\theta+\Delta_\theta) - \sin(\theta), \tag{2.53}$$

$$= [\sin(\theta) + \Delta_\theta\cos(\theta) - \frac{1}{2}\Delta_\theta^2\sin(\theta) + \dots] - \sin(\theta), \tag{2.54}$$

$$\cong [\sin(\theta) + \Delta_\theta\cos(\theta)] - \sin(\theta), \tag{2.55}$$

$$\cong \Delta_\theta\cos(\theta), \tag{2.56}$$

where Taylor expansion is utilized and higher order components are neglected, assuming a small Δ_θ. According to this expression, the Δ_θ corresponding to $\phi = \frac{2}{N_a}$ is

$$\Delta_\theta = \frac{2}{N_a \cos(\theta)}, \tag{2.57}$$

in radians. This value determines the angular coherence region, sweeping an angle of Δ_θ from $\theta - \frac{\Delta_\theta}{2}$ to $\theta + \frac{\Delta_\theta}{2}$. As seen, the sensitivity, or selectivity, increases with larger N_a and at broadside, where $\cos(\theta)$ is maximum. The mentioned region is 12.6 and 3.1 meters wide (tangential) for $N_a = 8$ and $N_a = 32$ antennas, at a 50 meters radial distance and broadside ($\theta = 0$).

2.1.5 Power

The power of a received signal coming from a direct path is determined by the Friis transmission equation. It is written as [102]

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 \tag{2.58}$$

where P_R and P_T are the received and transmitted powers, G_R and G_T are the gains of the receive and transmit antennas. In addition, λ is the wavelength, and d is the distance between the Tx and Rx antennas. Assuming isotropic antennas with unity gains, the path loss is

$$P_R/P_T = \left(\frac{\lambda}{4\pi d}\right)^2. \tag{2.59}$$

On the other hand, the path loss according to the channel model in (2.11) is $|h|^2 = \gamma^2$. In this case, the scalar γ is found to be $\lambda/(4\pi d)$, and the channel expression can be rewritten as

$$h = \frac{\lambda}{4\pi d}\, e^{-j2\pi d/\lambda}. \tag{2.60}$$

Friis' transmission equation governs the power-wise effects for direct channels between a transmitter and receiver. In full-duplex communications, the environmental echoes are also important, since they constitute far-field self-interference. The transmitted signal is scattered from an object and returns to the transceiver to be received. This signal is identical to the target echo in monostatic radars. The power of this

signal is described by the well-known radar range equation, written as

$$P_R = P_T G_T G_R \frac{\sigma \lambda^2}{(4\pi)^3 R^4} \tag{2.61}$$

where R is the range and σ is the radar cross section (RCS).

2.1.5.1 Element and Array Gain

In a half-wavelength array, individual antenna elements are very small. Hence, they are almost isotropic, and the related gain (on the Tx or Rx side) can be taken as unity. Then, the array gain can be found by considering each antenna separately. This consideration yields the same equation in the far field, but differences in the distance might be effective in the near field, especially for NFSI. Apart from that, the received and transmitted powers should be utilized with caution, since they belong to each wave component, not the combination.

For the transmit side, the power division between the elements should be considered. For the receive side, the effect of beamforming weights on the noise power should be taken into account. Then, the mechanism of the combination of N_a rays is dependent on the angle and beamforming weights. Considering these effects, the Friis transmission equation and the radar range equation can be utilized for the antenna arrays with the gain expression

$$G = \left| \boldsymbol{a}^H(\theta) \boldsymbol{c} \right|^2 \tag{2.62}$$

where $\boldsymbol{c}$ is the unit-norm beamformer vector but the steering vector $\boldsymbol{a}(\theta)$ is not unit-norm as defined in 2.44. Note that the gain in the beamformed direction is N_a, assuming that $\boldsymbol{c} = \frac{1}{\sqrt{N_a}} \boldsymbol{a}^H(\theta)$. If multiple streams are transmitted, the Tx power P_T should be divided by the number of streams D_T, or the norm of $\boldsymbol{c}$ should be $\frac{1}{\sqrt{D_T}}$.

2.1.6 Fast-Time and Slow-Time Parameters

The "fast-time" and "slow-time" parameters of the channel will be widely mentioned in the thesis. Such a separation of the channel components is studied in [73], which reports a 1000-fold difference between the channel coherence time (FT-CPI) and the

beam coherence time (ST-CPI). In this section, the rates of variation of channel parameters will be shown in an example.

The angular and temporal coherence are derived in (2.57) and (2.34). According to these, a user can move $\pm\frac{1}{N_a\cos(\theta)}$ radians tangentially, and $\pm\frac{c}{2W}$ radially, according to a base station, without a significant loss in angular and temporal coherence. During this movement, the phase and power would also change, which are discussed in Sections 2.1.5 and 2.1.2.

Consider that the transmitter moves toward a random direction, with approximately one wavelength λ displacement in Δt time for each of x and y directions. Assume that the bandwidth is 100 MHz, carrier frequency 30 GHz ($\lambda = 1$ cm), $N_a = 32$, and $d = 50$ m. This constructs a coherence region with approximately 3×3 meter squares. While the user moves a wavelength (1 cm) in Δ_t time, the total time needed for

- a phase rotation of π radians in the complex channel gain α is $\Delta_t/2$,
- a 0.1 dB change in the average incident power is $57\Delta_t$,
- the loss of the temporal coherence is $150\Delta_t$,
- the loss of the angular coherence is $155\Delta_t$.

This example shows that the complex gain is a fast-time parameter, while AoA, power, and delay are slow-time parameters. The time for the phase rotation is directly the definition of the channel coherence time given in (2.16). The difference with others verifies a second definition of coherence time for slow-time parameters. This fact can be utilized in statistical processing.

2.1.7 Multipath Channel

A single propagation path is considered in the previous sections. In practice, there are multiple paths due to reflections from the environment. In addition, each path is slightly diffused in temporal and spatial domains due to various effects, such as extended reflector surfaces. Therefore, the overall channel consists of clusters of single-

ray channels. To cover these effects, the signal model in (2.42) can be extended. Firstly, the multipath channel without diffusion is written as

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} h_k \boldsymbol{a}(\theta_k) x(t - t_{p,k}), \tag{2.63}$$

where θ_k and $t_{p,k}$ are the angle and propagation delay of the k^{th} path. In the diffused case, each of these parameters is diffused within a range about a center value. In this case, the signal can be written as

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} \sum_{l=1}^{L} h_{k,l} \boldsymbol{a}(\theta_{k,l}) x(t - t_{p,k,l}), \tag{2.64}$$

where $\theta_{k,l} \in [\theta_k - \Delta_{\theta,k}, \theta_k + \Delta_{\theta,k}]$ and $t_{p,k,l} \in [t_{p,k} - \Delta_{t,k}, t_{p,k} + \Delta_{t,k}]$. As seen, each path k constitutes a cluster with the L associated rays.[2] For an arbitrary cluster k, the diffusion in the delay might be ineffective for the information signal $x(t)$, but $h_{k,l}$ is sensitive to delay variations. As a first step, assumption of $x(t - t_{p,k,l}) \cong x(t - t_{p,k})$ yields

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} \sum_{l=1}^{L} h_{k,l} \boldsymbol{a}(\theta_{k,l}) x(t - t_{p,k}), \tag{2.65}$$

$$= \sum_{k=1}^{K} \boldsymbol{h}_k x(t - t_{p,k}), \tag{2.66}$$

$$\boldsymbol{h}_k \triangleq \sum_{l=1}^{L} h_{k,l} \boldsymbol{a}(\theta_{k,l}). \tag{2.67}$$

Since channels $h_{k,l}$ experience phase rotation for $l = \{1, \ldots, L\}$, each entry of the channel $\boldsymbol{h}_k$ tends to be a zero-mean random variable. To incorporate this, the channel can be written as $\sum_l \alpha_k \boldsymbol{a}(\theta_{k,l})$, where α_k is a zero-mean random variable. However, this definition forces a constant distribution over the angular spread at all times. In addition, $\theta_{k,l}$ is said to be in $[\theta_k - \Delta_{\theta,k}, \theta_k + \Delta_{\theta,k}]$, but their distribution is not exactly known. To reach an inclusive model, the channel is defined as

$$\boldsymbol{h}_k \triangleq \sum_{l=1}^{L} \alpha_{k,l} \boldsymbol{a}(\theta_{k,l}) \tag{2.68}$$

where L is now interpreted as a model parameter rather than a physical quantity, and $\theta_{k,l}$ are uniformly distributed inside $[\theta_k - \Delta_{\theta,k}, \theta_k + \Delta_{\theta,k}]$. Consequently, a cluster k has a common delay $t_{p,k}$, random complex gains $\alpha_{k,l}$, a center angle θ_k, and an

[2] This structure is illustrated in Fig. 5.2.

angular spread of $2\Delta_{\theta,k}$. The complex gains $\alpha_{k,l}$ are taken as a circularly symmetric complex Gaussian random variable with a variance γ^2 (i.e., $\alpha_k \sim \mathcal{CN}(0, \gamma^2)$) in this thesis, which is called a Rayleigh channel.

Since $\boldsymbol{h}_k$ is defined as a random vector, a second definition arises for the channel through the channel covariance matrix (CCM). This matrix is defined as $\boldsymbol{R}_k \triangleq \mathbb{E}\{\boldsymbol{h}_k \boldsymbol{h}_k^H\}$, and it can be written in two alternative forms:

$$\boldsymbol{R}_k = \gamma^2 \sum_{l=1}^{L} \boldsymbol{a}(\theta_{k,l}) \boldsymbol{a}^H(\theta_{k,l}), \tag{2.69}$$

$$\boldsymbol{R}_k = \gamma^2 \int_{\theta_k - \Delta_{\theta,k}}^{\theta_k + \Delta_{\theta,k}} \boldsymbol{a}(\theta) \boldsymbol{a}^H(\theta) d\theta \tag{2.70}$$

Based on the CCM, the channel can also be expressed as

$$\boldsymbol{h}_k = (\boldsymbol{R}_k)^{\frac{1}{2}} \boldsymbol{v} \tag{2.71}$$

where $\boldsymbol{v}$ is an L' dimensional random vector following $\mathcal{CN}(0, \boldsymbol{I}_{L'})$. The advantage of the second definition is that L' is a small number given that the angular spread is narrower than the angular resolution, whereas L should be large for accurate modeling.

2.1.7.1 Millimeter-Wave Effects

In mmWave channels, both the number of paths K and the number of rays in a cluster L are smaller, referred to as the angle/delay sparsity. This is due to propagation characteristics. In the mmWave bands, the electromagnetic waves experience larger path loss, and the reflectivity characteristics are changed. These reasons cause non-line-of-sight (NLOS) paths to fade out. Therefore, the simple model

$$\boldsymbol{y}(t) = \alpha \boldsymbol{a}(\theta) x(t - t_p) \tag{2.72}$$

is suitable for mmWave communications, although it is not perfectly accurate.

2.2 Processing

In this section, several processing design approaches will be introduced, building on the previous discussions of channel characteristics. These approaches provide the foundation for the design choices explored in the following chapters.

2.2.1 Parameter-Based Processing With Slow-Time Parameters

Parameter-based processing provides performance gains, but requires knowledge. It can be applied in spatial, temporal, or spectral domains. Actually, a straightforward but important example is in the spectral domain, implemented by preselection bandpass filters. It is known that the noise power is proportional to the bandwidth of this bandpass filter, and it should be designed to capture the signal with minimal bandwidth. In this process, the knowledge of the signal spectrum is used, which provides gain compared to a bandpass filter with unnecessarily wider bandwidth. This section will demonstrate this gain on a channel estimation example in the temporal domain. Consider the received signal in Eq. 2.66 under noise, such as

$$\boldsymbol{y}(t) = \sum_{k=1}^{K} \boldsymbol{h}_k x(t - t_{p,k}) + \boldsymbol{n}(t) \tag{2.73}$$

where $\boldsymbol{n}(t)$ is the additive white Gaussian noise (AWGN) following $\mathcal{CN}(\boldsymbol{0}, N_0\boldsymbol{I}_{N_a}$, where N_a is the number of antennas, and also the size of $\boldsymbol{y}(t)$, $\boldsymbol{h}(t)$, and $\boldsymbol{n}(t)$. After the required receive processing, such as matched filtering, consider that the signal is sampled at symbol-rate, and the obtained discrete-time signal is written inside $N_a \times N_c$ matrix $\boldsymbol{Y}$, where N_c is the number of collected symbols. This matrix can be expressed as

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{2.74}$$

where $N_a \times K$ matrix $\boldsymbol{H}$ contains channels in its columns for multiple delays, $K \times N_c$ matrix $\boldsymbol{X}$ contains signals corresponding to these delays in its rows, and $\boldsymbol{N}$ contains related noise samples. The LS channel estimate is

$$\hat{\boldsymbol{H}} = \boldsymbol{Y}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1} = \boldsymbol{H} + \boldsymbol{N}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1} \tag{2.75}$$

The expectation of $\boldsymbol{X}\boldsymbol{X}^H$ is $N_c\boldsymbol{I}_K$ due to the independence of symbols in time. Assuming N_c is large enough, equality is assumed according to the law of large numbers. Then, the error term $\boldsymbol{N}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1}$ is equal to $\frac{1}{N_c}\boldsymbol{N}\boldsymbol{X}^H$. Also, the variance of each term in $\boldsymbol{N}\boldsymbol{X}^H$ is N_cN_0. Therefore, each entry of the error term $\hat{\boldsymbol{H}} - \boldsymbol{H}$ is independent and follows $\mathcal{CN}(0, N_0/N_c)$. In this case, the expectation of the squared Frobenius norm of the error term is

$$\mathbb{E}\{||\hat{\boldsymbol{H}} - \boldsymbol{H}||_F\} = KN_a\frac{N_0}{N_c} \tag{2.76}$$

Note that the error power is proportional to the number of paths K, but it is not necessarily known. In this case, an approach might use all the possible delays, which is limited by N_c to satisfy invertibility. In this approach, the columns of $\boldsymbol{H}$ for non-existing delays are all zero; however, $\hat{\boldsymbol{H}}$ suffers from the error term also for these non-existing delays.

However, if the delays are known and they are utilized in the channel estimation, these unnecessary error terms are eliminated. For example, for millimeter-wave channels, the number of paths is few, and it is one in the model given in Eq. 2.72. In this case, an N_c-fold (hundreds, thousands) error difference arises between these cases ($k = 1$ with the parameter-based method and $K = N_c$ for the alternative method). This knowledge (on delays in this example) can be acquired by sensing methods. Once acquired, it can be used for a long time, according to the fast-time/slow-time parameter discussion in Section 2.1.6. Note that Sections 2.1.4.1 and 2.1.3.2 derive a coherence region for the validity of acquired angle and delay parameters, respectively. This region should be considered together with possible speeds of the transceivers and scatterers in the selection of parameter update rate to utilize this method safely and robustly, which is studied in Chapter 5 for the angular domain. Also, the abrupt changes would be problematic, such as emerging channel components. For this issue, Chapter 4 studies emerging scatterers and the detection of the SI components they produce.

2.2.2 Beamforming

For transceivers equipped with antenna arrays, beamforming is a linear transformation between information signals and those at the antennas. In reception, signals from antennas are linearly combined with optimal weights to reach a single desired signal. In transmission, a single signal is precoded to be transmitted from the antennas; that is, each antenna transmits a scaled version of the signal. Benefiting from linearity, multiple information signals can be multiplexed, extending 1-to-N_a relation to a D-to-N_a relation, where D is the number of information signals.

In this thesis, receive beamforming is discussed in detail. This is because the self-interference is formed in the receivers. In addition, mostly, the signal processing in the base station is addressed. The base station should manage both self-interference

and multi-user interference in full-duplex communications. Beamforming is a viable tool for these tasks.

The receive beamforming can be expressed as

$$z(t) = \boldsymbol{c}^H \boldsymbol{y}(t), \tag{2.77}$$

$$z(t) = \boldsymbol{c}^H [\sum_{k=1}^{K} \boldsymbol{h}_k x(t - t_{p,k}) + \boldsymbol{n}(t)], \tag{2.78}$$

$$= \sum_{k=1}^{K} (\boldsymbol{c}^H \boldsymbol{h}_k) x(t - t_{p,k}) + \boldsymbol{c}^H \boldsymbol{n}(t), \tag{2.79}$$

where $\boldsymbol{c}$ is the beamforming vector that contains beamforming weights, $z(t)$ is the obtained signal, and $\boldsymbol{n}(t) \sim \mathcal{CN}(0, N_0 \boldsymbol{I}_{N_a})$. The obtained signal is desired to be equal to (or easily convertible to) the information signal $x(t)$.

Note that the result will not be easily convertible to the $x(t)$ due to the sum of the multiple delayed versions. Therefore, a channel equalizer is needed to invert the temporal channel. In addition, the beamformer $\boldsymbol{c}$ cannot be optimized efficiently to yield high values for all the scalars $(\boldsymbol{c}^H \boldsymbol{h}_k)$ at once due to the different angular regions of paths. Instead, separate beamformers can be designed for each path, denoted by $\boldsymbol{c}_k$, yielding multiple outputs. This is written as

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_K(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1^H \\ \boldsymbol{c}_2^H \\ \vdots \\ \boldsymbol{c}_K^H \end{bmatrix} [\sum_{k=1}^{K} \boldsymbol{h}_k x(t - t_{p,k}) + \boldsymbol{n}(t)], \tag{2.80}$$

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_K(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1^H \boldsymbol{h}_1 & \boldsymbol{c}_1^H \boldsymbol{h}_2 & \dots & \boldsymbol{c}_1^H \boldsymbol{h}_K \\ \boldsymbol{c}_2^H \boldsymbol{h}_1 & \boldsymbol{c}_2^H \boldsymbol{h}_2 & \dots & \boldsymbol{c}_2^H \boldsymbol{h}_K \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{c}_K^H \boldsymbol{h}_1 & \boldsymbol{c}_K^H \boldsymbol{h}_2 & \dots & \boldsymbol{c}_K^H \boldsymbol{h}_K \end{bmatrix} \begin{bmatrix} x(t - t_{p,1}) \\ x(t - t_{p,2}) \\ \vdots \\ x(t - t_{p,K}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{c}_1^H \\ \boldsymbol{c}_2^H \\ \vdots \\ \boldsymbol{c}_K^H \end{bmatrix} \boldsymbol{n}(t), \tag{2.81}$$

To reach the information signals $x(t - t_{p,k})$ separately in $z_k(t)$, the diagonal elements in the matrix $(\boldsymbol{c}_k^H \boldsymbol{h}_k)$ should be one. In addition, off-diagonal elements, and also the noise component, should be suppressed as much as possible. This is an exemplary beamformer task where there is a signal of interest and also multi-cluster interferences to be suppressed. Those interferences can be extended to multi-user interference and self-interference.

2.2.2.1 Instantaneous and Statistical Beamforming Approaches

The weights can be optimized following different approaches. A widely adopted approach is instantaneous beamforming, which is robust but highly complex. The channel vectors $\boldsymbol{h}_k$ should be estimated and used in the beamforming vector expression. The related optimization problem can be written as

$$\boldsymbol{c} = \arg\min_{\boldsymbol{c}} \mathbb{E}\{| \sum_{\substack{k'=1 \\ k' \neq k}}^{K} \boldsymbol{c}_k^H \boldsymbol{h}_{k'} x(t - t_{p,k'}) + \boldsymbol{c}_k^H \boldsymbol{n}(t)|^2 \mid \boldsymbol{h}_{k'}\} \quad \text{subject to} \quad \boldsymbol{c}_k^H \boldsymbol{h}_k = 1 \tag{2.82}$$

$$= \arg\min_{\boldsymbol{c}} \sum_{\substack{k'=1 \\ k' \neq k}}^{K} \boldsymbol{c}_k^H \boldsymbol{h}_{k'} \boldsymbol{h}_{k'}^H \boldsymbol{c}_k + \boldsymbol{c}_k^H \boldsymbol{c}_k N_0 \quad \text{subject to} \quad \boldsymbol{c}_k^H \boldsymbol{h}_k = 1 \tag{2.83}$$

$$= \arg\min_{\boldsymbol{c}} \boldsymbol{c}_k^H \left(\sum_{\substack{k'=1 \\ k' \neq k}}^{K} \boldsymbol{h}_{k'} \boldsymbol{h}_{k'}^H + N_0 \boldsymbol{I}_{N_a} \right) \boldsymbol{c}_k \quad \text{subject to} \quad \boldsymbol{c}_k^H \boldsymbol{h}_k = 1 \tag{2.84}$$

The solution to this problem is

$$\boldsymbol{c} = \beta \left(\sum_{\substack{k'=1 \\ k' \neq k}}^{K} \boldsymbol{h}_{k'} \boldsymbol{h}_{k'}^H + N_0 \boldsymbol{I}_{N_a} \right)^{-1} \boldsymbol{h}_k \triangleq \beta \boldsymbol{\Psi}^{-1} \boldsymbol{h}_k \tag{2.85}$$

where $\beta = 1/(\boldsymbol{h}_k^H \boldsymbol{\Psi}^{-1} \boldsymbol{h}_k)$ with $\boldsymbol{\Psi} \triangleq \sum_{k'=1, k' \neq k}^{K} \boldsymbol{h}_{k'} \boldsymbol{h}_{k'}^H + N_0 \boldsymbol{I}_{N_a}$. Note that, without interference, the beamformer drops to a matched filter with $\boldsymbol{c} = \beta' \boldsymbol{h}_k$. As seen, all the channels $\boldsymbol{h}_k$ should be known and utilized for an instantaneous beamformer. This approach is more useful for the cases where angular spread is huge, and the number of unknowns in (2.68) is larger than the length of the channel vector, i.e., the number of antennas N_a. This happens at low frequencies, where the bandwidth is narrow and delayed components are not separated enough, and the electromagnetic waves reflect more.

On the other hand, a statistical beamformer attempts to utilize angular information. Let us repeat the derivation without the conditional expectation given the channel. Neglecting the angular spread for this initial discussion and assuming $\boldsymbol{h}_k = \alpha_k \boldsymbol{a}(\theta_k)$,

the problem is written as

$$\boldsymbol{c} = \arg\min_{\boldsymbol{c}} \mathbb{E}\{|\sum_{\substack{k'=1 \\ k'\neq k}}^{K} \boldsymbol{c}_k^H \boldsymbol{h}_{k'} x(t-t_{p,k'}) + \boldsymbol{c}_k^H \boldsymbol{n}(t)|^2\} \quad \text{subject to } \boldsymbol{c}_k^H \boldsymbol{a}(\theta_k) = 1, \tag{2.86}$$

$$= \arg\min_{\boldsymbol{c}} \sum_{\substack{k'=1 \\ k'\neq k}}^{K} \boldsymbol{c}_k^H \boldsymbol{R}_{k'} \boldsymbol{c}_k + \boldsymbol{c}_k^H \boldsymbol{c}_k N_0 \quad \text{subject to} \quad \boldsymbol{c}_k^H \boldsymbol{a}(\theta_k) = 1, \tag{2.87}$$

where $\boldsymbol{R}_{k'}$ are CCMs. Then, the optimal $\boldsymbol{c}$ is given by

$$\boldsymbol{c} = \beta \left(\sum_{\substack{k'=1 \\ k'\neq k}}^{K} \boldsymbol{R}_{k'} + N_0 \boldsymbol{I}_{N_a} \right)^{-1} \boldsymbol{a}(\theta_k), \tag{2.88}$$

$$= \beta \left(\sum_{\substack{k'=1 \\ k'\neq k}}^{K} \gamma_{k'}^2 \boldsymbol{a}(\theta_{k'}) \boldsymbol{a}^H(\theta_{k'}) + N_0 \boldsymbol{I}_{N_a} \right)^{-1} \boldsymbol{a}(\theta_k). \tag{2.89}$$

Note that the statistical beamformer does not need the channels $\boldsymbol{h}_k$ but the angles θ_k. As discussed in Section 2.1.6, $\boldsymbol{h}_k$ needs frequently updated estimates because it includes the fast-time parameter α, but the statistical beamformer stays valid for a longer time since it depends on the slow-time parameter θ_k.

For the case with angular spread, this form of derivation is inadequate because the subject becomes invalid. Assuming the absence of angular spread only for the intended channel yields (2.88), but it is suboptimal. For the case with angular spread, the generalized eigenbeamformer (GEB) provides a solution, which will be utilized in Chapter 5. It supports multiple beamforming vectors to address the spread of the intended channel. Also, it allows the usage of angular spread as a design parameter for robustness. The operation becomes dimension reduction, and GEB maximizes the mutual information between the full-dimensional observation and reduced-dimensional output.

2.2.3 Nonlinear Modeling

Due to the hardware impairments (e.g., power amplifier nonlinearity, I/Q imbalance), the intended transmitted signal $s(t)$ is distorted. In this section, the nonlinear effects

will be analyzed, the severity of these effects will be discussed, and the countermeasures will be introduced.

The distorted transmitted signal $s_{NL}(t)$ can be written as

$$s_{NL}(t) = f_{NL}(s(t)) = \sum_{k=0}^{\infty} a_k s^k(t) \tag{2.90}$$

where $f_{NL}(\cdot)$ is the nonlinear function describing the effects of impairments. This function can also be expressed as a polynomial, yielding

$$s_{NL}(t) = \sum_{k=0}^{\infty} a_k s^k(t) \tag{2.91}$$

where $a_k \in \mathcal{R}$ are the polynomial coefficients. To proceed further, the expression for $s(t)$ is modified below.

$$s(t) = a(t)\cos(2\pi f_c t) - b(t)\sin(2\pi f_c t), \tag{2.92}$$

$$s(t) = a(t)\left[\frac{1}{2}e^{j2\pi f_c t} + \frac{1}{2}e^{-j2\pi f_c t}\right] - b(t)\left[-\frac{j}{2}e^{j2\pi f_c t} + \frac{j}{2}e^{-j2\pi f_c t}\right], \tag{2.93}$$

$$s(t) = \frac{1}{2}[a(t) + jb(t)]e^{j2\pi f_c t} + \frac{1}{2}[a(t) - jb(t)]e^{-j2\pi f_c t}, \tag{2.94}$$

$$s(t) = \frac{1}{2}x(t)e^{j2\pi f_c t} + \frac{1}{2}x^*(t)e^{-j2\pi f_c t}. \tag{2.95}$$

Using this expression, powers of $s(t)$ can be written as

$$s^0(t) = 1 \tag{2.96}$$

$$s^1(t) = \frac{1}{2}x^*(t)e^{-j2\pi f_c t} + \frac{1}{2}x(t)e^{j2\pi f_c t}, \tag{2.97}$$

$$s^2(t) = \frac{1}{4}(x^*(t))^2 e^{-j4\pi f_c t} + \frac{1}{2}x(t)x^*(t) + \frac{1}{4}x^2(t)e^{j4\pi f_c t}, \tag{2.98}$$

$$s^3(t) = \frac{1}{8}(x^*(t))^3 e^{-j6\pi f_c t} + \frac{3}{8}x(t)(x^*(t))^2 e^{-j2\pi f_c t} + \frac{3}{8}x^2(t)x^*(t)e^{j2\pi f_c t} + \frac{1}{8}x^3(t)e^{j6\pi f_c t}. \tag{2.99}$$

These can be generalized to arbitrary k^{th} power as

$$s^k(t) = \sum_{n=-k}^{k} g_{k,n}(x(t), x^*(t))e^{jn2\pi f_c t}. \tag{2.100}$$

A further investigation yields the signals $g_{k,n}(\cdot)$ as

$$g_{k,n}(x(t), x^*(t)) = \begin{cases} 2^{-k}\binom{k}{m}x^m(t)(x^*(t))^{k-m}, & \text{if } m \text{ is integer} \\ 0, & \text{otherwise} \end{cases} \tag{2.101}$$

where $m = \frac{k+n}{2}$. Note that $g_{k,n}(\cdot) = g^*_{k,-n}(\cdot)$ since $s^k(t)$ is real. Then, $s_{NL}(t)$ can be rewritten as

$$s_{NL}(t) = \sum_{k=0}^{\infty} a_k \sum_{n=-k}^{k} g_{k,n}(x(t), x^*(t)) e^{jn2\pi f_c t} \tag{2.102}$$

As seen, the signal has components at frequencies $\{0,\ f_c,\ 2f_c,\ 3f_c,\ \dots\}$. However, these frequencies are apart from each other, and the receiver will have a band-pass filter around f_c. Therefore, with an abuse of notation, the out-of-band components are removed from the expression as

$$s_{NL}(t) = \sum_{k=0}^{\infty} a_k \left[g_{k,1}(x(t), x^*(t)) e^{j2\pi f_c t} + g^*_{k,1}(x(t), x^*(t)) e^{-j2\pi f_c t} \right] \tag{2.103}$$

$$s_{NL}(t) = \left[\sum_{k=0}^{\infty} a_k g_{k,1}(x(t), x^*(t)) \right] e^{j2\pi f_c t} + \left[\sum_{k=0}^{\infty} a_k g_{k,1}(x(t), x^*(t)) \right]^* e^{-j2\pi f_c t} \tag{2.104}$$

Then, the distorted signal can be expressed similarly to (2.95) as

$$s_{NL}(t) = \frac{1}{2} x_{NL}(t) e^{j2\pi f_c t} + \frac{1}{2} x^*_{NL}(t) e^{-j2\pi f_c t}, \tag{2.105}$$

$$x_{NL}(t) \triangleq 2 \sum_{k=0}^{\infty} a_k g_{k,1}(x(t), x^*(t)). \tag{2.106}$$

Using (2.101), $x_{NL}(t)$ can be written as

$$x_{NL}(t) = a_1 x(t) + \frac{3}{2} a_3 |x(t)|^2 x(t) + \frac{5}{2} a_5 |x(t)|^4 x(t) + \frac{35}{8} a_7 |x(t)|^6 x(t) + \dots \tag{2.107}$$

Full-duplex transceivers try to regenerate and subtract the SI, which is called SI cancellation. To do that, the nonlinear effects are learned using known $x_{NL}(t)$ and $x(t)$ pairs. While the signal $x(t)$ is inherently known, there are methods to observe $x_{NL}(t)$ or $s_{NL}(t)$. In addition, (2.107) presents a memoryless polynomial form. However, memory effects also exist due to hardware-related reasons and the smearing effect. These are covered by more comprehensive models, such as the memory polynomial (MP) and the generalized memory polynomial (GMP). In general, the nonlinear processing in the SIC procedure often boils down to estimation of model coefficients and generating a signal using the estimated coefficients. These issues will be discussed in Chapter 3. However, before Chapter 3, the remainder of this section establishes when nonlinear modeling is required and when it is not.

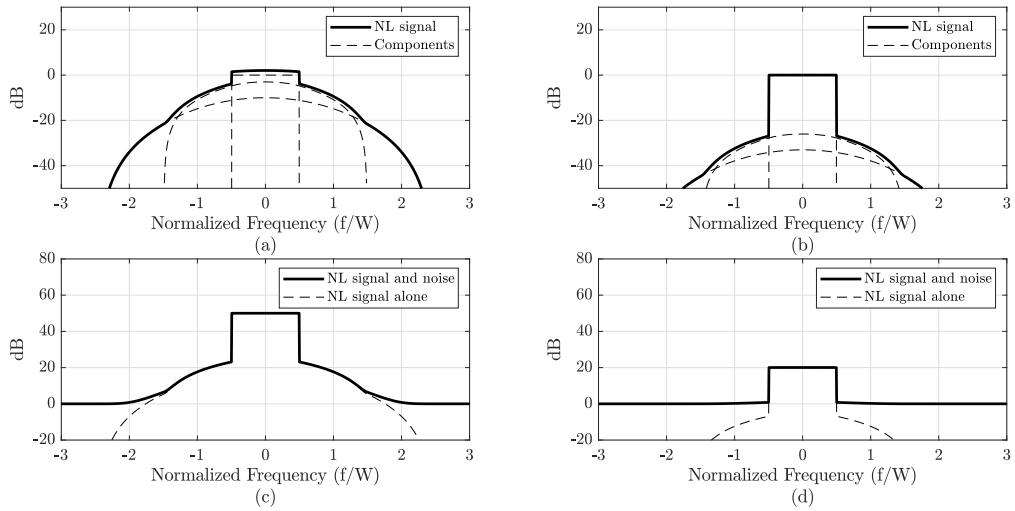


Figure 2.2: The spectrum of the nonlinear (NL) information signal. (a) Severe nonlinearity, (b) Mild nonlinearity, (c,d) The mildly nonlinear signal when received at (c) high SNR and (d) low SNR. The vertical axis is normalized to the linear component in (a) and (b), and to the noise level in (c) and (d).

Based on (2.107), exemplary spectra (i.e., power spectral densities (PSDs)) are plotted in Fig. 2.2 considering only the first three terms: $x(t)$, $|x(t)|^2x(t)$, and $|x(t)|^4x(t)$. The PSD of the linear information signal $x(t)$ is assumed to be uniform and limited between $-W/2$ and $W/2$, where W is the bandwidth. For the further components, the bandwidth gets wider and the maximum shrinks due to convolutive behavior in the frequency domain. Figure 2.2 shows severe and mild nonlinear effects in subplots (a) and (b). One should expect a mild nonlinear effect from a transceiver, which is an engineering product. The nonlinear effect should be regarded as a secondary effect, encountered in special cases. The subplots (c) and (d) show the mildly nonlinear signal in (b) under noise at different SNRs. As seen, the nonlinear effect disappears at low SNR.

The cases in subplots (c) and (d) directly relate to NFSI, FFSI, and SoI. The strong NFSI requires nonlinear modeling, but weaker FFSI and SoI might not need it.

2.2.4 Adaptive Filtering and Aging

One of the self-interference mitigation methods is regeneration and subtraction, namely DSIC. It needs an accurate SI channel estimation. For the time-varying characteristics due to the FFSI, adaptive filtering might be beneficial. In this section, some mechanisms affecting its performance will be demonstrated.

The observation for the SI channel estimation will contain the UL signal, SI, and noise. For this introductory part, the analysis will omit the presence of the UL signal. Consider the observation $\boldsymbol{z}_p$, written as

$$\boldsymbol{z}_p = \alpha_p \boldsymbol{s}_p + \boldsymbol{n}_p \tag{2.108}$$

where $\boldsymbol{s}_p$ is the transmitted signal vector, $\alpha_p \sim \mathcal{CN}(0, \sigma_\alpha^2)$ is a scalar SI channel, $\boldsymbol{n}_p$ is the noise vector. The vectors $\boldsymbol{z}_p$, $\boldsymbol{s}_p$, and $\boldsymbol{n}_p$ are of size $1 \times N_c$, where N_c is the number of subcarriers, and p indicates the OFDM symbol. A correlated time-variance is introduced on α_p using an AR(1) model as shown in Section 2.1.2.3, yielding $\mathbb{E}\{\alpha_p \alpha_k^*\} = c^{|p-k|}\sigma_\alpha^2$ with correlation c. The block form of the recursive least squares (RLS) filter is adopted as an adaptive method, which is described by

$$\hat{\alpha}_p = t_p r_p^{-1} \tag{2.109}$$

where

$$r_p = \sum_{k=1}^{p} \lambda^{p-k} \boldsymbol{s}_k \boldsymbol{s}_k^H \cong \sum_{k=1}^{p} \lambda^{p-k} N_c = \frac{1-\lambda^p}{1-\lambda} N_c \tag{2.110}$$

$$t_p = \sum_{k=1}^{p} \lambda^{p-k} \boldsymbol{z}_k \boldsymbol{s}_k^H \tag{2.111}$$

$$= \sum_{k=1}^{p} \lambda^{p-k} [\alpha_k \boldsymbol{s}_k + \boldsymbol{n}_k] \boldsymbol{s}_k^H \tag{2.112}$$

$$= \sum_{k=1}^{p} \lambda^{p-k} [N_c \alpha_k + \boldsymbol{n}_k \boldsymbol{s}_k^H] \tag{2.113}$$

$$= \frac{1-\lambda^p}{1-\lambda} N_c \alpha_p + \sum_{k=1}^{p} \lambda^{p-k} [N_c(\alpha_k - \alpha_p) + \boldsymbol{n}_k \boldsymbol{s}_k^H] \tag{2.114}$$

Note that, in the expressions above, $\boldsymbol{s}_k \boldsymbol{s}_k^H = N_c$ is assumed. Then, the RLS estimate is written as

$$\hat{\alpha}_p = \alpha_p + \frac{1-\lambda}{(1-\lambda^p)N_c} \sum_{k=1}^{p} \lambda^{p-k} [N_c(\alpha_k - \alpha_p) + \boldsymbol{n}_k \boldsymbol{s}_k^H] \tag{2.115}$$

The variance of the error term $e \triangleq \hat{\alpha}_p - \alpha_p$ gives the MSE (var(e)). It is derived below, where σ_N^2 is the variance of $\boldsymbol{n}_k \boldsymbol{s}_k^H$ and $f(k_1, k_2, p) \triangleq \mathbb{E}\{(\alpha_{k_1} - \alpha_p)(\alpha_{k_2} - \alpha_p)^*\}$. Note that $f(k_1, k_2, p) = \sigma_\alpha^2[c^{|k_1 - k_2|} + 1 - c^{p-k_1} - c^{p-k_2}]$ according to $\mathbb{E}\{\alpha_p \alpha_k^*\} = c^{|p-k|}\sigma_\alpha^2$.

$$\text{MSE} = \frac{(1-\lambda)^2}{(1-\lambda^p)^2 N_c^2} \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} \lambda^{p-k_1} \lambda^{p-k_2} \big[N_c^2 f(k_1, k_2, p) + N_c \sigma_N^2 \delta[k_1 - k_2] \big] \tag{2.116}$$

$$= \frac{(1-\lambda)^2}{(1-\lambda^p)^2} \left(\sum_{k_1=1}^{p} \sum_{k_2=1}^{p} \lambda^{p-k_1} \lambda^{p-k_2} f(k_1, k_2, p) + \frac{\sigma_N^2}{N_c} \sum_{k_1=1}^{p} \lambda^{2(p-k_1)} \right) \tag{2.117}$$

$$= \frac{(1-\lambda)^2}{(1-\lambda^p)^2} \left(\sum_{k_1=1}^{p} \sum_{k_2=1}^{p} \lambda^{p-k_1} \lambda^{p-k_2} f(k_1, k_2, p) + \frac{1-\lambda^{2p}}{1-\lambda^2} \frac{\sigma_N^2}{N_c} \right) \tag{2.118}$$

To reach a closed-form expression and easier interpretation, let p go to infinity, which yields

$$\text{MSE} = (1-\lambda)^2 \sigma_\alpha^2 \left(\frac{1 + c\lambda}{(1-\lambda^2)(1-c\lambda)} + \frac{1}{(1-\lambda)^2} - \frac{2}{(1-\lambda)(1-c\lambda)} \right) + \frac{(1-\lambda)^2}{1-\lambda^2} \frac{\sigma_N^2}{N_c} \tag{2.119}$$

$$= \frac{2\lambda^2(1-c)}{(1+\lambda)(1-c\lambda)} \sigma_\alpha^2 + \frac{1-\lambda}{1+\lambda} \frac{\sigma_N^2}{N_c} \tag{2.120}$$

where the first and second terms are the aging-based and the noise-based errors, respectively. The expression is valid for $\lambda \leqslant c < 1$. For $\lambda \leqslant c = 1$, the aging-based term is zero while the noise-based term is as it is. The scalar λ is the forgetting factor of RLS, which can be treated as the level of adaptivity. (Note that RLS drops to LS (nonadaptive estimator) when $\lambda = 0$). The noise-based error term is decreasing with λ, but the aging-based one is increasing. The obtained MSE expression is plotted against λ for different correlation (c) and INR ($\frac{\sigma_\alpha^2}{\sigma_N^2/N_c}$) cases in Fig. 2.3. To be able to provide easier comparison with LS, which yields the MSE σ_N^2/N_c, this ratio is set to unity.

In Fig. 2.3, it is seen that the optimum λ (the level of adaptivity) is generally small for high INR and it is higher for low INR cases. In fact, the optimum λ is zero (i.e., the optimum method is nonadaptive LS) for the high INR and low correlation (fast variation) case. The usage of adaptivity is beneficial (optimum λ is greater than zero) for the low INR cases, which are not significant for SIC tasks. For slow cases, the improvement depends on the selection of the forgetting factor, and the performance is very sensitive to slight deviations from the optimal value.

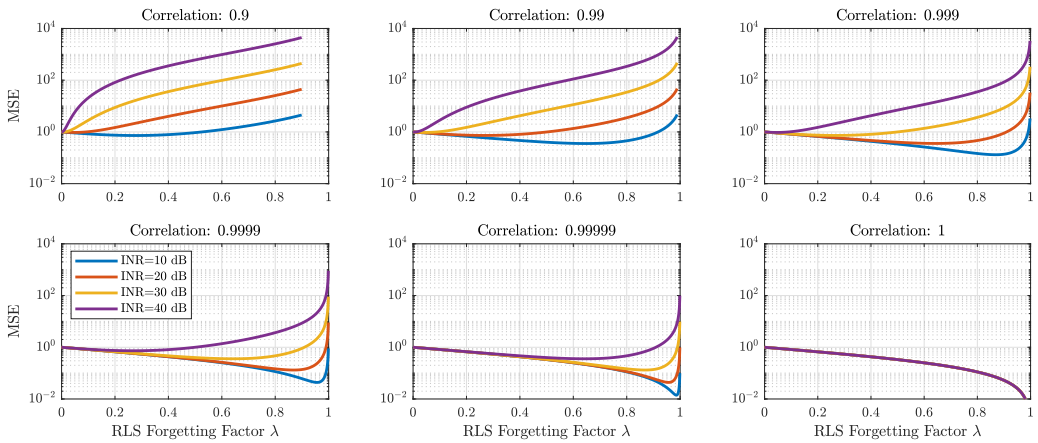


Figure 2.3: Analytic MSE versus RLS forgetting factor plots.

2.3 Power Analysis for Self-Interference

In this section, powers of SI types will be analyzed on a multipanel MIMO setting, illustrated in Fig. 2.4. Tx and Rx panels consist of half-wavelength uniform linear arrays (ULAs) with N_a antenna elements, where a hybrid beamformer structure is considered. The length of the related dimension is $D_x = \frac{N_a \lambda}{2}$, and the distance between the units is called D_{gap}. The positions of the Rx and Tx antenna elements are

$$x_m^{Rx} = +\frac{D_{gap}}{2} - \frac{\lambda}{4} + m\frac{\lambda}{2} \tag{2.121}$$

$$x_n^{Tx} = -\frac{D_{gap}}{2} + \frac{\lambda}{4} - n\frac{\lambda}{2} \tag{2.122}$$

respectively, for $m, n = 1, \ldots, N_a$.

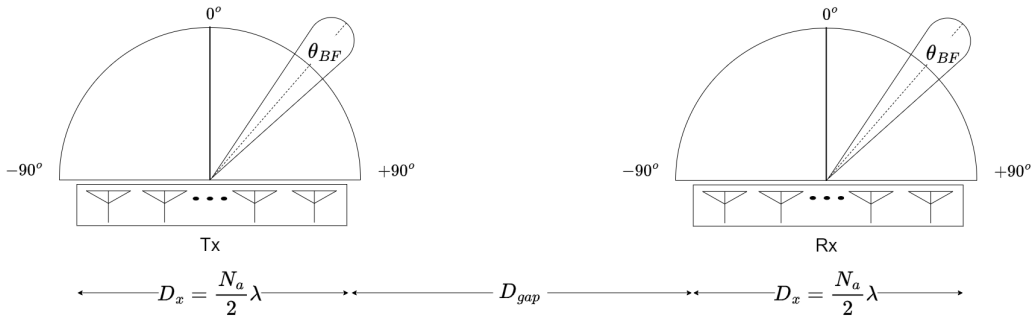


Figure 2.4: Considered setting of multipanel FD MIMO transceiver.

Table 2.1: Boundaries for Radiation Regions

	D_x	ULA ($D = D_x$)		URA ($D = \sqrt{2}D_x$)	
N_a	(f_c=30 GHz)	D_n/D_x	D_f/D_x	D_n/D_x	D_f/D_x
8	4 cm	1.24	8	2.09	16
16	8 cm	1.75	16	2.95	32
32	16 cm	2.48	32	4.17	64
64	32 cm	3.51	64	5.90	128

2.3.1 Power Equations

Friis transmission equation and radar range equation will be used in the analysis, which are introduced in Section 2.1.5. Friis' transmission equation is utilized for NFSI, adopting the spherical wave model, since it is a direct path from a transmitter antenna to a receiver antenna. The radar range equation is utilized for the FFSI, which is an echo from an object.

For the receiver side, if the hybrid beamformer is considered, the power received at any RF chain can be calculated separately using the related set of beamforming weights. The presence of multiple RF chains does not affect the calculations. For fully digital beamformers, the power observed at antenna elements can be calculated separately as described before.

For the transmitter side applying multi-stream transmission, if each stream is associated with a linearly independent set of beamformer weights illuminating distinct angular sectors, the power observed at a location is the sum of powers due to different streams. At one of the illuminated angular sectors, the related stream is dominant, while others are weak. One can approximate the observed power by focusing on the most dominant one, neglecting the others. In this case, the gain should be calculated using only the related beamforming weights. Also, if streams are allocated with equal powers, the power should be taken as P_T/D_T where D_T is the number of streams.

The NFSI is an unintended leakage from the Tx unit to the Rx unit. Therefore, the receiving end is not in the beamformed direction of the Tx array. Then, in the case of multi-stream transmission, it is unlikely to have a distinctly dominant one among

the leakages due to different sets of beamforming weights. The correct calculation is summing up the NFSI terms caused by all streams. However, assuming the beamformed directions are consecutive, the NFSI leakages can be assumed equal. In this case, the power should be taken as P_T, without the division by D_T.

2.3.2 Radiation Regions and Spherical-wave Model

Three radiation regions are defined in the literature, namely the reactive near field, the radiative near field, and the far field [53, 102]. The reactive near field starts from the antenna and ends at the distance $D_n = 0.62\sqrt{\frac{D^3}{\lambda}}$, where D is the maximum distance of two points inside the panel. The radiative near field covers the range between the distance D_n and the distance $D_f = \frac{2D^2}{\lambda}$. Farther ranges are classified as the far field. These distances are tabulated in Table 2.1 for ULA and uniform rectangular array (URA).

The propagation characteristics in the reactive near field cannot be modeled accurately. The spherical radiation model holds in the radiative near field, where a given phase of the electromagnetic wave forms a spherical surface with the transmitter at its center. In the far field, the mentioned sphere is so large that the observed portion can be assumed as a plane, which is called a plane wave. To sum up, model-based calculations are possible only for the radiative near field and the far field. Therefore, D_n/D_x values in Table 2.1 are the minimum values for D_{gap}/D_x to have a valid calculation.

In the radiative near-field, the MIMO channel can be written using the spherical radiation model [52, 103] as

$$[\boldsymbol{H}]_{mn} = \frac{\gamma}{d_{mn}} \exp\left(-j2\pi \frac{d_{mn}}{\lambda}\right), \qquad m, n = 1, \ldots, N_a \tag{2.123}$$

where $d_{mn} = |x_m^{Rx} - x_n^{Tx}|$ is the distance between the m^{th} Rx and n^{th} Tx antenna element, λ is the wavelength, and γ is a scaling constant. Then, the received power at the m^{th} Rx antenna element due to transmission from the n^{th} Tx antenna element is

$$P_{R,mn} = P_{T,n} |\, [\boldsymbol{H}]_{mn} \,|^2 = P_{T,n} \left(\frac{\gamma}{d_{mn}}\right)^2 \tag{2.124}$$

The power $P_{R,mn}$ can also be written via Friis' transmission equation as

$$P_{R,mn} = P_{T,n} \left(\frac{\lambda}{4\pi d_{mn}}\right)^2 \tag{2.125}$$

where omnidirectional antenna elements are assumed with $G_T = G_R = 1$. Then, setting

$$\gamma = \frac{\lambda}{4\pi} \tag{2.126}$$

ensures that the Friis transmission equation in (2.58) is automatically satisfied for any pair of Tx/Rx antenna elements. Then, the NFSI power after Rx beamforming can be written as

$$P_{NFSI} = P_T \left| \boldsymbol{c}_r^H \boldsymbol{H} \boldsymbol{c}_t \right|^2 \tag{2.127}$$

where $P_T \triangleq \sum_n P_{T,n}$ is the total Tx power fed to the Tx unit, which satisfies $P_T = N_a P_{T,n}$ for constant-modulus Tx beamforming weights such as DFT beams. In addition, $\boldsymbol{c}_r$ contains the Rx beamforming weights, and $\boldsymbol{c}_t$ contains the Tx beamforming weights. As a practical case, DFT beams are considered as[3]

$$\boldsymbol{c}_r = \boldsymbol{c}_t = \frac{1}{\sqrt{N_a}} \boldsymbol{a}(\theta_{BF}) \tag{2.128}$$

for both units where θ_{BF} is the steered direction.

2.3.3 Approximated NFSI Power Expression with Far-Field Model

The NFSI power is expressed in (2.127) using the spherical wave model, which is more suitable for the interested radiative near-field region. However, it does not explicitly show the effects of system parameters, such as the number of antennas, the steered direction, the gap between the units, etc. On the other hand, the plane wave model can build such connections explicitly with reasonable approximations. However, since it is more suitable for the far-field, it should be compared to the more accurate expression in (2.127) to evaluate its accuracy.

To reach an approximated NFSI power expression, Friis transmission equation in (2.58) can be used as

$$P_{NFSI} = P_T G_T G_R \left(\frac{\lambda}{4\pi D_{gap}} \right)^2 \tag{2.129}$$

[3] Note that the factor $\frac{1}{\sqrt{N_a}}$ normalizes the total Tx power for the Tx array, and the noise power for the Rx array. That is, the total Tx power stays as P_T while individual powers for elements are adjusted as P_T/N_a for any choice of N_a. In addition, the noise power is P_N both for a single Rx antenna element and after Rx beamforming; therefore, the noise power calculated for an antenna element can be used directly for the output of the Rx beamformer.

considering the setting in Fig. 2.4. Let us relate D_{gap} with $D_x = \frac{N_a\lambda}{2}$ using the ratio $\frac{D_{gap}}{D_x}$, which yields

$$P_{NFSI} = P_T G_T G_R \left(\frac{1}{2\pi \frac{D_{gap}}{D_x} N_a} \right)^2 . \tag{2.130}$$

This NFSI power expression can be improved by relating G_T and G_R to physical parameters, considering the scenario shown in Fig. 2.4. Each antenna element inside the arrays is assumed omnidirectional with unity gain. The overall array gain is constituted by beamforming. When the array beam is steered towards θ_{BF} with weights $\frac{1}{\sqrt{N_a}}\boldsymbol{a}(\theta_{BF})$, the total power observed at θ_{obs} can be written as

$$G(\theta_{BF}, \theta_{obs}) \triangleq \left| \boldsymbol{a}^H(\theta_{obs}) \left[\frac{1}{\sqrt{N_a}} \boldsymbol{a}(\theta_{BF}) \right] \right|^2 \tag{2.131}$$

$$= \begin{cases} \dfrac{1}{N_a} \dfrac{\sin^2\left(\frac{\pi}{2} N_a \phi\right)}{\sin^2\left(\frac{\pi}{2}\phi\right)}, & \phi \neq 0 \\ N_a, & \phi = 0 \end{cases} \tag{2.132}$$

where

$$\phi \triangleq \sin(\theta_{BF}) - \sin(\theta_{obs}). \tag{2.133}$$

Note that the gains in (2.130) are expressed in terms of function $G(\theta_{BF}, \theta_{obs})$ as

$$G_T = G(\theta_{BF}, \theta_{obs} = +90^\circ), \tag{2.134}$$

$$G_R = G(\theta_{BF}, \theta_{obs} = -90^\circ), \tag{2.135}$$

for the setting in Fig. 2.4, however, these expressions will be simplified further before being substituted into (2.130).

The beampattern $G(\theta_{BF}, \theta_{obs})$ is plotted in Fig. 2.5. As seen, it is a function taking a maximum value of N_a at $\theta_{obs} = \theta_{BF}$, and a minimum value of $1/N_a$ when $\phi = 1$, i.e. at $\theta_{obs} = \sin^{-1}(\sin(\theta_{BF}) - 1)$.[4] Contrary to intuition, the minimum does not occur at the farthest angle from θ_{BF}, except for the $\theta_{BF} = 0^\circ$ case with minima at +90° and -90°. In addition, the beampattern has the same value at $\theta_{obs} = +90^\circ$ and $\theta_{obs} = -90^\circ$ regardless of θ_{BF}.

[4] Note that the y-axis in Fig. 2.5 is in dB scale. The maximum and the minimum ± 15 dB correspond to $N_a = 32$ and $1/N_a = 1/32$.

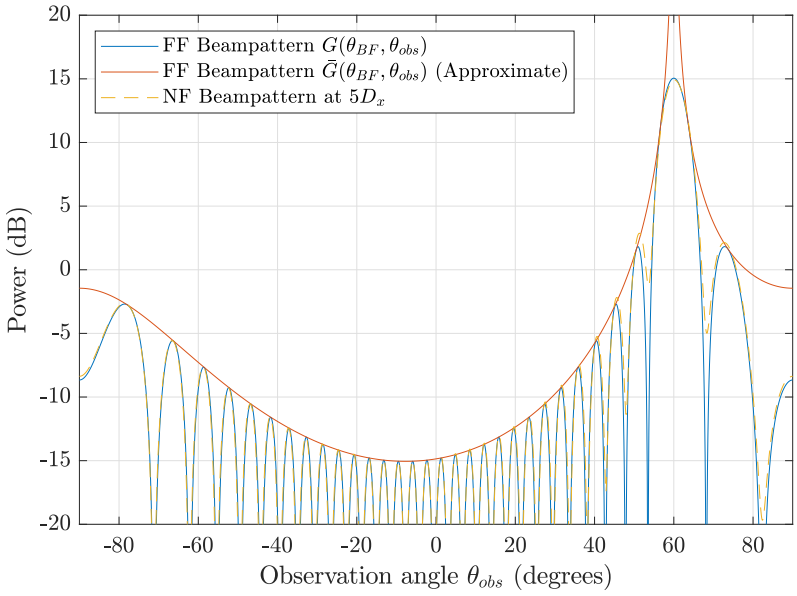


Figure 2.5: DFT beampattern. ($N_a = 32$, $\theta_{BF} = 60°$.)

Note that (2.127) was a more accurate NFSI power expression based on the near-field model, and a more useful but possibly less accurate expression is being derived now, based on the far-field model. The accuracy of the latter model is related to the dominance of the near-field effects, to be assessed in the numerical results section. However, the severity of these effects can also be assessed from the near-field beampattern plotted in Fig. 2.5 for $5D_x$ distance, which is a suitable distance for D_{gap}.[5] It is seen that the near-field effect at $5D_x$ is almost negligible. It only fills in the nulls slightly, preventing them from reaching zero.

The numerator in (2.132) governs the rapid changes in the beampattern, while the denominator controls the decreasing trend seen as the observer moves away from the beamformed direction. It is clear that the array gain is N_a at $\theta_{obs} = \theta_{BF}$, but the behavior rapidly changes far from the beamformed angle. That is, a slight change in θ_{BF} or θ_{obs} would yield a great deviation in calculations and conclusions outside the beamformed area. To avoid this sensitivity for NFSI power calculations, the general

[5] The near-field beampattern is calculated by taking Tx antenna element positions as in (2.122) and considering a single Rx antenna (observer) at $x = c + d\sin(\theta_{obs})$ and $y = d\cos(\theta_{obs})$ at xy-plane, where c is the center of the Tx unit and $d = 5D_x$. Following similar procedures explained for the NFSI channel, $1 \times N_a$ channel $\boldsymbol{H}$ is obtained. Then P_{NFSI} in (2.127) gives the beampattern with $\boldsymbol{c}_r = 1$ and $\boldsymbol{c}_t = \frac{1}{\sqrt{N_a}}\boldsymbol{a}(\theta_{BF})$.

trend is focused on by defining

$$\bar{G}(\theta_{BF}, \theta_{obs}) \triangleq \frac{1}{N_a \sin^2\left(\frac{\pi}{2}\phi\right)} \tag{2.136}$$

for the cases where $|\theta_{BF} - \theta_{obs}|$ is large. As seen in Fig. 2.5, $\bar{G}(\theta_{BF}, \theta_{obs})$ is a smooth function that follows the peaks of sidelobes of the beampattern.

Then, according to Fig. 2.4, the Tx and Rx gains for the NFSI power expression in (2.130) are

$$G_T = \bar{G}(\theta_{BF}, \theta_{obs} = +90^\circ) \tag{2.137}$$

$$G_R = \bar{G}(\theta_{BF}, \theta_{obs} = -90^\circ) \tag{2.138}$$

which yields

$$G_T = G_R = \frac{1}{N_a \cos^2\left(\frac{\pi}{2}\sin(\theta_{BF})\right)}. \tag{2.139}$$

Finally, substituting these gain expressions into (2.130), the NFSI power expression is obtained as

$$P_{NFSI} = \frac{P_T}{(2\pi)^2 \left(\frac{D_{gap}}{D_x}\right)^2 N_a^4 \cos^4\left(\frac{\pi}{2}\sin(\theta_{BF})\right)}, \tag{2.140}$$

which shows that the NFSI power is inversely proportional to the square of the gap between the units and the fourth power of the number of antennas. This approximated NFSI power expression is an alternative to the more accurate one in (2.127), to be compared in the numerical results section.

The NFSI power based on the spherical wave model in (2.127) and the approximated NFSI power expression in (2.140) are calculated for a Tx power of $P_T = 1$ Watt and $D_{gap} = 5D_x$. In addition, the noise power is calculated as

$$P_N = 10^{-20.4} \times W \quad \text{Watts} \tag{2.141}$$

$$= -204 + 10\log W \quad \text{dBW} \tag{2.142}$$

where W is the bandwidth. $W = 100$ MHz is taken, which yields -124 dBW noise power.

For FFSI, the radar range equation in (2.61) can be used. Both arrays are beamformed towards the interested direction as in Fig. 2.4, yielding $G_T = G_R = N_a$ as shown in

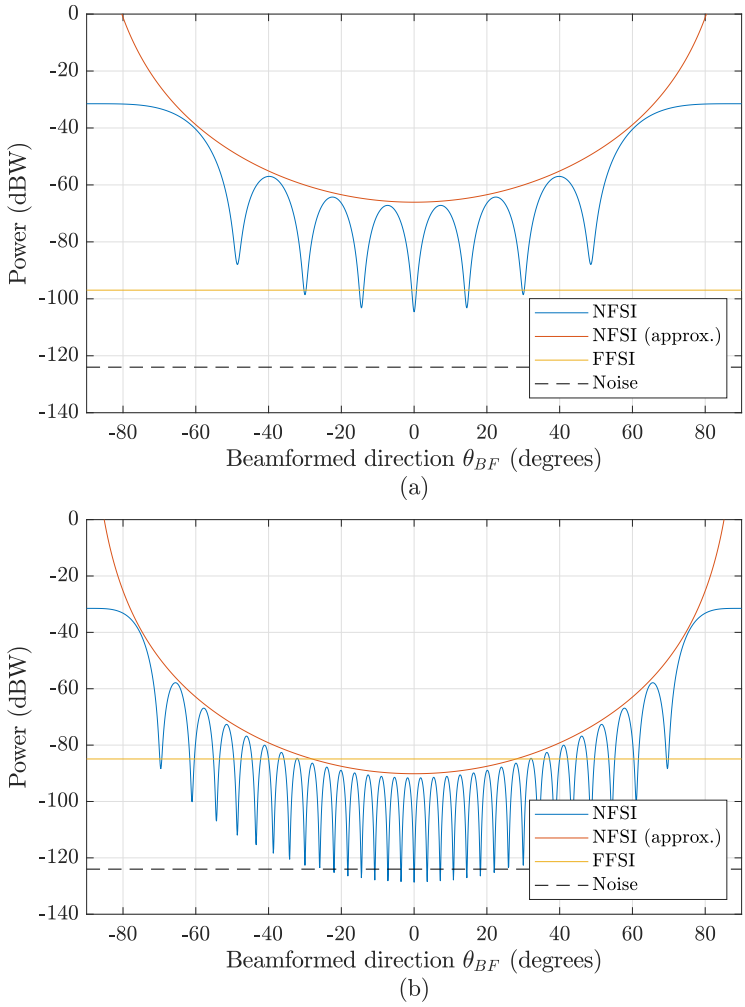


Figure 2.6: Calculated powers against the beamformed angle θ_{BF} with (a) $N_a = 8$, (b) $N_a = 32$.

Fig. 2.5. Then, the received FFSI power is

$$P_{FFSI} = \frac{P_T}{D_T} \frac{N_a^2 \sigma \lambda^2}{(4\pi)^3 R^4}. \tag{2.143}$$

For this expression, the number of streams, RCS, wavelength, and range are chosen as $D_T = 1$, $\sigma = 10$ m^2, $\lambda = 1$ cm for 30 GHz carrier frequency, and $R = 20$ m, respectively.

Using these settings, the powers are plotted in Fig. 2.6. It is seen that both SI components are at least 30 dB above the noise level, showing the severity of the effect. NFSI power dramatically changes according to the beamformed angle; it decreases as the beamformed angle gets closer to the broadside ($\theta = 0^\circ$). On the other hand, the FFSI

Table 2.2: INR Levels

	Maximum INR (NFSI) P_{NFSI}/P_N (dB)					INR (FFSI)
N_a	$\lvert\theta_{BF}\rvert < 30^\circ$	$\lvert\theta_{BF}\rvert < 45^\circ$	$\lvert\theta_{BF}\rvert < 60^\circ$	$\lvert\theta_{BF}\rvert < 75^\circ$	$\lvert\theta_{BF}\rvert < 90^\circ$	P_{FFSI}/P_N
1	92.47	92.47	92.47	92.47	92.47	8.98
2	86.47	90.59	92.11	92.48	92.50	15.00
4	**69.98**	81.84	90.52	92.38	92.51	21.02
8	**59.74**	**67.00**	83.39	91.97	92.51	27.04
16	**48.93**	**57.62**	**66.30**	90.29	92.51	33.06
32	**37.61**	**43.97**	**57.10**	82.75	92.51	39.09
64	**25.95**	**34.24**	**46.76**	**66.09**	92.51	45.11

power is constant due to the scenario description, where the scatterer is always at the beamformed direction. Also, the Tx and Rx beams are directed in the same direction, which is the worst case for the FFSI. Finally, the figure shows that the antenna arrays and beamforming alleviate the effect of NFSI, while they enhance the FFSI.

2.3.4 Analog-to-Digital Conversion and Dynamic Range

In practical systems, beamforming around endfire regions ($\theta = \pm 90^\circ$) is avoided, that is, $|\theta_{BF}|$ is limited by a maximum value, for example, 60°. Therefore, the maximum NFSI power inside the utilized angular interval should be considered as the worst case. Accordingly, the maximum INR values are tabulated for different θ_{BF} values in Table 2.2, together with the FFSI powers.

To be able to omit the complex analog SIC structure, the INR levels should be inside the dynamic range, which is $6.02 \times q + 1.76$ dB for a q-bit ADC [7]. Using a 12-bit ADC with 74 dB dynamic range, analog SIC can be omitted for the cases shown with bold numbers. The results show that as the number of antennas increases, the need for ASIC decreases. For example, the (-60°,$+60^\circ$) beam steering range can be used only with $N_a = 16$ and above without ASIC. For smaller arrays, the angular sector (-60°,$+60^\circ$) cannot be used completely and should be limited. While an array with $N_a = 4$ can use a very limited region around broadside without ASIC, $N_a = 2$ requires ASIC. A single-antenna system ($N_a = 1$) also needs the ASIC stage even

with an advantageous separate-antenna structure, implying that ASIC is inevitable for shared-antenna structures (see Section 1.1.1). These results show that the analog SIC can be omitted depending on the number of antennas and the dynamic range. It is worth noting that the Tx power P_T increases the SI level but not the noise; therefore, the dynamic range requirement also depends on the Tx power.

CHAPTER 3

AN ADAPTIVE-ITERATIVE NONLINEAR INTERFERENCE CANCELLATION IN TIME-VARYING FULL-DUPLEX CHANNELS

In this chapter, first, the sensitivity of traditional Full-Duplex (FD) communication systems to time variation in the self-interference (SI) channel is demonstrated via performance analysis. It is seen that conventional schemes are not capable of providing efficient operation regarding practical concerns such as spectral efficiency, SI channel aging and learning accuracy, and the contamination of signal-of-interest (SoI). Then, regarding the aforementioned concerns, a practical FD operation together with a novel iterative SoI-contamination-aware digital SI cancellation (DSIC) algorithm that relieves the sensitivity against time variation considerably is proposed. The proposed scheme, namely *Turbo* DSIC, is practical in terms of training and computational complexity thanks to its two-stage structure such that complex nonlinear parameter learning (NPL) is conducted very rarely in the first stage, and linear processing is performed adaptively at the symbol rate in the second stage, where SoI is iteratively estimated at each data block and eliminated from the received signal to improve SI channel estimation accuracy. Finally, the performance of *Turbo* DSIC, under hardware impairments and time-varying (TV) propagation, is evaluated via both computer simulations and actual experimental hardware setup in terms of bit error rate (BER) and achievable information rate (AIR) metrics.

3.1 Signal Model

This study considers in-band FD communication between two parties over a TV channel, performing block transmission of single-carrier (SC) modulation. Parties are

equipped with a single antenna each, and the transceiver is the shared-antenna type (see Section 1.1.1). As discussed in Sections 2.3 and 2.2.3, this type suffers from very strong NFSI and requires nonlinear modeling in SI cancellation.

The channels, namely the SI channel and SoI channel, are assumed to be constant within a transmission block, which will be called a coherent processing interval (CPI). The channels are denoted by $h_n^{SI}[k]$ and $h_n^{SoI}[k]$, respectively, where k indicates the CPI and n is the delay index.

3.1.1 SC Transmission Model

In discrete time, the signal to be transmitted in the k^{th} CPI can be expressed as

$$x_n[k] = \rho_x \sqrt{\mu} \sum_{m=0}^{N-1} a_m[k] p_{n-\mu m}, \tag{3.1}$$

where p_n is the unit-energy pulse shaping filter with upsampling factor μ, ρ_x is magnitude, and $a_n[k]$ are symbols taken from unit-power M-ary quadrature amplitude modulation (M-QAM) alphabet for $n = 0, \ldots, N-1$. Note that $x_n[k]$ is a zero-mean, cyclo-stationary signal with average power of ρ_x^2. Then, the transmitted sequence $\{x_n[k]\}$ is fed into PA $\Psi(\cdot)$[1], which outputs the nonlinearly distorted sequence $\Psi(\{x_n[k]\})$, whose sample at the n^{th} time instant is given as

$$\psi_n[k] = \Psi(\{x_n[k]\}). \tag{3.2}$$

Simultaneously, the second party transmits the desired signal $s_n[k]$ during active FD communication. It is expressed as

$$s_n[k] = \rho_s \sqrt{\mu} \sum_{m=0}^{N-1} d_m[k] p_{n-\mu m}, \tag{3.3}$$

where ρ_s is the magnitude and $d_n[k]$ are desired M-QAM symbols. PA output $\psi_n[k]$ echoes back to the transceiver through the unit-energy SI channel $h_n^{SI}[k]$. The SoI is $r_n[k] \triangleq s_n[k] \star h_n^{SoI}$, where $\star$ is the convolution operator, and h_n^{SoI} is the SoI channel, which is assumed to be time-invariant.[2] Therefore, before any processing, the total

[1] Although the main objective of a PA is amplification, $\Psi(\cdot)$ function only represents the nonlinearity effects with almost identical input/output power levels. The amplification is governed by ρ_x, together with all the effects on SI power such as path loss and passive suppression to provide easier discussion.

[2] SoI detection under a TV SoI channel is evaluated only in Fig. 3.9, which does not show a significant effect.

received signal $y_n^{total}[k]$ in the k^{th} CPI is

$$y_n^{total}[k] = r_n[k] + \psi_n[k] \star h_n^{SI}[k] + \eta_n[k], \tag{3.4}$$

where $\eta_n[k]$ is complex additive white Gaussian noise (AWGN) with $\eta_n[k] \sim \mathcal{CN}(0, \mu\sigma_\eta^2)$, which yields SI signal-to-noise ratio (SNR) as ρ_x^2/σ_η^2 and SoI SNR as ρ_s^2/σ_η^2 after the matched filter (MF) and downsampling (DS).

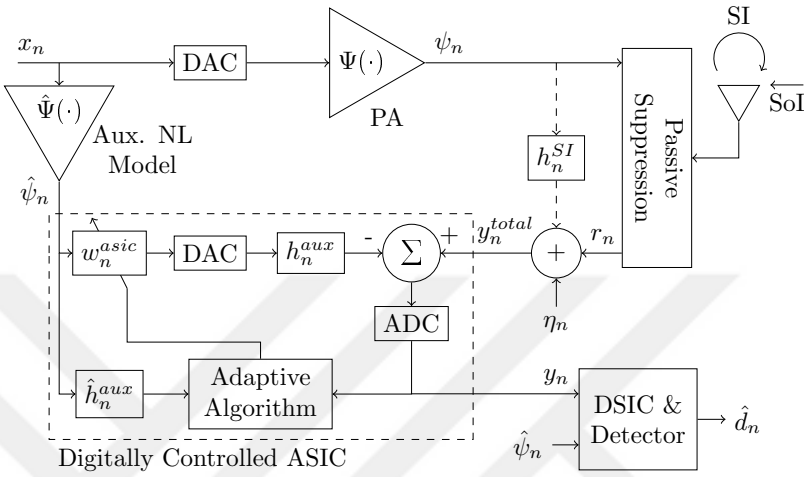


Figure 3.1: Digitally controlled ASIC architecture.

3.1.2 Time-Varying SI Channel Model

The Rician distributed SI channel in the k^{th} CPI is denoted by $h_n^{SI}[k]$ for $n = 0, 1, \ldots, L-1$. It consists of a constant mean component $\bar{h}_n$ for NFSI, and a TV Rayleigh component $\tilde{h}_n[k]$ for FFSI, as

$$h_n^{SI}[k] = \bar{h}_n + \tilde{h}_n[k]. \tag{3.5}$$

The ratio of the powers of the components is the Rician factor $f_n^{Rice} \triangleq |\bar{h}_n|^2/\sigma_n^2$, where σ_n^2 is the variance of the Rayleigh component. The power delay profile (PDP) can be defined as $\rho_n \triangleq |\bar{h}_n|^2 + \sigma_n^2$, which satisfies $\sum_{n=0}^{L-1} \rho_n = 1$. Given ρ_n and f_n^{Rice}, $|\bar{h}_n|^2$ and σ_n^2 can be calculated.

Based on a correlated block fading model, the TV Rayleigh component $\tilde{h}_n[k]$ is assumed to be a first-order autoregressive process [104], which is obtained as

$$\tilde{h}_n[k] = \alpha\, \tilde{h}_n[k-1] + \sqrt{1-\alpha^2}\, \sigma_n v_n[k], \tag{3.6}$$

where $0 \leqslant \alpha < 1$, and the innovation $v_n[k]$ is i.i.d. with $v_n[k] \sim \mathcal{CN}(0,1)$. Consequently, $h_n^{SI}[k]$ is a wide-sense stationary process over CPIs with mean $\bar{h}_n$ and covariance

$$E\left\{\tilde{h}_n^{SI}[k]\left(\tilde{h}_{n'}^{SI}[l]\right)^*\right\} = \sigma_n^2 \alpha^{|k-l|}\delta_{n-n'}, \tag{3.7}$$

where $0 \leqslant n, n' < L$. Doppler spread f_D is the reciprocal of the coherence time, which is assumed to be the time spent until the correlation between CPIs is halved according to (3.7). Then the normalized Doppler spread is defined as $f_D^{norm} \triangleq f_D/W$, where W is the operational bandwidth, which can be expressed as

$$f_D^{norm} = -\frac{\log_2(\alpha)}{N}, \tag{3.8}$$

which will be presented in parts per million (ppm). In order to give an idea about f_D^{norm} and make comparisons from real life, the $f_D = 2vf_c/c$ relation can be used, where v is the velocity of a scatterer, f_c is the carrier frequency, and c is the speed of light. Assuming $f_c/W = 100$, f_D^{norm} and α values are given in Table 3.1 for exemplary velocities.

Table 3.1: Time Variation and Normalized Doppler Spread

Velocity (m/s)	f_D^{norm} (f_c/W = 100)		α (N=100)	α (N=1000)
0.1	6.7×10^{-8}	0.067 ppm	0.999995	0.99995
0.3	2×10^{-7}	0.2 ppm	0.999986	0.99986
1	6.7×10^{-7}	0.67 ppm	0.999954	0.99954
3	2×10^{-6}	2 ppm	0.99986	0.9986
10	6.7×10^{-6}	6.7 ppm	0.99954	0.9954
30	2×10^{-5}	20 ppm	0.9986	0.9862

The reason for the Doppler spread in the SI channel is mainly the moving scatterers at the far-field of the antenna. Although the Rician component of the SI channel is not affected by the far-field scatterers, the Rayleigh component is. This effect significantly reduces SIC performance if it is not taken care of, as shown in this work. In the SI channel, Doppler spread values around 40 Hz, corresponding to a velocity of 2 m/s with a 3 GHz carrier frequency, can be seen in standard outdoor applications [46]. Depending on the CPI length and operational bandwidth, this Doppler frequency can

be detrimental to the SIC performance, as shown in the simulation results. Besides, the effects of the TV SI channel are shown to be more significant as the velocity increases by comparing the performances of FD systems for vehicle velocities of 3 m/s and 27 m/s [45].

3.1.3 Operation Modes

The proposed system employs three training phases with different periods, whose details will be explained in subsequent sections. These are called Mode-0, Mode-1, and Mode-2, in which SoI is absent. Mode-0 is a very-slow-time learning phase in which nonlinear parameter learning (NPL) is performed for very slowly varying hardware characteristics. It is performed infrequently, and noise is neglected due to the signal integration opportunity. Mode-1 and Mode-2 are two slow-time learning modes of K_1 and K_2 CPIs. After Mode-1 and Mode-2, a data transmission mode of K CPIs takes place with $K \gg (K_1 + K_2)$, which performs an adaptive fast-time SI prediction within its iterative structure to cope with the TV SI channel due to the mobility of scatterers. The cycle of Mode-1, Mode-2, and the data mode is repeated until NPL in Mode-0 is required again.

3.1.4 Digitally Controlled ASIC for Full-Duplex Radios

In the FD scheme, ASIC is generally inadequate to cancel the interference enough to establish successful communication. However, it is still needed to prevent the loss of SoI in quantization and to prevent the LNA from operating near saturation. The block diagram for the employed digitally controlled ASIC architecture is shown in Fig. 3.1. In Mode-0, the PA $\Psi(\cdot)$ is modeled by projecting ψ_n into GMP bases to obtain $\hat{\Psi}(\cdot)$, which is called the auxiliary nonlinear (NL) model. GMP with parameters K_{aux}, L_{aux}, and M_{aux} is used [51]. In Mode-1, which is responsible for filter adaptation in the ASIC stage, the signal

$$\hat{\psi}_n \triangleq \hat{\Psi}(\{x_n\}) \tag{3.9}$$

is adaptively filtered to cancel the received SI signal $y_n^{SI} \triangleq \psi_n \star h_n^{SI} + \eta_n$. Due to the existence of the auxiliary channel h_n^{aux} in the transmit chain, the FxLMS algorithm

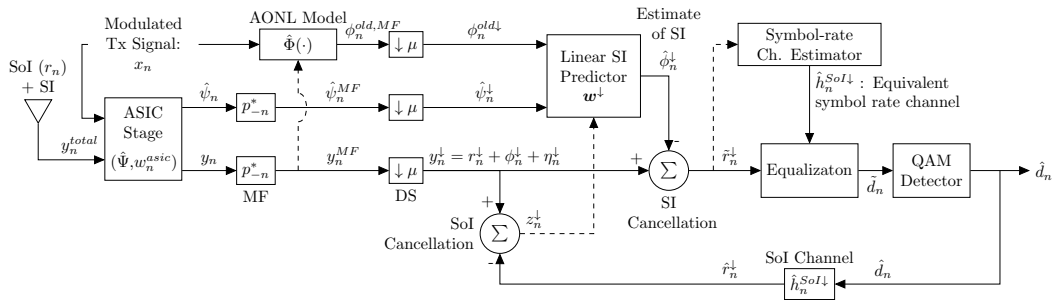


Figure 3.2: System diagram for Turbo-DSIC.

can be used efficiently, which is detailed in [51]. Adaptive filter learning is performed to obtain w_n^{asic} for $n = -L_{asic}, \ldots, +L_{asic}$, which will be called the ASIC filter henceforth. In Mode-2 and data mode, the filter w_n^{asic} is used, and the ASIC output $y_n[k]$ is obtained as

$$y_n[k] = y_n^{total}[k] - \hat{\psi}_n[k] \star w_n^{asic} \star h_n^{aux}. \tag{3.10}$$

3.2 Nonlinear Adaptive DSIC with Iterative SoI Detection: Turbo-DSIC

After the ASIC stage, the DSIC unit uses digital signal processing techniques, which offer more powerful tools to achieve the required SIC performance, such as iterative block processing. The system diagram for the proposed Turbo-DSIC scheme is illustrated in Fig. 3.2. The input of the DSIC stage $y_n[k]$ can be expressed in terms of three independent random signal components from (3.4) and (3.10) as

$$y_n[k] = \phi_n[k] + r_n[k] + \eta_n[k], \tag{3.11}$$

where residual SI $\phi_n[k]$ is defined as

$$\phi_n[k] \triangleq \psi_n[k] \star h_n^{SI}[k] - \hat{\psi}_n[k] \star w_n^{asic} \star h_n^{aux}. \tag{3.12}$$

Residual SI $\phi_n[k]$ and SoI $r_n[k]$ originated from known transmit (Tx) signal $x_n[k]$ and unknown receive (Rx) signal $s_n[k]$, respectively. The input $y_n[k]$ is passed through MF to obtain $y_n^{MF}[k] \triangleq y_n[k] \star p_{-n}^*$ and downsampled to symbol rate to obtain

$$y_n^{\downarrow}[k] \triangleq \sum_m y_m[k] p_{m-\mu n}^*. \tag{3.13}$$

Similarly, $\hat{\psi}_n^{MF}[k]$ and $\hat{\psi}_n^{\downarrow}[k]$ are obtained from $\hat{\psi}_n[k]$, as can be seen in Fig. 3.2. The signal $y_n^{\downarrow}[k]$ can be written as

$$y_n^{\downarrow}[k] = \phi_n^{\downarrow}[k] + r_n^{\downarrow}[k] + \eta_n^{\downarrow}[k], \tag{3.14}$$

whose terms correspond to the terms in (3.11) after MF and DS.

3.2.1 Two-Stage Signal Decomposition for NL SI Prediction

This stage aims to predict the residual SI ϕ_n (or $\phi_n^{\downarrow}$), which is nonlinearly related to the Tx signal x_n. This task can be performed in various topologies, which will be discussed considering the TV SI channel and NPL complexity. In Fig. 3.3, four different topologies are given. Training rate (fast/slow) and observation signal are indicated below each relevant block, where observation might contain SoI or not, depending on the system design. The SI channel can be modeled as shown by number 1 in Fig. 3.3, where NPL should be repeated frequently due to the TV SI channel, which is highly complex. Although it is shown as a single-stage, it represents various NL modeling techniques in the literature regardless of the number of stages. In number 2, NPL is repeated in slow time, which results in the aging of the NL model. Later in this section, it is shown that this aging can be compensated by an additional two-input linear filter, which processes the NL model output and the output of the auxiliary NL model from the ASIC stage. Predictions from these two topologies are used in SIC and given to the symbol detector after MF and DS.

The third and fourth topologies move the MF and DS operations before the linear filter, reducing the linear filter's complexity with symbol-rate operation with less memory. The SoI processing is also easier, which makes turbo processing more practical. The fourth one trains the NL model with the received signal after MF, which has a further advantage in complexity in data mode that will be explained later. This work follows the fourth topology for its low-complexity structure and suitability for TV SI channels.

In the proposed system in Fig. 3.2, the fourth topology is adopted where NL and linear models are referred to as ASIC output NL (AONL) model and linear SI predictor, respectively. NPL for the NL model is performed in Mode-0, as indicated in Fig. 3.2

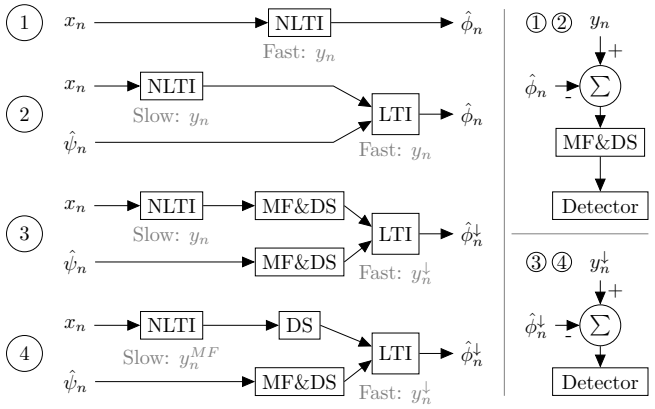


Figure 3.3: SI estimation topologies. (LTI: linear time-invariant system, NLTI: nonlinear time-invariant system.)

together with training details of other blocks. The linear model is equipped with RLS, and its learning is repeated in turbo iterations. In the remaining part of this section, details for topologies 2,3,4 will be explained. Then, the utilization of the first and second stages is described.

Only for this section, the variables related to the Mode-0 (NPL phase of the AONL model) are indicated by the superscript $M0$. Then, based on (3.12), ϕ_n^{M0} is written as

$$\phi_n^{M0} = \psi_n^{M0} \star h_n^{SI,M0} - \hat{\psi}_n^{M0} \star w_n^{asic,M0} \star h_n^{aux}, \tag{3.15}$$

where ψ_n^{M0}, $\hat{\psi}_n^{M0}$ and consequently ϕ_n^{M0} are nonlinearly related to Tx signal x_n^{M0}, such as $\phi_n^{M0} = \Phi\left(\left\{x_n^{M0}\right\}\right)$. Assume that $\Phi\left(\cdot\right)$ is learned perfectly in Mode-0, and after some time, it is fed with the Tx signal $x_n[k]$ in the data mode. Due to time variation, it cannot perfectly predict the current SI signal shown in (3.12), but it outputs

$$\phi_n^{old}[k] = \psi_n[k] \star h_n^{SI,M0} - \hat{\psi}_n[k] \star w_n^{asic,M0} \star h_n^{aux}. \tag{3.16}$$

Using (3.16), (3.12) can be rearranged as

$$\begin{aligned} \phi_n[k] &= \phi_n^{old}[k] \star \left(h_n^{SI,M0}\right)^{-1} \star h_n^{SI}[k] \\ &+ \hat{\psi}_n[k] \star w_n^{asic,M0} \star h_n^{aux} \star \left(h_n^{SI,M0}\right)^{-1} \star h_n^{SI}[k] \\ &- \hat{\psi}_n[k] \star w_n^{asic} \star h_n^{aux}. \end{aligned} \tag{3.17}$$

Simplifying (3.17) as

$$\phi_n[k] = \phi_n^{old}[k] \star w_n^1[k] + \hat{\psi}_n[k] \star w_n^2[k], \tag{3.18}$$

it is seen that $\phi_n[k]$ is linearly dependent on $\phi_n^{old}[k]$ and $\hat{\psi}_n[k]$ through the filters $w_n^1[k]$ and $w_n^2[k]$, which explains the two-input structure of the linear filters in Fig 3.3. This relation is still valid when each signal passes through MF. It can approximately be written in the downsampled domain as

$$\phi_n^{\downarrow}[k] \cong \phi_n^{old\downarrow}[k] \star w_n^{dsic,1\downarrow}[k] + \hat{\psi}_n^{\downarrow}[k] \star w_n^{dsic,2\downarrow}[k] \tag{3.19}$$

for the filters $w_n^{dsic,1\downarrow}[k]$ and $w_n^{dsic,2\downarrow}[k]$, which is the 3rd topology. The approximation is valid because signals are narrowband due to MF. Therefore, the selected samples can compensate for the effects of discarding highly correlated samples between the sampling points.

In the considered system and the fourth topology, the NL relation between x_n^{M0} and $\phi_n^{MF,M0} \triangleq \phi_n^{M0} \star p_{-n}^*$ (instead of ϕ_n^{M0}) is modeled as $\hat{\Phi}(\cdot)$. This model produces $\phi_n^{old,MF}[k] = \hat{\Phi}\left(\{x_n[k]\}\right)$ in the data mode. Since $\phi_n^{old,MF}[k]$ will be downsampled immediately to obtain $\phi_n^{old\downarrow}[k]$, the NL model does not need to produce the outputs that will be omitted, which reduces the computational complexity.

3.2.1.1 First Stage - Slow-Time NPL

Let y_n^{MF} in Fig. 3.2 be denoted as $y_n^{MF,M0}$ during Mode-0, in which SoI is absent and noise is neglected. Then, $y_n^{MF,M0} = \phi_n^{MF,M0}$ is observed for NPL training sequence x_n^{M0}. AONL model $\hat{\Phi}(\cdot)$ is fitted via GMP, NN, or GPR.

GMP: The GMP basis for triplet (k, l, m) is $x_{n-l}|x_{n-l-m}|^k$ for an arbitrary x_n signal. GMP coefficients are found by

$$\hat{\boldsymbol{c}} = (\boldsymbol{X}_{M0}^H \boldsymbol{X}_{M0})^{-1} \boldsymbol{X}_{M0}^H \boldsymbol{\phi}^{MF,M0}, \tag{3.20}$$

where $\boldsymbol{\phi}^{MF,M0} \triangleq [\phi_0^{MF,M0} \quad \dots \quad \phi_{\mu N_{M0}-1}^{MF,M0}]^T$. The matrix $\boldsymbol{X}_{M0}$, which is an $\mu N_{M0} \times (K_{GMP}+1)(2L_{NL}+1)(2M_{GMP}+1)$ matrix, involves nonlinear GMP bases in its columns created with x_n^{M0} for $n = 0, \dots, \mu N_{M0} - 1$, where N_{M0} is the number of symbols in Mode-0. These bases are for $k = 0, \dots, K_{GMP}$; $l = -L_{NL}, \dots, L_{NL}$; and $m = -M_{GMP}, \dots, M_{GMP}$; where K_{GMP}, L_{NL} and M_{GMP} are processing parameters.

In data mode, AONL model output is generated for the input $x_n[k]$, to obtain $\boldsymbol{\phi}^{old\downarrow}[k] \triangleq [\phi_0^{old\downarrow}[k] \quad \dots \quad \phi_{N-1}^{old\downarrow}[k]]^T$ as

$$\boldsymbol{\phi}^{old\downarrow}[k] = \boldsymbol{X}^{\downarrow}[k]\hat{\boldsymbol{c}}, \tag{3.21}$$

where $(\boldsymbol{X}^{\downarrow}[k])_{(i,:)} = (\boldsymbol{X}[k])_{(\mu i,:)}$ for $i = 1, \dots, N$ and $\boldsymbol{X}[k]$ is the GMP basis matrix created from $x_n[k]$, similar to $\boldsymbol{X}_{M0}$.

NN: The NN estimate based on the augmented real-valued time-delay neural network (ARVTDNN) architecture with one hidden layer of K_{NN} neurons [105] can be written as

$$\begin{bmatrix} \text{Re}\{\phi_n^{old\downarrow}[k]\} \\ \text{Im}\{\phi_n^{old\downarrow}[k]\} \end{bmatrix} = \mathbf{W}_2^T g(\mathbf{W}_1^T \mathbf{x}_{\mu n}^{NN}[k] + \mathbf{b}_1) + \mathbf{b}_2, \tag{3.22}$$

for $n = 0, \dots, N-1$ where

$$\begin{aligned} \mathbf{x}_n^{NN}[k] \triangleq \Big[&\text{Re}\{x_{n+L_{NL}}[k]\} \quad \dots \quad \text{Re}\{x_{n-L_{NL}}[k]\} \\ &\text{Im}\{x_{n+L_{NL}}[k]\} \quad \dots \quad \text{Im}\{x_{n-L_{NL}}[k]\} \Big]^T, \end{aligned} \tag{3.23}$$

and $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are obtained from $(\{x_n^{M0}\}, \{\phi_n^{MF,M0}\})$ pair during NPL, and $g(\cdot)$ is the activation function. Sizes of $\mathbf{W}_1$ and $\mathbf{W}_2$ are $(4L_{NL}+2) \times K_{NN}$ and $K_{NN} \times 2$, respectively. As seen, the network works in the real field, which decouples the input/output relation of real and imaginary parts. It helps to compensate for I/Q imbalance due to hardware impairments while working on actual hardware data [31].

Gaussian Process Regression (GPR): The final NL estimation method is the well-known GPR, a nonparametric Bayesian estimator in function space [106]. In GPR, a prior probability density function (PDF) is expressed in terms of hyperparameters, and posterior distributions are found by assuming Gaussian priors. The assignment of a priori PDFs is fulfilled during NPL. GPR finds the linear minimum mean-square error (LMMSE) estimate $\hat{\Phi}(\cdot)$ in function space by considering both the observations and the a priori PDFs. Error variance decreases with more training data, but the computational burden increases as more training is used, limiting the practical usage of GPR. The training data of the GPR method is created precisely like in the NN method to make a fair comparison between all NPL techniques.

3.2.1.2 Second Stage - Fast-Time Adaptive Linear SI Estimation

The residual SI estimate $\phi_n^{old\downarrow}[k]$ from the first stage is aged due to the TV SI channel. The second stage estimates the true residual SI $\phi_n^{\downarrow}[k]$ in (3.14) by using the aged estimate $\phi_n^{old\downarrow}[k]$ through (3.19) in an iterative manner from the observations contaminated by SoI. As seen from Fig. 3.2, in the i^{th} iteration of k^{th} CPI, the contaminated observation for $\phi_n^{\downarrow}[k]$ is

$$z_n^{\downarrow}[k,i] = \phi_n^{\downarrow}[k] + v_n^{SoI}[k,i], \tag{3.24}$$

for $i = 1, \ldots, I$, where

$$v_n^{SoI}[k,i] \triangleq r_n^{\downarrow}[k] - \hat{r}_n^{\downarrow}[k,i] + \eta_n^{\downarrow}[k] \tag{3.25}$$

is the residual contamination after SoI cancellation. Note that $z_n^{\downarrow}[k,i] = y_n^{\downarrow}[k] - \hat{r}_n^{\downarrow}[k,i]$ where the regeneration of estimated SoI $\hat{r}_n^{\downarrow}[k,i]$ will be explained in Section 3.2.2. In each iteration, $\hat{r}_n^{\downarrow}[k,i]$ is refined to weaken $v_n^{SoI}[k,i]$, and block RLS technique is utilized to estimate $\phi_n^{\downarrow}[k]$ from $z_n^{\downarrow}[k,i]$. According to (3.19) and (3.24), the system to be solved to calculate the fast-time linear SI prediction filter $\boldsymbol{w}^{\downarrow}[k,i]$ is

$$\boldsymbol{z}^{\downarrow}[k,i] = \boldsymbol{\Phi}[k]\boldsymbol{w}^{\downarrow}[k,i], \tag{3.26}$$

where $\boldsymbol{z}^{\downarrow}[k,i] \triangleq [z_0^{\downarrow}[k,i] \ldots z_{N-1}^{\downarrow}[k,i]]^T$ and

$$\boldsymbol{\Phi}[k] \triangleq \begin{bmatrix} \boldsymbol{\Phi}_1[k] & \boldsymbol{\Phi}_2[k] \end{bmatrix} \tag{3.27}$$

is the joint convolution matrix with $(\boldsymbol{\Phi}_1[k])_{pq} \triangleq \phi_{L_{dsic}+p-q}^{old\downarrow}[k]$ and $(\boldsymbol{\Phi}_2[k])_{pq} \triangleq \hat{\psi}_{L_{dsic}+p-q}^{\downarrow}[k]$ for $p = 1, \ldots, N$ and $q = 1, \ldots, 2L_{dsic}+1$. Therefore, the $2(2L_{dsic}+1) \times 1$ vector $\boldsymbol{w}^{\downarrow}[k,i]$, which will be called linear SI predictor henceforth, contains concatenated two vectors corresponding to $w_n^{dsic,1\downarrow}$ and $w_n^{dsic,2\downarrow}$ in (3.19). Then, block RLS recursions are given as

$$\boldsymbol{R}[k] = \lambda \boldsymbol{R}[k-1] + \boldsymbol{\Phi}^H[k]\boldsymbol{\Phi}[k], \tag{3.28}$$

$$\tilde{\boldsymbol{y}}[k,i] = \lambda \boldsymbol{y}[k-1] + \boldsymbol{\Phi}^H[k]\boldsymbol{z}^{\downarrow}[k,i], \tag{3.29}$$

for $k = 1, \ldots, K$ in data mode and $i = 1, \ldots, I$ where $\boldsymbol{y}[k] = \tilde{\boldsymbol{y}}[k, i = I]$, and $\boldsymbol{R}[0]$ and $\boldsymbol{y}[0]$ are provided by a training mode of K_2 CPIs, namely Mode-2, which

is an SoI-free initialization stage for RLS recursions. In Mode-2, $I = 1$ and steady-state $\boldsymbol{R}[0]$ and $\boldsymbol{y}[0]$ are obtained by running the recursions with SoI-free observation starting from $\boldsymbol{R}[-K_2] = \mathbf{0}$ and $\boldsymbol{y}[-K_2] = \mathbf{0}$. After RLS recursions at each iteration, the SI predictor is calculated by

$$\boldsymbol{w}^{\downarrow}[k,i] = (\boldsymbol{R}[k])^{-1}\tilde{\boldsymbol{y}}[k,i]. \tag{3.30}$$

Note that the matrix inversion is needed only once for a CPI, regardless of the number of iterations. When $\lambda = 0$, the block RLS method reduces to the least-squares (LS) method, which gives

$$\boldsymbol{w}^{\downarrow}[k,i] = (\boldsymbol{\Phi}^H[k]\boldsymbol{\Phi}[k])^{-1}\boldsymbol{\Phi}^H[k]\boldsymbol{z}^{\downarrow}[k,i]. \tag{3.31}$$

As will be shown in Sections 3.3 and 3.5, FD performance is very sensitive to time variation, which shortens the CPI. Since the required number of samples cannot be collected in a single CPI, the estimation accuracy of the SI predictor is limited. Therefore, the predictor obtained in the last iteration of the previous CPI is used directly in the first iteration of the current CPI. This prevents divergence due to turbo processing and provides a satisfactory initialization for the next iterations to track the SI channel variation successfully. The estimate $\hat{\boldsymbol{\phi}}^{\downarrow}[k,i] \triangleq [\hat{\phi}_0^{\downarrow}[k,i] \dots \hat{\phi}_{N-1}^{\downarrow}[k,i]]^T$ is calculated by

$$\hat{\boldsymbol{\phi}}^{\downarrow}[k,i] = \begin{cases} \boldsymbol{\Phi}[k]\boldsymbol{w}_0^{\downarrow}, & k = 1, i = 1, \\ \boldsymbol{\Phi}[k]\boldsymbol{w}^{\downarrow}[k-1,I], & k > 1, i = 1, \\ \boldsymbol{\Phi}[k]\boldsymbol{w}^{\downarrow}[k,i-1], & i > 1, \end{cases} \tag{3.32}$$

where $\boldsymbol{w}_0^{\downarrow}$ is the SI predictor calculated at the end of Mode-2, in the absence of SoI contamination. Assuming $\boldsymbol{w}_0^{\downarrow}$ is an accurate predictor and considering the opportunity of adaptation through iterations, usage of the previously calculated SI predictor in the first iteration sustains uncontaminated and advantageous conditions of Mode-2 throughout the data mode, despite the time variation and SoI contamination. Finally, SIC for i^{th} iteration is completed by

$$\tilde{r}_n^{\downarrow}[k,i] = y_n^{\downarrow}[k] - \hat{\phi}_n^{\downarrow}[k,i]. \tag{3.33}$$

3.2.2 Receiver Processing for SoI Detection and Regeneration

The receiver processing is applied on the signal $\tilde{r}_n^{\downarrow}[k,i]$ in (3.33). Using (3.14) and (3.33), it can be rewritten as

$$\tilde{r}_n^{\downarrow}[k,i] = r_n^{\downarrow}[k] + v_n^{SI}[k,i], \tag{3.34}$$

where $v_n^{SI}[k,i] \triangleq \phi_n^{\downarrow}[k] - \hat{\phi}_n^{\downarrow}[k,i] + \eta_n^{\downarrow}[k]$ is the residual SI after DSIC.

In this chapter, the SoI channel equalization and detection are out of the scope; however, in the literature, there are many algorithms proposed. For instance, LMMSE-based frequency domain equalization (FDE) or iterative blockwise decision feedback equalizers [107] can be employed to obtain the tentative symbol decisions, $\hat{d}_n[k,i]$, by using soft estimates $\tilde{d}_n[k,i]$ in Fig. 3.2.

Then, $\hat{d}_n[k,i]$ is passed through the estimated symbol-rate SoI channel $\hat{h}_n^{SoI\downarrow}$, which is the estimate of $h_n^{SoI\downarrow} \triangleq \left(h_m^{SoI} \star p_m \star p_{-m}^{*}\right)|_{m=\mu n}$, to obtain the regenerated SoI $\hat{r}_n^{\downarrow}[k,i]$ as

$$\hat{r}_n^{\downarrow}[k,i] = \hat{d}_n[k,i] \star \hat{h}_n^{SoI\downarrow}, \tag{3.35}$$

which is used to cancel the SoI contamination for the SI estimation process as

$$z_n^{\downarrow}[k,i] = y_n^{\downarrow}[k] - \hat{r}_n^{\downarrow}[k,i]. \tag{3.36}$$

The signal $z_n^{\downarrow}[k,i]$ is a contaminated observation of the residual SI signal $\phi_n^{\downarrow}[k]$, as stated in (3.24). The contamination $v_n^{SoI}[k,i]$ in (3.25) gradually weakens through iterations depending on the accuracy of $\hat{r}_n^{\downarrow}[k,i]$. As a result, SI estimate $\hat{\phi}_n^{\downarrow}[k,i]$ becomes more accurate and $v_n^{SI}[k,i]$ decreases. Together with the good start on $\hat{\phi}_n^{\downarrow}[k,i]$ provided by the SoI-free Mode-2 phase and usage of the filter from the previous CPI, this cycle of iterations improves the performance significantly.

3.2.3 Online Turbo-DSIC Summary

Throughout Section 3.2, all the steps of the Turbo-DSIC algorithm, which is held in the digital stage and illustrated in Fig. 3.2, are described. From these steps, the linear SI predictor (second stage of SI estimation) and SoI processing are repeated

Algorithm 1 Turbo-DSIC Algorithm

Input: K_1, K_2, K, I, N, μ, λ, L_{asic}, L_{dsic}, $\hat{\Psi}(\cdot)$, $\hat{\Phi}(\cdot)$

Mode-1 ($r_n[k] = 0$)

1: calculate w_n^{asic} for L_{asic} (Section 3.1.4) in K_1 CPIs

Mode-2 ($r_n[k] = 0$)

2: initialize $\boldsymbol{R}[-K_2] = \mathbf{0}$ and $\boldsymbol{y}[-K_2] = \mathbf{0}$

3: **for** $k = -K_2 + 1$ to 0 **do**

4: obtain $y_n^{\downarrow}[k]$ and $\hat{\psi}_n^{\downarrow}[k]$ by (3.13)

5: obtain $\phi_n^{old\downarrow}[k]$ by (3.21) or (3.22)

6: create $\boldsymbol{\Phi}[k]$ by (3.27) for memory L_{dsic}

7: obtain $\boldsymbol{R}[k]$ by (3.28)

8: obtain $\boldsymbol{y}[k] = \lambda \boldsymbol{y}[k-1] + \boldsymbol{\Phi}^H[k] y_n^{\downarrow}[k]$

9: **end for**

10: calculate $\boldsymbol{w}_0^{\downarrow} = (\boldsymbol{R}[0])^{-1} \boldsymbol{y}[0]$

Data Mode

11: **for** $k = 1$ to K **do**

12: obtain $y_n^{\downarrow}[k]$ and $\hat{\psi}_n^{\downarrow}[k]$ by (3.13)

13: obtain $\phi_n^{old\downarrow}[k]$ by (3.21) or (3.22)

14: create $\boldsymbol{\Phi}[k]$ by (3.27) for memory L_{dsic}

15: obtain $\boldsymbol{R}[k]$ by (3.28) and $(\boldsymbol{R}[k])^{-1}$

16: **for** $i = 1$ to I **do**

17: obtain $\hat{\phi}_n^{\downarrow}[k, i]$ by (3.32)

18: obtain $\tilde{r}_n^{\downarrow}[k, i]$ after DSIC by (3.33)

19: obtain tentative decisions $\hat{d}_n[k, i]$ (Section 3.2.2)

20: obtain regenerated SoI $\hat{r}_n^{\downarrow}[k, i]$ by (3.35)

21: obtain $z_n^{\downarrow}[k, i]$ after SoI cancellation by (3.36)

22: obtain $\tilde{\boldsymbol{y}}[k, i]$ by running the recursion in (3.29)

23: calculate $\boldsymbol{w}^{\downarrow}[k, i]$ as in (3.30)

24: **end for**

25: output $\boldsymbol{y}[k] = \tilde{\boldsymbol{y}}[k, i = I]$

26: output $\hat{d}_n[k, I]$ as final symbol decisions for k^{th} CPI

27: **end for**

28: return to step 1

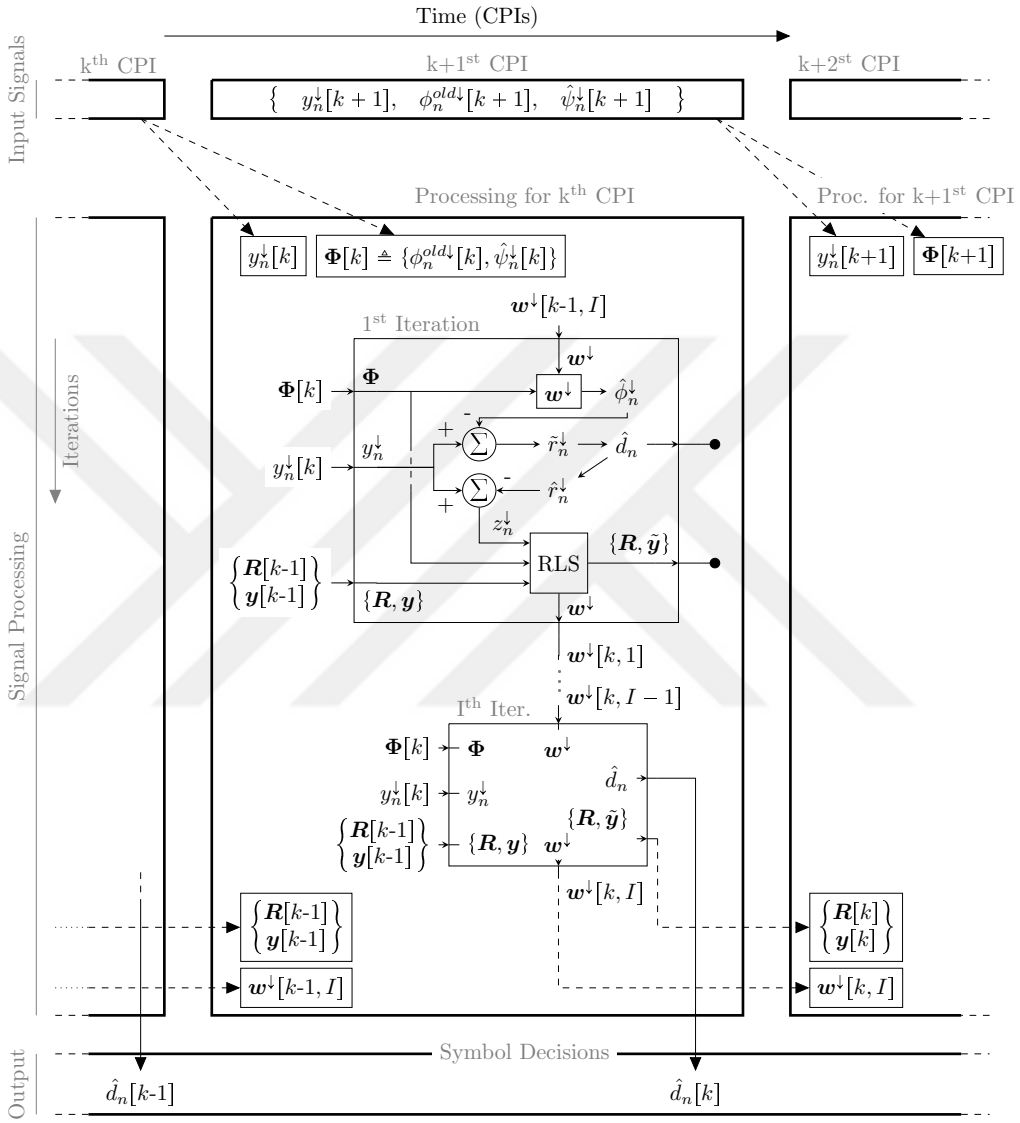


Figure 3.4: Progress of Turbo-DSIC algorithm in time.

iteratively. This iterative structure can be followed more easily from Algorithm 1 and Fig. 3.4.

In very-slow-time training, namely mode-0, auxiliary NL model $\hat{\Psi}(\cdot)$ and the AONL model $\hat{\Phi}(\cdot)$ are extracted via NPL as described in Sections 3.1.4 and 3.2.1.1. Then, Algorithm 1 is applied for mode-1, mode-2, and data mode, which are repeated until mode-0 is required again.

Fig. 3.4 illustrates the steps in the data mode, which are numbered from 11 to 27 in Algorithm 1. However, MF and DS (steps 12 and 13) are assumed as given. In addition, the fact that the inverse of the RLS recursion matrix $\boldsymbol{R}[k]$ is taken only once for a CPI (step 15) cannot be seen. From top to bottom in Fig. 3.4, the input signal stream is processed through iterations, and final decisions on symbols are output. Steps performed in an iteration are grouped in boxes, and progress through the iterations is clearly seen. The processing block for a CPI takes inputs also from the previous CPI, which are RLS variables, and these variables are conveyed to the next CPI similarly. Through iterations, most of the inputs of an iteration box stay the same, but the linear SI predictor $\boldsymbol{w}^{\downarrow}$ evolves.

3.3 Analysis on SIC Performance under Time-Varying SI Channel and SoI Contamination

In order to exhibit the challenges of FD communication under TV channels and verify the motivation behind the Turbo-DSIC algorithm, a theoretical analysis is performed for a simplified scenario, where PA nonlinearity and iterative processing are omitted. Consider that a continuous FD transmission is partitioned into blocks of N symbols, which defines CPIs to sustain coherency under a TV channel. After the k^{th} CPI, the received signal is

$$y_n[k] = x_n[k] \star h_n^{SI}[k] + s_n[k] + \eta_n[k], \tag{3.37}$$

where $x_n[k]$ is the SI signal, $h_n^{SI}[k]$ is the SI channel, $s_n[k]$ is the SoI, and $\eta_n[k] \sim \mathcal{CN}(0, \mu\sigma_\eta^2)$ is the AWGN. SI signal and SoI are expressed as in (3.1) and (3.3) for transmitted and received symbols $a_n[k]$ and $d_n[k]$, and magnitudes ρ_x and ρ_s, respectively. The signals $x_n[k]$, $s_n[k]$, $\eta_n[k]$ are defined for $n = -\mu N_p, \ldots, \mu(N + N_p - 1)$

where $2N_p$ is the span of p_n. The received signal can also be expressed in vector-matrix notation using vectors involving related signals for $n = -\mu N_p, \ldots, \mu(N + N_p - 1)$ as

$$\boldsymbol{y}[k] = \boldsymbol{s}[k] + \boldsymbol{X}[k]\boldsymbol{h}^{SI}[k] + \boldsymbol{\eta}[k], \tag{3.38}$$

where $(\boldsymbol{X}[k])_{ij} \triangleq x_{-\mu N_p+i-j}[k]$ for $i = 1, \ldots, N'$ and $j = 1, \ldots, L$, which makes $\boldsymbol{X}[k]$ an $N' \times L$ matrix, where

$$N' \triangleq \mu(N + 2N_p - 1) + 1. \tag{3.39}$$

The SI channel $\boldsymbol{h}^{SI}[k]$ is estimated by LS method, using the observation $\boldsymbol{y}^{obs}[k_{obs}]$ and SI signal $\boldsymbol{x}[k_{obs}]$, or equivalently the convolution matrix $\boldsymbol{X}[k_{obs}]$ as

$$\hat{\boldsymbol{h}}^{SI}[k] = (\boldsymbol{X}^H[k_{obs}]\boldsymbol{X}[k_{obs}])^{-1}\boldsymbol{X}^H[k_{obs}]\boldsymbol{y}^{obs}[k_{obs}], \tag{3.40}$$

where $k_{obs} \leqslant k$ and

$$\boldsymbol{y}^{obs}[k] \triangleq \boldsymbol{s}^{obs}[k] + \boldsymbol{X}[k]\boldsymbol{h}^{SI}[k] + \boldsymbol{\eta}[k], \tag{3.41}$$

where $\boldsymbol{s}^{obs}[k]$ is weakened SoI of power $\rho^2_{s,obs} \leqslant \rho^2_s$, to take the effect of SoI detection and cancellation via turbo processing into account. In the analysis, $\boldsymbol{s}^{obs}[k] = \frac{\rho_{s,obs}}{\rho_s}\boldsymbol{s}[k]$ is assumed after SoI cancellation [108]. SI is regenerated and cancelled to obtain the following signals.

$$\hat{\boldsymbol{s}}[k] \triangleq \boldsymbol{s}[k] + \boldsymbol{e}_s[k] \tag{3.42}$$

$$\boldsymbol{e}_s[k] \triangleq \boldsymbol{X}[k](\boldsymbol{h}^{SI}[k] - \hat{\boldsymbol{h}}^{SI}[k]) + \boldsymbol{\eta}[k] \tag{3.43}$$

After MF and DS, the vector $\tilde{\boldsymbol{d}}[k] \triangleq [\tilde{d}_0[k] \ldots \tilde{d}_{N-1}[k]]^T$ containing detector inputs is

$$\tilde{\boldsymbol{d}}[k] = \frac{1}{\rho_s\sqrt{\mu}}\boldsymbol{P}^H\hat{\boldsymbol{s}}[k], \tag{3.44}$$

where the $N' \times N$ matrix $\boldsymbol{P}$ conducts convolution with MF p^*_{-n} and downsampling by μ. It is assumed that $\mathbf{P}^H\mathbf{P} = \mathbf{I}_N$ due to the Nyquist-1 property of the pulse shaping filter. Then, the error vector $\boldsymbol{e}_d[k] \triangleq \tilde{\boldsymbol{d}}[k] - \boldsymbol{d}[k]$ of detector input and the average error power $P \triangleq E\{|\tilde{d}_n[k] - d_n[k]|^2\}$ are given as

$$\boldsymbol{e}_d[k] = \frac{1}{\rho_s\sqrt{\mu}}\boldsymbol{P}^H\boldsymbol{e}_s[k], \tag{3.45}$$

$$P = \frac{1}{N}\text{tr}\left(E\{\boldsymbol{e}_d[k](\boldsymbol{e}_d[k])^H\}\right). \tag{3.46}$$

3.3.1 Residual SI Power after DSIC

Total power P in (3.46) of the residual SI and the noise before detection, which will be called the residual SI power (RSIP), is calculated in the analysis provided in the Appendix B. From (B.17) and (B.25), and defining $\Delta_k \triangleq k - k_{obs}$, P is found as

$$\begin{aligned} P(\Delta_k, \rho_{s,obs}) =& C_x^{off}(1-\alpha^{\Delta_k})\frac{\rho_x^2}{\rho_s^2} \\ &+ C_s^{off}\frac{\rho_{s,obs}^2}{\rho_s^2} + (1+C_\eta^{off})\frac{\sigma_\eta^2}{\rho_s^2}, \end{aligned} \tag{3.47}$$

when $\Delta_k > 0$ and

$$P(\Delta_k, \rho_{s,obs}) = C_s^{on}\frac{\rho_{s,obs}^2}{\rho_s^2} + (1-C_\eta^{on})\frac{\sigma_\eta^2}{\rho_s^2}, \tag{3.48}$$

when $\Delta_k = 0$, for the power scaling constants C_x^{off}, C_s^{off}, C_η^{off}, C_s^{on}, C_η^{on}; definitions of which can be tracked from (B.17) and (B.25). These constants can be calculated based on the system parameters. Numerical calculations result in $C_x^{off} = 8.00e-4$, $C_s^{off} = 3.62e-2$, $C_\eta^{off} = 3.96e-2$, $C_s^{on} = 3.51e-2$, $C_\eta^{on} = 3.82e-2$ for 64-QAM, $N = 100$, $\mu = 8$, root raised cosine (RRC) p_n with 0.3 roll-off factor and span of 10 symbols, and the channel model described in Section 3.5. Since the power of detector inputs is normalized by (3.44), signal-to-interference-plus-noise ratio (SINR) before detection is $1/P(\Delta_k, \rho_{s,obs})$. Then, the spectral efficiency can be approximated as

$$R(\Delta_k, \rho_{s,obs}) \cong \log_2\left(1 + (P(\Delta_k, \rho_{s,obs}))^{-1}\right), \tag{3.49}$$

by assuming that the desired symbols $\boldsymbol{d}[k]$ and error signal $\boldsymbol{e}_d[k]$ are uncorrelated and Gaussian distributed [109].

3.3.2 Offline/Online Learning Schemes

In this analysis, two learning schemes are considered. In the offline learning scheme, the system estimates the SI channel in a training phase of a single CPI in the absence of SoI ($\rho_{s,obs} = 0$). Then, it cancels SI based on the estimated channel for K CPIs. This cycle of $K+1$ CPIs is repeated indefinitely. Then, average spectral efficiency (ASE) can be written as

$$R_{avg} = \frac{1}{K+1}\sum_{k'=1}^{K} R(\Delta_k = k', \rho_{s,obs} = 0). \tag{3.50}$$

In online learning, the SI channel is estimated directly from the current CPI in data mode, and $\Delta_k = 0$. There is no training phase, and SoI presence yields $\rho_{s,obs} = \rho_s$ conventionally, but $\rho_{s,obs} < \rho_s$ if SoI cancellation is applied as in the case of Turbo-DSIC. Then, ASE is

$$R_{avg} = R(\Delta_k = 0, \rho_{s,obs}). \tag{3.51}$$

Table 3.2: Training Modes

Block	Mode of Training	Training Frequency	Training Input → Observation
Aux. NL Model ($\hat{\Psi}$)	Mode-0	Very slow	$(x_n \rightarrow \psi_n)$
Lin. ASIC Filter (w_n^{asic})	Mode-1	Slow	$(\hat{\psi}_n \star \hat{h}_n^{aux} \rightarrow y_n^{total})$
AONL Model ($\hat{\Phi}$)	Mode-0	Very slow	$(x_n \rightarrow y_n^{MF})$
Lin. SI Predictor ($\boldsymbol{w}^{\downarrow}$)	Mode-2	Slow	$(\phi_n^{old\downarrow}, \hat{\psi}_n^{\downarrow} \rightarrow y_n^{\downarrow})$
	Data Mode	Fast	$(\phi_n^{old\downarrow}, \hat{\psi}_n^{\downarrow} \rightarrow z_n^{\downarrow})$

3.4 Complexity Analysis

In this section, the computational complexity of the proposed algorithm will be discussed on a time-average basis. Therefore, the proposed operational structure is important, which is designed carefully considering the effects of computationally complex model training. In Table 3.2, training objectives of the main blocks are detailed. Thanks to the proposed two-stage structure, highly complex NL model training operations do not dominate the overall complexity. The complexity of processing steps in the data mode dominates those in all training modes, thanks to their slow repetition rate. The numbers of multiplications per CPI of these steps are given in Table 3.3 together with the total DSIC complexity per CPI using $\mathcal{O}(\cdot)$ notation, for which dominant terms are chosen according to N. The first term in total DSIC complexity is due to the generation of $\phi_n^{old\downarrow}$ by the AONL model. It is the most dominant term because modeling upsampled signals requires larger memory. It is computationally complex; however, the necessity of upsampled NL modeling will be shown in Section 3.5 via an example of symbol-rate NL modeling. Note that the complexity has been reduced by a factor of μ since the model produces downsampled outputs; for example, GMP

uses $\boldsymbol{X}^{\downarrow}[k]$ instead of $\boldsymbol{X}[k]$ in (3.21). For GMP, $M_{GMP} = 0$ is assumed since it is the case for results in Section 3.5. Although the second and third terms are dominated by the first one, they show the effects of important variables of the algorithm, L_{dsic} and I. The second term shows the effect of L_{dsic} and it is due to the multiplication $\boldsymbol{\Phi}^H[k]\boldsymbol{\Phi}[k]$ in (3.28). The third term is due to the iterations through I, and it is the least effective, which shows that the Turbo-DSIC algorithm requires repetition of operations whose complexity is dominated by the others.

Table 3.3: Complexity per CPI

Operation	# of Mult.	Operation	# of Mult.
$\phi_n^{old\downarrow}$ (GMP)	$2K_{GMP}L_{NL}N$	$\hat{\phi}_n^{\downarrow}[k]$	$4L_{dsic}IN$
$\phi_n^{old\downarrow}$ (NN)	$4K_{NN}L_{NL}N$	Vector Recursion	$4L_{dsic}IN$
Matrix Recursion	$16L_{dsic}^2N$	$\boldsymbol{w}^{\downarrow}[k,i]$	$16L_{dsic}^2I$
Matrix Inverse	$64L_{dsic}^3$		
Total SIC (GMP)	$\mathcal{O}(K_{GMP}L_{NL}N) + \mathcal{O}(L_{dsic}^2N) + \mathcal{O}(L_{dsic}IN)$		
Total SIC (NN)	$\mathcal{O}(K_{NN}L_{NL}N) + \mathcal{O}(L_{dsic}^2N) + \mathcal{O}(L_{dsic}IN)$		

It is important to highlight the complexity-reducing effect of the proposed processing topology given in Fig. 3.3. Table 3.3 does not include the complexity of NL learning due to its infrequent repetition within our slow-time approach. Although complexities of NN and GPR are difficult to express, the complexity of NL learning for GMP is $\mathcal{O}(K_{GMP}^2L_{NL}^2\mu N_{M0})$ according to Eq. 3.20, but it repeated once in a very long time, not per-CPI. Considering per-CPI complexity in conventional schemes, $N_{M0} = N$ for online learning, and it is less than but proportional to N for offline learning (Fig. 3.7 evaluates $N_{M0} = N/2$ and $N_{M0} = N/11$ cases for offline learning). Therefore, the dominant complexity term in conventional processing is $\mathcal{O}(K_{GMP}^2L_{NL}^2\mu N)$, while that in the proposed one is $\mathcal{O}(K_{GMP}L_{NL}N)$. Consequently, the proposed processing scheme reduces complexity by $K_{GMP}L_{NL}\mu$ compared to the conventional ones. Note that, the considered values are $K_{GMP} = 5$, $L_{NL} = 16$ and $\mu = 8$ in Table 3.5.

3.5 Performance Evaluation

3.5.1 Discussion on the feasibility of FD communications under TV SI channels

Some of the main concerns about full-duplex communications under a TV SI channel are exhibited based on the analysis framework constructed in Section 3.3 via Fig. 3.5 and Fig. 3.6, and on simulation results for several representative SIC schemes via Fig. 3.7 and Table 3.4.

SIC in full-duplex systems is a task of learning/modeling the nonlinear SI channel, reproducing the SI signal and its cancellation. In the literature, different nonlinear modeling techniques are proposed to improve the modeling accuracy of the SI channel. The performance is generally measured by the decrease in the power of the residual SI signal after cancellation, which essentially measures the modeling quality. Rather than modeling quality, this work mainly investigates the effects of time variation in the SI channel on the performance of full-duplex systems and proposes solutions accordingly.

In the presence of a small variation in the SI channel, complications that can be faced are tried to be explained on a simple analytical model in Section 3.3. This model is constructed on a generic basis and does not depend on the details of the proposed algorithm. That is, this model and the results derived in this section are valid for any kind of algorithm, thanks to its simplicity. It lacks the coverage of the nonlinear effects, which is helpful to satisfy this generality. The results do not depend on the nature of hardware-related nonlinearity or nonlinear modeling techniques that can be used in an SIC algorithm. The main goal of this section is to emphasize the severity of time variation in the SI channel by demonstrating the failure of a linear setting in time-varying environments. Since a linear system is assumed, they constitute an upper bound for any system or algorithm that involves nonlinearity. Therefore, the severity of time variation is generic and valid for any system in the literature.

Online learning scheme is assessed in Fig. 3.5 by using (3.48), (3.49) and (3.51). It can be seen that it needs a long observation time N due to SoI contamination. However, N is limited by the need for coherence, whose effect is not covered by the

analysis. Residual SoI power (normalized by noise power) variations demonstrate that SoI cancellation provides a significant advantage. However, SoI cancellation is an iterative process, and online learning starts by observing the received SoI with $\rho_{s,obs} = \rho_s$, which is 20 dB in Fig. 3.5. Reduction in $\rho_{s,obs}$ is only possible if the initial accuracy of SoI detection is good enough. Therefore, achieving $\rho_{s,obs} < \rho_s$ cases, 0 and 10 dB residual SoI cases in Fig. 3.5, through turbo iterations is unlikely for online learning with small N due to poor starting SIC performance. On the other hand, if the process starts with the SI predictor obtained in the previous CPI admitting a channel aging brought by one CPI duration, it is almost certain that $\rho_{s,obs} < \rho_s$. In fact, if there is a single initialization phase with $\rho_{s,obs} = 0$, the first CPI detects and cancels SoI successfully. Then, subsequent CPIs start their processes with $\rho_{s,obs} < \rho_s$. Hence, achieving the performances for reduced $\rho_{s,obs}$ is more likely when the previously calculated SI predictor is used for the first iteration of SoI cancellation.

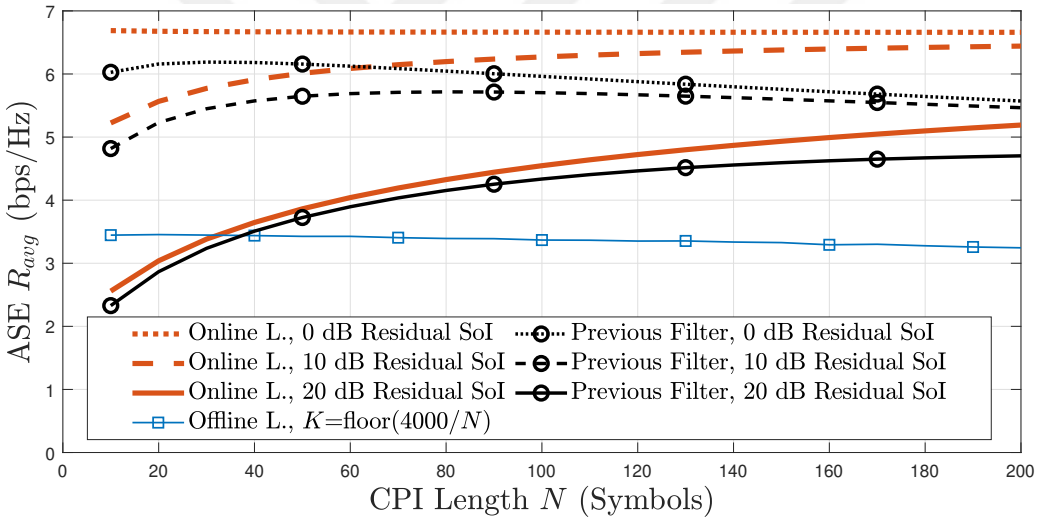


Figure 3.5: ASE versus N. (SI SNR=70dB, SoI SNR=20dB, f_D^{norm}=1 ppm.)

ASE performance of offline learning is investigated in Fig. 3.6 by using (3.47), (3.49) and (3.50). The effect of training overhead is observed as K decreases, where the SI channel estimate is updated frequently. When K is large and the previously estimated SI channel is used for a long time, channel aging becomes dominant. Under this trade-off, the offline learning scheme cannot reach the performance of the other options with any choice of K, unless the time variation is extremely slow. It is seen that SoI contamination and channel aging are two major challenges for FD communi-

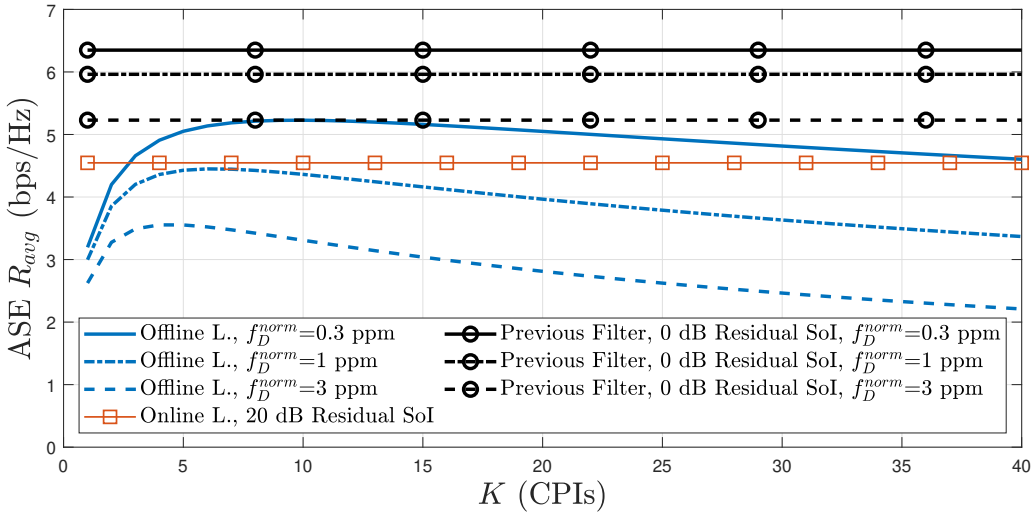


Figure 3.6: ASE versus K. (SI SNR=70dB, SoI SNR=20dB, N=100 Symbols.)

cation, even for negligible time variations in the SI channel. Offline learning scheme avoids the SoI contamination problem but suffers from aging and training overhead. Online learning scheme overcomes the aging problem; however, SoI contamination cannot be eliminated in a *coherent* processing interval. Unlike these two schemes, which entirely bypass one challenge and suffer from the other, the use of a previously established SI predictor and applying iterative SoI cancellation compromise these two challenges effectively.

Upon the findings from the analysis of a basic linear system, a numerical study is also conducted with simulations to better express the concerns about full-duplex communications under a TV SI channel and demonstrate the robustness of the proposed system. This study assumes a nonlinear system where the Saleh model is synthetically adopted for PA, and parameters in Table 3.5 are used. In Fig. 3.7, online [15, 31, 33–39] and offline [32, 45] learning schemes are compared with the proposed Turbo-DSIC algorithm. In these schemes, SI modeling is basically accomplished via the GMP (5,16,0) model simply as a single-stage, without any details such as DS, two-stage processing, turbo processing, or adaptiveness, to provide generality. In offline schemes, the SI channel is learned in an isolated (SoI-free) training period of x CPIs, and then the acquired model is used in a data period of y CPIs, which is indicated in the legend as 'Offline x-y'. In online schemes, there is no SoI-free training period. The model is learned in z SoI-contaminated CPIs in data mode and used

for the same CPIs, which is indicated as 'Online z'.

It is seen in Fig. 3.7 that offline schemes are generally better than the online schemes due to SoI-free learning when they are compared by training durations x and z. In time-invariant conditions, on the left side of the graph, performance can be enhanced by increasing these training durations. However, they are both vulnerable to the time variation in the SI channel, towards the right side of the graph. In the offline scheme, the learned channel in the training stage is aged in the test stage. This limits the 'exploit' duration of the 'explored' channel, as seen from 'Offline 10-10' and 'Offline 10-100' curves, which equivalently limits the spectral efficiency. The performance of the online scheme is limited by the intra-block time-coherency of the SI channel. The proposed Turbo-DSIC algorithm moves the sensitive region of 0.1 ppm to the 1-3 ppm without limiting the spectral efficiency. The aforementioned concerns can be summarized as spectral efficiency, SI channel learning accuracy, SoI contamination awareness, and robustness against channel aging. These schemes are also assessed in terms of these concerns in Table 3.4, where tick, tilde, and cross mean decent, moderate, and poor, respectively.

An SIC technique can model SI very accurately, achieving very low bit error rate (BER) in the left-hand side of Fig. 3.7, which is the time-invariant region. However, this work focuses more on how long it stays robust as the time variation in the SI channel increases. To the authors' knowledge, this factor is not studied explicitly in the literature. Thanks to the generality and simplicity of their system definition, the offline and online curves in Fig. 3.7 can be considered to represent the works in the literature. This study contributes to the literature by pointing out the sensitivity of full-duplex systems to the time variation in the SI channel, and then proposing a solution that moves the sensitive region from 0.1 ppm to 1-3 ppm. In addition, this solution is practical in terms of computational complexity, thanks to its two-stage structure, where computationally complex nonlinear learning is performed very rarely in the first stage and linear processing in the second stage is performed at the symbol rate.

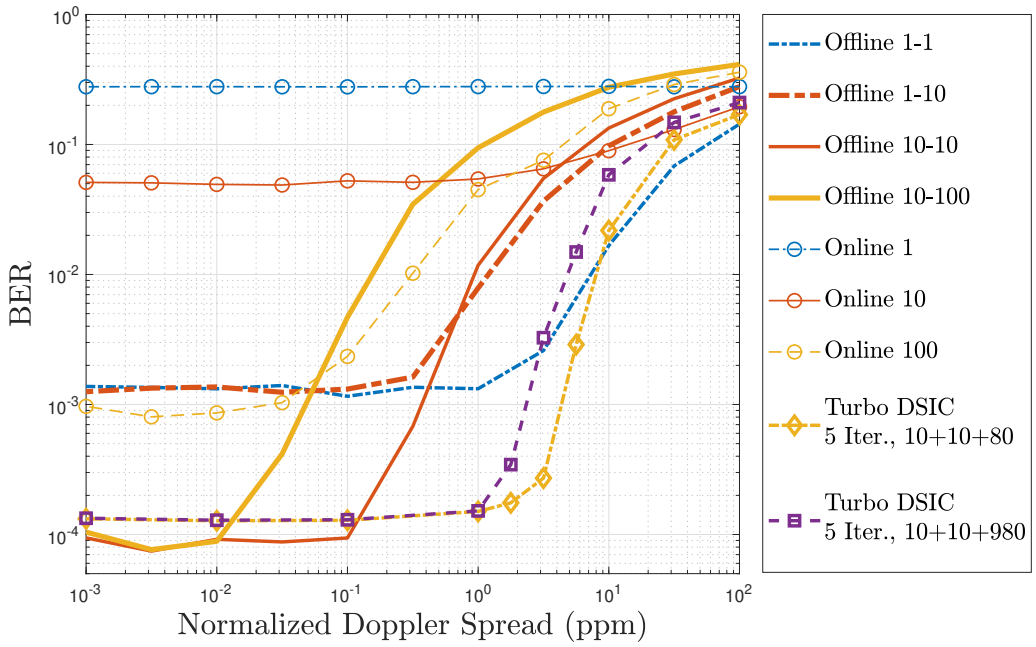


Figure 3.7: BER Performances of different SIC schemes against time variation in the SI channel. ($N = 100$, GMP with (5,16,0) for online and offline schemes.)

Table 3.4: Various SIC Schemes

Scheme	Spectral Efficiency	SI Channel Learning Accuracy	SoI Contamination Awareness	Robustness Against Channel Aging
Offline 1-1	50%	~	NA	✓
Offline 1-10	91%	~	NA	~
Offline 10-10	50%	✓	NA	~
Offline 10-100	91%	✓	NA	×
Online 1	100%	×	×	✓
Online 10	100%	×	~	~
Online 100	100%	~	✓	×
Turbo DSIC	80% / 98%	✓	✓	✓

3.5.2 Parameter Selection for Proposed System

List of parameter settings is given in Table 3.5. Parameters subjected to variation through simulations are indicated by $\sim$. The PDP ρ_n of the SI channel is adapted from [16]. Algorithm parameters are selected carefully to respond to some requirements. Auxiliary NL model parameters are selected to yield low complexity since it works in real-time. Nonlinearity parameters of the AONL model (K_{GMP}, K_{NN}) should be adequate to represent nonlinearity successfully. AONL model memory L_{NL} and $\mu \times L_{dsic}$ should be comparable with $L + L_{asic}$ since $w_n^{(asic)}$ might sweep the nonlinearity to larger delays. Selection of λ and $h_n^{(SoI)}$ is explained in the next paragraphs.

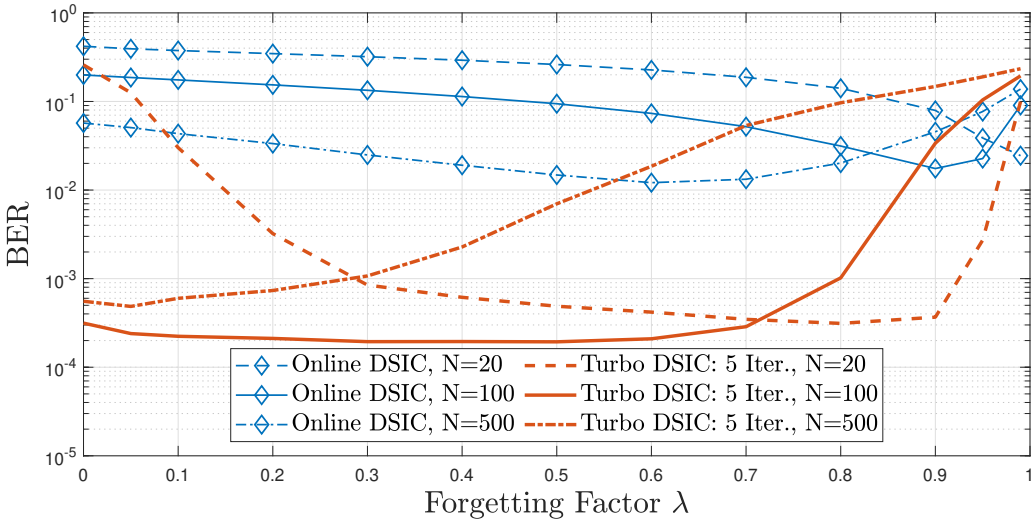


Figure 3.8: Performance against forgetting factor λ. (Saleh model, 30 dB SoI SNR, GMP for AONL, $L_{NL} = 16$, $L_{dsic} = 3$, f_D^{norm}=1 ppm.)

Performances of DSIC algorithms with respect to the forgetting factor λ are shown in Fig. 3.8 for different CPI lengths N. Note that α, K_1, K_2, and K are adjusted for $N = 20$ and $N = 500$ to keep the Doppler spread and time spent in each mode constant. Generally, it is seen that the optimum λ value increases as N decreases, and the optimum λ is less for the Turbo-DSIC than that of online DSIC. Turbo-DSIC outperforms online DSIC, showing the importance of iterative cancellation of SoI contamination via turbo processing. For $N = 100$, it is observed that λ is ineffective to performance until 0.7, which enables the choice $\lambda = 0$, reducing RLS to LS. RLS should be avoided for longer CPI lengths because the observation time is enough, and

Table 3.5: Simulation Parameters

Parameter	Description	Value / Details
M	QAM size	64
p_n	Pulse shaping filter	RRC (Span: 10 symbols, Roll-off factor: 0.3)
μ	Upsampling ratio	8
ρ_x^2/σ_η^2	SI SNR	70 dB
ρ_s^2/σ_η^2	SoI SNR	$\sim$ 30 dB
N	Symbols in a CPI	$\sim$ 100
N_{M0}	Symbols in mode-0	10000
K_1, K_2	CPIs in mode-1&2	10, 10
K	CPIs in data mode	980
ρ_n	SI channel PDP (before normalization)	0, -25, -30, ..., -70, -75 dB for $n = 0, \ldots, 11$ [16]
L	SI channel size	12
f_n^{Rice}	SI channel Rician factor	10 for $0 < n < L$, ∞ for $n = 0$
f_D^{norm}	Normalized Doppler spread	$\sim$ 1 ppm
$\Psi(\cdot)$	PA model	Saleh model with $(\alpha_a, \beta_a) = (2, 1)$ and $(\alpha_\phi, \beta_\phi) = (\pi/4, 0.25)$ [110] & 5 dB input backoff [111]
$\hat{\Psi}(\cdot)$	Aux. NL model	GMP
K_{aux}	GMP Parameters for Auxiliary NL model	3
L_{aux}		0
M_{aux}		0
L_{asic}	Linear ASIC filter memory	5
h_n^{aux}	Auxiliary line channel	Same as in [51]
$\hat{\Phi}(\cdot)$	AONL Model	GMP/NN/GPR
K_{GMP}	GMP parameters for AONL	5
M_{GMP}		0
K_{NN}	# of neurons in NN for AONL	30
L_{NL}	Common memory for AONL	$\sim$ 16
L_{dsic}	Linear SI predictor memory	$\sim$ 3
λ	RLS forgetting factor	0

information from the past is obsolete. Small CPI length requires the usage of RLS with larger λ values because the lack of adequate observation should be compensated by some additional information from the past, even if it is aged. Note that the time variation limits the choice of coherent duration N. Therefore, the LS method ($\lambda = 0$) can be used if an adequate CPI is available. If CPI is very short, the system can still work with an appropriate choice of λ for the block RLS technique. For this reason, the performance of the proposed algorithm will be evaluated for settings of CPI length $N = 100$ and forgetting factor $\lambda = 0$.

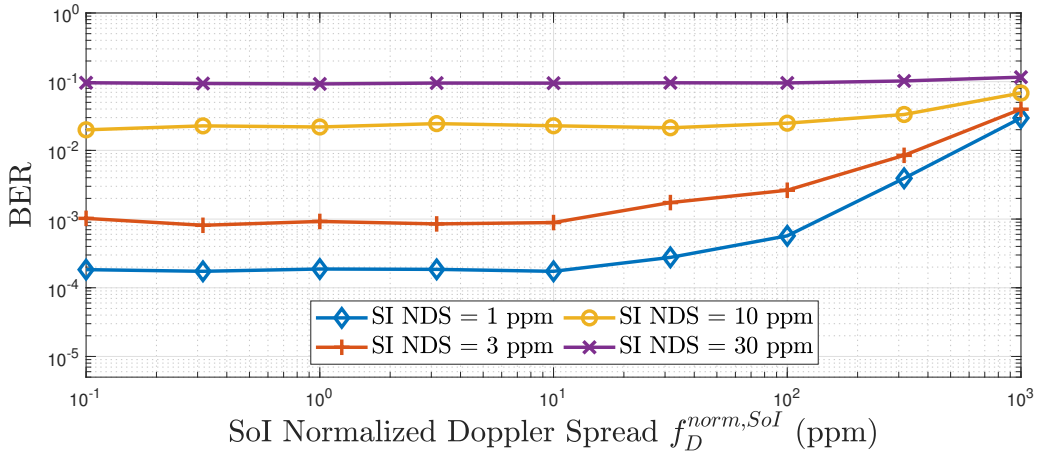


Figure 3.9: BER performance against time variation in SoI Channel. (Saleh Model, Turbo-DSIC with 5 iterations, GMP for AONL.)

BER performance against time variation in the SoI channel is plotted in Fig. 3.9. For this plot, the SoI channel is modeled similarly to the SI channel, which is described in Section 3.1.2. Only the Rician factor for the 0$^{\text{th}}$ tap, f_0^{Rice}, is also taken as 10 dB because the zeroth tap does not represent the hardware leakage for the SoI channel. SoI and SI channels are independent and uncorrelated. SoI channel is equalized in the frequency domain using a symbol-rate LMMSE equalizer. In equalization, simply the channel from the previous CPI is used as the estimate. Therefore, the SoI channel experiences a channel aging of one CPI duration.

In Fig. 3.9, it is seen that the time variation in the SoI channel starts to affect the performance slightly at around 30 ppm. However, BER increases from 10^{-4} to 10^{-3} when the normalized Doppler spread (NDS) for the SI channel increases from 1 ppm to 3 ppm. While 10 ppm of time variation in the SoI channel does not affect the

performance at all, the same amount of time variation in the SI channel limits BER to 10^{-2}.

As is seen from Fig. 3.9, the time variation in the SI channel is more detrimental than the time variation in the SoI channel. Due to all those facts, time variation in the SoI channel is kept out of the scope of this chapter, and time variation in the SI channel is focused on. SoI channel is assumed as AWGN with $h_n^{SoI} = \delta_n$; and equalization, together with normalization, is simply based on $\hat{h}_n^{SoI} = \sqrt{\mu}\rho_s\delta_n$.

3.5.3 Synthetic Case with Saleh PA Model

In this section, numerical results will be shared to evaluate the performance of the proposed algorithm. Performance of the Turbo-DSIC algorithm will be compared with different processing schemes, namely conventional receiver without DSIC, online DSIC, and online DSIC with SoI-free observation (genie-aided operation). In online DSIC, there is no adaptive and iterative SI prediction; instead, the SI is estimated directly via (3.19) using fully contaminated data with SoI. A hypothetical online DSIC with SoI-free operation utilizes $y_n^{\downarrow}[k] - r_n^{\downarrow}[k]$ while constructing the linear SI predictor $\boldsymbol{w}^{\downarrow}$. It is given as the benchmark for Turbo-DSIC performance.

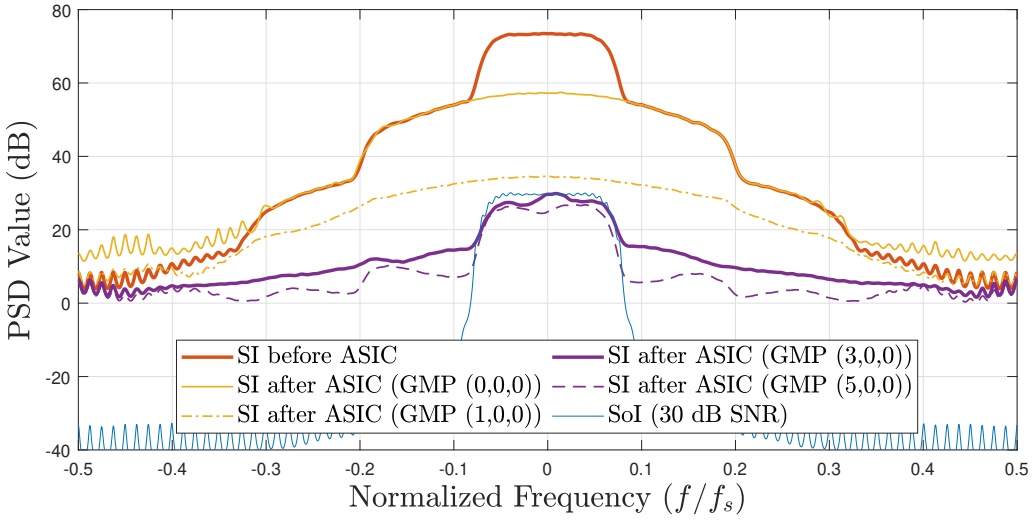


Figure 3.10: Spectra of signals before and after the ASIC stage for the Saleh model.

Spectra of the signals in the ASIC stage are shown in Fig. 3.10 for 70 dB SI signal power. The level of nonlinearity brought by the Saleh model with selected parameters

can be seen from the *SI before ASIC* spectrum. A polynomial of order K_{aux} (GMP model with $L_{aux} = M_{aux} = 0$) followed by the linear ASIC filter w_n^{asic} with $L_{asic} = 5$ performs the ASIC operation. They are trained in the training phases as explained in Section 3.2.3 and kept fixed during data transmission. Then, the spectrum of the *SI after ASIC* is obtained for different K_{aux} values. These results include the effect of aging of the linear ASIC filter during data transmission. Nonlinearity order for auxiliary NL model $K_{aux} = 3$ is selected for the rest of the results, which suppresses the SI to some extent to prevent the loss of SoI after quantization. However, there is still a leak of nonlinear SI whose power is about 30 dB. For comparison, the power spectral density (PSD) of an SoI with 30 dB SNR is also given.

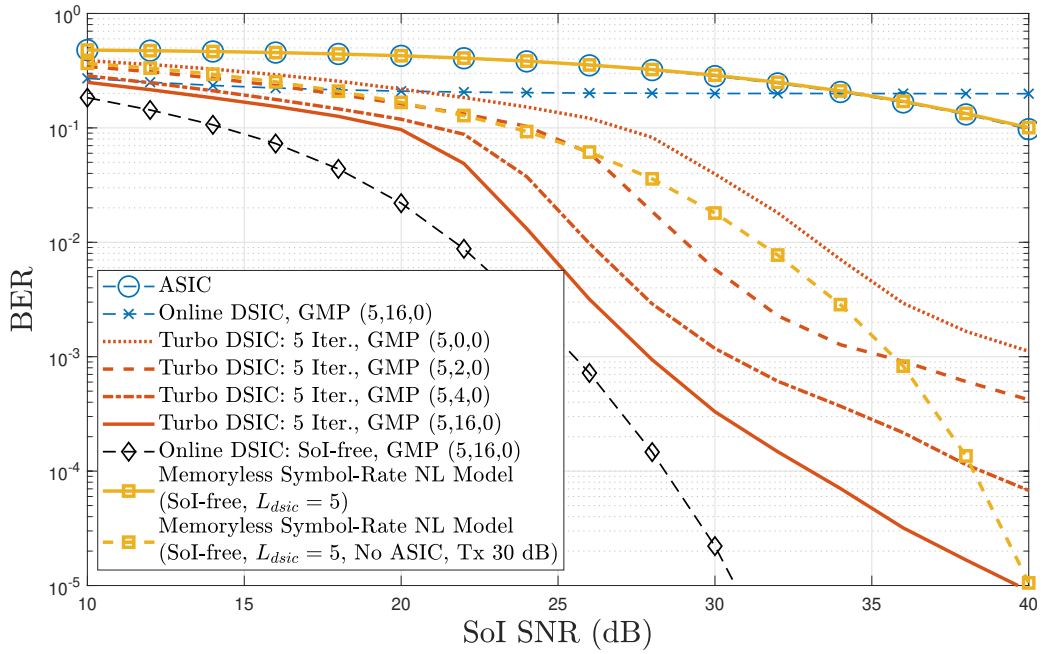


Figure 3.11: BER performance against SoI SNR. (Saleh model, GMP for AONL.)

Uncoded SoI detection performance of the Turbo-DSIC algorithm against SoI SNR can be seen in Fig. 3.11. The online DSIC fails even at high SoI SNR due to SoI contamination. Performances for the Turbo-DSIC algorithm are given for different GMP parameters of the AONL model, which are denoted by "GMP $(K_{GMP}, L_{NL}, M_{GMP})$". The performance of the Turbo-DSIC improves with the memory L_{NL}, getting closer to the genie-aided SoI-free training case. Under SoI-contaminated training, performance improvement tends to saturate at larger SoI power levels since the residual interference signal is correlated with SoI. Note that such a relatively low error floor

level of Turbo-DSIC in BER can be effectively eliminated by using forward error correction.

In addition, performances of two methods with memoryless symbol-rate NL model are given by using GMP (5,0,0) for AONL model with input a_n and output $\phi_n^{old\downarrow}$ in Fig. 3.11. This approach ignores the memory effects of nonlinearity and the violation of the Nyquist-1 property of pulse shaping. Moreover, only the first input $\phi_n^{old\downarrow}$ is used by the linear SI predictor, ignoring the second term in (3.19), which assumes only passive suppression (without employing active RF cancellation) as considered in the majority of recent DSIC studies such as [34, 50]. $L_{dsic} = 3 + 16/\mu = 5$ is chosen and performances for online DSIC with SoI-free observation case are given to compare it with the proposed Turbo-DSIC (GMP (5,16,0) and $L_{dsic} = 3$). This processing performs similar to direct detection after ASIC without providing any improvement. When the ASIC stage is removed and SI power is set to the ASIC output power level of the previous case (30 dB in Fig. 3.10) to meet the ASIC assumption of the approach in [50], performance improves deceptively. It shows that the scaling assumption on ASIC is not realistic, and claimed performances would not hold for a practical system model under TV conditions. In addition, the system with a memoryless NL model requires 5-10 dB more SNR to achieve the performance of the proposed Turbo-DSIC algorithm, showing the importance of the upsampled NL modeling with memory.

For the cases in Fig. 3.11, lower bounds on achievable information rate (AIR) are given in Fig. 3.12, where AIR values for FD cases are multiplied by two to signify two-way communication. In this study, the generalized mutual information (GMI) metric is used to find a lower bound on AIR [112]. Based on GMI, the mismatched decoding capacity, namely AIR, of the proposed FDE scheme can be expressed as

$$C_M = \log_2 M - E_{d,\tilde{d}}\left[\log_2\left(\frac{\sum_{d'\in\mathcal{X}} \tilde{p}(\tilde{d}|d')}{\tilde{p}(\tilde{d}|d)}\right)\right], \tag{3.52}$$

where $\tilde{p}(\tilde{d}|d)$ is the mismatched PDF approximated as Gaussian, based on Bussgang decomposition of FDE output $\tilde{d}$ in Section 3.2.2. It can be noted that the mismatched capacity expression in (3.52) can also be interpreted as the spectral efficiency in (3.49) for a finite-size constellation alphabet. There are memory options that outperform the HD case for any SoI SNR; a 2-fold AIR increase can be achieved for SoI SNR larger than 25 dB under 70 dB SI power.

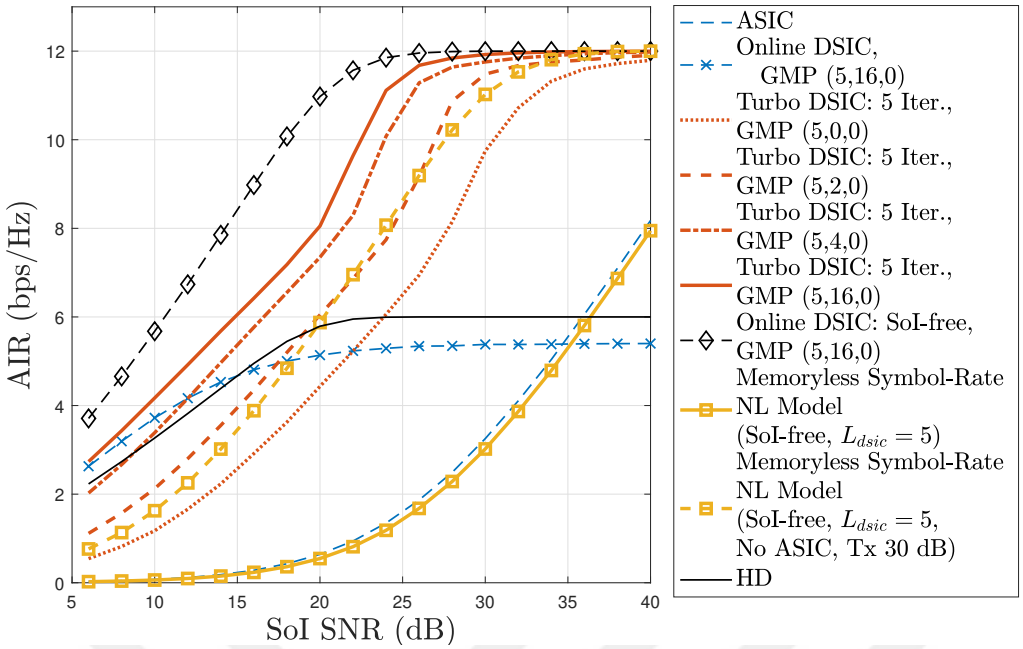


Figure 3.12: AIR performance against SoI SNR. (Saleh model, GMP for AONL.)

3.5.4 Hardware Implementation

The proposed Turbo-DSIC algorithm is implemented on the hardware setup shown in Fig. 3.13. This setup is based on the digitally controlled ASIC architecture given in Fig. 3.1, where FxLMS algorithm is utilized for the linear ASIC. Utilized settings are provided in Table 3.6. This measurement setup includes a Xilinx ZC706 carrier board with a programmable AD9375 transceiver, which is equipped with two RF transmitter chains to generate main and auxiliary transmitter RF signals, and an observation receiver chain to capture the PA output signal. The transmit signal is fed through a low-cost SKY66297-11 PA before the passive 25 dB SI suppression is achieved in the RF circulator. Note that, since a directive antenna with low return loss is connected to the setup, the ASIC algorithm must deal with the TV SI signal due to the reflections along the main beam. After a total of 75 dB analog cancellation is achieved in the ASIC stage, the remaining signal is captured by the AD9375 receiver for further DSIC.

The ASIC output y_n is measured to have -64 dBm total power, and its noise floor is around -80 dBm. The SoI SNR is determined by assuming -80 dBm noise power. SoI is synthetically added to the hardware ASIC output, and the Turbo-DSIC algorithm

Table 3.6: Hardware Settings and Measurements

System Parameters		Power Measurements	
Carrier Frequency	2630 MHz	Output Power @3	10 dBm
Signal Bandwidth	15.36 MHz	SI Power @2	-15 dBm
Sampling Rate	122.88 Msps	SI Power @1	-65 dBm
		Transmit Noise Floor	-80 dBm

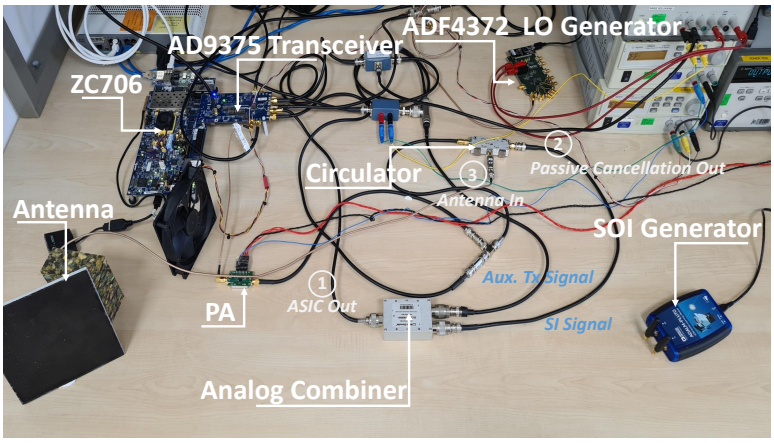


Figure 3.13: Experimental RF setup for in-band full-duplex transceiver.

is performed on this signal. Nonlinear modeling based on NN and GPR work in the real field, which enables the SI predictor to overcome the I/Q imbalance. BER performances are shown for different NL memory orders L_{NL} in Table 3.7, and SoI SNR levels in Fig. 3.14. Table 3.7 is constructed for 30 dB SoI SNR level, $L_{dsic} = 3$, $\lambda = 0$ and $N = 100$. Table 3.7 also shows SINR after DSIC, where SINR is 14 dB after performing ASIC. The Turbo-DSIC performance improves with increasing memory order of the AONL model. Among the modeling techniques, it is seen that NN is attractive compared to GMP and GPR-based NL modeling techniques, thanks to its superior SINR performance and complexity advantage.

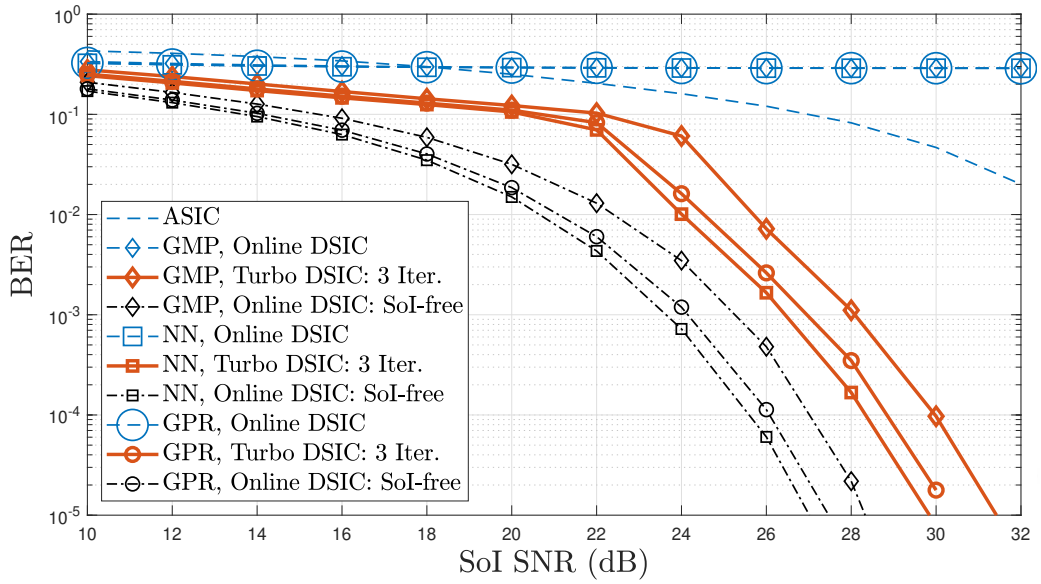


Figure 3.14: BER performance against SoI SNR. (HW data, $L_{NL} = 4$.)

Table 3.7: BER and SINR for Hardware Data

AONL Model	Metric	L_{NL}				
		0	1	2	4	8
GMP	BER	3.1e-04	1.6e-04	1.7e-04	9.6e-05	2.2e-05
	SINR	26.08 dB	26.61 dB	26.56 dB	26.87 dB	27.54 dB
NN	BER	2.8e-04	2.8e-05	1.4e-05	5.4e-06	4.9e-06
	SINR	25.99 dB	27.49 dB	28.01 dB	28.61 dB	28.83 dB
GPR	BER	2.8e-04	1.0e-04	6.5e-05	1.5e-05	6.3e-06
	SINR	25.95 dB	26.80 dB	27.27 dB	28.06 dB	28.59 dB

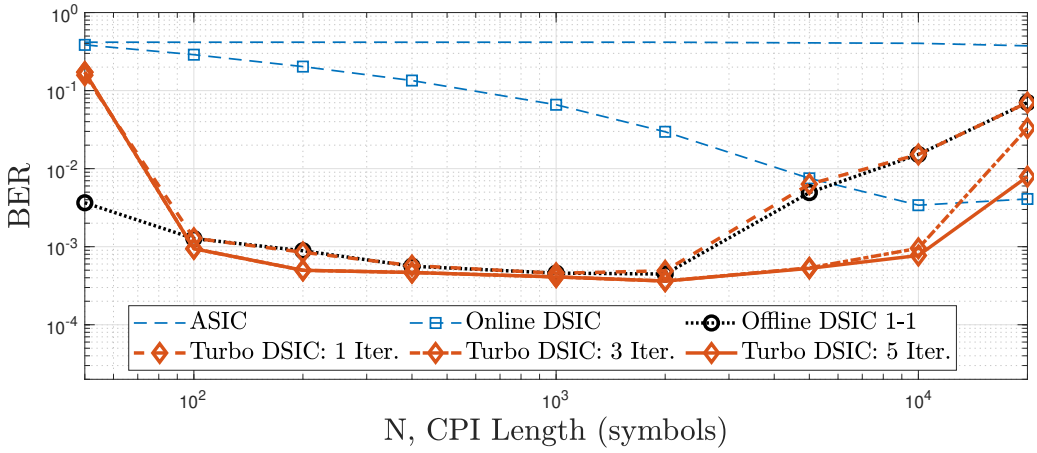


Figure 3.15: BER performance against assumed CPI length. (HW Data, GMP for AONL, $L_{NL} = 16$.)

Lastly, the performances of the algorithms with respect to the choice of CPI length are compared in Fig. 3.15. These results demonstrate the time-varying characteristics of the SI channel. For lower CPI lengths, DSIC with offline training outperforms DSIC with online and turbo training methods since the SI channel does not change significantly through small data packets due to high CPI to CPI SI channel correlation. However, channel acquisition before transmitting each small data packet decreases spectral efficiency in offline learning methods (with SoI free training). As the CPI length increases, the performance of the offline learning scheme significantly degrades due to the decrease in CPI-to-CPI SI channel correlation. However, the proposed iterative turbo DSIC algorithm exhibits more robust characteristics against SI channel aging. It can be seen from Fig. 3.15 that as the CPI length increases, the performances of all algorithms start to degrade since the channel decorrelates within a CPI duration. In other words, the SI channel at the beginning of the CPI block varies throughout the CPI block, and the SI channel at the end of the CPI block becomes significantly different, which yields performance degradation.

CHAPTER 4

SENSING AND MITIGATION OF MULTI-SCATTERER SELF-INTERFERENCE FOR FULL-DUPLEX MIMO COMMUNICATIONS

In this chapter, the joint use of digital self-interference cancellation (DSIC) and spatial suppression is investigated to mitigate far-field self-interference (SI) in full-duplex multiple-input multiple-output (MIMO) systems. The methods and techniques derived in this chapter are complementary to those in the Chapter 3. While this chapter focuses on time-varying far-field SI (FFSI), it frees the techniques in Chapter 3 from the time-variance-related concerns, allowing them to focus solely on NFSI and utilize more time integration for accuracy.

FFSI, caused by echoes from environmental scatterers, is modeled based on the scatterers' angle and delay parameters, stored in a scatterer map. For each scatterer, the most suitable action regarding communication is selected from transmit beamforming, receive beamforming, DSIC, and no action. This selection is based on simple metrics that show the expected uplink and downlink communication performance. In addition, emerging scatterers that deteriorate the communication are detected, and their delay and angles are acquired, providing an up-to-date scatterer map and presenting a *sensing for communication* case. The proposed selection policy is compared with the individual implementations of DSIC and spatial suppression, highlighting the failure cases for each. It is shown that the proposed policy stays unaffected in these problematic cases and achieves SI-free performance.

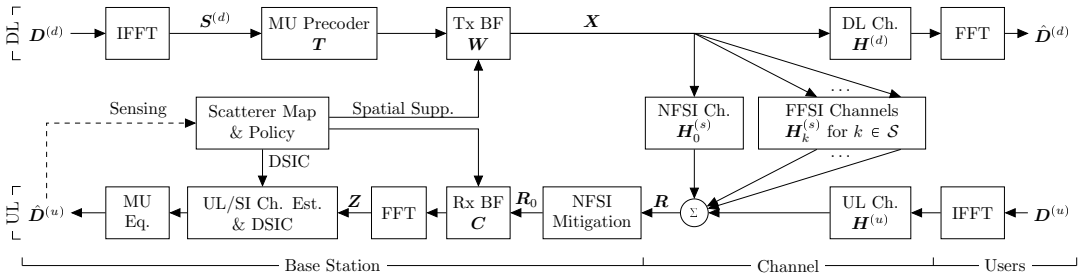


Figure 4.1: General system diagram. (Ch.: Channel, Est.: Estimator, MU: Multi-user, Eq.: Equalizer, FFT: Fast Fourier transform, IFFT: Inverse FFT).

4.1 System Model

The considered FD system features a multi-panel MIMO base station (BS), orthogonal frequency-division multiplexing (OFDM), and mmWave communication. K_u uplink (UL) users, K_d downlink (DL) users, and K_s scatterers are considered as shown in Fig. 4.2, forming the sets $\mathcal{U}_u \triangleq \{1, \ldots, K_u\}$, $\mathcal{U}_d \triangleq \{1, \ldots, K_d\}$, and $\mathcal{S} \triangleq \{1, \ldots, K_s\}$, respectively. Also, a UL/DL user pair in the same position describes an FD user equipment. While the users have single-antenna systems, the BS consists of collocated Tx and Rx units with uniform linear arrays (ULAs) with N_a antenna elements, where the multi-panel structure (i.e., separate-antenna type transceiver, see Section 1.1.1) weakens the NFSI. The BS employs a two-stage fully digital beamformer, with angle-based slow-time and reduced-dimensional fast-time beamforming stages. Variables related to UL/DL users and scatterers are indicated by u, d, and s, respectively. The system diagram in Fig. 4.1 illustrates signals, channels, and operations for BS and user equipments (UE) that will be detailed in the next subsections.

4.1.1 Transmitted Signals

Information symbols are denoted by $d_{k,m}^{(u)}$ for UL user $k \in \mathcal{U}_u$, and $d_{k,m}^{(d)}$ for DL user $k \in \mathcal{U}_d$, for subcarrier $m \in \{0, \ldots, N_c - 1\}$. UL users and the BS implement the inverse fast Fourier transform (IFFT) to obtain the time-domain signals $s_{k,n}^{(u)}$ and $s_{k,n}^{(d)}$

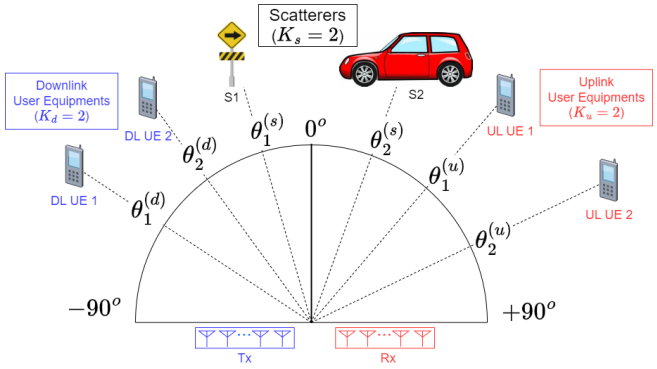


Figure 4.2: Exemplary scenario illustration and angular convention.

for time n. Defining $(\boldsymbol{S}^{(x)})_{kn} \triangleq s_{k,n}^{(x)}$, and $(\boldsymbol{D}^{(x)})_{km} \triangleq d_{k,m}^{(x)}$, it is shown as[1]

$$\boldsymbol{S}^{(x)} = \boldsymbol{D}^{(x)}\boldsymbol{F}^H, \qquad x \in \{u, d\}, \tag{4.1}$$

where $\boldsymbol{F}$ and $\boldsymbol{F}^H$ are FFT and IFFT matrices with $\boldsymbol{F}\boldsymbol{F}^H = \boldsymbol{I}_{N_c}$, and $(\boldsymbol{F})_{ab} \triangleq \frac{1}{\sqrt{N_c}} e^{-j\frac{2\pi}{N_c}(a-1)(b-1)}$. While UL users $k \in \mathcal{U}_u$ transmit $s_{k,n}^{(u)}$ directly, the BS transmits the beamformed signal $\boldsymbol{X}$, obtained as

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{T}\boldsymbol{S}^{(d)}, \tag{4.2}$$

where $\boldsymbol{W} \triangleq [\boldsymbol{w}_1 \cdots \boldsymbol{w}_{K_d}]$ is $N_a \times K_d$ statistical Tx beamforming matrix and $\boldsymbol{T}$ is $K_d \times K_d$ multi-user zero-forcing precoder with $\mathrm{tr}(\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}^H\boldsymbol{W}^H) = 1$. While $\boldsymbol{W}$ is responsible for directing beams towards DL users and partial suppression of FFSI, $\boldsymbol{T}$ forces the multi-user interference between DL users to zero. Expressing these two operations separately will be helpful in discussions in the next sections.

4.1.2 Channels

Channels have temporal and spatial characteristics. To describe the spatial aspects, the steering vector $\boldsymbol{a}(\theta)$ for a half-wavelength ULA with N_a elements will be used as

$$\boldsymbol{a}(\theta) \triangleq \frac{1}{\sqrt{N_a}} \begin{bmatrix} 1 & e^{j\pi\sin\theta} & \dots & e^{j(N_a-1)\pi\sin\theta} \end{bmatrix}^T, \tag{4.3}$$

where θ is the angle as shown in Fig. 4.2. When the angle is for scatterers or users,

$$\boldsymbol{q}_k^{(x)} \triangleq \boldsymbol{a}(\theta_k^{(x)}), \qquad x \in \{u, d, s\}, \tag{4.4}$$

[1] Cyclic prefix and oversampling are not shown for notational simplicity.

will be used. The spatial channels for the k^{th} DL user and the k^{th} UL user are denoted by $\boldsymbol{h}_k^{(d)}$ and $\boldsymbol{h}_k^{(u)}$, respectively, written as[2]

$$\boldsymbol{h}_k^{(x)} = \alpha_k^{(x)} \boldsymbol{q}_k^{(x)}, \qquad x \in \{u, d\}, \tag{4.5}$$

where $\alpha_k^{(x)} \sim \mathcal{CN}(0, P_k^{(x)})$. Reduced-dimensional effective channels $\boldsymbol{g}_k^{(d)}$ and $\boldsymbol{g}_k^{(u)}$ include the effect of beamformers as

$$\boldsymbol{g}_k^{(u)} \triangleq \boldsymbol{C}^H \boldsymbol{h}_k^{(u)}, \qquad \boldsymbol{g}_k^{(d)} \triangleq \boldsymbol{T}^H \boldsymbol{W}^H \boldsymbol{h}_k^{(d)}, \tag{4.6}$$

where $\boldsymbol{C} \triangleq [\boldsymbol{c}_1 \cdots \boldsymbol{c}_{K_u}]$ is the $N_a \times K_u$ Rx beamforming matrix. While $\boldsymbol{h}_k^{(x)}$ has length N_a, $\boldsymbol{g}_k^{(x)}$ has length $K_x < N_a$ for $x \in \{u, d\}$.

There are also SI channels to describe leakages from the Tx antennas to the Rx antennas, as shown in Fig. 4.1. While the NFSI channel $\boldsymbol{H}_0^{(s)}$ describes the direct leakage, FFSI is due to the scatterers in the environment. For the k^{th} scatterer, the $N_a \times N_a$ FFSI channel is

$$\boldsymbol{H}_k^{(s)} \triangleq \alpha_k^{(s)} \boldsymbol{q}_k^{(s)} (\boldsymbol{q}_k^{(s)})^H, \tag{4.7}$$

where $\alpha_k^{(s)} \sim \mathcal{CN}(0, P_k^{(s)})$. The FFSI channel depends on the scatterer angle $\theta_k^{(s)}$ through $\boldsymbol{q}_k^{(s)}$. Also, it depends on the distance of the scatterer through a delay component, $l_k^{(s)}$, whose effect will be shown later. Finally, the effective FFSI channel that includes the effects of beamformers is defined as

$$\boldsymbol{G}_k^{(s)} = \alpha_k^{(s)} \boldsymbol{C}^H \boldsymbol{q}_k^{(s)} (\boldsymbol{q}_k^{(s)})^H \boldsymbol{W} \boldsymbol{T}, \tag{4.8}$$

which is a system from K_d DL streams to K_u Rx chains.

4.1.3 Received Signals

The total signal received by the BS is shown by the matrix $\boldsymbol{R}$ of size $N_a \times N_c$ for antennas and time. It is written as

$$\boldsymbol{R} = \boldsymbol{H}^{(u)} \boldsymbol{S}^{(u)} + \boldsymbol{H}_0^{(s)} \boldsymbol{W} \boldsymbol{T} \boldsymbol{S}^{(d)} + \sum_{k=1}^{K_s} \boldsymbol{H}_k^{(s)} \boldsymbol{W} \boldsymbol{T} \boldsymbol{S}_{l_k^{(s)}}^{(d)} + \boldsymbol{N}, \tag{4.9}$$

[2] Temporal spread (multipath) and angular spread effects are not considered due to the well-known angle-delay sparsity in the mmWave channels [96–98]. In addition, possible delays between all users are considered to be compensated [113] as shown in (4.11), to focus on the effects of SI.

where $(\boldsymbol{H}^{(u)})_{:,k} = \boldsymbol{h}_k^{(u)}$, and $(\boldsymbol{N})_{:,n} \sim \mathcal{CN}(\boldsymbol{0}, N_0 \boldsymbol{I}_{N_a})$. In addition, $\boldsymbol{S}^{(d)}_{l_k^{(s)}}$ indicates the effect of delay $l_k^{(s)}$, shifting the columns of $\boldsymbol{S}^{(d)}$ to the right in a cyclic manner. Then, the BS applies NFSI cancellation using the NFSI channel estimate $\hat{\boldsymbol{H}}_0^{(s)}$ to obtain $\boldsymbol{R}_0 = \boldsymbol{R} - \hat{\boldsymbol{H}}_0^{(s)} \boldsymbol{W}\boldsymbol{T}\boldsymbol{S}^{(d)}$, expressed as

$$\boldsymbol{R}_0 = \boldsymbol{H}^{(u)}\boldsymbol{S}^{(u)} + \boldsymbol{E}_0^{(s)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{S}^{(d)} + \sum_{k=1}^{K_s} \boldsymbol{H}_k^{(s)}\boldsymbol{W}\boldsymbol{T}\boldsymbol{S}^{(d)}_{l_k^{(s)}} + \boldsymbol{N}, \tag{4.10}$$

where $\boldsymbol{E}_0^{(s)} = \boldsymbol{H}_0^{(s)} - \hat{\boldsymbol{H}}_0^{(s)}$ is the channel estimation error. Then, the BS applies Rx beamforming and fast Fourier transform (FFT) to obtain the matrix $\boldsymbol{Z} \triangleq \boldsymbol{C}^H \boldsymbol{R}_0 \boldsymbol{F}$, of size $K_u \times N_c$ for UL users and subcarriers. It is written as

$$\boldsymbol{Z} = \boldsymbol{G}^{(u)}\boldsymbol{D}^{(u)} + \boldsymbol{Z}_0^{(s)} + \sum_{k=1}^{K_s} \boldsymbol{G}_k^{(s)}\boldsymbol{D}^{(d)}\boldsymbol{E}_{l_k^{(s)}} + \boldsymbol{N}^{(z)}, \tag{4.11}$$

where $(\boldsymbol{G}^{(u)})_{:,k} = \boldsymbol{g}_k^{(u)}$, $\boldsymbol{E}_l \triangleq \text{diag}\{e^{-j\frac{2\pi}{N_c}ml}\}_m$, and $(\boldsymbol{N}^{(z)})_{:,n} \sim \mathcal{CN}(\boldsymbol{0}, N_0 \boldsymbol{C}^H\boldsymbol{C})$ is the noise term. In addition, $\boldsymbol{Z}_0^{(s)} \triangleq \boldsymbol{C}^H \boldsymbol{E}_0^{(s)} \boldsymbol{W}\boldsymbol{T}\boldsymbol{S}^{(d)}\boldsymbol{F}$ is the residual NFSI term. Thus, the observation consists of SoI, SI, and noise terms.

For DL, the k^{th} user receives $(\boldsymbol{h}_k^{(d)})^H \boldsymbol{X} = (\boldsymbol{g}_k^{(d)})^H \boldsymbol{S}^{(d)}$ and applies FFT to reach information in N_c subcarriers. The information obtained by each user is written into the rows of a $K_d \times N_c$ matrix $\hat{\boldsymbol{D}}^{(d)}$, expressed as

$$\hat{\boldsymbol{D}}^{(d)} = (\boldsymbol{H}^{(d)})^H \boldsymbol{W}\boldsymbol{T}\boldsymbol{D}^{(d)} + \boldsymbol{N}^{(d)}, \tag{4.12}$$

$$= (\boldsymbol{G}^{(d)})^H \boldsymbol{D}^{(d)} + \boldsymbol{N}^{(d)}, \tag{4.13}$$

where $\boldsymbol{G}^{(d)} \triangleq [\boldsymbol{g}_1^{(d)} \cdots \boldsymbol{g}_{K_d}^{(d)}]$, $\boldsymbol{H}^{(d)} \triangleq [\boldsymbol{h}_1^{(d)} \cdots \boldsymbol{h}_{K_d}^{(d)}]$, and $\boldsymbol{N}^{(d)}$ is the noise term.

4.1.4 Multi-User Equalization and Precoding

Multi-user equalization for the UL communication can be implemented via the least-squares (LS) method [114] to obtain the soft information symbol estimates $\hat{\boldsymbol{D}}^{(u)}$ as

$$\hat{\boldsymbol{D}}^{(u)} = \left((\hat{\boldsymbol{G}}^{(u)})^H \hat{\boldsymbol{G}}^{(u)}\right)^{-1} (\hat{\boldsymbol{G}}^{(u)})^H (\boldsymbol{Z} - \hat{\boldsymbol{Z}}_{si}). \tag{4.14}$$

The UL channel estimate $\hat{\boldsymbol{G}}^{(u)}$ and the regenerated SI $\hat{\boldsymbol{Z}}_{si}$ are the products of DSIC stage preceding this equalization as shown in Fig. 4.1, to be detailed in Section

4.2.2.2. The obtained symbol estimates $\hat{\boldsymbol{D}}^{(u)}$ are related to the actual ones as

$$\hat{\boldsymbol{D}}^{(u)} = \boldsymbol{M}^{(u)}\boldsymbol{D}^{(u)} + \boldsymbol{\Xi}^{(u)}, \tag{4.15}$$

$$\boldsymbol{M}^{(u)} \triangleq \left((\hat{\boldsymbol{G}}^{(u)})^H \hat{\boldsymbol{G}}^{(u)}\right)^{-1} (\hat{\boldsymbol{G}}^{(u)})^H \boldsymbol{G}^{(u)}, \tag{4.16}$$

where $\boldsymbol{\Xi}^{(u)}$ contains the residual SI and the noise. For the UL user k, the information symbol estimates can be written as

$$\hat{\boldsymbol{d}}_k^{(u)} = m_{kk}^{(u)} \boldsymbol{d}_k^{(u)} + \sum_{j=1, j\neq k}^{K_u} m_{kj}^{(u)} \boldsymbol{d}_j^{(u)} + \boldsymbol{\xi}_k^{(u)}, \tag{4.17}$$

where $m_{ij}^{(u)} \triangleq \left(\boldsymbol{M}^{(u)}\right)_{ij}$. The terms are for intended symbols, multi-user interference, and SI and noise, respectively. In addition, $\hat{\boldsymbol{d}}_i^{(u)}$, $\boldsymbol{d}_i^{(u)}$, and $\boldsymbol{\xi}_i^{(u)}$ are the i^{th} row of $\hat{\boldsymbol{D}}^{(u)}$, $\boldsymbol{D}^{(u)}$, and $\boldsymbol{\Xi}^{(u)}$, respectively. Note that $\boldsymbol{M}^{(u)} = \boldsymbol{I}_{K_u}$ and $m_{ij}^{(u)} = \delta[i-j]$ for the perfect estimation case where $\hat{\boldsymbol{G}}^{(u)} = \boldsymbol{G}^{(u)}$. So, the objective of the equalization is to mitigate the multi-user interference.

The corresponding objective on the DL signal $\hat{\boldsymbol{D}}^{(d)}$ in (4.12) is called multi-user precoding, implemented by the BS via the matrix $\boldsymbol{T}$. Note that, $\boldsymbol{H}^{(d)} = \boldsymbol{Q}^{(d)}\boldsymbol{\Lambda}_d$ where $\boldsymbol{Q}^{(d)} \triangleq [\boldsymbol{q}_1^{(d)} \cdots \boldsymbol{q}_{K_d}^{(d)}]$ and $\boldsymbol{\Lambda}_d \triangleq \text{diag}\{\alpha_k^{(d)}\}_k$, yielding $\boldsymbol{\Lambda}_d^H (\boldsymbol{Q}^{(d)})^H \boldsymbol{W}\boldsymbol{T}\boldsymbol{D}^{(d)}$ for the first term in (4.12). Then, $\boldsymbol{T}$ is designed as a statistical zero-forcing precoder that omits $\boldsymbol{\Lambda}_d$, written as

$$\boldsymbol{T} \triangleq \gamma_t \left((\boldsymbol{Q}^{(d)})^H \boldsymbol{W}\right)^{-1}, \tag{4.18}$$

where γ_t satisfies $\text{tr}(\boldsymbol{W}\boldsymbol{T}\boldsymbol{T}^H\boldsymbol{W}^H) = 1$. It yields

$$\hat{\boldsymbol{D}}^{(d)} = \boldsymbol{M}^{(d)}\boldsymbol{D}^{(d)} + \boldsymbol{N}^{(d)}, \tag{4.19}$$

where $\boldsymbol{M}^{(d)} = \gamma_t \boldsymbol{\Lambda}_d^H$. As seen, the multi-user interference is forced to zero since $\boldsymbol{M}^{(d)}$ is diagonal, but the phase and gain correction is left to the users.

4.1.5 Power Expressions

In the signal model, transmit power and path loss are governed only by the channel coefficients $\alpha_k^{(x)}$ and their variances $P_k^{(x)}$ for $x \in \{u, d, s\}$ for simplicity. To relate to the physical parameters, the UL signal power at a single element of the BS Rx array:

$P_{R,UL}$, and the SI power at a single Rx element due to a single Tx element: $P_{R,SI}$, can be written as

$$P_{R,UL} = P_{UE}\frac{\lambda^2}{(4\pi)^2 d_u^2}, \qquad P_{R,SI} = \frac{P_{BS}}{N_a}\frac{\lambda^2\sigma}{(4\pi)^3 d_s^4}, \tag{4.20}$$

using Friis' transmission and radar range equations, respectively, for isotropic antenna elements with unity gains.[3] In these equations, λ is the wavelength, P_{UE} is the user Tx power, P_{BS} is total Tx power for BS, σ is the radar cross section (RCS) for scatterers, and d_u and d_s are the distances of the user and scatterer, respectively. In addition, $P_{R,UL}$ and $P_{R,SI}$ are $\frac{P_k^{(u)}}{N_a}$ and $\frac{P_k^{(s)}}{N_a^3}$, respectively, according to the signal model due to normalizations in the channel and Tx BF definitions. Consequently, the variances in the signal model are expressed in terms of physical parameters as

$$P_k^{(u)} = P_{UE}\frac{N_a\lambda^2}{(4\pi)^2 d_u^2}, \qquad P_k^{(s)} = P_{BS}\frac{N_a^2\lambda^2\sigma}{(4\pi)^3 d_s^4}, \tag{4.21}$$

which include the Tx powers, path loss, and scattering effects.

4.2 Two-stage SI Mitigation Scheme

Two-stage SI mitigation is common in the related literature, where the steps of the mitigation process are separated for numerous reasons, particularly complexity, but the SI is generally treated as a whole. Instead, a time-variance-based two-stage SI mitigation process can separately aim at the time-invariant NFSI component and the time-varying FFSI components. This separation provides a modular structure that can address the specific characteristics and requirements of these interference terms better.

The proposed time-variance-based separation overcomes the dilemma of the *accuracy* and *freshness* in channel estimation under time-variation, also known as the explore/exploit dilemma [58]. While accuracy requires long observation (decreasing the update rate), freshness requires fast update rates (decreasing the observation time). The first stage integrates as much observation as possible in an iterative manner, acquired from the previous communication signals. Using this long observation for accuracy, the time-invariant NFSI channel is estimated, neglecting the fast update

[3] Antenna gains for whole arrays would depend on beamforming weights and user/scatterer positions. Instead, the element-to-element definitions of $P_{R,UL}$ and $P_{R,SI}$ overcome this dependency and provide a general validity.

need stemming from the time-variance of the FFSI component. The separation of NFSI and FFSI components is handled naturally by the low-pass characteristics of this operation, based on their distinct time-variance characteristics. After NFSI is re-generated and subtracted, remaining time-varying FFSI components are mitigated in the second stage using *DSIC* and *spatial suppression* via transmit beamforming (Tx BF) and receive beamforming (Rx BF). Both stages are designed to be performed simultaneously during communication, without waiting or interruptions.

4.2.1 NFSI Cancellation

NFSI cancellation leverages the passive suppression thanks to the multi-panel structure. Further, the system utilizes an estimate of the NFSI channel $\boldsymbol{H}_0^{(s)}$ for SIC, continuously refined by the UL signal $\boldsymbol{R}$ in (4.9). To describe the estimation procedure with multiple OFDM symbols, the OFDM symbol index $p \in \{1, \ldots, P\}$ is indicated as a subscript for various previously defined variables, such as $\boldsymbol{R}_p$, $\boldsymbol{X}_p$, $\boldsymbol{W}_p$, $\boldsymbol{T}_p$, $\boldsymbol{S}_p^{(d)}$. At P^{th} OFDM symbol, the collected observation is

$$\bar{\boldsymbol{R}} \triangleq [\boldsymbol{R}_1 \ \cdots \ \boldsymbol{R}_P] = \boldsymbol{H}_0^{(s)} \bar{\boldsymbol{X}} + \boldsymbol{\Xi}_0, \tag{4.22}$$

where $\bar{\boldsymbol{X}} \triangleq [\boldsymbol{X}_1 \ \cdots \ \boldsymbol{X}_P]$, and $\boldsymbol{\Xi}_0$ contains the UL, FFSI, and noise terms. Then, the LS estimate for $\boldsymbol{H}_0^{(s)}$ is written as

$$\hat{\boldsymbol{H}}_0^{(s)} = \bar{\boldsymbol{R}} \bar{\boldsymbol{X}}^H \left(\bar{\boldsymbol{X}} \bar{\boldsymbol{X}}^H \right)^{-1} = \boldsymbol{H}_0^{(s)} + \boldsymbol{E}_0^{(s)}, \tag{4.23}$$

where $\boldsymbol{E}_0^{(s)} \triangleq \boldsymbol{\Xi}_0 \bar{\boldsymbol{X}}^H \left(\bar{\boldsymbol{X}} \bar{\boldsymbol{X}}^H \right)^{-1}$ is the estimation error. Note that the estimate can also be expressed as

$$\hat{\boldsymbol{H}}_0^{(s)} = \Big(\sum_{p=1}^{P} \boldsymbol{R}_p \boldsymbol{X}_p^H \Big) \Big(\sum_{p=1}^{P} \boldsymbol{X}_p \boldsymbol{X}_p^H \Big)^{-1}, \tag{4.24}$$

which is a block recursive least squares (RLS) filter implementation with a unity forgetting factor. After regeneration and subtraction using $\hat{\boldsymbol{H}}_0^{(s)}$, the residue shown in (4.10) is related to $\boldsymbol{E}_0^{(s)}$. That is, the residual SI power is proportional to the error variance, which will be analyzed to evaluate the performance of NFSI cancellation. First, Appendix C shows that

$$\sum_{p=1}^{P} \boldsymbol{X}_p \boldsymbol{X}_p^H \cong \frac{N_c P}{N_a} \boldsymbol{I}_{N_a}, \tag{4.25}$$

assuming that the space is spanned nearly uniformly by the transmit beamformer $\boldsymbol{W}_p$ during the considered interval of P OFDM symbols. Then, the estimation error matrix in (4.10) is

$$\boldsymbol{E}_0^{(s)} = \frac{N_a}{N_c P} \sum_{p=1}^{P} (\boldsymbol{R}_p^{FFSI} + \boldsymbol{R}_p^{UL} + \boldsymbol{N}_p) \boldsymbol{X}_p^H, \tag{4.26}$$

where $\boldsymbol{R}_p^{FFSI}$ and $\boldsymbol{R}_p^{UL}$ are FFSI and UL terms in $\boldsymbol{R}$ in (4.9) for the p^{th} OFDM symbol. The estimation error is zero-mean, and the FFSI, UL, and noise terms are uncorrelated by the definition of gains $\alpha_k^{(x)}$ of channels. Furthermore, each of these terms is uncorrelated within its own group in different OFDM symbols denoted by p.[4] Therefore, the variance $\sigma_{e,ab}^2$ of an arbitrary entry $(\boldsymbol{E}_0^{(s)})_{ab}$ in the error matrix can be written as the sum of individual variances as

$$\sigma_{e,ab}^2 = \frac{N_a^2}{N_c^2 P^2} \sum_{p=1}^{P} (\sigma_{FFSI,ab}^2 + \sigma_{UL,ab}^2 + \sigma_{N,ab}^2), \tag{4.27}$$

where $\sigma_{FFSI,ab}^2$ is the variance of $(\boldsymbol{R}_p^{FFSI} \boldsymbol{X}_p^H)_{ab}$, and $\sigma_{UL,ab}^2$ and $\sigma_{N,ab}^2$ are defined similarly. These variances are independent of both P and p, yielding[5]

$$\sigma_{e,ab}^2 = \frac{N_a^2}{N_c^2 P} \left(\sigma_{FFSI,ab}^2 + \sigma_{UL,ab}^2 + \sigma_{N,ab}^2 \right), \tag{4.28}$$

which shows the time-integration property that constitutes

$$\sigma_{e,ab}^2 \propto P^{-1}. \tag{4.29}$$

This time-integration property can be exploited to the fullest extent since the NFSI channel is time-invariant, leaving adaptiveness to FFSI variations to the second stage. The learning might continue indefinitely with $P \to \infty$, even during data communication with the suitable iterative structure given in (4.24). Note that P can reach hundreds of thousands in a second according to subcarrier spacings considered for

[4] This is guaranteed for UL and noise terms due to mismatched regression. For FFSI, channel correlation is directly reflected in the related term in the error variance. For slightly varying FFSI channels, P^{-1} fading rate in (4.29) is still valid, with a scaling in the related term in (4.28) by coherence time in terms of OFDM symbols. However, completely stationary FFSI components would act as NFSI and leak into the NFSI channel estimate, weakened by delay domain separation according to $\sin^2(\pi l_{k_s})/\sin^2(\frac{\pi}{N_c} l_{k_s})$. After NFSI cancellation, these leakages create weak FFSI ghosts in zero delay. They can be suppressed either in FFSI mitigation or can be included in NFSI cancellation by a multi-delay SIC design, which is out of scope of this chapter.

[5] The variances are proportional to the related powers $P_k^{(s)}$, $P_k^{(u)}$ and N_0, and they might depend on N_a, N_c, K_d, K_u, but not p and P.

mmWave frequencies. Therefore, assuming that the NFSI is suppressed to a negligible level, the signal model

$$\boldsymbol{Z} = \boldsymbol{G}^{(u)}\boldsymbol{D}^{(u)} + \sum_{k=1}^{K_s} \boldsymbol{G}_k^{(s)}\boldsymbol{D}^{(d)}\boldsymbol{E}_{l_k^{(s)}} + \boldsymbol{N}^{(z)}, \tag{4.30}$$

will be adopted in order to focus on the mitigation techniques for time-varying FFSI components, where the main contributions of this chapter lie.

This NFSI cancellation procedure might be extended to cover numerous factors, such as nonlinearity and time dispersion, as in Chapter 3, thanks to the modular structure provided by time-variance-based separation of NFSI and FFSI mitigation. In any case, the time-integration property remains solid, and it is useful only with the following FFSI mitigation stage within the proposed two-stage mitigation scheme.

4.2.2 FFSI Mitigation Methods

For the time-varying FFSI components, concurrent application of multiple methods is considered based on scatterer properties. For each scatterer, a method is selected among Rx BF, Tx BF, and DSIC. Let $\mathcal{S}_{Rx}$, $\mathcal{S}_{Tx}$, $\mathcal{S}_{DSIC}$ be the sets of scatterers for these methods, respectively. Also, some scatterers require no action, represented by $\mathcal{S}_{NA}$. These sets disjointly partition the set of all scatterers $\mathcal{S}$, expressed as

$$\bigcup_{a\in\mathcal{A}} \mathcal{S}_a = \mathcal{S}, \qquad\qquad \mathcal{S}_a \cap \mathcal{S}_{a'} = \varnothing, \tag{4.31}$$

for $a, a' \in \mathcal{A}$, $a \neq a'$, where $\mathcal{A} \triangleq \{Tx, Rx, DSIC, NA\}$ is the set of actions.

4.2.2.1 Spatial SI Suppression

This approach aims to suppress the FFSI by optimizing the beamformers $\boldsymbol{C}$ and $\boldsymbol{W}$ statistically (see Section 2.2.2.1) based on scatterer angles $\theta_k^{(s)}$, utilizing the spatial characteristic of the FFSI channels shown in (4.8). With the sets $\mathcal{S}_{Rx}$ and $\mathcal{S}_{Tx}$, the well-known maximum signal-to-interference-plus-noise ratio (SINR) and the maximum signal-to-leakage-plus-noise ratio (SLNR) beamformer [115] can be used for

the Rx and Tx beamformers, respectively. They are written as

$$\boldsymbol{c}_k = \beta_k^{(c)} \Big(N_0 \boldsymbol{I}_{N_a} + \sum_{k' \in \mathcal{S}_{Rx}} \rho_{k'}^{(s,c)} \boldsymbol{q}_{k'}^{(s)} (\boldsymbol{q}_{k'}^{(s)})^H \Big)^{-1} \boldsymbol{q}_k^{(u)}, \tag{4.32}$$

$$\boldsymbol{w}_k = \beta_k^{(w)} \Big(N_0 \boldsymbol{I}_{N_a} + \sum_{k' \in \mathcal{S}_{Tx}} \rho_{k'}^{(s,w)} \boldsymbol{q}_{k'}^{(s)} (\boldsymbol{q}_{k'}^{(s)})^H \Big)^{-1} \boldsymbol{q}_k^{(d)}, \tag{4.33}$$

where $\beta_k^{(c)}$ and $\beta_k^{(w)}$ normalizes norms of $\boldsymbol{c}_k$ and $\boldsymbol{w}_k$ to unity, respectively. In addition, $\rho_{k'}^{(s,c)} \triangleq P_{k'}^{(s)} \sum_{k''} |(\boldsymbol{q}_{k''}^{(d)})^H \boldsymbol{q}_{k'}^{(s)}|^2$ and $\rho_{k'}^{(s,w)} \triangleq P_{k'}^{(s)} \sum_{k''} |(\boldsymbol{q}_{k''}^{(u)})^H \boldsymbol{q}_{k'}^{(s)}|^2$.

4.2.2.2 Digital SI Cancellation (DSIC)

Channel estimation for DSIC is implemented using the parameter-based processing approach discussed in Section 2.2.1. UL and DL communications are performed using the designed beamformers, which suppress the related scatterers. However, the received UL signal $\boldsymbol{Z}$ is still contaminated by the SI due to the remaining scatterers. Using the pilots on $N_{c,tr} \ll N_c$ subcarriers,[6] the UL channel is estimated jointly[7] with the SI channels for the scatterers in $\mathcal{S}_{DSIC}$. For this, the LS estimator is written as

$$\hat{\bar{\boldsymbol{G}}} \triangleq \Big[\hat{\boldsymbol{G}}^{(u)} \hat{\boldsymbol{G}}_{k_1^{DSIC}}^{(s)} \hat{\boldsymbol{G}}_{k_2^{DSIC}}^{(s)} \dots \Big] = \boldsymbol{Z} \bar{\boldsymbol{D}}^H (\bar{\boldsymbol{D}} \bar{\boldsymbol{D}}^H)^{-1}, \tag{4.34}$$

where k_i^{DSIC} for $i = 1, \dots, |\mathcal{S}_{DSIC}|$ denotes the scatterers in the set $\mathcal{S}_{DSIC}$. In addition, $(K_u + K_d |\mathcal{S}_{DSIC}|) \times N_{c,tr}$ matrix $\bar{\boldsymbol{D}}$ contains the related pilots as

$$\bar{\boldsymbol{D}} = \Big[[\boldsymbol{D}^{(u)}]^T \quad [\boldsymbol{D}^{(d)} \boldsymbol{E}_{l_1}]^T \quad [\boldsymbol{D}^{(d)} \boldsymbol{E}_{l_2}]^T \dots \Big]^T, \tag{4.35}$$

where l_i is the delay for the scatterer k_i^{DSIC}. Note that these delays are slow-time parameters, as discussed in Section 2.1.6. Then, the FFSI is regenerated for all the subcarriers as

$$\hat{\boldsymbol{Z}}_{si} = \sum_{k \in \mathcal{S}_{DSIC}} \hat{\boldsymbol{G}}_k^{(s)} \boldsymbol{D}^{(d)} \boldsymbol{E}_{l_k^{(s)}}, \tag{4.36}$$

to be subtracted from the received signal $\boldsymbol{Z}$ as in (4.14).

[6] Usage of pilots or all subcarriers in training or data modes changes some variables such as $\boldsymbol{Z}$, $\boldsymbol{D}^{(u)}$, $\boldsymbol{D}^{(d)}$, $\boldsymbol{E}_l$, although not indicated for notational simplicity. The distinction will be clear from the context.

[7] Joint estimation outperforms separate estimation and does not require iterations [60].

4.3 Selection Policy for FFSI Mitigation

This section aims to determine a mitigation method for each scatterer by constructing the sets $\mathcal{S}_{Rx}$, $\mathcal{S}_{Tx}$, $\mathcal{S}_{DSIC}$, and $\mathcal{S}_{NA}$. Each scatterer is handled with the best action based on prior performance metrics of communication, which are approximations of signal-to-noise ratio (SNR) for DL and SINR for UL. Deriving these, each scatterer is considered separately with the assumed single-scatterer model

$$\boldsymbol{Z} = \boldsymbol{G}^{(u)}\boldsymbol{D}^{(u)} + \boldsymbol{G}^{(s)}_{k_s}\boldsymbol{D}^{(d)}\boldsymbol{E}_{l^{(s)}_{k_s}} + \boldsymbol{N}, \tag{4.37}$$

for the scatterer k_s, where other scatterers in (4.30) are neglected. Although $\boldsymbol{c}_k$ and $\boldsymbol{w}_k$ will be expressed based on (4.37) in derivation, beamformer implementations remain as described in (4.32) and (4.33). Since the scatterer k_s will be focused, this index is dropped from variables $\boldsymbol{G}^{(s)}_{k_s}$, $\alpha^{(s)}_{k_s}$, $P^{(s)}_{k_s}$, $l^{(s)}_{k_s}$ and $\boldsymbol{q}^{(s)}_{k_s}$ until (4.56) for simplicity. According to (4.37), the UL SINR $\gamma^{(u)}_k$ and DL SNR $\gamma^{(d)}_k$ for the k^{th} UL or DL user are written as[8]

$$\gamma^{(u)}_k = \frac{P^{(u)}_k |\boldsymbol{c}^H_k \boldsymbol{q}^{(u)}_k|^2}{P^{(s)} \sum_{k'} |\boldsymbol{w}^H_{k'} \boldsymbol{q}^{(s)}|^2 |\boldsymbol{c}^H_k \boldsymbol{q}^{(s)}|^2 + N_0}, \qquad k \in \mathcal{U}_u \tag{4.38}$$

$$\gamma^{(d)}_k = \frac{P^{(d)}_k |\boldsymbol{w}^H_k \boldsymbol{q}^{(d)}_k|^2}{N_0}, \qquad k \in \mathcal{U}_d \tag{4.39}$$

where $\boldsymbol{c}^H_k \boldsymbol{c}_k = 1$, $\boldsymbol{w}^H_k \boldsymbol{w}_k = 1$ and $\boldsymbol{T}$ is assumed as $\boldsymbol{I}_{K_d}$. In the next sections, $\gamma^{(u)}_k$ and $\gamma^{(d)}_k$ will be written for each case $a \in \mathcal{A}$, denoted by $\gamma^{(u,a)}_k$ and $\gamma^{(d,a)}_k$. Note that these metrics show the effects of a single scatterer, and they will be calculated for each scatterer in Section 4.3.4.

4.3.1 No-SI and No-Action Cases

In these cases, there are no mitigation efforts with conventional Rx and Tx beamformers $\boldsymbol{c}_k = \boldsymbol{q}^{(u)}_k$ and $\boldsymbol{w}_k = \boldsymbol{q}^{(d)}_k$. For the no-SI case, the metrics are

$$\gamma^{(u,0)}_k = \frac{P^{(u)}_k}{N_0}, \qquad \gamma^{(d,0)}_k = \frac{P^{(d)}_k}{N_0}, \tag{4.40}$$

[8] The multi-user interference is ignored because it will be handled by the multi-user equalizer after FFSI mitigation, as shown in Fig. 4.1.

respectively. The no-action case is different only in the presence of the unattended FFSI, with metrics

$$\gamma_k^{(u,NA)} = \frac{\gamma_k^{(u,0)}}{1 + \frac{P^{(s)}}{N_0} \sum_{k'=1}^{K_d} Q_{k'}^{(ds)} Q_k^{(us)}}, \tag{4.41}$$

and $\gamma_k^{(d,NA)} = \gamma_k^{(d,0)}$, where $Q_k^{(ds)}$ and $Q_k^{(us)}$ govern the effects of user/scatterer positioning, defined as

$$Q_k^{(ds)} \triangleq |(\boldsymbol{q}_k^{(d)})^H \boldsymbol{q}^{(s)}|^2, \qquad Q_k^{(us)} \triangleq |(\boldsymbol{q}_k^{(u)})^H \boldsymbol{q}^{(s)}|^2. \tag{4.42}$$

4.3.2 Spatial Suppression via Rx or Tx Beamforming

In the Rx BF case, (4.32) is written for a single scatterer as

$$\boldsymbol{c}_k = \beta_k^{(c)} \left(N_0 \boldsymbol{I} + \rho_s \boldsymbol{q}^{(s)} (\boldsymbol{q}^{(s)})^H \right)^{-1} \boldsymbol{q}_k^{(u)}, \tag{4.43}$$

where $\rho_s \triangleq P^{(s)} \sum_{k'} |(\boldsymbol{q}_{k'}^{(d)})^H \boldsymbol{q}^{(s)}|^2$, and the Tx beamformer is $\boldsymbol{w}_k = \boldsymbol{q}_k^{(d)}$. Then, the DL SNR is simply $\gamma_k^{(d,Rx)} = \gamma_k^{(d,0)}$ from (4.39), and the UL SINR can be calculated from (4.38). For a closed-form expression, $\boldsymbol{c}_k$ can be rewritten using Woodbury's identity as

$$\boldsymbol{c}_k = \beta_k^{\prime(c)} \left(\boldsymbol{q}_k^{(u)} - \frac{(\boldsymbol{q}^{(s)})^H \boldsymbol{q}_k^{(u)}}{1 + N_0/\rho_s} \boldsymbol{q}^{(s)} \right), \tag{4.44}$$

$$\cong \beta_k^{\prime(c)} \left(\boldsymbol{q}_k^{(u)} - [(\boldsymbol{q}^{(s)})^H \boldsymbol{q}_k^{(u)}] \boldsymbol{q}^{(s)} \right), \tag{4.45}$$

where $\beta_k^{\prime(c)} \triangleq \frac{\beta_k^{(c)}}{N_0}$, and (4.45) implements a high INR assumption ($\rho_s/N_0 \to \infty$). Then, $||\boldsymbol{c}_k||^2 = 1$ is satisfied by $\beta_k^{\prime(c)} = (1 - Q_k^{(us)})^{-\frac{1}{2}}$, which yields $|\boldsymbol{c}_k^H \boldsymbol{q}_k^{(u)}|^2 = 1 - Q_k^{(us)}$ and $|\boldsymbol{c}_k^H \boldsymbol{q}^{(s)}|^2 = 0$ for $\gamma_k^{(u,Rx)}$ in (4.38). Therefore, the metrics are calculated as

$$\gamma_k^{(u,Rx)} = \gamma_k^{(u,0)} (1 - Q_k^{(us)}), \qquad \gamma_k^{(d,Rx)} = \gamma_k^{(d,0)}. \tag{4.46}$$

For the Tx BF case, the Rx beamformer is $\boldsymbol{c}_k = \boldsymbol{q}_k^{(u)}$, and the Tx beamformer $\boldsymbol{w}_k$ in (4.33) is modified. Following the similar procedures, the metrics can be found as

$$\gamma_k^{(u,Tx)} = \gamma_k^{(u,0)}, \qquad \gamma_k^{(d,Tx)} = \gamma_k^{(d,0)} (1 - Q_k^{(ds)}). \tag{4.47}$$

4.3.3 Digital SI Cancellation (DSIC)

For the DSIC metric, the conventional beamformers $\boldsymbol{c}_k = \boldsymbol{q}_k^{(u)}$ and $\boldsymbol{w}_k = \boldsymbol{q}_k^{(d)}$ are assumed, which directly sets the DL SNR as $\gamma_k^{(d,DSIC)} = \gamma_k^{(d,0)}$. Then, the assumed form of $\boldsymbol{Z}$ in (4.37) can also be written as

$$\boldsymbol{Z} = \left[\boldsymbol{G}^{(u)}\boldsymbol{G}^{(s)}\right]\bar{\boldsymbol{D}} + \boldsymbol{N}, \tag{4.48}$$

where $\bar{\boldsymbol{D}} = \left[[\boldsymbol{D}^{(u)}]^T \quad [\boldsymbol{D}^{(d)}\boldsymbol{E}_l]^T\right]^T$. The LS estimates for channel matrices are given by $\boldsymbol{Z}\bar{\boldsymbol{D}}^H(\bar{\boldsymbol{D}}\bar{\boldsymbol{D}}^H)^{-1}$, which yields

$$\left[\hat{\boldsymbol{G}}^{(u)}\hat{\boldsymbol{G}}^{(s)}\right] = \left[\boldsymbol{G}^{(u)}\boldsymbol{G}^{(s)}\right] + \boldsymbol{N}\bar{\boldsymbol{D}}^H(\bar{\boldsymbol{D}}\bar{\boldsymbol{D}}^H)^{-1}. \tag{4.49}$$

The matrix $\bar{\boldsymbol{D}}$ is of size $(K_u + K_d) \times N_{c,tr}$, which carries training sequences for UL and DL users, each of length $N_{c,tr}$. Each information symbol is unit-power, and the sequences are linearly independent. With the further assumption that they are orthogonal, $\bar{\boldsymbol{D}}\bar{\boldsymbol{D}}^H$ becomes $N_{c,tr}\boldsymbol{I}_{K_u+K_d}$, and the error term becomes $N_{c,tr}^{-1}\boldsymbol{N}\bar{\boldsymbol{D}}^H$. Focusing on the related matrix block, the error for the SI channel estimate is found as

$$\boldsymbol{\Delta}_s \triangleq \hat{\boldsymbol{G}}^{(s)} - \boldsymbol{G}^{(s)} = \frac{1}{N_{c,tr}}\boldsymbol{N}\left[\boldsymbol{D}^{(d)}\boldsymbol{E}_l\right]^H. \tag{4.50}$$

Columns of $\boldsymbol{N}$ are independent and identically distributed (i.i.d.) with $\mathcal{CN}(\boldsymbol{0}, N_0\boldsymbol{C}^H\boldsymbol{C})$. Since the DL pilot sequences are assumed orthogonal, the error matrix $\boldsymbol{\Delta}_s$ has i.i.d. columns and each column follows $\mathcal{CN}(\boldsymbol{0}, \frac{N_0}{N_{c,tr}}\boldsymbol{C}^H\boldsymbol{C})$. When SI is re-generated as $\hat{\boldsymbol{Z}}_{si} = \hat{\boldsymbol{G}}^{(s)}\boldsymbol{D}^{(d)}\boldsymbol{E}_l$ and subtracted, it yields

$$\boldsymbol{Z} - \hat{\boldsymbol{Z}}_{si} = \boldsymbol{G}^{(u)}\boldsymbol{D}^{(u)} - \boldsymbol{\Delta}_s\boldsymbol{D}^{(d)}\boldsymbol{E}_l + \boldsymbol{N}, \tag{4.51}$$

which is different from (4.37) only in the SI term. Hence, the SI power term in (4.41) should be modified. The new term, the residual SI with power P_{SI}^{DSIC}, can be expressed for the UL user k_u and the subcarrier m as

$$(\boldsymbol{\Delta}_s\boldsymbol{D}^{(d)}\boldsymbol{E}_l)_{k_u,m} = \sum_{k_d=1}^{K_d}(\boldsymbol{\Delta}_s)_{k_u,k_d}\, d_{k_d,m}^{(d)}\, e^{-j\frac{2\pi}{N_{c,tr}}ml}. \tag{4.52}$$

Then, the variance of this term gives P_{SI}^{DSIC} as

$$P_{SI}^{DSIC} = \sum_{k_d=1}^{K_d}\left(\frac{N_0}{N_{c,tr}}\boldsymbol{C}^H\boldsymbol{C}\right)_{k_u,k_u}|d_{k_d,m}^{(d)}e^{-j\frac{2\pi}{N_{c,tr}}ml}|^2, \tag{4.53}$$

$$= \sum_{k_d=1}^{K_d}\frac{N_0}{N_{c,tr}}\boldsymbol{c}_{k_u}^H\boldsymbol{c}_{k_u}|d_{k_d,m}^{(d)}|^2 = \frac{K_dN_0}{N_{c,tr}}, \tag{4.54}$$

using the properties of $\mathbf{\Delta}_s$ described under (4.50), and assumed unit average energy for pilots. Finally, the UL SINR is written similarly to (4.41) as

$$\gamma_k^{(u,DSIC)} = \frac{\gamma_k^{(u,0)}}{1 + \frac{K_d}{N_{c,tr}}}. \tag{4.55}$$

4.3.4 Determining Actions for Scatterers Based on the Metrics

The approximated UL SINRs and DL SNRs are tabulated in Table 4.1. When focused on a single scatterer $k_s \in \mathcal{S}$ and a single action $a \in \mathcal{A}$, there is a metric for each user, such as $\gamma_1^{(u,a)}, \ldots, \gamma_{K_u}^{(u,a)}, \gamma_1^{(d,a)}, \ldots, \gamma_{K_d}^{(d,a)}$. Their arithmetic mean gives a single metric for this scatterer/action pair as

$$\bar{\gamma}_{k_s}^{(a)} = \frac{1}{K_u + K_d} \left(\sum_{k=1}^{K_u} \frac{\gamma_k^{(u,a)}}{\gamma_k^{(u,0)}} + \sum_{k=1}^{K_d} \frac{\gamma_k^{(d,a)}}{\gamma_k^{(d,0)}} \right), \tag{4.56}$$

where the scatterer index k_s is reinstated, which was dropped after (4.37) for simplicity. To determine the actions, this average performance metric is calculated for each scatterer/action pair. Then, each scatterer is associated with the best action, as detailed in Algorithm 2.

4.4 Slow-Time and Fast-Time Operations

The proposed SI mitigation scheme consists of spatial SI suppression and DSIC with parameter-based designs, as discussed in Section 2.2.1. These slowly-changing parameters of scatterers are delay $l_k^{(s)}$ for DSIC, and angle $\theta_k^{(s)}$ and power $P_k^{(s)}$ for spatial suppression, stored in the *scatterer map* shown in Fig. 4.1. DSIC also depends

Table 4.1: Prior Performance Metrics

User\Action	**No Action**	**Rx BF**	**Tx BF**	**DSIC**
UL user k_u $\left(\gamma_{k_u}^{(u,a)}/\gamma_{k_u}^{(u,0)}\right)$	$\frac{1}{1 + \frac{P^{(s)}}{N_0} \sum_{k_d} Q_{k_d}^{(ds)} Q_{k_u}^{(us)}}$	$1 - Q_{k_u}^{(us)}$	1	$\frac{1}{1 + \frac{K_d}{N_{c,tr}}}$
DL user k_d $\left(\gamma_{k_d}^{(d,a)}/\gamma_{k_d}^{(d,0)}\right)$	1	1	$1 - Q_{k_d}^{(ds)}$	1

Algorithm 2 Construction of Action Sets.

Input: $\theta_{k_u}^{(u)}$, $\theta_{k_d}^{(d)}$, $\theta_{k_s}^{(s)}$, $P_{k_s}^{(s)}$ for $k_u \in \mathcal{U}_u$, $k_d \in \mathcal{U}_d$, $k_s \in \mathcal{S}$, and $\mathcal{A}$.

1: $\mathcal{S}_a = \varnothing$ for $a \in \mathcal{A}$.
2: **for** $k_s = 1$ to K_s **do**
3: set $\boldsymbol{q}^{(s)} = \boldsymbol{q}_{k_s}^{(s)}$ and $\boldsymbol{P}^{(s)} = \boldsymbol{P}_{k_s}^{(s)}$.
4: calculate $Q_{k_u}^{(us)}$, $Q_{k_d}^{(ds)}$ for $k_u \in \mathcal{U}_u$, $k_d \in \mathcal{U}_d$. (Eq. (4.42))
5: calculate $\bar{\gamma}_{k_s}^{(a)}$ for $a \in \mathcal{A}$. (Eq. (4.56) and Table 4.1)
6: $a^* = \arg\max_{a \in \mathcal{A}} \bar{\gamma}_{k_s}^{(a)}$
7: $\mathcal{S}_{a^*} \leftarrow \mathcal{S}_{a^*} \cup \{k_s\}$
8: **end for**

Output: $\mathcal{S}_a$ for $a \in \mathcal{A}$.

on fast-time channel parameters, $\alpha_k^{(x)}$, which is a disadvantage compared to spatial suppression in terms of complexity. The parameter-based design removes the need to estimate the SI channel for all subcarriers (or delays) for DSIC. Also, it enables angle-based beamformer design, rather than one based on instantaneous channel estimates, which provides advantages both in performance and complexity.

Delay, angle, and power are called slow-time parameters, in comparison to Doppler-related fast-time parameters represented by gains $\alpha_k^{(x)}$. Their slowly varying characteristics introduce a separate and longer coherent processing interval (CPI), valid only for the systems that are related exclusively to these parameters [116]. This longer coherence time is shown to be on the order of seconds in [73], i.e., thousands of mmWave OFDM symbol durations. Hence, a division of operations into slow-time and fast-time is enabled. Slow-time operations are those creating (or maintaining) and utilizing the parameters in the scatterer map, namely (i) slow-time sensing, (ii) selection policy, and (iii) calculation of beamforming weights. Fast-time operations are all the other communication-related operations, DSIC, and also detection-triggered sensing for emerging scatterers. While fast-time operations are performed for each OFDM symbol, slow-time operations are performed rarely, as shown in Fig. 4.3, once in thousands of OFDM symbol durations.

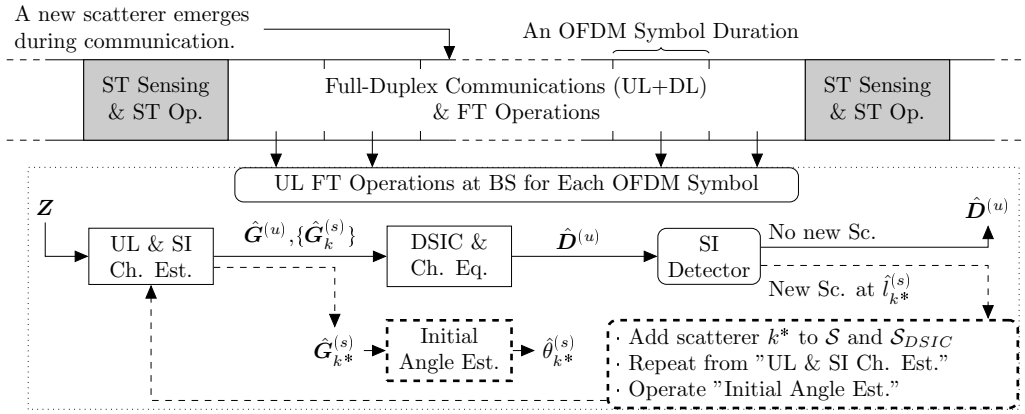


Figure 4.3: Regular FD communication and slow-time sensing intervals, complemented by detection-triggered fast-time sensing for emerging scatterers using UL communication signals within fast-time Rx processing flow. (Detection-triggered operations are dashed. ST: Slow-Time, FT: Fast-Time, Op.: Operations, Sc.: Scatterer.)

4.4.1 Sensing for Scatterer Map

The parameter-based processing proposed for FD communications needs sensing for up-to-date scatterer map information, which can be categorized as *sensing for communication*. Since a time-shared form of ISAC is considered, as in Fig. 4.3, sensing requirements of the proposal are very generic. Therefore, the methods proposed in the ISAC literature are generally suitable for this task. In addition, methods for slow-time sensing are studied also in Chapter 5, which reduces the time spent for angular estimation by utilizing the channel estimates collected during communications. The contribution of this chapter will be to provide useful prior information to slow-time sensing procedures and simplify them.

Possible changes in the scatterer map are emergence, disappearance, and gradual variations. Also, significant variations can be regarded as disappearance and emergence at neighboring locations. Sensing workload for disappearance and variation is easier thanks to the available prior information. Scanning the whole angle/delay space is unnecessary, since verification beams and delay-specific regression according to the prior information would be enough for sensing. In terms of FD communications, disappearance would naturally enhance the performance, and variations would have a limited effect with an optimized repetition rate for sensing.

On the other hand, emerging scatterers are the heaviest burden on sensing mechanisms, as they require scanning of the entire angle/delay space. Also, they significantly degrade FD communication performance as unattended SI sources, which can cause packet losses. For emerging scatterers, this work proposes a detection, mitigation, and parameter estimation scheme. In addition to recovering performance and preventing packet losses, it will reduce the sensing workload from an exhaustive search to verification and refinement of the estimation results.

Therefore, this work considers (i) a slow-time sensing for the verification and tracking of the prior information using specific sensing beams within a negligibly small time interval, and (ii) a detection-triggered fast-time sensing for emerging scatterers and significant variations, detailed in the next subsection.

4.4.2 Detection-Triggered Fast-Time Sensing for Emerging Scatterers

To formulate the detection problem, hypotheses $H_1(l)$ and $H_0(l)$ describe the presence and absence of an emerging scatterer at delay l, respectively. After the SI mitigation for known scatterers, the information symbol estimates in (4.15) can be rewritten for these cases as

$$H_0(l): \quad \hat{\boldsymbol{D}}^{(u)} = \boldsymbol{D}^{(u)} + \tilde{\boldsymbol{N}}, \tag{4.57}$$

$$H_1(l): \quad \hat{\boldsymbol{D}}^{(u)} = \boldsymbol{D}^{(u)} + \boldsymbol{M}^{(s)}\boldsymbol{D}^{(d)}\boldsymbol{E}_{l^*} + \tilde{\boldsymbol{N}}, \tag{4.58}$$

where $\tilde{\boldsymbol{N}} \triangleq (\boldsymbol{M}^{(u)} - \boldsymbol{I})\boldsymbol{D}^{(u)} + \boldsymbol{\Xi}^{(u)}$ contains the effects of UL channel estimation error, residual SI (from known scatterers), and the thermal noise. In addition, k^* and l^* are the index and delay of the emerging scatterer, respectively, and $\boldsymbol{M}^{(s)}$ is the modified version of the new SI channel $\boldsymbol{G}^{(s)}_{k^*}$ seen after the multi-user equalizer, which is written as

$$\boldsymbol{M}^{(s)} \triangleq \left((\hat{\boldsymbol{G}}^{(u)})^H \hat{\boldsymbol{G}}^{(u)}\right)^{-1} (\hat{\boldsymbol{G}}^{(u)})^H \boldsymbol{G}^{(s)}_{k^*}. \tag{4.59}$$

4.4.2.1 Detection and Delay Estimation

Observing hypotheses (4.57) and (4.58), a detector can be designed based on an estimate of $\boldsymbol{M}^{(s)}$. It seems suitable to process $\hat{\boldsymbol{D}}^{(u)} - \boldsymbol{D}^{(u)}$ for this estimate. The r^{th} row

of that matrix is written as

$$(\hat{\boldsymbol{D}}^{(u)} - \boldsymbol{D}^{(u)})_{r,:} = \sum_{t=1}^{K_d} (\boldsymbol{M}^{(s)})_{r,t} (\boldsymbol{D}^{(d)})_{t,:} \boldsymbol{E}_l + (\tilde{\boldsymbol{N}})_{r,:}. \tag{4.60}$$

Among the estimators investigated in the published work [60] for similar purposes, it is seen that the suboptimal *per-user SI channel estimator* performs similarly with the other complex estimators in detection, although it fails in SI regeneration. This simple estimator is adapted for this study since complexity might be an issue for this task. It estimates the entries of $\boldsymbol{M}^{(s)}$ separately as

$$\left(\hat{\boldsymbol{M}}_l^{(s)}\right)_{r,t} = \text{IFFT}\{N_{c,tr}^{-\frac{1}{2}} (\hat{\boldsymbol{D}}^{(u)} - \boldsymbol{D}^{(u)})_{r,:} \oslash (\boldsymbol{D}^{(d)})_{t,:}\}_l, \tag{4.61}$$

for $r = 1, \ldots, K_u$ and $t = 1, \ldots, K_d$, where $\oslash$ is elementwise division. Candidate estimates $\hat{\boldsymbol{M}}_l^{(s)}$ are obtained for each delay $l \in \mathcal{L}$, where $\mathcal{L}$ is the set of possible delays. Then, the Frobenius norm squares of these matrices are tested as

$$\rho_l \underset{H_0(l)}{\overset{H_1(l)}{\gtrless}} \alpha, \qquad \rho_l \triangleq ||\hat{\boldsymbol{M}}_l^{(s)}||_F^2, \tag{4.62}$$

which decides $H_1(l)$ or $H_0(l)$ for each $l \in \mathcal{L}$. The threshold α is determined by the constant false alarm rate (CFAR) method [117], which fixes the false alarm probability $P_{FA} \triangleq P(\rho_l > \alpha \,|H_0(l))$ to a desired value. However, the cumulative distribution function (CDF) of ρ_l given $H_0(l)$ is needed, which is unknown due to the complex structure of $\tilde{\boldsymbol{N}}$.

Alternatively, a *cell-averaging* CFAR test [117] requires only the type of the distribution, where the unknown variance is empirically calculated from the metric ρ_l for $l \in \mathcal{L}$. Assuming each entry of $\hat{\boldsymbol{M}}_l^{(s)}$ is i.i.d. and follows $\mathcal{CN}(0, \sigma^2)$ given $H_0(l)$ with an unknown σ^2, the test is written as [60]

$$\rho_l \underset{H_0(l)}{\overset{H_1(l)}{\gtrless}} \frac{g(P_{FA})}{2K_u K_d} \left(\frac{1}{(|\mathcal{L}| - 1)} \sum_{l' \in \mathcal{L}, l' \neq l} \rho_{l'} \right), \tag{4.63}$$

where $g(P_{FA})$ sets the CDF of χ^2 distribution with $2K_u K_d$ degrees of freedom to $1 - P_{FA}$. The scalings convert the mean of this distribution, $2K_u K_d$, to the empirical mean.

4.4.2.2 Repetition of Channel Estimation and DSIC

In the case of detection, the system immediately mitigates the related SI by DSIC. The detected scatterer k^* at delay l^* is added to the DSIC set $\mathcal{S}_{DSIC}$. The channel estimation and DSIC operations are repeated with the new set as in Section 4.2.2.2. With this repetition, besides having the new SI channel estimate $\hat{\boldsymbol{G}}_{k*}^{(s)}$, previously estimated channels of users and scatterers are refined. Therefore, the multi-user channel equalization is also repeated as in Section 4.1.4 with the new UL channel estimate. After mitigation via DSIC, the cleared observation passes through the detector once more, enabling detection of multiple scatterers. As shown with the loop structure in Fig. 4.3, iterations continue until there are no new detections.

4.4.2.3 Initial Angle Estimation

Having the SI channel estimate $\hat{\boldsymbol{G}}_{k*}^{(s)}$ for the emerging scatterer from the previous step, its angle can be estimated too. Using the general definition in (4.8), the SI channel can be rewritten as $\text{vec}\{\boldsymbol{G}_k^{(s)}\} = \alpha_k^{(s)}\boldsymbol{v}(\theta_k^{(s)})$, where $\text{vec}\{\cdot\}$ is the vectorization operator and $\boldsymbol{v}(\theta) \triangleq \text{vec}\{\boldsymbol{C}^H\boldsymbol{a}(\theta)\boldsymbol{a}^H(\theta)\boldsymbol{W}\}$. The *nonlinear LS* estimator [114] can be written for $\theta_{k*}^{(s)}$ using the estimate $\hat{\boldsymbol{G}}_{k*}^{(s)}$ as [60]

$$\hat{\theta}_{k*}^{(s)} = \arg\min_{\theta} ||\text{vec}\{\hat{\boldsymbol{G}}_{k*}^{(s)}\} - \frac{\boldsymbol{v}(\theta)\boldsymbol{v}^H(\theta)}{\boldsymbol{v}^H(\theta)\boldsymbol{v}(\theta)}\text{vec}\{\hat{\boldsymbol{G}}_{k*}^{(s)}\}||^2, \tag{4.64}$$

which substitutes the LS estimate for $\alpha_k^{(s)}$ into the maximum likelihood (ML) metric $||\text{vec}\{\hat{\boldsymbol{G}}_{k*}^{(s)}\} - \alpha_k^{(s)}\boldsymbol{v}(\theta)||^2$. This angular estimate can be directly used, or it can be taken as primary information to be verified or refined. In this way, the scatterer map is updated for the new scatterer with its delay $l_{k*}^{(s)}$ and angle $\theta_{k*}^{(s)}$. Using these, the power $P_{k*}^{(s)}$ can also be estimated through a longer observation.

4.5 Complexity Analysis

The required number of multiplications for various UL side BS operations are presented in Table 4.2 using the $\mathcal{O}$ notation, where generally $N_c \gg N_{c,tr} > |\mathcal{S}_{DSIC}| \cong$

K_u. The DL side operations in BS are fewer as shown in Fig. 4.1, and their complexities are similar to their duals present in the UL side.

As explained in Section 4.4, determining actions via the proposed policy (metric calculation and the algorithm), calculation of beamforming weights for spatial suppression, and general sensing procedures are slow-time operations. These operations are performed once in thousands of OFDM symbol durations, and their complexities per OFDM symbol are negligible. Also, the complexity of the initial angular estimator is negligible since it works only if a detection occurs.

Inspecting Table 4.2, the numbers of scatterers associated with beamforming, $|\mathcal{S}_{Tx}|$ and $|\mathcal{S}_{Rx}|$, do not have any impact for fully digital beamformers. However, the number of scatterers for DSIC, $|\mathcal{S}_{DSIC}|$, increases the complexity. It is a key difference between DSIC and spatial suppression for complexity originating from their requirements. DSIC depends on the transmitted signal and fast-time channel parameters, represented by $\alpha_k^{(x)}$, while beamforming depends only on slowly-varying angles, as also discussed in Section 2.1.6.

The SI detector also adds to the complexity. This operation provides awareness, prevents packet losses, and eases the slow-time sensing procedures with prior information. However, emergence is a low-probability event. Hence, the complexity can be optimized by adjusting the number of utilized subcarriers. For example, overall detection and false-alarm probabilities can still be achieved with fewer subcarriers by using decision-based integration over OFDM symbols, i.e., the "m out of n" method, in exchange for a symbol-level latency.

Table 4.2: Complexities of Operations per OFDM Symbol

Operation	**# Multiplications**
Beamforming	$\mathcal{O}\left(K_u N_a N_c\right)$
FFT/IFFT	$\mathcal{O}\left(K_u N_c \log N_c\right)$
Channel Estimation	$\mathcal{O}\left((K_u + K_d\lvert\mathcal{S}_{DSIC}\rvert)^2 N_{c,tr}\right)$
DSIC	$\mathcal{O}\left(K_u K_d \lvert\mathcal{S}_{DSIC}\rvert N_c\right)$
Multi-user Equalization	$\mathcal{O}\left(K_u^2 N_c\right)$
SI Detector	$\mathcal{O}(K_u K_d N_{c,tr}^2 \log N_{c,tr})$

Compared to the state-of-the-art, the proposed policy provides an educated method to choose the set of scatterers for each action. In terms of complexity, the advantage is having a reduced number of scatterers ($|\mathcal{S}_{DSIC}|$) for DSIC, rather than all the scatterers (K_s) as in studies that implement only DSIC. Also, the usage of the scatterer map enables a parameter-based processing (angle & delay), significantly reducing the complexity and providing extra gain as discussed in Section 2.2.1. Otherwise, the channel should have been estimated for all subcarriers, and the beamformer weights should have been renewed for each OFDM symbol based on channel estimates.

4.6 Performance Evaluation

For the simulations, considered system parameters are $N_a = 32$ antenna elements both in Tx and Rx arrays of BS, $N_{c,tr} = 64$ subcarriers for training, $P_k^{(x)}/N_0 = 30$ dB input SNR for $x \in \{u, d\}$, and $P_k^{(s)}/N_0 = 46$ dB maximum INR.[9] For each Monte-Carlo trial in the simulations, the UL SINR $\Gamma_k^{(u)}$ and DL SNR $\Gamma_k^{(d)}$ are measured separately for each user as

$$\Gamma_k^{(x)} = \frac{||m_{kk}^{(x)} \boldsymbol{d}_k^{(x)}||^2}{||\hat{\boldsymbol{d}}_k^{(x)} - m_{kk}^{(x)} \boldsymbol{d}_k^{(x)}||^2}, \tag{4.65}$$

for $x \in \{u, d\}$. The worst-case UL SINR $\Gamma^{(u,w)}$ and the worst-case DL SNR $\Gamma^{(d,w)}$ are defined as

$$\Gamma^{(u,w)} = \min_{k \in \mathcal{U}_u} \mathbb{E}\{\Gamma_k^{(u)}\}, \qquad \Gamma^{(d,w)} = \min_{k \in \mathcal{U}_d} \mathbb{E}\{\Gamma_k^{(d)}\}, \tag{4.66}$$

which show the worst performance among multiple UL and DL users. The achievable information rate (AIR) is

$$I_k^{(u)} = \mathbb{E}\left\{\log_2\left(1 + \Gamma_k^{(u)}\right)\right\}, \quad I_k^{(d)} = \mathbb{E}\left\{\log_2\left(1 + \Gamma_k^{(d)}\right)\right\}. \tag{4.67}$$

In addition, the LS channel estimator in (4.34) will be compared with the block RLS estimator in some figures, to discuss the impact of adaptivity. For RLS, (4.34) is modified to

$$\hat{\bar{\boldsymbol{G}}} = (\sum_{p=1}^{P} \lambda^{P-p} \boldsymbol{Z}_p \bar{\boldsymbol{D}}_p^H)(\sum_{p=1}^{P} \lambda^{P-p} \bar{\boldsymbol{D}}_p \bar{\boldsymbol{D}}_p^H)^{-1}, \tag{4.68}$$

[9] These ratios are found by utilizing (4.21) for $P_{BS} = 37$ dBm, $P_{UE} = 23$ dBm, $d_u = 100$ m, $d_s = 20$ m, $\sigma = 10$ m^2 RCS (a medium value between automobiles and pedestrians), $\lambda = 1$ cm and 100 MHz bandwidth.

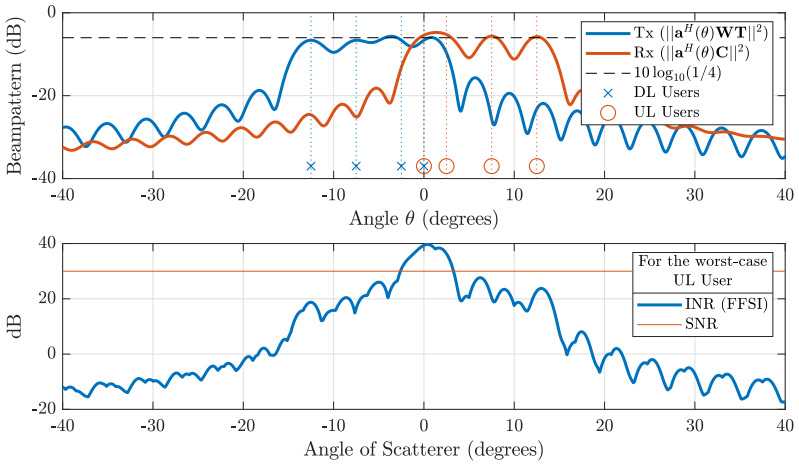


Figure 4.4: Beampatterns and exposed far-field SI power. ($K_d = 4$, $K_u = 4$, $K_s = 1$. User positions are as shown with markers.)

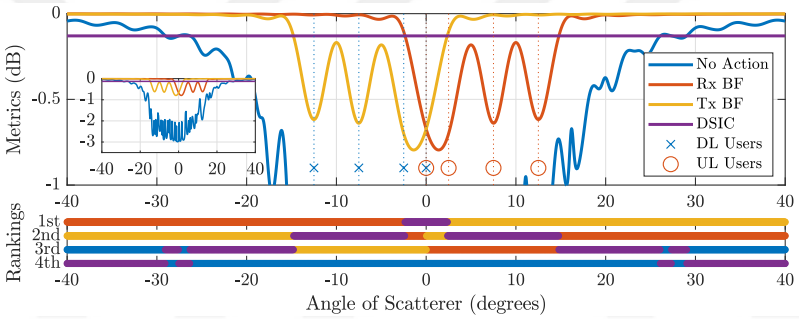


Figure 4.5: Calculated action metrics and color-coded rankings of actions according to the angle of the considered scatterer. ($K_d = 4$, $K_u = 4$, $K_s = 1$. User positions are as shown with markers.)

where p is the OFDM symbol index, P is the current OFDM symbol, and the forgetting factor $\lambda = 0.9$. A first-order autoregressive (AR(1)) correlation [58] is implemented on all the channels via their complex gains $\alpha_k^{(x)}$. The correlation values are the same for all channels.

4.6.1 Initial Insights for the Single-Scatterer Case

The beampatterns and exposed far-field SI power are plotted in Fig. 4.4. As seen, due to the combined effect of the Tx and Rx beampatterns, the exposed FFSI power depends on the angles of the users and the scatterer. For example, the FFSI might be

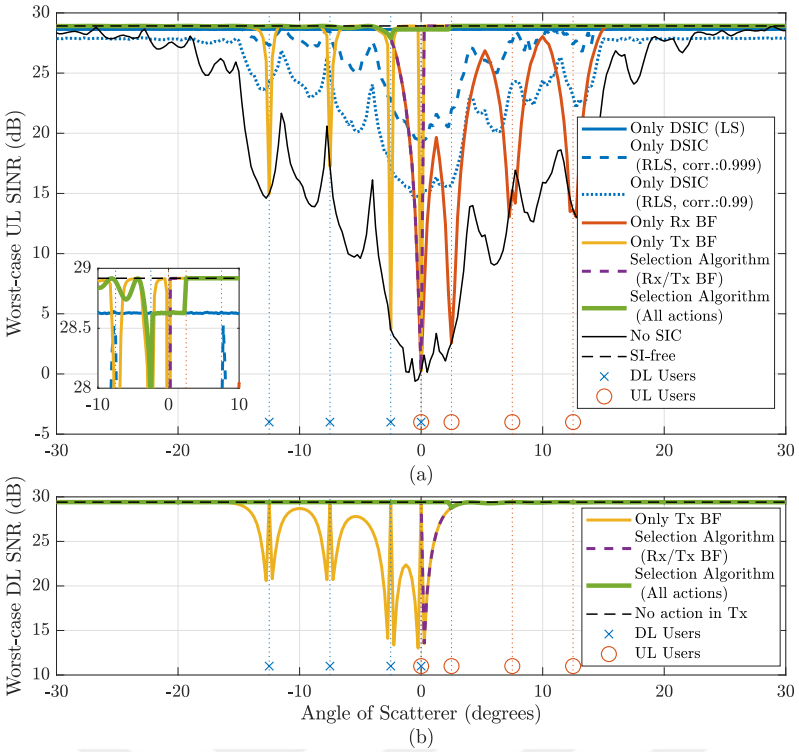


Figure 4.6: Worst-case UL SINR and DL SNR against the angle of a single scatterer. ($K_d = 4$, $K_u = 4$, $K_s = 1$. Markers show the user positions.)

stronger than the UL signal if the scatterer is close to the users, or weaker than the noise if it is distant.

The action metrics are shown in Fig. 4.5 according to the angle of the considered scatterer. For this figure, and also for Figures 4.6 and 4.11, users are at 0° , $\pm 2.5^\circ$, $\pm 7.5^\circ$, $\pm 12.5^\circ$. Generally, it is seen that the scatterers near UL users require Tx BF, and those near DL users require Rx BF. When a scatterer is close to both UL and DL users, or an FD user, spatial suppression fails and DSIC is required. Otherwise, taking no action for a distant scatterer has negligible disadvantages.

Figure 4.6 shows simulated UL and DL performances with different mitigation methods. "No action in Tx" in subplot (b) represents all variants in subplot (a) without Tx BF. As seen, the proposed selection is the best and closest to the SI-free case. For this figure, the metrics for the selection algorithm can be explicitly seen in Fig. 4.5, and the chosen action can be followed from the "1st" row in the subplot below.

On the other hand, DSIC with LS is very successful for this single-scatterer case.

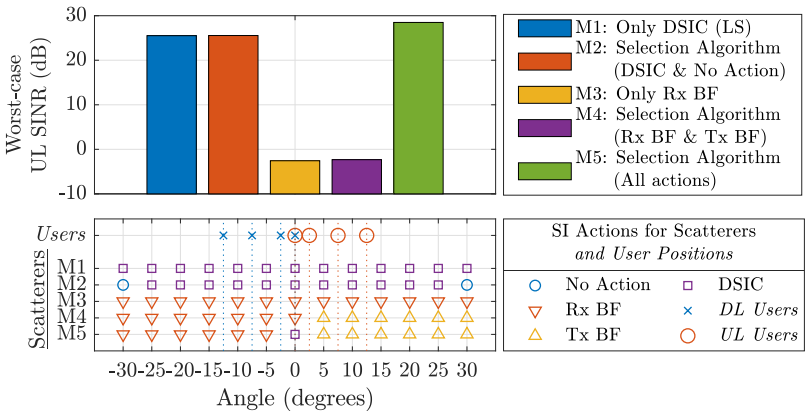


Figure 4.7: Action choices and performances on an exemplary scenario. ($K_d = 4$, $K_u = 4$, $K_s = 13$. User and scatterer positions are as shown with markers.)

As expected, DSIC performance does not depend on the scatterer angle. In the next figures, DSIC performance will degrade in the multi-scatterer case. The adaptive DSIC with RLS is included in the evaluation only to check if it can recover that degradation. In contrast, it deteriorates the performance even in the single-scatterer case due to the aging of the channels, which becomes more pronounced as the SI gets stronger or the channel varies faster.

Failure Mechanisms of Spatial Suppression: As seen, Rx BF degrades the UL performance when the scatterer is close to the UL users, and Tx BF degrades both the UL and the DL performance when the scatterer is close to the DL users. When the proposed algorithm is run with $\mathcal{A} = \{Tx, Rx\}$ to address this, failure persists in the vicinity of the FD user. These results introduce the basics of failure mechanisms of spatial suppression, and the first factor constituting the contributions of the proposed joint usage of DSIC and spatial suppression.

4.6.2 Performance Evaluation for the Multi-Scatterer Case

4.6.2.1 Illustration of Selected Actions for Scatterers

In the multi-scatterer case, each scatterer is associated with an action by the proposed policy. Fig. 4.7 shows these associations for 13 scatterers in the bottom subplot for five different methods (M1 to M5 in the top legend). While scatterers are uniformly

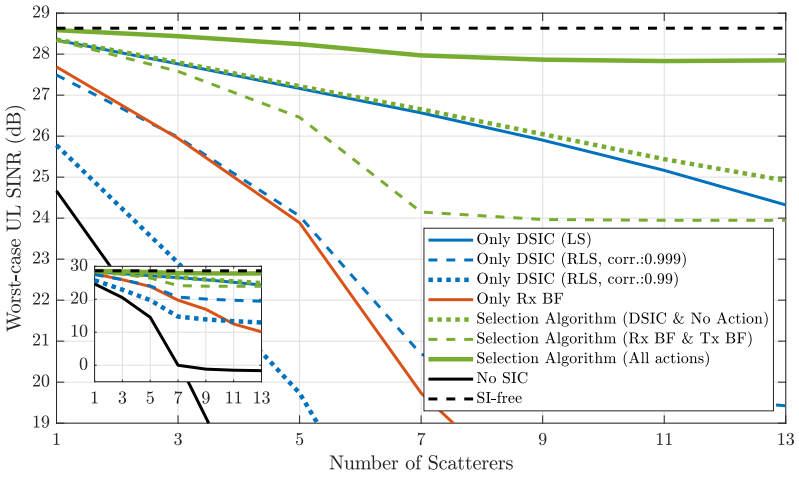


Figure 4.8: Worst-case UL SINR against the number of scatterers. ($K_d = 4$, $K_u = 4$, with the scenario described in Section 4.6.2.2.)

placed between -30° and +30°, user positions are the same as the previous plots. In addition, the resultant performances are also shown in the bar plot. In general, scatterers tend to require DSIC, Rx BF, or Tx BF when they are close to FD, DL, or UL users, respectively.

4.6.2.2 Overall Communication Performance on Randomized Scenarios

To allow more inclusive performance comparison for the multiscatterer case, the following figures will show results from simulations where user and scatterer positions are random between -30° and +30°, except one DL user and one UL user are always at $\theta = 0°$ as an FD user. SI mitigation performance is shown against the number of scatterers in Fig. 4.8, and against SI power in Fig. 4.9. It is seen that the proposed selection policy is robust against both factors and outperforms conventional implementations of DSIC and spatial suppression. Spatial suppression alternatives initially lose performance against these factors, but they converge to a performance metric as RCS increases, since the beamformer designs account for the RCS level.

Failure Mechanisms of DSIC: DSIC loses its almost perfect performance in Fig. 4.6 when scatterers are multiple, since the FFSI channel gets diffused in the delay domain as each scatterer has a different delay. These different delays decorrelate the

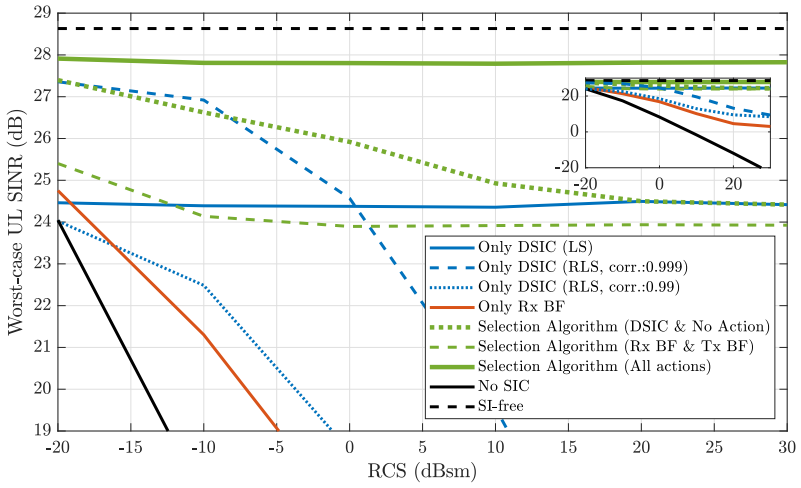


Figure 4.9: Worst-case UL SINR against RCS σ. ($K_d = 4$, $K_u = 4$, $K_s = 13$, with the scenario described in Section 4.6.2.2. 10 dBsm RCS is used for all the other simulations. Adding 36 dB to RCS gives the maximum INR $P_k^{(s)}/N_0$.)

observation noise for each scatterer during channel estimation, and each scatterer injects its own statistically independent error into the regenerated SI. Consequently, the performance monotonically decreases with the number of scatterers. This can be validated from the metric in (4.55), for which K_d uncorrelated sequences (intended for DL users) return from a single scatterer. The number is multiplied by K_s since each scatterer causes an uncorrelated echo due to different delays. This mechanism is directly related to the parameter-based processing discussed in Section 2.2.1. For fewer scatterers, the performance leverages the extra gain provided by the parameter-based processing, but each path brings its own error as the number of scatterers increases.

It is also interesting that the DSIC performance is independent of the SI power, as also shown in (4.55). Although the channel estimate gets more accurate for stronger SI, it is in the normalized mean squared error (NMSE) sense. The residual SI power is related directly to the power of the estimation error (without normalization), which is related to noise properties. Therefore, DSIC should not be applied for the scatterers yielding a weak SI, since the introduced error-based term might be stronger than the actual SI.

These disadvantages of DSIC constitute the second factor of the proposed joint usage policy. As seen, the proposed selection policy with all actions is robust against

the number of scatterers and the SI power, and superior to all conventional methods, represented by only DSIC, only Tx BF, and only Rx BF options. This is because the policy provides an intelligent and educated method to balance the mentioned advantages and disadvantages of DSIC and spatial suppression.

For constrained systems, enhanced alternatives to only DSIC and only BF schemes are also provided, for which the algorithm is run for $\mathcal{A} = \{DSIC, NA\}$ and $\mathcal{A} = \{Tx, Rx\}$, respectively. Although they are inferior to the case where all the actions are available, they outperform their conventional counterparts that represent the state-of-the-art. This is because they provide some level of flexibility to address the aforementioned failure mechanisms of DSIC and spatial suppression.

4.6.2.3 Full-Duplex Communication Rate

Full-duplex communication rates are shown in Fig. 4.10 in terms of AIR for UL and DL communications. In terms of median AIR, the proposed policy loses only 2.2% in UL and 1.2% in DL from SI-free FD cases, which is the theoretical limit, while conventional only DSIC, only Tx BF, and only Rx BF lose 16%, 45%, and 59% in UL, respectively. In other words, while Tx and Rx BF have no apparent improvement over the half-duplex case, only DSIC improves the rate by a factor of 1.67. The spatial suppression fails due to the one FD user in the scenario. On the other hand, the improvement of the proposed method almost achieves the theoretical limit by a factor of 1.95 in UL and 1.97 in DL communication.

4.6.3 Sensing Performance

Figure 4.11 is about the detection, mitigation, and initial angular estimation for emerging scatterers, detailed in Section 4.4.2. It shows the case where a new scatterer emerges while multiple scatterers are already known, and they are mitigated within the proposed policy, with the same positioning and actions in Fig. 4.7.

The emerging scatterers degrade the performance up to 30 dB, as shown in subplot (a). They are detected with the probabilities shown in subplot (b). While the detec-

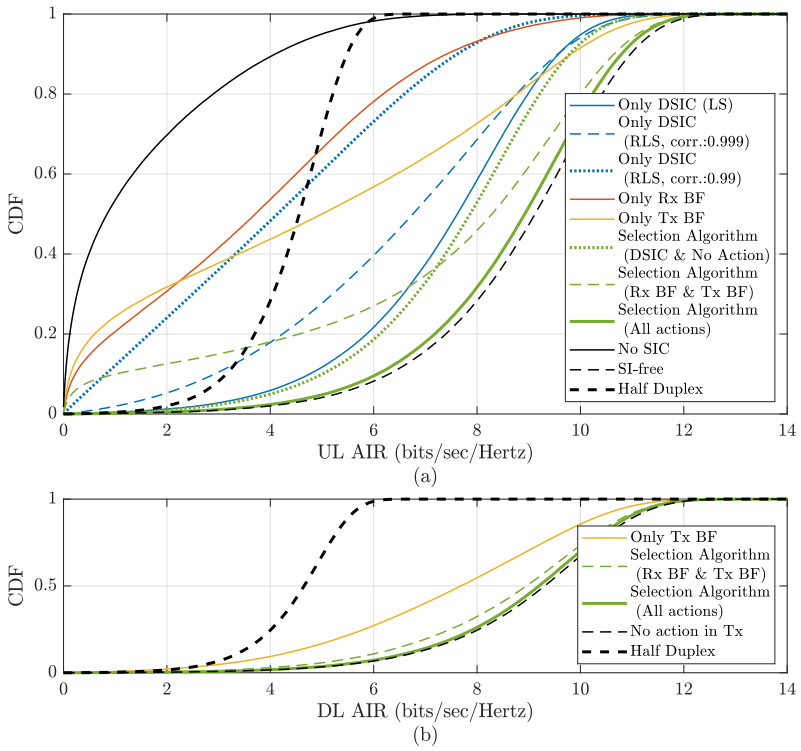


Figure 4.10: CDF of UL/DL AIR where user and scatterer positions are random. ($K_d = 4$, $K_u = 4$, $K_s = 13$, with the scenario described in Section 4.6.2.2.)

tion is guaranteed if they degrade the performance, they might be missed if they are weak due to being outside the beams, or their angular vicinity is suppressed due to a known scatterer. That is, the missed cases generally involve ineffective scatterers. Then, detected scatterers are mitigated by DSIC as described in Section 4.4.2, yielding almost an SI-free performance shown in subplot (a). Finally, the angles of the scatterers are estimated with error distributions as shown in subplot (c). This estimation is very accurate if they are inside the beams (between -15° and +15°), with errors almost always below 0.2°. However, rare cases of gross errors occur outside the beams, preventing blue and yellow curves from reaching one. Being very accurate except for rare outliers, this initial angular estimate can convert an exhaustive search into a simple verification process.

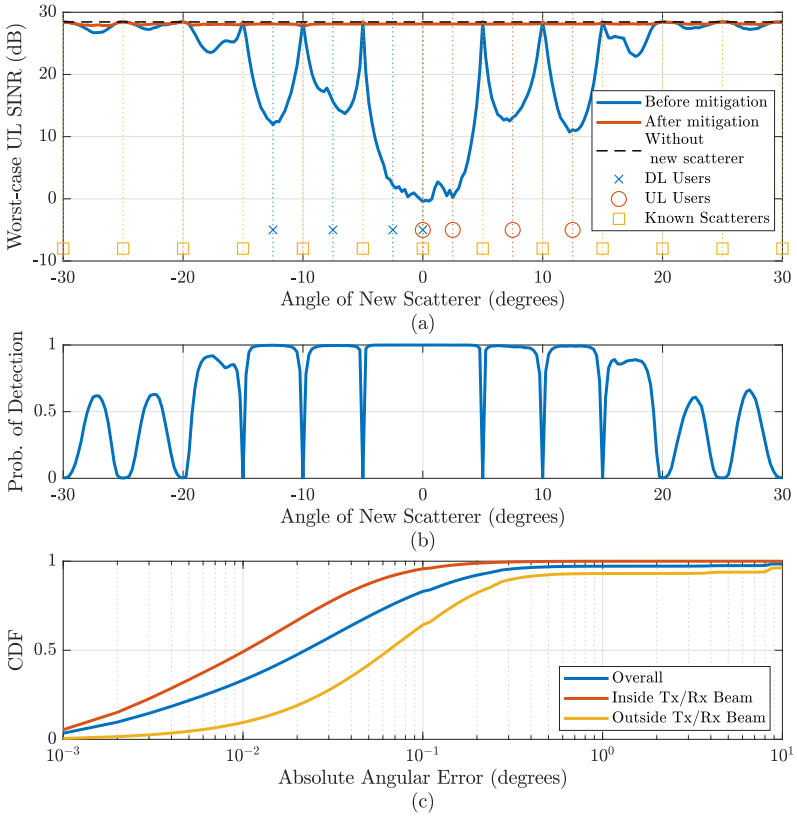


Figure 4.11: Performance of detection and mitigation mechanisms and initial angular estimation in case of an emerging single scatterer. ($K_d = 4$, $K_u = 4$, and $K_s = 13$ known scatterers. Markers show user and scatterer positions.)

CHAPTER 5

ROBUST STATISTICAL BEAMFORMING WITH MULTI-CLUSTER TRACKING FOR TIME-VARYING MASSIVE MIMO

In this chapter, a joint design of instantaneous channel estimation, beam tracking, and adaptive beamformer construction is proposed for a millimeter-wave massive multiple-input multiple-output (MIMO) system. This design focuses on efficiency in terms of performance and computational complexity under the adverse effects of time variation and mobility of sources, the presence of multi-user and multipath components, or simply multi-clusters, and the near-far effect. The design is also suitable for hybrid beamforming and frequency-selective channels. In the proposed system, channel parameters are estimated in time-division duplex (TDD) uplink mode using a per-cluster approach rather than a joint approach, which significantly reduces the complexity. Per-cluster estimation is possible thanks to the proposed interference-aware statistical beamforming method, namely reduced-dimensional Generalized Eigen-beamformer (RD-GEB), which is suitable for hybrid beamformer (HBF) structures. RD-GEB undertakes the computational load of interference mitigation and enables a simpler design for the remaining stages. In addition, the overall design is based on the separation of channel parameters as *fast-time* and *slow-time*, leaving only the instantaneous channel estimation and channel-matched filtering as fast-time operations, which are handled inside cluster-specific reduced-dimensional subspaces. Beam tracking and beamformer construction are held in slow time rarely, which reduces the time-averaged complexity. Furthermore, beam tracking is performed by leveraging a batch of instantaneous channel estimates, which removes the need for an additional training process. The proposed low-complexity design is shown to outperform the conventional methods.

In the remainder of the chapter, subscripts and superscripts generally indicate the time and cluster affiliations, respectively. Also, $\bar{x}$ and $\tilde{x}$ indicate the transform of an arbitrary variable x after the analog beamformer (ABF) and the digital beamformer (DBF), respectively.

5.1 Utilization in Full-Duplex Systems

The methods and techniques in this chapter can be used in the slow-time sensing interval mentioned in Section 4.4 to acquire angles of users and scatterers required in spatial suppression. This chapter classifies the duplexing as TDD for user equipments (UEs), and the base station (BS) might adopt full-duplex or TDD operation to serve multiple HD UEs. However, frequency-division duplexing (FDD) is out of scope due to the difference between downlink and uplink channels. If the BS is full-duplex, some of the UL signal clusters (to be discussed in this chapter) can be interpreted as FFSI, which carry similar modulated signals. In this case, the BS receives FFSI and UL signals, each with its own angular and temporal characteristics, similar to the scenario definition in Chapter 4. The advantage of estimating the angle for the FFSI (compared to a UE) is that the BS knows the whole signal coming from the scatterers, instead of a portion allocated for training. In this regard, this chapter complements the previous chapters about the acquisition of required angular information. In addition, the HBF consideration also covers fully digital beamformers as in Chapter 4, by applying the ABF digitally. This two-stage digital approach is known to reduce complexity with infrequent updates and spatial implementation of the FFT algorithm in the first stage.

5.2 System Model

This study considers a massive MIMO system where U single-antenna mobile UEs are simultaneously communicating with a BS with a uniform linear array (ULA) of N antenna elements in uplink mode using single-carrier modulation. Signals arrive at the BS as M angularly resolved signal clusters, where $M \geqslant U$. Parameters of MIMO channels are grouped into two, namely, FT and ST parameters. The BS esti-

Table 5.1: Nomenclature for Chapter 5

Variable	Description
n, p, k, m, u	Index for symbol, FT-CPI, ST-CPI, cluster, user
N, R, D_m	# "Antennas, RFCs (ABF outputs), DBF outputs" $\triangleq$ *
P, M, U	# FT-CPIs in an ST-CPI, clusters, users
N_{F}, N_{S}	# Symbols in (training, data) mode in an FT-CPI
$\boldsymbol{S}$, $\boldsymbol{W}^{(m)}$, $\boldsymbol{T}^{(m)}$	ABF, DBF, Total BF matrices
$\boldsymbol{y}_{n,p}$, $\boldsymbol{r}_{n,p}$, $\boldsymbol{z}_{n,p}^{(m)}$	Received Signal (at *)
$\boldsymbol{h}_p^{(m)}$, $\bar{\boldsymbol{h}}_p^{(m)}$, $\tilde{\boldsymbol{h}}_p^{(m,m)}$	Channel (at *)
$s_{n,p}^{(m)}$, $b_{n,p}^{(u)}$	Symbols from clusters and users
$E^{(m)}$, N_0	Cluster power, noise variance
$\alpha_{l,p}^{(m)}$, $\boldsymbol{a}(\theta)$	Channel complex gain, ULA steering vector
$\boldsymbol{R}_p^{(m)}$, $\bar{\boldsymbol{R}}_p^{(m)}$, $\tilde{\boldsymbol{R}}_p^{(m,m)}$	Channel covariance matrix (at *)
$\boldsymbol{\Psi}_p$, $\bar{\boldsymbol{\Psi}}_p$, $\tilde{\boldsymbol{\Psi}}_p^{(m)}$	Total covariance matrix (at *)
$\boldsymbol{x}_p^{(m)}$, $\boldsymbol{\nu}_p^{(m)}$, $\boldsymbol{A}$	Mobility state vector, innovation, transition matrix
$\theta_p^{(m)}$, $\Delta_p^{(m)}$	Mean AoA, AS
$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}$, $\boldsymbol{\xi}_p^{(m)}$	IEC estimate, estimation error
$\boldsymbol{R}_k^{(\mathrm{ST},m)}$, $\boldsymbol{\Psi}_k^{(\mathrm{ST})}$, $\bar{\boldsymbol{R}}_k^{(\mathrm{ST},m)}$, $\bar{\boldsymbol{\Psi}}_k^{(\mathrm{ST})}$, $\boldsymbol{x}_k^{(\mathrm{ST},m)}$, $\theta_k^{(\mathrm{ST},m)}$	Some ST model variables ($(\cdot)_k^{(\mathrm{ST})}$) corresponding to the previously defined variables ($(\cdot)_p$)

mates these parameters in separate repetition intervals, namely FT-CPI and ST-CPI. The estimated channel parameters can also be used in downlink mode, leveraging the channel reciprocity. The most important variables that will be defined in the subsequent sections are listed in Table 5.1.

5.2.1 Signal Model for Single Carrier Uplink Transmission

In uplink mode, the received signal vector $\boldsymbol{y}_{n,p} \in \mathbb{C}^{N\times 1}$ in n^{th} discrete time in p^{th} FT-CPI is

$$\boldsymbol{y}_{n,p} = \sum_{m=1}^{M} \sqrt{E^{(m)}}\boldsymbol{h}_p^{(m)} s_{n,p}^{(m)} + \boldsymbol{\eta}_{n,p} \tag{5.1}$$

where $s_{n,p}^{(m)}$ are zero-mean unit-variance symbols, $E^{(m)}$ are cluster powers, and $\boldsymbol{h}_p^{(m)} \in \mathbb{C}^{N\times 1}$ is the channel vector for m^{th} cluster. Also, $\boldsymbol{\eta}_{n,p} \sim \mathcal{CN}(\boldsymbol{0}, N_0\boldsymbol{I}_N)$ is the AWGN vector.

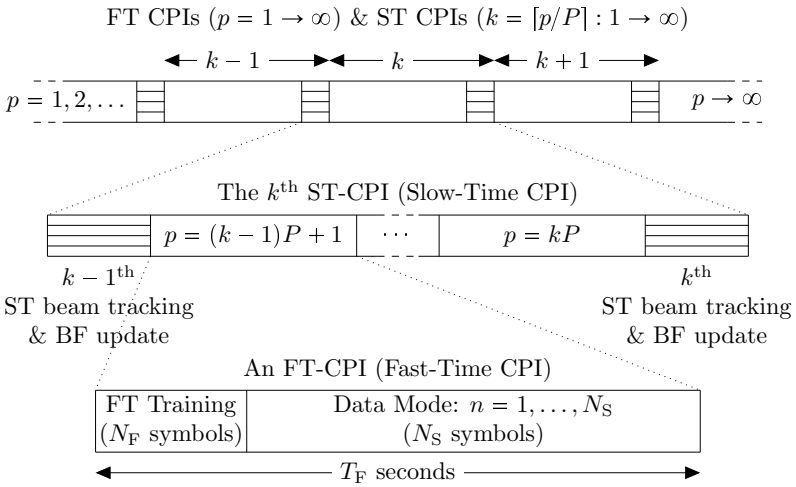


Figure 5.1: Signaling schemes and coherent processing intervals.

Signal clusters might belong to U different users with $U = M$, or some clusters might be multipath components from a smaller number of users with $U < M$. This fact can be shown as

$$s_{n,p}^{(m)} = b_{n-l_m,p}^{(\mathcal{U}(m))}, \tag{5.2}$$

where $b_{n,p}^{(u)}$ is the transmitted symbol from u^{th} user, and l_m indicates the delay for m^{th} cluster. $\mathcal{U}(m) = u$ is the user-cluster association function from the set of clusters $\{1, \ldots, M\}$ to the set of users $\{1, \ldots, U\}$. It is simply an identity function $\mathcal{U}(m) = m$ for the case $U = M$. The considered user-cluster association is illustrated in Fig. 5.2 for an example scenario.

5.2.2 Multi-Cluster Massive MIMO Channel Model

The ray-tracing-based Saleh-Valenzuela channel model is considered for angularly sparse millimeter wave channels [72, 118], which is written as

$$\boldsymbol{h}_p^{(m)} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \alpha_{l,p}^{(m)} \boldsymbol{a}(\tilde{\theta}_{l,p}^{(m)}) \tag{5.3}$$

for the m^{th} cluster in the p^{th} FT-CPI, where $\boldsymbol{a}(\theta) \triangleq \frac{1}{\sqrt{N}}[1\, e^{j\pi \sin\theta} \cdots e^{j(N-1)\pi \sin\theta}]$ with half-wavelength element spacing, and L is the number of rays. $\alpha_{l,p}^{(m)} \sim \mathcal{CN}(0, 1)$ are complex gains which are i.i.d. in l and p. In addition, $\tilde{\theta}_{l,p}^{(m)}$ are angles between the

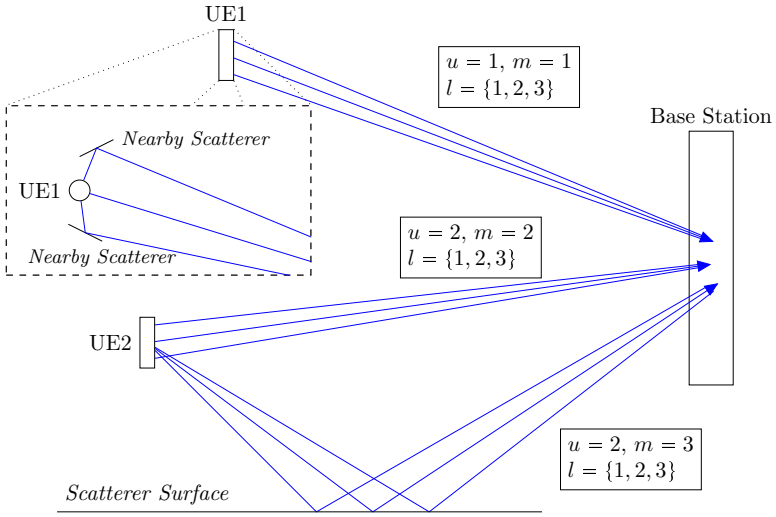


Figure 5.2: Illustration of users and signal clusters. ($U = 2$, $M = 3$, $L = 3$; $\mathcal{U}(m) = u$ is defined by $\mathcal{U}(1) = 1, \mathcal{U}(2) = 2, \mathcal{U}(3) = 2$).

incident rays inside the m^{th} cluster and the perpendicular to the array.[1] These angles are uniformly placed in the interval $(\theta_p^{(m)} - \Delta_p^{(m)}/2, \theta_p^{(m)} + \Delta_p^{(m)}/2)$ for m^{th} cluster in p^{th} FT-CPI. This interval will be described in the next sections by the mean AoA $\theta_p^{(m)}$ and the angular spread (AS) $\Delta_p^{(m)}$. The clustered channel structure can also be seen in Fig. 5.2. As depicted, clusters may occur due to nearby scatterers and scatterer surfaces, resulting in AS. On the other hand, the effects of AS can be experienced under mobility, even if the channel does not exhibit clustered signals instantaneously. Therefore, AS can also be an assumed parameter in the system design for robustness under mobility, affecting the analytical models of the channel and CCM. As L goes to infinity, the cluster CCM $\boldsymbol{R}_p^{(m)} \triangleq \mathbb{E}\{\boldsymbol{h}_p^{(m)}(\boldsymbol{h}_p^{(m)})^H\}$ becomes

$$\boldsymbol{R}_p^{(m)} = \frac{1}{\Delta_p^{(m)}} \int_{\theta_p^{(m)} - \Delta_p^{(m)}/2}^{\theta_p^{(m)} + \Delta_p^{(m)}/2} \boldsymbol{a}(\theta)\boldsymbol{a}^H(\theta) d\theta \tag{5.4}$$

[1] In three-dimensional space, this definition corresponds to the complementary angle of the angle between the incident wave and the direction of the ULA. As explained in [119], the definition (a) removes the dependence on the orientation of the reference coordinate system, and resulting products of sine or cosine of the elevation and azimuth angles, (b) represents the spatial sensing and beamforming ability of the ULA completely, (c) makes the angle of arrival (AoA) represent *angular* effects of mobility to the full extent. If the BS is equipped with a uniform planar array (UPA), the channel model with this AoA definition is valid for any set of elements that forms a ULA. By selecting ULAs in two perpendicular directions, the proposed system can be operated in parallel for each ULA, which extends the proposed system to UPAs for 2-D beam tracking.

at the p^{th} FT-CPI. Therefore, from (5.1), the total covariance matrix $\boldsymbol{\Psi}_p \triangleq \mathbb{E}\{\boldsymbol{y}_{n,p}\boldsymbol{y}_{n,p}^H\}$ is expressed as

$$\boldsymbol{\Psi}_p = \sum_{m=1}^{M} E^{(m)}\boldsymbol{R}_p^{(m)} + N_0\boldsymbol{I}_N. \tag{5.5}$$

5.2.3 Time Variation Model

The block-fading model [101] is adopted to model the time variation in the channel. The channel $\boldsymbol{h}_p^{(m)}$ is assumed to be invariant within FT-CPIs of T_{F} seconds, which consists of $N_{\text{F}} + N_{\text{S}}$ symbols as seen in Fig. 5.1. The channel varies after each FT-CPI as p progresses.

In the channel definition given in (5.3), complex gains $\alpha_{l,p}^{(m)}$ are *FT parameters*, and mean AoA $\theta_p^{(m)}$ and AS $\Delta_p^{(m)}$ are *ST parameters*. Time variation is mainly governed by the uncorrelatedness of $\alpha_{l,p}^{(m)}$ in FT-CPI index p, which is shown as

$$\mathbb{E}\{\alpha_{l,p}^{(m)}(\alpha_{l',p'}^{(m')})^*\} = \delta[l-l']\delta[m-m']\delta[p-p'], \tag{5.6}$$

while ST parameters also vary through p in a slower and correlated manner.

In this work, variation of $\theta_p^{(m)}$ is modeled via a linear Gaussian state-space model while $\Delta_p^{(m)}$ is assumed to be constant. $\theta_p^{(m)}$ and $\omega_p^{(m)}$ being the angular position and velocity, respectively; the state vector in the p^{th} FT-CPI is denoted by $\boldsymbol{x}_p^{(m)} \triangleq [\theta_p^{(m)} \quad \omega_p^{(m)}]^T$. The state equation is

$$\boldsymbol{x}_{p+1}^{(m)} = \boldsymbol{A}\boldsymbol{x}_p^{(m)} + \boldsymbol{\nu}_p^{(m)}, \qquad \boldsymbol{A} = \begin{bmatrix} 1 & T_{\text{F}} \\ 0 & 1 \end{bmatrix}, \tag{5.7}$$

where $\boldsymbol{\nu}_p^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\nu)$ is the innovation with $\boldsymbol{\Sigma}_\nu = \text{diag}\{\sigma_\theta^2, \sigma_\omega^2\}$. The system initializes with a given $\boldsymbol{x}_0^{(m)}$.

On the other hand, variations of the incident power $E^{(m)}$ and cluster delay l_m due to the movement of sources are related to the relative change in the distance. Therefore, their variations are very slow, and they are neglected.

Table 5.2: FT-CPI and ST-CPI

Number of Symbols in an FT-CPI				Number of FT-CPIs in an ST-CPI			
	Speed v (m/s)				N		
f_c/W	0.1	1	10	d/λ	16	64	128
30	10×10^6	1×10^6	100×10^3	1000	1250	313	156
100	3×10^6	300×10^3	30×10^3	3000	3750	938	470
300	1×10^6	100×10^3	10×10^3	10000	12500	3100	1600
1000	300×10^3	30×10^3	3×10^3	30000	37500	9400	4700

5.2.4 Practical Two-Stage Training Modes for Time-Varying Massive MIMO

The proposed structure of training and data transmission phases is illustrated in Fig. 5.1. FT-CPI and ST-CPI are defined to express the coherent interval in terms of FT and ST parameters. FT-CPIs consist of $N_{\mathrm{F}} + N_{\mathrm{S}}$ symbols, whereas ST-CPIs are longer and take P FT-CPIs. In literature, they are also called gain and angle coherence time [72], and channel coherence time and beam coherence time [73]. They determine the repetition time for the training modes *FT training* and *ST beam tracking* as described in Section 5.4, where beam tracking is performed more rarely. CPIs also determine the update times of different system blocks as described in Section 5.3, where statistical blocks are updated more rarely.

This two-stage structure reduces the average computational complexity of beam tracking per time and enables low-complexity techniques for FT training with the aid of approximately known AoA, which is called *beam-awareness*. However, the design parameters should be chosen such that the duration for $N_{\mathrm{S}} + N_{\mathrm{F}}$ symbols and P FT-CPIs are shorter than the actual FT-CPI and ST-CPI, respectively. As a design guideline, Table 5.2 provides these intervals for different hardware settings, where f_c, W, v, d, and λ are carrier frequency, bandwidth, speed, distance, and wavelength, respectively. FT-CPI is calculated as a tenth of the reciprocal of the Doppler frequency due to radial velocity ($0.1W/(vf_c/c)$ symbols), while beamwidth ($\sim 2/N$ radians) and tangential velocity (v/d radians/second) are considered for the ST-CPI, yielding $20(d/\lambda)/N$ FT-CPIs. Note that fc/W is the inverse normalized bandwidth and d/λ determines the path loss and power-related requirements [120].

5.3 General System Structure

In this section, the general structure of the beamforming and processing system will be introduced, leaving the details of design procedures to Sections 5.4 and 5.5. The HBF structure is adopted as depicted in Fig. 5.3 for its practicality and energy efficiency. ABF produces inputs for R RFCs via analog phase shifters and combiners, selecting all the interested clusters.

After the ABF, different than conventional systems, statistical beamforming is employed in cluster-specific DBF blocks via RD-GEB in Section 5.5. DBFs create cluster-specific processing blocks, each of which includes an ICS-CMF and a per-cluster estimation block. The DBF is responsible for the formation of a cluster-specific subspace that mitigates the multi-cluster interference. Therefore, ICS-CMF and estimators are designed by ignoring the multi-user interference, yielding low complexity. ICS-CMF combines the outputs from the DBF as a matched filter. The chain of ABF, DBF, and ICS-CMF gradually reduces the signal dimension from N to R/M, D_m, and 1, where $N > R/M \geqslant D_m \geqslant 1$. The output of DBF is also the place where the per-cluster estimation is performed, which consists of FT training and ST beam tracking, which will be detailed in Section 5.4.

Gains and phases of the outputs for clusters are corrected, and they are also combined by the digital cluster combiner (DCC) to reach symbol estimates for users, in the presence of multipath components ($U < M$) as described by (5.2). In this case, DCC is the block that brings the suitability for frequency-selective channels to the system.[2]

In this structure, ABF and DBF are updated after each ST beam tracking rarely, while ICS-CMF and DCC are updated after each FT training more frequently.

In the following subsections, the notation of variables will be introduced regarding the beamformer stages and their outputs, which is also illustrated in Fig. 5.4.

[2] Performance evaluation of DCC is left out of scope since it requires a more detailed scenario description. However, the quality of the symbol estimates at its input, measured via signal-to-interference-plus-noise ratio (SINR) defined in Section 5.7.2, determines its performance directly [120].

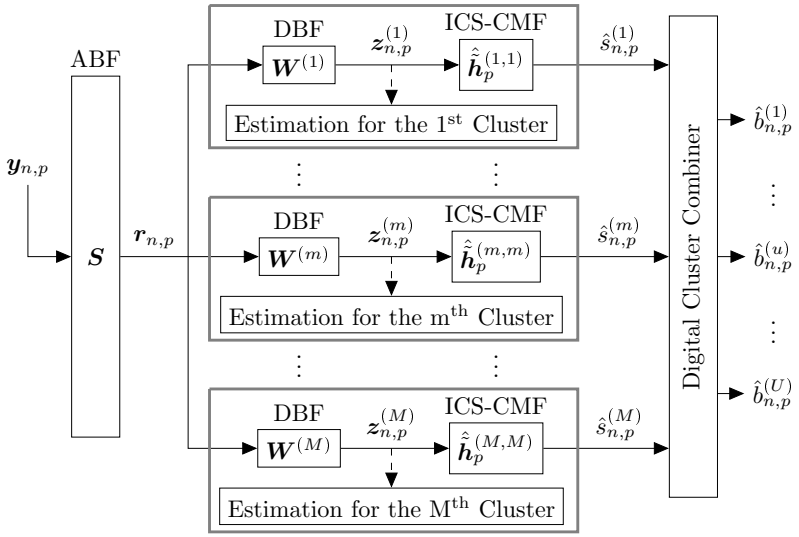


Figure 5.3: Beamforming structure.

5.3.1 DFT-Based Analog Beamformer for Full Cluster Coverage

The analog beamforming matrix is denoted by $\boldsymbol{S} \in \mathbb{C}^{N \times R}$ for R RFCs, whose columns are selected among DFT bases, resulting in a practical implementation via analog phase shifters. Also, the columns are orthonormal such that $\boldsymbol{S}^H \boldsymbol{S} = \boldsymbol{I}_R$. The output of the ABF is

$$\boldsymbol{r}_{n,p} \triangleq \boldsymbol{S}^H \boldsymbol{y}_{n,p} = \sum_{m=1}^{M} \sqrt{E^{(m)}} \bar{\boldsymbol{h}}_p^{(m)} s_{n,p}^{(m)} + \bar{\boldsymbol{\eta}}_{n,p} \tag{5.8}$$

where $\bar{\boldsymbol{h}}_p^{(m)} \triangleq \boldsymbol{S}^H \boldsymbol{h}_p^{(m)}$ and $\bar{\boldsymbol{\eta}}_{n,p} \triangleq \boldsymbol{S}^H \boldsymbol{\eta}_{n,p}$. After the analog stage, covariance matrices $\boldsymbol{R}_p^{(m)}$ and $\boldsymbol{\Psi}_p$ are modified as

$$\bar{\boldsymbol{R}}_p^{(m)} = \boldsymbol{S}^H \boldsymbol{R}_p^{(m)} \boldsymbol{S}, \tag{5.9}$$

$$\bar{\boldsymbol{\Psi}}_p = \boldsymbol{S}^H \boldsymbol{\Psi}_p \boldsymbol{S} = \sum_{m=1}^{M} E^{(m)} \bar{\boldsymbol{R}}_p^{(m)} + N_0 \boldsymbol{I}_R, \tag{5.10}$$

where $\bar{\boldsymbol{\Psi}}_p \triangleq \mathbb{E}\{\boldsymbol{r}_{n,p} \boldsymbol{r}_{n,p}^H\}$ and $\bar{\boldsymbol{R}}_p^{(m)} \triangleq \mathbb{E}\{\bar{\boldsymbol{h}}_p^{(m)} (\bar{\boldsymbol{h}}_p^{(m)})^H\}$.

5.3.2 Statistical Digital Beamformer for Inter-Cluster Separation

The digital stage starts with a bank of DBFs. DBFs are cluster-specific digital combiners for which statistical beamforming is employed. The DBF for the m^{th} cluster

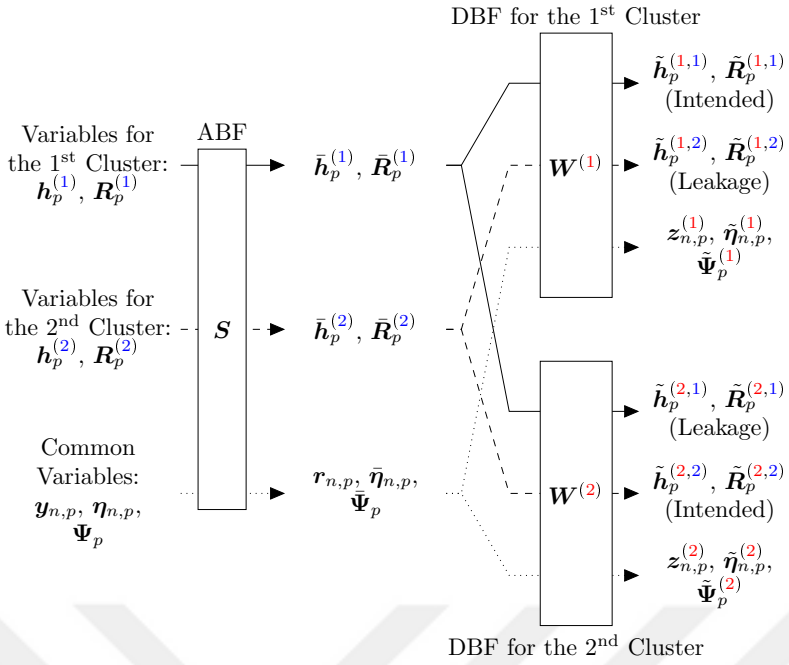


Figure 5.4: Notation of variables according to the processing stage for M=2.

$\boldsymbol{W}^{(m)} \in \mathbb{C}^{R\times D_m}$ has orthonormal columns such that $(\boldsymbol{W}^{(m)})^H\boldsymbol{W}^{(m)} = \boldsymbol{I}_{D_m}$. It processes the ABF output $\boldsymbol{r}_{n,p}$ and outputs $\boldsymbol{z}_{n,p}^{(m)} \triangleq \left(\boldsymbol{W}^{(m)}\right)^H \boldsymbol{r}_{n,p}$. Defining the total beamforming matrix $\boldsymbol{T}^{(m)} \triangleq \boldsymbol{S}\boldsymbol{W}^{(m)}$, the DBF output is

$$\boldsymbol{z}_{n,p}^{(m)} = (\boldsymbol{T}^{(m)})^H\boldsymbol{y}_{n,p} = \sum_{m'=1}^{M} \sqrt{E^{(m')}}\tilde{\boldsymbol{h}}_p^{(m,m')}s_{n,p}^{(m')} + \tilde{\boldsymbol{\eta}}_{n,p}^{(m)} \tag{5.11}$$

where $\tilde{\boldsymbol{h}}_p^{(m,m')} \triangleq (\boldsymbol{T}^{(m)})^H\boldsymbol{h}_p^{(m')}$ is the IEC and $\tilde{\boldsymbol{\eta}}_{n,p}^{(m)} \triangleq (\boldsymbol{T}^{(m)})^H\boldsymbol{\eta}_{n,p} \sim \mathcal{CN}(\boldsymbol{0}, N_0\boldsymbol{I}_{D_m})$. After the digital stage, covariance matrices of $\tilde{\boldsymbol{h}}_p^{(m,m')}$ and $\boldsymbol{z}_{n,p}^{(m)}$, denoted by $\tilde{\boldsymbol{R}}_p^{(m,m')}$ and $\tilde{\boldsymbol{\Psi}}_p^{(m)}$, respectively, are given in (5.12) and (5.13).

$$\tilde{\boldsymbol{R}}_p^{(m,m')} = (\boldsymbol{T}^{(m)})^H\boldsymbol{R}_p^{(m')}\boldsymbol{T}^{(m)} \tag{5.12}$$

$$\tilde{\boldsymbol{\Psi}}_p^{(m)} = (\boldsymbol{T}^{(m)})^H\boldsymbol{\Psi}_p\boldsymbol{T}^{(m)} = \sum_{m'=1}^{M} E^{(m')}\tilde{\boldsymbol{R}}_p^{(m,m')} + N_0\boldsymbol{I}_{D_m} \tag{5.13}$$

The design method for the statistical DBF will be given in Section 5.5, which will utilize the outputs of beam trackers that will be established in Section 5.4.

5.3.3 Intra-Cluster Spatial Channel Matched Filtering (ICS-CMF)

Different from temporal channel-matched filtering, ICS-CMF compensates for the indifference of statistically constructed slow-time DBF to fast-time variations in the channel. After beamforming, D_m outputs are obtained for the m^{th} cluster. ICS-CMF efficiently combines these outputs using the matched filtering method, assuming the multi-cluster interference has already been suppressed to a negligible level by beamforming.[3] The output of ICS-CMF is obtained by

$$\hat{s}_{n,p}^{(m)} \triangleq (\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H \boldsymbol{z}_{n,p}^{(m)} \tag{5.14}$$

where $\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}$ is the estimate for the channel $\tilde{\boldsymbol{h}}_p^{(m,m)}$, which will be detailed in Section 5.4.1. The output is expressed as

$$\hat{s}_{n,p}^{(m)} = \sum_{m'=1}^{M} \sqrt{E^{(m')}} (\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H \tilde{\boldsymbol{h}}_p^{(m,m')} s_{n,p}^{(m')} + (\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H \tilde{\boldsymbol{\eta}}_{n,p}^{(m)} \tag{5.15}$$

5.4 Two-Stage Parameter Estimation Procedure

The task of channel estimation is held in two stages in the proposed system. They are named FT training and ST beam tracking, which are introduced in Section 5.2.4. To recite, there is an FT/ST separation in channel parameters, CPIs, estimation phases, and processing blocks in terms of their update rate.

The proposed *per-cluster* estimation approach employs separate estimators for each cluster as seen in Fig. 5.3. In Fig. 5.5, the estimation procedure is illustrated in more detail for one of the clusters. The fundamental aspect of the per-cluster estimation is that the estimation takes place after cluster-specific DBFs, which mitigate the interference better than conventional DFT beams. Also, the size of the effective channel decreases further. As a result, IEC estimators and beam trackers can be designed more simply, as if only a single user exists, in the presence of multiple users or clusters.

On the other hand, in a conventional estimation scheme, the channel is estimated jointly after ABF, constructed with DFT beams. DFT beams are weak against inter-

[3] If multiple users ($\leqslant D_m$) fall into the angular sector of the same cluster, joint intra-cluster processing can be applied [100] both in data and training phases in RD subspace (D_m).

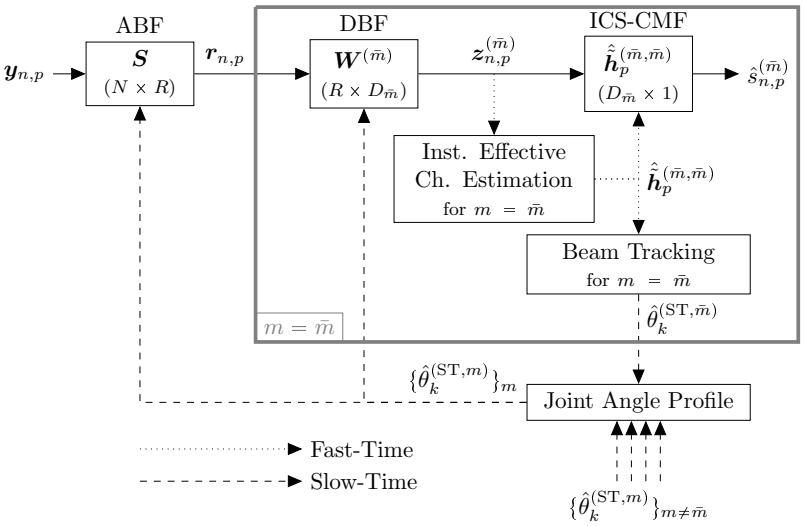


Figure 5.5: Estimation structure for the $\bar{m}^{\text{th}}$ signal cluster.

ference, which charges the estimator with the interference mitigation task through pilots or other methods. Also, in these systems, DBF depends on instantaneous channel estimates and needs more frequent updates, yet it is still responsible for interference mitigation for data transmission. Therefore, conventional methods suffer from performance loss, complexity increase, or a need for longer training overhead for channel estimation.

Beam trackers use batches of IEC estimates to estimate AoAs. It is very important that the channel estimator operates in fast-time and the beam tracker operates in slow-time, as shown by the dotted and dashed lines in Fig. 5.5. Therefore, ABF and DBF are updated in slow-time, while the ICS-CMF is updated in fast-time. Since DBF is slowly updated, a more complex design is tolerable, which increases the interference mitigation performance.

5.4.1 FT Estimation: Instantaneous Effective Channel (IEC) Estimation

The IEC estimator aims to estimate the rapidly changing (instantaneous) RD channel observed after DBF, which is the reason why it is called the *effective* channel. Training sequences of length N_{F} are transmitted, which are shown by the vector $\boldsymbol{s}^{(m)} = [s_{1+l_m,p}^{(m)} \cdots s_{N_{\mathrm{F}}+l_m,p}^{(m)}]^T = [b_{1,p}^{(\mathcal{U}(m))} \cdots b_{N_{\mathrm{F}},p}^{(\mathcal{U}(m))}]^T$, where l_m is the cluster de-

lay. The received signals from $n = 1 + l_m$ to $n = N_{\mathrm{F}} + l_m$ are collected, and the techniques below are applied. In the subsequent sections, $l_m = 0$ is assumed for the sake of simplicity.[4]

5.4.1.1 Beam-Aware Least Squares (BA-LS) Estimation

This technique observes the received signal after the ABF and DBF ($\boldsymbol{T}^{(m)} = \boldsymbol{S}\boldsymbol{W}^{(m)}$), which were set previously, to estimate the IEC denoted by $\tilde{\boldsymbol{h}}_p^{(m,m)}$. *Beam-awareness* implies that the subspace created by $\boldsymbol{T}^{(m)}$ is cluster-specific and interference mitigation is accomplished before by the beamformer, which yields a simple design. This technique also does not need synchronization between the sources since it operates per-cluster. Therefore, $\boldsymbol{s}^{(m)}$ for different m's are uncorrelated.

The received signal $\boldsymbol{z}_{n,p}^{(m)}$ after $\boldsymbol{T}^{(m)}$ is as shown in (5.11). Received signals during the training are concatenated to obtain $\boldsymbol{z}_{:,p}^{(m)} \triangleq [(\boldsymbol{z}_{1,p}^{(m)})^T \dots (\boldsymbol{z}_{N_{\mathrm{F}},p}^{(m)})^T]^T$, which is expressed as

$$\boldsymbol{z}_{:,p}^{(m)} = \sum_{m'=1}^{M} \sqrt{E^{(m')}} (\boldsymbol{s}^{(m')} \otimes \boldsymbol{I}_{D_m}) \tilde{\boldsymbol{h}}_p^{(m,m')} + \tilde{\boldsymbol{\eta}}_{:,p}^{(m)}, \tag{5.16}$$

where $\tilde{\boldsymbol{\eta}}_{:,p}^{(m)} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_{\mathrm{F}}} \otimes (N_0 \boldsymbol{I}_{D_m}))$, and $\otimes$ is the Kronecker product operator. Then, the received signal is processed by the LS estimator $\boldsymbol{Z}^{(m)}$ to obtain the LS estimate for the IEC $\tilde{\boldsymbol{h}}_p^{(m,m)}$ as

$$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} \triangleq (\boldsymbol{Z}^{(m)})^H \boldsymbol{z}_{:,p}^{(m)}, \qquad \boldsymbol{Z}^{(m)} = \frac{(\boldsymbol{s}^{(m)} \otimes \boldsymbol{I}_{D_m})}{\sqrt{E^{(m)}} N_{\mathrm{F}}} \tag{5.17}$$

where the expression for $\boldsymbol{Z}^{(m)}$ is simplified with the assumption of $(\boldsymbol{s}^{(m)})^H \boldsymbol{s}^{(m)} = N_{\mathrm{F}}$. Consequently, the channel estimate can be expressed as

$$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} = \tilde{\boldsymbol{h}}_p^{(m,m)} + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \frac{\sqrt{E^{(m')}} \big((\boldsymbol{s}^{(m)})^H \boldsymbol{s}^{(m')} \big)}{\sqrt{E^{(m)}} N_{\mathrm{F}}} \tilde{\boldsymbol{h}}_p^{(m,m')} + \boldsymbol{\xi}_p^{(m)} \tag{5.18}$$

where the second term is the inter-cluster interference and the last term is the error due to noise with $\boldsymbol{\xi}_p^{(m)} \sim \mathcal{CN}(\boldsymbol{0}, \frac{N_0}{E^{(m)} N_{\mathrm{F}}} \boldsymbol{I}_{D_m})$.

[4] The delay l_m is a slowly-changing parameter that is easy to track [120]. Therefore, it is not a big difficulty for per-cluster BA-LS. However, it increases the effective delay spread, hence the size of the observation vector, for conventional joint techniques in Section 5.4.1.2.

5.4.1.2 Conventional Joint Estimation Techniques

In contrast to per-cluster BA-LS, a conventional estimation method might process the signals directly after the ABF and estimate the multi-cluster channels jointly. In this case, the $N_{\mathrm{F}}R \times 1$ observation vector is $\boldsymbol{r}_{:,p} \triangleq \left[\boldsymbol{r}_{1,p}^T \cdots \boldsymbol{r}_{N_{\mathrm{F}},p}^T\right]^T$ and the $MR \times 1$ regressor vector would be $\bar{\boldsymbol{h}}_p^{(:)} \triangleq [(\bar{\boldsymbol{h}}_p^{(1)})^T \cdots (\bar{\boldsymbol{h}}_p^{(M)})^T]^T$. The observation can be expressed as

$$\boldsymbol{r}_{:,p} = \left(\left[\sqrt{E^{(1)}}\boldsymbol{s}^{(1)} \ \ldots \ \sqrt{E^{(M)}}\boldsymbol{s}^{(M)}\right] \otimes \boldsymbol{I}_R\right) \bar{\boldsymbol{h}}_p^{(:)} + \bar{\boldsymbol{\eta}}_{:,p} \tag{5.19}$$

where $\bar{\boldsymbol{\eta}}_{:,p}$ is the AWGN with covariance $N_0\boldsymbol{I}_{N_{\mathrm{F}}R}$. With these definitions and the ones in Section 5.3.1, well-known LS and minimum mean squared error (MMSE) methods are implemented to estimate the concatenated channels $\bar{\boldsymbol{h}}_p^{(:)}$ jointly from the observation $\boldsymbol{r}_{:,p}$ after ABF [114, 120]. Note that these joint methods require synchronization between users before the training phase.

5.4.2 ST Estimation: Beam Tracking

In the ST beam tracking phase, ST parameters of the channel, namely AoA $\theta_p^{(m)}$ and AS $\Delta_p^{(m)}$, should be estimated, which determine the second-order statistics of the channel as seen in (5.4). However, in this work, AS $\Delta_p^{(m)}$ is assumed to be constant and known.[5] AoA of each cluster is estimated separately by cluster-specific estimators by leveraging the IEC estimates from the latest P FT-CPIs. Therefore, beam tracking is actually a calculation phase, and transmission of a new training sequence is not needed. After the beam tracking phase, beamformers are updated with the new information.

5.4.2.1 Assumed Slow-Time Model

The overall design of ST operations, namely beam tracking and beamformer update, assumes a *coherent interval* in terms of ST parameters, namely *ST-CPI*, which is longer than FT-CPIs as shown in Fig. 5.1. Therefore, these ST operations are repeated

[5] AS estimation is studied in literature [72] and it is out of the scope of this chapter. It is shown in [100] that the performance is not very sensitive to the exact value of AS.

after a long time compared to FT operations. The length of ST-CPI is determined by the number of FT-CPIs within, denoted by P, as a design parameter, which should yield a similarity in terms of channel statistics such as $\boldsymbol{R}_p^{(m)} \cong \boldsymbol{R}_{p+P}^{(m)}$.

Due to the assumption of a different CPI, ST operations work under a mismatched channel model. Based on the aforementioned similarity, *assumed* variables related to previously defined ones are needed, indicated by the superscript ST. The first example of ST variables is

$$\boldsymbol{R}_k^{(\mathrm{ST},m)} \cong \boldsymbol{R}_p^{(m)}, \qquad \text{for } p = (k-1)P+1, \ldots, kP. \tag{5.20}$$

As seen, ST variables are indexed by the ST-CPI index k, which is related to FT-CPI indices through $k = \lceil p/P \rceil$. Several ST variables will be used in the next sections without an explicit definition such as $\boldsymbol{\Psi}_k^{(\mathrm{ST})}$, $\bar{\boldsymbol{R}}_k^{(\mathrm{ST},m)}$, $\bar{\boldsymbol{\Psi}}_k^{(\mathrm{ST})}$, $\boldsymbol{x}_k^{(\mathrm{ST},m)}$, $\theta_k^{(\mathrm{ST},m)}$, but they imply a similar relation.

The state-space model for angular variation defined in Section 5.2.3 is modified for ST operations as

$$\boldsymbol{x}_{k+1}^{(\mathrm{ST},m)} \triangleq \boldsymbol{A}^{(\mathrm{ST})} \boldsymbol{x}_k^{(\mathrm{ST},m)} + \boldsymbol{\nu}_k^{(\mathrm{ST},m)}, \tag{5.21}$$

$$\theta_k^{(\mathrm{ST},m)} \triangleq \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{x}_k^{(\mathrm{ST},m)}, \tag{5.22}$$

where $\boldsymbol{A}^{(\mathrm{ST})} \triangleq \boldsymbol{A}^P$, $\boldsymbol{\nu}_k^{(\mathrm{ST},m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\nu^{(\mathrm{ST})})$, and $\boldsymbol{\Sigma}_\nu^{(\mathrm{ST})} = \sum_{i=1}^{P} \boldsymbol{A}^i \boldsymbol{\Sigma}_\nu (\boldsymbol{A}^i)^T$.

Furthermore, it is assumed that the overall beamformer structure has suppressed the multi-cluster interference to a negligible level. Therefore, the channel estimates expressed in (5.18) are suboptimally assumed as

$$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} \cong \tilde{\boldsymbol{h}}_p^{(m,m)} + \boldsymbol{\xi}_p^{(m)} \tag{5.23}$$

with $\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}(\theta_k^{(\mathrm{ST},m)}) + \frac{N_0}{E^{(m)} N_{\mathrm{F}}} \boldsymbol{I}_{D_m})$ in the design of estimators. The parametric RD CCM $\boldsymbol{R}(\theta)$ is expressed as

$$\boldsymbol{R}(\theta) \triangleq (\boldsymbol{T}^{(m)})^H \boldsymbol{R}^{\mathrm{FD}}(\theta) \boldsymbol{T}^{(m)}, \tag{5.24}$$

where $\boldsymbol{R}^{\mathrm{FD}}(\theta) \triangleq \frac{1}{\Delta} \int_{\theta-\Delta/2}^{\theta+\Delta/2} \boldsymbol{a}(\theta') \boldsymbol{a}^H(\theta') d\theta'$ is the full-dimensional (FD)[6] CCM for mean AoA θ and AS Δ. The approximation $\boldsymbol{R}^{\mathrm{FD}}(\theta) \cong \mathrm{diag}(\boldsymbol{a}(\theta)) \boldsymbol{D} \, \mathrm{diag}(\boldsymbol{a}(\theta))^H$

[6] Although FD was used for *full-duplex* previously, it will be used for *full-dimensional* only in this chapter, often compared to reduced-dimensional (RD) entities and operations.

is validated in the prior work [89], where $(\boldsymbol{D})_{a,b} = \text{sinc}\left((a-b)\cos(\theta)\sin(\Delta/2)\right)$ for $a, b = 1, \ldots, N$. Further inspired by [89] for a simpler calculation of $\boldsymbol{R}(\theta)$, it is approximated by

$$\boldsymbol{R}(\theta) \cong \boldsymbol{E}(\theta)\boldsymbol{E}^H(\theta) \tag{5.25}$$

$$\boldsymbol{E}(\theta) \triangleq (\boldsymbol{T}^{(m)})^H \text{diag}(\boldsymbol{a}(\theta))\boldsymbol{E}^{\text{FD}} \tag{5.26}$$

where $\boldsymbol{E}^{\text{FD}}$ has $\sqrt{\lambda_d}\boldsymbol{e}_d$ in its columns where λ_d and $\boldsymbol{e}_d$ are eigenvalues and eigenvectors of $\boldsymbol{D}$. Noting $\boldsymbol{R}^{\text{FD}}(\theta)$ and $\boldsymbol{D}$ are effectively low-rank for practical Δ values, $\boldsymbol{E}^{\text{FD}}$ can be constructed with the most dominant D_m eigenvalues and eigenvectors, yielding a size of $N \times D_m$ where $D_m \ll N$.[7] $\boldsymbol{E}^{\text{FD}}$ is calculated and stored once, and $\boldsymbol{R}(\theta)$ can be calculated with $\mathcal{O}(ND_m^2)$ multiplications via (5.25), instead of $\mathcal{O}(N^2 D_m)$ via (5.24), leveraging also the diagonal form of the $N \times N$ matrix $\text{diag}(\boldsymbol{a}(\theta))$ inside the $\boldsymbol{E}(\theta)$ expression.[8]

5.4.2.2 Beam-Aware Maximum-Likelihood (BA-ML) Estimator

The observation vector used by the BA-ML estimator for the m^{th} cluster in the k^{th} ST beam tracking phase is

$$\boldsymbol{f}_k^{(m)} \triangleq \left[\left[\hat{\tilde{\boldsymbol{h}}}_{(k-1)P+1}^{(m,m)}\right]^H \cdots \left[\hat{\tilde{\boldsymbol{h}}}_{kP}^{(m,m)}\right]^H \right]^H. \tag{5.27}$$

In order to parameterize it on θ, (5.23) is rewritten as

$$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} \cong \boldsymbol{E}(\theta_k^{(\text{ST},m)})\boldsymbol{\beta}_p + \boldsymbol{\xi}_p^{(m)}, \tag{5.28}$$

where $\boldsymbol{\beta}_p \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{D_m})$ is the basis coefficient vector for the matrix $\boldsymbol{E}(\theta_k^{(\text{ST},m)})$ defined in (5.26), whose columns span the range space of $\boldsymbol{R}(\theta)$ in (5.24). The observation in (5.27) is parameterized on the sought unknown $\theta_k^{(\text{ST},m)}$ with the model in (5.28). However, $\boldsymbol{\beta}_p$ is also an unknown. In this case, the *nonlinear least squares method*, which is equivalently the ML method for Gaussian cases [99], yields the

[7] The number of columns of $\boldsymbol{E}^{\text{FD}}$ and the number of DFT outputs are equal since both are determined by the effective rank of the CCM.

[8] $\boldsymbol{D}$ depends on θ through $\cos(\theta)\sin(\Delta/2)$, which is the transformed AS divided by 2π, obtained from $\pi\sin(\theta + \Delta/2) - \pi\sin(\theta - \Delta/2)$ [89]. [100] shows the tolerance to AS errors, therefore dependency on θ can be neglected by selecting $\theta = 0$. Nevertheless, one can store more than one $\boldsymbol{D}$ and $\boldsymbol{E}^{\text{FD}}$ matrices for quantized values of θ for a more accurate approximation.

estimated $\theta_k^{(\mathrm{ST},m)}$ from the observation $\boldsymbol{f}_k^{(m)}$ as

$$\hat{\theta}_{k+1}^{(\mathrm{ST},m)} = arg \max_{\theta} \max_{\{\boldsymbol{\beta}_p\}} p(\boldsymbol{f}_k^{(m)} | \theta, \{\boldsymbol{\beta}_p\}), \tag{5.29}$$

which firstly finds and sets the maximizing set of $\{\boldsymbol{\beta}_p\}$ for each given θ, then applies classical ML procedure on θ. However, the given problem has PD_m+1 unknowns but PD_m equations. Therefore, the size of $\boldsymbol{\beta}_p$ is reduced to $D'_m < D_m$, and $\boldsymbol{E}(\theta_k^{(\mathrm{ST},m)}) \in \mathbb{C}^{D_m \times D_m}$ is modified as $\boldsymbol{E}'(\theta_k^{(\mathrm{ST},m)}) \in \mathbb{C}^{D_m \times D'_m}$ by simply removing the weakest eigenvectors from $\boldsymbol{E}^{\mathrm{FD}}$ in (5.26). This modifies (5.28) as

$$\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} \cong \boldsymbol{E}'(\theta_k^{(\mathrm{ST},m)}) \boldsymbol{\beta}'_p + \boldsymbol{\xi}_p^{(m)}. \tag{5.30}$$

Substituting (5.30) in (5.27), it is found in Appendix D.1 that the AoA estimate can be calculated as

$$\hat{\theta}_{k+1}^{(\mathrm{ST},m)} = arg \min_{\theta} \mathrm{tr} \left(\boldsymbol{M}(\theta) \left(\sum_{p=(k-1)P+1}^{kP} \hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} (\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H \right) \right), \tag{5.31}$$

where $\boldsymbol{M}(\theta) \triangleq \boldsymbol{I}_{D_m} - \boldsymbol{E}'(\theta) \left(\boldsymbol{E}'^H(\theta) \boldsymbol{E}'(\theta) \right)^{-1} \boldsymbol{E}'^H(\theta)$.

5.4.2.3 Statistical Extended Kalman Filter (SEKF)

A Bayesian method, particularly the Kalman filter, can exploit the correlation of AoA through time better due to slow variation. Therefore, the second-order channel statistics are fed to the EKF as observations, since it is a nonlinear function of AoA. The statistics are obtained via a sample-mean covariance matrix, and the observation vector for EKF is obtained as

$$\boldsymbol{f}_k^{(m)} = \mathrm{vec} \left\{ \frac{1}{P} \sum_{p=(k-1)P+1}^{kP} \hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} (\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H \right\} \tag{5.32}$$

where $\mathrm{vec}\{\cdot\}$ is the vectorization operator. Assuming angular coherence, the observation can be rewritten as

$$\boldsymbol{f}_k^{(m)} = \mathrm{vec} \left\{ \boldsymbol{R}_f(\theta_k^{(\mathrm{ST},m)}) \right\} + \boldsymbol{q}_k^{(m)} \tag{5.33}$$

whose first term is the mean with $\boldsymbol{R}_f(\theta_k^{(\mathrm{ST},m)}) \triangleq \boldsymbol{R}(\theta_k^{(\mathrm{ST},m)}) + \frac{N_0}{E^{(m)} N_{\mathrm{F}}} \boldsymbol{I}_{D_m}$, where (5.25) can be used for $\boldsymbol{R}(\theta)$. The second term $\boldsymbol{q}_k^{(m)}$ is the zero-mean error vector with

covariance $\boldsymbol{Q}_k^{(m)}$, which is found in Appendix D.2 as

$$\boldsymbol{Q}_k^{(m)} = \frac{1}{P}\left(\boldsymbol{R}_f(\theta_k^{(\mathrm{ST},m)})\right)^* \otimes \boldsymbol{R}_f(\theta_k^{(\mathrm{ST},m)}) \tag{5.34}$$

To implement EKF, $\mathrm{vec}\{\boldsymbol{R}_f(\theta_k^{(\mathrm{ST},m)})\}$ in (5.33) is linearized using the first-order Taylor expansion, which yields

$$\boldsymbol{f}_k^{(m)} \cong \boldsymbol{B}_k^{(m)}\boldsymbol{x}_k^{(\mathrm{ST},m)} + \boldsymbol{q}_k^{(m)} + \{known\,\&\,constant\} \tag{5.35}$$

where $\boldsymbol{B}_k^{(m)}$ is the Jacobian matrix of $\mathrm{vec}\{\boldsymbol{R}_f(\theta)\}$ with respect to $\boldsymbol{x}_k^{(\mathrm{ST},m)}$, to be evaluated at $\hat{\boldsymbol{x}}_{k|k-1}^{(\mathrm{ST},m)} = [\hat{\theta}_{k|k-1}^{(\mathrm{ST},m)} \quad \hat{\omega}_{k|k-1}^{(\mathrm{ST},m)}]$. It is found as

$$\boldsymbol{B}_k^{(m)} = \left[\left(\frac{\partial \mathrm{vec}\{\boldsymbol{R}(\theta)\}}{\partial\theta}\right)\Big|_{\theta=\hat{\theta}_{k|k-1}^{(\mathrm{ST},m)}} \qquad \boldsymbol{0}\right] \tag{5.36}$$

whose second column, derivative with respect to speed, is zero due to the angular coherence assumption. The derivative with respect to the angular position can be calculated numerically or analytically using the simple structure in (5.25).

Using these definitions, well-known iterations of EKF [114] are applied for each cluster m. These iterations are shown below for the k^{th} beam tracking phase.

$$\boldsymbol{K} = \boldsymbol{\Sigma}_{k|k-1}^{(\mathrm{ST},m)}(\boldsymbol{B}_k^{(m)})^H\left(\boldsymbol{B}_k^{(m)}\boldsymbol{\Sigma}_{k|k-1}^{(\mathrm{ST},m)}(\boldsymbol{B}_k^{(m)})^H + \boldsymbol{Q}_k^{(m)}\right)^{-1} \tag{5.37}$$

$$\hat{\boldsymbol{x}}_{k|k}^{(\mathrm{ST},m)} = \hat{\boldsymbol{x}}_{k|k-1}^{(\mathrm{ST},m)} + \boldsymbol{K}\left(\boldsymbol{f}_k^{(m)} - \mathrm{vec}\{\boldsymbol{R}_f(\hat{\theta}_{k|k-1}^{(\mathrm{ST},m)})\}\right) \tag{5.38}$$

$$\boldsymbol{\Sigma}_{k|k}^{(\mathrm{ST},m)} = \boldsymbol{\Sigma}_{k|k-1}^{(\mathrm{ST},m)} - \boldsymbol{K}\boldsymbol{B}_k^{(m)}(\boldsymbol{\Sigma}_{k|k-1}^{(\mathrm{ST},m)})^H \tag{5.39}$$

$$\hat{\boldsymbol{x}}_{k+1|k}^{(\mathrm{ST},m)} = \boldsymbol{A}^{(\mathrm{ST})}\hat{\boldsymbol{x}}_{k|k}^{(\mathrm{ST},m)} \tag{5.40}$$

$$\boldsymbol{\Sigma}_{k+1|k}^{(\mathrm{ST},m)} = \boldsymbol{A}^{(\mathrm{ST})}\boldsymbol{\Sigma}_{k|k}^{(\mathrm{ST},m)}(\boldsymbol{A}^{(\mathrm{ST})})^T + \boldsymbol{\Sigma}_{\nu}^{(\mathrm{ST})} \tag{5.41}$$

After the k^{th} ST beam tracking, beamformers are updated with predicted AoAs $\hat{\theta}_{k+1|k}^{(\mathrm{ST},m)}$ for $m = 1, \ldots, M$, which are the first elements of $\hat{\boldsymbol{x}}_{k+1|k}^{(\mathrm{ST},m)}$ for $m = 1, \ldots, M$.

5.5 Statistical Beamformer Construction

5.5.1 Full Dimensional GEB (FD-GEB)

GEB is a statistical beamformer whose performance is widely evaluated in [100] and [121]. Similar to the well-known Capon beamformer [99], GEB minimizes in-

terference while the intended signal power is kept constant. However, GEB accomplishes this task over variances as

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^H \boldsymbol{\Psi}_k^{(\mathrm{ST})} \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^H \boldsymbol{R}_k^{(\mathrm{ST},m)} \boldsymbol{w} = c \tag{5.42}$$

where $\boldsymbol{w}$ is a beamformer vector, and the terms are variances of $\boldsymbol{w}^H \boldsymbol{y}_{n,p}$ and $\boldsymbol{w}^H \boldsymbol{h}_p^{(m)}$, respectively. The covariance matrices can be calculated parametrically, and c is a constant. The solution to this problem requires generalized eigendecomposition of the matrix pair $(\boldsymbol{R}_k^{(\mathrm{ST},m)}, \boldsymbol{\Psi}_k^{(\mathrm{ST})})$, expressed as

$$\boldsymbol{R}_k^{(\mathrm{ST},m)} \boldsymbol{e} = \lambda \boldsymbol{\Psi}_k^{(\mathrm{ST})} \boldsymbol{e}. \tag{5.43}$$

Choosing $\boldsymbol{w}_{\text{opt}} = \boldsymbol{e}_{\max}$ solves the problem in (5.42) and yields the maximum SINR, where the generalized eigenvector $\boldsymbol{e}_{\max}$ corresponds to the maximum generalized eigenvalue λ_{max}. In addition, instead of using only the most dominant generalized eigenvector to reach a scalar output, the most dominant D_m eigenvectors can be used to reach a multidimensional output, i.e., a vector. In this case, the mutual information between the output and the intended part in the input is maximized for a given D_m, which means optimal dimension reduction [121].

FD-GEB is a single-stage beamformer where $\boldsymbol{T}^{(m)} \in \mathbb{C}^{N \times D_m}$ is constructed with the most dominant D_m generalized eigenvectors, using either a fully digital or a fully analog beamformer (with the double phase shifter structure).

5.5.2 Reduced Dimensional GEB (RD-GEB)

For HBFs, GEB can be implemented in the subspace created by an ABF $\boldsymbol{S}$ using the effective CCM pair $(\bar{\boldsymbol{R}}_k^{(\mathrm{ST},m)}, \bar{\boldsymbol{\Psi}}_k^{(\mathrm{ST})})$, given in (5.9) and (5.10), instead of full-dimensional CCM pair $(\boldsymbol{R}_k^{(\mathrm{ST},m)}, \boldsymbol{\Psi}_k^{(\mathrm{ST})})$. However, a slightly different technique is proposed, which is more robust and integrated with the design of the ABF $\boldsymbol{S}$. After the AoA estimates $\hat{\theta}_k^{(\mathrm{ST},m)}$ for $m = 1, \ldots, M$ are produced in $k-1^{\text{st}}$ ST beam tracking phase, ABF is formed by the selection of the R DFT bases which are directed to the closest angles to $\hat{\theta}_k^{(\mathrm{ST},m)}$. The DFT frequencies are

$$\phi_k = \frac{2\pi}{N} k, \quad k = 1, \ldots, N \tag{5.44}$$

and the k^{th} DFT basis is $\boldsymbol{u}(\phi_k)$ where $\boldsymbol{u}(\phi) \triangleq \frac{1}{\sqrt{N}}[1 \quad e^{j\phi} \quad \dots \quad e^{j(N-1)\phi}]^T$. Then, R/M DFT frequencies are selected for each m such that $\sum_{r=1}^{R/M} |\phi_{k_r^{(m)}} - \pi \sin(\hat{\theta}_k^{(\text{ST},m)})|$ is minimized, where $k_r^{(m)}$ are the indices for the selected set.[9] Then the ABF $\boldsymbol{S}$ is constructed as

$$\boldsymbol{S} = [\tilde{\boldsymbol{S}}^{(1)} \quad \dots \quad \tilde{\boldsymbol{S}}^{(M)}], \tag{5.45}$$

$$\tilde{\boldsymbol{S}}^{(m)} = [\boldsymbol{u}(\phi_{k_1^{(m)}}) \quad \dots \quad \boldsymbol{u}(\phi_{k_{R/M}^{(m)}})]. \tag{5.46}$$

It is assumed that each possible DFT basis selects an angular sector, such that $\boldsymbol{u}(\phi_k)$ selects the angular sector from θ_1 to θ_2, where $\pi \sin(\theta_1) = \phi_k - \pi/N$ and $\pi \sin(\theta_2) = \phi_k + \pi/N$. These angular sectors form a codebook of covariance matrices $\{\boldsymbol{C}_k\}$ for $k = 1, \dots, N$, where

$$\boldsymbol{C}_k \triangleq \frac{N}{2\pi} \int_{\phi_k - \pi/N}^{\phi_k + \pi/N} \boldsymbol{u}(\phi)\boldsymbol{u}^H(\phi) d\phi. \tag{5.47}$$

Using this codebook, approximated CCMs for the clusters are calculated as

$$\hat{\tilde{\boldsymbol{R}}}_k^{(\text{ST},m)} = \boldsymbol{S}^H \left(\frac{M}{R} \sum_{r=1}^{R/M} \boldsymbol{C}_{k_r^{(m)}} \right) \boldsymbol{S} \tag{5.48}$$

$$\hat{\tilde{\boldsymbol{\Psi}}}_k^{(\text{ST})} = \sum_{m=1}^{M} E^{(m)} \hat{\tilde{\boldsymbol{R}}}_k^{(\text{ST},m)} + N_0 \boldsymbol{I}_R \tag{5.49}$$

Then, the DBF $\boldsymbol{W}^{(m)}$ is obtained using the most dominant D_m generalized eigenvectors of the matrix pair $(\hat{\tilde{\boldsymbol{R}}}_k^{(\text{ST},m)}, \hat{\tilde{\boldsymbol{\Psi}}}_k^{(\text{ST})})$, which yields the total beamformer as $\boldsymbol{T}^{(m)} = \boldsymbol{S}\boldsymbol{W}^{(m)}$.

DBF is an intermediate stage of digital processing, which produces D_m-dimensional outputs. Therefore, the value of D_m might seem ineffective because it does not determine any hardware requirements like N and R, and it is neither about the input nor the output of the digital stage. However, it is effective on performance because it is the dimension of the output produced by the last statistical and slowly-updated part of the system, namely DBF. After DBF, ICS-CMF processes the D_m-dimensional signal in a more adaptive, fast-updated manner. Large values of D_m provide better coverage of the angular spread of a cluster, a longer ST-CPI and robustness due to the increased

[9] Note that $k_r^{(m)}$ are consecutive in r for an arbitrary m, that is $k_r^{(m)} = \bar{k}^{(m)} + r$ for some $\bar{k}^{(m)}$. Therefore, the selection is straightforward.

coverage, and a better representation of the angular characteristics of the channel for beam tracking. On the other hand, small values of D_m provide a better multi-user interference mitigation (higher SINR), and less complexity for IEC estimation and ICS-CMF due to the reduced dimension.

Columns of $\boldsymbol{W}^{(m)}$ (generalized eigenvectors) are not necessarily orthogonal, therefore they are orthogonalized via QR decomposition, assigning the obtained vectors as the new $\boldsymbol{W}^{(m)}$ such that $(\boldsymbol{W}^{(m)})^H \boldsymbol{W}^{(m)} = (\boldsymbol{T}^{(m)})^H \boldsymbol{T}^{(m)} = \boldsymbol{I}_{D_m}$. This process does not bring any information loss, and the range space of $\boldsymbol{W}^{(m)}$ is preserved. In addition, it has less computational complexity than eigendecomposition, and it provides numerical stability and notational simplicity.

5.6 Computational Complexity

Computational complexities of all the alternatives are tabulated in Table 5.3 in terms of the number of multiplications using $\mathcal{O}(\cdot)$ notation [120]. The dominant operations resulting in these complexities are also given in the third column, where MI, MM, and ED mean matrix inversion, matrix multiplication, and eigendecomposition, respectively. Also, * means a size reduction in inversion is possible and applied in favor of related techniques.

Roughly assume that $M \cong D_m \cong R/M < N_{\mathrm{F}} \ll N_\theta < N < P$ (see Table 5.4), where N_θ is the number of AoAs to be searched per cluster for search-based beam trackers. Accordingly, the computational complexity of techniques can be sorted as BA-LS < Joint LS $\ll$ Joint MMSE for instantaneous channel estimation, SEKF < BA-ML $\ll$ orthogonal matching pursuit (OMP) for beam tracking, and RD-GEB $\ll$ FD/RD MMSE BF < FD-GEB for beamformer construction.

Consequently, it is clearly seen that the proposed per-cluster estimation approach with statistical beamforming, represented by BA-LS, BA-ML, SEKF, and GEB, yields very low computational complexity compared with the conventional joint estimation approach with instantaneous beamforming, represented by joint LS, joint MMSE, OMP, and MMSE BF. Especially for beam tracking and beamformer construction, the proposed techniques (BA-LS, BA-ML, and SEKF) reduce the computational com-

Table 5.3: Average Computational Complexity per FT-CPI

Task	Technique	# of Multiplications	Dominant Op.
Inst. Eff. Channel Estimation	BA-LS	$\mathcal{O}(MN_{\text{F}}D_m)$	(5.17)
	Joint LS	$\mathcal{O}(M^3) + \mathcal{O}(MN_{\text{F}}R)$	MI* & MM
	Joint MMSE	$\mathcal{O}(M^3R^3) + \mathcal{O}(M^2N_{\text{F}}R^3)$	MI* & MM
Beam Tracking	BA-ML	$\mathcal{O}(MN_\theta ND_m^2/P)$ $+\mathcal{O}(MD_m^2 + MN_\theta D_m^3/P)$	(5.26), (5.31)
	SEKF	$\mathcal{O}(MND_m^2/P)$ $+\mathcal{O}(MD_m^2) + \mathcal{O}(MD_m^4/P)$	(5.26), (5.32), (5.37)*
	OMP [120]	$\mathcal{O}(MN_\theta RN_{\text{F}})$	Step 6 in Alg. 1 [120]
Beamformer Construction	FD-GEB	$\mathcal{O}(MN^3/P)$	ED
	RD-GEB	$\mathcal{O}(MR^3/P)$	ED
	FD-MMSE BF	$\mathcal{O}(M^3 + NM^2)$	MI* & MM
	RD-MMSE BF	$\mathcal{O}(M^3 + RM^2)$	MI* & MM

plexity by at least a factor of P (with P typically set to 1000 in the results section) compared to conventional alternatives.

5.7 Numerical Evaluations

5.7.1 Auxiliary Definitions

For comparison purposes in numerical results, some auxiliary definitions are needed. Firstly, the prefix *RD* (reduced dimensional) implies that the DBF is working after an ABF in an HBF system, whereas *FD* (full dimensional, or fully digital) refers to a system where the beamformer is fully digital and there is no ABF, therefore $\boldsymbol{S} = \boldsymbol{I}_N$ and there are $R = N$ RFCs.

DFT beamformer has the same ABF $\boldsymbol{S}$ as RD-GEB. However, the DBF $\boldsymbol{W}^{(m)}$ is a matrix of ones and zeros, only selecting the RFCs (columns of $\boldsymbol{S}$) for the m^{th} cluster in (5.45) so that $\boldsymbol{T}^{(m)} = \boldsymbol{S}\boldsymbol{W}^{(m)} = \tilde{\boldsymbol{S}}^{(m)}$. This setting of DBF can be seen as the most primitive attempt to create cluster-specific subspaces after ABF. The improvement brought by RD-GEB can be seen from its difference from DFT BF.

The *periodogram* is a spectral estimation method [99], which is the implicit basis for a variety of techniques in the beam tracking literature. It is applied in the range space of the DFT beamformer in the case of HBF structures. The peak of $\rho(\theta) \triangleq \sum_{n,p} |\boldsymbol{a}(\theta)^H \tilde{\boldsymbol{S}}^{(m)} \boldsymbol{z}_{n,p}^{(m)}|^2$ is searched for the m^{th} cluster, where $\boldsymbol{z}_{n,p}^{(m)} = (\tilde{\boldsymbol{S}}^{(m)})^H \boldsymbol{y}_{n,p}$. Inspecting (5.46), note that $\rho(\theta)$ measures the power at the RFC outputs one by one at certain θ values. Therefore, it represents the related studies in Section 1.2.1. Considering all the training phases in an ST-CPI are used, the summation above is taken from PN_{F} samples.

MMSE beamformer is another beamforming method, different than MMSE channel estimator, for which an MMSE estimator [114] is constructed where symbols $s_{n,p}^{(m)}$ for $m = 1, \ldots, M$ are treated as the multiple unknown parameters with the observation $\boldsymbol{y}_{n,p}$ in (5.1) or $\boldsymbol{r}_{n,p}$ in (5.8), depending on whether it is FD or RD. The channels $\boldsymbol{h}_p^{(m)}$ or $\bar{\boldsymbol{h}}_p^{(m)}$ are combiners of unknown parameters and should be known. That is why the MMSE beamformer is categorized as an *instantaneous beamformer* since it should be updated with each new instantaneous channel estimate at the rate of channel decorrelation (i.e., once in an FT-CPI), after each FT training.

There are two modes related to beam tracking, the *Self-Driven* mode and the *Genie-Aided BF* mode. In the proposed system, IEC estimators and beam trackers operate after the beamformers, which are constructed via the previous AoA estimates. Therefore, all the performance measures depend on the previous AoA estimate, which describes the *Self-Driven* mode. On the other hand, the *Genie-Aided BF* mode removes the effect of the previous AoA estimate for performance analysis purposes. The beamformers are constructed with true AoAs at the beginning of each ST-CPI, and all the performance measures are collected at the end of each ST-CPI, including angular root mean squared error (RMSE) from beam trackers, although their AoA estimates will not be used in the next ST-CPI. Note that this does not mean that the mismatches due to the movement inside the ST-CPIs are eliminated.

5.7.2 Performance Measures

5.7.2.1 Average NMSE for BA-LS

The performance of IEC estimation is measured by normalized mean squared error (NMSE), which is defined for the m^{th} cluster as[10]

$$\text{NMSE}_p^{(m)} \triangleq \frac{\mathbb{E}\{||\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} - \tilde{\boldsymbol{h}}_p^{(m,m)}||_2^2\}}{\mathbb{E}\{||\tilde{\boldsymbol{h}}_p^{(m,m)}||_2^2\}}. \tag{5.50}$$

It is shown in Appendix D.3 that the NMSE can be calculated as

$$\text{NMSE}_p^{(m)} = \frac{\text{tr}\left(\tilde{\boldsymbol{\Psi}}_p^{(m)} - E^{(m)}\tilde{\boldsymbol{R}}_p^{(m,m)}\right)}{E^{(m)}N_{\text{F}}\,\text{tr}\left(\tilde{\boldsymbol{R}}_p^{(m,m)}\right)}. \tag{5.51}$$

Then, the *average NMSE* is found by averaging $\text{NMSE}_p^{(m)}$ through all the possible p instants in Monte-Carlo experiments.

5.7.2.2 Angular RMSE for Beam Tracking

Angular error is defined as $\hat{\theta}_k^{(\text{ST},m)} - \theta_{(k-1)P+1}^{(m)}$, and the angular RMSE is

$$\text{RMSE}^{(m)} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\hat{\theta}_k^{(\text{ST},m)} - \theta_{(k-1)P+1}^{(m)}\right)^2} \tag{5.52}$$

for the m^{th} cluster, where K is the number of all the possible slow-time instants in Monte-Carlo experiments. Note that the error definition neglects the variation inside the ST-CPI and focuses only on the estimation performance.

5.7.2.3 Average SINR after ICS-CMF

The overall performance of the ABF, DBF, beam tracker, and IEC estimator will be measured via SINR at the output of ICS-CMF, which is the symbol estimate $\hat{s}_{n,p}^{(m)}$ given in (5.15). Since the estimated channel is known to the detector rather than the true channel, the true signal term is taken as $S_p^{(m)} \triangleq \sqrt{E^{(m)}}(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}s_{n,p}^{(m)}$,

[10] Note that the expectations both in SINR and NMSE calculations are FT expectations that are conditioned on the settings at the related FT-CPI, such as beamformers and positions.

Table 5.4: Simulation Parameters

Parameter	Description	Value / Details
N	Number of antennas	128
M	Number of clusters	4
U	Number of users	4
N_{F}	FT training sequence length	10
N_{S}	Data mode length in an FT-CPI	990
T_{F}	Duration of an FT-CPI	1e-5 seconds
P	Number of FT-CPIs	$\sim$ 1000
R	Number of RFCs	16
D_m	Number of DBF outputs	3
$\theta_0^{(m)}$	Initial AoAs	$\{10^\circ,20^\circ,-10^\circ,-20^\circ\}$
$\Delta^{(m)}$	ASs	$\{3^\circ,3^\circ,3^\circ,3^\circ\}$
$E^{(m)}/N_0$	Cluster SNRs	{10,40,30,30} dB

and the interference-plus-noise term $N_p^{(m)} = \hat{s}_{n,p}^{(m)} - S^{(m)}$ is the remaining part. With these definitions, SINR for the m^{th} cluster is defined as

$$\mathrm{SINR}_p^{(m)} \triangleq \mathbb{E}\left\{\frac{|S_p^{(m)}|^2}{|N_p^{(m)}|^2}\right\}. \tag{5.53}$$

Then, the *average SINR* is found by averaging $\mathrm{SINR}_p^{(m)}$ through all the possible p instants in Monte-Carlo experiments.

5.7.3 Simulation Settings

Selected parameters for simulations are listed in Table 5.4. One FT-CPI consists of $N_{\mathrm{F}} + N_{\mathrm{S}} = 1000$ symbols, which is suitable according to Table 5.2, and it takes $T_{\mathrm{F}} = 10\mu s$ assuming a bandwidth of 100 MHz. Selection of $N_{\mathrm{F}} = 10$ and $N_{\mathrm{S}} = 990$ yields 1% training overhead.[11]

There are $M = 4$ signal clusters. The first cluster, which will be the most interested one, suffers from a significant near-far effect since it is surrounded by 20 to 30 dB

[11] $T_{\mathrm{F}} = 10\mu s$ approximately corresponds to one OFDM symbol duration with 120 kHz subcarrier spacing described by 3GPP standards for millimeter wave channels [122].

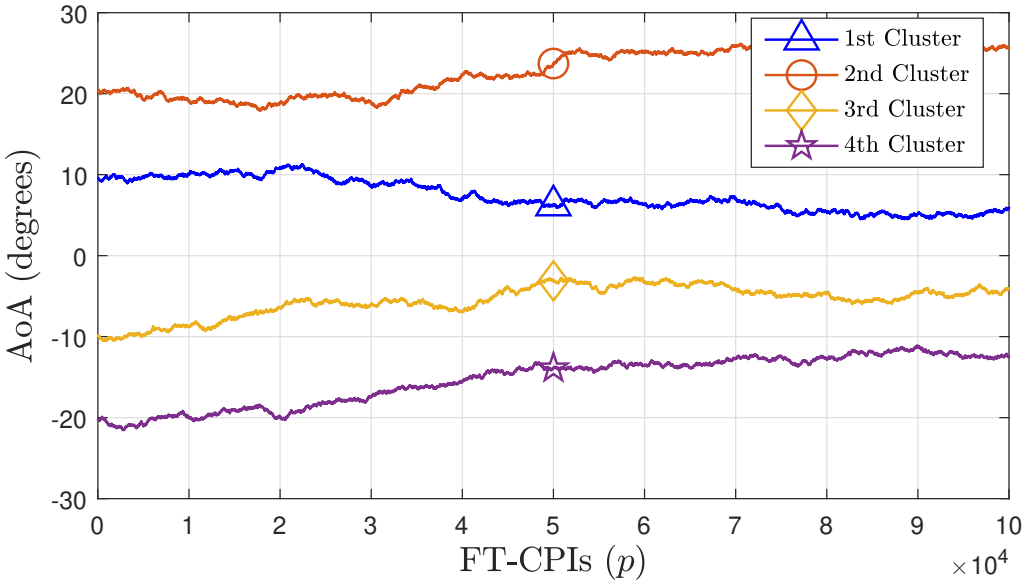


Figure 5.6: One realization of AoAs from the movement model.

stronger clusters. The movement model given in Section 5.2.3 is implemented with $\sigma_\theta^2 = 1.45e-4$, $\sigma_\omega^2 = 1.46e-6$, and initial AoAs given in Table 5.4. This setting of variances corresponds to 10 m/s speed and 1 m/s^2 acceleration assuming 150 meters distance from the BS after 1 second of random movement.[12] One example of the resultant movement process is plotted in Fig. 5.6 for a duration of 1 second.[13]

Parameters D_m and R/M should be comparable with the practical rank of effective CCMs $\tilde{\boldsymbol{R}}_p^{(m,m)}$, which is determined by the AS $\Delta^{(m)}$ and the number of antennas N. If the ratio R/N is limited, the number of active users can be reduced by user scheduling, or D_m and R/M can be reduced by treating the AS $\Delta^{(m)}$ as a design parameter. The effects of the value of D_m on the beampattern are shown in Fig. 5.7. As small values prevent adequate coverage of the intended sector, large values cause interference leakage. The selected value $D_m = 3$ provides 50 dB interference mitigation.

For each cluster, only the selected region by ABF $\tilde{\boldsymbol{S}}^{(m)}$, which is discussed below (5.45), is searched for BA-ML, OMP [120], and the periodogram with a resolution of

[12] The standard deviation for the angular position (speed) divided by time is interpreted as speed (acceleration), in the innovation covariance matrix after 100.000 FT-CPIs (1 second), which is computed similarly to $\boldsymbol{\Sigma}_\nu^{(\mathrm{ST})}$ in Section 5.4.2.1.

[13] Note that the selection of $T_\mathrm{F} = 10\mu s$ with the assumption of 100 MHz bandwidth is just to give an insight into a practical operation. T_F only affects the movement model, and variances would be changed for different T_F and bandwidth just to have a similar movement to the one in Fig. 5.6.

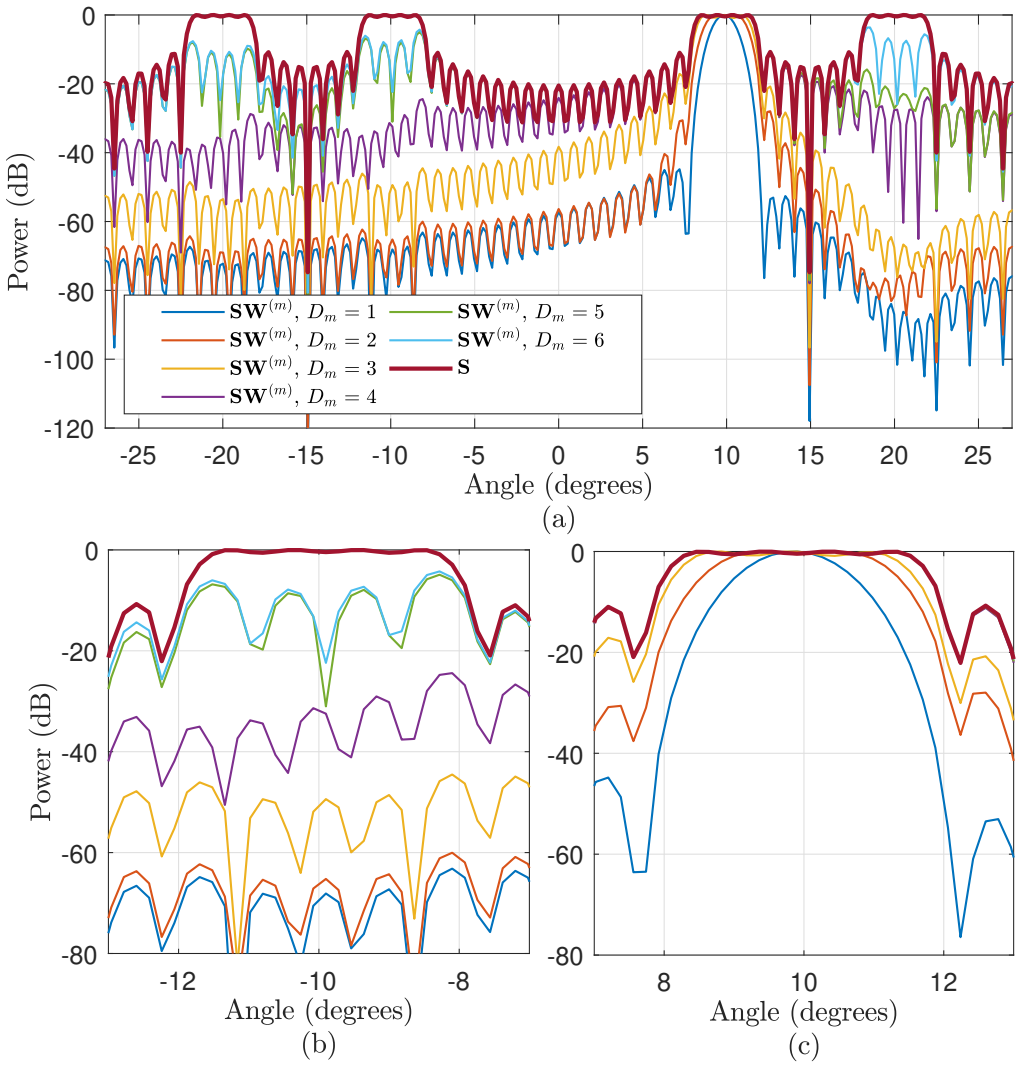


Figure 5.7: Effect of D_m on the beam pattern after DBF for the first cluster ($m = 1$). (a) All clusters, (b) interference area, (c) intended area.

0.1°, which yields $N_\theta \cong 40$. $D'_m = 2$ is chosen for BA-ML.

5.7.4 Numerical Results

In this section, numerical results from the simulations will be shared. In the simulations, $p_{\max}$ FT-CPIs of movement period, whose one realization for $p_{\max} = 1e5$ is given in Fig. 5.6, is repeated many times to have a Monte-Carlo experiment.[14] In

[14] If any two clusters get closer than 3°, or any cluster gets outside of $[-60^\circ, 60^\circ]$ region, related realization is ended before $p_{\max}$ FT-CPIs.

the simulations, the parameters given in Section 5.7.3 are used, and the performance measures given in Section 5.7.2 are collected.

As a beginning, the motivation behind the proposed per-cluster estimation approach with statistical beamforming will be clarified via comparisons with conventional techniques in Figures 5.8 and 5.9. In these figures, mobility and beam tracking are temporarily left out of focus with the settings of Genie-Aided BF mode and $p_{\max} = 1000$, for which mobility is almost absent as seen in Fig. 5.6. In Fig. 5.8, statistical beamforming (FD-GEB, RD-GEB, and DFT BF) and instantaneous beamforming (FD/RD MMSE BF) are compared according to beamformer update time P, which work with second-order statistics (CCMs) and actual channel estimates, respectively. As expected, slow-time alternatives are robust against P, while MMSE beamformers fail immediately when $P > 1$, since the actual channel decorrelates, but statistics remain almost the same. Although the best performance is reached by MMSE beamformers using MMSE channel estimates with $P = 1$, the optimum choice in terms of computational complexity per time is statistical beamformers with a large P. The first scheme needs MMSE channel estimates for each FT-CPI, which is computationally complex as seen from Table 5.3, and the second scheme needs channel estimates from BA-LS, which is simpler. In addition, RD-GEB has less time-averaged complexity than the RD MMSE beamformer. The simpler alternative of channel estimation for instantaneous beamforming, joint LS, causes a performance loss. Finally, the RD alternatives, which are suitable for HBFs, exhibit minimal performance loss compared with FD alternatives, which are structurally and computationally more complex and needy.

In Fig. 5.9, instantaneous channel estimation methods are compared against signal-to-noise ratio (SNR). As discussed in Section 5.6, BA-LS is computationally simpler than the conventional techniques of joint LS and joint MMSE. Although simpler, it is seen in Fig. 5.9 that BA-LS does not lose performance compared with the superior technique of joint MMSE.[15] This high performance of BA-LS arises from the prior processing via DBF, where GEB is employed. The role of GEB is seen from the performance loss of BA-LS with DFT BF, compared to BA-LS with the RD-GEB case.

[15] The vertical axis is *normalized* MSE, and the reason for BA-LS to seem unexpectedly better than joint MMSE is the difference in the sizes of the channels, which are $D_m, D_m, R/M, N, N, R, R$ in the order of the legend.

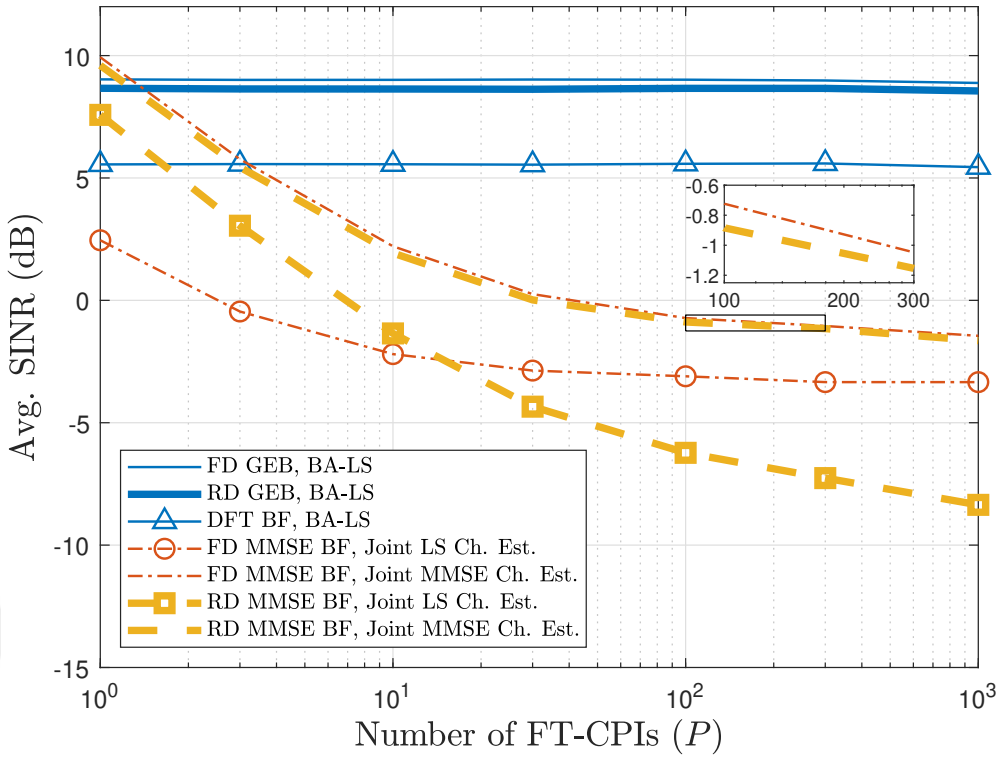


Figure 5.8: Comparison of instantaneous and statistical beamforming. ($m = 1$, $p_{\max} = 1000$, Genie-Aided mode.)

In conclusion, BA-LS with GEB is a very efficient technique in terms of performance and computational complexity.

The Self-Driven mode, which is defined in Section 5.7.1, might suffer from instability. Fig. 5.10 depicts the angular RMSE when the beamformers are constructed with a given error in the AoA. In this regard, Fig. 5.10 constitutes a basis for the stability of self-driven operation. For a stable self-driven performance, the next angular error is desired to be smaller than the given error, which causes a given error to gradually get smaller and become zero. The region where this behavior is expected is called the *lock-in region*, as shown in the figure. First of all, it is seen that the error in the next estimate is step-like against the previously given error, which is due to discrete angles in the definition of ABF and discretized angular sectors in the definition of RD-GEB. Second, note that the AS is 3 degrees, and $R/M = 4$ RFCs of ABF create a beamwidth of approximately 3.5 degrees. Therefore, a given error larger than 3 or less than -3 means total beam loss. It is seen in Fig. 5.10 that BA-ML and OMP offer

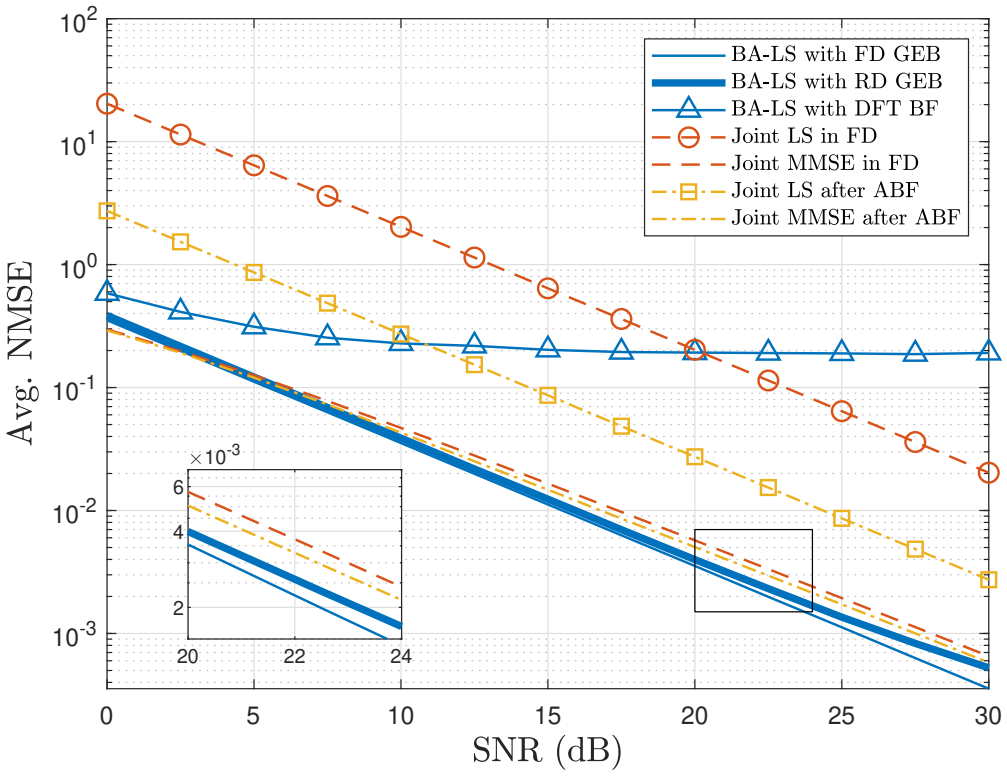


Figure 5.9: Comparison of instantaneous channel estimation techniques. ($m = 1$, $p_{\max} = 1000$, Genie-Aided mode.)

the most robust performance when there is no beam loss. SEKF also remains in the lock-in region; however, the error due to linearization in (5.35) and (5.36) increases with the given angular error. Periodogram fails due to the near-far effect because the cluster $m = 2$ is at 20 degrees, which is 30 dB stronger, and estimates deviate towards this cluster. Lastly, the OMP shows nearly 0.5 degrees of RMSE when the given angular error is zero. This is because OMP actually searches for a point target, but BA-ML and SEKF search for a cluster with some AS through their modeling by (5.30) and (5.33). Therefore, OMP appoints the first powerful AoA candidate inside the AS as an estimate, but others find the center of the observed cluster.

In Figures 5.11, 5.12, and 5.13, the performances of various angular estimators, the proposed IEC estimator BA-LS, and the overall performance measure SINR will be shared. Although some results for the Genie-Aided BF mode are also given for comparison, the main focus is on the performance of the Self-Driven mode with the setting of $p_{\max} = 100000$, which leads to more angular variation. In Fig. 5.11, tracking be-

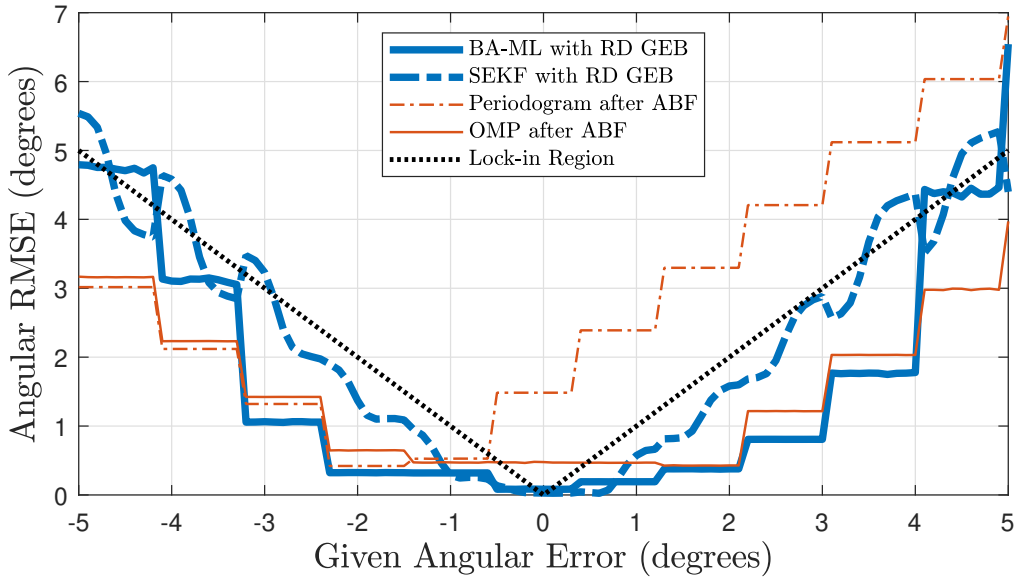


Figure 5.10: RMS error of angular estimates at the end of an ST-CPI depending on a given error upon which beamformers are constructed at the beginning of the ST-CPI. ($m = 1$, $P = 1000$.)

haviors of different angular estimators are given. This figure shows more clearly that the actual AoAs are changing after each FT-CPI, but estimates are produced after every P FT-CPIs, or an ST-CPI. Inside an ST-CPI, beamformers $\boldsymbol{S}$ and $\boldsymbol{W}^{(m)}$, which are constructed upon these estimates, remain unchanged. It is also seen that the periodogram immediately deviates towards the stronger neighboring cluster, while others are robust to the near-far effect.

In Fig. 5.12, cumulative distribution functions (CDFs) of angular error for different beam trackers are given. It is seen that the proposed estimators SEKF and BA-ML outperform the alternatives, with 90% of estimates being smaller than 0.5 degrees when operated with FD or RD-GEB. Comparing these with OMP after ABF, the contribution of applying cluster-specific DBFs before the estimation (i.e., the per-cluster approach) is seen. With this approach, the complexity is decreased and performance is enhanced.

In Fig. 5.13, the CDF of NMSE of the proposed instantaneous channel estimation technique BA-LS is plotted. In this figure, the effects of various beamforming and beam tracking techniques on the instantaneous channel estimation accuracy are shown.

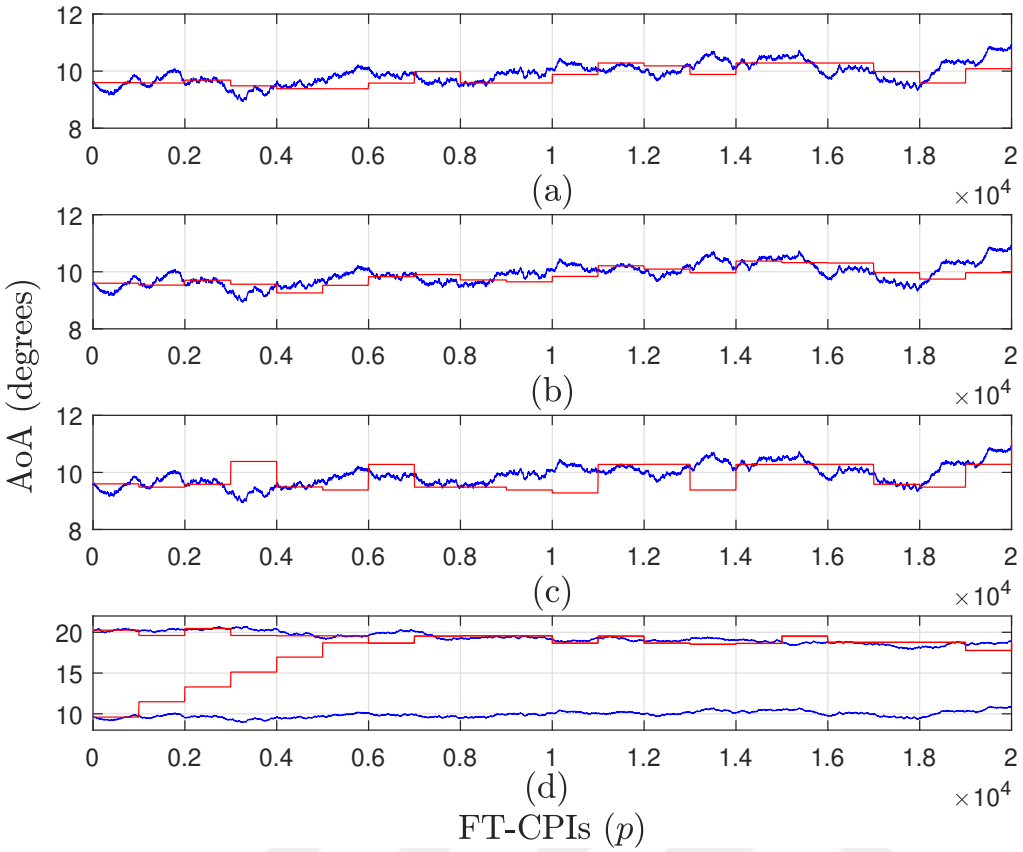


Figure 5.11: Tracking behaviors of angular estimation techniques: (a) BA-ML with RD-GEB, (b) SEKF with RD-GEB, (c) OMP with ABF, (d) Periodogram with ABF (Blue lines: true AoA, red lines: estimates, $p_{\max} = 100000$, $P = 1000$, Self-Driven mode.)

It is seen that the GEB alternatives used together with the proposed BA-ML and SEKF beam trackers, which are the representatives of the per-cluster estimation approach, result in NMSE values between 0.035 and 0.05. Note that these numbers are lower-bounded by $(\text{SNR} \times N_{\text{F}})^{-1}$, which is 0.01 for $m = 1$ and $N_{\text{F}} = 10$. It is seen that RD-GEB works with minimal performance loss compared with FD-GEB, and proposed BA-ML and SEKF beam trackers yield NMSE performance similar to Genie-Aided BF mode. Finally, the huge advantage of per-cluster operation is seen from the difference between RD-GEB and DFT BF, which are different only in the design of $\boldsymbol{W}^{(m)}$.

Next, Fig. 5.14 shows angular RMSE and SINR performances against the design parameter P, the number of FT-CPIs in an ST-CPI. There is a performance loss at large

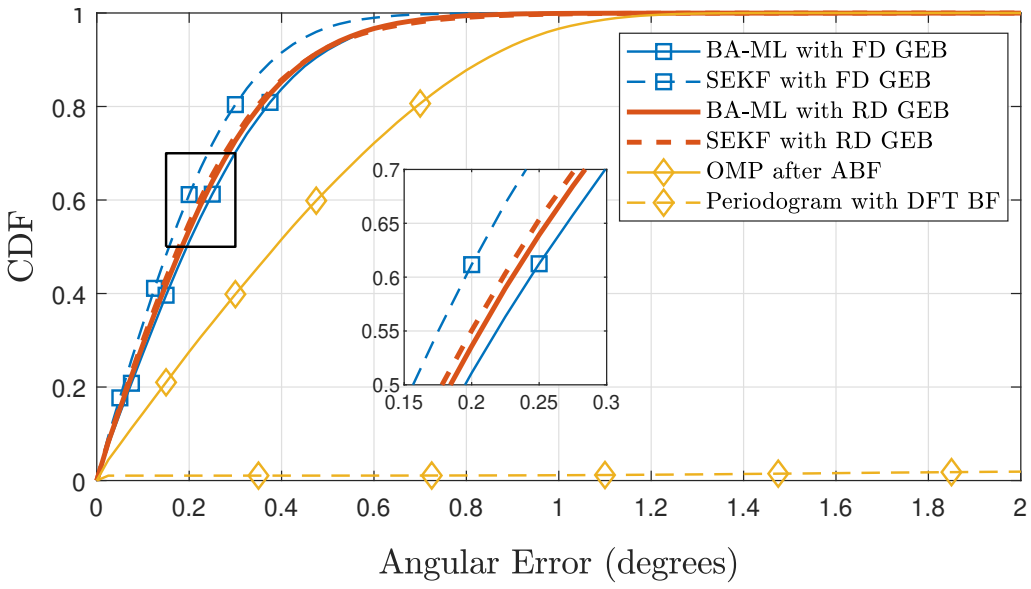


Figure 5.12: CDF of error in the angular estimates. ($m = 1$, $p_{\max} = 100000$, $P = 1000$, Self-Driven mode.)

P for all the alternatives due to the loss of angular coherency because of mobility. For small P, BA-ML performance degrades due to the decrease in the observation period. However, it is seen that SEKF is not affected by this factor since it is a Kalman filter variant that can collect information from the previous ST-CPIs. Therefore, SEKF can succeed even with the primitive DFT BF, but DFT BF fails in beamforming and channel estimation as seen in subplot (b) and also in Fig. 5.13. Finally, it is seen that the BA-ML and SEKF methods in Self-Driven mode perform similarly to the Genie-Aided mode. Also, the RD alternatives with hybrid beamforming perform with minimal loss compared with the structurally complex and needy FD alternatives. In the literature, there are studies that take the FT-CPI longer about 0.125 ms, considering the slot definition in 3GPP standards [122], which includes 14 OFDM symbols. Fig. 5.15 shows the performance of the system for $T_{\mathrm{F}} = 100\mu s$. For this longer FT-CPI with the same speed, there are fewer IEC estimates that can be used before the angular coherency is lost. Therefore, a slight increase in the angular RMSE is seen for BA-ML at the optimum P value, which is 100. However, this increase does not degrade the overall performance.

Finally, the possible effects of the speed of the sources are investigated in Fig. 5.16. For this figure, the system operates with an FT-CPI and ST-CPI, which establishes

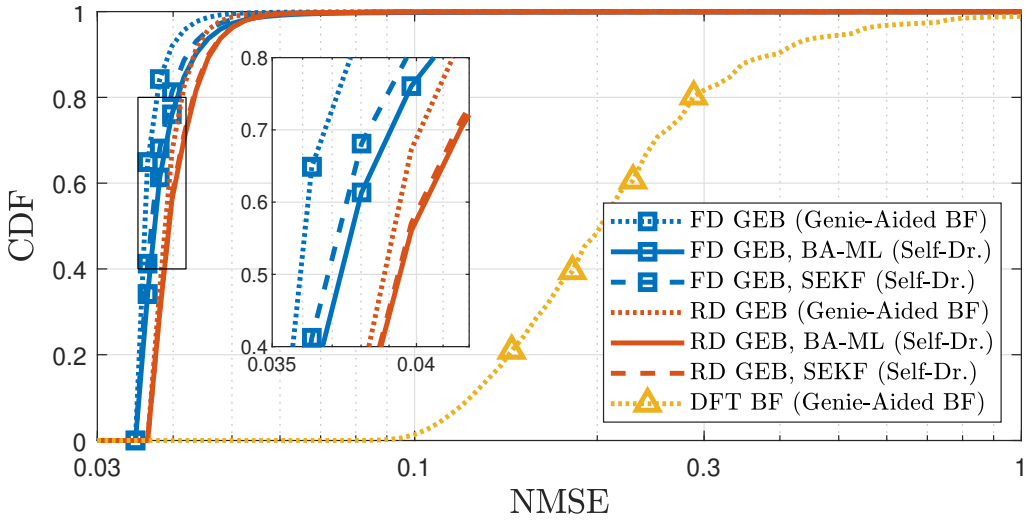


Figure 5.13: CDF of NMSE for BA-LS with different beamformers. ($m = 1$, $p_{\max} = 100000$, $P = 1000$.)

an assumption on the speed v, and some sources do not fit into this assumption. In addition, SEKF also assumes a speed due to its model-based structure. Note that the mentioned speed is the standard deviation due to random movement, which means a range of speeds up to $3v$ can be considered for a v value within the 3-sigma rule. The relation to the angular variation is established considering a 150 m distance. Fig. 5.16 shows that the ST-CPI length can be decreased in more rapidly varying scenarios, which increases the complexity. For $P = 100$ cases, the performance is highly robust to the change in the speeds of the users. On the other hand, the SEKF cases with different assumptions of speed perform almost the same. Therefore, SEKF is not very sensitive to the assumption of the speed, although it is model-based.

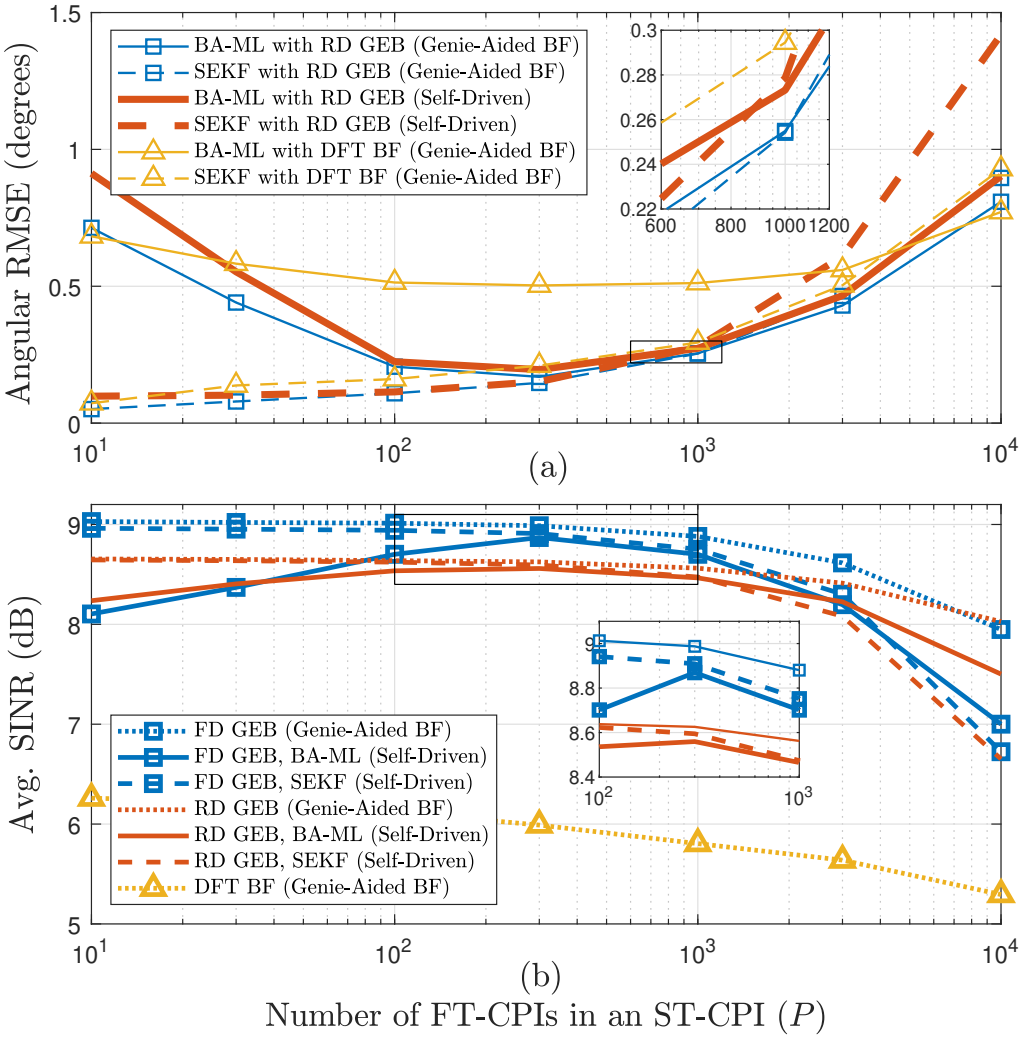


Figure 5.14: Performances against the length of an ST-CPI: (a) RMS error for angular estimators, (b) Average SINR. ($m = 1$, $p_{\max} = 100000$.)

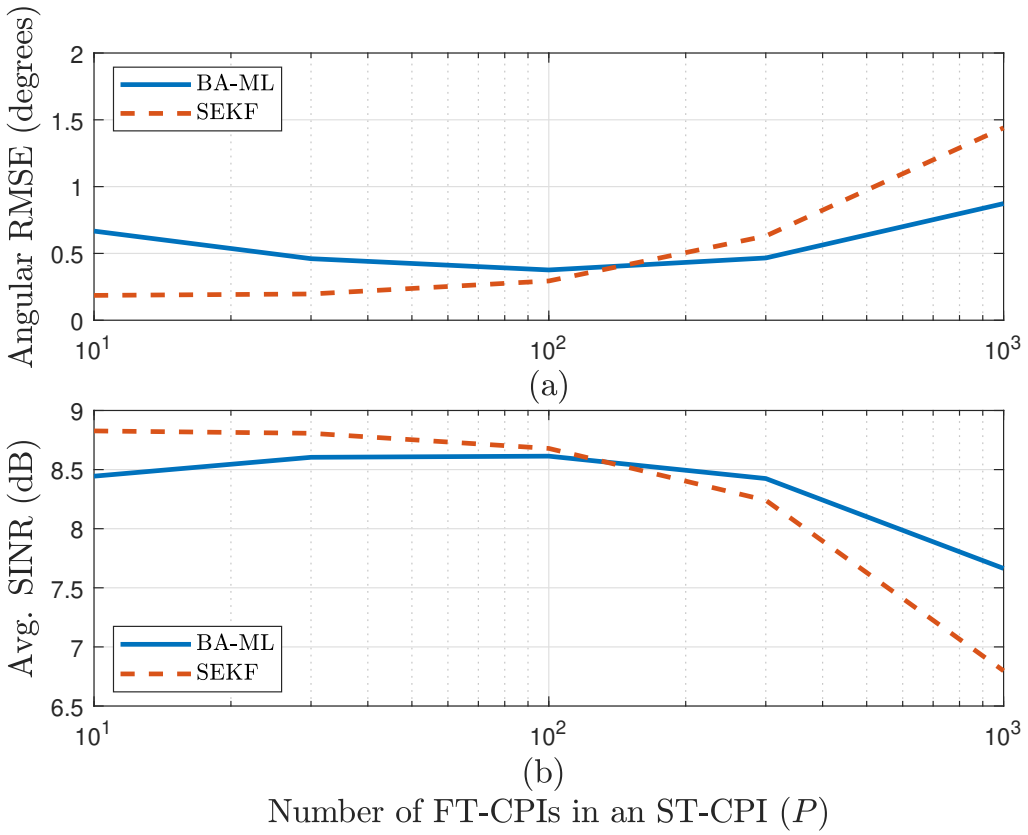


Figure 5.15: Performances against the length of an ST-CPI: (a) RMS error for angular estimators, (b) Average SINR. ($m = 1$, $p_{\text{max}} = 100000$, $T_{\text{F}} = 100\mu s$, RD-GEB, Self-Driven BF.)

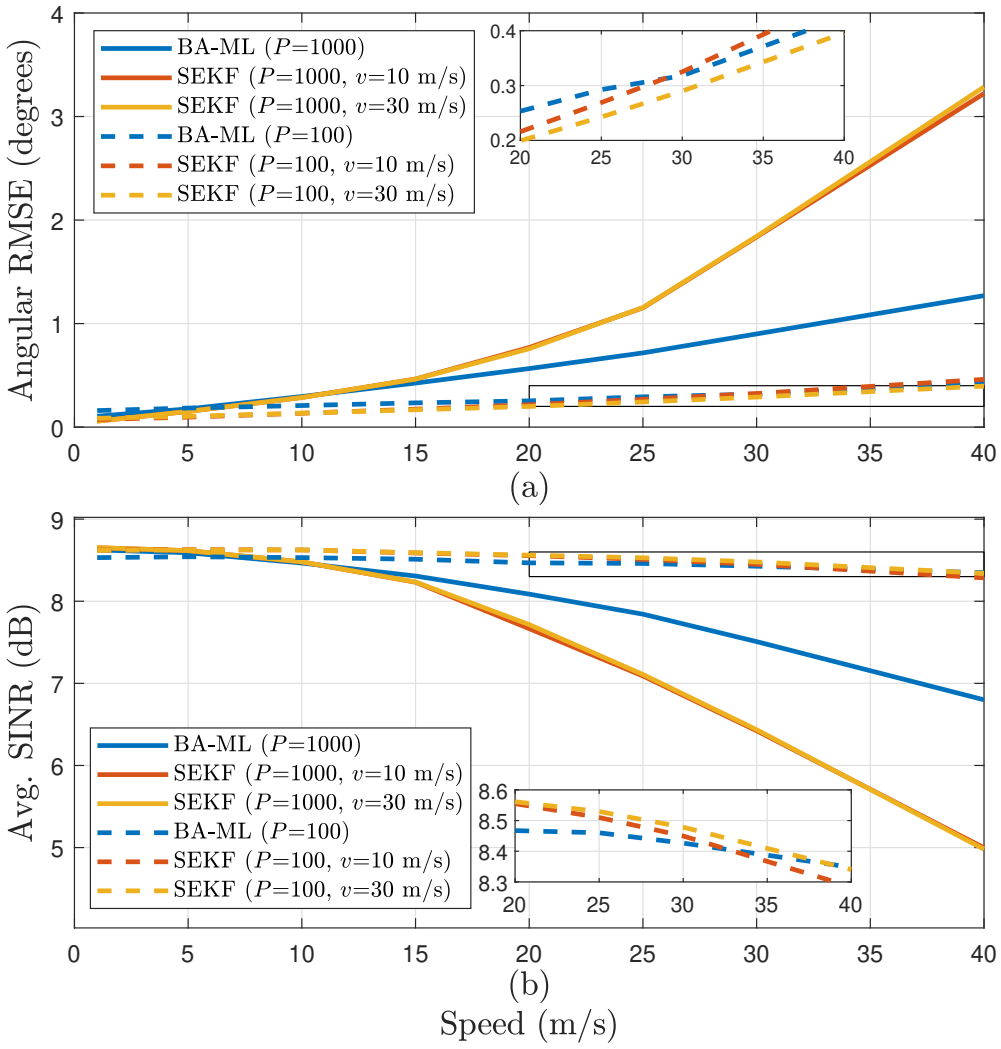


Figure 5.16: Performances against the speed of the sources: (a) RMS error for angular estimators, (b) Average SINR. ($m = 1$, $p_{\max} = 100000$, RD-GEB, Self-Driven BF.)

CHAPTER 6

CONCLUSIONS

In this thesis, sensing, beamforming, and self-interference mitigation for full-duplex communications are investigated. In Chapter 3, the SIC operation of FD communication is assessed under time varying SI channels. While power asymmetry between SoI and SI prompts high accuracy of SIC, the need for NL modeling for SI prediction creates a significant complexity burden. These challenges are amplified under time variation, which introduces the problem of SI estimation under high SoI contamination together with the trade-offs between observation time and coherence, namely, channel aging and training overhead. It is shown that FD communication is very sensitive to time variation, such that even a small amount of variation degrades the performance severely. These challenges are presented and discussed based on analytical and experimental results. As a remedy, an efficient, adaptive, and iterative Turbo-DSIC algorithm is proposed, and its performance is evaluated via both synthetic and hardware data. It is shown that the proposed algorithm solves the SI prediction under the SoI contamination problem by iterative SoI cancellation. Furthermore, it overcomes the aforementioned trade-offs by introducing a new processing strategy, and reduces the DSIC complexity through a novel two-stage signal decomposition approach for NL SI prediction.

The effect of FFSI on FD MIMO communications is investigated in Chapter 4. It is observed that the FFSI also degrades the performance significantly, while the NFSI is more extensively studied in the literature. For SI mitigation, DSIC and spatial suppression are examined, and their disadvantages are highlighted. The proposed selection policy bypasses the disadvantages of both by selecting the most suitable mitigation action for each scatterer based on derived metrics. Furthermore, models

and methods are based on the angle and delay parameters of scatterers, collected in a scatterer map. The maintenance of the scatterer map is addressed, with a particular focus on newly emerging scatterers. It is shown that if an emerging scatterer degrades the performance, besides being detected and mitigated successfully, its delay and angle are accurately estimated to be added to the scatterer map, exhibiting a sensing-for-communication case.

In Chapter 5, a novel and low-complexity channel estimation framework is proposed, which includes IEC estimation and beam tracking. This framework, namely the per-cluster estimation with statistical beamforming, leverages the angular sparsity in multi-user channels that is typical for millimeter-wave communications. If the number of RF chains is adequate, each signal cluster is processed individually by angular division, which provides a huge advantage in terms of interference mitigation. Proposed beam tracking methods BA-ML and SEKF use only a collection of readily available channel estimates from the recent estimation phases, without any additional sounding. Therefore, the study has shown that the beam trackers can operate at much slower rates than instantaneous channel estimation. The proposed scheme is also shown to be stable in self-driven operation so that it can recover from instantaneous beam errors. Due to its angular sensing and interference-mitigation ability, the proposed system is promising for integrated sensing and communications (ISAC), and also for the self-interference cancellation (SIC) task in full-duplex communications.

REFERENCES

[1] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, 2014.

[2] G. Liu, F. R. Yu, H. Ji, V. C. M. Leung, and X. Li, "In-Band Full-Duplex Relaying: A Survey, Research Issues and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 500–524, 2015.

[3] H. Li, J. Van Kerrebrouck, O. Caytan, H. Rogier, J. Bauwelinck, P. Demeester, and G. Torfs, "Self-Interference Cancellation Enabling High-Throughput Short-Reach Wireless Full-Duplex Communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6475–6486, 2018.

[4] H. Şenol, X. Li, and C. Tepedelenlioğlu, "Rapidly Time-Varying Channel Estimation for Full-Duplex Amplify-and-Forward One-Way Relay Networks," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 3056–3069, 2018.

[5] T. Miyajima, H. Teshiromori, and Y. Sugitani, "Adaptive Self-Interference Suppression for Full Duplex Filter-and-Forward Relaying," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1701–1704, 2020.

[6] Z. Liu, S. Aditya, H. Li, and B. Clerckx, "Joint transmit and receive beamforming design in full-duplex integrated sensing and communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2907–2919, 2023.

[7] B. Smida, A. Sabharwal, G. Fodor, G. C. Alexandropoulos, H. A. Suraweera, and C.-B. Chae, "Full-duplex wireless for 6G: Progress brings new opportunities and challenges," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2729–2750, 2023.

[8] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks:

challenges and research advances," *IEEE Network*, vol. 28, no. 6, pp. 6–11, 2014.

[9] C. K. Sheemar, S. Chatzinotas, D. Slock, E. Lagunas, and J. Querol, "Parallel and distributed hybrid beamforming for multicell millimeter wave MIMO full duplex," *IEEE Trans. Veh. Technol.*, pp. 1–16, 2024.

[10] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surv. Tut.*, vol. 17, no. 4, pp. 2017–2046, 2015.

[11] M. Sattari, H. Guo, D. Gündüz, A. Panahi, and T. Svensson, "Full-duplex millimeter wave MIMO channel estimation: A neural network approach," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 1093–1108, 2024.

[12] D. Korpi, T. Huusari, Y.-S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 286–290, 2015.

[13] M. Heino, D. Korpi, T. Huusari, E. Antonio-Rodriguez, S. Venkatasubramanian, T. Riihonen, L. Anttila, C. Icheln, K. Haneda, R. Wichman, and M. Valkama, "Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 91–101, 2015.

[14] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous Transmission and Reception: Algorithm, Design and System Level Performance," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 5992–6010, 2013.

[15] D. Korpi, Y.-S. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, "Adaptive Nonlinear Digital Self-Interference Cancellation for Mobile Inband Full-Duplex Radio: Algorithms and RF Measurements," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–7, 2015.

[16] D. Liu, Y. Shen, S. Shao, Y. Tang, and Y. Gong, "On the analog self-interference cancellation for full-duplex communications with imperfect channel state information," *IEEE Access*, vol. 5, pp. 9277–9290, 2017.

[17] A. Kiayani, M. Z. Waheed, L. Anttila, M. Abdelaziz, D. Korpi, V. Syrjälä, M. Kosunen, K. Stadius, J. Ryynänen, and M. Valkama, "Adaptive nonlinear RF cancellation for improved isolation in simultaneous transmit–receive systems," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 5, pp. 2299–2312, 2018.

[18] D. Korpi, J. Tamminen, M. Turunen, T. Huusari, Y.-S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Full-duplex mobile device: pushing the limits," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 80–87, 2016.

[19] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, 2014.

[20] Y. Chen, C. Ding, Y. Jia, and Y. Liu, "Antenna/propagation domain self-interference cancellation (SIC) for in-band full-duplex wireless communication systems," *Sensors*, vol. 22, no. 5, 2022.

[21] J. Ding, Z. Zhou, W. Li, C. Wang, L. Lin, and B. Jiao, "Movable antenna-enabled co-frequency co-time full-duplex wireless communication," *IEEE Commun. Lett.*, vol. 28, no. 10, pp. 2412–2416, 2024.

[22] T. Riihonen, D. Korpi, O. Rantula, H. Rantanen, T. Saarelainen, and M. Valkama, "Inband Full-Duplex Radio Transceivers: A Paradigm Shift in Tactical Communications and Electronic Warfare?," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 30–36, 2017.

[23] R. Askar, J. Chung, Z. Guo, H. Ko, W. Keusgen, and T. Haustein, "Interference handling challenges toward full duplex evolution in 5G and beyond cellular networks," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 51–59, 2021.

[24] H. Alves, T. Riihonen, and H. A. Suraweera, eds., *Full-duplex communications for future wireless networks*. Springer Singapore, 2020.

[25] A. T. Le, L. C. Tran, X. Huang, and Y. J. Guo, "Beam-based analog self-interference cancellation in full-duplex MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2460–2471, 2020.

[26] J. W. Kwak, M. S. Sim, I.-W. Kang, J. Park, K.-K. Wong, and C.-B. Chae, "Analog self-interference cancellation with practical RF components for full-duplex radios," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4552–4564, 2023.

[27] H. Kim, K. Ko, K. Kwon, I.-C. Hwang, and S. Park, "Fast electrical balance duplexer tuning using neural networks for RF self-interference cancellation in in-band full-duplex systems," *IEEE Access*, vol. 12, pp. 151805–151824, 2024.

[28] Y. Liu, X. Quan, W. Pan, and Y. Tang, "Digitally assisted analog interference cancellation for in-band full-duplex radios," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1079–1082, 2017.

[29] F. J. Soriano-Irigaray, J. S. Fernandez-Prat, F. J. Lopez-Martinez, E. Martos-Naya, O. Cobos-Morales, and J. T. Entrambasaguas, "Adaptive self-interference cancellation for full duplex radio: Analytical model and experimental validation," *IEEE Access*, vol. 6, pp. 65018–65026, 2018.

[30] E. Ahmed and A. M. Eltawil, "All-Digital Self-Interference Cancellation Technique for Full-Duplex Systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, 2015.

[31] D. Korpi, L. Anttila, V. Syrjälä, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion full-duplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, 2014.

[32] M. Yilan, O. Gurbuz, and H. Ozkan, "Nonlinear digital self-interference cancellation for full duplex communication," *Phys. Commun.*, vol. 35, p. 100698, 2019.

[33] K. Komatsu, Y. Miyaji, and H. Uehara, "Iterative Nonlinear Self-Interference Cancellation for In-Band Full-Duplex Wireless Communications Under Mixer Imbalance and Amplifier Nonlinearity," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4424–4438, 2020.

[34] P. P. Campo, L. Anttila, D. Korpi, and M. Valkama, "Cascaded Spline-Based Models for Complex Nonlinear Systems: Methods and Applications," *IEEE Trans. Signal Process.*, vol. 69, pp. 370–384, 2021.

[35] A. Balatsoukas-Stimming, "Non-Linear Digital Self-Interference Cancellation for In-Band Full-Duplex Radios Using Neural Networks," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 1–5, 2018.

[36] P. P. Campo, L. Anttila, D. Korpi, and M. Valkama, "Adaptive Cancellation of Nonlinear Self-Interference in Wireless Full-Duplex: Cascaded Spline-Interpolated Methods," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pp. 1265–1271, 2020.

[37] L. Anttila, V. Lampu, S. A. Hassani, P. P. Campo, D. Korpi, M. Turunen, S. Pollin, and M. Valkama, "Full-Duplexing With SDR Devices: Algorithms, FPGA Implementation, and Real-Time Results," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2205–2220, 2021.

[38] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low Complexity Neural Network Structures for Self-Interference Cancellation in Full-Duplex Radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, 2021.

[39] V. Tapio and M. Juntti, "Non-Linear Self-Interference Cancelation for Full-Duplex Transceivers Based on Hammerstein-Wiener Model," *IEEE Commun. Lett.*, pp. 1–1, 2021.

[40] A. Chowdhury, R. Chopra, C. R. Murthy, and H. A. Suraweera, "On the Achievable Rates of Full-Duplex Massive MIMO Systems Under Channel Aging," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 1–5, 2019.

[41] F. Chen, R. Morawski, and T. Le-Ngoc, "Self-Interference Channel Characterization for Wideband 2×2 MIMO Full-Duplex Transceivers Using Dual-Polarized Antennas," *IEEE Trans. Antennas Propag.*, vol. 66, no. 4, pp. 1967–1976, 2018.

[42] J. Tamminen, M. Turunen, D. Korpi, T. Huusari, Y.-S. Choi, S. Talwar, and M. Valkama, "Digitally-controlled RF self-interference canceller for full-duplex radios," in *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 783–787, 2016.

[43] B. Lee, J.-B. Lim, C. Lim, B. Kim, and J.-y. Seol, "Reflected Self-Interference Channel Measurement for mmWave Beamformed Full-Duplex System," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Dec.*, pp. 1–6, 2015.

[44] C. Campolo, A. Molinaro, A. O. Berthet, and A. Vinel, "Full-duplex radios for vehicular communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 182–189, 2017.

[45] D. H. Kong, Y.-S. Kil, and S.-H. Kim, "Neural Network aided Digital Self-Interference Cancellation for Full-Duplex Communication over Time-Varying Channels," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2022.

[46] S. N. Venkatasubramanian, C. Zhang, L. Laughlin, K. Haneda, and M. A. Beach, "Geometry-Based Modeling of Self-Interference Channels for Outdoor Scenarios," *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 3297–3307, 2019.

[47] Y. He, H. Zhao, W. Guo, S. Shao, and Y. Tang, "A Time-Robust Digital Self-Interference Cancellation in Full-Duplex Radios: Receiver Design and Performance Analysis," *IEEE Access*, vol. 8, pp. 185021–185031, 2020.

[48] C. Zhang and X. Luo, "Adaptive Digital Self-Interference Cancellation for Millimeter-Wave Full-Duplex Backhaul Systems," *IEEE Access*, vol. 7, pp. 175542–175553, 2019.

[49] S. Chakraborty and D. Sen, "Iterative SAGE-Based Joint MCFOs and Channel Estimation for Full-Duplex Two-Way Multi-Relay Systems in Highly Mobile Environment," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7379–7394, 2018.

[50] H. Vogt, G. Enzner, and A. Sezgin, "State-Space Adaptive Nonlinear Self-Interference Cancellation for Full-Duplex Communication," *IEEE Trans. Signal Process.*, vol. 67, no. 11, pp. 2810–2825, 2019.

[51] A. Kurt, M. B. Salman, H. A. Satana, and G. M. Guvensen, "An Efficient Iterative SIC for Full-Duplex SC-FDE Radio Under Hardware Impairments," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, 2021.

[52] I. P. Roberts, H. B. Jain, S. Vishwanath, and J. G. Andrews, "Millimeter wave analog beamforming codebooks robust to self-interference," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, 2021.

[53] I. P. Roberts, A. Chopra, T. Novlan, S. Vishwanath, and J. G. Andrews, "Beamformed self-interference measurements at 28 GHz: Spatial insights and angular spread," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9744–9760, 2022.

[54] C. B. Barneto, T. Riihonen, S. D. Liyanaarachchi, M. Heino, N. González-Prelcic, and M. Valkama, "Beamformer design and optimization for joint communication and full-duplex sensing at mm-Waves," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8298–8312, 2022.

[55] E. Everett, C. Shepard, L. Zhong, and A. Sabharwal, "Softnull: Many-antenna full-duplex wireless via digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8077–8092, 2016.

[56] S.-M. Kim, Y.-G. Lim, L. Dai, and C.-B. Chae, "Performance analysis of self-interference cancellation in full-duplex massive MIMO systems: Subtraction versus spatial suppression," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 642–657, 2023.

[57] E. Balti, S. Akoum, I. Alfalujah, and B. L. Evans, "Hybrid beamforming design for full-duplex millimeter wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 11, pp. 17041–17058, 2024.

[58] A. Kurt, M. B. Salman, U. B. Sarac, and G. M. Guvensen, "An adaptive-iterative nonlinear interference cancellation in time-varying full-duplex channels," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 1862–1878, 2023.

[59] A. Koc and T. Le-Ngoc, "Full-duplex mmWave massive MIMO systems: A joint hybrid precoding/combining and self-interference cancellation design," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 754–774, 2021.

[60] A. Kurt and G. M. Guvensen, "Sensing and mitigation of far-field self-interference for full-duplex MIMO systems," in *Proc. IEEE 35th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2024.

[61] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[62] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[63] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, 2015.

[64] R. Jia, X. Chen, C. Zhong, D. W. K. Ng, H. Lin, and Z. Zhang, "Design of non-orthogonal beamspace multiple access for cellular internet-of-things," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 538–552, 2019.

[65] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, 2014.

[66] S. Fortunati, L. Sanguinetti, F. Gini, M. S. Greco, and B. Himed, "Massive MIMO radar for target detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 859–871, 2020.

[67] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.

[68] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, 2020.

[69] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[70] C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and Y. Hao, "6G wireless channel measurements and models: Trends and challenges," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 22–32, 2020.

[71] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, 2016.

[72] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TD-D/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, 2018.

[73] V. Va, J. Choi, and R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5014–5029, 2017.

[74] Y. Ma, S. Ren, Z. Quan, and Z. Feng, "Data-driven hybrid beamforming for uplink multi-user MIMO in mobile millimeter-wave systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9341–9350, 2022.

[75] S. Shaham, M. Ding, M. Kokshoorn, Z. Lin, S. Dang, and R. Abbas, "Fast channel estimation and beam tracking for millimeter wave vehicular communications," *IEEE Access*, vol. 7, pp. 141104–141118, 2019.

[76] J. Palacios, D. De Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," in *Proc. Conf. Comput. Commun.*, pp. 1–9, 2017.

[77] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, pp. 2909–2913, 2015.

[78] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, "Millimeter wave beam alignment: Large deviations analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1619–1631, 2017.

[79] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, 2017.

[80] D. Zhang, A. Li, M. Shirvanimoghaddam, P. Cheng, Y. Li, and B. Vucetic, "Codebook-based training beam sequence design for millimeter-wave tracking

systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5333–5349, 2019.

[81] D. Zhang, A. Li, C. Pradhan, J. Li, B. Vucetic, and Y. Li, "Training beam sequence design for multiuser millimeter wave tracking systems," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6939–6955, 2021.

[82] M. Hussain and N. Michelusi, "Learning and adaptation for millimeter-wave beam tracking and training: A dual timescale variational framework," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 37–53, 2022.

[83] J. Li, X. Li, L. Xiao, and S. Zhou, "Joint multi-beam and channel tracking for mmWave hybrid beamforming multi-user systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1513–1517, 2021.

[84] C. Liu, M. Li, L. Zhao, P. Whiting, S. V. Hanly, I. B. Collings, and M. Zhao, "Robust adaptive beam tracking for mobile millimetre wave communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1918–1934, 2021.

[85] D. Zhu, J. Choi, Q. Cheng, W. Xiao, and R. W. Heath, "High-resolution angle tracking for mobile wideband millimeter-wave systems with antenna array calibration," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7173–7189, 2018.

[86] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, pp. 743–747, 2016.

[87] S. Jayaprakasam, X. Ma, J. W. Choi, and S. Kim, "Robust beam-tracking for mmWave mobile communications," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2654–2657, 2017.

[88] P. Zhu, H. Lin, J. Bao, J. Li, and D. Wang, "Beam tracking for distributed millimeter-wave massive MIMO systems based on the unscented Kalman filter," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 712–716, 2022.

[89] A. Kurt and G. M. Guvensen, "An adaptive hybrid beamforming scheme for time-varying wideband massive MIMO channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–7, 2020.

[90] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, 2017.

[91] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2016.

[92] S. Shaham, M. Kokshoorn, M. Ding, Z. Lin, and M. Shirvanimoghaddam, "Extended kalman filter beam tracking for millimeter wave vehicular communications," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020.

[93] A. Kurt and G. M. Guvensen, "Robust statistical beamforming with multi-cluster tracking for time-varying massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 3, pp. 3499–3515, 2024.

[94] M. F. Keskin, H. Wymeersch, and V. Koivunen, "Monostatic sensing with OFDM under phase noise: From mitigation to exploitation," *IEEE Trans. Signal Process.*, vol. 71, pp. 1363–1378, 2023.

[95] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, 2022.

[96] Z. Chen and C. Yang, "Pilot decontamination in wideband massive MIMO systems by exploiting channel sparsity," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5087–5100, 2016.

[97] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[98] W. Sloane, C. Gentile, M. Shafi, J. Senic, P. A. Martin, and G. K. Woodward, "Measurement-based analysis of millimeter-wave channel sparsity," *IEEE Antennas Wireless Propag. Lett.*, vol. 22, no. 4, pp. 784–788, 2023.

[99] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[100] A. Kurt and G. M. Guvensen, "An efficient hybrid beamforming and channel acquisition for wideband mm-Wave massive MIMO channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–7, 2019.

[101] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[102] C. Balanis, *Antenna Theory: Analysis and Design*. Wiley, 2016.

[103] J.-S. Jiang and M. Ingram, "Spherical-wave model for short-range MIMO," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1534–1541, 2005.

[104] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot Beam Pattern Design for Channel Estimation in Massive MIMO Systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, 2014.

[105] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, 2019.

[106] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA, 2006.

[107] F. Pancaldi, G. M. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 37–56, 2008.

[108] S. A. Hassani, V. Lampu, K. Parashar, L. Anttila, A. Bourdoux, B. v. Liempd, M. Valkama, F. Horlin, and S. Pollin, "In-Band Full-Duplex Radar-Communication System," *IEEE Systems Journal*, vol. 15, no. 1, pp. 1086–1097, 2021.

[109] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, 2000.

[110] A. A. M. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1715–1720, 1981.

[111] C.-P. Liang, J.-H. Jong, W. Stark, and J. East, "Nonlinear amplifier effects in communications systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1461–1466, 1999.

[112] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, 2000.

[113] S. Kurma, K. Singh, S. Mumtaz, T. A. Tsiftsis, and C.-P. Li, "Resource optimization in active-STAR-RIS-aided THz ISAC systems with DDA modulation: A machine-learning approach," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 15291–15307, 2024.

[114] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. USA: Prentice-Hall, Inc., 1993.

[115] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, 2007.

[116] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8315–8328, 2019.

[117] M. Richards, *Fundamentals of Radar Signal Processing, Second Edition*. McGraw Hill LLC, 2013.

[118] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2018.

[119] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "3-D beamforming for flexible coverage in millimeter-wave UAV communications," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 837–840, 2019.

[120] A. Kurt and G. M. Guvensen, “Robust statistical beamforming with multi-cluster tracking for time-varying massive MIMO (extended version),” DOI: 10.48550/ARXIV.2303.00457, 2023.

[121] M. Bayraktar and G. M. Guvensen, “An efficient interference-aware constrained massive MIMO beamforming for mm-wave JSDM,” *IEEE Access*, vol. 9, pp. 87877–87897, 2021.

[122] 3GPP, “NR Physical Channels and Modulation Release 17,” Standard TS 38.211, 2023. 17.5.0.

APPENDICES

A Information Decorrelation

The correlations $c_{SC}(\Delta_t)$ and $c_{OFDM}(\Delta_t)$ defined in Section 2.1.3.2 can be computed as below.

$$c_{SC}(\Delta_t) = \mathbb{E}\Big\{ \int_{-\infty}^{+\infty} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} s_{n_1} s_{n_2}^* p_{SC}(t - n_1 T) p_{SC}^*(t - n_2 T - \Delta t) dt \Big\} \quad \text{(A.1)}$$

$$= \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \mathbb{E}\{s_{n_1} s_{n_2}^*\} \int_{-\infty}^{+\infty} p_{SC}(t - n_1 T) p_{SC}^*(t - n_2 T - \Delta t) dt \quad \text{(A.2)}$$

$$= \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \delta[n_1 - n_2] \int_{-\infty}^{+\infty} p_{SC}(t - n_1 T) p_{SC}^*(t - n_2 T - \Delta t) dt \quad \text{(A.3)}$$

$$= \sum_{n_1=0}^{N-1} \int_{-\infty}^{+\infty} p_{SC}(t - n_1 T) p_{SC}^*(t - n_1 T - \Delta t) dt \quad \text{(A.4)}$$

$$= N \int_{-\infty}^{+\infty} p_{SC}(t) p_{SC}^*(t - \Delta t) dt \quad \text{(A.5)}$$

$$c_{OFDM}(\Delta_t) = \mathbb{E}\Big\{ \int_{-\infty}^{+\infty} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} s_{n_1} s_{n_2}^* e^{j2\pi\Delta_f(n_1 t - n_2 t + n_2 \Delta_t)} p_{OFDM}(t) \times p_{OFDM}^*(t - \Delta t) dt \Big\} \quad \text{(A.6)}$$

$$= \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \mathbb{E}\{s_{n_1} s_{n_2}^*\} \int_{-\infty}^{+\infty} e^{j2\pi\Delta_f(n_1 t - n_2 t + n_2 \Delta_t)} p_{OFDM}(t) \times p_{OFDM}^*(t - \Delta t) dt \quad \text{(A.7)}$$

$$= \sum_{n_1=0}^{N-1} e^{j2\pi n_1 \Delta_f \Delta_t} \int_{-\infty}^{+\infty} p_{OFDM}(t) p_{OFDM}^*(t - \Delta t) dt \quad \text{(A.8)}$$

$$= e^{j\pi(N-1)\Delta_f \Delta_t} \frac{\sin(\pi N \Delta_f \Delta_t)}{\sin(\pi \Delta_f \Delta_t)} \int_{-\infty}^{+\infty} p_{OFDM}(t) p_{OFDM}^*(t - \Delta t) dt \quad \text{(A.9)}$$

B Calculation of RSIP After Linear SIC

B.1 Channel Aging

On the SI channel description given in Section 3.1.2, define the K-step aging of the n^{th} delay as

$$e_n^h[k,K] \triangleq h_n^{SI}[k] - h_n^{SI}[k-K], \tag{B.1}$$

where $K \geqslant 0$. Note that the aged channel is $h_n^{SI}[k-K] = \bar{h}_n + \tilde{h}_n[k-K]$ and the current channel can be written as

$$\begin{aligned} h_n^{SI}[k] = \bar{h}_n &+ \alpha^K \tilde{h}_n[k-K] \\ &+ \sqrt{1-\alpha^2}\sigma_n \sum_{k'=0}^{K-1} \alpha^{k'} v_n[k-k']. \end{aligned} \tag{B.2}$$

Then, the K-step aging becomes

$$\begin{aligned} e_n^h[k,K] =&(\alpha^K - 1)\tilde{h}_n[k-K] \\ &+ \sqrt{1-\alpha^2}\sigma_n \sum_{k'=0}^{K-1} \alpha^{k'} v_n[k-k']. \end{aligned} \tag{B.3}$$

Using uncorrelatedness of $\tilde{h}_n[k-K]$ and $v_n[k-k']$ for $k' < K$, $E\{|\tilde{h}_n[k-K]|^2\} = \sigma_n^2$, $E\{|v_n[k-k']|^2\} = 1$ and the identity $\sum_{k=0}^{K-1} c^k = (1-c^K)/(1-c)$, the power of the K-step aging is found as

$$E\left\{\left|e_n^h[k,K]\right|^2\right\} = 2(1-\alpha^K)\sigma_n^2. \tag{B.4}$$

K-step aging for different delays n are uncorrelated. Then,

$$E\left\{\boldsymbol{e}^h[k,K](\boldsymbol{e}^h[k,K])^H\right\} = 2(1-\alpha^K)\text{diag}(\{\sigma_n^2\}_{n=0}^{L-1}), \tag{B.5}$$

where $\boldsymbol{e}^h[k,K] \triangleq \boldsymbol{h}^{SI}[k] - \boldsymbol{h}^{SI}[k-K]$.

B.2 Derivation of RSIP Expressions for $k_{obs} < k$

Upon the signal model given in Section 3.3, define the unit-power single-carrier $N' \times L$ convolution matrix $\boldsymbol{X}_{sc}$ to be identically distributed to $\boldsymbol{X}[k]/\rho_x$ for any k. $\boldsymbol{X}_{sc}$ will

be helpful in obtaining expectations independent of k and ρ_x by replacing $\boldsymbol{X}[k]$. In addition, $N' \times N'$ autocorrelation matrices for SoI and the noise are

$$E\{\boldsymbol{s}^{obs}[k](\boldsymbol{s}^{obs}[k])^H\} = (\rho_s^{obs})^2 \boldsymbol{R}_{sc}, \tag{B.6}$$

$$E\{\boldsymbol{\eta}[k](\boldsymbol{\eta}[k])^H\} = \mu\sigma_\eta^2 \boldsymbol{I}_{N'}, \tag{B.7}$$

where $\boldsymbol{R}_{sc}$ is the normalized SoI covariance matrix. Also, the operator $\mathcal{L}(\cdot)$ is defined as

$$\mathcal{L}(\boldsymbol{G}) \triangleq \left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H \tag{B.8}$$

where $\boldsymbol{G}$ is an arbitrary matrix. Using (3.40) and (3.41), the estimated SI channel is

$$\begin{aligned}\hat{\boldsymbol{h}}^{SI}[k] = \boldsymbol{h}^{SI}[k_{obs}] &+ \mathcal{L}(\boldsymbol{X}[k_{obs}])\boldsymbol{s}^{obs}[k_{obs}] \\ &+ \mathcal{L}(\boldsymbol{X}[k_{obs}])\boldsymbol{\eta}[k_{obs}].\end{aligned} \tag{B.9}$$

Then, according to (3.43), residual signal after SIC is

$$\begin{aligned}\boldsymbol{e}_s[k] = \boldsymbol{X}[k]\boldsymbol{e}^h[k, k-k_{obs}] &- \boldsymbol{X}[k]\mathcal{L}(\boldsymbol{X}[k_{obs}])\boldsymbol{s}^{obs}[k_{obs}] \\ &- \boldsymbol{X}[k]\mathcal{L}(\boldsymbol{X}[k_{obs}])\boldsymbol{\eta}[k_{obs}] + \boldsymbol{\eta}[k].\end{aligned} \tag{B.10}$$

Through (3.45) and (3.46), the RSIP P can be rewritten as

$$P = \frac{1}{\rho_s^2\mu N}\text{tr}\left(E\{\boldsymbol{P}^H\boldsymbol{e}_s[k](\boldsymbol{e}_s[k])^H\boldsymbol{P}\}\right). \tag{B.11}$$

The four terms of $\boldsymbol{e}_s[k]$ expression in (B.10) are uncorrelated. The contributions of these terms to P in (B.11) can be found separately and summed up to obtain the total RSIP as $P = P_1 + P_2 + P_3 + P_4$, where P_k is the contribution of the k^{th} term in (B.10) to the power expression in (B.11). The first term P_1 that represents the channel variation is calculated as

$$\begin{aligned}P_1 \triangleq \frac{1}{\rho_s^2\mu N}\text{tr}\Big(E\Big\{&\boldsymbol{P}^H\boldsymbol{X}[k]\boldsymbol{e}^h[k, k-k_{obs}] \\ &(\boldsymbol{e}^h[k, k-k_{obs}])^H\boldsymbol{X}^H[k]\boldsymbol{P}\Big\}\Big),\end{aligned} \tag{B.12}$$

$$\begin{aligned}P_1 = \frac{2(1-\alpha^{k-k_{obs}})}{\mu N}\frac{\rho_x^2}{\rho_s^2}\text{tr}\Big(&E\{\boldsymbol{X}_{sc}^H\boldsymbol{P}\boldsymbol{P}^H\boldsymbol{X}_{sc}\} \\ &\text{diag}(\{\sigma_n^2\}_{n=0}^{L-1})\Big).\end{aligned} \tag{B.13}$$

The second term P_2, which is responsible for the contamination of SoI in the channel estimation, is calculated as

$$P_2 \triangleq \frac{1}{\rho_s^2 \mu N} \text{tr}\Big(E\Big\{ \boldsymbol{P}^H \boldsymbol{X}[k] \mathcal{L}\left(\boldsymbol{X}[k_{obs}]\right) \boldsymbol{s}^{obs}[k_{obs}] \boldsymbol{s}^{obs}[k_{obs}]^H \left(\mathcal{L}\left(\boldsymbol{X}[k_{obs}]\right)\right)^H \boldsymbol{X}^H[k] \boldsymbol{P} \Big\} \Big), \tag{B.14}$$

$$P_2 = \frac{\rho_{s,obs}^2}{\rho_s^2 \mu N} \text{tr}\Big(E\{\boldsymbol{X}_{sc}^H \boldsymbol{P} \boldsymbol{P}^H \boldsymbol{X}_{sc}\} E\{(\boldsymbol{X}_{sc}^H \boldsymbol{X}_{sc})^{-1} \boldsymbol{X}_{sc}^H \boldsymbol{R}_{sc} \boldsymbol{X}_{sc} (\boldsymbol{X}_{sc}^H \boldsymbol{X}_{sc})^{-1}\} \Big). \tag{B.15}$$

In (B.13) and (B.15), $\rho_x \boldsymbol{X}_{sc}$ is substituted instead of $\boldsymbol{X}[k]$ after being sure that the surrounding expectation does not involve both $\boldsymbol{X}[k]$ and $\boldsymbol{X}[k_{obs}]$ by using trace property, since they are uncorrelated. Then, substitutions are made according to (B.5), (B.6) and (B.7). The expressions for the next terms are also derived similarly, together with the Nyquist-1 property $\boldsymbol{P}^H \boldsymbol{P} = \boldsymbol{I}_N$ to be used for the noise terms. Therefore, expressions for the next terms will be given more directly. P_3 is found as

$$P_3 = \frac{\sigma_\eta^2}{\rho_s^2 N} \text{tr}(E\{\boldsymbol{X}_{sc}^H \boldsymbol{P} \boldsymbol{P}^H \boldsymbol{X}_{sc}\} E\{(\boldsymbol{X}_{sc}^H \boldsymbol{X}_{sc})^{-1}\}). \tag{B.16}$$

which represents the contamination of noise in the estimated channel. P_4 is the average power of the noise in the received signal and it is found as $P_4 = \sigma_\eta^2 / \rho_s^2$. Then, the total RSIP after MF when $k_{obs} < k$ is

$$P = \frac{2(1-\alpha^{k-k_{obs}})}{\mu N} \frac{\rho_x^2}{\rho_s^2} \text{tr}\left(\boldsymbol{A}\boldsymbol{D}\right) + \frac{\rho_{s,obs}^2}{\rho_s^2 \mu N} \text{tr}(\boldsymbol{A}\boldsymbol{B}) + \frac{\sigma_\eta^2}{\rho_s^2 N} \text{tr}(\boldsymbol{A}\boldsymbol{C}) + \frac{\sigma_\eta^2}{\rho_s^2}, \tag{B.17}$$

where

$$\boldsymbol{A} \triangleq E\{\boldsymbol{X}_{sc}^H \boldsymbol{P} \boldsymbol{P}^H \boldsymbol{X}_{sc}\}, \tag{B.18}$$

$$\boldsymbol{B} \triangleq E\{(\boldsymbol{X}_{sc}^H \boldsymbol{X}_{sc})^{-1} \boldsymbol{X}_{sc}^H \boldsymbol{R}_{sc} \boldsymbol{X}_{sc} (\boldsymbol{X}_{sc}^H \boldsymbol{X}_{sc})^{-1}\}, \tag{B.19}$$

$$\boldsymbol{C} \triangleq E\{(\boldsymbol{X}_{sc}^H \boldsymbol{X}_{sc})^{-1}\}, \quad \boldsymbol{D} \triangleq \text{diag}\left(\left\{\sigma_n^2\right\}_{n=0}^{L-1}\right). \tag{B.20}$$

B.3 Derivation of RSIP Expressions for $k_{obs} = k$

According to (3.40), (3.41) and (3.43), estimated SI channel and residual signal after SIC when $k_{obs} = k$ are

$$\hat{\boldsymbol{h}}^{SI}[k] = \boldsymbol{h}^{SI}[k] + \mathcal{L}\left(\boldsymbol{X}[k]\right)\left(\boldsymbol{s}^{obs}[k] + \boldsymbol{\eta}[k]\right), \tag{B.21}$$

$$\begin{aligned}\boldsymbol{e}_s[k] = &- \boldsymbol{X}[k]\mathcal{L}\left(\boldsymbol{X}[k]\right)\boldsymbol{s}^{obs}[k] \\ &+ \left(\boldsymbol{I} - \boldsymbol{X}[k]\mathcal{L}\left(\boldsymbol{X}[k]\right)\right)\boldsymbol{\eta}[k].\end{aligned} \tag{B.22}$$

Note that there is no effect of channel variation when $k_{obs} = k$. The two terms in (B.22) are uncorrelated. Then, (B.22) is substituted into (B.11) and the total power can be written as $P = P_1 + P_2$, where numerations on the subscripts of the terms follow the order in (B.22). The simplified expressions for the terms are given as

$$\begin{aligned}P_1 = \frac{\rho_{s,obs}^2}{\rho_s^2 \mu N}\mathrm{tr}\Big(E\Big\{&\boldsymbol{X}_{sc}(\boldsymbol{X}_{sc}^H\boldsymbol{X}_{sc})^{-1}\boldsymbol{X}_{sc}^H\boldsymbol{P}\boldsymbol{P}^H \\ &\boldsymbol{X}_{sc}(\boldsymbol{X}_{sc}^H\boldsymbol{X}_{sc})^{-1}\boldsymbol{X}_{sc}^H\Big\}\boldsymbol{R}_{sc}\Big),\end{aligned} \tag{B.23}$$

$$P_2 = \frac{\sigma_\eta^2}{\rho_s^2} - \frac{\sigma_\eta^2}{\rho_s^2 N}\mathrm{tr}\left(\boldsymbol{P}^H E\{\boldsymbol{X}_{sc}(\boldsymbol{X}_{sc}^H\boldsymbol{X}_{sc})^{-1}\boldsymbol{X}_{sc}^H\}\boldsymbol{P}\right). \tag{B.24}$$

Then, the total RSIP after MF when $k_{obs} = k$ is

$$P = \frac{\rho_{s,obs}^2}{\rho_s^2 \mu N}\mathrm{tr}(\boldsymbol{E}\boldsymbol{R}_{sc}) + \frac{\sigma_\eta^2}{\rho_s^2} - \frac{\sigma_\eta^2}{\rho_s^2 N}\mathrm{tr}(\boldsymbol{P}^H\boldsymbol{F}\boldsymbol{P}), \tag{B.25}$$

where

$$\boldsymbol{E} \triangleq E\{\boldsymbol{X}_{sc}(\boldsymbol{X}_{sc}^H\boldsymbol{X}_{sc})^{-1}\boldsymbol{X}_{sc}^H\boldsymbol{P}\boldsymbol{P}^H\boldsymbol{X}_{sc}(\boldsymbol{X}_{sc}^H\boldsymbol{X}_{sc})^{-1}\boldsymbol{X}_{sc}^H\}, \tag{B.26}$$

$$\boldsymbol{F} \triangleq E\{\boldsymbol{X}_{sc}(\boldsymbol{X}_{sc}^H\boldsymbol{X}_{sc})^{-1}\boldsymbol{X}_{sc}^H\}. \tag{B.27}$$

C Derivations Related to NFSI Cancellation

An approximate expression for $\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^H = \sum_{p=1}^{P} \boldsymbol{X}_p \boldsymbol{X}_p^H$ will be derived considering P is very large. From (4.2),

$$\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^H = \sum_{p=1}^{P} \boldsymbol{W}_p \boldsymbol{T}_p \boldsymbol{S}_m^{(d)} (\boldsymbol{S}_m^{(d)})^H \boldsymbol{T}_p^H \boldsymbol{W}_p^H, \tag{C.1}$$

where $\boldsymbol{S}_m^{(d)} (\boldsymbol{S}_m^{(d)})^H \cong N_c \boldsymbol{I}_{K_d}$ by the law of large numbers. Also, $\boldsymbol{W}_p \boldsymbol{T}_p \cong \frac{1}{\sqrt{K_d}} \boldsymbol{Q}_p^{(d)}$ is assumed, which is accurate if the DL users are not closer than angular resolution, and exact when orthogonal discrete Fourier transform (DFT) beams are used, yielding

$$\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^H \cong \frac{N_c}{K_d} \sum_{p=1}^{P} \boldsymbol{Q}_p^{(d)} (\boldsymbol{Q}_p^{(d)})^H. \tag{C.2}$$

Assuming that $\boldsymbol{Q}_p^{(d)}$ is constituted by uniformly random K_d DFT beams for each p, groups of N_a/K_d OFDM symbols can be found for which the DFT codebook is completed. That is, the whole space is spanned completely PK_d/N_a times within P OFDM symbols. Indexing each p with a unique (p_1, p_2) pair, this can be shown as

$$\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^H = \frac{N_c}{K_d} \sum_{p_1=1}^{\frac{PK_d}{N_a}} \sum_{p_2=1}^{\frac{N_a}{K_d}} \boldsymbol{Q}_{p_1,p_2}^{(d)} (\boldsymbol{Q}_{p_1,p_2}^{(d)})^H = \frac{N_c}{K_d} \sum_{p_1=1}^{\frac{PK_d}{N_a}} \boldsymbol{I}_{N_a}. \tag{C.3}$$

This relation holds exactly if DFT beams are considered and accurately otherwise, given that P is large and the space is spanned randomly and uniformly during P OFDM symbols. Finally, it yields

$$\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^H \cong \frac{N_c P}{N_a} \boldsymbol{I}_{N_a}. \tag{C.4}$$

D Expressions for Robust Statistical Beamforming

D.1 Angular Estimate for BA-ML

In this appendix, (5.31) will be derived using (5.29) and the definitions in 5.4.2.2. Since the channel estimates in (5.27) are taken from different FT-CPIs, they are independent and $p(\boldsymbol{f}_k^{(m)}|\theta,\{\boldsymbol{\beta}'_p\}) = \prod_{p=(k-1)P+1}^{kP} p(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}|\theta,\boldsymbol{\beta}'_p)$ can be used. Then,

$$\hat{\theta}_{k+1}^{(\mathrm{ST},m)} = arg\max_{\theta} \prod_{p=(k-1)P+1}^{kP} \max_{\boldsymbol{\beta}'_p} p(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}|\theta,\boldsymbol{\beta}'_p). \tag{D.1}$$

It is assumed that $p(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}|\theta,\boldsymbol{\beta}'_p)$ is expressed by $\mathcal{CN}(\boldsymbol{E}'(\theta)\boldsymbol{\beta}'_p, \frac{N_0}{E^{(m)}N_{\mathrm{F}}}\boldsymbol{I}_{D_m})$. Then, the likelihood $p(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}|\theta,\boldsymbol{\beta}'_p)$ is maximized for given $\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}$ and θ by the LS estimate $\hat{\boldsymbol{\beta}}_p$ of $\boldsymbol{\beta}'_p$. Then, the mean $\boldsymbol{E}'(\theta)\hat{\boldsymbol{\beta}}_p$ satisfies $\boldsymbol{E}'(\theta)\hat{\boldsymbol{\beta}}_p = \boldsymbol{P}(\theta)\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}$, where

$$\boldsymbol{P}(\theta) \triangleq \boldsymbol{E}'(\theta)\left(\boldsymbol{E}'^H(\theta)\boldsymbol{E}'(\theta)\right)^{-1}\boldsymbol{E}'^H(\theta). \tag{D.2}$$

Then, the function can be converted to log-likelihood to obtain

$$\hat{\theta}_{k+1}^{(\mathrm{ST},m)} = arg\min_{\theta} \sum_{p=(k-1)P+1}^{kP} \left(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} - \boldsymbol{P}(\theta)\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}\right)^H \left(\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} - \boldsymbol{P}(\theta)\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}\right), \tag{D.3}$$

$$\hat{\theta}_{k+1}^{(\mathrm{ST},m)} = arg\min_{\theta} \sum_{p=(k-1)P+1}^{kP} (\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)})^H \boldsymbol{M}(\theta)\hat{\tilde{\boldsymbol{h}}}_p^{(m,m)}, \tag{D.4}$$

where $\boldsymbol{M}(\theta) \triangleq (\boldsymbol{I}_{D_m} - \boldsymbol{P}(\theta))^H(\boldsymbol{I}_{D_m} - \boldsymbol{P}(\theta)) = \boldsymbol{I}_{D_m} - \boldsymbol{P}(\theta)$. This expression can be implemented as in (5.31) utilizing the properties of the trace operation, which reduces complexity since the second matrix is calculated only once during the search.

D.2 Covariance Matrix in SEKF

The observation $\boldsymbol{f}_k^{(m)}$ in (5.32) can also be written as

$$\boldsymbol{f}_k = \frac{1}{P}\sum_{p=(k-1)P+1}^{kP} \hat{\tilde{\boldsymbol{h}}}_p^{*} \otimes \hat{\tilde{\boldsymbol{h}}}_p, \tag{D.5}$$

where the superscripts are dropped for the sake of simplicity. Its mean is $\mathbb{E}\{\boldsymbol{f}_k\} = \mathrm{vec}\{\boldsymbol{R}_f(\theta_k)\}$, and covariance matrix $\boldsymbol{Q}_k$ is

$$\boldsymbol{Q}_k = \mathbb{E}\left\{\boldsymbol{f}_k\boldsymbol{f}_k^H\right\} - \mathbb{E}\{\boldsymbol{f}_k\}\mathbb{E}\{\boldsymbol{f}_k\}^H. \tag{D.6}$$

While the second term is known, the first term should be calculated. Using (D.5), it can be written as

$$\mathbb{E}\left\{\boldsymbol{f}_k \boldsymbol{f}_k^H\right\} = \frac{1}{P^2} \sum_{p_1} \sum_{p_2} \boldsymbol{X}(p_1, p_2) \tag{D.7}$$

where $\boldsymbol{X}(p_1, p_2)$ can be written in two equivalent forms:

$$\boldsymbol{X}(p_1, p_2) = \mathbb{E}\left\{\left(\hat{\tilde{\boldsymbol{h}}}_{p_1}^* \otimes \hat{\tilde{\boldsymbol{h}}}_{p_1}\right)\left(\hat{\tilde{\boldsymbol{h}}}_{p_2}^* \otimes \hat{\tilde{\boldsymbol{h}}}_{p_2}\right)^H\right\} \tag{D.8}$$

$$\boldsymbol{X}(p_1, p_2) = \mathbb{E}\left\{\left(\hat{\tilde{\boldsymbol{h}}}_{p_1} \hat{\tilde{\boldsymbol{h}}}_{p_2}^H\right)^* \otimes \left(\hat{\tilde{\boldsymbol{h}}}_{p_1} \hat{\tilde{\boldsymbol{h}}}_{p_2}^H\right)\right\} \tag{D.9}$$

Note that each entry of $\boldsymbol{X}(p_1, p_2)$ is a product of four variables, without summations. Since the channel is complex Gaussian, the identity $\mathbb{E}\{aa^*bb^*\} = \mathbb{E}\{aa^*\}\mathbb{E}\{bb^*\} + \mathbb{E}\{ab^*\}\mathbb{E}\{a^*b\}$ can be used, where a and b are complex Gaussian random variables. The essence of this identity is that different groups are assumed as independent for each term, e.g., a and b^* are assumed as independent of a^* and b in the second term. The two equivalent forms of $\boldsymbol{X}(p_1, p_2)$ in (D.8) and (D.9) can be used for these two terms. Consequently,

$$\boldsymbol{X}(p_1, p_2) = \mathbb{E}\{\boldsymbol{f}_k\}\mathbb{E}\{\boldsymbol{f}_k\}^H + \delta[p_1 - p_2]\boldsymbol{R}_f^*(\theta_k) \otimes \boldsymbol{R}_f(\theta_k) \tag{D.10}$$

Using (D.10), (D.7), and (D.6), the expression in (5.34) is obtained.

D.3 Analytical NMSE Expression for BA-LS

For the denominator in the NMSE expression in Section 5.7.2.1, $\mathbb{E}\{||\tilde{\boldsymbol{h}}_p^{(m,m)}||_2^2\} = \text{tr}(\tilde{\boldsymbol{R}}_p^{(m,m)})$ by definition. For the numerator, the error vector is defined as $\boldsymbol{e} \triangleq \hat{\tilde{\boldsymbol{h}}}_p^{(m,m)} - \tilde{\boldsymbol{h}}_p^{(m,m)}$, which is evident from (5.18). Its terms are independent, and the total energy of $\boldsymbol{e}$ is equal to the sum of energies of its terms. It is assumed that the symbols have unit power, and $(\boldsymbol{s}^{(m)})^H \boldsymbol{s}^{(m)}$ is deterministic and equal to N_F. In addition, $\boldsymbol{s}^{(m')}$ are uncorrelated random data for $m' \neq m$, which yields $\mathbb{E}\{|(\boldsymbol{s}^{(m)})^H \boldsymbol{s}^{(m')}|^2\} = N_\text{F}$. Then, the energy $\mathbb{E}\{||\boldsymbol{e}||_2^2\} = \text{tr}(\mathbb{E}\{\boldsymbol{e}\boldsymbol{e}^H\})$ of the error is written as (D.11) using (5.18), and simplified as (D.12) by inspecting (5.13).

$$\mathbb{E}\{||\boldsymbol{e}||_2^2\} = \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \frac{E^{(m')} N_\text{F}}{E^{(m)} N_\text{F}^2} \text{tr}\left(\tilde{\boldsymbol{R}}_p^{(m,m')}\right) + \text{tr}\left(\frac{N_0 \boldsymbol{I}_{D_m}}{E^{(m)} N_\text{F}}\right) \tag{D.11}$$

$$= \frac{1}{E^{(m)} N_\text{F}} \text{tr}\left(\tilde{\boldsymbol{\Psi}}_p^{(m)} - E^{(m)} \tilde{\boldsymbol{R}}_p^{(m,m)}\right) \tag{D.12}$$

CURRICULUM VITAE

Kurt, Anıl

Research Assistant, Dept. of Electrical and Electronics Engineering
Middle East Technical University, Ankara, Turkey

Education

- **Ph.D.**, Electrical and Electronics Engineering, METU, Ankara, Turkey (2019 – September 2025)
  - Thesis: *Sensing, Beamforming and Self-Interference Mitigation Methods for Full-Duplex Communications*
- **M.Sc.**, Electrical and Electronics Engineering, METU, Ankara, Turkey (2016 – 2019)
  - Thesis: *A General Framework on Adaptive Hybrid Beamforming and Channel Acquisition for Wideband mm-Wave Massive MIMO Systems*
- **B.Sc.**, Electrical and Electronics Engineering, METU, Ankara, Turkey (2011 – 2016)

Work Experience

- **Research Assistant**, METU (May 2019–Present)
- **Researcher**, 5G and Beyond Graduate Support Program with METU, Vodafone, BTK (Dec. 2018–Apr. 2019)
- **Project Engineer**, PARGE R&D (Apr. 2017–May 2018)
- **Candidate Engineer**, Aselsan Inc. (Jan. 2016–Jun. 2016)

Publications

A. Journal Articles

- **A. Kurt** and G. M. Guvensen, "Sensing and Mitigation of Multi-Scatterer Self-Interference for Full-Duplex MIMO Communications," arXiv:2412.02838, Dec. 2024. *(Submitted to IEEE Trans. Wireless Commun.)*
- **A. Kurt** and G. M. Guvensen, "Robust Statistical Beamforming With Multi-Cluster Tracking for Time-Varying Massive MIMO," *IEEE Trans. Veh. Tech.*, Mar. 2024.
- **A. Kurt**, M. B. Salman, U. B. Sarac and G. M. Guvensen, "An Adaptive-Iterative Nonlinear Interference Cancellation in Time-Varying Full-Duplex Channels," *IEEE Trans. Veh. Tech.*, Feb. 2023.

B. Conference Papers

- **A. Kurt** and G. M. Guvensen, “Sensing and Mitigation of Far-Field Self-Interference for Full-Duplex MIMO Systems,” *IEEE PIMRC*, Valencia, Spain, 2024.
- **A. Kurt**, M. B. Salman, H. A. Satana and G. M. Guvensen, "An Efficient Iterative SIC for Full-Duplex SC-FDE Radio Under Hardware Impairments," *IEEE ICC*, Montreal, 2021.
- **A. Kurt** and G. M. Guvensen, "An Adaptive Hybrid Beamforming Scheme for Time-Varying Wideband Massive MIMO Channels," *IEEE ICC*, Dublin, 2020.
- **A. Kurt** and G. M. Guvensen, "An Efficient Hybrid Beamforming and Channel Acquisition for Wideband mm-Wave Massive MIMO Channels," *IEEE ICC*, Shanghai, 2019.