THE UNIVERSITY *of* EDINBURGH

The University of Edinburgh

Ph.D Thesis

On The Feasibility and Applications of In-Band Full-Duplex Radios for Future Wireless Networks

Author: Haifeng Luo

Supervisor: Prof Tharmalingam Ratnarajah

A thesis submitted in fulfilment of the requirements for the degree of Ph.D.

with Institute for Digital Communications

in the

School of Engineering

September 2023

Declaration of Authorship

I declare that this thesis was composed by myself, that the work contained therein is my own except where explicitly stated otherwise, and that none of the work has been submitted for any other degree or professional qualification.

Haifeng Luo

Abstract

Due to the continuous increase of the demands for the wireless network's capacity, in-band full-duplex (IBFD) has recently become a key research topic due to its potential to double spectral efficiency, reduce latency, enhance emerging applications, etc., by transmitting and receiving simultaneously over the same channel. Meanwhile, many studies in the literature experimentally demonstrated the feasibility of IBFD radios, which leads to the belief that it is possible to introduce IBFD in the standard of the next-generation networks. Therefore, in this thesis, we timely study the feasibility of IBFD and investigate its advantages for emerging applications in future networks.

In the first part, we investigate the interference suppression methods to maximize the IBFD gain by minimizing the effects of self-interference (SI) and co-channel interference (CCI). To this end, we first study a 3-step self-interference cancellation (SIC) scheme. We focus on the time domain-based analog canceller and nonlinear digital canceller, explaining their rationale, demonstrating their effectiveness, and finding the optimal design by minimizing the residual effects. To break the limitation of conventional electrical radio frequency (RF) cancellers, we study the photonic-assisted canceller (PAC) and propose a new design, namely a fiber array-based canceller. We propose a new low-complexity tuning algorithm for the PAC. The effectiveness of the proposed fiber array canceller is demonstrated via simulations. Furthermore, we construct a prototype of the fiber array canceller with two taps and carry out experiments in real-world environments. Results show that the 3-step cancellation scheme can bring the SI close to the receiver's noise floor. Then, we consider the multiple-input multiple-output (MIMO) scenarios, proposing to employ hybrid RF-digital beamforming to reduce the implementation cost and studying its effects on the SIC design. Additionally, we propose a user allocation algorithm to reduce the CCI from the physical layer. A heterogeneous industrial Internet of Things (IIoT) scenario is considered, while the proposed algorithm can be generalized by modifying the parameters to fit any other network.

In the second part, we study the beamforming schemes for IBFD multi-cell multi-user (IBFD-MCMU) networks. The transceiver hardware impairments (HWIs) and channel uncertainty are considered for robustness. We first enhance zero-forcing (ZF) and maximum ratio transmission and combining (MRTC) beamforming to be compatible with IBFD-MCMU networks in the presence of multi-antenna users. Then, we study beamforming for SIC, which is challenging for MCMU networks due to the limited antennas but complex interference. We propose a minimum mean-squared error (MMSE)-based scheme to enhance the SIC performance while minimizing its effects on the sum rate. Furthermore, we investigate a robust joint power allocation and beamforming (JPABF) scheme, which approaches the performance of existing optimal designs with reduced

complexity. Their performance is evaluated and compared through 3GPP-based simulations.

In the third part, we investigate the advantages of applying IBFD radios for physical layer security (PLS). We focus on a channel frequency response (CFR)-based secret key generation (SKG) scheme in MIMO systems. We formulate the intrinsic imperfections of IBFD radios (e.g., SIC overheads and noise due to imperfect SIC) and derive their effects on the probing errors. Then we derive closed-form expressions for the secret key capacity (SKC) of the SKG scheme in the presence of a passive eavesdropper. We analyze the asymptotic behavior of the SKC in the high-SNR regime and reveal the fundamental limits for IBFD and half-duplex (HD) radios. Based on the asymptotic SKC, numerical results illustrate that effective analog self-interference cancellation (ASIC) is the basis for IBFD to gain benefits over HD. Additionally, we investigate essential processing for the CFR-based SKG scheme and verify its effectiveness via simulations and the National Institute of Standards and Technology (NIST) test.

In the fourth part, we consider a typical application of IBFD radios: integrated sensing and communication (ISAC). To provide reliable services in high-mobility scenarios, we introduce orthogonal time frequency space (OTFS) modulation and develop a novel framework for OTFS-ISAC. We give the channel representation in different domains and reveal the limitations and disadvantages of existing ISAC frameworks for OTFS waveforms and propose a novel radar sensing method, including a conventional MUSIC algorithm for angle estimation and a delay-time domain-based range and velocity estimator. Additionally, we study the communication design based on the estimated radar sensing parameters. To enable reliable IBFD radios in high-mobility scenarios, a SIC scheme compatible with OTFS and rapidly-changing channels is proposed, which is lacking in the literature. Numerical results demonstrate that the proposed ISAC waveform and associated estimation algorithm can provide both reliable communications and accurate radar sensing with reduced latency, improved spectral efficiency, etc.

Acknowledgements

First of all, I am very grateful to my supervisor Prof. Tharmalingam Ratnarajah who has given me a lot of guidance and help in this process. Without his guidance, I would not have been here. He not only teaches me with his erudite knowledge but also influences me with his patient and rigorous attitude and encourages me with confidence, which makes me enjoy the process pleasurable and peacefully.

I would also like to thank Dr. Navneet Garg and Dr. Abhijeet Bishnu for their generous help while doing the research.

I want to thank my parents, Luo Jianguo and Guo Dechun, for their unconditional trust and support, which allow me to focus on my study and finish this thesis smoothly.

Finally, I would like to thank my girlfriend, Ding Yi, who has been with me and inspired me all the time.

Yi, will you marry me?

Contents

List of Figures

List of Tables

List of Abbreviations

3GPP	3rd Generation Partnership Project
AP	Access point
AoA	Angle of arrival
AoD	Angle of departure
ADC	Analog-to-digital converter
ASIC	Analog self-interference cancellation
AWGN	Additive white Gaussian noise
BS	Base station
BER	Bit error ratio
BPF	Bandpass filter
CP	Cyclic prefix
CCI	Co-channel interference
CSI	Channel state information
CFR	Channel frequency response
CSI-RS	Channel state information reference signal
DL	Downlink
DP	Direct path
DAC	Digital-to-analog converter
DSP	Digital signal processing
DoFs	Degrees of freedom
DSIC	Digital self-interference cancellation
EOM	Electro-optic modulator
EVM	Error vector magnitude
eMBB	Enhanced mobile broadband
FR	Frequency range

FBG	Fiber Bragg grating
FIR	Finite impulse response
FFT	Fast Fourier transform
FDD	Frequency-division duplex
FDC	Frequency domain-based canceller
FSPL	Free-space path loss
FMCW	Frequency-modulated continuous wave
HD	Half-duplex
HWI	Hardware impairment
i.i.d.	Independent and identically distributed
IAB	Integrated access and backhaul
InH	Indoor hotspot
ICI	Inter-channel interference
IDI	Inter-Doppler interference
ISI	Inter-symbol interference
ISD	Inter-site distance
IIoT	Industrial Internet of Things
IBFD	In-band full-duplex
IFFT	Inverse fast Fourier transform
ISFFT	Inverse symplectic fast Fourier transform
ISAC	Integrated sensing and communication
JPABF	Joint power allocation and beamforming
JPAIM	Joint power allocation and interference management
KPI	Key performance indicator
KKT	Karush–Kuhn–Tucker
KGR	Key generation rate
KDR	Key disagreement rate
LS	Least square
LOS	Line of sight
LNA	Low-noise amplifier
ML	Maximum likelihood
MP	Massage passing
MAC	Media access control

mMTC	Massive machine-type communications
mmWave	Millimeter wave
MCMU	Multi-cell multi-user
MMSE	Minimum mean-squared error
MRTC	Maximum ratio transmission and combining
MIMO	Multiple-input multiple-output
MWSR	Maximum weighted sum rate
MUSIC	Multiple signal classification
NN	Neural network
NR	New radio
NSP	Null-space projection
NLOS	Non-line of sight
NIST	National institute of standards and technology
OMP	Orthogonal matching pursuit
OFDM	Orthogonal frequency division multiplexing
OTFS	Orthogonal time frequency space
PA	Power amplifier
PD	Photo-detector
PAC	Photonics-assisted canceller
PLS	Physical layer security
PSD	Power spectral density
PSS	Primary synchronization signal
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
RP	Reflection path
RF	Radio frequency
RMa	Rural macrocell
RMS	Root mean-square
RSI	Residual self-interference
RSS	Residual signal strength
RSSI	Residual signal strength indicator
SE	Spectral efficiency
SI	Self-interference

SDR	Software-defined radio
SoI	Signal of interest
SIC	Self-interference cancellation
SKC	Secret key capacity
SKG	Secret key generation
SNR	Signal-to-noise ratio
SINR	Signal-to-interference-plus-noise ratio
SISO	Single-input single-output
SFFT	Symplectic fast Fourier transform
TDC	Time domain-based canceller
TDD	Time-division duplex
TDL	Tapped delay line
UE	User equipment
UL	Uplink
UMi	Urban-micro
UMa	Urban-macro
URLLC	Ultra-reliable low-latency communications
V2V	Vehicle-to-vehicle
VOA	Variable optical attenuator
VODL	Variable optical delay line
WSR	Weighted sum rate
WLAN	Wireless local area networks
ZF	Zero-forcing

List of Notations

a	Scalar
$\mathbf{a}$	Vector
$\mathbf{A}$	Matrix
$\mathbf{I}$	Identity matrix
$\{a_i\}_{i=1}^{N}$	Set
$\delta(\cdot)$	Dirac delta function
$\lvert a \rvert$	Absolute value
$\lvert \mathbf{A} \rvert$	Determinant
$\lVert \mathbf{A} \rVert$	Euclidean norm
$\lVert \mathbf{A} \rVert_F$	Frobenius norm
$(\mathbf{A})^{-1}$	Inverse
$(\mathbf{A})^{\dagger}$	Hermitian
$(\mathbf{A})^{T}$	Transpose
$\bar{a}$	Complex conjugate
$(\cdot)^{*}$	Optimal solution
$tr(\cdot)$	Trace
$Cov(\cdot)$	Covariance
$Var(\cdot)$	Variance
$\mathbb{E}\{\cdot\}$	Expectation operator
$\mathcal{D}(\mathbf{A})$	Diagonal matrix containing the elements along the diagonal of $\mathbf{A}$
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}$	Set of real numbers
$\mathcal{R}\{\cdot\}$	Real
$\mathcal{I}\{\cdot\}$	Imaginary
max	Maximize

min	Minimize
lim	Limit
$\lceil \cdot \rceil$	Ceiling
$\lfloor \cdot \rfloor$	Floor
$\lceil \cdot \rfloor$	Round
$a * b$	Convolution
$[\mathbf{A}]_{i,j}$ or $\mathbf{A}[i,j]$	The element on the i^{th} row and j^{th} colounm of $\mathbf{A}$
$\mathcal{U}(a,b)$	Uniform distribution within the range of $[a,b]$
$\mathcal{N}(a,b)$	Real Gaussian random variable with mean a and variance b
$\mathcal{CN}(a,b)$	Complex Gaussian random variable with mean a and variance b
$\int_x (\cdot) dx$	Integral over x
$\frac{\partial}{\partial x}$	Partial derivative
$\mathcal{O}(\cdot)$	Order of magnitude
$U(\cdot)$	Utility
$\log_2$	Base 2 logarithm
$\log_{10}$	Base 10 logarithm
$\arctan(\cdot)$	Arctangent
$\angle$	Angle
$\underline{\nu}_{1:x}[\mathbf{A}]$	Eigenvectors associated with the x^{th} smallest eigenvalues of $\mathbf{A}$.
$\bar{\nu}_{1:x}[\mathbf{A}]$	Eigenvectors associated with the x^{th} largest eigenvalues of $\mathbf{A}$.
$\mathbb{N}[\cdot]$	Normalization operator
$\text{vec}(\mathbf{A})$	Vector obtained by stacking columns of $\mathbf{A}$
$I(a;b)$	Mutual information
$I(a;b\|c)$	Conditional mutual information
$H(\cdot)$	Entropy
$\odot$	Element-wise multiplication
$./$	Element-wise division

List of Symbols

c	Light speed
e	Euler's number
f_c	Carrier frequency
f_D	Doppler frequency
Δf	Subcarrier spacing
Δ_τ	Probing interval
σ^2	Variance
Δ, $\mathbf{\Delta}$	Error
$\bigtriangledown$	Gradient
s, $\mathbf{s}$	Transmitted baseband symbol
r, $\mathbf{r}$	Received baseband symbol
x, $\mathbf{x}$, $\mathbf{X}$	Transmitted signal
y, $\mathbf{y}$, $\mathbf{Y}$	Received signal
d_{tx}, $\mathbf{d}_{\text{tx}}$, $\mathbf{D}_{\text{tx}}$	Transmitter's HWIs
d_{rx}, $\mathbf{d}_{\text{rx}}$, $\mathbf{D}_{\text{rx}}$	Receiver's HWIs
n, $\mathbf{n}$, $\mathbf{N}$	Receiver's AWGN
κ	Transmitter HWI factor
β	Receiver HWI factor
ϱ	Pathloss
l	Insertion loss
w	Tunable weight
τ	Delay
ς_m	Loss of the m^{th} tap of canceller
ϕ	Tuanle phase in Chapter 2
	AoD in Chapter 6

θ	Phase shifting in Chapter 2
	AoA in Chapter 6
φ	Phase noise
ξ	RF chain gain
ρ	Peak-to-average power ratio in Chapter 2
	Correlation coefficient in Chapter 5
ρ'	Correlation coefficient
b	Effective number of bits of the ADC
g, G[1], $\mathbf{G}$	SI channel or canceller response
$g_{\text{tx}}/g_{\text{rx}}$	Tx/Rx pulse shaping filter
h	Channel coefficient
ψ_p	p^{th} order basis function
P[2]	Power
μ	Learning rate
Λ_h	SINR for channel estimation
ϖ	Lagrange multiplier in Chapter 4
	Probing error plus channel strength in Chapter 5
$\mathbf{F}_{\text{RF}}$	RF precoder
$\mathbf{W}_{\text{RF}}$	RF combiner
$\mathbf{H}$	Actual CSI
$\hat{\mathbf{H}}$	Estimated CSI
$\mathbf{V}$	Digital precoder
$\mathbf{U}$	Digital combiner
$\mathbf{T}$	Covariance matrix of transmitted signal
$\mathbf{C}$	Covariance matrix of received signal
$\mathbf{\Omega}$	Covariance matrix of interference plus noise
α	Coupling factor in Chapters 3
	Power coefficient for DL user in Chapter 4
	Ratio of SIC overheads in Chapter 5
γ	Power coefficient for UL user
R	Achievable rate
λ	Lagrange multiplier

[1] G also denotes the number of cells but in Chapter 4 only.
[2] P denotes the nonlinearity order for the DSIC but in Section 2.5 only.

ν	Lagrange multiplier in Chapter 4
	Doppler shift in Chapter 6
ϑ	SIC depth in dB
η	SIC depth in scale
ε	MSE
ϵ	RSI power
C_k	Secret key capacity
$\mathcal{M}$	Set of binary bits
$\mathbf{a}$	Steering vector

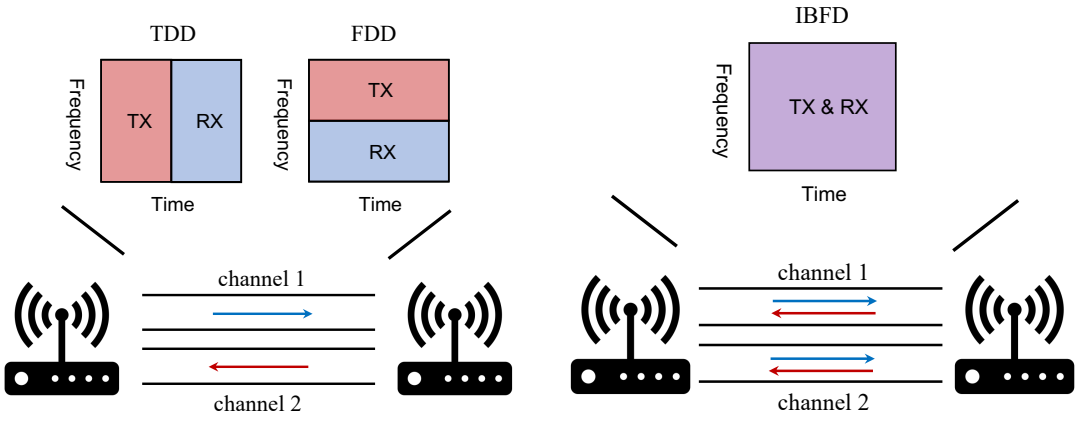
Chapter 1

Introduction

1.1 Background

Since wireless communication became a topic of study in the 1960s, there has been a consensus that it is generally not possible for radios to transmit and receive simultaneously over the same frequency [6]. Thus, radios have been working in half-duplex (HD) mode that they either transmit and receive in different time slots (i.e., time division duplex (TDD)) or over orthogonal channels (i.e., frequency division duplex (FDD)), as Figure 1.1(a) depicts, since the first generation (1G) analog radio communication systems. However, the wireless communication society never gives up pursuing in-band full-duplex (IBFD) radios (i.e., transmit and receive simultaneously over the same channel, as Figure 1.1(b) depicts). The main driving force for researching IBFD radios is their potential to double the spectral efficiency compared to existing HD radios.

Higher throughput has been an important goal in the evolution of wireless networks. Looking back at the previous generations of wireless network standards, the peak data bit rate has increased from 2.4 Kbps in 1G to 20 Gbps in the fifth generation (5G). In addition, wireless communication techniques are being employed in many vertical industries, not just limited to voice service and short messaging services. For instance, 5G has been utilized in the vehicle to everything, e-health, smart utilities, Industry 4.0 manufacturing, etc. [7]. With the concept of connected things and the emergence of many new applications, ultra-low latency has gradually become an important factor of wireless communications, e.g., 1 ms in 5G. Future sixth generation (6G) wireless communication



(a) Half-duplex (HD) (b) In-band full-duplex (IBFD)

FIGURE 1.1: Schematic figures of HD and IBFD radios.

networks intend to further transform connected things into connected intelligence. For this purpose, a continuously increasing throughput, shortening latency, and improving capacity to serve the explosive growth of devices in the connected intelligence world are required. In addition to the spectral efficiency and latency, the requirements for security, mobility, connectivity density, energy efficiency, etc., are moving towards higher key performance indicators (KPI) in the process of network evolution. It is envisioned that 6G will increase the respective capability by a factor of 10-100 to the previous mobile generation upgrades [8], as Figure 1.2 shows [1]. Meanwhile, it is predicted that the

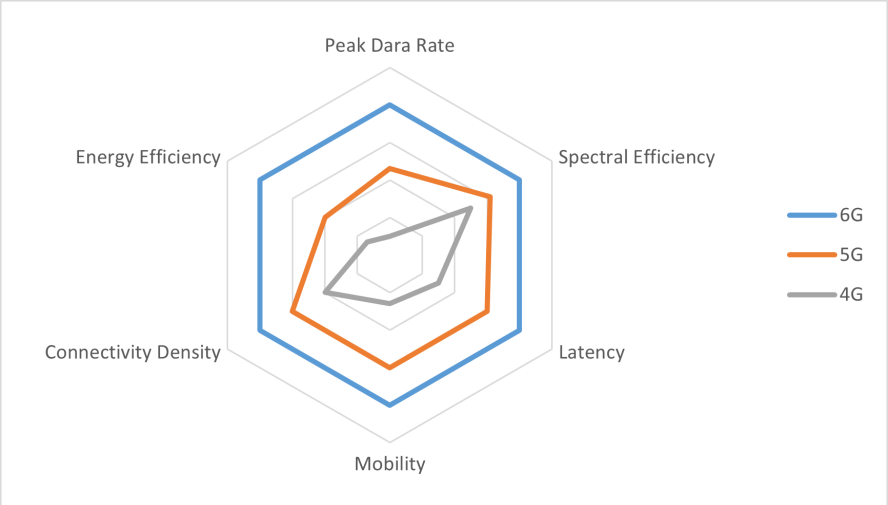


FIGURE 1.2: Key performance indicators of 6G compared to 4G and 5G [1].

global mobile traffic volume will increase exponentially in the next decade, and Figure 1.3 shows the predicted growth of global mobile connectivity during 2020-2030 [2].

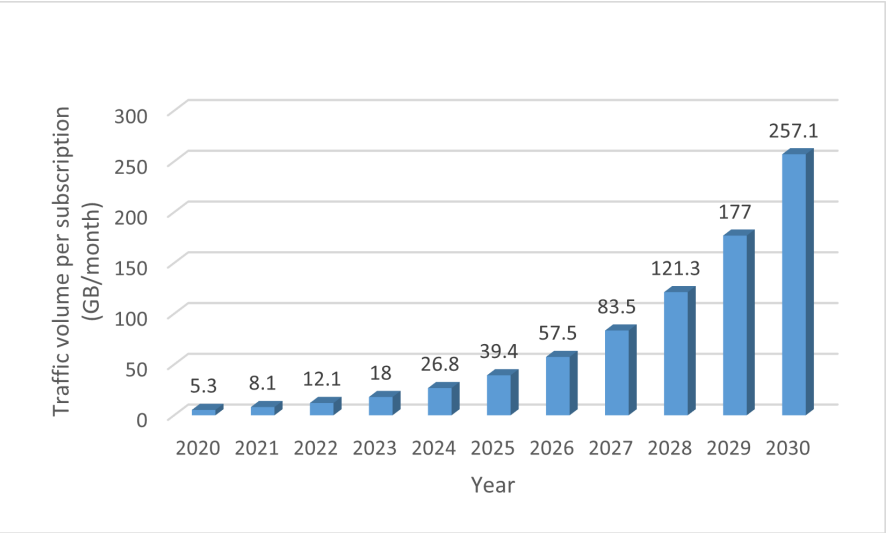


FIGURE 1.3: The predicted growth of global mobile connectivity [2].

These requirements require sufficient spectral resources to be satisfied, which leads to the continuous movement of the wireless network's operating bands to higher frequencies, i.e., from up to 3.4 kHz in 1G to up to 50 GHz in 5G, and 6G is moving to THz communications. Although the communication spectrum is expected to move to the THz band (i.e., within the range of 275 GHz-3 THz) that has not yet been allocated for any purpose worldwide [2], it is still critical to enhancing the spectrum utilization for these stringent requirements and explosive growth of mobile connectivity. It is envisioned that 6G requires up to 100 bps/Hz [2] of spectral efficiency (SE), which could be 5-10 times higher than the requirement of 5G [1]. To meet the improved SE and connectivity requirements, IBFD research is motivated to complement and sustain the evolution of 5G toward future networks. IBFD allows uplink and downlink users to simultaneously use all time-frequency resources, providing the potential to double spectral efficiency, reduce latency, and specific benefits to some applications, etc., which have been demonstrated by many studies. However, most of the existing IBFD research in the literature is based on 4G or wireless local area networks (WLAN) protocols, which are out-of-date even for existing 5G networks. Therefore, it is necessary and timely to study IBFD based on the requirements of future networks in terms of both implementation and applications.

1.2 Contributions

Although IBFD has not been included in current standards of wireless networks, many studies have verified its feasibility through experiments and revealed its advantages in terms of spectral efficiency improvement, latency reduction, relay efficiency, etc. Such results lead us to believe that in the near future, it is possible to introduce IBFD into the wireless network standards to obtain its various benefits. However, with the evolution of the wireless network, the parameters and requirements (e.g., operating frequencies, bandwidth, etc.) will be greatly improved compared to those considered in the existing literature. In addition, the advantages of IBFD for many emerging technologies have not been thoroughly investigated. While previous research has largely demonstrated the feasibility and advantages of IBFD in specific contexts, there are significant gaps that this thesis aims to address. We focus on the emerging challenges in future wireless networks and the application of IBFD to novel technologies. Below are our distinct contributions:

- *Interference mitigation in IBFD systems:* Introduction of a 3-step self-interference cancellation (SIC) technique specifically tailored for both single-input single-output (SISO) and multiple-input multiple-output (MIMO) systems, and a game theory-based resource allocation algorithm to suppress co-channel interference (CCI). We start with SISO systems and review the SIC techniques in the propagation, radio frequency (RF), and digital domains, focusing on the time domain multi-tap canceller and nonlinear digital cancellation. We derive the optimal analog SIC (ASIC) depth by minimizing the residual self-interference (RSI) power. Moving to MIMO systems, we propose to utilize RF beamforming to reduce the SIC implementation cost and analyze its effects on the RF canceller design and overall cancellation performance. Furthermore, we investigate digital beamforming for SIC and propose a minimum mean-squared error (MMSE)-based scheme, which could enhance the SIC capacity while minimizing the sum rate loss. To reduce the CCI, we propose a game theoretic user allocation algorithm to assign user pairs that may cause significant CCI into orthogonal resources, minimizing it from the physical layer.

- *Photonics-assisted RF cancellation:* Pioneering the use of off-the-shelf photonics components for wideband RF cancellation. We give the general architecture and corresponding tuning algorithms of the photonics-assisted cancellers (PAC) and

propose a fiber array canceller design whose effectiveness is verified by mathematical derivation and hardware simulations. Additionally, we carry out a 2-tap prototype of such a design and combine it with the antenna isolation and linear frequency domain-based digital canceller to verify and evaluate their performance in real-world environments.

- *Beamforming in IBFD multi-cell multi-user (MCMU) networks:* Adaptation and enhancement of existing linear beamforming schemes like zero-forcing (ZF), maximum ratio transmission and combining (MRTC), and minimum mean-squared error (MMSE) to suit IBFD-MCMU networks, and a cross-designed power allocation policy and beamforming scheme optimized for MMSE. We compare them in terms of complexity and achievable spectral efficiency. Their performance is evaluated under various conditions and 3rd Generation Partnership Project (3GPP)-specified scenarios to provide practical insights. In order to further improve the performance, we crossly design the power allocation policy and beamforming scheme since they affect each other. We formulate a joint optimization problem based on MMSE and derive the solutions by decomposing it into convex sub-problems.

- *Physical layer security (PLS) in IBFD:* First comprehensive study focusing on the application of IBFD radios for enhancing the performance of physical layer-based secret key generation (SKG) schemes. We focus on a channel frequency response (CFR)-based SKG scheme in MIMO orthogonal frequency division multiplexing (OFDM) systems and derive the secret key capacity (SKC) for IBFD and HD radios, respectively. We reveal the fundamental limitations on the key capacity through asymptotic behavior analysis and derive the conditions on which IBFD operation can offer a higher SKC than its HD counterpart. For practical implementation, we study the properties of the collected samples and propose an averaging preprocessing and segmental quantization scheme to guarantee the effectiveness of the generated key.

- *Integrated sensing and communication (ISAC) framework:* Development of a new framework for ISAC specifically designed for high-mobility scenarios in future networks. To support high-mobility scenarios that emerging applications in the future network may need to consider, we introduce orthogonal time frequency space (OTFS) modulation as an alternative to existing OFDM. We propose a

low-complexity radar sensing design to acquire the angle, velocity, and range information based on the fact that the ISAC receiver has full knowledge of the transmitted signal. Then, we propose SIC schemes based on the obtained radar sensing parameters and propose a variant of the multi-tap canceller to support large Doppler shifts. Additionally, we discuss the transmission scheme by crossly considering radar sensing and communication to improve spectral efficiency and reduce latency.

Through these contributions, we not only extend the applicability of IBFD to future network requirements but also lay the groundwork for its integration into emerging technologies. This thesis offers novel techniques and frameworks that push the boundaries of current IBFD research, providing a comprehensive guide for the design and deployment of IBFD in next-generation wireless networks.

1.3 Thesis overview

The rest of this thesis is organized as follows.

- Chapter 2 studies SIC techniques. It first considers SISO systems and presents an overview of antenna interface designs, a time domain multi-tap canceller, and a nonlinear digital canceller. Then, we move to MIMO systems and discuss the SIC cost, providing a solution to a feasible implementation. Finally, we propose a game theoretic user allocation algorithm to reduce the CCI from the physical layer.

- Chapter 3 focuses on the photonics-assisted RF canceller design to break the limitation of conventional electrical delay lines. We first give a general architecture and two tuning algorithms of this canceller. Then, it focuses on a fiber array-based canceller and demonstrates the experiment results of a 3-step scheme that consists of antenna isolation, a 2-tap fiber array-based canceller, and digital cancellation.

- Chapter 4 considers MCMU networks with multi-antenna users and IBFD base stations and studies the beamforming schemes for such networks. We investigate

linear beamforming schemes and compare their complexity and performance, followed by a beamforming cancellation discussion. Then, we crossly design the power allocation and beamforming to obtain a jointly optimal scheme.

- Chapter 5 investigates the advantages of IBFD for PLS. We consider a CFR-based SKG scheme and derive the secret key capacity via measurable metrics with practical imperfections. Then, we analyze its asymptotic behavior in the high signal-to-noise ratio (SNR) region. Finally, we give implementation details of such schemes and verify their effectiveness through simulations.

- Chapter 6 develops a novel OTFS-ISAC framework. We give the mathematical derivations of OTFS modulation and present the wireless channel in different domains. Then, we introduce a low-complexity radar sensing scheme followed by an IBFD communication design, which includes the SIC techniques and transmission schemes.

- Chapter 7 summarizes the contributions of our works and provides directions for future work.

Chapter 2

Interference Suppression for IBFD Radios

2.1 Introduction

Although IBFD has tremendous advantages over existing HD (either TDD or FDD) systems, as highlighted earlier in Chapter 1, it has not been employed in any current wireless networks yet. The reason is that IBFD operation introduces many additional interferences since it does not separate the uplink and downlink transmissions in the time or frequency domain as in HD systems. Therefore, when an IBFD radio transmits signals, they will be heard by its own receiver, which is known as self-interference (SI). In addition, its transmission will be received by all receivers within the range, causing CCI. The additional interference will severely compromise the performance of IBFD radios either in the point-to-point system or cellular network.

SI is regarded as the main obstacle to effective IBFD radios. The transmitted RF waveforms are rapidly attenuated with distance in the air, as the free-space path loss (FSPL) formula given in [6] suggests, so the received SI is much stronger than the signal of interest (SoI). Figure 2.1 shows the pathloss of wireless channels, providing an intuitive understanding. The unwanted SI suffers a very small attenuation due to the proximity of the co-located transmitter and receiver, while the desired SoI suffers a large attenuation

Work in this chapter has been published in China Communications October 2021 [10] and IEEE Access, December 2021 [9], with a preliminary version presented at IEEE WCNC April 2022 [11] and IEE ICC May 2022 [12].

since it comes from a relatively far transmitter. The difference between the power of received SI and SoI could be 100 dB higher. The limited dynamic range (usually < 100 dB) of practical receivers restricts SI from being eliminated in the digital domain, which could be much easier. In contrast, the received SoI is usually within the limited dynamic range, so it could be dealt with by digital processing.

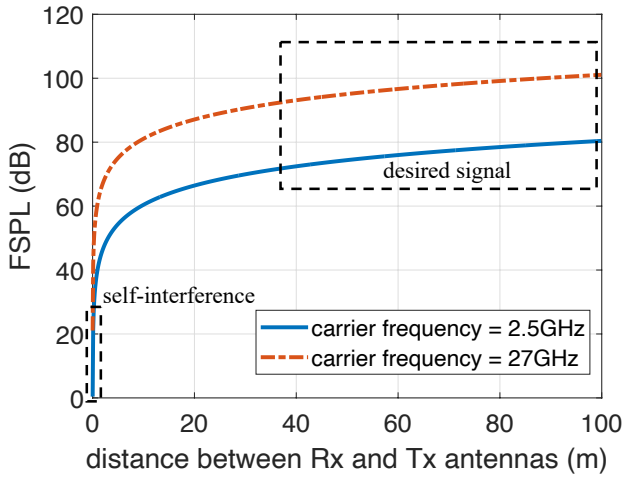


FIGURE 2.1: Pathloss of wireless channels calculated from the formula in 3GPP documents [3] (urban micro scenarios).

Without suppressing SI before the receiver, it will be saturated and introduce distortions that are challenging to be eliminated in the digitized signal, leaving significant SI effects even though the SI itself may have been eliminated. Thus, it is necessary to suppress SI sufficiently in the analog domain. However, eliminating SI in the analog domain, if achievable, is complex and energy- and finance-costly. Therefore, SI is usually only suppressed to be within the receiver's dynamic range in the analog domain, while the RSI will be further mitigated by digital SIC (DSIC) to approach the receiver's noise floor. The SI is usually suppressed through a 3-step SIC scheme for effectiveness and cost consideration, where the SI power variant is depicted in Figure 2.2. Most of the reported real-world experiments used such a scheme [4, 13–16][1].

The first step of SIC is realized by well-designed antenna interfaces, which provide Tx-Rx isolation in the propagation domain. There are multiple passive cancellation mechanisms proposed, which have been summarized for separate antennas in [17] and for shared antennas in [18]. The reported separation configurations, polarization designs, near-field

[1]Some studies did not use analog SIC but successfully demonstrated effective SIC since they transmit a relatively low power and the receiver has a large dynamic range.

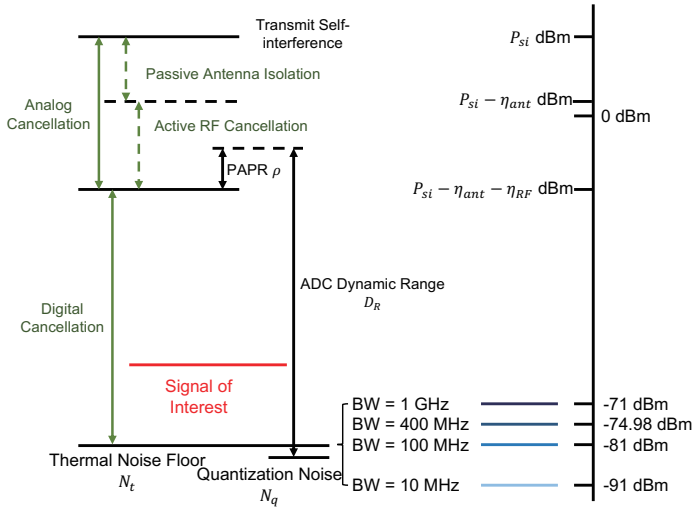


FIGURE 2.2: Power budget for an in-band full-duplex node.

cancellation with various numbers of antennas, isolation feed network, and decoupling structure are listed with operating bans and cancellation depths in Tables 1-4 of [19], then these techniques are compared in Table 5 in terms of antenna size, pros, and cons. Among this wide variety of techniques, physical separation is the most fundamental approach, which could provide up to 45.3 dB cancellation within a 50 cm distance. The directional isolation, absorptive shielding, and cross-polarisation mechanisms have been extensively studied in [20] through experimental measures, and a combination of these techniques can provide up to 73.8 dB of direct path isolation.

The second step happens in the RF domain before the received signal is digitized. Additional circuits are required to generate the RF cancellation signal to subtract it from the received signal. Depending on the tapping location of the reference signal, it can be realized by the auxiliary chain-assisted and multi-tap cancellers, which are also known as *Rice* and *Stanford* architectures [21]. The auxiliary chain-assisted method (i.e., *Rice* architecture) doubles the transmitter chains and needs complex pre-processing to guarantee performance, so it is relatively rarely used. Nevertheless, authors in [22] have demonstrated $>$ 50 dB of cancellation depth over 20 MHz using this canceller. Most of the studies focus on the time-domain multi-tap canceller design. In [23], authors investigate the RF components for practical implementation, where the impacts of components' resolutions and estimation errors are analyzed. In addition, the multi-tap canceller can also be constructed in the frequency domain. The two types of multi-tap cancellers have been realized and demonstrated in practical systems using off-the-shelf

components [10, 15, 17] and integrated circuits [18] with tradeoffs between various performance metrics, such as achievable cancellation depth, Tx power handling, antenna loss, noise figure, power consumption, etc.

Finally, the RSI is digitized and eliminated in the digital domain. Traditional DSIC is model-based, which estimates the effective SI coupling coefficients and regenerates the baseband cancellation signal. The performance is determined by the adopted models. To capture the nonlinearity of transceivers, nonlinear models are usually utilized, which show improvement of cancellation depth with enhancing transmit power compared to linear models [24]. The capability of non-linear models and estimation approaches are summarized in [25]. In the era of artificial intelligence, neural networks (NN) are widely studied for DSIC. NN-DSIC is data-driven, and it learns the nonlinear effects from data. Thus, it is not necessary to use a high-order and long-length memory model, reducing the complexity. A single-layer feed-forward NN with 6 input, 5 hidden, and 2 output nodes can approach the performance of a polynomial memory model-based canceller with 36% multiplications reduction [26]. Researchers in the signal processing and wireless communications communities further develop low-complexity NN-DSIC to make them suitable for implementation in mobile devices [27, 28], and it is envisioned that the auto-encoder is a promising technique to enhance the DSIC in terms of complexity, convergence speed, and cancellation depth [25].

With a combination of these techniques, the SI could be eliminated. For instance, with a combination of antenna isolation, multi-tap RF canceller, and nonlinear digital canceller, over 100 dB and 110 dB cancellation depths are realized within 80 MHz and 20 MHz bandwidth centered at 2.56 GHz in [29]. However, there are many techniques for the three steps, and it is not clear how to choose the most appropriate method. Additionally, these studies consider only SISO or very small-scale MIMO systems (e.g., $\leqslant$ 2 antennas), while the scalability of the SIC with antenna arrays remains to be studied.

The main contribution of our work is that we review existing cancellation techniques and analyze their advantages and disadvantages, focusing on the most promising techniques. In addition, we investigate the inter-effect of the three steps and find the optimal ASIC depth. Then, we consider the implementation of the 3-step cancellation scheme in MIMO systems and propose a feasible solution through RF beamforming, analyzing its effects on the canceller design and performance. In addition to the SIC, we also consider the CCI

effects and propose to minimize its effects from the physical layer via a game theoretic user allocation algorithm.

The rest of this chapter is organized as follows. In Section 2.2, we give preliminaries that are used in the later sections, including the IBFD transceiver architecture, SI channel model, and hardware impairment models. Section 2.3 provides an overview of the antenna interface designs to provide passive SIC in the propagation domain. Then, Section 2.4 introduces three RF canceller architectures and focuses on the time domain-based multi-tap canceller, detailing the cancellation rationale and implementation method. Section 2.5 studies a nonlinear digital canceller and derives the optimal ASIC depth by minimizing the RSI power. In Section 2.6, a hybrid beamforming architecture is introduced to provide an alternative implementation for SIC in MIMO systems, and its feasibility and performance are analyzed. Finally, a user allocation algorithm is proposed in Section 2.7 to reduce the CCI. Numerical results are given in Section 2.8, and conclusions are drawn in Section 2.9.

2.2 Preliminaries

2.2.1 System model

In this section, the received SI is modeled in detail to provide a mathematical perspective of the cancellation mechanism. Figure 2.3 depicts an IBFD transceiver with SIC configurations [2]. The transmitter and receiver can share a single antenna through different interfaces (i.e., shared antennas) or connect to separate antennas. For shared antennas, the limited isolation between interfaces causes the Tx signal to leak to the receiver, yielding the direct path (DP) coupling component. In addition, the antenna receives its own Tx signals reflected by objectives in the surrounding environment, yielding reflection path (RP) components. For separate antennas, the DP component comes from the line of sight (LOS) over-the-air propagation between Tx and Rx antennas, and RP components also come from the reflectors.

[2]Figure 2.3 depicts a MIMO transceiver, while we consider the single antenna system for the 3-step SIC design. The SIC for MIMO systems will be discussed later in Section 2.6.

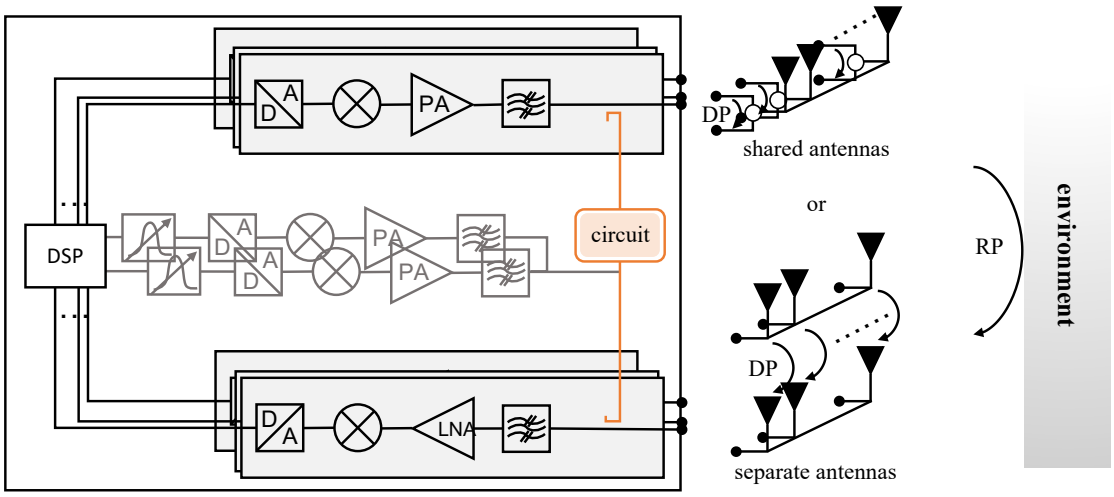


FIGURE 2.3: The architecture of a multi-antenna IBFD transceiver, which is configurable to cope with SI.

2.2.2 Self-interference channel model

Generally, the self-interference coupling channel consists of the LOS component and non-LOS (NLOS) components due to scattering, and it can be described by the tapped delay line (TDL) model as

$$g_{\mathrm{si}}(t) = \sum_{l=0}^{L} h_{\mathrm{si},l}(t) \cdot \delta\left(t - \tau_{\mathrm{si},l}\right), \tag{2.1}$$

where L is the number of reflection paths, and $h_{si,l}$ and $\tau_{si,l}$ denote the channel coefficient and delay of the l^{th} path, respectively. The channel coefficient and delay are related as [30]

$$h_{\mathrm{si},l} = \sqrt{f(c\tau_{\mathrm{si},l})}, \quad \text{where} \quad f(x) = \min\left\{\zeta, k\left\|x\right\|^{-\alpha}\right\}, \tag{2.2}$$

where c is the light speed, $\alpha > 2$, ζ and k are constants with appropriate dimension.

2.2.3 Self-interference representation

Let $s(n)$ represent the baseband modulated symbols at the transmitter, which is converted to the analog signal $s(t)$ via the digital-to-analog converter (DAC). Then, it is converted into the RF domain through the transmitter chain, yielding the transmitted

RF signal as[3]

$$x(t) = \xi_{\text{tx}} \cdot s(t) e^{j2\pi(f_c t + \varphi_{\text{tx}}(t))} + d_{\text{tx}}(t) + n_{\text{tx}}(t), \tag{2.3}$$

where f_c denotes the carrier frequency; ξ_{tx} is the total gain of the RF transmitter chain; $n_{\text{tx}}(t)$, $\varphi_{\text{tx}}(t)$, and $d_{\text{tx}}(t)$ represent the time-varying additive white Gaussian noise (AWGN), phase noise, and distortion of the transmitter chain, respectively. According to the wireless SI channel model, the SI perceived by the co-located receiver is given as

$$y_{\text{si}}(t) = x(t) * g_{\text{si}}(t) = \sum_{l=0}^{L} h_{\text{si},l}(t) \cdot x(t - \tau_{\text{si},l}). \tag{2.4}$$

Meanwhile, the IBFD transceiver will receive SoI from the intended terminals. Thus, the signal perceived by the Rx antenna can be written as

$$\tilde{y}(t) = y_{\text{si}}(t) + y_{\text{soi}}(t) = x(t) * g_{\text{si}}(t) + y_{\text{soi}}(t), \tag{2.5}$$

where $y_{\text{soi}}(t)$ denotes the received SoI. Regardless of the cancellation mechanism, we assume the analog canceller can be described by $h_{\text{canc}}(t)$, then the input signal to the receiver chain is given as

$$\begin{aligned} y(t) &= x(t) * [h_{\text{si}}(t) - h_{\text{canc}}(t)] + y_{\text{soi}} \\ &= y_{\text{rsi}}(t) + y_{\text{soi}}(t) \end{aligned} \tag{2.6}$$

where $y_{\text{rsi}}(t)$ denotes the RSI after analog cancellation. The RF signal is down-converted into the baseband analog domain via the receive chain, and digitalized by the analog-to-digital converter (ADC), yielding the received baseband symbols as

$$\tilde{r}(n) = \xi_{\text{rx}} \cdot y(n) e^{j2\pi(-f_c n + \varphi_{\text{rx}}(n))} + d_{\text{rx}}(n) + n_{\text{rx}}(n), \tag{2.7}$$

where ξ_{rx}, $d_{\text{rx}}(n)$, and $n_{\text{rx}}(n)$ denote the total gain, phase noise, AWGN, and distortion of the RF receiver chain, respectively. Assume the digital canceller generates a baseband cancellation signal $r_{\text{dsi}}(n)$ regardless of the mechanism, then the clean SI-free symbols

[3]The nonlinear transceiver distortion is described by $d_{\text{tx}}(t)$ here for simplicity, which is a nonlinear function of the transmitted signal that can be characterized by the memory polynomial model given in the later Section 2.5.

to be decoded are represented as

$$\begin{aligned} r(n) &= \tilde{r}(n) - r_{\text{dsi}}(n) \\ &= \left(g_{rx} y^b_{rsi}(n) - s_c(n)\right) + g_{rx} y^b_{soi}(n) + d_{rx}(n) + n_{rx}(n). \end{aligned} \tag{2.8}$$

2.2.4 Transceiver hardware impairments

The linear and nonlinear distortions of the transceiver typically depend on the properties of the electronic components and are related to the power of the input signal, transceiver gains, and input intercept points of electronic components detailed in [31]. Actually, its power is proportional to the power of transmitted or received signals [32]. Suppose κ is the transmitter distortion factor and β is the receiver distortion factor to describe these effects, then the transceiver hardware impairments (HWIs) can be modeled as $d_{tx} \sim \mathcal{CN}\left(0, \kappa \, \mathrm{E}\left(|g_{tx} s|^2\right)\right)$ and $d_{rx} \sim \mathcal{CN}\left(0, \beta \, \mathrm{E}\left(|y|^2\right)\right)$. We assume the limited dynamic range of ADCs is the dominating factor of the receiver for the SI without loss of generality. The quantization noise of the ADC has a well-established theory that the signal-to-interference-plus-noise ratio (SINR) only depends on the number of bits of the ADC and the distribution of the input signal. So, the power of the quantization noise measured in dBm can be modeled as [31]:

$$P_q = -6.02b - 4.76 + \rho + P_{in} \tag{2.9}$$

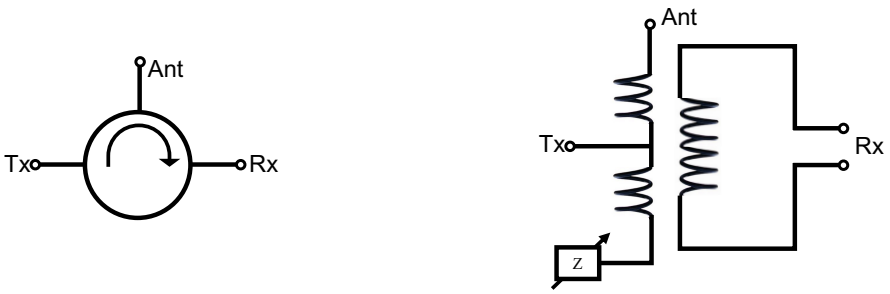
where b is the effective number of bits of the ADC, P_{in} is the input power of the ADC in dBm, and ρ is the peak-to-average power ratio of the input signal in dB.

2.3 Antenna domain cancellation

This section provides an overview of existing antenna domain cancellation designs, which suppress SI in the propagation domain through antenna interface designs or beamforming schemes (for MIMO systems).

2.3.1 Shared antennas

For shared antennas, *circulators* and *electrical-balance duplexers* are commonly utilized to provide the Tx-Rx isolation. The circulator has only three ports connected to the transmitter, receiver, and the shared antenna, as Figure 2.4(a) depicts. The signal circulates in a single direction within the circulator circle, preventing the Tx signal from propagating to Rx. The electrical-balance duplexer has four ports, as Figure 2.4(b) depicts, where three ports are connected to the transmitter, receiver, and the shared antenna, while the fourth port is terminated by a tunable impedance to match the impedance of the antenna. The two interface designs can be implemented in a variety of ways to meet different performance requirements, e.g., operating frequency, isolation depth, power consumption, antenna loss, area, efficiency, etc. The readers are referred to [18] for more details of shared antenna interfaces, where the state-of-art integrated circuit implementations are comprehensively reviewed.



(a) Circulator (b) Electrical balanced duplexer

FIGURE 2.4: Shared antenna interface designs to provide Tx-Rx isolation.

2.3.2 Separate antennas

For separate antennas, the SIC is implemented through the following four approaches.

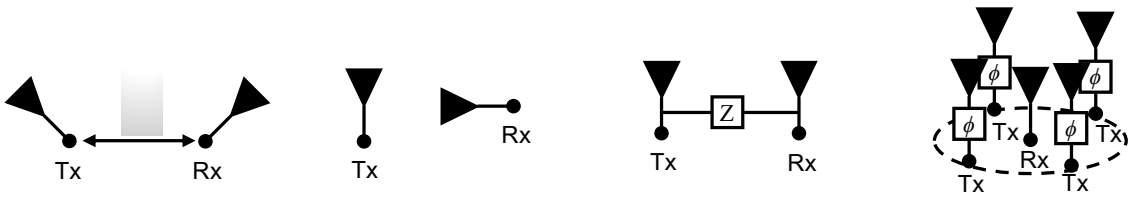
- *Separation* aims to improve the isolation between Tx and Rx antennas by increasing their distance or placing a special object between them, as Figure 2.5(a) depicts. This approach has the advantage of being easy to implement, but its performance is limited by the size of the terminal devices. Nevertheless, it is compatible with other methods, and it is always desired to separate the Tx and Rx antennas as

much as possible while designing the transceiver. The object could be an absorber, a reflector, or a high-impedance surface, which attenuates the radiated signal flow.

- *Cross-polarization* makes the Tx and Rx antennas operate on orthogonal polarizations as Figure 2.5(b) depicts. Taking advantage of the fact that different polarizations do not couple into each other, the Rx antenna can isolate the Tx signal.

- *Coupling network* mitigates SI by inserting a match impedance between Tx and Rx antennas as Figure 2.5(c) depicts. For passive coupling networks, the impedance is fixed to cope solely with the fairly constant direct path coupling. This method can be extended to be active by adaptively tuning the impedance elements, further enhancing the mitigation.

- *Spatial nulling* requires multiple Tx antennas with phase shifters to form a spatial null space where the Rx antenna should be located [4]. The Tx antennas are configured circularly, as Figure 2.5(d) shows. This method may affect the far-field radiation pattern. Although more Tx antennas can compensate for this effect, the cost and power consumption could be problematic.



(a) Separation (b) Cross-polarization (c) Coupling network (d) Spatial nulling

FIGURE 2.5: Separate antenna interface designs to provide Tx-Rx isolation.

2.3.3 Antenna arrays

With antenna arrays, the beam pattern can be adjusted to form highly-directional beams by tuning the amplitude and phase of each element, which is known as beamforming. Therefore, the Tx antenna array can be adjusted to form spatial nulls at a subset of Rx antennas to minimize the SI leakage in the propagation domain, as Figure 2.6 depicts.

[4] Two Tx antennas with phase shifts of 0 and 180 degrees, respectively, could be employed to provide the null space in the middle of the two antennas, which is a specific case of the spatial nulling method and is also known as *anti-phase*.

It could be realized by digital beamforming or RF beamforming. Digital beamforming has high flexibility in signal processing that can adjust the beam pattern on each subcarrier. Thus, digital beamforming can cope with frequency selectivities caused by multipath effects, i.e., it can suppress RP components. It is attractive because it means the complex and costly RF cancellers may no longer be necessary. In contrast, RF beamforming is frequency-independent, so that cannot deal with frequency selectivity. The RF beamformer is usually chosen through beam management according to the location of communication nodes [33]. It forms highly directional narrow beams towards the intended receiver so that the DP coupling will be reduced. Beamforming cancellation has the advantage of low implementation complexity in antenna arrays, but it consumes spatial degrees of freedom (DoFs) that could have been used for the payload of intended terminals, compromising the IBFD gain in terms of throughput.

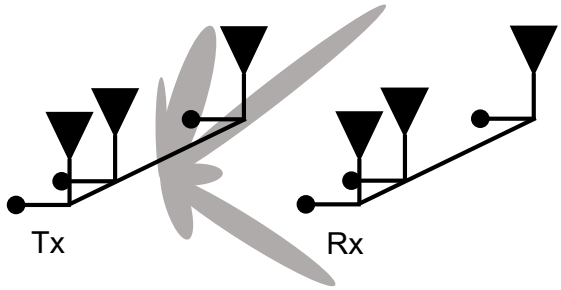


FIGURE 2.6: The formed directional transmitted beams provide spatial nulling.

When multiple DoFs are available in the Rx SI (e.g., with multiple Rx antennas), one can perform Rx beamforming in the digital domain to mitigate the SI effects. The output signals from each Rx channel can be independently weighted and then combined to form a null space of the SI, i.e., the Rx beam pattern is steered to point to the direction that nulls the Rx SI. Still, this will consume DoFs that could have been used for the payload from the intended terminals, reducing the throughput. However, it can achieve extremely low-complexity SIC with antenna arrays, so it is a direction worth studying, especially for large-scale antenna arrays.

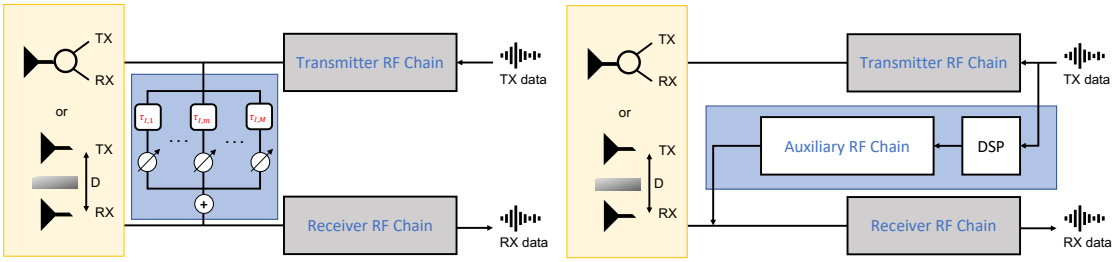
2.4 RF domain cancellation

Antenna domain cancellation designs usually deal with the DP component only, while the RP components are challenging to be mitigated. Therefore, the received SI could

still saturate the receiver if there are strong reflection paths. In order to enable IBFD radios in the real-world environment, the transceiver must be configurable to suppress these RP components, which requires RF cancellers.

2.4.1 Two architectures of RF cancellers

Based on the reference tapping locations, RF cancellers can be classified into *Stanford* architecture and *Rice* architecture, as Figure 2.7 depicts. *Rice* architecture taps the base-

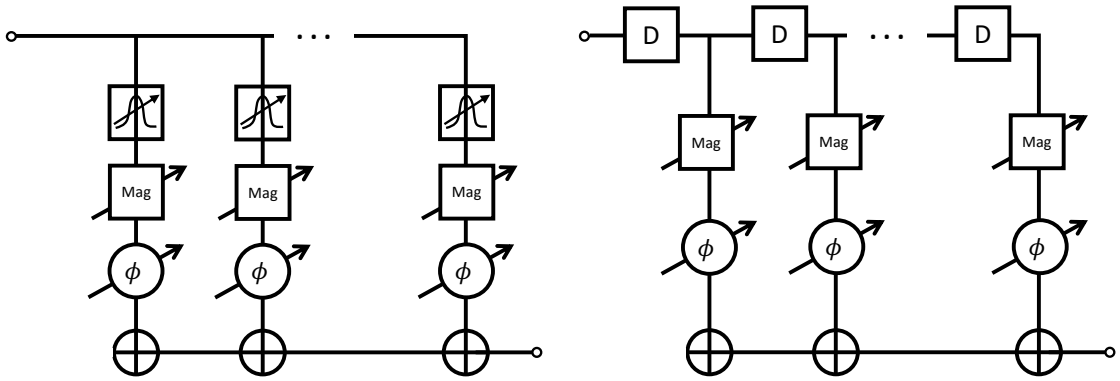


(a) *Stanford* architecture (b) *Rice* architecture

FIGURE 2.7: Two architectures of the RF cancellation design.

band signal as the reference to generate an equivalent baseband replica of the received SI and converts it into the RF domain through auxiliary RF chains for cancellation. Therefore, this method takes advantage of the flexibility of digital signal processing (DSP) at the cost of doubling the transmitter RF chain. In addition, both the ordinary and auxiliary RF chains introduce noise and nonlinearity while converting the baseband signals into the RF domain, so this method needs pre-distortion algorithms to compensate for these effects, yielding high computational complexity.

Stanford architecture taps the reference signal directly in the RF domain and generates the cancellation signal via a multi-tap canceller, which can be regarded as an RF finite impulse response (FIR) filter. The design objective of the multi-tap canceller is to mimic the wireless SI channel. Since the reference signal includes the hardware impairments of transmitter RF chains, it is not necessary to compensate for their effects by pre-processing. Furthermore, due to the propagation attenuation nature of wireless channels, the multi-tap canceller can be built with passive components, saving power consumption. Therefore, Stanford architecture is usually the focus of research. The multi-tap canceller can be implemented in the time or frequency domain as Figure 2.8 depicts. Frequency domain-based canceller (FDC) processes the reference signal at different frequency points



(a) Frequency domain-based design (b) Time domain-based design

FIGURE 2.8: The two architectures of multi-tap cancellers.

selected by bandpass filters, as Figure 2.8(a) depicts. Each tap has a filter with different center frequencies and quality factors, followed by gain and phase controllers. The outputs of these taps are combined to emulate the SI channel over the operation band. The filters can be programmable using a multi-path bandpass filter design, i.e., filters' quality factor and center frequency can be controlled [34].

2.4.2 Time domain-based canceller

Our study focuses on the time domain-based canceller (TDC) design due to its easier implementation. The TDC can be generally described as

$$g_{\text{tdc}}(t) = \sum_{m=0}^{M-1} w_m e^{j2\pi\phi_m} \delta(t - \tau_m), \tag{2.10}$$

where w_m, ϕ_m, and τ_m denote the tunable weight, tunable phase, and fixed delay of the m^{th} tap. The TDC representation has an identical form as the TDL channel model given in Equation (2.1). Thus, the SI can be eliminated by the TDC if their coefficients are perfectly matched [5], i.e., $w_i e^{j2\pi\phi_i} = h_{\text{si},i}$, and $\tau_{\text{si},i} = \tau_i$ for all $i \in [0, L-1]$ (with $M = L - 1$), resulting in $g_{\text{tdc}}(t) = g_{\text{si}}(t)$, as Equation (2.6) suggests. However, this is not feasible in practice due to the varying real-world environment. The number of taps M and the delay of each tap τ_m are fixed as long as the canceller is constructed, which limits its flexibility to adapt to the varying environment. Assuming the maximum delay spread of the SI channel D_s is known while constructing the canceller, the best choice is

[5] The hardware impairments of the TDC are ignored here to demonstrate its basic rationale, while the HWIs will be included in the next Chapter to explore practical limitations.

to uniformly distribute the delays of the M taps within the range of $[0, D_s]$. The weights and phases of each tap can be obtained by solving the optimization problem as

$$\min_{\{w_m,\phi_m\}_{m=0}^{M-1}} \int_t |g_{\text{tdc}}(t) - g_{\text{si}}(t)|^2 \, dt. \tag{2.11}$$

However, we want to highlight that any error in the impulse response will distort all frequency components, as the Fourier transform indicates. Therefore, minimizing the difference between the impulse responses of the TDC and SI channel actually aims at minimizing their difference over the entire frequency domain, while only a specific bandwidth is the band of interest (BoI). With limited degrees of freedom (i.e., number of taps), it is for sure that optimizing over the BoI outperforms optimizing over the entire band [9, 10]. So we tune the canceller in the frequency domain by minimizing the frequency components of the TDC and SI channel within the BoI. Perform Fourier transform to $g_{\text{tdc}}(t)$ and $g_{\text{si}}(t)$, the frequency response of the canceller and SI channel can be obtained as

$$G_{\text{tdc}}(f) = \sum_{m=0}^{M-1} w_m e^{-j2\pi(f\tau_m - \phi_m)}, \tag{2.12}$$

$$G_{\text{si}}(f) = \sum_{l=0}^{L} h_{\text{si},l} e^{-j2\pi f \tau_{\text{si},l}}. \tag{2.13}$$

The optimal weights and phases can be obtained by solving the minimization problem as

$$\min_{\{w_m,\phi_m\}_{m=0}^{M-1}} \int_{f_c-\frac{B_W}{2}}^{f_c+\frac{B_W}{2}} |G_{\text{tdc}}(f) - G_{\text{si}}(f)|^2 \, df. \tag{2.14}$$

Sampling the frequency response at rate $\frac{1}{\Delta f}$, then $\frac{B_W}{\Delta f} + 1$ samples can characterize the canceller and SI channel within the BoI, yielding two vectors as

$$\mathbf{g}_{\text{tdc}} = \left[G_{\text{tdc}}(f_c - \frac{B_W}{2}), G_{\text{tdc}}(f_c - \frac{B_W}{2} + \Delta f), \ldots, G_{\text{tdc}}(f_c + \frac{B_W}{2}) \right]^T, \tag{2.15}$$

$$\mathbf{g}_{\text{si}} = \left[G_{\text{si}}(f_c - \frac{B_W}{2}), G_{\text{si}}(f_c - \frac{B_W}{2} + \Delta f), \ldots, G_{\text{si}}(f_c + \frac{B_W}{2}) \right]^T. \tag{2.16}$$

According to Equation (2.12), $\mathbf{g}_{\text{tdc}}$ can be written in matrix form as

$$\mathbf{g}_{\text{tdc}} = \mathbf{\Pi}_{\text{tdc}} \mathbf{W}_{\text{tdc}}, \tag{2.17}$$

where

$$\mathbf{\Pi}_{\text{tdc}} = \begin{bmatrix} e^{-j2\pi\left(f_c-\frac{B_W}{2}\right)\tau_0} & e^{-j2\pi\left(f_c-\frac{B_W}{2}\right)\tau_1} & \dots & e^{-j2\pi\left(f_c-\frac{B_W}{2}\right)\tau_{M-1}} \\ e^{-j2\pi\left(f_c-\frac{B_W}{2}+\Delta f\right)\tau_0} & e^{-j2\pi\left(f_c-\frac{B_W}{2}+\Delta f\right)\tau_1} & \dots & e^{-j2\pi\left(f_c-\frac{B_W}{2}+\Delta f\right)\tau_{M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi\left(f_c+\frac{B_W}{2}\right)\tau_0} & e^{-j2\pi\left(f_c+\frac{B_W}{2}\right)\tau_1} & \dots & e^{-j2\pi\left(f_c+\frac{B_W}{2}\right)\tau_{M-1}} \end{bmatrix}, \tag{2.18}$$

$$\mathbf{W}_{\text{tdc}} = \left[w_0 e^{j2\pi\phi_0}, w_1 e^{j2\pi\phi_1}, \dots, w_{M-1} e^{j2\pi\phi_{M-1}}\right]^T. \tag{2.19}$$

$\mathbf{g}_{\text{si}}$ can be directly obtained from the channel estimator in the conventional OFDM systems with the sampling interval Δf equal to the subcarrier space. Thus, the minimization problem in Equation (2.14) can be written in the matrix form as

$$\min_{\mathbf{W}_{\text{tdc}}} \left\|\mathbf{\Pi}_{\text{tdc}}\mathbf{W}_{\text{tdc}} - \mathbf{g}_{\text{si}}\right\|^2, \tag{2.20}$$

which can be solved by the Wiener solution given as

$$\overline{\mathbf{W}}_{\text{tdc}} = \left(\mathbb{E}\left\{\mathbf{\Pi}_{\text{tdc}}^{\dagger}\mathbf{\Pi}_{\text{tdc}}\right\}\right)^{-1}\mathbb{E}\left\{\mathbf{\Pi}_{\text{tdc}}^{\dagger}\mathbf{g}_{\text{si}}\right\}. \tag{2.21}$$

However, it is challenging to obtain the statistical knowledge of the canceller and SI channel in varying environments. Thus, the optimal weights are usually obtained via the adaptive learning method with the help of a feedback loop, which consists of a residual signal strength indicator (RSSI), an ADC, and a field programmable gate array (FPGA) to measure the SI channel and RSI power. The instant RSI power is a function of the weights such that $P(\mathbf{W}_{\text{tdc}}^{(t)}) = \left\|\mathbf{\Pi}_{\text{tdc}}\mathbf{W}_{\text{tdc}}^{(t)} - \mathbf{g}_{\text{si}}^{(t)}\right\|_F^2$. The weights could be randomly initialized and be updated based on the instant SI channel until converge, i.e., the RSI power does not decrease with updated weights, as

$$\mathbf{W}_{\text{tdc}}^{(t+1)} = \mathbf{W}_{\text{tdc}}^{(t)} - \mu\nabla_{\mathbf{W}}^{(t)}, \tag{2.22}$$

where $\nabla_{\mathbf{W}}^{(t)} = \frac{\partial P(\mathbf{W}_{\text{tdc}}^{(t)})}{\partial \mathbf{W}_{\text{tdc}}^{(t)}} = \left(\mathbf{\Pi}_{\text{tdc}}^{\dagger}\mathbf{\Pi}_{\text{tdc}}\right)^{-1}\mathbf{\Pi}_{\text{tdc}}^{\dagger}\mathbf{g}_{\text{si}}^{(t)}$ is the gradient, and μ is the learning rate.

2.5 Digital domain cancellation

In this section, we will study the model-based digital cancellation scheme. The model-based method requires appropriate models to describe the baseband input-output relationship of the Tx-Rx link, which includes the transceiver nonlinearity, wireless channel, antenna interfaces, etc. The models can be classified into linear models and nonlinear models. Linear models implement discrete FIR filters, which have a similar architecture as the TDC but have the advantage of high flexibility due to digital processing. However, liner models cannot handle the nonlinearity of transceivers, e.g., IQ imbalance, amplifier nonlinearity, and phase noise. To address this issue, nonlinear models are exploited. We assume the nonlinearity is dominated by the power amplifier (PA) without loss of generality. The nonlinearity effects of actual power amplifiers can be effectively described by the memory polynomial model, so the digitalized baseband RSI symbols in Equation (2.7) can be alternatively written as [35]

$$\tilde{r}(n) = \sum_{p=1,\, odd}^{P} \sum_{k=0}^{K-1} h_{\text{eff},p}(k) \psi_p \left(s\left(n-k\right)\right) + \tilde{z}(n), \tag{2.23}$$

where $\psi_p(s(n)) = |s(n)|^{p-1}\, s(n)$ is the p^{th} order basis function, P is the highest nonlinearity order, K represents the memory length of the power amplifier, and $h_{\text{eff},p}(k)$ is the coefficient of the p-th order effective SI channel that includes the effects of the overall baseband Tx-Rx link, and $\tilde{z}(n)$ represents the interference[6] and noise (i.e., the residual terms in Equation (2.7)). The effective channel coefficients can be estimated with the received (i.e., $\bar{r}(n)$) and know transmitted symbols (i.e., $s(n)$), which can be given by the common least square (LS) estimator as [35]

$$\hat{\mathbf{h}}_n = \left(\Psi^{\dagger}(n) \Psi(n) \right)^{-1} \cdot \Psi^{\dagger}(n) \cdot \tilde{r}(n), \tag{2.24}$$

where $\hat{\mathbf{h}}_n$ and $\Psi(n)$ denote the stacked estimated coefficient and basis function vectors for the n^{th} time sample given as

$$\hat{\mathbf{h}}_n = [h_{\text{eff},1}(0)\ h_{\text{eff},1}(1)\ \cdots\ h_{\text{eff},1}(K-1)\ \cdots\ h_{\text{eff},P}(0)\ h_{\text{eff},P}(1)\ \cdots\ h_{\text{eff},P}(K-1)]^T, \tag{2.25}$$

[6] The received SoI is regarded as interference for effective SI channel estimation.

$$\begin{aligned}\Psi(n) = [&\psi_1(s(n))\ \psi_1(s(n-1))\ \cdots\ \psi_1(s(n-K+1))\ \cdots \\ &\psi_P(s(n))\ \psi_P(s(n-1))\ \cdots\ \psi_P(s(n-K+1))]^T.\end{aligned} \tag{2.26}$$

The estimated effective channel coefficients $\hat{\mathbf{h}}_n$ and the actual effective channel coefficients $\mathbf{h}_n$ are related as

$$\hat{\mathbf{h}}_n = \mathbf{h}_n + \mathbf{e}_n, \tag{2.27}$$

where $\mathbf{e}_n \sim \mathcal{CN}(\mathbf{0}, \frac{1}{\Lambda_h}\cdot\mathbf{I})$ is the LS estimation error with Λ_h denoting the SINR. Consider a point-to-point communication scenario, the SINR of effective SI channel estimation can be given as

$$\Lambda_h = \frac{10^{(P_{\text{si}}-\vartheta_{\text{ant}}-\vartheta_{\text{rf}}-30)/10}}{10^{(P_{\text{int}}-\varrho_p^{\text{int}}-30)/10} + \sigma^2_{\text{canc}} + \sigma^2_{N_f} + 10^{(P_{\text{si}}-\vartheta_{\text{ant}}-\vartheta_{\text{rf}}-30+\beta)/10}}, \tag{2.28}$$

where P_{si} and P_{int} denotes the transmit power of the SI and SoI in dBm, ϱ_p^{int} represents the pathloss of the channel from the intended node to the IBFD node in dB, ϑ_{ant} and ϑ_{rf} represent the antenna isolation depth and RF cancellation depth in dB, $\sigma^2_{N_f} = 10^{(N_f-30)/10}$ is the receiver noise floor with $N_f = -174 + 10\times\log_{10}(B_W) + 13$ in dBm assuming -174 dBm/Hz of thermal noise density and 13 dB of noise figure, and σ^2_{canc} and σ^2_{rx} denote the power of canceller noise and receiver distortions, respectively. Usually, an interference-free period could be provided by the media access control (MAC) protocol to remove the term $10^{(P_{\text{int}}-\varrho_p^{\text{int}}-30)/10}$ and enhance the SIC performance.

Then, the baseband cancellation signal is generated as

$$r_{\text{dsi}}(n) = \sum_{p=1,\,odd}^{P}\sum_{k=0}^{K-1}\hat{h}_{\text{eff},p}(k)\psi_p\left(s\left(n-k\right)\right). \tag{2.29}$$

The received signal subtracts the generated cancellation signal, leaving SoI, noise, and errors due to imperfect channel estimation. Equation (2.8) can be written as

$$r(n) = r_{\text{soi}}(n) + \sum_{p=1,\,odd}^{P}\sum_{k=0}^{K-1}e_p(k)\psi_p\left(s\left(n-k\right)\right) + n_{\text{rx}}(n). \tag{2.30}$$

After the final digital canceller, the RSI comes from effective channel estimation errors and receiver distortions caused by SI, whose power can be given as

$$P_{\text{rsi}} = \frac{10^{(P_{\text{si}}-30)/10}/g_{\text{tx}}}{\tilde{\Lambda}_h} + 10^{(P_{\text{si}}-\vartheta_{\text{alg}}-30+\beta)/10}, \tag{2.31}$$

where $\vartheta_{\text{alg}} = \vartheta_{\text{ant}} + \vartheta_{\text{rf}}$ denotes the total ASIC depth in dB, and $\tilde{\Lambda}_h$ removes the term $10^{(P_{\text{int}} - \varrho_p^{\text{int}} - 30)/10}$ from Λ_h in Equation (2.28). It is obvious that insufficient RF cancellation depth will increase the receiver distortion power, leaving a large noise to SoI. In contrast, too deep RF cancellation will reduce the SINR for effective channel estimation, increasing the errors. Thus, an appropriate ASIC depth is required to minimize the SI effects, which can be obtained by solving the optimization problem $\max_{\eta_{\text{alg}}} P_{\text{rsi}}$ via the "*CVX*" toolbox on *MATLAB*. Regardless of the performance degradation, a deep RF cancellation is not desired from the view of implementation cost. Thus, it is necessary to find the optimal ASIC depth to determine the number of taps in the canceller.

2.6 SIC in MIMO systems

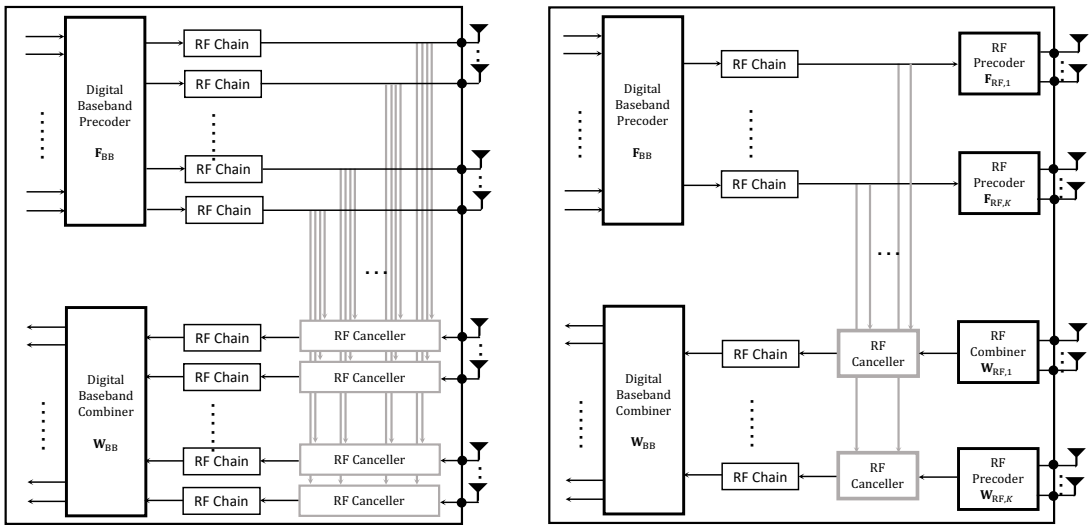
MIMO systems have been employed in current wireless networks to enhance the system capacity. By exploring the spatial DoFs provided by antenna arrays, MIMO systems can improve the throughput, increase the number of connected users, etc. In addition, it is regarded as the key technology to enable frequency range 2 (FR2) spectrum ($\geq$ 24.250 GHz), i.e., millimeter wave (mmWave), which can provide abundant spectrum resources and will be utilized in beyond 5G and 6G networks [33, 36, 37]. Communications in the FR2 band suffer from weak diffraction ability and susceptibility to blockages due to the short carrier wavelength, resulting in high path loss of communication links. To provide reliable mmWave links, large-scale antenna arrays (i.e., massive MIMO) are explored to form highly-directional beams. Thus, it is necessary to consider SIC in MIMO and massive MIMO systems.

2.6.1 Direct extension of existing methods

The existing SIC schemes can be easily extended into MIMO systems by placing a canceller between each Tx-Rx antenna pair to represent associated SI coupling[7]. However, this results in very high complexity in terms of both computation and implementation. For an IBFD transceiver with N_{ant} Tx antennas and M_{ant} Rx antennas, a total of

[7] Although beamforming could utilize spatial DoFs to null the SI, it affects the far-field beam pattern, so it may compromise the quality of intended channels, which will be detailed in the later Section 4.4. In contrast, the 3-step SIC scheme does not affect the far-field beam pattern, so it is worth studying this scheme.

$N_{\text{ant}}M_{\text{ant}}$ cancellers are required, as Figure 2.9(a) shows. Such a large number of cancellers introduce challenges in fabrication processes to enable packageable and low-power canceller manufacturing. Digital cancellation is free of manufacturing concerns but has extremely high computation overheads. The parameters that digital cancellers need to calculate increase almost by the cube of the number of Rx antennas M_{ant} with a nonlinearity order of 3 [35]. These challenges make directly extending SIC schemes in massive MIMO systems practically prohibited. For instance, with 128×128 Tx/Rx antenna arrays, $128 \times 128 \approx 1.64 \times 10^4$ RF canceller circuits are required, and $128^3 \approx 2.10 \times 10^6$ parameters need to be estimated by digital cancellers, which look like an unattainable number.



(a) MIMO transceiver without RF beamforming (b) MIMO transceiver with RF beamforming

FIGURE 2.9: The IBFD transceivers with large-scale antenna array (i.e., massive MIMO).

2.6.2 With hybrid beamforming architecture

Recent studies explore a hybrid RF-digital beamforming architecture to reduce the cost of massive MIMO transceivers [36, 37]. This architecture utilizes analog beamforming to reduce the number of transmitter RF chains. Taking advantage of the reduced RF chains, RF canceller circuits can be inserted between each transmitter and receiver RF chain pair. By doing so, the number of canceller circuits can be significantly reduced, e.g., from 128×128 to 4×4, making it feasible with large-scale antenna arrays. In this

section, we will study the effects of such hybrid beamforming architecture on the RF canceller circuit design and cancellation performance.

2.6.2.1 RF beamforming matrix properties

Figure 2.9(b) shows the transceiver architecture applying hybrid beamforming, where N_{ant} Tx antennas and M_{ant} Rx antennas are equally divided into N_{sub} and M_{sub} subarrays, respectively. There are N_{RF} and M_{RF} RF chains at the transmitter and receiver, and $N_{\text{cR},i}$ and $M_{\text{cR},j}$ RF chains are connected to the i^{th} and j^{th} subarray of the transmitter and receiver, respectively. Considering the common OFDM system, the received SI at the k^{th} subcarrier can be described as:

$$\mathbf{y}[k] = \mathbf{W}_{\text{RF}}^{\dagger}\mathbf{G}_{\text{si}}[k]\left(\xi\mathbf{F}_{\text{RF}}\mathbf{s}[k] + \mathbf{d}_{\text{tx}}[k]\right) + \mathbf{n}[k] + \mathbf{d}_{\text{rx}}[k], \tag{2.32}$$

where $\mathbb{E}\left\{\mathbf{s}[k]\mathbf{s}^{\dagger}[k]\right\} = \frac{P_k}{N_{\text{RF}}}\mathbf{I}_{N_{\text{RF}}}$ with P_k denoting the transmit power budget of the k^{th} subcarrier; $\mathbf{G}_{\text{si}}[k] \in \mathbb{C}^{M_{\text{ant}}\times N_{\text{ant}}}$ denotes the wireless SI channel; $\mathbf{F}_{\text{RF}}$ and $\mathbf{W}_{\text{RF}}$ are RF beamforming matrices; $\mathbf{d}_{\text{tx}}[k]$ and $\mathbf{d}_{\text{rx}}[k]$ represent the transmitter and receiver distortions, respectively; ξ is the scaling factor (i.e., RF chain gain) to satisfy the power constraints; and $\mathbf{n}[k]$ is the AWGN given as

$$\mathbf{n}[k] \sim \mathcal{CN}\left(\mathbf{0}, \sigma_n^2 \cdot \mathbf{I}_{M_{\text{RF}}}\right). \tag{2.33}$$

The digital beamforming effects have been included in $\mathbf{s}[k]$. We have $N_s = N_{\text{RF}} \ll N_{\text{ant}}$ and $M_s = M_{\text{RF}} \ll M_{\text{ant}}$ to allow the fully-digital beamforming being decomposed into the digital beamforming followed by RF beamforming without penalty [38]. The RF beamforming is implemented via phase shifters in practice. Therefore, only the phase of the signals is adjusted, while the amplitude is fixed. It also suggests that the RF beamforming is independent of subcarriers (i.e., it is frequency-flat). The RF precoder matrix is given as

$$\mathbf{F}_{\text{RF}} = \mathcal{D}\left(\mathbf{f}_{\text{RF},1}, \mathbf{f}_{\text{RF},2}, \ldots, \mathbf{f}_{\text{RF},N_{\text{sub}}}\right), \tag{2.34}$$

where $\mathbf{f}_{\text{RF},i} \in \mathbb{C}^{N_{\text{aS}}\times N_{\text{cR},i}}$ represents the beamformer for the i^{th} subarray, which can be written as

$$\mathbf{f}_{\mathrm{RF},i} = \begin{bmatrix} e^{j\theta_{i,11}} & e^{j\theta_{i,12}} & \dots & e^{j\theta_{i,1N_{\mathrm{cR},i}}} \\ e^{j\theta_{i,21}} & e^{j\theta_{i,22}} & \dots & e^{j\theta_{i,2N_{\mathrm{cR},i}}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\theta_{i,N_{\mathrm{aS}}1}} & e^{j\theta_{i,N_{\mathrm{aS}}2}} & \dots & e^{j\theta_{i,N_{\mathrm{aS}}N_{\mathrm{cR},i}}} \end{bmatrix}, \tag{2.35}$$

where $N_{\mathrm{aS}} = \frac{N_{\mathrm{ant}}}{N_{\mathrm{sub}}}$ is the number of antennas per subarray, and θ is the phase-shifting induced by corresponding phase shifters, which is subjected to uniform distribution within 0 to 2π, such that $\theta \sim U(0, 2\pi)$. Assume $f_m = e^{j\theta_m}$ and $f_n = e^{j\theta_n}$ are two non-zero elements of the precoder matrix $\mathbf{F}_{\mathrm{RF}}$, we have

$$\mathbb{E}\left\{f_m \bar{f}_n\right\} = \mathbb{E}\left\{e^{j(\theta_m - \theta_n)}\right\} = \begin{cases} 1, & \text{if } m = n \\ \mathbb{E}\left\{e^{jx}\right\}, & \text{otherwise} \end{cases} \tag{2.36}$$

where $x = \theta_m - \theta_n$ that has the possibility density function (PDF) as

$$f(x) = \begin{cases} \frac{1}{4\pi^2}x + \frac{1}{2\pi}, & \text{if } -2\pi \le x < 0 \\ -\frac{1}{4\pi^2}x + \frac{1}{2\pi}, & \text{if } 0 \le x \le 2\pi \end{cases} \tag{2.37}$$

Thus, we will have

$$\begin{aligned} \mathbb{E}\left\{e^{j(\theta_m - \theta_n)}\right\} &= \int_{-2\pi}^{2\pi} e^{jx} f(x) dx \\ &= \int_{-2\pi}^{0} e^{jx}\left(\frac{1}{4\pi^2}x + \frac{1}{2\pi}\right) dx + \int_{0}^{2\pi} e^{jx}\left(-\frac{1}{4\pi^2}x + \frac{1}{2\pi}\right) dx \\ &= \left[\frac{x}{j}e^{jx} - e^{jx} + \frac{1}{2\pi j}e^{jx}\right]_{-2\pi}^{0} + \left[-\frac{x}{j}e^{jx} + e^{jx} + \frac{1}{2\pi j}e^{jx}\right]_{0}^{2\pi} \\ &= 0. \end{aligned} \tag{2.38}$$

Therefore, we can rewrite Equation (2.36) as

$$\mathbb{E}\left\{f_m \bar{f}_n\right\} = \begin{cases} 1, & \text{if } m = n \\ 0, & \text{otherwise} \end{cases} \tag{2.39}$$

Based on this, the covariance matrix of the RF precoder can be derived as

$$\mathbb{E}\left\{\mathbf{F}_{\mathrm{RF}}\left(\mathbf{F}_{\mathrm{RF}}\right)^{\dagger}\right\} = \mathcal{D}\left(N_{\mathrm{cR},1}\mathbf{I}_{N_{\mathrm{aS}}}, N_{\mathrm{cR},2}\mathbf{I}_{N_{\mathrm{aS}}}, \dots, N_{\mathrm{cR},N_{\mathrm{sub}}}\mathbf{I}_{N_{\mathrm{aS}}}\right), \tag{2.40}$$

which yields

$$tr\left(\mathbb{E}\left\{\mathbf{F}_{\mathrm{RF}}\left(\mathbf{F}_{\mathrm{RF}}\right)^{\dagger}\right\}\right)=\frac{N_{\mathrm{ant}}N_{\mathrm{RF}}}{N_{\mathrm{sub}}}. \tag{2.41}$$

Similarly, the RF combiner is given as

$$\mathbf{W}_{\mathrm{RF}}=\mathcal{D}\left(\mathbf{w}_{\mathrm{RF},1},\mathbf{w}_{\mathrm{RF},2},\ldots,\mathbf{w}_{\mathrm{RF},M_{\mathrm{sub}}}\right), \tag{2.42}$$

where $\mathbf{w}_{\mathrm{RF},i}\in\mathbb{C}^{M_{\mathrm{aS}}\times M_{\mathrm{cR},i}}$ is the RF combiner for the i^{th} subarray with $M_{\mathrm{aS}}=\frac{M_{\mathrm{ant}}}{M_{\mathrm{sub}}}$ denoting the number of Rx antennas per subarray. Following a similar derivation, we will have

$$\mathbb{E}\left\{\left(\mathbf{W}_{\mathrm{RF}}\right)^{\dagger}\mathbf{W}_{\mathrm{RF}}\right\}=\frac{M_{\mathrm{ant}}}{M_{\mathrm{sub}}}\mathbf{I}_{M_{\mathrm{RF}}}, \tag{2.43}$$

$$tr\left(\mathbb{E}\left\{\left(\mathbf{W}_{\mathrm{RF}}\right)^{\dagger}\mathbf{W}_{\mathrm{RF}}\right\}\right)=\frac{M_{\mathrm{ant}}M_{\mathrm{RF}}}{M_{\mathrm{sub}}}. \tag{2.44}$$

2.6.2.2 RF cancellation

We consider the common TDC design. As aforementioned, we would like to tap the reference signal from the transmitter RF chain and insert the generated cancellation signal back into the receiver RF chain. Assume the output RF signal of the u^{th} transmitter RF chain is $x_u(t)$, the v^{th} receiver RF chain receives signal caused by $x_u(t)$ can be represented as

$$\begin{aligned}y_v(t)&=\sum_{m=1}^{M_{\mathrm{ant},v}}e^{j\theta_m(t)}\sum_{n=1}^{N_{\mathrm{ant},u}}g_{\mathrm{si},m,n}(t)*\left(e^{j\theta_n(t)}x_u(t)\right)\\&=\tilde{h}_{vu}(t)*x_u(t),\end{aligned} \tag{2.45}$$

where $M_{\mathrm{ant},u}$ and $N_{\mathrm{ant},v}$ denote the number of Tx and Rx antennas connected to the u^{th} transmitter RF chain and v^{th} receiver RF chain; $\theta_m(t)$ and $\theta_n(t)$ are the phase-shifting induced by the RF combiner and precoder at the m^{th} Rx and n^{th} Tx antennas, respectively; $g_{\mathrm{si},m,n}(t)$ is the wireless SI channel that can be modeled by the TDL channel model in Equation (2.1)[8]; and $\tilde{g}_{vu}(t)$ is the effective SI channel the canceller needs to mimic, which can be written as

$$\tilde{g}_{vu}(t)=\sum_{m=1}^{M_{\mathrm{ant},v}}\sum_{n=1}^{N_{\mathrm{ant},u}}e^{j\theta_m(t)}e^{j\theta_n(t)}g_{\mathrm{si},m,n}(t)$$

[8]Studies on FR2 communications usually adopt the 3D channel model to describe the wireless channels, which is identical to the TDL model in nature, as detailed in Appendix 2.10.

$$= \sum_{m=1}^{M_{\text{ant},v}} \sum_{n=1}^{N_{\text{ant},u}} \sum_{l=0}^{L_{mn}} h_{mn,l}\delta(t-\tau_{mn,l})e^{j(\theta_m(t)+\theta_n(t))}, \tag{2.46}$$

where $h_{mn,l}$ and $\tau_{mn,l}$ denote the channel coefficient and delay of the l^{th} path between the m^{th} Rx antenna and the n^{th} Tx antenna, respectively. For a specific operation bandwidth and target cancellation depth, the required number of taps of the multi-tap canceller is related to the frequency-selectivity of the channel it mimics. It is demonstrated that the required number of taps is almost proportional to the root-mean-square (RMS) delay spread of the effective SI channel [10], which is given as

$$\tau_{\text{RMS}}^{(v,u)} = \sqrt{\frac{\sum_{n=1}^{M_{\text{ant},v}} \sum_{m=1}^{N_{\text{ant},u}} \sum_{l=0}^{L_{mn}} \chi_{mn,l}\tau_{mn,l}^2}{\sum_{n=1}^{M_{\text{ant},v}} \sum_{m=1}^{N_{\text{ant},u}} \sum_{l=0}^{L_{mn}} \chi_{mn,l}} - \tau_0^2}, \tag{2.47}$$

where $\chi_{mn,l} = \left|h_{mn,l}e^{j(\theta_m(t)+\theta_n(t))}\right|^2 = |h_{mn,l}|^2$ captures the power profile of associated rays and $\tau_0^{(v,u)}$ represents the mean delay of the effective SI channel that can be described as

$$\tau_0^{(v,u)} = \frac{\sum_{n=1}^{M_{\text{ant},v}} \sum_{m=1}^{N_{\text{ant},u}} \sum_{l=0}^{L_{mn}} \chi_{mn,l}\tau_{mn,l}}{\sum_{n=1}^{M_{\text{ant},v}} \sum_{m=1}^{N_{\text{ant},u}} \sum_{l=0}^{L_{mn}} \chi_{mn,l}}. \tag{2.48}$$

Let $\rho_t\left(B_W, \bar{\eta}_{\text{RF}}\right)$ describe the portion relationship between the RMS delay spread and the required number of taps, which is a function of operational bandwidth B_W and desired cancellation depth $\bar{\eta}_{\text{RF}}$, the total required number of taps is given as

$$\Sigma_{\text{hyb}} = \sum_{v=1}^{M_{\text{RF}}} \sum_{u=1}^{N_{\text{RF}}} \left\lceil \rho_t(B_W, \bar{\eta}_{\text{RF}}) \cdot \tau_{\text{RMS}}^{(v,u)} \right\rceil \tag{2.49}$$

Remarking: we also derive the required number of taps for conventional architecture (i.e., without RF beamforming, as Figure 2.9(a) shows). The RMS delay spread of the wireless SI channel between the r^{th} Rx antenna and s^{th} Tx antenna is given as

$$\bar{\tau}_{\text{RMS}}^{(r,s)} = \sqrt{\frac{\sum_{l=0}^{L_{rs}} h_{rs,l}\tau_{rs,l}^2}{\sum_{l=0}^{L_{rs}} h_{rs,i}} - \left(\frac{\sum_{l=0}^{L_{rs}} h_{rs,l}\tau_{rs,l}}{\sum_{l=0}^{L_{rs}} h_{rs,l}}\right)^2}. \tag{2.50}$$

The total required number of taps is given as

$$\Sigma_{\text{con}} = \sum_{r=1}^{M_{\text{ant}}} \sum_{s=1}^{N_{\text{ant}}} \left\lceil \rho_t(B_W, \bar{\eta}_{\text{RF}}) \cdot \bar{\tau}_{\text{RMS}}^{(r,s)} \right\rceil. \tag{2.51}$$

Usually, $M_{ant} \gg M_{\mathrm{RF}}$ and $N_{\mathrm{ant}} \gg N_{\mathrm{RF}}$, while $\tau_{\mathrm{RMS}}^{(v,u)}$ and $\bar{\tau}_{\mathrm{RMS}}^{(r,s)}$ are close ($\tau_{\mathrm{RMS}}^{(v,u)} \approx \bar{\tau}_{\mathrm{RMS}}^{(r,s)}$) if the wireless SI channels have similar RMS delay, spreads due to the proximity of antennas. Thus, the hybrid architecture can significantly reduce the total number of taps for RF cancellation. The RF cancellation performance can be described by the interference suppression ratio, i.e., cancellation depth, which is given as

$$\eta^{\mathrm{RF}} = \mathbb{E}_k \left\{ \frac{\left(\xi \mathbf{F}_{\mathrm{RF}} \mathbf{s}[k] + \mathbf{d}_{\mathrm{tx}}[k]\right)^{\dagger} \mathbf{G}_{\mathrm{si}}^{\dagger}[k] \mathbf{G}_{\mathrm{si}}[k] \left(\xi \mathbf{F}_{\mathrm{RF}} \mathbf{s}[k] + \mathbf{d}_{\mathrm{tx}}[k]\right)}{\left(\xi \mathbf{F}_{\mathrm{RF}} \mathbf{s}[k] + \mathbf{d}_{\mathrm{tx}}[k]\right)^{\dagger} \mathbf{G}_{\mathrm{rsi}}^{\dagger}[k] \mathbf{G}_{\mathrm{rsi}}[k] \left(\xi \mathbf{F}_{\mathrm{RF}} \mathbf{s}[k] + \mathbf{d}_{\mathrm{tx}}[k]\right)} \right\}, \tag{2.52}$$

where $\mathbf{G}_{\mathrm{rsi}}[k] = \mathbf{G}_{\mathrm{si}}[k] - \mathbf{G}_{\mathrm{rfc}}[k]$ denotes the RSI channel after the RF cancellation with $\mathbf{G}_{\mathrm{rfc}}[k]$ representing the filter coefficient at the k^{th} subcarrier formed by RF cancellers. Thus, the RSI and SI channel coefficients can be related as

$$\mathbb{E}_k \left\{ \mathbf{G}_{\mathrm{rsi}}[k] \mathbf{G}_{\mathrm{rsi}}^{\dagger}[k] \right\} = \frac{1}{\eta^{\mathrm{RF}}} \mathbb{E}_k \left\{ \mathbf{G}_{\mathrm{si}}[k] \mathbf{G}_{\mathrm{si}}^{\dagger}[k] \right\}. \tag{2.53}$$

2.6.2.3 Digital cancellation

We consider a linear frequency-domain digital canceller here. At the receiver, the RSI suppressed within the receiver's dynamic range is digitized with trivial quantization noise. The frequency-domain digital RSI signal at the pilot subcarrier k_p is given as

$$\begin{aligned} \bar{\mathbf{y}}[k_p] &= \mathbf{W}_{\mathrm{RF}}^{\dagger} \left(\mathbf{G}_{\mathrm{si}}[k_p] + \mathbf{G}_{\mathrm{rfc}}[k_p]\right) \left(\xi \mathbf{F}_{\mathrm{RF}} \mathbf{s}[k_p] + \mathbf{d}_{\mathrm{tx}}[k_p]\right) + \mathbf{d}_{\mathrm{rfc}}[k_p] + \mathbf{d}_{\mathrm{rx}}[k_p] + \mathbf{n}[k_p] \\ &= \mathbf{G}_{\mathrm{eff}}[k_p] \mathbf{s}[k_p] + \tilde{\mathbf{z}}[k_p], \end{aligned} \tag{2.54}$$

where $\mathbf{G}_{\mathrm{eff}}[k_p] = \xi \mathbf{W}_{\mathrm{RF}}^{\dagger} \mathbf{G}_{\mathrm{rsi}}[k_p] \mathbf{F}_{\mathrm{RF}}$ denotes the effective SI channel that includes the effects of RF beamforming, wireless SI channel, and RF cancellers, $\mathbf{d}_{\mathrm{rfc}}[k_p]$ represents the noise and distortions caused by cancellers that can be described as $\mathbf{d}_{\mathrm{rfc}}[k_p] \sim \mathcal{CN}\left(\mathbf{0}, \sigma_c^2 \cdot \mathbf{I}\right)$, and $\tilde{\mathbf{z}}[k_p] = \mathbf{W}_{\mathrm{RF}}^{\dagger} \mathbf{G}_{\mathrm{rsi}}[k_p] \mathbf{d}_{\mathrm{tx}}[k_p] + \mathbf{d}_{\mathrm{rfc}}[k_p] + \mathbf{d}_{\mathrm{rx}}[k_p] + \mathbf{n}[k_p]$ denotes all the transceiver and canceller noise and distortions. Interference from other nodes is not present here since an interference-free period is usually provided by the MAC protocol via carrier sense to achieve accurate SI channel estimation. Adopting the MMSE channel estimator, the estimated coefficients at the pilot subcarrier k_p are given as

$$\hat{\mathbf{G}}_{\mathrm{eff}}[k_p] = \tilde{\mathbf{y}}[k_p] \left(\mathbf{s}^{\dagger}[k_p] \mathbf{s}[k_p]\right)^{-1} \mathbf{s}^{\dagger}[k_p] = \mathbf{G}_{\mathrm{eff}}[k] - \mathbf{\Delta}_{\mathrm{ch}}[k], \tag{2.55}$$

where $\mathbf{\Delta}_{\text{ch}}[k]$ is the error due to imperfect channel estimation[9]. With known $\mathbf{s}[k]$, the digital baseband cancellation signal is reconstructed as

$$\hat{\mathbf{y}}[k] = \hat{\mathbf{G}}_{\text{eff}}[k]\mathbf{s}[k] = \mathbf{G}_{\text{eff}}[k]\mathbf{s}[k] - \mathbf{\Delta}_{\text{ch}}[k]\mathbf{s}[k]. \tag{2.56}$$

Then, the RSI can be subtracted from the received signal, yielding the overall receiver noise as

$$\mathbf{\Psi}[k] = \tilde{\mathbf{y}}[k] - \hat{\mathbf{y}}[k] = \mathbf{\Delta}_{\text{ch}}[k]\mathbf{s}[k] + \tilde{\mathbf{z}}[k]. \tag{2.57}$$

The variance matrix of $\mathbf{\Psi}[k]$ is given as

$$\mathbb{E}\left\{\mathbf{\Psi}[k]\mathbf{\Psi}^{\dagger}[k]\right\} = \left(1 + \frac{1}{N_{\text{RF}}}\right)\left(\frac{(\kappa+\beta)\varrho_{\text{si}}P_k M_{\text{ant}}}{(1+\kappa)\eta_{\text{RF}}M_{\text{sub}}} + \sigma_c^2 + \sigma_n^2\right)\mathbf{I}_{M_{\text{RF}}}, \tag{2.58}$$

which is detailed in Appendix 2.11 with ϱ_{si} denoting the SI channel strength (i.e., the variance of the elements of matrix $\mathbf{G}_{\text{si}}$). The expression indicates the hybrid architecture, i.e., the number of transmitter RF chains and the number of subarrays at the receiver, will affect the digital cancellation performance. Besides, too deep RF cancellation (i.e., ϑ^{RF} is large) will also degrade the digital cancellation performance. The degradation comes from the fact that the RF cancellation decreases the SNR of the SI signal for channel estimation in digital cancellation. This linear canceller cannot process with the transceiver nonlinearity, so its performance may be compromised by the increasing distortion factor κ and β. However, our proposed linear frequency-domain canceller is easy to implement in OFDM systems and has low computational complexity, reducing processing time to reduce the end-to-end latency.

2.7 User allocation for CCI reducation

In addition to the SI, IBFD operation introduces CCI as highlighted earlier in Section 2.1. The CCI usually comes from a relatively far transmitter, so it is usually within the receiver's dynamic range due to the propagation loss. Thus, it can be mitigated by digital processing, which is much easier than SIC. However, to minimize the latency, it is desired to combat the CCI from the physical layer instead of digital processing. Thus, we study a user allocation method to minimize the CCI effects.

[9]The channel coefficients at the remaining subcarriers are calculated by interpolation on the estimated channel coefficients at pilot subcarriers. We ignore the interpolation errors here for simplicity.

It is worth noting that IBFD operation and the abundant spectrum resources in FR2 are promising to provide ultra-reliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC) services simultaneously, which is attractive for the industrial Internet of Things (IIoT). IBFD radios allow uplink eMBB devices and downlink URLLC devices to use any time slot and subcarrier. Thus, it is not necessary to reserve time slots for URLLC devices as in TDD systems, reducing the latency. The micro-cell nature of mmWave communications yields a relatively low channel delay due to short propagation distance, so it can support higher numerology (e.g., 30 kHz, 60 kHz, 120 kHz, and 240 kHz of subcarrier spacing) and reduce the symbol duration [39]. For instance, expanding the subcarrier spacing from 15 kHz to 60 kHz can reduce the symbol duration time from 72 μs to 18 μs. Thus, we consider an IIoT scenario and propose the user allocation algorithm accordingly, while it can be generalized by modifying the parameters to fit any other wireless network.

2.7.1 Game theoretic algorithm

In IIoT networks, IBFD base stations (e.g., gNB: the 5G new radio (NR) term of the base station) need to provide service to uses with various configurations and requirements, e.g., different power budgets, operation bandwidth, etc. Figure 2.10 depicts a typical IIoT scenario, where an IBFD-enabled access point (AP) services 4 uplink eMBB devices (closed-circuit television (CCTV) cameras) and 8 downlink URLLC devices (sensors and robotics arms). In order to cope with such a heterogeneous environment, we propose a user allocation algorithm through a game-theoretic approach. Assume the whole communication bandwidth is separated into T_{DL} and T_{UL} orthogonal sub-bands for DL and UL users, as Figure 2.11 shows. The sub-bands can have different bandwidths to provide different performance profiles for a variety of user needs in IIoT networks. Let $\mathcal{X} = \{\chi^{\mathrm{DL}}, \chi^{\mathrm{UL}}\}$ denote user allocation policy which is the collections of all DL and UL indicators. χ^{DL} and χ^{UL} have dimension of $T_{\mathrm{DL}} \times K_{\mathrm{DL}}$ and $T_{\mathrm{UL}} \times K_{\mathrm{UL}}$ respectively, where K_{DL} and K_{UL} represent the number of DL and UL users. The allocation matrices are given as

$$[\chi^{\mathrm{DL}}]_{t_d,k_d} = \begin{cases} 1, & \text{if DL user } k_d \text{ is allocated into the } t_d^{th} \text{ DL sub-band } \mathrm{sub}_t^{\mathrm{DL}} \\ 0, & \text{otherwise} \end{cases} \tag{2.59}$$

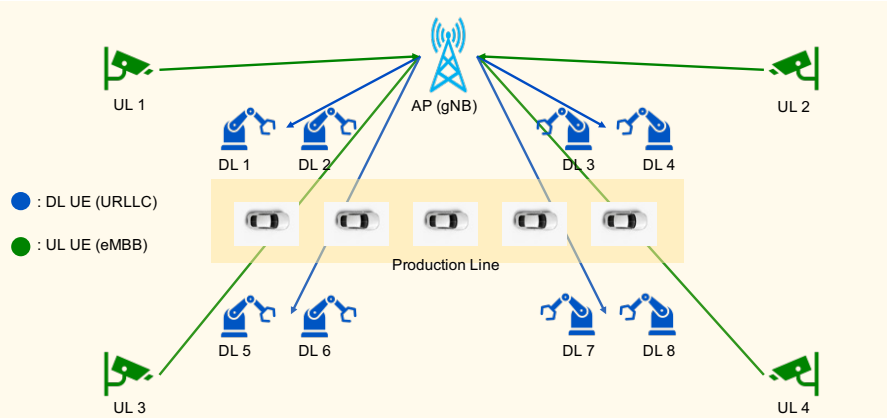


FIGURE 2.10: A typical IIoT scenario with 4 UL eMBB devices and 8 DL URLLC devices served by a 5G NR AP.

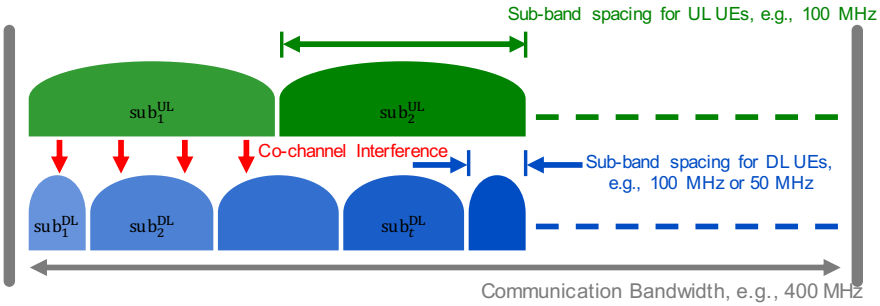


FIGURE 2.11: Orthogonal sub-bands division for UL and DL UEs.

$$[\chi^{\text{UL}}]_{t_u,k_u} = \begin{cases} 1, & \text{if UL user } k_u \text{ is allocated into the } t_u^{th} \text{ UL sub-band } \text{sub}_t^{\text{UL}} \\ 0, & \text{otherwise} \end{cases} \tag{2.60}$$

There will be CCI as long as the DL sub-bands and UL sub-bands overlap. The essence of user allocation is to allocate the DL and UL user pairs that may cause significant CCI to orthogonal sub-bands while leaving the user pairs with small CCI in the overlapping sub-bands. The achievable sum rate of DL user i and UL user j served by BS g can be described as

$$R_i^{\text{DL}} = \mathcal{G}_i^{\text{DL}} \cdot \log_2 \left(1 + \frac{M_{\text{ant},i} P_i^{\text{DL}} |\varrho_{i,g}^{\text{DL}}|^2}{\Sigma_i^{\text{DL}}} \right), \tag{2.61}$$

$$R_j^{\text{UL}} = \mathcal{G}_j^{\text{UL}} \cdot \log_2 \left(1 + \frac{N_{\text{ant},j} P_j^{\text{UL}} |\varrho_{g,j}^{\text{UL}}|^2}{\Sigma_j^{\text{UL}}} \right), \tag{2.62}$$

where $\mathcal{G}_i^{\text{DL}} = \min\{M_{\text{ant},i}, N_{\text{ant},g}\}$ and $\mathcal{G}_j^{\text{UL}} = \min\{N_{\text{ant},g}, M_{\text{ant},g}\}$ are the degrees of freedom gains for the sum rate; $\varrho_{i,g}^{\text{DL}}$ and $\varrho_{g,j}^{\text{UL}}$ represent the path losses form BS g to DL user i and form UL user j to BS g, respectively; P_i^{DL} and P_j^{UL} denote the transmit power of corresponding users; Σ_i^{DL} and Σ_j^{UL} represent the power of interference plus

noise for the i^{th} DL user and j^{th} UL user given as

$$\Sigma_i^{\mathrm{DL}} = M_{\mathrm{ant},i} \sum_{b \neq g}^{K_{\mathrm{BS}}} \zeta_i^{\mathrm{DL}} P_b^{\mathrm{BS}} |\varrho_{i,b}^{\mathrm{DL}}|^2 + M_{\mathrm{ant},i} \sum_{j_i=1}^{K_{j_i}^{\mathrm{ol}}} \zeta_{i,j_i}^{\mathrm{ol}} P_{j_i}^{\mathrm{UL}} |\varrho_{i,j_i}^{\mathrm{DU}}|^2 + P_{n,i}^{\mathrm{DL}}, \tag{2.63}$$

$$\Sigma_j^{\mathrm{UL}} = \zeta_j^{\mathrm{UL}} \left(\sum_{b \neq g}^{K^{\mathrm{BS}}} N_{\mathrm{ant},b} P_b^{\mathrm{BS}} |l_{g,b}^{\mathrm{BS}}|^2 + P_{\mathrm{rsi},g} + P_{\mathrm{n},g}^{\mathrm{BS}} \right), \tag{2.64}$$

where $\varrho_{i,j}^{\mathrm{DU}}$, and $\varrho_{g,b}^{\mathrm{BS}}$ represent the path losses from UL user j to DL user i and from BS b to BS g, respectively; P_b^{BS} represents the transmit power of BS b; $P_{\mathrm{rsi},g}$ is the RSI power after SIC at BS g to capture the SIC effects; $P_{n,i}^{\mathrm{DL}}$ and $P_{\mathrm{n},g}^{\mathrm{BS}}$ are AWGN power at DL user i and BS g; $K_{j_i}^{\mathrm{ol}}$ is the number of UL users which have overlapping bandwidth with DL user i; ζ_i^{DL} and ζ_j^{UL} capture the ratio of the sub-band of corresponding users to the whole bandwidth; $\zeta_{i,j_i}^{\mathrm{ol}}$ is the ratio of the overlapping portion of sub-bands of DL user i and UL user j_i to the sub-band of DL user i. The optimal user allocation policy is achieved if the sum rate of the network is maximized, which can be cast as

$$\max_{\mathcal{X}} \sum_{i=1}^{K_{\mathrm{DL}}} R_i^{\mathrm{DL}} + \sum_{j=1}^{K_{\mathrm{UL}}} R_j^{\mathrm{UL}}, \tag{2.65}$$

$$s.t. \sum_{t=1}^{T_{\mathrm{DL}}} [\chi^{\mathrm{DL}}]_{t,i} = 1, \ \forall \ i \in [1, K_{\mathrm{DL}}] \tag{2.66}$$

$$\sum_{t=1}^{T_{\mathrm{UL}}} [\chi^{\mathrm{UL}}]_{t,j} = 1, \ \forall \ j \in [1, K_{\mathrm{UL}}] \tag{2.67}$$

where the two constraints force each user to be allocated to only one sub-band. This problem can be solved by a game theoretic approach, where the sum utility of all users in this network is given as

$$U(\mathcal{X}) = \sum_{i=1}^{K_{\mathrm{DL}}} R_i^{\mathrm{DL}} + \sum_{j=1}^{K_{\mathrm{UL}}} R_j^{\mathrm{UL}}. \tag{2.68}$$

The user allocation game is a characteristic formation game with non-transferable utility, which can be solved based on a preference relation as

$$\chi \prec \chi'|_{t \leftarrow i} \Leftrightarrow U(\mathcal{X}) < U(\mathcal{X}'), \tag{2.69}$$

Algorithm 1 User allocation algorithm

1: Randomly allocate the UL users to associated sub-bands, e.g., $[\chi^{\mathrm{UL}}]_{t,j} = 1,\ \forall\, t = j$ and $[\chi^{\mathrm{UL}}]_{t,j} = 0,\ \forall\, t \neq j$, where $j \in [1, K_{\mathrm{UL}}]$.
2: Allocate all DL users to into a single sub-band, e.g., $[\chi^{\mathrm{DL}}]_{t,i} = 1,\ \forall\ t = 1$, and $[\chi^{\mathrm{DL}}]_{t,i} = 0,\ \forall\, t \neq 1$, where $i \in [1, K_{\mathrm{DL}}]$.
3: Record current user allocation policy as $\mathcal{X}$
4: **for all** DL user $i \in [1, K_{\mathrm{DL}}]$ **do**
5: DL user i moves into all other sub-band except its current sub-band.
6: Record the new allocation policy after the UE moves as $\mathcal{X}'$.
7: Calculate the sum utilities under the two policy $U(\mathcal{X})$ and $U(\mathcal{X}')$.
8: Compare the preference of the i^{th} DL user as Equation (2.69).
9: **if** the preference relation is satisfied **then**
10: Update the current policy as $\mathcal{X} \leftarrow \mathcal{X}'$.
11: **else**
12: Go back to policy $\mathcal{X}$.
13: **end if**
14: **end for**

where $\chi \prec \chi'|_{t \leftarrow i}$ means that user i is preferred to be allocated into sub-band t instead of its current sub-band. The moving operation will be done if the sum utility increases after user i moves to sub-band t. We consider each sub-band for UL users will only be occupied by a single user to maximize its capacity since UL eMBB devices require high data throughput. Hence, the UL users can be randomly allocated to these sub-bands at first. Then, each DL user compares the preference with being allocated into all other sub-bands and executes the moving operation if the condition in Equation (2.69) is satisfied. It should be noted that the indicator of sub-band only determines the order of these sub-bands, while its bandwidth varies with associated users. Performing the compare-and-moving operation for all DL users, the optimal user allocation policy $\bar{\mathcal{S}}$ is obtained. The user allocation algorithm is summarized as Algorithm 1.

2.7.2 Complexity analysis

The algorithm is based on the compare-and-moving operations of each DL user. There is one time of computations when each DL user compares the sum utility and determines whether the user is moving. In order to obtain the optimal policy, each of all K_{DL} DL users should move to all other $T_{\mathrm{DL}} - 1$ sub-bands and executes the compare-and-moving operation, which results in a total of $K_{\mathrm{DL}}(T_{\mathrm{DL}} - 1)$ computation times in one cycle. Given a number of cycle times C, the computational complexity of the proposed algorithm is $\mathcal{O}(CK_{\mathrm{DL}}(T_{\mathrm{DL}} - 1))$.

2.7.3 Convergence behavior

Let $\mathcal{X}_0$ and $\mathcal{X}_f$ denote the initial and final user allocation policy, respectively. During the game, the policy is changed as follows

$$\mathcal{X}_0 \to \mathcal{X}_1 \to \mathcal{X}_2 \to \cdots \to \mathcal{X}_f. \tag{2.70}$$

Algorithm 1 indicates that the user allocation policy $\mathcal{X}$ will only be changed if the sum utility increases, which suggests that the sum utility is strictly increased with the policy changes in sequence (2.70) such that

$$U(\mathcal{X}_0) < U(\mathcal{X}_1) < U(\mathcal{X}_2) < \cdots < U(\mathcal{X}_f). \tag{2.71}$$

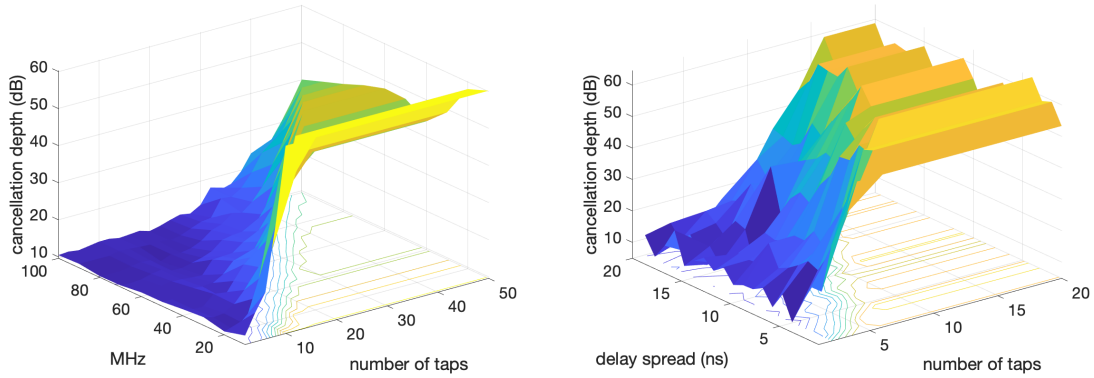
Since the number of all possible user allocation policies is finite due to a finite number of sub-bands and users, the policy in sequence (2.70) is guaranteed to converge to the local optimal policy.

2.8 Simulation results

In this section, we will show some simulation results of interference suppression in IBFD systems. The SI channel is generated using the "TDL-D" channel model given in the 3GPP document [3], and the RMS delay spread is set as 20 ns. Other settings are given in corresponding sections.

2.8.1 RF cancellation

The flexibility and cancellation capacity of the TDC strongly depend on the available number of taps, which determines the available DoFs for solving the minimization problem given in Equation (2.20). With an enlarging operation bandwidth in future wireless networks, more taps are required to emulate the SI channel at more frequency points. Figure 2.12(a) shows the achievable cancellation depth of the TDC with increasing operation bandwidths and numbers of taps. It verifies the analysis that a large number of taps is essential to achieve sufficient cancellation depth with increasing bandwidth. The



(a) Enlarging bandwidth (20 ns of delay spread) (b) Increasing delay spread (20 MHz of bandwidth)

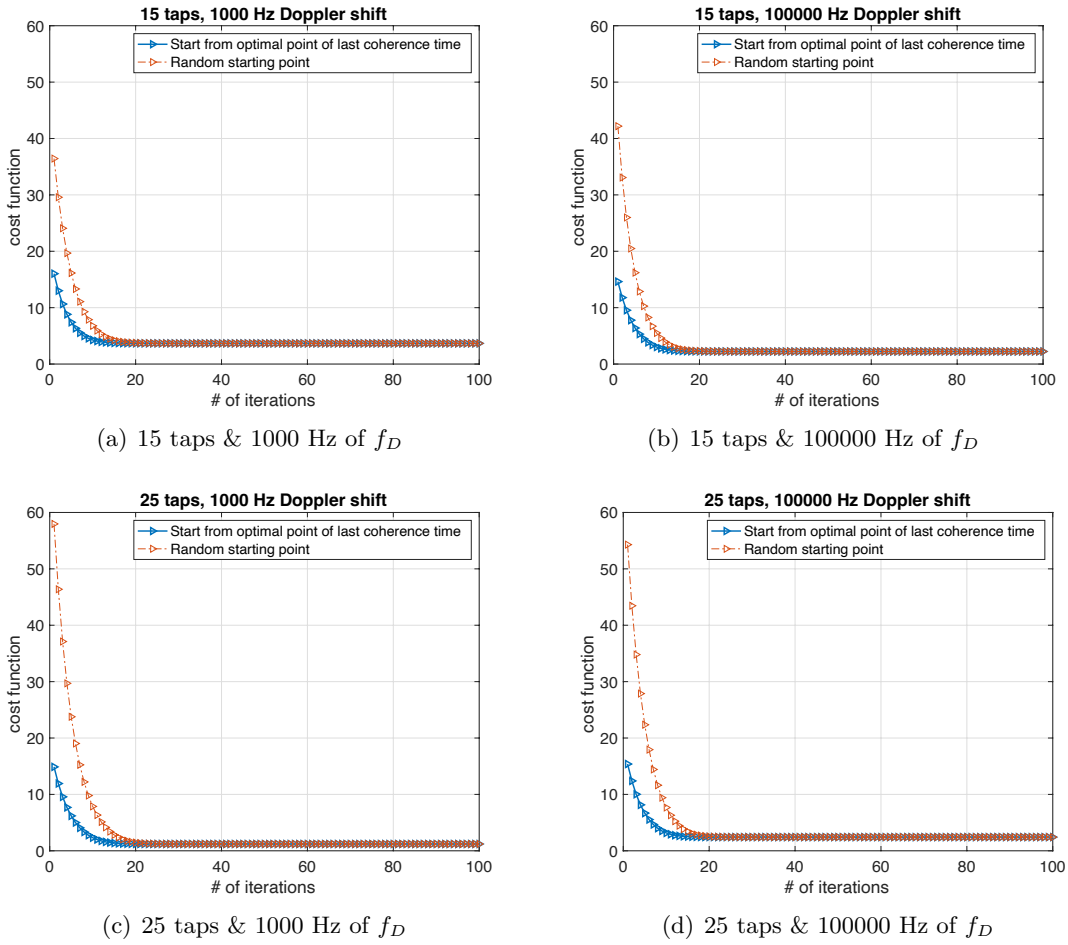
FIGURE 2.12: Achievable cancellation depth of the TDC versus the number of taps.

TDC requires 50 taps to achieve 40 dB of cancellation depth over a 100 MHz bandwidth, while < 10 taps are sufficient for < 20 MHz bandwidth.

Besides, the frequency selectivity of the SI channel will also affect the cancellation performance. In a reflection-rich environment, the RMS delay spread is relatively large, yielding a more selective frequency response. Thus, the SI channel characteristics need more frequency samples to be presented, requiring higher DoFs to solve the minimization problem. Figure 2.12(b) shows the achievable cancellation depth of the TDC with increasing RMS delay spreads and numbers of taps. It illustrates that the required number of taps is almost linearly increased with respect to the operation bandwidth and delay spread of the channel.

As illustrated earlier, the canceller is tuned by the adaptive learning method in practice due to the lack of statistical knowledge of the canceller and SI channel in varying environments, yielding implementation concerns about the tuning speed. Instead of starting at random points, we start at the optimal point of the last coherence time to save the tuning time. This benefit comes from the fact that two consecutive coherent time slots are highly correlated, so the optimal points of the two time slots are very close. Figure 2.13 shows the cost function with iterations of the tuning algorithm under various conditions. It can be seen that starting at the alternative point we proposed can save the tuning time compared with random start points, especially when the Doppler shift of the SI channel (f_D) is small. Besides, the number of taps does not significantly increases the tuning time, and nearly optimal performance can be achieved within 100 iterations

with the alternative start point. Therefore, the tuning time could be much shorter than the coherence time of the SI channel with advanced FPGAs.



(a) 15 taps & 1000 Hz of f_D

(b) 15 taps & 100000 Hz of f_D

(c) 25 taps & 1000 Hz of f_D

(d) 25 taps & 100000 Hz of f_D

FIGURE 2.13: Convergence behavior of the tuning algorithm.

2.8.2 Digital cancellation

We consider an OFDM system where the simulation parameters are listed in Table 2.1. The effective baseband SI channel coefficients $h_{\text{eff},p}(k)$ are randomly generated with $P = 5$ and $K = 3$ since the performance of the nonlinear digital canceller does not rely on specific parameters. The effects of nonideal filters are included in the simulations.

Figure 2.14 shows the variation of the Rx SI during the processing of SIC. We adopt physical separation for antenna isolation (SIC 1), TDC for RF cancellation (SIC 2), and polynomial memory-based canceller for digital cancellation (SIC 3) with different antenna isolation (AI), the dynamic range of receivers (DR), and the number of canceller

TABLE 2.1: Parameters for the 3-step SIC simulations.

Parameters	**Values**
Number of subcarriers	128
Signal bandwidth	400 MHz
Subcarrier modulation	4QAM
Carrier frequency	2.5 GHz
Transmit power (IBFD node)	34 dBm (8 dBm/MHz)
Transmit power (intended node)	33 dBm (7 dBm/MHz)
Phase noise	0.5°
Effective number of bits of ADC	12
Pulse shaping filter	Root-raised-cosine (0.25)
LPF/BPF (Kaiser window)	Transit band = 100/10 MHz

taps (NT). It should be noted that the dynamic range involves the effects of ADCs here, and is a more comprehensive metric to describe the receiver's quality. In a receiver context, a higher resolution ADC will allow for a larger dynamic range, meaning it can handle a wider range of input signal amplitudes without distortion. For simplicity, we approximate the saturation distortion using Gaussian noise that has power linearly related to the input signal power. The results illustrate the properties of different SIC schemes. Antenna isolation mainly suppresses the direct path, so it does not significantly change the characteristics of the frequency response, and its performance is limited by reflection paths. Further improving the isolation level may not help reduce the Rx SI power. With sufficient taps, the RF canceller can effectively suppress the SI over the operation band, and the digital canceller can bring the RSI to approach the noise floor, proving the effectiveness of the nonlinear digital canceller. In contrast, when the Rx SI is not sufficiently suppressed in the analog domain due to limited taps, the digital canceller cannot effectively suppress the RSI due to the receiver saturation. Thus, RF cancellation is critical to achieving effective overall SIC with practical dynamic range-limited receivers.

Figure 2.15 shows the RSI power variation with enhancing ASIC depths. It can be seen that either insufficient or excessive RF cancellation depth will result in a high RSI power, while an appropriate ASIC depth could minimize the SI effects and bring the RSI close to the receiver's noise floor. The results are consistent with our analysis and system-level simulations demonstrated above.

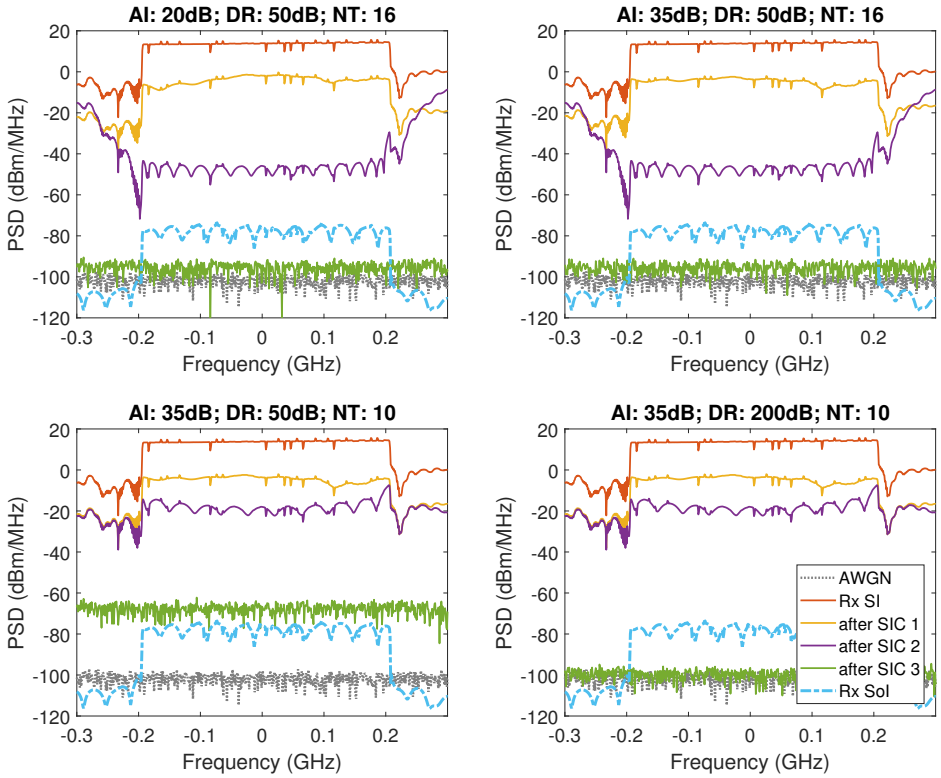


FIGURE 2.14: Power spectral density (PSD) of the RSI signals at each stage of the 3-step SIC process.

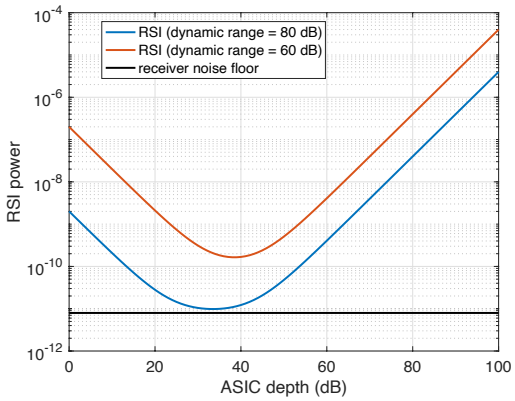


FIGURE 2.15: The power of residual noise after the final digital canceller versus ASIC depth.

2.8.3 Hybrid architecture effects

We consider a 256×256 massive MIMO system with 12-bit ADCs and 24 dBm of transmit power. An OFDM system with 1920 subcarriers and 60 kHz subcarrier spacing is considered, and 20 dB of antenna isolation is assumed.

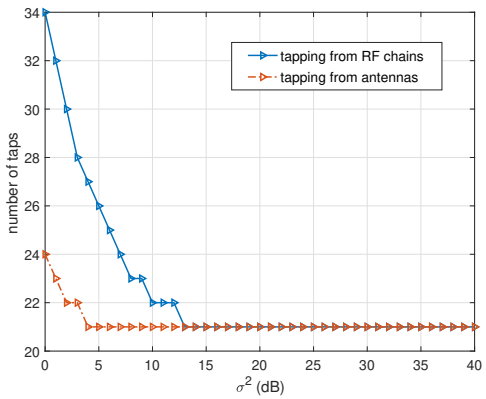


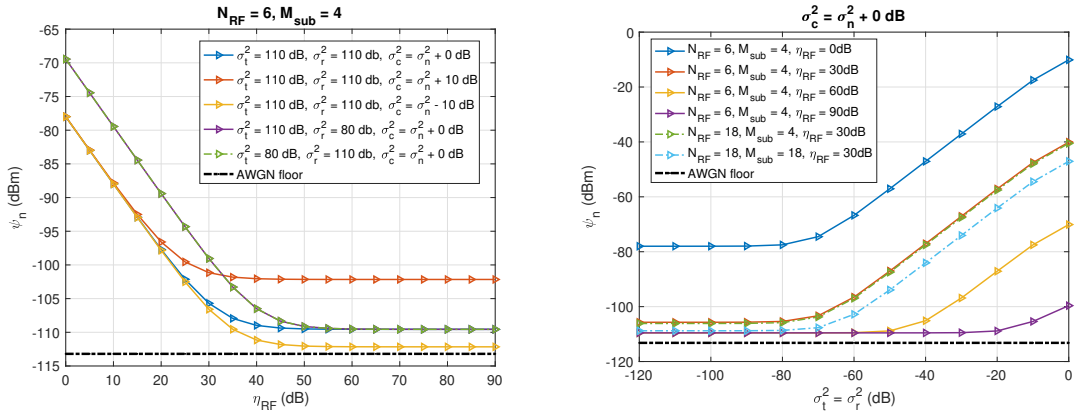
FIGURE 2.16: Number of required taps of a single canceller with normal and hybrid MIMO transceiver architectures under different channel conditions.

Figure 2.16 shows the number of required taps of a single canceller[10] with normal MIMO transceiver and hybrid beamforming architecture, i.e., tapping reference signals from antennas and RF chains, respectively. The x-axis σ^2 captures the difference between RMS delay spreads of the wireless SI channels between the Tx and Rx antenna arrays, i.e., assume $\tau_{\text{RMS}}^{(r,s)} = \bar{\tau}_{\text{RMS}} + \Delta_{\tau_{\text{RMS}},rs}$, where $\Delta_{\tau_{\text{RMS}},rs} \sim \mathcal{N}\left(0, \frac{\bar{\tau}_{\text{RMS}}^2}{\sigma^2}\right)$ and $\bar{\tau}_{\text{RMS}} = 20$ ns. The results suggest that the RF beamforming may increase the delay spread of the effective SI channel if the wireless SI channels have a significant RMS delay spread difference, resulting in more taps required by a single canceller. However, the total number of taps will still be significantly reduced due to the much smaller number of required cancellers. For instance, a total of $34 \times 6 \times 6 = 1224$ taps are required with RF beamforming even at $\sigma^2 = 0$ dB, while the normal MIMO transceiver needs $20 \times 256 \times 256 = 1310720$ taps. Usually, the difference between these wireless SI channels will not be so significant due to their proximity. For scenarios $\sigma^2 > 12$ dB, RF beamforming does not increase the RMS delay spread.

Figures 2.17(a) and 2.17(b) show the overall residual noise level ψ_n after the SIC with different hybrid architectures. The results suggest more subarrays at the receiver can improve the SIC performance and can somehow compensate for the insufficient RF cancellation, relieving the burden of each RF canceller. However, more subarrays require more RF chains, resulting in a higher implementation cost of transceivers. In contrast, the number of RF chains at the transmitter does not have significant effects. In addition, it can be seen that the canceller's noise and distortions will limit the overall SIC

[10] We assume $\rho_t(B_W, \eta_{\text{RF}}) = 1$ for simplicity while it is actually determined by the operational bandwidth and target cancellation depth.

performance. The self-interference can only be suppressed to be close to the receiver AWGN floor if the noise induced by RF cancellers is much lower than the AWGN (i.e., < -10 dB). Deeper RF cancellation depths (i.e., > 35 dB) do not provide significant benefits for the overall SIC, but they enhance tolerance for larger transceiver distortions.



(a) ASIC depth (b) HWIs factor

FIGURE 2.17: Overall residual noise vs RF cancellation depth and transceiver distortions.

2.8.4 User allocation for CCI reduction

We assume all user equipment (UE) and the AP are equipped with an identical number of transmitting and receiving antennas, and the number of antennas at 5G NR access point, eMBB devices, and URLCC devices is 256, 16, and 8, respectively. The access point transmits signals with a total power of 24 dBm that is averagely allocated to the 8 DL users, and the transmit power at the UL users is 23 dBm. The whole bandwidth is equally divided into 4 orthogonal sub-bands for both DL and UL users. The path loss model is taken from [3]. After performing the proposed user allocation algorithm, the UEs are allocated to sub-bands as Table 2.2 shows. It actually allocated the UL users and DL users farthest away from this UL user to the identical sub-band in this simple case. Fig. 2.18 shows the sum rate variation of this network during the user allocation algorithm with three different random initial user allocation policies, which proves the convergence behavior. Different initial policies may result in different policy-changing sequences but will converge to the same final optimal policy.

TABLE 2.2: Optimal user allocation policy

Sub-bands	UL UE	DL UE
sub-band 1 (27.8-27.9 GHz)	4	1, 2
sub-band 2 (27.9-28.0 GHz)	3	3, 4
sub-band 3 (28.0-28.1 GHz)	2	5, 6
sub-band 4 (28.1-28.2 GHz)	1	7, 8

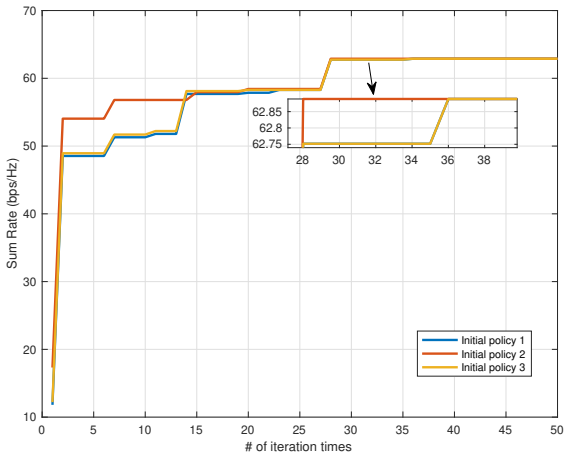


FIGURE 2.18: Sum rate variation with iterations of the user allocation algorithm.

2.9 Conclusions

In this chapter, we studied the interference mitigation schemes to deal with the SI and CCI introduced by IBFD operations. It is revealed by analysis and simulations that effectively suppressing SI before the receiver, which can be realized by the antenna interface designs and RF canceller circuits, is critical to eliminate the SI effects. Appropriate antenna interfaces can provide considerable passive direct path isolation. However, it cannot handle the reflection components caused by objects in the surrounding environment, so it may not prevent the receiver from saturation in real-world scatter-rich environments. Therefore, an RF canceller is necessary for IBFD transceivers. Depending on the tapping location of the reference signal, RF cancellers can be categorized into *Stanford* and *Rice* architectures, whose performance is limited by the number of available taps and transmitter distortions due to practical hardware imperfections, respectively. Furthermore, there exists an optimal ASIC depth, which can minimize the SI effects, depending on the receiver's dynamic range. Finally, the digitized RSI is subtracted by the nonlinear digital canceller. With a sufficient number of taps for RF cancellation, the SI can be suppressed to approach the receiver's noise floor by this 3-step SIC scheme.

Then, we discussed extending the method to MIMO systems and revealed the practical issues of implementing RF cancellation for mmWave communications, where the large-scale antenna arrays will be utilized to compensate for the large pathloss but require a physically-prohibited number of cancellers. By leveraging the hybrid beamforming architecture, we proposed to insert cancellers between RF chain pairs instead of antenna pairs. It is demonstrated that RF beamforming does not significantly increase the frequency-selectivity of the effective SI channel, so the required number of taps for a single canceller does not increase. Thus, the total required true tunable delay lines (i.e., taps) will be significantly reduced to a feasible number due to the much fewer RF chains. In addition, more antenna subarrays at the receiver will improve the performance of digital cancellers. In addition, we proposed a user allocation algorithm to mitigate the CCI from the physical layer. This algorithm allocates users that could cause strong interference into orthogonal sub-bands while leaving users with weak interference in overlapping sub-bands, which is realized by a game theory method. The effects of CCI can be minimized from the physical layer via this algorithm and numerical results indicate significant throughput improvement due to appropriate user allocation policy.

Appendix 2.10 3D FR2 channel model

The SI channel consists of the LOS and NLOS components and is given as [38]

$$g_{\text{si}}(t) = \sqrt{\frac{K}{(1+K)}} h^{\text{LOS}} + \sqrt{\frac{1}{(K+1)}} \sum_{n=1}^{N} \sum_{m_n=1}^{M_n} h_{n,m_n}^{\text{NLOS}}(t - \tau_n - \tau_{m_n}), \tag{2.72}$$

where N is the number of clusters, M_n is the number of rays in n^{th} cluster, τ_n is the delay of the n^{th} cluster, τ_{m_n} is the delay of m_n^{th} ray in the n^{th} cluster, and K is the Rician factor. h^{LOS} and h_{n,m_n}^{NLOS} denote the complex channel gains for LOS and NLOS paths that can be found in [33, 38]. This 3D channel model can be simplified to the TDL model as

$$g_{\text{si}}(t) = h_{rs,0}\delta(t - \tau_{rs,0}) + \sum_{l=1}^{L} h_{rs,l}\delta(t - \tau_{rs,l}), \tag{2.73}$$

where h and τ describe the attenuation and delay of associated paths (e.g., $h_{rs,0} = \sqrt{\frac{K}{(1+K)}} h^{\text{LOS}}$ and $\tau_{rs,0} = 0$ for the LOS path, and $h_{rs,l} = \sqrt{\frac{1}{(K+1)}} h_{n,m_n}^{\text{NLOS}}$ and $\tau_{rs,l} = \tau_n + \tau_{m_n}$ for NLOS paths), $L = \sum_{n=1}^{N} M_n$ is the total number of NLOS paths.

Appendix 2.11 Receiver noise covariance matrix derivation

The performance of digital cancellation strongly depends on the accuracy of channel coefficients estimation, so we will first derive the statistics of the estimation error $\mathbf{\Delta}_{\text{ch}}$. The subcarrier index k is omitted here since the statistics are identical for all subcarriers.

$$\mathbb{E}\left\{\mathbf{\Delta}_{\text{ch}}\right\} = \mathbb{E}\left\{\mathbf{G}_{\text{eff}} - \hat{\mathbf{G}}_{\text{eff}}\right\} = \mathbb{E}\left\{\mathbf{H}_{\text{eff}} - \left(\mathbf{H}_{\text{eff}}\mathbf{s} + \tilde{\mathbf{z}}\right)\left(\mathbf{s}^{\dagger}\mathbf{s}\right)^{-1}\mathbf{s}^{\dagger}\right\} = \mathbf{0}. \tag{2.74}$$

The covariance matrix is given as

$$\begin{aligned} Var\left(\mathbf{\Delta}_{\text{ch}}\right) &= \mathbb{E}\left\{\left(\mathbf{\Delta}_{\text{ch}} - \mathbb{E}\left\{\mathbf{\Delta}_{\text{ch}}\right\}\right)\left(\mathbf{\Delta}_{\text{ch}} - \mathbb{E}\left\{\mathbf{\Delta}_{\text{ch}}\right\}\right)^{\dagger}\right\} \\ &= \mathbb{E}\left\{\left(\mathbf{G}_{\text{eff}} - \hat{\mathbf{G}}_{\text{eff}}\right)\left(\mathbf{G}_{\text{eff}} - \hat{\mathbf{G}}_{\text{eff}}\right)^{\dagger}\right\} = \frac{1}{P_k}\mathbb{E}\left\{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\dagger}\right\}, \end{aligned} \tag{2.75}$$

where the covariance matrix of the noise and distortions are given as

$$\begin{aligned} \mathbb{E}\left\{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\dagger}\right\} &= \mathbb{E}\left\{\mathbf{G}_{\text{eff}}\mathbf{d}_{\text{tx}}\mathbf{d}_{\text{tx}}^{\dagger}\mathbf{G}_{eff}^{\dagger}\right\} + \mathbb{E}\left\{\mathbf{d}_{\text{rfc}}\mathbf{d}_{\text{rfc}}^{\dagger}\right\}\mathbb{E}\left\{\mathbf{d}_{\text{rx}}\mathbf{d}_{\text{rx}}^{\dagger}\right\} + \mathbb{E}\left\{\mathbf{n}\mathbf{n}^{\dagger}\right\} \\ &= \kappa\mathbb{E}\left\{\mathbf{G}_{\text{eff}}\mathcal{D}\left(\mathbf{F}_{\text{RF}}\mathbf{s}\mathbf{s}^{\dagger}\mathbf{F}_{\text{RF}}^{\dagger}\right)\mathbf{G}_{\text{eff}}^{\dagger}\right\} \\ &\quad + \beta\cdot\mathbb{E}\left\{\mathcal{D}\left(\left(\tilde{\mathbf{y}} - \mathbf{r}\right)\left(\tilde{\mathbf{y}} - \mathbf{r}\right)^{\dagger}\right)\right\} + (\sigma_c^2 + \sigma_n^2)\mathbf{I}_{M_{\text{RF}}}. \end{aligned} \tag{2.76}$$

We can assume that the residual received signal $\tilde{\mathbf{y}}$ is dominated by the RSI without loss of generality (i.e., $\tilde{\mathbf{y}} \approx \mathbf{G}_{\text{eff}}\,\mathbf{s}$ and $\tilde{\mathbf{y}} - \mathbf{r} \approx \mathbf{G}_{\text{eff}}\,\mathbf{s}$), then Equation (2.76) can be written as

$$\begin{aligned} \mathbb{E}\left\{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\dagger}\right\} &\approx \frac{\kappa P_k}{N_{\text{RF}}}\mathbb{E}\left\{\mathbf{G}_{\text{eff}}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{RF}}^{\dagger}\mathbf{G}_{\text{eff}}^{\dagger}\right\} + \left(\sigma_c^2 + \sigma_n^2 + \frac{\beta P_k \varrho_{\text{si}}}{(1+\kappa)\eta_{\text{RF}}}\frac{M_{\text{ant}}}{M_{\text{sub}}}\right)\mathbf{I}_{M_{\text{RF}}} \\ &= \frac{\kappa P_k}{N_{\text{RF}}}\cdot\frac{\varrho_{\text{si}}N_{\text{RF}}M_{\text{ant}}}{(1+\kappa)\eta_{\text{RF}}M_{\text{sub}}}\mathbf{I}_{M_{\text{RF}}} + \left(\sigma_c^2 + \sigma_n^2 + \frac{\beta P_k \varrho_{\text{si}}}{(1+\kappa)\eta_{\text{RF}}}\frac{M_{\text{ant}}}{M_{\text{sub}}}\right)\mathbf{I}_{M_{\text{RF}}} \\ &= \underbrace{\left(\frac{(\kappa+\beta)\varrho_{\text{si}}P_k M_{\text{ant}}}{(1+\kappa)\eta_{\text{RF}}M_{\text{sub}}} + \sigma_c^2 + \sigma_n^2\right)}_{\varepsilon_z}\mathbf{I}_{M_{\text{RF}}} = \varepsilon_z\cdot\mathbf{I}_{M_{\text{RF}}}. \end{aligned} \tag{2.77}$$

Thus, we can represent the channel estimation error covariance matrix as $Var\left(\mathbf{\Delta}_{\text{ch}}\right) = \frac{\varepsilon_z}{P_k}\cdot\mathbf{I}_{M_{\text{RF}}}$. The covariance matrix of the overall residual noise level of the receiver is

given as

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{\Psi}\boldsymbol{\Psi}^{\dagger}\right\} &= \mathbb{E}\left\{\left(\boldsymbol{\Delta}_{\mathrm{ch}}\mathbf{s}+\tilde{\mathbf{z}}\right)\left(\boldsymbol{\Delta}_{ch}\mathbf{s}+\tilde{\mathbf{z}}\right)^{\dagger}\right\} \\ &= \frac{P_k}{N_{\mathrm{RF}}}\mathbb{E}\left\{\boldsymbol{\Delta}_{\mathrm{ch}}\boldsymbol{\Delta}_{\mathrm{ch}}^{\dagger}\right\}+\mathbb{E}\left\{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\dagger}\right\} \\ &= \underbrace{\left(1+\frac{1}{N_{\mathrm{RF}}}\right)\varepsilon_z}_{\psi_n}\cdot\mathbf{I}_{M_{\mathrm{RF}}} = \psi_n\mathbf{I}_{M_{\mathrm{RF}}}.\end{aligned} \tag{2.78}$$

Chapter 3

Photonics-Assisted RF Canceller

3.1 Introduction

A large number of taps (i.e., true tuneable delay lines) is critical to enable wideband IBFD radios. However, it is demonstrated that the hardware imperfections limit the capacity of the TDC to provide sufficient taps. It should be noted that Equation (2.10) describes the TDC without practical imperfections, while a practical TDC should be modeled as

$$g_{\text{tdc}}(t) = \sum_{m=0}^{M-1} \varsigma_m w_m e^{j2\pi(\phi_m + \Delta_{\phi,m})} \delta(t - \tau_m - \Delta_{\tau,m}), \tag{3.1}$$

where ς_m is the attenuation coefficient of the m^{th} tap due to the insertion loss caused by real electrical components, $\Delta_{\phi,m}$ and $\Delta_{\tau,m}$ describe the phase noise and delay error of the m^{th} tap, respectively. It is experimentally demonstrated that the insertion loss increases significantly with increasing taps for the electrical component (e.g., microstrips and coaxial cables) based canceller. Due to the large insertion loss, the reference signals flowed into later taps will be significantly attenuated by ς_m, making them too small to have effects on the accumulated output of the TDC. Thus, these taps are invalid even though they have been created. In addition, electrical components usually have a narrow frequency coverage range, which indicates that the electrical-based canceller is subjected to both the operating frequency and bandwidth. In contrast, photonics has a much lower insertion loss, which is promising to provide a large number of effective taps.

Work in this chapter has been published in IEEE Transactions on Wireless Communications, November 2021 [38] and IEEE Access, December 2021 [9], with a preliminary version presented at Asilomar Conference on Signals, Systems, and Computers November 2020 [40] and IEEE ICC June 2021 [37].

The insertion loss of traditional electrical delay lines and photonic-assisted delay lines are compared in [4], which is shown in Figure 3.1. It demonstrates that the elaborate photonics-assisted canceller (PAC) design (i.e., the fiber Bragg grating (FBG)-based canceller) has the potential to create hundreds of effective taps, while this number is limited to be within 20 for electrical delay lines.

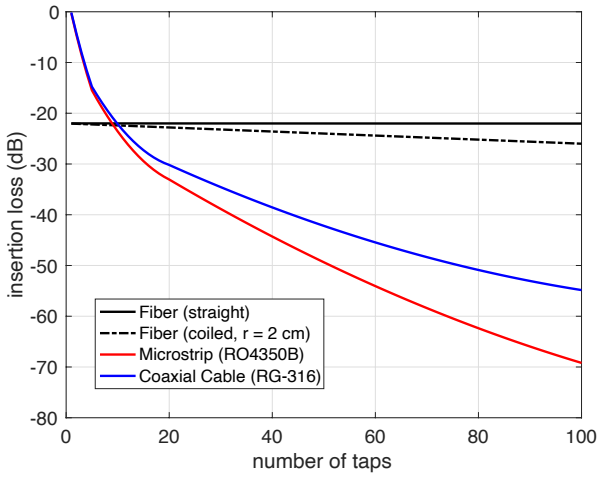


FIGURE 3.1: Canceller insertion loss comparison that includes traditional electrical delay lines and FBG delay lines for both straight and coiled cases (fetched from Figure 8 in [4]).

The trend of the wireless network evaluation is moving to higher frequency bands to enhance the system capacity in terms of throughput, user connections, etc., as stated earlier in Chapter 1. In contrast, existing electrical delay lines reported in the literature usually consider $<$ 100 MHz bandwidth within $<$ 6 GHz frequencies, which cannot meet the requirements of future networks. Therefore, a high-flexibility canceller capable of high operating frequencies is critical for future IBFD radios, which motivates PAC studies due to the intrinsic superiorities of large instantaneous bandwidth, broad frequency coverage range, precise wide time-delay tuning, and immunity to electromagnetic interference of microwave photonics [15]. Furthermore, the PAC is also widely studied by the Society of Radio-over-fiber (RoF), which wants to enable IBFD RoF for higher spectral efficiency.

There are two types of PAC designs in general, depending on the domain where the cancellation happens. The first type cancels the SI in the optical domain, which requires converting the received signal into the optical domain in addition to the reference signal [41, 42]. The second type can be regarded as a variant of the TDC but constructs

the adaptive FIR in the optical domain, so it cancels the SI in the RF domain and only converts the reference signal into the optical domain [4]. Both types have been experimentally demonstrated to provide effective wideband RF cancellation.

In [41], a tunable multipath photonics delay line is utilized for SIC, where each path consists of a variable optical attenuator (VOA) and variable optical delay line (VODL). Thanks to the large adjusting arrangement and high tuning precision of photonic components, 30.18 dB and 20.88 dB cancellation depths are realized within the bandwidth of 100 MHz and 2 GHz centered at 16.6 GHz. In [43], an optical spectrum processer (OSP) is leveraged to finely adjust the amplitude and phase instead of VOAs and VODLs, achieving 30 dB cancellation over 10 GHz bandwidth (from 13 GHz to 23 GHz). A more elaborate design is proposed in [4], which utilizes fiber Bragg gratings to cause uniformly-distributed delays in a single fiber so that the VODLs are not needed, and an average of 20 dB cancellation over 1 GHz bandwidth centered at 2.5 GHz is achieved. In [42], the SI is canceled in the optical domain, where an array of tunable lasers with a shared electro-optic modulator (EOM) is used to process the reference signal. An ultrawideband RF cancellation with SoI recovery is implemented for 16QAM symbols, where 20 dB cancellation over 9 GHz bandwidth (from 0.5 GHz to 9.5 GHz) and 20 dB cancellation over 5.8 GHz bandwidth (from 10 GHz to 15.8 GHz) are achieved, respectively.

To reduce the hardware complexity, studies recently explored advanced digital filtering to pre-process the reference signal. In [44], a digital-assisted design is proposed to cancel the multi-path SI components using only a single delay line with the help of a digital adaptive filter. The direct path coupling is canceled through an analog reference signal, while the reflection path coupling is mitigated by the adaptive filter to address the issue of limited DAC resolution. Similarly, authors in [45] also construct and adaptively adjust the reference signal in the digital domain and then convert both the reference and received signals into the optical domain to cancel via a dual-drive Mach–Zehnder modulator (DD-MZM) showing 20 dB cancellation for up to 9 GHz single-tone SI signals.

The feasibility and effectiveness of PAC have been verified by discrete fiber-optic components in previous studies. However, there are still concerns related to size, production, energy consumption, etc., for practical applications, which could be addressed by integrated photonics chips. Thanks to the mature fabrication processes from the

CMOS industry, it is promising to implement the proposed designs on a chip, realizing a portable and energy-efficient PAC that can be mass manufactured [15]. An integrated dual-parallel Mach-Zehnder modulator (DP-MZM) based chip scale canceller is demonstrated in [46], which provides 30 dB cancellation for up to 20 GHz frequencies.

The main contribution of our work is that we focus on the photonics-assisted TDC design and propose the tuning algorithm. Then, we propose a new architecture of the PAC, namely fiber array-based canceller, which is similar to the one in [41] but we published earlier. We verify its effectiveness through mathematical derivations, hardware simulations, and experiments in order. Based on the experiment measuring, we analyze its scalability to more taps. In addition to the PAC, we also implement a metal-based isolation and frequency domain-based digital canceller using 5G NR protocols on the software-defined radio (SDR) platform along with the proposed PAC, evaluating the performance of this 3-step SIC scheme.

The rest of this chapter is organized as follows. In Section 3.2, we give a general architecture of the photonics-assisted TDC and study the tuning algorithms accordingly. Section 3.3 introduces two implementations of the photonics-assisted TDC, including a Fiber Bragg gratings-based design and our proposed fiber array canceller. Simulation results are given in Section 3.4 to demonstrate the superiority of PAC and verify the effectiveness of the proposed architecture. Then, we give the experimental works in Section 3.5, including the setup details, chosen components, and measuring results. Finally, conclusions are drawn in Section 3.6.

3.2 The general photonics-assisted TDC

3.2.1 Canceller representation

Figure 3.2 depicts a general architecture of the photonics-assisted canceller, where an EOM and a photo-detector (PD) are employed to switch the signal between the RF and optical domain. It should be noted that it is generally difficult to tune the phase of the modulated RF signal in the optical domain, so four identical branches are alternatively utilized to tune a complex coefficient. Each branch has the architecture as Figure 3.2, but their input RF signals are phase shifted, and the phase shifters on the FIR filter

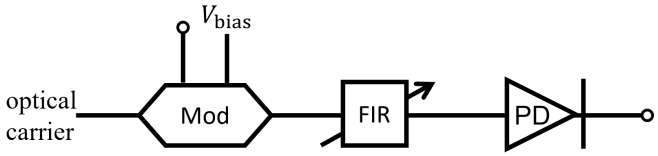


FIGURE 3.2: A general model of the photonics-assisted canceller.

taps (as shown in Figure 2.8(b)) are removed.

$$\begin{aligned} g_{\text{pac}}(t) = \sum_{m=0}^{M-1} \big(& \varsigma_{I+,m} w_{I+,m} e^{j\phi_{I+}} \delta(t-\tau_{I+,m}) + \varsigma_{I-,m} w_{I-,m} e^{j\phi_{I-}} \delta(t-\tau_{I-,m}) \\ & + \varsigma_{Q+,m} w_{Q+,m} e^{j\phi_{Q+}} \delta(t-\tau_{Q+,m}) + \varsigma_{Q-,m} w_{Q-,m} e^{j\phi_{Q-}} \delta(t-\tau_{Q-,m}) \big), \end{aligned} \tag{3.2}$$

where $\phi_{I+} = 0$, $\phi_{I-} = \pi$, $\phi_{Q+} = \frac{\pi}{2}$, and $\phi_{Q-} = \frac{3\pi}{2}$ are the phase shifting of the four branches, which are implemented directly to the RF signal before the optical modulator. $w_{I+,m}$, $w_{I+,m}$, $w_{I+,m}$, and $w_{I+,m}$ are real positive weights that can be adjusted by VOAs. Ignore the minor delay and loss differences of the m^{th} tap on the four branches, Equation (3.2) can be written as

$$g_{\text{pac}}(t) = \sum_{m=0}^{M-1} \varsigma_m \left(w_{I+,m} - w_{I-,m} + j w_{Q+,m} - j w_{Q-,m} \right) \delta(t-\tau_m), \tag{3.3}$$

which is identical to the TDC given in Equation (3.1) (phase noise and delay error ignored) with

$$w_m = \sqrt{\left(w_{I+,m} - w_{I-,m}\right)^2 + \left(w_{Q+,m} - w_{Q-,m}\right)^2}, \tag{3.4}$$

$$\phi_m = \arctan\left(\frac{w_{Q+,m} - w_{Q-,m}}{w_{I+,m} - w_{I-,m}} \right). \tag{3.5}$$

3.2.2 Tuning algorithm

As Equation (3.3) suggests, the amplitude and phase are adjusted by these real weights. Usually, we utilize passive optical tunable attenuators on each tap to reduce the power consumption and nonlinearity of the canceller. Thus, the tuning problem can be cast as

$$\begin{aligned} & \min_{\{w_{I+,m}, w_{I-,m}, w_{Q+,m}, w_{Q-,m}\}_{m=0}^{M-1}} \int_t \left| g_{\text{pac}}(t) - g_{\text{si}}(t) \right|^2 dt, \\ & s.t. \; -1 \leqslant w_{p,m} \leqslant 1, \; \forall \; m \in [0, M-1], \; p \in \{I+, I-, Q+, Q-\} \end{aligned} \tag{3.6}$$

where the constraints come from the passive VOAs. This could also be converted into the frequency domain and written in the matrix form as the TDC. However, due to the real-valued nature of the tuning weights of the photonics-assisted canceller, it cannot be tuned in the complex form as the TDC does. In order to obtain real-valued solutions, we need to formulate a real-valued optimization problem. Thus, we divide the complex channel and canceller frequency samples into real and imaginary parts as

$$\widetilde{\mathbf{G}}_{\text{pac}} = \begin{bmatrix} \mathcal{R}\{\mathbf{g}_{I+}\} & \mathcal{R}\{\mathbf{g}_{I-}\} & \mathcal{R}\{\mathbf{g}_{Q+}\} & \mathcal{R}\{\mathbf{g}_{Q-}\} \\ \mathcal{I}\{\mathbf{g}_{I+}\} & \mathcal{I}\{\mathbf{g}_{I-}\} & \mathcal{I}\{\mathbf{g}_{Q+}\} & \mathcal{I}\{\mathbf{g}_{Q-}\} \end{bmatrix}, \tag{3.7}$$

$$\mathbf{G}_{\text{si}} = \left[\mathcal{R}\left\{\mathbf{g}_{\text{si}}^T\right\}, \mathcal{I}\left\{\mathbf{g}_{\text{si}}^T\right\}\right]^T, \tag{3.8}$$

where $\mathbf{g}_p = \left[G_p(f_c - \frac{B_W}{2}), G_p(f_c - \frac{B_W}{2} + \Delta f), \ldots, G_p(f_c + \frac{B_W}{2})\right]^T$ with p denoting the place holder for $I+$, $I-$, $Q+$, and $Q-$ and $G_p(f)$ denoting the measured frequency response of corresponding branch, and $\mathbf{g}_{\text{si}}$ is the frequency sample vector of SI channel given in Equation (2.16). Then, a real-valued tuning problem can be formulated as

$$\begin{aligned} &\min_{\mathbf{W}_{\text{pac}}} \quad \left\| \widetilde{\mathbf{G}}_{\text{pac}} \mathbf{W}_{\text{pac}} - \mathbf{G}_{\text{si}} \right\|^2, \\ &s.t. -1 \leqslant \mathbf{W}_{\text{pac}} \leqslant 1 \end{aligned} \tag{3.9}$$

where $\mathbf{W}_{\text{pac}} = [\mathbf{w}_{I+}, \mathbf{w}_{I-}, \mathbf{w}_{Q+}, \mathbf{w}_{Q-}]^T$ with $\mathbf{w}_p = [w_{p,0}, w_{p,1}, \cdots, w_{p,M-1}]$. The optimal weights can also be given by the Wiener solution as

$$\overline{\mathbf{W}}_{\text{pac}} = \left(\mathbb{E}\left\{\widetilde{\mathbf{G}}_{\text{pac}}^{\dagger} \widetilde{\mathbf{G}}_{\text{pac}}\right\}\right)^{-1} \mathbb{E}\left\{\widetilde{\mathbf{G}}_{\text{pac}}^{\dagger} \mathbf{G}_{\text{si}}\right\}. \tag{3.10}$$

Remarking: The size of the matrices $\widetilde{\mathbf{G}}_{\text{pac}}$ and $\mathbf{G}_{\text{si}}$ are of $2\left(\frac{B_W}{\Delta f} + 1\right) \times 4M$ and $2\left(\frac{B_W}{\Delta f} + 1\right) \times 1$, respectively. Evaluating the computational complexity by multiplication operations, the complexity of calculating $\overline{\mathbf{W}}_{\text{pac}}$ requires

$$32M^2\left(\frac{B_W}{\Delta f} + 1\right) + 16M^2 + 8M\left(\frac{B_W}{\Delta f} + 1\right)$$

multiplications plus an inverse operation to a matrix of $4M \times 4M$, where the complexity of the inverse operation is given as $\mathcal{O}\left(64M^3\right)$. In contrast, the size of matrices $\mathbf{\Pi}_{\text{tdc}}$ and $\mathbf{g}_{\text{si}}$ in Equation (2.21) are of $\left(\frac{B_W}{\Delta f} + 1\right) \times M$ and $\left(\frac{B_W}{\Delta f} + 1\right) \times 1$, respectively. Thus, the

complexity of calculating $\overline{\overline{\mathbf{W}}}_{\text{tdc}}$ requires

$$M^2\left(\frac{B_W}{\Delta f}+1\right)+M^2+M\left(\frac{B_W}{\Delta f}+1\right)$$

multiplications plus an inverse operation to a matrix of $M \times M$, where the complexity of the inverse operation is given as $\mathcal{O}\left(M^3\right)$.

Thus, the enlarged matrix sizes will significantly increase the computational complexity of the tuning algorithm. Especially for wideband operations, i.e., large Δf, which increase both dimensions of the 2D matrix since wideband operation requires a large number of taps as illustrated above. In order to have a more intuitive understanding, Figure 3.3 shows the time required to calculate the weights with different tap numbers and bandwidths.

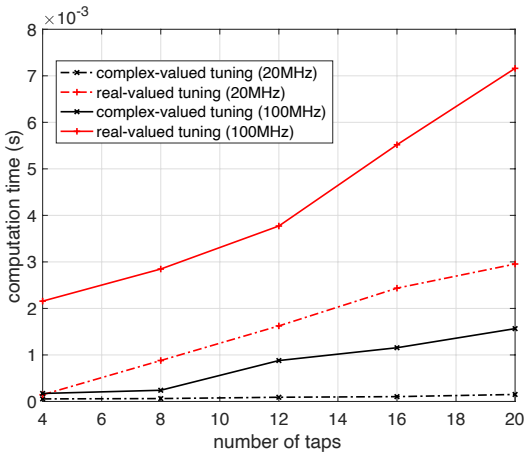


FIGURE 3.3: Time taken to calculate the weight vectors as Equations (2.21) and (3.10).

To reduce the computation overheads, we propose to obtain the weights in an alternative way. Assuming the four branches have almost identical responses that it can be described by Equation (3.3) with trivial mismatches, we can calculate the weights in the complex form and assign them to the four branches. The canceller can be tuned in the complex form as introduced in Section 2.4.2[1], yielding the complex weight vector $\mathbf{W}_{\text{tdc}} = \left[w_0 e^{j2\pi\phi_0}, w_1 e^{j2\pi\phi_1}, \ldots, w_{M-1} e^{j2\pi\phi_{M-1}}\right]^T$. Then, the real-valued weights of the four branches can be given as

$$w_{I+,m} = \max\left\{\mathcal{R}\left\{w_m e^{j2\pi\phi_m}\right\}, 0\right\}$$

[1]The hardware imperfections (e.g., insertion loss, phase error, delay mismatch, etc.) could be included through corresponding coefficients, which could be acquired by measuring the canceller.

$$
\begin{aligned}
w_{I-,m} &= \max\left\{-\mathcal{R}\left\{w_m e^{j2\pi\phi_m}\right\}, 0\right\} \\
w_{Q+,m} &= \max\left\{\mathcal{I}\left\{w_m e^{j2\pi\phi_m}\right\}, 0\right\} \\
w_{Q-,m} &= \max\left\{-\mathcal{I}\left\{w_m e^{j2\pi\phi_m}\right\}, 0\right\}.
\end{aligned} \tag{3.11}
$$

Simulation results demonstrate that the same cancellation depth can be achieved if the four branches are identical, but the computation overheads can be significantly reduced by doing so.

3.3 Implementation architectures

3.3.1 Fiber Bragg grating-based design

The FBG-based design is proposed in [4], which attracts our interest due to its ability to construct multiple taps within a single fiber cable. To understand how FBGs can construct the FIR filter, we will first introduce the concept of the FBG. The FBG is the periodic refractive index variation along the core of a fiber optic cable, which can reflect signals at its associated grating wavelength. A fiber with FBGs connected to a 3-port optical circulator is shown in Figure 3.4, where the signal flows from port 1 to port 2 and enters the fiber with FBGs. The gratings inside the fiber will reflect the signal at specific wavelengths, and the reflected signal goes back to port 2 and flows to port 3. Thus, the link from port 1 to port 3 can be regarded as a bandpass filter (BPF). Figure

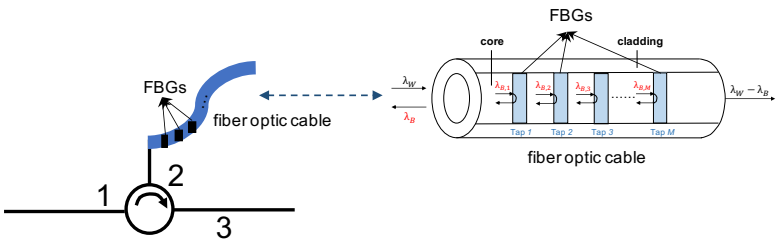


FIGURE 3.4: A fiber with FBGs connected to a 3-port optical circulator.

3.5 shows the frequency response of a fiber with 8 FBGs, where the red line is obtained through observing from port 3 and the black line is obtained through observing from the other end of the cable.

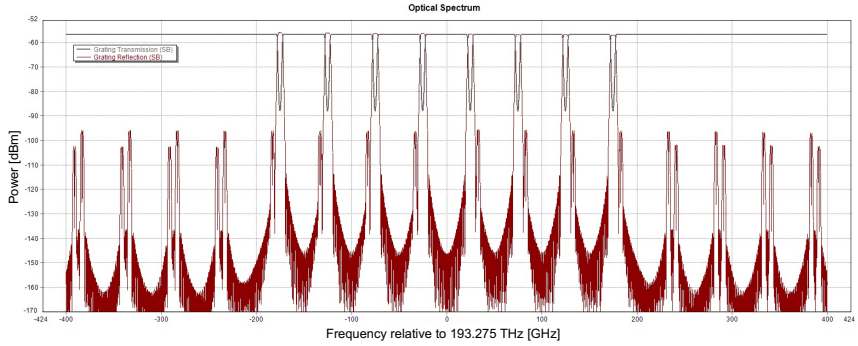


FIGURE 3.5: The optical spectrum observed from port 2 and the other end of the fiber.

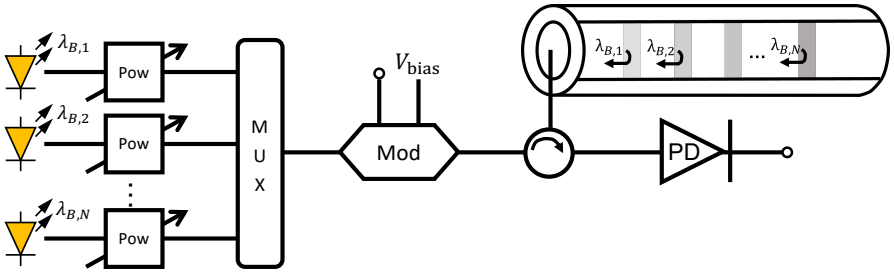


FIGURE 3.6: The fiber Bragg gatings-based canceller.

Taking advantage of this feature, an elaborate optical FIR filter can be constructed by a single fiber cable as Figure 3.6[2]. The coupled transmitted RF signal is modulated onto M optical carriers generated by tunable lasers with wavelengths uniformly distributed in the range of $[\lambda_{B,1}, \lambda_{B,M}]$ at first. Signals modulated onto these optical carriers are reflected at the corresponding gratings. The signals are delayed differently while propagating through the fiber due to different grating locations, and the tunable weights are implemented by adjusting the power of the optical carriers, realizing an FIR filter.

Remarking: Equation (3.11) indicates that half of the weights will be 0 using the alternative tuning algorithm. The weights are 0 means that corresponding optical carriers are not generated. Therefore, this algorithm can reduce the usage of optical carriers by half compared to tuning in the real-valued form. So, the proposed algorithm can save the implementation cost and power of the FBG-based canceller.

3.3.2 Fiber array-based design

The FBG-based design requires multiple laser sources to generate multiple different-wavelength optical carriers to be processed in a single fiber, which yields cost and size

[2]This figure shows the architecture of arbitrary one of the four $I+$, $I-$, $Q+$, $Q-$ branches since they are identical. The difference between the four branches is the phase of the input RF reference signal, which will be detailed later in Figure 3.7

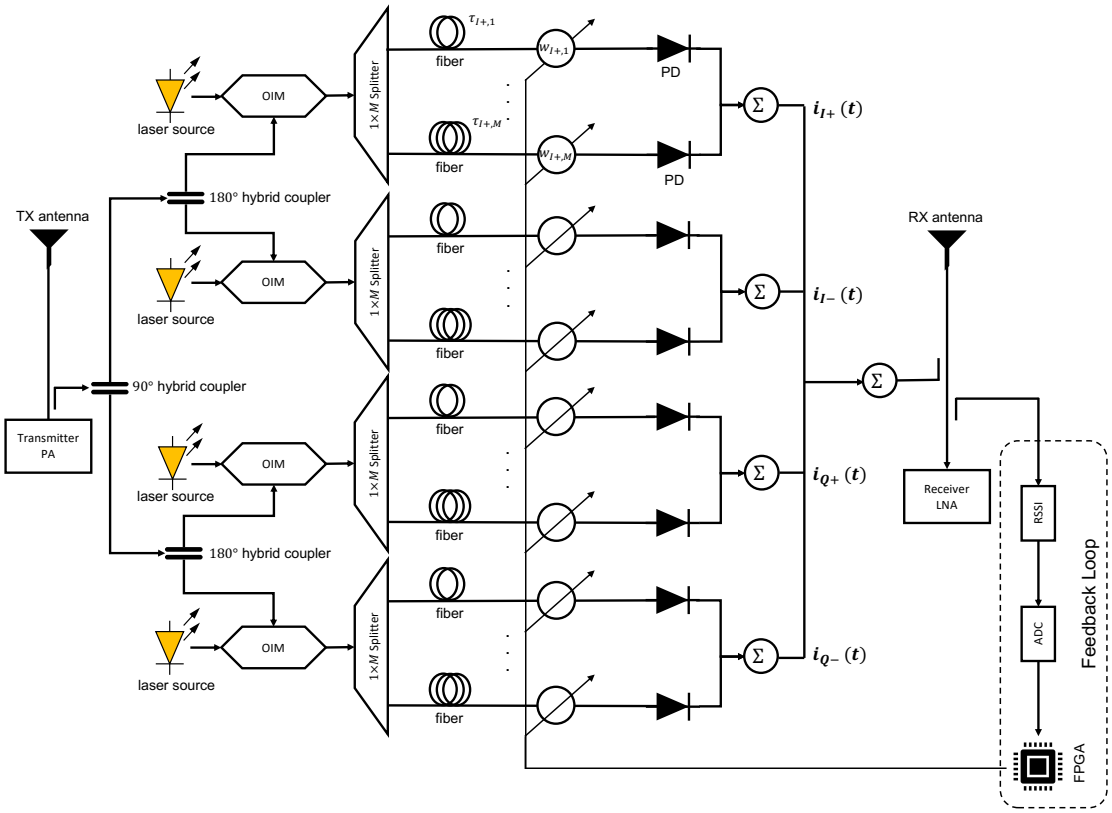


FIGURE 3.7: A schematic diagram of the fiber array-based canceller.

concerns. In addition, the elaborate FBG itself increases the cost and introduces integrating implementation concerns. In this section, we propose a more general fiber array-based design using common optical components, which is shown in Figure 3.7. This figure provides a complete architecture of this design, including the four branches with phase-shifted RF input and the feedback loop to tune the canceller.

To understand how it forms the FIR filter, we derive the mathematical signal processing representation. First, the transmitted RF signal $x(t)$ is tapped into the canceller via a 90° hybrid coupler with coupling factor α_1, which equally splits the RF reference signal with a resultant 90° phase shift between two output ports as

$$\begin{cases} x_I(t) = \sqrt{\frac{1}{2}}\alpha_1 e^{0j\pi} \cdot x(t) = \sqrt{\frac{1}{2}}\alpha_1 \cdot x(t) \\ x_Q(t) = \sqrt{\frac{1}{2}}\alpha_1 e^{j\pi/2} \cdot x(t) = j\sqrt{\frac{1}{2}}\alpha_1 \cdot x(t) \end{cases} \tag{3.12}$$

Then the two signals are equally split by a 180° hybrid coupler, which introduces α_2 coupling factor and 180° phase shift between two output ports as

$$\begin{cases} x_{I+}(t) = \sqrt{\frac{1}{2}}\alpha_2 e^{0j\pi} \cdot x_I(t) = \frac{1}{2}\alpha_1\alpha_2 \cdot x(t) \\ x_{I-}(t) = \sqrt{\frac{1}{2}}\alpha_2 e^{j\pi} \cdot x_I(t) = -\frac{1}{2}\alpha_1\alpha_2 \cdot x(t) \\ x_{Q+}(t) = \sqrt{\frac{1}{2}}\alpha_2 e^{0j\pi} \cdot x_Q(t) = \frac{j}{2}\alpha_1\alpha_2 \cdot x(t) \\ x_{Q-}(t) = \sqrt{\frac{1}{2}}\alpha_2 e^{j\pi} \cdot x_Q(t) = -\frac{j}{2}\alpha_1\alpha_2 \cdot x(t) \end{cases}. \tag{3.13}$$

The four phase-shifted reference signals are independently processed by the four branches, which have identical architectures to perform identical functions. Thus, we assume they have identical mathematical representations without loss of generality. Take the $I+$ branch as an instance, the RF signal $x_{I+}(t)$ is converted into the optical domain through an EOM with the optical carrier generated by the laser source, yielding the optical signal as

$$\begin{aligned} p_{I+}(t) &= l_{\text{eom}} A_o \left(1 + x_{I+}(t)\right) \cos(2\pi f_o t) \\ &= l_{\text{eom}} A_o \left(1 + \frac{1}{2}\alpha_1\alpha_2 \cdot x(t)\right) \cos(2\pi f_o t), \end{aligned} \tag{3.14}$$

where A_o and f_o denote the amplitude and frequency of the optical carrier, and l_{eom} is the insertion loss of the EOM. The modulated optical signal is then equally split by the M-tap splitter, and each output port of the splitter can be described as

$$p_{I+,m}(t) = \sqrt{\frac{1}{M}} l_{\text{sp}} \cdot p_{I+}(t) = \sqrt{\frac{1}{M}} l_{\text{sp}} l_{\text{eom}} A_o \left(1 + \frac{1}{2}\alpha_1\alpha_2 \cdot x(t)\right) \cos(2\pi f_o t), \tag{3.15}$$

where l_{sp} is the insertion loss of the splitter. These optical signals flow into M different-length fibers with VOAs and photo-diodes, which is referred to as a fiber array. The fiber array independently delays and weights the M signals, and the input signal to the m^{th} photo-diode is given as

$$\begin{aligned} \tilde{p}_{I+,m}(t) &= l l_{\text{fb}}(m) l_{\text{voa}} \cdot a_{I+,m} p_{I+,m}(t - \tau_{I+,m}) \\ &= l_{\text{oc}} l_{\text{fb}}(m) w_{I+,m} A_0 \left(1 + \frac{1}{2}\alpha_1\alpha_2 x(t - \tau_{I+,m})\right) \cos(2\pi f_0 (t - \tau_{I+,m})), \end{aligned} \tag{3.16}$$

where $l_{\text{fb}}(m)$ and $\tau_{I+,m}$ denote the fiber length-dependent propagation loss and delay, $l_{\text{oc}} = \sqrt{\frac{1}{M}} l_{\text{voa}} l_{\text{sp}} l_{\text{eom}}$ is the fixed insertion loss of optical components. Finally, these delayed and weighted signals are converted back to the RF domain by photo-diodes through direct detection and combined together to form the output as

$$i_{I+}(t) = R_{\text{pd}} \sum_{m=1}^{M} \frac{\alpha_1 \alpha_2 \left(l_{\text{oc}} l_{\text{fb}}(m) w_{I+,m} A_o \right)^2}{2} x(t - \tau_{I+,m}) = \sum_{m=1}^{M} \varsigma_m w_{I+,m} x(t - \tau_{I+,m}), \tag{3.17}$$

where R_{pd} is the responsivity of the photo-diode; $\varsigma_{I+,m} = \frac{1}{2} R_{\text{pd}} \alpha_1 \alpha_2 l_{\text{oc}}^2 l_{\text{fb}}^2(m) A_o^2$ describes the fixed loss of the m^{th} tap of the $I+$ branch; and $w_{I+,m} = \frac{a_{I+,m}^2}{2}$ is the tunable weight. We assume the fixed loss and delay of the m^{th} taps are equal for the four branches due to identical architectures, so we omit the branch in the subscript of β and τ. The output of the four branches is combined to form the final output of the canceller, so its input-output relation can be written as

$$\begin{aligned} y_{\text{fac}}(t) &= i_{I+}(t) + i_{I-}(t) + i_{Q+}(t) + i_{Q-}(t) \\ &= \sum_{m=1}^{M} \varsigma_m ((w_{I+,m} - w_{I+,m}) + j(w_{Q+,m} - w_{Q-,m})) x(t - \tau_m), \end{aligned} \tag{3.18}$$

which describes the desired FIR filter as Equation (3.3).

3.4 Simulation results

The effectiveness of our proposed fiber-array design is firstly demonstrated by hardware simulations on the photonics simulation software *VPIphotonics*. An OFDM signal over 600 MHz is used as the transmitted SI and the SI channel is generated using the "TDL-D" channel model given in the 3GPP document [3] with 1ns of RMS delay spread. The insertion losses for the modulator, VOAs, and the 1×8 optical power splitter are 4.5 dB, 1.5 dB, and 10.7 dB, respectively. The propagation loss for the single-mode fiber is set as 0.461 dB/m as it is assumed to be coiled into a 2 cm ring for packing. The VOAs have an attenuation range from 0 dB to 30 dB with a 0.1 dB tuning step. Figure 3.8 shows the RSI channel with the 8-taps canceller compared to the original SI channel. It can be seen that the SI is efficiently suppressed over the band from 2.4 GHz to 2.6 GHz, and the additional noise and distortions introduced by the canceller are close to the AWGN of the receiver (about 5 dB higher than the receiver AWGN).

Then, we demonstrate how insertion loss compromises the cancellation performance by comparing the photonics-assisted and conventional electrical delay lines, whose insertion

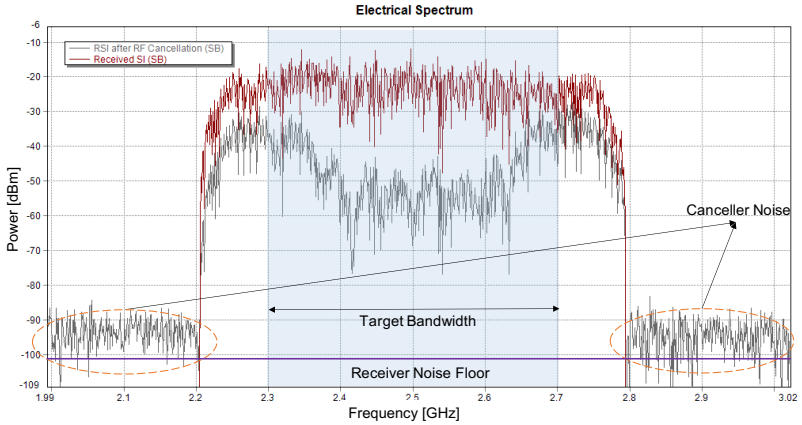
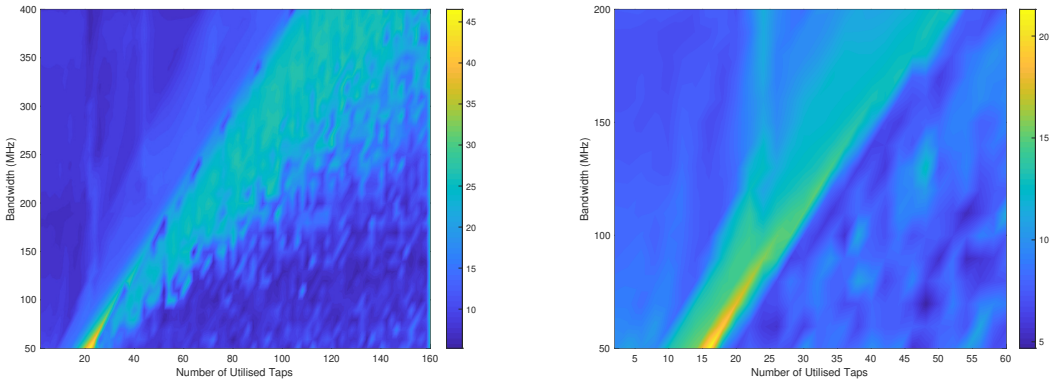


FIGURE 3.8: Cancellation results of an 8-taps fiber array canceller simulated on *VPIphotonics* software.

losses are taken from Figure 3.1[3]. For simplicity, the nonlinear effects and the noise while switching the signal between the RF and optical domains are ignored here, and we assume the perfect CSI of the SI channel is available. It can be seen that no matter how many taps are constructed using electrical components (e.g., microstrips), desired performance cannot be achieved over wide bands due to a large insertion loss penalty, which invalidates the later taps. In contrast, the photonics-assisted design can achieve efficient wideband RF cancellation, e.g., more than 30 dB within bandwidth beyond 400 MHz, since all the constructed taps are effective, enhancing the flexibility of the canceller.



(a) Phonics-assisted canceller (b) Electrical canceller

FIGURE 3.9: Cancellation depth versus the number of taps and operation bandwidth.

[3]This simulation result is also valid for the fiber array design although we use the insertion loss of the FBG-based delay lines. The reason is that the two designs share the same benefits of extremely low losses of photonic components and have similar loss properties with increasing taps, which will be explained later in Section 3.5.5

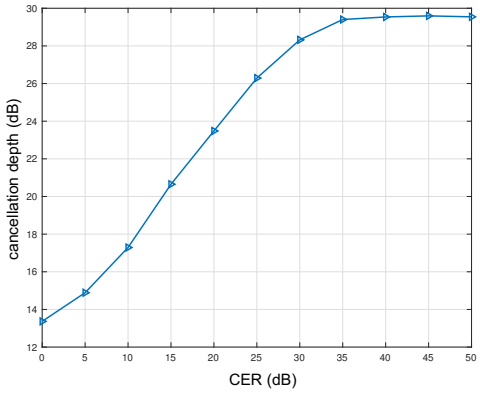


FIGURE 3.10: Performance versus practical measurement errors.

As we stated earlier, our proposed alternative low-complexity tuning method is developed based on the assumption that the four branches are almost identical and the phase shifting of the four branches is ideal. However, there will be mismatches between the four branches, and there may be phase errors in the phase shifters. Figure 3.10 shows how the mismatches and phase errors affect the cancellation performance, where the x-axis is the coefficient-to-error ratio (ECR) that describes the magnitude of errors. It can be seen that the mismatches and phase errors will degrade the performance. However, as long as the error is 35 dB smaller than the coefficients, the performance will not be significantly affected.

3.5 Experiment work

In this section, we demonstrate our experimental works of the fiber array-based canceller. Radios require hardware (e.g., modulators/demodulators, filters, mixers, amplifiers, etc.) to generate the transmitted RF signal and convert the received RF signal into the baseband domain. These hardware components have different functions to process the signal. Software-defined radio (SDR) is a configurable radio communication system that allows these functions to be implemented by software on a personal computer (PC). We use the high-performance SDR platform - "*Ettus Research USRP X310*" as the IBFD transceiver for our experiments. By adopting 5G NR protocols to generate the symbols on *MATLAB*, it can be regarded as a 5G terminal device. *LabVIEW* Communications System Design Suite offers a design environment to program processors and FPGAs of the SDR hardware. Thus, the RF signal can be generated as follows.

1) Generate the baseband modulated OFDM symbols based on 5G NR protocol on *MATLAB*.

2) Set the carrier frequency, sampling rate, Tx gain, and Tx channel on *LabVIEW* using the "niUSRP Configure Signal" module and import the generated OFDM symbols as the baseband signal.

3) The USRP will convert the baseband signal to the desired RF signal according to the configuration.

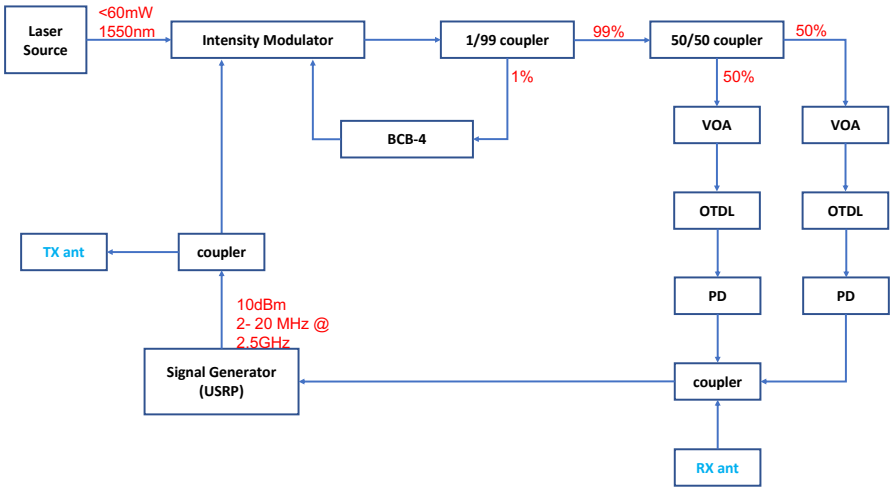
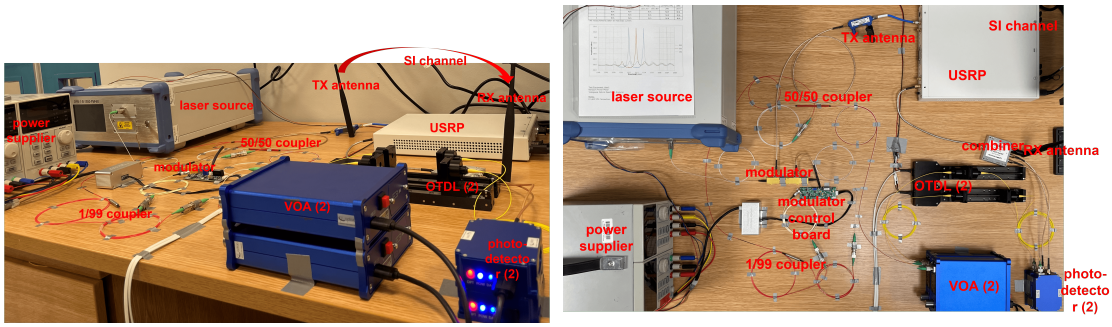


FIGURE 3.11: Block diagram of the setup for the 2-tap fiber array canceller.

For the sake of implementation cost, we use an optical tuneable delay line (OTDL) (i.e., VODL) to adjust the phase of the signal instead of the four-branch architecture. The OTDL tunes the delay of each tap by changing the propagation distance. Figure 3.11 shows the block diagram of the 2-tap canceller connection, and Figure 3.12 shows the photos of the corresponding experimental canceller demo. The key components utilized and their specifications are listed in Table 3.1.

3.5.1 Transfer function measure

An appropriate bias voltage is required to set the optical intensity modulator to work in the desired mode. To this end, the transfer function of the modulator is needed. The transfer function describes the bias voltage-optical output power relationship, which is usually a sine wave. We measure it by continuously increasing the bias voltage and recording the corresponding optical output power, as depicted in Figure 3.13. Figure



(a) Front view (b) Top view

FIGURE 3.12: Photos of the experiment work setup of the 2-tap prototype.

TABLE 3.1: Specifications of the key components utilized to construct the prototype.

COMPONENT	PRODUCT NO.	SPEC.	MFR.
Laser source	DFB-1-B-1550	1550nm, 60mW	OPTILAB
Optical intensity modulator	IM-1550-20-A	1550nm, 20GHz BW	OPTILAB
Modulator bias control board	DCB-4	Four bias mode	OPTILAB
VOA	VOA-C-M	1550nm, 38dB range, 0.02dB resolution, RS485 remote interface	OPTILAB
Photodiode	PD-30-M-S-DC	30GHz, 20mW, SMA, DC coupled	OPTILAB
OTDL	ODL-300	1550nm, SM fiber, servo motor	OZ OPTICS
Directional coupler	ZUDC6-0283-S+	6dB coupling, 2-8GHz, 27dB directivity	MINI-CIRCUITS
RF power combiner	ZN3PD-622W-S+	3 ways, 1800-6200MHz	MINI-CIRCUITS

3.14 shows the measured transfer function of the modulator with a step of 0.1 V. Setting the bias voltage at the quadrature point of the negative slope of the transfer function can inverse the phase of the RF signal without magnitude distortions, which is desired for our application.

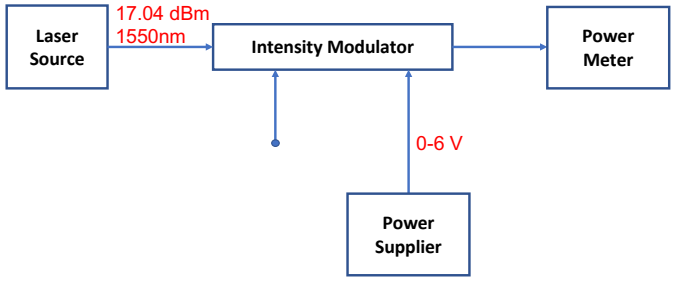


FIGURE 3.13: Block diagram of the setup for the transfer function measurement.

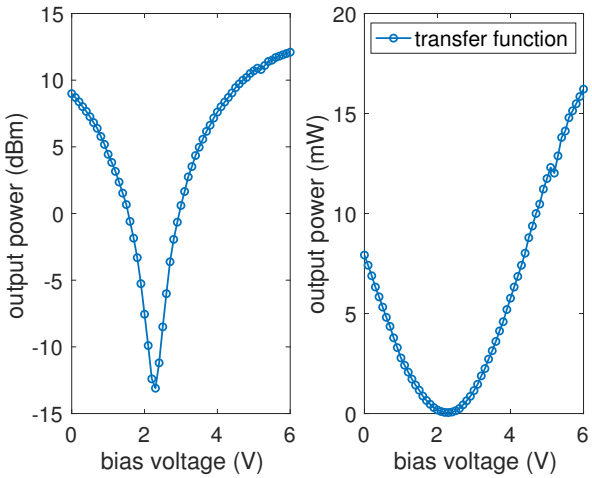


FIGURE 3.14: Measured transfer function of the optical intensity modulator.

Remarking: It should be noted that environmental changes (e.g., temperature) may affect the modulator and shift the transfer function. To maintain a stable linear operation over long periods of time, a bias voltage control board is employed in practical work.

3.5.2 Experiment setup

We adopt the 5G NR protocol for the experiments, and the parameters are listed in Table 3.2. At the receiver end, we fetch the received streams from the USRP buffer on

TABLE 3.2: Parameters for generating the Tx signal based on the 5G NR protocol.

Parameters	**Values**
Subcarrier spacing	15 kHz
Sampling rate	5 MHz
Bandwidth	2 MHz
Carrier frequency	2.5 GHz
Symbols per frame	14(1 ms)
Baseband modulation	QPSK
Cyclic prefix (CP)	"normal"
Synchronization signal	primary synchronization signal (PSS)
Pilots	channel state information reference signal (CSI-RS)
USRP Tx gain	31.5 dB

LabVIEW. Then, we process to acquire the frequency samples as follows.

1) Perform frame detection to find the boundary of a frame with the help of PSS and compensate for the carrier frequency offset (CFO).

3) Extract the 2^{nd} OFDM symbol in the frame, which consists of the nonzero CSI-RS, then remove the CP and convert it into the frequency domain.

4) Extract the received CSI-RS according to the known inserted locations in the grid.

5) Perform the LS channel estimation to obtain the channel frequency response of the Tx-Rx link[4].

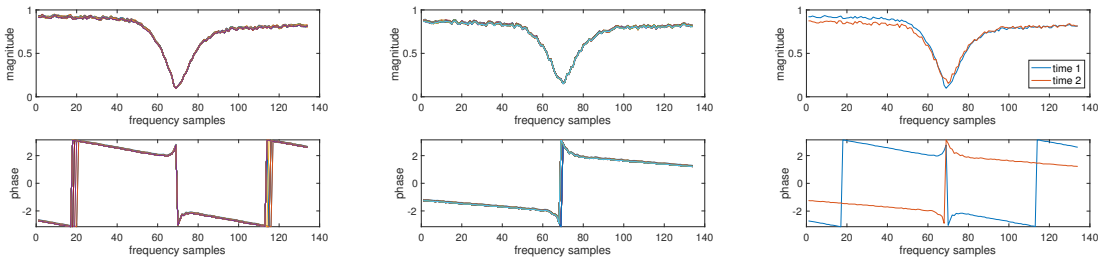
3.5.3 Canceller measuring

As illustrated earlier in Section 3.2.2, the frequency response of the canceller is required for the tuning algorithm. Although the canceller can be mathematically described by a multi-tap representation, it is challenging to acquire the exact coefficients of each tap due to hardware imperfections, e.g., phase noise, non-flat frequency response, etc. Thus, we directly measure the response of the whole canceller circuit to eliminate the mismatch between the actual response and the mathematical model. To this end, we remove the Rx antenna from the combiner connected to the USRP, so it will only receive the signal propagated through the canceller circuit. Performing the receiver processing, the response of the whole canceller circuit can be obtained.

It should be noted that we can measure one of the two taps individually by disconnecting the other one from the combiner. By doing so, we obtain the channel frequency response (CFR) of the first tap as shown in Figure 3.15, where Figure 3.15(a) and 3.15(b) show the estimated CFRs from multiple consecutive frames with an interval of 3 minutes. There is a depression in the middle of the CFR since we didn't perform "fftshift" operation. It can be seen that the CFRs of the canceller do not change significantly in a short period of time (at least within 100 OFDM frames, i.e., 0.1s). We average the hundreds of CFRs measured at "time 1" and "time 2" and compare them in Figure 3.15(c), which shows that the magnitude and shape of the CFR will not change significantly, but its phase may shift at different times.

Then, we connect the second tap and measure the 2-tap canceller with similar processing, and Figure 3.16 shows the results. It seems that the 2-tap canceller is more stable than a single tap that its CFR doesn't change over a fairly long period of time. It should

[4]That could be the wireless SI channel, RF canceller, effective SI channel with antenna isolation or RF cancellation, etc., depending on the physical connection, which will be detailed later.



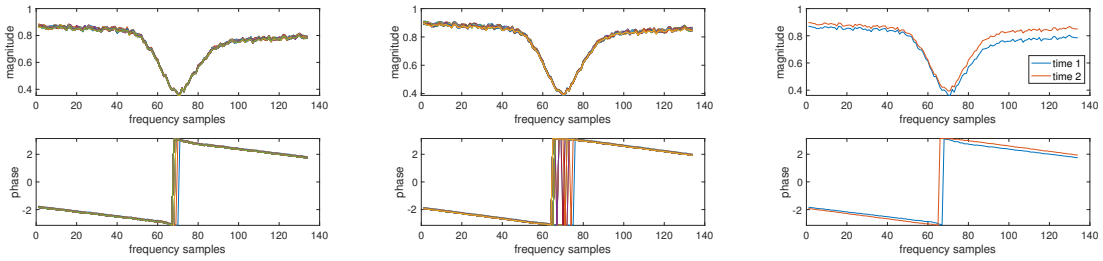
(a) Multiple measurements within the time of consecutive OFDM frames (time 1).

(b) Multiple measurements within the time of consecutive OFDM frames (time 2).

(c) Averaged measurements at time 1 and time 2.

FIGURE 3.15: Measured CFRs of the 1-tap Canceller.

be noted that this may be because the phase of both taps is changing, causing them to cancel each other out when looking at the overall response, but the phase of a single tap is actually changing.



(a) Multiple measurements within the time of consecutive OFDM frames (time 1).

(b) Multiple measurements within the time of consecutive OFDM frames (time 2).

(c) Averaged measurements at time 1 and time 2.

FIGURE 3.16: Measured CFRs of the 2-tap Canceller.

3.5.4 Cancellation results

We adopt the following processes to test the cancellation performance.

1) Measure the wireless SI channel with the canceller circuit disabled (disconnected from the combiner).

2) Calculate the optimal weight vectors and tune corresponding VOAs and OTDLs.

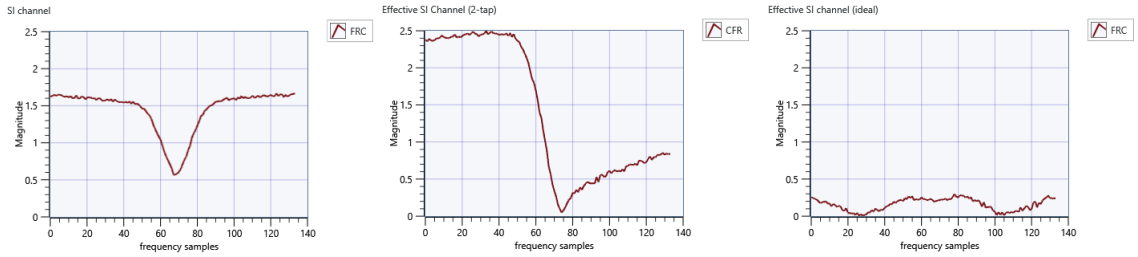
3) Connect the canceller back to the combiner and measure the effective SI channel.

With the calculated optimal complex weights, we can adjust the attenuation factors of VOAs and the delays of OTDLs of the M taps as

$$\begin{aligned} -10 \times \log_{10}\left(|[\mathbf{W}_{tdc}]_m|\right) &\rightarrow m^{th} \text{ VOA}, \\ \frac{\angle [\mathbf{W}_{tdc}]_m}{2\pi f_c} &\rightarrow m^{th} \text{ OTDL}. \end{aligned} \tag{3.19}$$

The control is implemented by sending corresponding commands to the VOAs and OTDLs through RS232 interfaces, making the canceller adaptive to deal with wireless SI channel changes.

Figure 3.17 shows the wireless SI channel and the effective SI channel after cancellation, where Figure 3.17(b) shows the real effective SI channel measured from the combiner output and Figure 3.17(c) shows the ideal effective SI channel, i.e., the best achievable result by solving the optimization problem, which could be achieved if the canceller has stable responses, unlimited tuning resolution and the canceller and wireless SI channel remain unchanged during the tuning period. It can be seen that the real effective SI channel cannot achieve the ideal case, mainly due to the long tuning time, where the phase of the canceller may have changed. The OTDL takes a relatively long time to adjust the delays since it is achieved by moving the stage by a servo motor. We can see considerable cancellation depths at half of the bands, while the SI is not suppressed over the other half bands due to hardware imperfections, e.g., phase shifting of canceller as shown in Figure 3.15(c), long tuning period, etc.

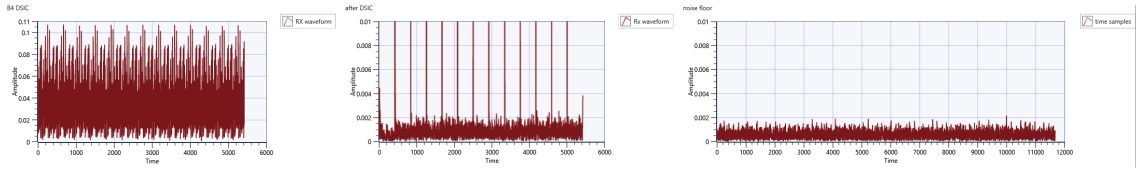


(a) Measured SI channel. (b) Measured effective SI channel. (c) Ideal effective SI channel.

FIGURE 3.17: Wireless SI channel and the effective SI channel after cancellation.

After the RF cancellation, we perform the linear digital self-interference cancellation. With the estimated effective SI channel and known Tx signal, we can easily regenerate a replica of the Rx RSI in the digital domain and cancel it out. Figure 3.18 shows the Rx waveform before and after DSIC, which shows the RSI is effectively suppressed by

DSIC. The residual signal after DSIC has already approached the receiver noise floor. The receiver noise is measured with the combiner disconnected from the USRP, i.e., the USRP does not receive any signal.



(a) Before DSIC. (b) After DSIC. (c) Receiver noise floor.

FIGURE 3.18: Rx waveform before and after digital self-interference cancellation.

Table 3.3 summarizes the experimental cancellation results, where test results are experimentally achieved, and ideal results mean the achievable results with the ideal hardware (i.e., unlimited tuning speed and stable responses), which are calculated according to the measured CFR of the canceller and wireless SI channel. With a well-designed canceller circuit, i.e., a feedback loop to stabilize the canceller and a phase control unit with a fast response speed, the ideal results should be achievable.

TABLE 3.3: Experimental RF cancellation results

Demo	**Test Results**	**Ideal Results**
1-tap canceller	1.71 dB over 1 MHz	0.89 dB over 2 MHz
2-tap canceller	8.38 dB over 1 MHz	17.89 dB over 2 MHz
digital canceller	28.81 dB over 2 MHz	N/A

To demonstrate the effects of analog SIC, we measure the dynamic range of the receiver (i.e., USRP). We connect the transmitter RF chain to the receiver RF chain directly using a short cable, whose loss could be ignored. The receiver performs the DSIC algorithm. Then, we change the analog gain of the transmitter, i.e., change the Tx power and measure the power of the RSI after the DSIC. If the Rx power (which is equal to the Tx power ignoring the cable loss) is within the dynamic range of the receiver, the RSI should approach the receiver noise floor. Otherwise, the receiver introduces large distortions, compromising the performance of the digital canceller and yielding a high RSI power. The measurement results are plotted in Figure 3.19. It can be inferred that at least 36.5 dB of ASIC depth is required to suppress the SI (with Tx power of -27.5 dBm) to be within the range of the receiver's dynamic range, i.e., minimize the receiver distortions and suppress the RSI to close the noise floor.

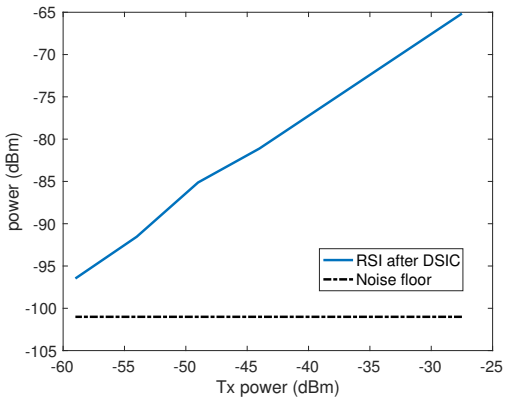


FIGURE 3.19: RSI power after DSCI versus Tx power.

Figure 3.20 shows the variation of the SI power during the SIC steps of our experiments with the receiver's noise floor and dynamic range denoted. The SI is firstly passively mitigated in the propagation domain by the physical separation (i.e., increase the Tx/Rx antenna distance and place a metal plate between Tx/Rx antennas), suppressing the SI from -27.5 dBm to -59.29 dBm. Then, a 2-tap fiber array canceller can provide 8.38 dB of cancellation depth within 1 MHz bandwidth in the RF domain, bringing the SI to the receiver's dynamic range. Finally, the digital canceller processes the RSI and brings it close to the receiver's noise floor. It is still around 4 dB higher than the noise floor since the fiber array canceller will introduce additional noise, slightly degrading the DSIC performance. Nevertheless, we successfully demonstrated effective SIC through experiments and real-world SI channels.

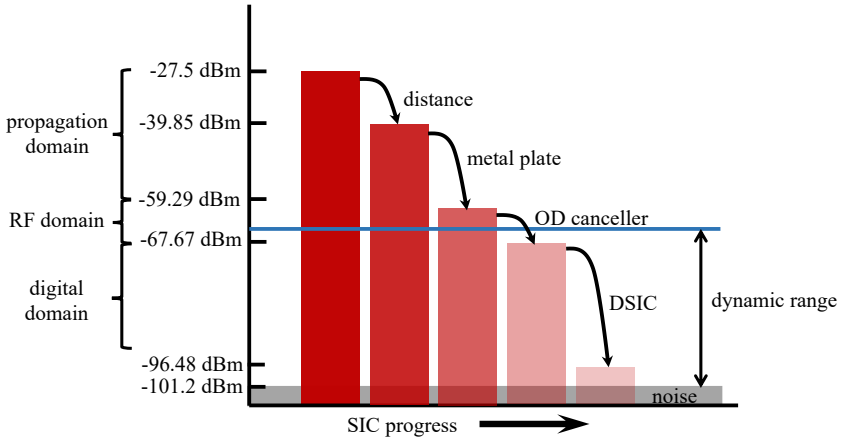


FIGURE 3.20: RSI power variation during the 3-step SIC experiment (OD: optical domain, i.e., fiber array canceller).

3.5.5 Extension to more taps

In this section, we investigate the ability of the fiber array-based canceller to increase the number of taps. As illustrated above, the main limiting factor of constructing a large number of taps is the increasing insertion loss with taps. We measure the output power at different locations of the demo, as Figure 3.21 shows. The measurement shows that splitting the modulated signal into two branches only introduces 3.06 dB of insertion loss due to splitting the modulated signal, while the extremely low insertion loss of the optical components has trivial effects. Splitting the signal in half itself halves the signal energy, i.e., introduce $10 \times \log_{10}(2) = 3.01$ dB of loss, while the components only cause $0.05(= 5.07 - 3.01 - 2.01)$ dB of insertion loss. Therefore, we can infer that using high-performance optical components to create multiple taps would increase the insertion loss of a single tap by $10 \times \log_{10}(M)$ dB due to splitting the signal, ignoring the trivial loss of components.

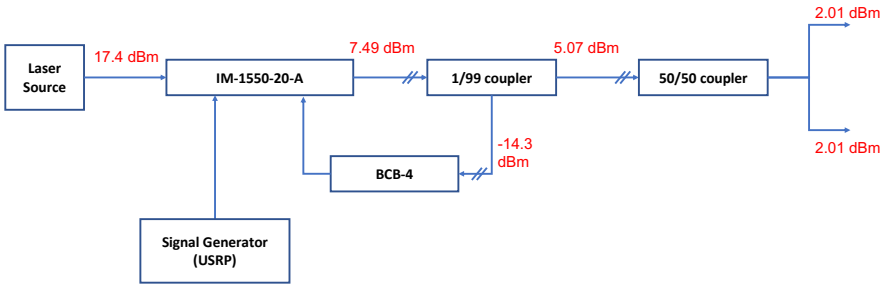


FIGURE 3.21: The output power measured at different locations of the canceller prototype.

It seems that the fiber-array canceller has a much higher insertion loss than the FBG-based design, whose insertion loss is almost constant with taps, as shown in Figure 3.1. It should be noted, however, that FBG-based design requires M laser sources to generate M optical carriers, thus avoiding splitting the modulated optical signal, yielding the constant insertion loss caused by optical components only. That suggests the FBG-based design has M times more power consumption compared to the fiber-array design. In fact, if the fiber-array design uses an optical carrier of M times the power, it can also compensate for the loss due to splitting the signal, leaving the constant loss caused by optical components. This indicates that the two designs are essentially the same in terms of increasing the number of taps.

3.6 Conclusions

In this chapter, we studied a photonics-assisted RF canceller design motivated by the enlarging operation bands of future wireless networks. The large insertion loss of electrical components limits the flexibility of multi-tap cancellers so that electrical delay lines cannot support wideband operations. Therefore, photonic delay lines are motivated due to their extremely low loss and flat frequency response at high frequencies. The photonics-assisted canceller converts the RF signal into the optical domain to process, where it is challenging to tune the phase of the RF signal. Therefore, a variant of the multi-tap canceller, which consists of four branches, is usually utilized, and we studied the tuning algorithm for this architecture, proposing a new low-complexity tuning method. Simulations demonstrated that photonics-assisted cancellers could provide effective wideband cancellation due to their nearly constant insertion loss. We first introduced the reported FBG-based design. Then, we designed and analyzed a fiber array-based canceller. A set of mathematical derivations of the canceller's processing proves it can form the desired FIR filter as the typical multi-tap canceller, and hardware simulations on *VPIphtonics* verified its effectiveness. We further verified the 3-step SIC scheme in the real-world environment by building up an experimental demonstration that consists of antenna isolation, a 2-tap fiber array-based canceller, and a linear digital canceller. Our canceller can be adaptively tuned to match the real wireless SI channel changes, and we showed successful SIC using 5G NR protocol signals, where the SI is suppressed close to the receiver's noise floor. In contrast, existing studies usually utilize a simple single-frequency signal (e.g., sine wave) or use a channel emulator instead of the real-time-varying channel. Furthermore, we illustrated that our design has the ability to construct up to 100 taps by analyzing the insertion loss.

Chapter 4

Beamforming Schemes for IBFD Multi-cell Multi-user Networks

4.1 Introduction

MIMO systems have been employed in practical networks by equipping antenna arrays at base stations and user terminals to meet the growth of users and increase data-rate demands. In MIMO systems, beamforming is a powerful tool for transmitting or receiving the signal in a spatially selective manner, providing reliable communication links in the presence of interference and noise, which have been proved by massive studies. To improve spectral efficiency and reduce latency, it is considered to employ IBFD base stations in future cellular networks, i.e., multi-cell multi-user (MCMU) networks. IBFD operations in cellular networks introduce additional CCI, where uplink transmissions interfere with downlink users, in addition to the SI at base stations. Therefore, the interference is much more complex in the IBFD cellular networks, making beamforming design more challenging.

It is revealed in [48] that conventional adaptive beamforming techniques are not robust against the mismatch in the desired signal steering vector. In contrast, convex optimization could provide effective and robust beamforming designs through rigorous formulation and solution. Based on this idea, many studies have developed effective

Work in this chapter has been published in IEEE Transactions on Vehicular Technology March 2023 [47] and submitted to IEEE Transactions on Vehicular Technology March 2023, with a preliminary discussion given in China Communications October 2021 [10].

beamforming designs to improve the spectral or energy efficiency of various networks, where a complex design problem is formulated and recast into tractable convex forms to be solved. Practical wireless networks usually employ linear beamforming techniques such as ZF, MRTC, and MMSE beamforming schemes. In [49], the performance of ZF and MRTC beamforming schemes are evaluated and compared to the optimal scheme for sum rate maximization in the IBFD single-cell multi-user network, demonstrating the IBFD gain and revealing that increasing the number of antennas at IBFD nodes can improve the IBFD gain. It is illustrated in [50] that a ZF/MRT processing can guarantee a balance between sum rate maximization and maintaining the fairness between uplink and downlink rate and improve the performance of the IBFD cloud radio access network. The performance of the ZF precoding and maximum ratio transmission (MRT) precoding is analyzed in MCMU networks in [51] and [52], respectively. Existing studies only perform beamforming at IBFD nodes or consider the single-cell scenario, so they are not compatible with future cellular networks, i.e., MCMU networks with multi-antenna users.

In contrast, many studies have extended the MMSE beamforming to be compatible with IBFD-MCMU MIMO networks. Authors in [53] propose a beamforming scheme by minimizing the sum of mean-squared errors (MSE) of an IBFD multi-user network and show a significant performance improvement over HD. In [54], the IBFD-MCMU network is further studied, and a weighted sum rate (WSR) maximization beamforming design is proposed with the power constraints by exploiting the relationship between WSR and weighted MMSE. For massive MIMO systems, an MMSE-based hybrid beamforming scheme is proposed in [38] to achieve the IBFD gain with reduced RF chains. These studies have demonstrated the considerable performance of MMSE-based beamforming schemes for IBFD-MCMU networks. However, they all assume effective SIC has been realized in both the analog and digital domains.

As illustrated earlier in Section 2.6, implementing the subtractive SIC techniques in MIMO systems has extremely high complexity and cost, which could be physically prohibited in massive MIMO systems. Therefore, it is strongly motivated to explore the spatial DoFs provided by antenna arrays for SIC. A ZF-based beamforming design is proposed in [55], which maximizes the target channel while limiting the SI channel strength to zero. The design eliminates the SI thanks to the ZF constraint. An orthogonal RF

beamformer design is proposed in [56] to enhance the SIC quality by suppressing especially the reflection path coupling, which is realized by designing orthogonal beams through the intended direction. The SI channel strength is reduced by up to 110 dB via the RF beamforming with 80 dB of antenna isolation, where the no-orthogonal beamformer provides 18.04-28.76 dB of SIC. A similar orthogonal beamformer design is also proposed in [57], but it derives the orthogonal beamformer first, followed by decomposing it into digital and RF beamformers. Many studies also investigate null-space projection (NSP) to suppress SI in the propagation domain [58]. The NSP steers the transmitted beams to the direction orthogonal to the SI channel, minimizing the SI coupling but distorting the original precoders and introducing precoding errors. The trade-off between the precoding errors and SIC capability of the NSP-based methods is analyzed in [59]. It illustrates that achieving effective SIC may severely sacrifice the downlink capacity. Furthermore, the authors in [60] compare the performance of the two SIC methods in massive MIMO systems, i.e., the typical subtractive method and NSP-based spatial suppression, by considering the channel estimation error and spatial correlation of the SI channel. They give the desired SIC techniques under given system conditions according to the ergodic sum rate. However, the limited dynamic range of receivers and implementation complexity are not considered, and they require extra DoF (i.e., sufficient transmitting antennas) for the NSP.

In addition to the beamforming, an appropriate power allocation policy can also manage the interference since the strength of interference depends on the transmit power of associated nodes. In MIMO systems, power allocation and beamforming should be crossly designed for optimal performance since they will affect each other. Studies illustrate that an appropriate joint power allocation and beamforming (JPABF) scheme can significantly improve the system capacity in MIMO systems. A sum rate maximization problem with respect to joint power allocation and beamforming is formulated for a single-cell 2-user network in [61]. It is solved using an iterative algorithm, and simulation results show that it achieves close-to-bound sum rate performance. The main challenge of the JPABF design is that the joint optimization problem is non-convex, so it is difficult to derive closed-form solutions [62]. The JPABF scheme has not yet been intensively studied for IBFD-MCMU networks, where a large amount of CCI and significant SI pose even greater challenges.

Despite the abundance of research on beamforming, detailed evaluations of the performance and complexity of beamforming for IBFD-MCMU networks are still lacking. The main contribution of our work is the provided insights for practical IBFD cellular networks. We derive a generalized form of the common three linear beamforming schemes and compare them comprehensively in terms of performance and complexity. In order to achieve a low-complexity implementation of IBFD base stations, we study the transmitting beamforming for SIC and propose an MMSE-based design that is compatible with MCMU networks and does not have requirements on the antenna array size. In addition, we crossly consider the power allocation and beamforming to maximize the network's capacity, deriving an iterative 2-step JPABF algorithm, which is called as joint power allocation and interference management (JPAIM) algorithm.

The rest of this chapter is organized as follows. In Section 4.2, we give preliminaries for the beamforming design, including the transmitted and received signal models, channel models with channel uncertainty, and performance metrics, i.e., achievable sum rate. Then, we derive the enhanced ZF and MRTC beamforming schemes in Section 4.3, followed by the beamforming cancellation discussions with the complexity comparisons in Section 4.4. Then, we further consider the power allocation policy in Section 4.5 and derive the JPABF design. Finally, 3GPP-related simulation settings are given in Section 4.6 and numerical results are demonstrated accordingly, and conclusions are drawn in Section 4.7.

4.2 Preliminaries

4.2.1 System model

We consider a G-cell network, where the base station in the g^{th} cell serves K_g^d downlink (DL) users and K_g^u uplink (UL) users. Assume all base stations (BS) have N_{bs} transmitting antennas and M_{bs} receiving antennas, while UL user equipment (UE) has N_{ue} transmitting antennas and DL UE has M_{ue} receiving antennas. Fig. 4.1 shows the interference between the nodes in such an IBFD-MCMU network, where the additional interference due to IBFD operation is highlighted in red.

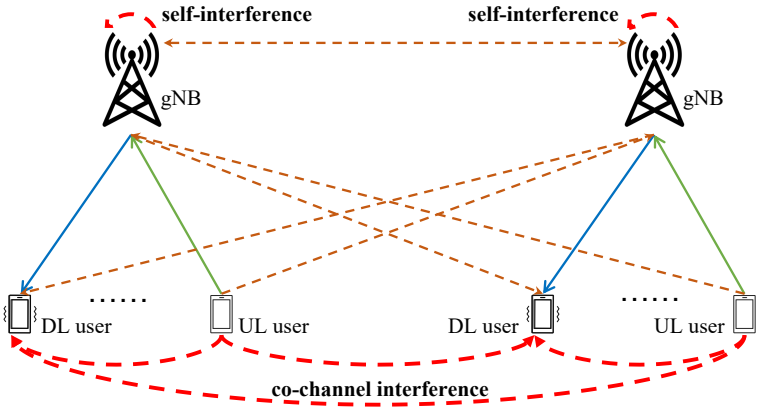


FIGURE 4.1: Interference between nodes in an IBFD multi-cell multi-user network.

4.2.2 Transmitted signals

Let $\mathbf{s}_{k_g^d} \in \mathbb{C}^{b_d \times 1}$ represent the data symbols intended to the k^{th} downlink user in the g^{th} cell with statistics $\mathbb{E}\left\{\mathbf{s}_{k_g^d}\mathbf{s}_{k_g^d}^{\dagger}\right\} = \mathbf{I}_{b_d}$; $\mathbf{V}_{k_g^d} \in \mathbb{C}^{N_{bs} \times b_d}$ denotes the associated precoding matrix; and $\alpha_{k_g^d}$ is the power coefficient that reflects the transmit power allocated to the corresponding user. The signal transmitted at the g^{th} BS can be denoted as

$$\mathbf{x}_g = \sum_{k=1}^{K_g^d} \left(\alpha_{k_g^d}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d} + \mathbf{d}_{tx,k_g^d}\right), \tag{4.1}$$

where $\mathbf{d}_{tx,k_g^d}$ represents the hardware impairments due to the limited dynamic range of practical transmitters. The limited dynamic range is a natural consequence of imperfect DACs, oscillators, and PAs. Experimental measurements demonstrate that the transmitter hardware impairments are independent of the transmitted signals and can be closely described by the circular complex Gaussian model as [32]

$$\mathbf{d}_{tx,g} \sim \mathcal{CN}\left(\mathbf{0}, \kappa_{bs}\mathbb{E}\left\{\mathcal{D}\left(\mathbf{V}_g\mathbf{A}_g\mathbf{A}_g^{\dagger}\mathbf{V}_g^{\dagger}\right)\right\}\right), \tag{4.2}$$

where $\kappa_{bs} \ll 1$ characterizes the dynamic range of the transmitters at base stations. The base station transmits an accumulation of signals for DL users within the g^{th} cell in nature, so the intended transmitted signal, i.e., without transmitter distortions, can be written as $\tilde{\mathbf{x}}_g = \mathbf{V}_g\mathbf{A}_g\mathbf{s}_g$, where $\mathbf{s}_g = \left[\mathbf{s}_{1_g^d}^T, \mathbf{s}_{2_g^d}^T, \ldots, \mathbf{s}_{K_g^d}^T\right]^T$, $\mathbf{A}_g = \mathcal{D}\big(\alpha_{1_g^d}\mathbf{I}_{b_d}, \alpha_{2_g^d}\mathbf{I}_{b_d}, \ldots, \alpha_{K_g^d}\mathbf{I}_{b_d}\big)$, and $\mathbf{V}_g = \left[\mathbf{V}_{1_g^d}, \mathbf{V}_{2_g^d}, \ldots, \mathbf{V}_{K_g^d}\right]$. The transmit power of the g^{th} base station is given as

$$\begin{aligned} P_g = \mathbb{E}\left\{tr\left(\tilde{\mathbf{x}}_g\tilde{\mathbf{x}}_g^{\dagger}\right)\right\} &= tr\left(\mathbb{E}\left\{\mathbf{V}_g\mathbf{A}_g\mathbf{s}_g\mathbf{s}_g^{\dagger}\mathbf{A}_g^{\dagger}\mathbf{V}_g^{\dagger}\right\}\right) \\ &= \sum_{k=1}^{K_g^d}\alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right). \end{aligned} \tag{4.3}$$

Similarly, let $\mathbf{s}_{k_g^u} \in \mathbb{C}^{b_u \times 1}$ denote the uplink payload symbols with statistics $\mathbb{E}\left\{\mathbf{s}_{k_g^u}\mathbf{s}_{k_g^u}^{\dagger}\right\} = \mathbf{I}_{b_u}$; $\mathbf{V}_{k_g^u}$ represents the associated precoding matrix; $\gamma_{k_g^u}$ represents the power coefficient that reflects the allocated transmit power; and $\kappa_{ue} \ll 1$ characterizes the dynamic range of the transmitters at the user equipment. The transmitted signal of the k^{th} uplink user in the g^{th} cell (i.e., k_g^u) can be denoted as

$$\mathbf{x}_{k_g^u} = \underbrace{\gamma_{k_g^u}\mathbf{V}_{k_g^u}\mathbf{s}_{k_g^u}}_{\tilde{\mathbf{x}}_{k_g^u}} + \mathbf{d}_{tx,k_g^u}, \tag{4.4}$$

where $\mathbf{d}_{tx,k_g^u}$ denotes the transmitter hardware impairments at this uplink user that can be described as

$$\mathbf{d}_{tx,k_g^u} \sim \mathcal{CN}\left(\mathbf{0}, \kappa_{ue}\gamma_{k_g^u}^2\mathbb{E}\left\{\mathcal{D}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right)\right\}\right). \tag{4.5}$$

The transmit power of the k^{th} UE in the g^{th} cell is given as

$$P_{k_g^u} = \mathbb{E}\left\{tr\left(\tilde{\mathbf{x}}_{k_g^u}\tilde{\mathbf{x}}_{k_g^u}^{\dagger}\right)\right\} = \gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right). \tag{4.6}$$

The covariance matrices of the transmitted signals can be written as

$$\begin{aligned} \mathbf{T}_g = \mathbb{E}\left\{\mathbf{x}_g\mathbf{x}_g^{\dagger}\right\} &= \sum_{k=1}^{K_g^d}\underbrace{\alpha_{k_g^d}^2\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H + \kappa_{bs}\mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H\right)\right)}_{\mathbf{T}_{k_g^d}} \\ &= \alpha_{k_g^d}^2\left(\mathbf{V}_g\mathbf{V}_g^{\dagger} + \kappa_{bs}\mathcal{D}\left(\mathbf{V}_g\mathbf{V}_g^{\dagger}\right)\right), \\ \mathbf{T}_{k_g^u} = \mathbb{E}\left\{\mathbf{x}_{k_g^u}\mathbf{x}_{k_g^u}^{\dagger}\right\} &= \gamma_{k_g^u}^2\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger} + \kappa_{ue}\mathcal{D}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right)\right). \end{aligned} \tag{4.7}$$

Remarking: For the full transmit power case, the transmitted signals can be represented with $\alpha_{k_g^d} = \sqrt{\frac{P_{bs}}{K_g^d}}$ assuming average power allocation and $\gamma_{k_g^u} = \sqrt{P_{ue}}$. It should be noted that the norm of the precoders must be constrained to satisfy with the transmit power conditions with fixed coefficients $\alpha_{k_g^d}$ and $\gamma_{k_g^u}$ such that $\left\|\mathbf{V}_{k_g^d}\right\|_F^2 = \left\|\mathbf{V}_{k_g^u}\right\|_F^2 = 1,\ \forall\ k, g.$

4.2.3 Received signals

Let $\mathbf{H}_{r,t}$ represent the coefficients matrix of the wireless MIMO channel from transmitting node t to receiving node r throughout this chapter. The signal received by the g^{th} BS can be denoted as

$$\begin{aligned}\mathbf{y}_g &= \underbrace{\sum_{i=1}^{G}\sum_{j=1}^{K_i^u}\mathbf{H}_{g,i_j^u}\mathbf{x}_{i_j^u} + \mathbf{H}_{g,g}\mathbf{x}_g + \sum_{j\neq g}^{G}\mathbf{H}_{g,j}\mathbf{x}_j}_{\tilde{\mathbf{y}}_g} + \mathbf{d}_{rx,g} + \mathbf{n}_g \\ &= \gamma_{k_g^u}\mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u}\mathbf{s}_{k_g^u} \\ &\quad + \underbrace{\mathbf{H}_{g,k_g^u}\mathbf{d}_{tx,k_g^u} + \sum_{(i,j)\neq(k,g)}^{G}\sum^{K_i^u}\mathbf{H}_{g,i_j^u}\mathbf{x}_{i_j^u} + \mathbf{H}_{g,g}\mathbf{x}_g + \sum_{j\neq g}^{G}\mathbf{H}_{g,j}\mathbf{x}_j + \mathbf{d}_{rx,g} + \mathbf{n}_g}_{\tilde{\mathbf{z}}_{k_g^d}},\end{aligned} \tag{4.8}$$

where $\tilde{\mathbf{y}}_g$ denotes the sum of received signals and $\tilde{\mathbf{z}}_{k_g^d}$ is the interference plus noise for the intended uplink payload from the user k_g^u; $\mathbf{n}_g$ denotes the additive white Gaussian noise (AWGN) of the receiver such that $\mathbf{n}_g \sim \mathcal{CN}\left(\mathbf{0}, \sigma_g^2\mathbf{I}\right)$ with σ_g^2 representing the AWGN power; and $\mathbf{d}_{rx,g}$ denotes the hardware impairments due to the limited dynamic range of practical receivers. The limited dynamic range is a natural consequence of imperfect low-noise amplifiers (LNAs), oscillators, and ADCs. Experimental measurements demonstrate that the receiver hardware impairments are independent of the received signals and can be closely described by the circular complex Gaussian model as [32]

$$\mathbf{d}_{rx,g} \sim \mathcal{CN}\left(\mathbf{0}, \beta_{bs}\mathbb{E}\left\{\mathcal{D}\left(\tilde{\mathbf{y}}_g\tilde{\mathbf{y}}_g^{\dagger}\right)\right\}\right), \tag{4.9}$$

where $\beta_{bs} \ll 1$ characterizes the dynamic range of the receivers at base stations. Similarly, the signal received by the k^{th} downlink UE in the g^{th} cell can be denoted as

$$\begin{aligned}\mathbf{y}_{k_g^d} &= \underbrace{\sum_{i=1}^{G}\sum_{j=1}^{K_i^d}\mathbf{H}_{k_g^d,j}\mathbf{x}_{i_j^d} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathbf{H}_{k_g^d,i_j^u}\mathbf{x}_{i_j^u}}_{\tilde{\mathbf{y}}_{k_g^d}} + \mathbf{d}_{rx,k_g^d} + \mathbf{n}_{k_g^d} \\ &= \alpha_{k_g^d}\mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d} \\ &\quad + \underbrace{\mathbf{H}_{k_g^d,g}\mathbf{d}_{tx,k_g^d} + \sum_{(i,j)\neq(k,g)}^{G}\sum^{K_i^d}\mathbf{H}_{k_g^d,j}\mathbf{x}_{i_j^d} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathbf{H}_{k_g^d,i_j^u}\mathbf{x}_{i_j^u} + \mathbf{d}_{rx,k_g^d} + \mathbf{n}_{k_g^d}}_{\tilde{\mathbf{z}}_{k_g^d}},\end{aligned} \tag{4.10}$$

where $\tilde{\mathbf{y}}_{k_g^d}$ denotes the sum of received signals and $\tilde{\mathbf{z}}_{k_g^d}$ is the interference plus noise for the intended downlink payload at the user k_g^d; $\mathbf{n}_{k_g^d}$ denotes the AWGN of the UE receiver such that $\mathbf{n}_{k_g^d} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_{k_g^d}^2 \mathbf{I}\right)$ with $\sigma_{k_g^d}^2$ representing corresponding AWGN power; and $\mathbf{d}_{k_g^d}$ denotes the hardware impairments of the UE receiver described as

$$\mathbf{d}_{rx,k_g^d} \sim \mathcal{CN}\left(\mathbf{0}, \beta_{ue}\mathbb{E}\left\{\mathcal{D}\left(\tilde{\mathbf{y}}_{k_g^d}\tilde{\mathbf{y}}_{k_g^d}^{\dagger}\right)\right\}\right), \tag{4.11}$$

where $\beta_{ue} \ll 1$ characterizes the dynamic range of the receivers at the user equipment.

Remarking: The values of κ_{bs}, κ_{ue}, β_{bs}, and β_{ue} are related to the measurable error vector magnitudes (EVMs) of corresponding RF transceivers. The HWIs model utilized is a verified model based on experiments [32] and has been adopted by many studies in the field of wireless communications (see [54] and references therein). Due to the power consumption problem in MIMO systems, the transceivers tend to use low-resolution DACs/ADCs. So we assume that the dynamic range of receivers is mainly limited by DACs/ADCs, which can be described by the additive quantization noise model (AQNM) given in [63].

4.2.4 Channel model

4.2.4.1 Wireless channels

In the multi-cell system, there are Rician and Rayleigh channels depending on the presence of LOS components, whose models are given as follows (each matrix $\mathbf{H}$ contains i.i.d zero-mean circularly symmetric complex Gaussian random variables of variance 0.5 per dimension).

- Rician channels: if there are LOS paths between the transmitter and receiver, the channel can be modeled by the Rician channel as [64]

$$\mathbf{H}_{r,t} = \sqrt{\varrho_{r,t}}\left(\sqrt{\frac{\kappa_{r,t}}{\kappa_{r,t}+1}}\mathbf{H}_{r,t}^{LOS} + \sqrt{\frac{1}{\kappa_{r,t}+1}}\mathbf{H}\right), \tag{4.12}$$

  where $\varrho_{r,t}$ is the pathloss between the r^{th} receiving node and the t^{th} transmitting node (the receiving and transmitting nodes could be any BS or UE); $\kappa_{r,t}$ is the associated Rician factor; $\mathbf{H}_{r,t}^{LOS}$ denotes the LOS component such that $\mathbf{H}_{r,t}^{LOS} =$

$\mathbf{a}_{M_r}^H(\theta_{\text{AOA},1})\mathbf{a}_{N_t}(\theta_{\text{AOD},1})$, where $\mathbf{a}_N(\theta) = \left[1, e^{j2\pi\frac{d}{\lambda_s}\sin\theta}, \cdots, e^{j2\pi\frac{d}{\lambda_s}(N-1)\sin\theta}\right]$ with θ, d, and λ_s representing the angle of departure (AoD) or angle of arrival (AoA), the distance between neighbouring antennas, and signal wavelength, respectively.

- Rayleigh channels: if there are no LOS paths between the transmitter and receiver, the channel can be modeled by the Rayleigh channel, which has i.i.d. elements as

$$\mathbf{H}_{r,t} = \sqrt{\varrho_{r,t}}\,\mathbf{H}. \tag{4.13}$$

4.2.4.2 Channel uncertainty

Accurate channel estimation is challenging in practice due to limited training resources, resulting in channel uncertainty in the obtained channel state information (CSI). Let $\hat{\mathbf{H}}_{r,t}$ denote the estimate associated with the actual wireless channel $\mathbf{H}_{r,t}$, they are related as [65]

$$\mathbf{H}_{r,t} = \hat{\mathbf{H}}_{r,t} + \mathbf{\Delta}_{r,t}, \tag{4.14}$$

where $\mathbf{\Delta}_{r,t}$ denotes the channel uncertainty (i.e., estimation errors). We adapt the stochastic error model, which describes the channel uncertainty as $\mathbf{\Delta}_{r,t} \sim \mathcal{CN}\left(\mathbf{0}, \tilde{\sigma}_{r,t}^2\mathbf{I}\right)$ [54, 65]. Using the accessible estimated CSI with the statistical channel uncertainty, the received signals given in Equations (4.8) and (4.10) can be written as

$$\mathbf{y}_g = \gamma_{k_g^u}\hat{\mathbf{H}}_{g,k_g^u}\mathbf{V}_{k_g^u}\mathbf{s}_{k_g^u} + \mathbf{z}_{k_g^u} + \mathbf{e}_g, \tag{4.15}$$

$$\mathbf{y}_{k_g^d} = \alpha_{k_g^d}\hat{\mathbf{H}}_{k_g^d,g}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d} + \mathbf{z}_{k_g^d} + \mathbf{e}_{k_g^d}, \tag{4.16}$$

where $\mathbf{z}_{k_g^u}$ and $\mathbf{z}_{k_g^d}$ are the variants of $\tilde{\mathbf{z}}_{k_g^u}$ and $\tilde{\mathbf{z}}_{k_g^d}$ with the actual channel matrices replaced by their estimates as[1]

$$\mathbf{z}_{k_g^u} = \hat{\mathbf{H}}_{g,k_g^u}\mathbf{d}_{tx,k_g^u} + \sum_{(i,j)\neq(k,g)}^{G}\sum^{K_i^u}\hat{\mathbf{H}}_{g,i_j^u}\mathbf{x}_{i_j^u} + \mathbf{H}_{g,g}\mathbf{x}_g + \sum_{j\neq g}^{G}\hat{\mathbf{H}}_{g,j}\mathbf{x}_j + \mathbf{d}_{rx,g} + \mathbf{n}_g, \tag{4.17}$$

$$\mathbf{z}_{k_g^d} = \hat{\mathbf{H}}_{k_g^d,g}\mathbf{d}_{tx,k_g^d} + \sum_{(i,j)\neq(k,g)}^{G}\sum^{K_i^d}\hat{\mathbf{H}}_{k_g^d,j}\mathbf{x}_{i_j^d} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}\mathbf{x}_{i_j^u} + \mathbf{d}_{rx,k_g^d} + \mathbf{n}_{k_g^d}, \tag{4.18}$$

[1] We use the perfect CSI of SI channels since it can be estimated with high SINR.

and $\mathbf{e}_g$ and $\mathbf{e}_{k_g^d}$ represent the noise due to imperfect CSI that can be represented as

$$\mathbf{e}_g = \sum_{i=1}^{G}\sum_{j=1}^{K_i^u} \mathbf{\Delta}_{g,i_j^u}\mathbf{x}_{i_j^u} + \sum_{j\neq g}^{G} \mathbf{\Delta}_{g,j}\mathbf{x}_j, \tag{4.19}$$

$$\mathbf{e}_{k_g^d} = \sum_{i=1}^{G}\sum_{j=1}^{K_i^d} \mathbf{\Delta}_{k_g^d,j}\mathbf{x}_{i_j^d} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \mathbf{\Delta}_{k_g^d,i_j^u}\mathbf{x}_{i_j^u}. \tag{4.20}$$

These errors can be described by the complex Gaussian model as $\mathbf{e}_g \sim \mathcal{CN}\left(\mathbf{0}, \hat{\sigma}_g^2\mathbf{I}_{M_{bs}}\right)$ and $\mathbf{e}_{k_g^d} \sim \mathcal{CN}\left(\mathbf{0}, \hat{\sigma}_{k_g^d}^2\mathbf{I}_{M_{ue}}\right)$ with $\hat{\sigma}_g^2$ and $\hat{\sigma}_{k_g^d}^2$ given as

$$\hat{\sigma}_g^2 = \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \tilde{\sigma}_{g,i_j^u}^2 tr\left(\mathbf{T}_{i_j^u}\right) + \sum_{j\neq g}^{G} \tilde{\sigma}_{g,j}^2 tr\left(\mathbf{T}_j\right), \tag{4.21}$$

$$\hat{\sigma}_{k_g^d}^2 = \sum_{j=1}^{G}\sum_{k=1}^{K_j^d} \tilde{\sigma}_{k_g^d,j}^2 tr\left(\mathbf{T}_{i_j^d}\right) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \tilde{\sigma}_{k_g^d,i_j^u}^2 tr\left(\mathbf{T}_{i_j^u}\right). \tag{4.22}$$

The proof is given in Appendix 4.8.

4.2.5 Achievable rate

The achievable sum rate is used as the metric to evaluate the performance throughout this chapter. With the antenna array, receivers can perform combining (i.e., receiving beamforming) to enhance the SINR for a higher achievable sum rate. The achievable downlink rate of the k^{th} downlink user in the g^{th} cell is given as

$$R_{k_g^d} = \log_2\left|\mathbf{I}_{b_d} + \mathbf{U}_{k_g^d}^{\dagger}\mathbf{C}_{k_g^d,S}\mathbf{U}_{k_g^d}\left(\mathbf{U}_{k_g^d}^{\dagger}\mathbf{C}_{k_g^d,X}\mathbf{U}_{k_g^d}\right)^{-1}\right|, \tag{4.23}$$

where $\mathbf{U}_{k_g^d} \in \mathbb{C}^{M_{ue}\times b_d}$ is the receiving beamforming matrix, $\mathbf{C}_{k_g^d,S}$ denotes the covariance matrix of the signal of interest, and $\mathbf{C}_{k_g^d,X}$ represents the covariance matrix of the interference plus noise terms, which are given as

$$\mathbf{C}_{k_g^d,S} = Cov\left(\alpha_{k_g^d}\hat{\mathbf{H}}_{k_g^d,g}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d}\right) = \alpha_{k_g^d}^2\hat{\mathbf{H}}_{k_g^d,g}\mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)\hat{\mathbf{H}}_{k_g^d,g}^{\dagger}, \tag{4.24}$$

$$\mathbf{C}_{k_g^d,X} = Cov\left(\mathbf{z}_{k_g^d} + \mathbf{e}_{k_g^d}\right) \overset{(a)}{=} \mathbf{C}_{k_g^d} + \beta_{bs}\cdot\mathcal{D}\left(\mathbf{C}_{k_g^d}\right) - \mathbf{C}_{k_g^d,S}, \tag{4.25}$$

where (a) uses the covariance matrix of the received signal $\mathbf{y}_{k_g^d}$ derived in Appendix 4.9.

Similarly, the achievable uplink rate of the k^{th} uplink user in the g^{th} cell can be given as

$$R_{k_g^u} = \log_2 \left| \mathbf{I}_{b_u K_g^u} + \mathbf{U}_{k_g^u}^{\dagger} \mathbf{C}_{k_g^u,S} \mathbf{U}_{k_g^u} \left(\mathbf{U}_{k_g^u}^{\dagger} \mathbf{C}_{k_g^u,X} \mathbf{U}_{k_g^u} \right)^{-1} \right|, \tag{4.26}$$

where $\mathbf{U}_{k_g^u} \in \mathbb{C}^{M_{bs} \times b_u}$ is the combining matrix for corresponding uplink payload; $\mathbf{C}_{k_g^u,S}$ and $\mathbf{C}_{k_g^u,X}$ are covariance matrices of the signal of interest, and interference plus noise terms given as

$$\mathbf{C}_{k_g^u,S} = Cov \left(\gamma_{k_g^u} \hat{\mathbf{H}}_{g,k_g^u} \mathbf{V}_{k_g^u} \mathbf{s}_{k_g^u} \right) = \gamma_{k_g^u}^2 \hat{\mathbf{H}}_{g,k_g^u} \mathcal{D} \left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^{\dagger} \right) \hat{\mathbf{H}}_{g,k_g^u}^{\dagger}, \tag{4.27}$$

$$\mathbf{C}_{k_g^u,X} = Cov \left(\mathbf{z}_{k_g^u} + \mathbf{e}_{k_g^u} \right) = \mathbf{C}_g + \beta_{bs} \cdot \mathcal{D} \left(\mathbf{C}_g \right) - \mathbf{C}_{k_g^u,S}. \tag{4.28}$$

4.3 Liner beamforming

In this section, we will study the performance of the linear beamforming schemes for IBFD-MCMU networks. Linear beamforming schemes usually have low implementation complexity and relatively considerable performance, so they are widely used in current wireless networks. However, IBFD operations complicates the already very complex interference in the MCMU network, while liner beamformers have limited capacity of interference management. Thus, it is desired to exploit the performance of linear beamformers for IBFD-MCMU networks.

Existing linear beamforming techniques for IBFD-MCMU networks are usually developed based on the MMSE beamformers. These designs are effective in maximizing the sum rate or energy efficiency but require high computation overheads due to the complex calculation of the interference covariance matrix. In contrast, the ZF and MRTC schemes have a much lower computational complexity but are not well studied for IBFD-MCMU networks. Thus, we investigate the two schemes for such networks in this section. We do not consider the power allocation but use the full transmit power schemes here, i.e., $\alpha_{k_g^d} = \sqrt{\frac{P_{bs}}{K_g^d}}$ and $\gamma_{k_g^u} = \sqrt{P_{ue}}$, since the focus oh this section is to investigate the interference management capacity of these beamformers. While any off-the-shelf techniques can be employed by changing the power coefficients of all nodes based on the allocated power allocation policy, further improving the sum rate.

4.3.1 Enhanced zero-forcing (eZF)

The objective of zero-forcing beamforming is to eliminate interference, which has stringent requirements on the number of transmitting and receiving antennas to leave null space, i.e., the number of antennas should be larger than the number of total interference. The stringent requirements may not be satisfied at each node in the IBFD-MCMU network. The users usually have very limited antennas, so they cannot perform the conventional zero-forcing precoding or combining, and base stations need a large number of antennas to eliminate the self-interference. Thus, the conventional ZF beamforming is not compatible with IBFD-MCMU networks. Based on the idea of ZF beamforming, we enhance it to be compatible with our networks by minimizing interference. At the base station, the precoder for downlink payload and the combiner for uplink payload should minimize the interference at associated users, which can be formulated as

$$\min_{\|\mathbf{V}_{k_g^d}\|_F^2=b_d} \left\|\mathbf{\Gamma}_{k_g^d}\mathbf{V}_{k_g^d}\right\|_F^2, \quad \min_{\|\mathbf{U}_{k_g^u}\|_F^2=b_u} \left\|\mathbf{U}_{k_g^u}^\dagger\mathbf{\Pi}_{k_g^u}\right\|_F^2, \tag{4.29}$$

where the constraints come from the power conditions as explained above; $\mathbf{\Gamma}_{k_g^d}$ and $\mathbf{\Pi}_{k_g^d}$ are stack matrices given as

$$\begin{aligned}\mathbf{\Gamma}_{k_g^d} = \big[&\mathbf{U}_{1_1^d}^\dagger\hat{\mathbf{H}}_{1_1^d,g};\ldots;\mathbf{U}_{(k-1)_g^d}^\dagger\hat{\mathbf{H}}_{(k-1)_g^d,g};\mathbf{U}_{(k+1)_g^d}^\dagger\hat{\mathbf{H}}_{(k+1)_g^d,g};\ldots;\\ &\mathbf{U}_{K_G^d}^\dagger\hat{\mathbf{H}}_{K_G^d,g};\mathbf{U}_{1_1^u}^\dagger\hat{\mathbf{H}}_{1,g};\ldots;\mathbf{U}_{K_G^u}^\dagger\hat{\mathbf{H}}_{G,g}\big],\end{aligned} \tag{4.30}$$

$$\begin{aligned}\mathbf{\Pi}_{k_g^u} = \big[&\hat{\mathbf{H}}_{g,1_1^u}\mathbf{V}_{1_1^u},\ldots,\hat{\mathbf{H}}_{g,(k-1)_g^u}\mathbf{V}_{(k-1)_g^u},\hat{\mathbf{H}}_{g,(k+1)_g^u}\mathbf{V}_{(k+1)_g^u},\ldots,\\ &\hat{\mathbf{H}}_{g,K_G^u}\mathbf{V}_{K_G^u},\hat{\mathbf{H}}_{g,1}\mathbf{V}_{1_1^d},\ldots,\hat{\mathbf{H}}_{g,G}\mathbf{V}_{K_G^d}\big].\end{aligned} \tag{4.31}$$

The solutions can be given as [66]

$$\mathbf{V}_{k_g^d}^* = \underline{\nu}_{1:b_d}\left[\mathbf{\Gamma}_{k_g^d}^\dagger\mathbf{\Gamma}_{k_g^d}\right], \quad \mathbf{U}_{k_g^u}^* = \underline{\nu}_{1:b_u}\left[\mathbf{\Pi}_{k_g^u}\mathbf{\Pi}_{k_g^u}^\dagger\right], \tag{4.32}$$

where $\underline{\nu}_{1:b}[\mathbf{A}]$ denotes the b eigenvectors associated with the b^{th} smallest eigenvalues of matrix $\mathbf{A}$. Similarly, the precoder for uplink payload and the combiner for downlink payload should minimize the interference at associated users, which can be described as

$$\min_{\|\mathbf{V}_{k_g^u}\|_F^2=1} \left\|\mathbf{\Gamma}_{k_g^u}\mathbf{V}_{k_g^u}\right\|_F^2, \quad \min_{\|\mathbf{U}_{k_g^d}\|_F^2=1} \left\|\mathbf{U}_{k_g^d}^\dagger\mathbf{\Pi}_{k_g^d}\right\|_F^2, \tag{4.33}$$

where $\mathbf{\Gamma}_{k_g^u}$ and $\mathbf{\Pi}_{k_g^u}$ are stack matrices given as

$$\begin{aligned}\mathbf{\Gamma}_{k_g^u} = \big[&\mathbf{U}_{1_1^u}^\dagger \hat{\mathbf{H}}_{1,k_g^u}; \ldots; \mathbf{U}_{(k-1)_g^u}^\dagger \hat{\mathbf{H}}_{g,k_g^u}; \mathbf{U}_{(k+1)_g^u}^\dagger \hat{\mathbf{H}}_{g,k_g^u}; \ldots; \\ &\mathbf{U}_{K_G^u}^\dagger \hat{\mathbf{H}}_{G,k_g^u}; \mathbf{U}_{1_1^d}^\dagger \hat{\mathbf{H}}_{1_1^d,k_g^u}; \ldots; \mathbf{U}_{K_G^d}^\dagger \hat{\mathbf{H}}_{K_G^d,k_g^u}\big],\end{aligned} \tag{4.34}$$

$$\begin{aligned}\mathbf{\Pi}_{k_g^d} = \big[&\hat{\mathbf{H}}_{k_g^d,1} \mathbf{V}_{1_1^d}, \ldots, \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{(k-1)_g^d}, \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{(k+1)_g^d}, \ldots, \\ &\hat{\mathbf{H}}_{k_g^d,G} \mathbf{V}_{K_G^d}, \hat{\mathbf{H}}_{k_g^d,1_1^u} \mathbf{V}_{1_1^u}, \ldots, \hat{\mathbf{H}}_{k_g^d,K_G^u} \mathbf{V}_{K_G^u}\big].\end{aligned} \tag{4.35}$$

These minimization problems have a similar form, so the solutions can be given as

$$\mathbf{V}_{k_g^u}^* = \underline{\nu}_{1:b_u}\left[\mathbf{\Gamma}_{k_g^u}^\dagger \mathbf{\Gamma}_{k_g^u}\right], \quad \mathbf{U}_{k_g^d}^* = \underline{\nu}_{1:b_d}\left[\mathbf{\Pi}_{k_g^d} \mathbf{\Pi}_{k_g^d}^\dagger\right]. \tag{4.36}$$

The loss function of the eZF beamforming design can be given by the sum of its objectives of these minimization problems as

$$\begin{aligned}\mathcal{L}_{\text{eZF}} = &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\left(\left\|\mathbf{\Gamma}_{k_g^d}\mathbf{V}_{k_g^d}\right\|_F^2 + \left\|\mathbf{U}_{k_g^d}^\dagger \mathbf{\Pi}_{k_g^d}\right\|_F^2\right) \\ &+ \sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left(\left\|\mathbf{\Gamma}_{k_g^u}\mathbf{V}_{k_g^u}\right\|_F^2 + \left\|\mathbf{U}_{k_g^u}^\dagger \mathbf{\Pi}_{k_g^u}\right\|_F^2\right).\end{aligned} \tag{4.37}$$

The optimized variables should always reduce the loss as it reduces one of these objective functions.

4.3.2 Enhanced maximum ratio transmission and combining (eMRTC)

The objective of the enhanced maximum ratio transmission and combining beamforming is to maximize the signal of interest, which can be described as

$$\begin{aligned}\max_{\mathbf{U}_{k_g^d},\mathbf{V}_{k_g^d},\mathbf{U}_{k_g^u},\mathbf{V}_{k_g^u}} \quad &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\mathbb{E}\left\{\left\|\mathbf{U}_{k_g^d}^\dagger \mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d}\right\|_F^2\right\} + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\mathbb{E}\left\{\left\|\mathbf{U}_{k_g^u}^\dagger \mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u}\mathbf{s}_{k_g^u}\right\|_F^2\right\}. \\ \text{s.t.} \quad &\left\|\mathbf{V}_{k_g^d}\right\|_F^2 = \left\|\mathbf{U}_{k_g^d}\right\|_F^2 = 1 \;\forall\; k,g \\ &\left\|\mathbf{V}_{k_g^u}\right\|_F^2 = \left\|\mathbf{U}_{k_g^u}\right\|_F^2 = 1 \;\forall\; k,g\end{aligned} \tag{4.38}$$

The objective function is convex and differentiable to a specific precoder or combiner. Fixing other beamforming matrices, the optimization problem with respect to a specific

precoder or combiner has the same form as the maximization problem $\max_{\|\mathbf{X}\|_F^2=b} \|\mathbf{AX}\|_F^2$. Thus, the solutions can be given as [66]

$$\mathbf{V}_{k_g^d}^* = \bar{\nu}_{1:b_d}\left[\mathbf{H}_{k_g^d,g}^\dagger \mathbf{U}_{k_g^d}\mathbf{U}_{k_g^d}^\dagger \mathbf{H}_{k_g^d,g}\right], \tag{4.39}$$

$$\mathbf{U}_{k_g^d}^* = \bar{\nu}_{1:b_d}\left[\mathbf{V}_{k_g^d}^\dagger \mathbf{H}_{k_g^d,g}^\dagger \mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d}\right], \tag{4.40}$$

$$\mathbf{V}_{k_g^u}^* = \bar{\nu}_{1:b_u}\left[\mathbf{H}_{g,k_g^u}^\dagger \mathbf{U}_{k_g^u}\mathbf{U}_{k_g^u}^\dagger \mathbf{H}_{g,k_g^u}\right], \tag{4.41}$$

$$\mathbf{U}_{k_g^u}^* = \bar{\nu}_{1:b_u}\left[\mathbf{V}_{k_g^u}^\dagger \mathbf{H}_{g,k_g^u}^\dagger \mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u}\right], \tag{4.42}$$

where $\bar{\nu}_{1:b}[\mathbf{A}]$ denotes the b eigenvectors associated with the b^{th} largest eigenvalues of matrix $\mathbf{A}$. The loss function of eMRTC beamforming can be given by the reciprocal of the objective function as it is a maximization problem, denoted as

$$\mathcal{L}_{\text{eMRTC}} = \frac{1}{\sum_{g=1}^{G}\sum_{k=1}^{K_g^d} P_{s,k_g^d} + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} P_{s,k_g^u}}, \tag{4.43}$$

where the denominator is equal to the objective function with $P_{s,k_g^d} = \left\|\mathbf{U}_{k_g^d}^\dagger \mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d}\right\|_F^2$ and $P_{s,k_g^u} = \left\|\mathbf{U}_{k_g^u}^\dagger \mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u}\right\|_F^2$. The optimized variables should reduce the loss as they maximize the objective.

4.3.3 Iterative algorithm

The expressions of solutions to the eZF and eMRTC beamformers show interdependence between precoders and combiners. To obtain the optimal beamforming schemes, we iteratively update one of the precoders or combiners with others fixed until it converges. The procedure is summarized as Algorithm 2. At each iteration, the optimized beamformer can reduce the loss function of associated beamforming schemes. Thus, the loss function is guaranteed to be reduced with iterations until it converges to a local minimum, which will be proven via simulations in later Section 4.6.

4.4 Beamforming for self-interference cancellation

As introduced earlier in Section 2.3.3, beamforming could be exploited for self-interference cancellation. In this section, we give details of three beamforming cancellation schemes.

Algorithm 2 Iterative algorithm for optimal linear beamforming.

Input: estimated CSI $\hat{\mathbf{H}}_{r,t}\ \forall\ r,t$
statistical knowledge of CSI error $\tilde{\sigma}_{r,t}\ \forall\ r,t$
convergence threshold ς_{min}
Output: $\mathbf{V}_{k_g^d}, \mathbf{U}_{k_g^d}, \mathbf{V}_{k_g^u}, \mathbf{U}_{k_g^u}\ \forall\ k,g$

```
Randomly initialize precoders V_{k_g^d} and V_{k_g^u} ∀ k, g.
for t = 1, ..., max_iter do
    Update combiners U*_{k_g^d} and U*_{k_g^d} ∀ k, g from their solutions.
    Calculate the associated loss function at this iteration L_alg^(t).
    Calculate the decreased loss as L_- = L_alg^(t-1) − L_alg^(t).
    if L_- < ς_min then
        break.
    else
        Update precoders V*_{k_g^d} and V*_{k_g^u} ∀ k, g from their solutions.
    end if
end for
```

It should be noted that the precoder suppresses the analog SI before the receiver to prevent receiver saturation. Thus the precoder-based cancellation is regarded as an ASIC technique, although it is implemented by digital signal processing.

4.4.1 Orthogonal Tx/Rx beamformers

The received SI at the g^{th} BS are combined through and yields the signal as

$$\tilde{\mathbf{y}}_{\mathrm{si},g} = \mathbf{U}_g^{\dagger}\mathbf{H}_{g,g}\sum_{k=1}^{K_g^d}\left(\alpha_{k_g^d}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d} + \mathbf{d}_{tx,k_g^d}\right), \tag{4.44}$$

where $\mathbf{U}_g$ is designed for the intended uplink payload such that $\mathbf{U}_g = \left[\mathbf{U}_{1_g^u}, \mathbf{U}_{2_g^u}, \ldots, \mathbf{U}_{K_g^u}\right]$. The transmitted and received beams can be steered in the desired direction through the transmitting or receiving beamformers. This indicates that the SI can be eliminated by solving the following problem as

$$\begin{aligned}\min_{\mathbf{V}_{k_g^d},\mathbf{U}_g}\ \mathbb{E}\left\{\|\tilde{\mathbf{y}}_{\mathrm{si},g}\|_F^2\right\} &= \sum_{k=1}^{K_g^d}\alpha_{k_g^d}^2\mathbf{U}_g^{\dagger}\mathbf{H}_{g,g}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger} + \mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)\right)\mathbf{H}_{g,g}^{\dagger}\mathbf{U}_g. \\ \text{s.t. } \|\mathbf{U}_{k_g^u}\|_F^2 &= \|\mathbf{V}_{k_g^d}\|_F^2 = 1\ \forall\ k\end{aligned} \tag{4.45}$$

The optimal solutions to this problem show interdependence between $\mathbf{V}_{k_g^d}$ and $\mathbf{U}_{k_g^u}$, which will require an iterative algorithm for the final solution and cause high complexity.

Alternatively, this could be realized by forming $\mathbf{U}_g^\dagger$ and $\mathbf{V}_{k_g^d}$ that span of orthogonal subspaces chosen from the right and left singular vectors of the SI channel. Perform singular value decomposition (SVD) to $\mathbf{H}_{g,g}$, we can write it as

$$\mathbf{H}_{g,g} = \mathbf{Q}_l \mathbf{\Lambda} \mathbf{Q}_r^\dagger, \tag{4.46}$$

where $\mathbf{Q}_l \in \mathbb{C}^{M_{bs} \times M_{bs}}$ consists of the M_{bs} left singular vectors, $\mathbf{Q}_r \in \mathbb{C}^{N_{bs} \times N_{bs}}$ consists of the N_{bs} right singular vectors, and $\mathbf{\Lambda} \in \mathbb{C}^{M_{bs} \times N_{bs}}$ consists of $\min\{M_{bs}, N_{bs}\}$ singular values along its diagonal. Assume $b_d + b_u \geqslant \min\{M_{bs}, N_{bs}\}$, the beamformers can be chosen as

$$\mathbf{U}_{k_g^u} = [\mathbf{Q}_l]_{:,1:b_u}\,,\ \mathbf{V}_{k_g^d} = [\mathbf{Q}_r]_{:,N_{bs}-b_d+1:N_{bs}}\,, \tag{4.47}$$

which leads to $\mathbf{U}_g^\dagger \mathbf{H}_{g,g} \mathbf{V}_{k_g^d} = \mathbf{0}$ due to the orthogonality of singular vectors, leaving trivial transmitter distortions since $\kappa_{bs} \ll 1$. It is worth noting that the precoder is formed by the singular vectors corresponding to the b_d^{th} smallest singular values here to minimize the received SI power, preventing the receiver saturation. However, this method utilizes all available spatial DoFs for eliminating the SI, so sacrifices the MIMO gain for uplink and downlink payloads. Alternatively, the beamforming matrices can be formed by choosing singular vectors of the intended channel (either uplink or downlink). By doing so, the intended channel strength could be maximized, but the spatial DoFs are utilized by only one user. Thus, this method only suits the point-to-point system.

4.4.2 Null space projection

In MCMU networks, we must consider all target channels and interfering channels to adaptively optimize the system with limited spatial DoFs, while we still want to enhance the ASIC capability of downlink precoders. Thus, we propose an NSP-based MMSE beamforming scheme. The processing consists of two steps: 1) obtain the desired MMSE precoder that minimizes the downlink MSE; 2) project the precoder to the null space of the SI channel to minimize the received RSI power. The first step could be done using the typical MMSE beamforming algorithm. Let $\widetilde{\mathbf{V}}_{k_g^d}$ denote the MMSE precoder, then we project it to the null space of the SI channel as

$$\mathbf{V}_{k_g^d} = \mathbf{P}_{g,D} \widetilde{\mathbf{V}}_{k_g^d}, \tag{4.48}$$

where $\mathbf{P}_{g,D}$ is the projection matrix consisting of subspace that spans the eigenvectors associated with the D smallest eigenvalues of $\mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g} + \kappa_{bs}\mathcal{D}\left(\mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g}\right)$[2] such that

$$\mathbf{P}_{g,D} = [\mathbf{Q}]_{:,M_{bs}-D+1:M_{bs}}\,, \tag{4.49}$$

where $\mathbf{Q}$ comes from the eigenvalue decomposition (EVD) as

$$\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\dagger} = \mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g} + \kappa_{bs}\mathcal{D}\left(\mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g}\right). \tag{4.50}$$

By making the projection, the received RSI will be reduced, and the reduced amount depends on D. It means that D strikes a balance between the SIC depth and precoding errors.

4.4.3 Enhanced minimum mean-squared error (eMMSE)

The objective of the conventional MMSE beamforming scheme is to minimize the sum of MSE of the network, where the MSE expression for a single user is given in Appendix 4.10. Minimizing the sum of MSE by optimizing precoders will include the effects of combiners, as the MSE expressions suggest. As a result, precoders may not provide a sufficient ASIC depth to prevent receiver saturation if combiners have effectively suppressed the RSI and distortions caused by it. However, the degrees of freedom of combiners occupied by SIC and uncorrelated distortions due to receiver saturation will naturally compromise the uplink communication. Thus, the existing MMSE-based beamforming schemes for IBFD networks [38, 53, 54] require sufficient ASIC realized by other techniques to guarantee their performance. As explained in Chapter 2, the analog cancellation techniques usually have very high implementation complexity, which is the main obstacle to bringing IBFD to current networks. Thus, it is desired to take advantage of the spatial DoFs provided by the transmitting antenna array to effectively suppress the SI in the propagation domain, which is expected to be realized by the precoder. For this purpose, we enhance the ASIC capability of MMSE precoders by adding a constraint on the perceived RSI power, formulating the eMMSE problem as

[2] This is obtained by derivating the received RSI power (without digital receiving beamforming) to the downlink precoder, i.e., $\frac{\partial \mathbb{E}\left\{\left\|\mathbf{y}_{\mathrm{si},g}\right\|_F^2\right\}}{\partial \mathbf{V}_{k_g^d}} = \mathbf{H}_{g,g}\mathbf{H}_{g,g}^{\dagger} + \kappa_{bs}\mathcal{D}\left(\mathbf{H}_{g,g}\mathbf{H}_{g,g}^{\dagger}\right)$. The receiving beamformers are not included here since we want to minimize the RSI power before the receiver to prevent it from saturation. The term $\kappa_{bs}\mathcal{D}\left(\mathbf{H}_{g,g}\mathbf{H}_{g,g}^{\dagger}\right)$ helps the projection with suppressing the transmitter distortions caused by the SI, further improving the performance

$$\min_{\{\alpha,\gamma,\mathbf{V},\mathbf{U}\}_{\forall k,g}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \varepsilon_{k_g^d}(\alpha,\gamma,\mathbf{V},\mathbf{U}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \varepsilon_{k_g^u}(\alpha,\gamma,\mathbf{V},\mathbf{U}), \tag{4.51}$$

$$\text{s.t. } \epsilon_{\text{rsi},g}(\mathbf{V}_g) \leq \bar{\epsilon}_{\text{rsi},g} \ \forall\ g, \tag{4.52}$$

where $\bar{\epsilon}_{\text{rsi},g}$ is the tolerable RSI power of the g^{th} BS that could be chosen according to the dynamic range of receivers, and $\epsilon_{\text{rsi},g}(\mathbf{V}_g)$ is the actual received RSI power that can be denoted as

$$\begin{aligned}\epsilon_{\text{rsi},g}(\mathbf{V}_g) &= \sum_{k=1}^{K_g^d} tr\left(\mathcal{D}\left(\mathbf{H}_{g,g}\left(\alpha_{k_g^d}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d}+\mathbf{d}_{tx,k_g^d}\right)\left(\alpha_{k_g^d}\mathbf{V}_{k_g^d}\mathbf{s}_{k_g^d}+\mathbf{d}_{tx,k_g^d}\right)^{\dagger}\mathbf{H}_{g,g}^{\dagger}\right)\right)\\ &= \sum_{k=1}^{K_g^d}\alpha_{k_g^d}^2 tr\left(\mathcal{D}\left(\mathbf{H}_{g,g}\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\mathbf{H}_{g,g}^{\dagger}+\kappa_{bs}\mathbf{H}_{g,g}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)\mathbf{H}_{g,g}^{\dagger}\right)\right).\end{aligned} \tag{4.53}$$

The Lagrange function of this problem is given as

$$\mathcal{L}_{\text{eMMSE}} = \sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\varepsilon_{k_g^d}(\alpha,\gamma,\mathbf{V},\mathbf{U}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\varepsilon_{k_g^u}(\alpha,\gamma,\mathbf{V},\mathbf{U}) + \sum_{g=1}^{G}\lambda_g\left(\epsilon_{\text{rsi},g}(\mathbf{V}_g)-\bar{\epsilon}_{\text{rsi},g}\right), \tag{4.54}$$

where $\lambda_g \geqslant 0$ is the Lagrange multiplier associated with the received RSI power constraint of the g^{th} BS. The Lagrange function is convex and differentiable to a single precoder or combiner, so we can differentiate it with respect to $\mathbf{V}_{k_g^d}$, $\mathbf{V}_{k_g^u}$, $\mathbf{U}_{k_g^d}$, and $\mathbf{U}_{k_g^u}$, respectively, and set the derivatives to zero, yielding the solutions as

$$\mathbf{V}_{k_g^d}^{*} = \mathbb{N}\left[\frac{1}{\alpha_{k_g^d}}\left(\mathbf{\Omega}_g+\lambda_g^{*}\bigtriangledown_{\epsilon,\mathbf{V}}\right)^{-1}\hat{\mathbf{H}}_{k_g^d,g}^{\dagger}\mathbf{U}_{k_g^d}\right], \tag{4.55}$$

$$\mathbf{V}_{k_g^u}^{*} = \mathbb{N}\left[\frac{1}{\gamma_{k_g^u}}\mathbf{\Omega}_{k_g^u}^{-1}\hat{\mathbf{H}}_{g,k_g^u}^{\dagger}\mathbf{U}_{k_g^u}\right], \tag{4.56}$$

$$\mathbf{U}_{k_g^d}^{*} = \mathbb{N}\left[\alpha_{k_g^d}\mathbf{C}_{k_g^d}^{-1}\hat{\mathbf{H}}_{k_g^d,g}\mathbf{V}_{k_g^d}\right], \tag{4.57}$$

$$\mathbf{U}_{k_g^u}^{*} = \mathbb{N}\left[\gamma_{k_g^u}\mathbf{C}_{g}^{-1}\hat{\mathbf{H}}_{g,k_g^u}\mathbf{V}_{k_g^u}\right], \tag{4.58}$$

where $\mathbb{N}[\cdot]$ denote the normalization operation to satisfy the full transmit power conditions, and $\bigtriangledown_{\epsilon,\mathbf{V}}$ denotes the gradient of the actual RSI power with respect to the precoder that can be given as

Algorithm 3 Bisection searching algorithm.

1: Initialising the searching range as $\underline{\lambda}_g = 0$ and $\bar{\lambda}_g \gg tr\left(\mathbf{\Omega}_g\right)/tr\left(\bigtriangledown_{\epsilon,\mathbf{V}}\right)$.
2: **repeat**
3: Calculate $\lambda^{(t)} = (\underline{\lambda}_g + \bar{\lambda}_g)/2$.
4: Update the precoding matrices from the solutions.
5: Calculate current RSI power $\epsilon_{\mathrm{rsi},g}^{(t)}\left(\mathbf{V}_g\right)$.
6: **if** $\epsilon_{\mathrm{rsi},g}^{(t)}\left(\mathbf{V}_g\right) < \bar{\epsilon}_{\mathrm{rsi},g}$ **then**
7: $\bar{\lambda}_g \leftarrow \lambda^{(t)}$;
8: **else**
9: $\underline{\lambda}_g \leftarrow \lambda^{(t)}$;
10: **end if**
11: **until** $\lambda^{(t)} - \lambda^{(t-1)} < \iota_{\mathrm{conv}}$, where ι_{conv} is the converging threshold.

$$\bigtriangledown_{\epsilon,\mathbf{V}} = \mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g} + \kappa_{bs}\mathcal{D}\left(\mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g}\right). \tag{4.59}$$

$\mathbf{\Omega}_g$ and $\mathbf{\Omega}_{k_g^u}$ denote the gradient of the interference with respect to the precoder that is given in Appendix 4.11. The optimal Lagrange multiplier should satisfy the Karush–Kuhn–Tucker (KKT) conditions with complementary slackness as [66]

$$\lambda_g^*\left(\epsilon_{\mathrm{rsi},g}\left(\mathbf{V}_g\right) - \bar{\epsilon}_{\mathrm{rsi},g}\right) = 0,\ \forall\ g \tag{4.60}$$

$$\epsilon_{\mathrm{rsi},g}\left(\mathbf{V}_g\right) \leqslant \bar{\epsilon}_{\mathrm{rsi},g},\ \forall\ g \tag{4.61}$$

$$\lambda_g^* \geqslant 0,\ \forall\ g \tag{4.62}$$

which could be found via the Bisection searching given in Algorithm 3. Theoretically, λ_g strikes a balance between minimizing the RSI power and minimizing the sum MSE. When λ_g is equal to 0, eMMSE degenerates to the traditional MMSE. With λ_g increasing, the downlink precoders will be biased towards minimizing the RSI power, thereby enhancing their ASIC ability.

The loss function of the eMMSE beamforming can be given by the Lagrange function given in (4.54), i.e., the sum of MSE plus RSI power. The optimized variables guarantee to reduce the loss at each iteration since the solution is given by the local minimum of the Lagrange function.

4.4.4 Computational complexity

We measure the computational complexity by accounting for the required multiplication operations. We assume $K_g^d = K_g^u = K, \forall\ g$, $N_{bs} = M_{bs} = A_b$, $N_{ue} = M_{ue} = A_u$, and

$b_d = b_u = N_s$ for simplicity. Usually, we will have $A_b \gg A_u \geqslant N_s$. Taking $\mathbf{V}_{k_g^d}$ as an instance, calculating the stack matrix $\mathbf{\Gamma}_{k_g^d}$ and $\mathbf{\Gamma}_{k_g^d}^{\dagger}\mathbf{\Gamma}_{k_g^d}$ for eZF needs a total of

$$(3KG-1)A_b^2N_s + (KG-1)A_bA_uN_s$$

multiplications[3]; calculating $\mathbf{H}_{k_g^d,g}^{\dagger}\mathbf{U}_{k_g^d}\mathbf{U}_{k_g^d}^{\dagger}\mathbf{H}_{k_g^d,g}$ for eMRTC requires only

$$A_b^2N_s + A_bA_uN_s$$

multiplications, while the Eigen decomposition has the same complexity (i.e., $\mathcal{O}\left(A_b^3\right)$). Thus, eZF has higher computational complexity than eMRTC with even $G = 1$ and $K = 1$, and the difference increases significantly with increasing G and K. For the eMMSE precoder, calculating $\mathbf{\Omega}_g + \lambda_g^* \bigtriangledown_{\epsilon,\mathbf{V}}$ needs a total of

$$KG(2A_b^3 + A_b^2A_u + A_u^2A_b + (A_b^2 + A_u^2)N_s + 2A_b^2) + A_b^3$$

multiplications and the inverse operation has the computational complexity of $\mathcal{O}\left(A_b^3\right)$, the rest of matrix multiplication requires

$$A_b^2A_u + A_bA_uN_s$$

multiplications. Thus, the eMMSE beamforming has a much higher computational complexity than eZF and eMRTC in MCMU networks, regardless of the complexity of Bi-directional searching for finding the Lagrange multiplier. The computational complexity of the three schemes can be given as

$$C_{\text{eZF}} = \mathcal{O}\left(A_b^3 + GKA_b^2N_s\right), \tag{4.63}$$

$$C_{\text{eMRTC}} = \mathcal{O}\left(A_b^3\right), \tag{4.64}$$

$$C_{\text{MMSE}} = \mathcal{O}\left(GKA_b^3\right). \tag{4.65}$$

It should be noted that although eMMSE, eZF, and eMRTC have the computational complexity of the same order with $G = K = 1$, eMMSE needs more multiplication, as

[3]The required multiplication operations of calculating the matrix multiplication of $\mathbf{A}_1\mathbf{A}_2$ is $a \times b \times c$, where $\mathbf{A}_1$ is of size $a \times b$ and $\mathbf{A}_2$ is of size $b \times c$.

analyzed above. The difference is more significant with enlarging antenna arrays and networks (i.e., increasing numbers of cells and users).

4.5 Joint power allocation and beamforming

In this section, we jointly optimize the power coefficients to further improve the system throughput.

4.5.1 Problem formulation

The power allocation and beamforming should have a common objective for the joint design. We use the MMSE as the design criteria since it can maximize the sum rate of the network with MMSE receivers, which are proved in [54]. Meanwhile, we would like the downlink precoder to suppress the SI in the propagation domain to relieve the burden of RF cancellation. Thus, we formulate the joint power allocation and beamforming problem as

$$(P.1) \quad \min_{\{\alpha,\gamma,\mathbf{V},\mathbf{U}\}_{\forall k,g}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \varepsilon_{k_g^d}(\alpha,\gamma,\mathbf{V},\mathbf{U}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \varepsilon_{k_g^u}(\alpha,\gamma,\mathbf{V},\mathbf{U}) \tag{4.66}$$

$$\text{s.t. } (C.1) \ \sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right) \leqslant P_{bs}, \ \forall\, g \tag{4.67}$$

$$(C.2) \ \gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right) \leqslant P_{ue}, \ \forall\, k,g \tag{4.68}$$

$$(C.3) \ \epsilon_{\text{rsi},g}(\mathbf{V}_g) \leqslant \bar{\epsilon}_{\text{rsi},g}, \ \forall\, g \tag{4.69}$$

where constraints $(C.1)$ and $(C.2)$ satisfy the transmit power budgets of the base stations and users, respectively, and constraint $(C.3)$[4] enhances the ASIC capability of the downlink precoder, as explained in Section 4.4.3.

Lemma 4.1. *The optimization problem* $(P.1)$ *is equivalent to the problem* $(P.2)$ *shown below with appropriate* ν_g *values.*

[4]We still denote the RSI power as the function of the downlink precoder only for the joint power allocation and beamforming problem since we want to suppress the SI solely by precoding schemes instead of reducing the transmit power for the downlink payload.

$$(P.2) \quad \min_{\{\alpha,\gamma,\mathbf{V},\mathbf{U}\}_{\forall k,g}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \varepsilon_{k_g^d}(\alpha,\gamma,\mathbf{V},\mathbf{U}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \varepsilon_{k_g^u}(\alpha,\gamma,\mathbf{V},\mathbf{U}) + \sum_{g=1}^{G} \nu_g \epsilon_{rsi,g}(\mathbf{V}_g) \tag{4.70}$$

$$s.t.\ (C.1) \ \sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right) \leqslant P_{bs},\ \forall\ g \tag{4.71}$$

$$(C.2)\ \gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right) \leqslant P_{ue},\ \forall\ k,g \tag{4.72}$$

Proof. See Appendix 4.12. □

According to Lemma 4.1, the optimal power allocation and beamforming schemes can be obtained from solving problem $(P.2)$. Still, an appropriate value of ν_g needs to be chosen at first to formulate the problem. As stated in Appendix 4.12, ν_g should be equal to $\varpi_{\text{rsi},g}$ to make th problems equivalent, while $\varpi_{\text{rsi},g}$ depends on the chosen tolerable RSI power $\bar{\epsilon}_{\text{rsi},g}$. It is challenging to mathematically derive the value of ν_g from given $\bar{\epsilon}_{\text{rsi},g}$ since the multiple variables are entangled. However, we can directly give the value of ν_g, and it will reflect a specific tolerable RSI power. The resulting solutions will maximize the sum rate under the corresponding RSI power constraint. We choose ν_g according to its properties. As (4.70) suggests, ν_g can adjust the preference of the precoder: the objective function will be dominated by the sum of MSE when ν_g is relatively small, and conversely, the objective function will be dominated by the RSI when ν_g is rather large. As a result, precoders will tend to minimize the sum of MSE or RSI accordingly. Therefore, we can set a relatively large ν_g to help suppress SI before the receiver if there is not sufficient ASIC depth and set a relatively small ν_g to minimize the precoding errors if there is sufficient ASIC depth. According to experiments, we can set the value of ν_g based on the realized ASIC depth η_g as $\nu_g = 10^{-\eta_g/5}$.

4.5.2 Sub-problems and solutions

From the MSE expressions, we can see that the main difficulty of the minimization problem $(P.2)$ is that the optimized variables are entangled, leading to a non-convex problem. The non-convex constrained optimization problem may not be converted to a convex form with simple manipulations. Thus, we propose a two-stage approach to

solve it, where the original joint optimization problem is decomposed into two convex sub-problems: interference management through beamforming and power allocation. The two sub-problems are relatively easy to solve through standard solutions, but the decomposition leads to a suboptimal solution to the joint optimization problem.

In the interference management stage, we will fix the power coefficients to form the first sub-problem regarding beamforming matrix optimization. However, the objective function in (4.70) is still not jointly convex to the transmitting and receiving beamforming matrices (i.e., precoders and combiners). Thus, we further divide this sub-problem into combiner updating and precoder updating stages.

4.5.2.1 Receiving beamforming

With fixed transmitting beamforming matrices and power coefficients, the combining matrices are optimized to minimize the sum of MSE. The analog RSI power is not affected by the combiners, so it is removed from the objective function, and the combiner optimization problem is formulated as

$$(S.1.1) \quad \min_{\{\mathbf{U}\}_{\forall k,g}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \varepsilon_{k_g^d}(\mathbf{U}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \varepsilon_{k_g^u}(\mathbf{U}), \tag{4.73}$$

which is equivalent to the conventional MMSE receiver. The objective function is convex and differentiable to the combining matrices with other variables fixed, and there is no constraint associated with combining matrices. Thus, we can differentiate the objective function with respect to $\mathbf{U}_{k_g^d}$ and $\mathbf{U}_{k_g^u}$ $\forall\ k, g$, and set the derivatives to zero, then the optimal solutions to the sub-problem $(S.1.1)$ are given as

$$\mathbf{U}_{k_g^d}^* = \alpha_{k_g^d} \mathbf{C}_{k_g^d}^{-1} \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d}, \tag{4.74}$$

$$\mathbf{U}_{k_g^u}^* = \gamma_{k_g^u} \mathbf{C}_{g}^{-1} \hat{\mathbf{H}}_{g,k_g^u} \mathbf{V}_{k_g^u}, \tag{4.75}$$

where $\mathbf{C}_{k_g^d}$ and $\mathbf{C}_g$ are covariance matrices of the received signals at the downlink user k_g^d and the g^{th} base station that has been derived in Appendix 4.9.

4.5.2.2 Transmitting beamforming

With fixed combining matrices and power coefficients, the sub-problem regarding precoder optimization is formulated as

$$(S.1.2) \quad \min_{\{\alpha,\gamma,\mathbf{V},\mathbf{U}\}_{\forall k,g}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \varepsilon_{k_g^d}(\mathbf{V}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \varepsilon_{k_g^u}(\mathbf{V}) + \sum_{g=1}^{G} \nu_g \epsilon_{\mathrm{rsi},g}(\mathbf{V}_g). \tag{4.76}$$

$$\text{s.t. } (C.1) \ \sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right) \leqslant P_{bs}, \ \forall\, g \tag{4.77}$$

$$(C.2) \ \gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right) \leqslant P_{ue}, \ \forall\, k,g \tag{4.78}$$

To solve the constrained optimization problem, we need to augment the objective function with a weighted sum of the constraint functions [66], yielding the Lagrange function as

$$\begin{aligned}\mathcal{L}(\mathbf{V},\nu,\varpi) = &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\left(\varepsilon_{k_g^d}(\mathbf{V}) + \varpi_g \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H\right) - \varpi_g P_{bs}\right)\\ &+ \sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left(\varepsilon_{k_g^u}(\mathbf{V}) + \varpi_{k_g^u}\left(\gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H\right) - P_{ue}\right)\right)\\ &+ \sum_{g=1}^{G}\nu_g\sum_{k=1}^{K_g^d}\alpha_{k_g^d}^2 tr\left(\mathbf{H}_{g,g}\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H\mathbf{H}_{g,g}^H + \kappa_{bs}\mathbf{H}_{g,g}\mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H\right)\mathbf{H}_{g,g}^H\right),\end{aligned} \tag{4.79}$$

where $\varpi_g \geqslant 0$ and $\varpi_{k_g^u} \geqslant 0$ are Lagrange multipliers associated with power constraints of base stations and uplink users. Differentiate the Lagrange function with respect to $\mathbf{V}_{k_g^d}$ and $\mathbf{V}_{k_g^u}\ \forall\, k,g$, and set the derivatives to zero, the optimal solutions to the sub-problem $(S.1.2)$ are given as

$$\mathbf{V}_{k_g^d}^* = \frac{1}{\alpha_{k_g^d}}\left(\mathbf{\Omega}_g + \nu_g \bigtriangledown_{\epsilon,\mathbf{V}} + \varpi_g^*\mathbf{I}\right)^{-1}\mathbf{H}_{k_g^d,g}^{\dagger}\mathbf{U}_{k_g^d}, \tag{4.80}$$

$$\mathbf{V}_{k_g^u}^* = \frac{1}{\gamma_{k_g^d}}\left(\mathbf{\Omega}_{k_g^u} + \varpi_{k_g^u}^*\mathbf{I}\right)^{-1}\mathbf{H}_{g,k_g^u}^{\dagger}\mathbf{U}_{k_g^u}, \tag{4.81}$$

where $\bigtriangledown_{\epsilon,\mathbf{V}} = \mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g} + \kappa_{bs}\mathcal{D}\left(\mathbf{H}_{g,g}^{\dagger}\mathbf{H}_{g,g}\right)$ as given earlier in Section 4.4.3, and $\mathbf{\Omega}_{k_g^d}$ and $\mathbf{\Omega}_{k_g^u}$ represent the interference-plus-distortions matrices for associated users given in Appendix 4.11. The optimal Lagrange multipliers should satisfy the KKT conditions

with complementary slackness as [66]

$$\varpi_g^* \left(\sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}^* \left(\mathbf{V}_{k_g^d}^* \right)^\dagger \right) - P_{bs} \right) = 0, \forall\ g \tag{4.82}$$

$$\sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}^* \left(\mathbf{V}_{k_g^d}^* \right)^\dagger \right) \leqslant P_{bs}, \forall\ g \tag{4.83}$$

$$\varpi_{k_g^u}^* \left(\gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}^* \left(\mathbf{V}_{k_g^u}^* \right)^\dagger \right) - P_{ue} \right) = 0, \forall\ k, g \tag{4.84}$$

$$\gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}^* \left(\mathbf{V}_{k_g^u}^* \right)^\dagger \right) \leqslant P_{ue}, \forall\ k, g \tag{4.85}$$

$$\varpi_g^* \geqslant 0,\ \varpi_{k_g^u}^* \geqslant 0, \forall\ k, g \tag{4.86}$$

They could also be found through Bisection searching. However, calculating the precoder and its norm with different values of the Lagrange multipliers for each iteration during the Bisection searching will yield high computational complexity. To reduce the complexity, we use SVD to convert the expressions of the transmit power into a scalar form. Performing SVD to $\mathbf{\Omega}_g + \nu_g \bigtriangledown_{\epsilon,\mathbf{V}}$[5] such that

$$\mathbf{\Omega}_g + \nu_g \bigtriangledown_{\epsilon,\mathbf{V}} = \mathbf{Q}_g \mathbf{D}_g \mathbf{Q}_g^\dagger, \tag{4.87}$$

where $\mathbf{Q}_g \mathbf{Q}_g^\dagger = \mathbf{I}$. Then, the optimal downlink precoder solution given in Equation (4.80) can be alternatively written as

$$\begin{aligned} \mathbf{V}_{k_g^d}^* &= \frac{1}{\alpha_{k_g^d}} \left(\mathbf{Q}_g \mathbf{D}_g \mathbf{Q}_g^\dagger + \mathbf{Q}_g \varpi_g^* \mathbf{Q}_g^\dagger \right)^{-1} \mathbf{H}_{k_g^d,g}^\dagger \mathbf{U}_{k_g^d} \\ &= \frac{1}{\alpha_{k_g^d}} \mathbf{Q}_g \left(\mathbf{D}_g + \varpi_g^* \mathbf{I} \right)^{-1} \mathbf{Q}_g^\dagger \mathbf{H}_{k_g^d,g}^\dagger \mathbf{U}_{k_g^d}. \end{aligned} \tag{4.88}$$

Let $\mathbf{B}_{k_g^d} = \mathbf{Q}_g^\dagger \mathbf{H}_{k_g^d,g}^\dagger \mathbf{U}_{k_g^d} \mathbf{U}_{k_g^d}^\dagger \mathbf{H}_{k_g^d,g} \mathbf{Q}_g$, the BS power constraint ($C.1$) can be written as

$$\sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}^* (\mathbf{V}_{k_g^d}^*)^\dagger \right) = \sum_{k=1}^{K_g^d} \sum_{n=1}^{N_{bs}} \frac{[\mathbf{B}_{k_g^d}]_{n,n}}{\left([\mathbf{D}_g]_{n,n} + \varpi_g \right)^2} \leqslant P_{bs}. \tag{4.89}$$

Similarly, performing SVD to $\mathbf{\Omega}_{k_g^u}$ such that $\mathbf{\Omega}_{k_g^u} = \mathbf{Q}_{k_g^u} \mathbf{D}_{k_g^u} \mathbf{Q}_{k_g^u}^\dagger$ and letting $\mathbf{B}_{k_g^u} = \mathbf{Q}_{k_g^u}^\dagger \mathbf{H}_{g,k_g^u}^\dagger \mathbf{U}_{k_g^u} \mathbf{U}_{k_g^u}^\dagger \mathbf{H}_{g,k_g^u} \mathbf{Q}_{k_g^u}$, the uplink UE power constraint ($C.2$) can be written as

[5] It is a Hermitian matrix since both the covariance matrix $\mathbf{\Omega}_g$ and the gradient $\bigtriangledown_{\epsilon,\mathbf{V}}$ are Hermitian, so it can be singular value decomposed.

$$\sum_{n=1}^{N_{ue}} \frac{[\mathbf{B}_{k_g^u}]_{n,n}}{\left([\mathbf{D}_{k_g^u}]_{n,n} + \varpi_{k_g^u}\right)^2} \leqslant P_{ue}. \tag{4.90}$$

Since $\mathbf{B}_g$, $\mathbf{D}_g$, $\mathbf{B}_{k_g^u}$, and $\mathbf{D}_{k_g^u}$ are fixed while updating the precoding matrices, we only need to calculate their values once before the Bisection searching, then only scalar calculations are required during the process of Bisection searching for finding the Lagrange multipliers.

4.5.2.3 Power allocation

In the second step, the beamforming matrices are fixed to update the power coefficients. The sub-problem with regard to power allocation policy optimization is formulated as

$$(S.2) \min_{\{\alpha,\gamma,\mathbf{V},\mathbf{U}\}_{\forall k,g}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \varepsilon_{k_g^d}(\alpha,\gamma) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \varepsilon_{k_g^u}(\alpha,\gamma). \tag{4.91}$$

$$\text{s.t. } (C.1) \sum_{k=1}^{K_g^d} \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right) \leqslant P_{bs},\ \forall\ g \tag{4.92}$$

$$(C.2)\ \gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right) \leqslant P_{ue},\ \forall\ k,g \tag{4.93}$$

To solve the constrained problem, we augment the objective function with a weighted sum of the constraint functions, yielding the Lagrange function as

$$\begin{aligned}\mathcal{L}(\alpha,\gamma,\lambda) = &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\left(\varepsilon_{k_g^d}(\alpha,\gamma) + \lambda_g \alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{H}\right) - \lambda_g P_{bs}\right)\\ &+ \sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left(\varepsilon_{k_g^u}(\alpha,\gamma) + \lambda_{k_g^u}\left(\gamma_{k_g^u}^2 tr\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{H}\right) - P_{ue}\right)\right),\end{aligned} \tag{4.94}$$

where $\lambda_g \geqslant 0$ and $\lambda_{k_g^u} \geqslant 0$ are Lagrange multipliers associated with power constraints for the power coefficients. The Lagrange function is convex to the power coefficients. Differentiate the Lagrange function with respect to $\alpha_{k_g^d}$ and $\gamma_{k_g^u}$ $\forall\ k,g$ and set the derivatives to zero, the optimal power coefficients for the sub-problem $(S.2)$ are given as

$$\alpha_{k_g^d}^* = \frac{\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^d}^{\dagger}\hat{\mathbf{H}}_{k_g^d,g}\mathbf{V}_{k_g^d}\right)\right\}}{\chi_{k_g^d} + \lambda_g^* tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{H}\right)}, \tag{4.95}$$

$$\gamma_{k_g^u}^* = \frac{\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^u}^\dagger \hat{\mathbf{H}}_{g,k_g^u} \mathbf{V}_{k_g^u}\right)\right\}}{\chi_{k_g^u} + \lambda_{k_g^u}^* tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^H\right)}, \tag{4.96}$$

where $\chi_{k_g^d}$ and $\chi_{k_g^u}$ denote the power of interference plus noise that are detailed in Appendix 4.13. The optimal Lagrange multipliers should satisfy the KKT conditions with complementary slackness as

$$\lambda_g^* \left(\sum_{k=1}^{K_g^d} (\alpha_{k_g^d}^*)^2 tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger\right) - P_{bs}\right) = 0, \forall\ g \tag{4.97}$$

$$\sum_{k=1}^{K_g^d} (\alpha_{k_g^d}^*)^2 tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger\right) \leqslant P_{bs}, \forall\ g \tag{4.98}$$

$$\lambda_{k_g^u}^* \left((\gamma_{k_g^u}^*)^2 tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right) - P_{ue}\right) = 0, \forall\ k, g \tag{4.99}$$

$$(\gamma_{k_g^u}^*)^2 tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right) \leqslant P_{ue}, \forall\ k, g \tag{4.100}$$

$$\lambda_g^* \geqslant 0,\ \lambda_{k_g^u}^* \geqslant 0, \forall\ k, g \tag{4.101}$$

To find the optimal value of Lagrange multiplier $\lambda_{k_g^u}$, we can differentiate the Lagrange function with respect to $\lambda_{k_g^u}$ and set the derivative to zero as

$$\begin{aligned} \frac{\partial \mathcal{L}(\alpha, \gamma, \lambda)}{\partial \lambda_{k_g^u}} &= (\gamma_{k_g^u}^*)^2 tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right) - P_{ue} \\ &= \left(\frac{\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^u}^\dagger \hat{\mathbf{H}}_{g,k_g^u} \mathbf{V}_{k_g^u}\right)\right\}}{\chi_{k_g^u} + \lambda_{k_g^u}^* tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right)}\right)^2 tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right) - P_{ue} = 0, \end{aligned} \tag{4.102}$$

yielding the optimal Lagrange multiplier as

$$\lambda_{k_g^u}^* = \max\left\{0, -\frac{\chi_{k_g^u}}{tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right)} + \frac{\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^u}^\dagger \hat{\mathbf{H}}_{g,k_g^u} \mathbf{V}_{k_g^u}\right)\right\}}{\sqrt{tr\left(\mathbf{V}_{k_g^u} \mathbf{V}_{k_g^u}^\dagger\right) P_{ue}}}\right\}, \tag{4.103}$$

where $\max\{\cdot\}$ guarantees $\lambda_{k_g^u}^* \geqslant 0$ to strictly satisfy the constraint. For λ_g, its optimal value can be obtained from the positive root of

$$\sum_{k=1}^{K_g^d} \left(\frac{\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^d}^\dagger \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d}\right)\right\}}{\chi_{k_g^d} + \lambda_g^* tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger\right)}\right)^2 - \frac{P_{bs}}{tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger\right)} = 0$$

according to the derivative, i.e., $\frac{\partial \mathcal{L}(\alpha,\gamma,\lambda)}{\partial \lambda_g}$, or $\lambda_g^* = 0$ if the positive root does not exist. However, it is difficult to derive the closed-form expression of its root since the variable is in the denominator. Thus, we employ the Bisection searching to find its optimal value satisfying the KKT conditions. The searching range can be given as $\left[0, \bar{\lambda}_g\right]$, where $\bar{\lambda}_g$ is the upper bound that can be derived as follows. The BS power constraint can be written as

$$\begin{aligned}
0 &= P_{bs} - \sum_{k=1}^{K_g^d} \left(\frac{\mathcal{R}\left\{ tr\left(\mathbf{U}_{k_g^d}^\dagger \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d} \right) \right\}}{\chi_{k_g^d} + \lambda_g tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right)} \right)^2 tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right) \\
&\leqslant P_{bs} - \frac{\sum_{k=1}^{K_g^d} \mathcal{R}\left\{ tr\left(\mathbf{U}_{k_g^d}^\dagger \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d} \right) \right\}^2 tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right)}{\sum_{k=1}^{K_g^d} \left(\chi_{k_g^d} + \lambda_g tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right) \right)^2} \\
&= P_{bs} - \frac{\sum_{k=1}^{K_g^d} \mathcal{R}\left\{ tr\left(\mathbf{U}_{k_g^d}^\dagger \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d} \right) \right\}^2 tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right)}{\lambda_g^2 \sum_{k=1}^{K_g^d} \left(tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right) \right)^2 + 2\lambda_g \sum_{k=1}^{K_g^d} tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right) \chi_{k_g^d} + \sum_{k=1}^{K_g^d} \chi_{k_g^d}^2},
\end{aligned} \tag{4.104}$$

which gives

$$\begin{aligned}
\lambda_g^2 \sum_{k=1}^{K_g^d} tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right) + 2\lambda_g \sum_{k=1}^{K_g^d} \chi_{k_g^d} + \sum_{k=1}^{K_g^d} \frac{\chi_{k_g^d}^2}{tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right)} \\
- \frac{1}{P_{bs}} \sum_{k=1}^{K_g^d} \mathcal{R}\left\{ tr\left(\mathbf{U}_{k_g^d}^\dagger \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d} \right) \right\}^2 \leqslant 0,
\end{aligned} \tag{4.105}$$

which is in the form of $a\lambda_g^2 + 2b\lambda_g + c \leq 0$, where $a = \sum_{k=1}^{K_g^d} tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right)$, $b = \sum_{k=1}^{K_g^d} \chi_{k_g^d}$, and $c = \sum_{k=1}^{K_g^d} \frac{\chi_{k_g^d}^2}{tr\left(\mathbf{V}_{k_g^d} \mathbf{V}_{k_g^d}^\dagger \right)} - \frac{1}{P_{bs}} \sum_{k=1}^{K_g^d} \mathcal{R}\left\{ tr\left(\mathbf{U}_{k_g^d}^\dagger \hat{\mathbf{H}}_{k_g^d,g} \mathbf{V}_{k_g^d} \right) \right\}^2$. Thus, we can derive the upper bound as

$$\lambda_g \leqslant -\frac{b}{a} + \sqrt{\frac{b^2}{a^2} - \frac{c}{a}} = \bar{\lambda}_g. \tag{4.106}$$

4.5.3 Iterative algorithm

The solutions to these sub-problems show interdependence on each other. Thus, we need an iterative algorithm to obtain the overall solution by updating the beamforming matrices and power coefficients until they converge to constants. We first initialize power coefficients to meet the power budgets of each node and randomly initialize beamforming

Algorithm 4 Iterative JPAIM algorithm

1: Initialize the power allocation coefficients as $\alpha_{k_g^d} = \sqrt{\frac{P_{bs}}{b_d K_g^d}}$ and $\gamma_{k_g^u} = \sqrt{\frac{P_{ue}}{b_u}}$ $\forall\, k, g$.
2: Randomly initialize and normalize the transmitting beamforming matrices.
3: **repeat**
4: Update the receiving beamforming matrices $\mathbf{U}_{k_g^d}$ and $\mathbf{U}_{k_g^u}$ $\forall\ k, g$.
5: Calculate and record the current sum MSE of the network as $\varepsilon^{(t)}$.
6: Update the precoding matrices $\mathbf{V}_{k_g^d}$ and $\mathbf{V}_{k_g^u}$ $\forall\ k, g$.
7: Update the power allocation coefficients $\alpha_{k_g^d}$ and $\gamma_{k_g^u}$ $\forall\ k, g$.
8: **until** $\varepsilon^{(t)} - \varepsilon^{(t-1)} < \iota_{\min}$ or $t > t_{\max}$, where $\iota_{\min}$ is the sum MSE threshold and $t_{\max}$ is the maximum iteration limit.

matrices. Then, we continuously update the transmitting beamforming matrices, power coefficients, and receiving beamforming matrices in order. The iterative procedure stops if the loss function, i.e., the sum MSE of the network, does not decrease with iterations (or the decreasing amount is less than the threshold) or the iteration time exceeds the limitation. This processing is summarized as Algorithm 4. Note that, since the joint optimization problem is not jointly convex to the power coefficients and beamforming matrices, the algorithm can only converge to the local minimal closest to the initial point. Thus, the performance strongly depends on the initial point, which is also the limiting factor of existing studies, and a global optimum of the joint power allocation and beamforming in the MCMU network is still challenging to be achieved [54].

4.5.3.1 Convergence behaviour

Let $\Sigma_M(\alpha, \beta, \mathbf{V}, \mathbf{U})$ denote the sum MSE of the network, which is the loss function of the joint optimization problem $(P.1)$ (or $(P.2)$) and the three sub-problems $(S.1.1)$, $(S.1.2)$, and $(S.2)$. At each iteration, one of the optimizing variables is updated with others fixed, and the Lagrange function is convex to the target variable under this condition. Thus, we always obtain the optimal value of the target variable with other variables fixed, i.e., reach the global minimum of corresponding sub-problem. So the sum MSE of the network is guaranteed to be reduced at each step as

$$\Sigma_M\left(\alpha, \gamma, \mathbf{V}, \mathbf{U}^{(t+1)}\right) \leqslant \Sigma_M\left(\alpha, \gamma, \mathbf{V}, \mathbf{U}^{(t)}\right), \tag{4.107}$$

$$\Sigma_M\left(\alpha, \gamma, \mathbf{V}^{(t+1)}, \mathbf{U}\right) \leqslant \Sigma_M\left(\alpha, \gamma, \mathbf{V}^{(t)}, \mathbf{U}\right), \tag{4.108}$$

$$\Sigma_M\left(\alpha^{(t+1)}, \gamma^{(t+1)}, \mathbf{V}, \mathbf{U}\right) \leqslant \Sigma_M\left(\alpha^{(t)}, \gamma^{(t)}, \mathbf{V}, \mathbf{U}\right). \tag{4.109}$$

After the t^{th} iteration, we have the joint power allocation and beamforming policy as $\alpha^{(t)}, \gamma^{(t)}, \mathbf{V}^{(t)}, \mathbf{U}^{(t)}$, and then update it to $\mathbf{U}^{(t+1)}, \mathbf{V}^{(t+1)}, \alpha^{(t+1)}, \gamma^{(t+1)}$ step by step. Since the loss function is reduced at each step of this iteration, the loss function is guaranteed to be reduced at each iteration until it converges to a local optimal.

4.5.3.2 Computational complexity

We also assume $K_g^d = K_g^u = K, \forall\ g$, $N_{bs} = M_{bs} = A_b$, $N_{ue} = M_{ue} = A_u$, and $b_d = b_u = N_s$ for simplicity and we usually have $A_b \gg A_u \geq N_s$. The total number of multiplications of calculating beamforming matrices (taking $\mathbf{V}_{k_g^d}$ as an instance) can be given as

$$M_v = \underbrace{GK\left[3A_b^3 + A_b^2(2A_u + 3N_s + 6) + A_b(A_u^2 + 2A_uN_s) + A_u^2N_s\right]}_{\mathbf{\Omega}_g} + \underbrace{A_b^3}_{\bigtriangledown_{\epsilon,\mathbf{V}}} + \underbrace{A_b^2A_u + A_bA_uN_s}_{\text{rest multiplication}},$$

plus an Eigen decomposition and an inverse to a matrix of size $A_b \times A_b$, where the calculation process is divided into calculating $\mathbf{\Omega}_g$, calculating $\bigtriangledown_{\epsilon,\mathbf{V}}$, and rest matrix multiplication, respectively. To update the corresponding power coefficient $\alpha_{k_g^d}$, the complexity mainly comes from calculating $\chi_{k_g^d}$ and the remaining matrix multiplication. Thus, the total number of multiplication operations required to compute $\alpha_{k_g^d}$ is

$$M_\alpha = \underbrace{GK[2A_b^3 + A_b^2(A_u + 5N_s + 2) + A_b(A_u^2 + 4A_uN_s + N_s^2) + A_u^2N_s + 2A_uN_s + 2N_s^2]}_{\chi_{k_g^d}} + \underbrace{A_b^2N_s + A_b(A_uN_s + N_s^2)}_{\text{remaining multiplication}}.$$

Therefore, the number of multiplication operations required for one iteration of the JPAIM algorithm is given as $M_v + M_\alpha$ plus an Eigen decomposition and inverse operation to a matrix of size $A_b \times A_b$, which is in the order of GKA_b^3[6] such that

$$C_{\text{JPAIM}} = \mathcal{O}\left(GKA_b^3\right). \tag{4.110}$$

[6]The complexity of calculating other beamforming matrices will not exceed the number of multiplications of calculating the downlink precoding matrices.

Remarking: We compare our algorithm with the maximum weighted sum rate (MWSR) algorithm to evaluate its performance. The complexity of the MWSR algorithm can be analyzed similarly based on the solution given in [54], which is given as $C_{\text{MWSR}} = \mathcal{O}\left(GKA_b^3\right)$. It should be noted that although the complexity of the two algorithms is in the same order, JPAIM needs $M_\alpha + A_b^3$ more multiplications, which almost double the time complexity of a single iteration. Therefore, the JPAIM algorithm needs at least half the convergence time to reduce the overall complexity.

4.6 Simulation results

Our simulations follow the 3GPP specifications [3, 67], and the related parameters are given in Table 4.1. We consider each cell to have the same number of downlink and uplink users in simulations, i.e., $K_g^d = K_g^u = K \ \forall \ g$. The ASIC depth realized by any other methods (e.g., antenna isolation or RF cancellers as introduced in Chapter 2) will be directly reflected by the pathloss of the SI channel regardless of the implementation methods. We assume the pathloss of the SI channel without any ASIC is 0 dB due to the proximity of the IBFD transmitter and receiver. Therefore, x dB of realized ASIC by other techniques means $\eta_g = x$ dB $\forall g$ in simulations. The network is constructed

TABLE 4.1: Simulation parameters for MCMU networks based on 3GPP specifications.

Parameters	**Values**
Transmit power	$P_{bs} = 23 + 10\log_{10} K$ dBm, $P_{ue} = 23$ dBm
Rician factor	20 dB for SI channel [7]
N_{bs} (M_{bs})	16 (if not specified)
N_{ue} (M_{ue})	2 (if not specified)
LOS probability	Table 7.4.2-1 in [3]
Pathloss	Table 7.4.1-1 in [3]
Thermal noise density	-174 dBm/Hz
Noise figure	13 dB (BS), 9 dB (UE)
ADC/DAC bits	12 (if not specified)
Channel uncertainty	$\hat{\sigma}_{B,A}^2 = \varrho \left\|\mathbf{H}_{B,A}\right\|_F^2 \ \forall A, B$, where $\varrho = -120$ dB

based on 3GPP specifications, e.g., urban-micro (UMi), urban-macro (UMa), rural-macro (RMa), and indoor hotspot (InH) scenarios. The parameters of the four scenarios are given in Table 4.2. We consider a common UMi scenario with 200 m of inter-site distance (i.e., the distance between two neighboring base stations) if not specified.

TABLE 4.2: Four typical 3GPP-specified scenarios (ISD: inter-site distance).

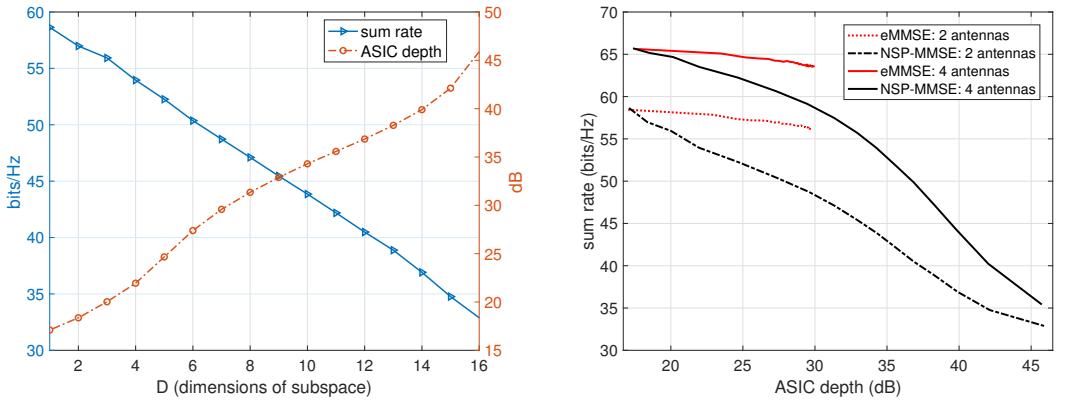
Parameters		**UMi - street canyon**	**UMa**
Cell layout		Hexagonal grid ISD = 200 m	Hexagnal grid ISD = 500 m
BS antenna height h_{bs}		10 m	25 m
UE location	Outdoor/Indoor	20% outdoor 80% indoor	20% outdoor 80% indoor
	LOS/NLOS	LOS and NLOS	LOS and NLOS
	Height h_{ue}	1.5-22.5 m	1.5-22.5 m
Minimum BS-to-UE distance		10 m	35 m
Parameters		**RMa**	**InH**
Cell layout		Hexagonal grid ISD = 1732 m	120 m×50 m×3 m ISD = 20 m
BS antenna height h_{bs}		35 m	3 m (ceiling)
UE location	Outdoor/Indoor	50% outdoor 50% indoor	Indoor
	LOS/NLOS	LOS and NLOS	LOS and NLOS
	Height h_{ue}	1.5 m	1 m
Minimum BS-to-UE distance		35 m	0 m

4.6.1 Beamforming cancellation

As explained earlier in Section 4.4, beamforming cancellation takes up a portion of the spatial DoFs available for the payload, so enhancing the SIC depth will inevitably reduce the sum rate. Figure 4.2 demonstrates the tradeoff between the two criteria. Figure 4.2(a) shows the variations of the sum rate and ASIC depth achieved by the NSP-MMSE precoding with projection matrix spanning of D dimension of the subspace. The projection steers the transmitted beams to the direction orthogonal to the SI channel, distorting the original MSME precoders and introducing precoding errors to the downlink payload. The trade-off between the precoding errors and SIC capability of the NSP-based methods is mathematically analyzed in [59]. D determines the projection degrees to the null space of the SI channel, and there is no projection when $D = 0$. Thus, increasing D can enhance the ASIC depth of precoders at the cost of communication efficiency and vice versa.

Figure 4.2(b) shows the achievable sum rate against the ASIC depth of the NSP-MMSE and our proposed eMMSE precoding. The tradeoff between the two criteria is made by adjusting the dimension of subspace D for the NSP-MMSE precoding as illustrated above, while it is made by adjusting the tolerable RSI power $\bar{\varepsilon}_{\text{rsi},g}$ for the eMMSE precoding. The results show that our proposed eMMSE precoding can reduce the sum

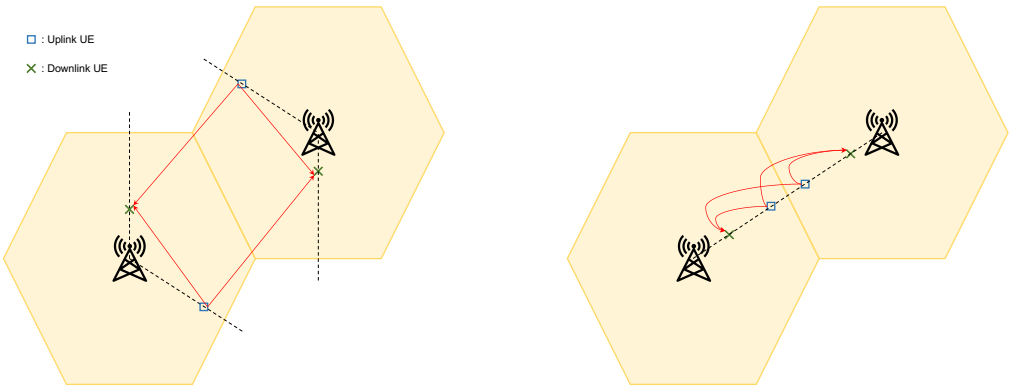
rate loss while enhancing the ASIC depth compared to the NSP-MMSE precoding. The reason is that the downlink rate maximization (i.e., downlink MSE minimization) and the ASIC depth enhancement (i.e., RSI power minimization) are jointly considered by eMMSE, while they are independently obtained by NSP-MMSE. Thus, eMMSE can utilize the spatial DoFs more efficiently to enhance the ASIC depth while minimizing the precoding errors for the downlink payload. However, the upper bound of achievable ASIC depth of eMMSE is lower than NSP-MMSE with the transmit power constraints.



(a) Sum rate and ASIC depth versus D (NSP-MMSE) (b) Achievable sum rate versus ASIC depth

FIGURE 4.2: Beamforming cancellation performance evaluation for the NSP-MMSE and eMMSE.

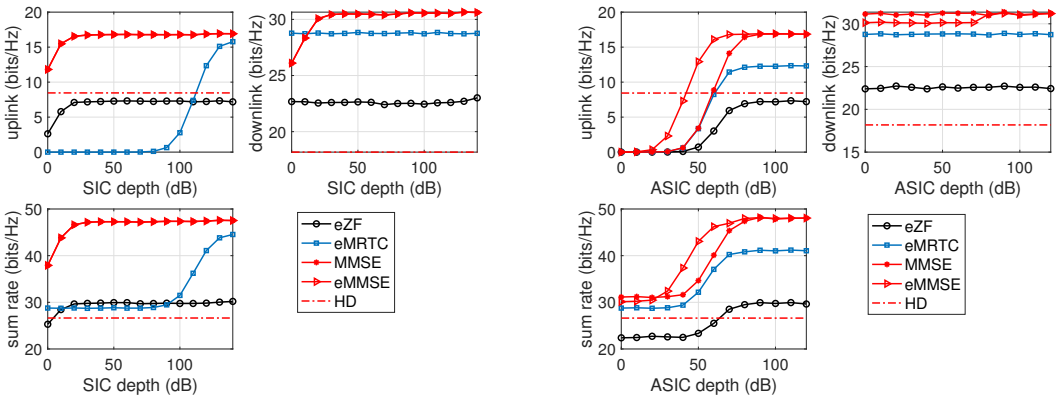
Then, we explore the SIC capability of the three linear beamforming schemes in MCMU networks. We start with a simple scenario with 2 cells, and each BS serves 2 users as Fig. 4.3, where the downlink users and uplink users are $\frac{4}{10}r$ and $\frac{6}{10}r$ away from the BS (r is the radius of the cell), respectively. The CCI is stronger in scenario 2 than in scenario 1, as the uplink and downlink users are closer. In order to show the effect of SIC, we use scenario 1 for simulations to minimize the effect of CCI. Figure 4.4(a) shows the achievable rate of this IBFD-MCMU network with increasing SIC depth realized at the IBFD base stations. The performance of the three proposed linear beamforming schemes is evaluated and compared to the typical MMSE beamforming. The achievable rate of its HD counterpart is achieved via the eMMSE beamforming. We assume an infinite dynamic range of receivers here. Under this condition, the SIC in the analog and digital domains have equal effects since the receiver will not be saturated. Thus, we set the tolerable RSI power to infinite for eMMSE under the infinite dynamic range assumption, where ASIC is not essential, and eMMSE will be equal to MMSE. The results



(a) Scenario 1: with weak CCI (b) Scenario 2: with strong CCI

FIGURE 4.3: Two scenarios of a 2-cell IBFD network with 2 uplink UEs and 2 downlink UEs with CCI highlighted.

demonstrate the ability of eMMSE (or MMSE) and eZF beamformers to suppress SI that full IBFD gain is obtained with only $\geqslant$ 30 dB of SIC realized by other techniques and the RSI is canceled by the beamformers. In contrast, the eMRTC beamformers cannot specifically suppress SI, although they can provide a little SIC depth when maximizing the target link due to the different directions of the target and SI channels. Therefore, it can enable uplink communications with $>$ 80 dB of SIC but requires $>$ 140 dB of SIC to achieve the full IBFD gain.



(a) infinite dynamic range of receivers (b) 60dB dynamic range of receivers

FIGURE 4.4: Achievable sum rate with different self-interference and transceiver conditions.

Then we consider practical receivers with a 60 dB dynamic range to demonstrate their ASIC capability. Figure 4.4(b) shows the achievable sum rate versus ASIC depth realized by other techniques, where a total of 140 dB SIC is assumed to be realized. It illustrates

the importance of ASIC that the desired IBFD gain cannot be realized without sufficient ASIC, even if a total of 140 dB of SIC is applied due to the receiver saturation. It also illustrates the advantages of our proposed eMMSE beamforming scheme that it can improve the uplink rate when ASIC is insufficient compared to other schemes. However, it decreases the downlink rate at the same time, which is caused by the downlink precoding errors, as explained above. Nevertheless, the downlink rate reduction is much less than the uplink rate improvement, so the sum rate is improved. Thus, our proposed eMMSE beamforming can reduce the requirement for the ASIC depth to obtain the desired IBFD gain. Numerical results show that 20dB lower ASIC depth is required by eMMSE, which can significantly reduce the complexity and cost of RF cancellers as demonstrated in Section 2.4, inspiring a low-complexity IBFD transceiver design.

4.6.2 Liner beamforming performance

In this section, we evaluate the performance of the three proposed linear beamforming schemes for IBFD-MCMU networks with multi-antenna users, providing guidance to the IBFD cellular network implementation.

Figure 4.5 shows the convergence behavior of the three algorithms, which plots the loss function variation with iterations. It should be noted that the three schemes have different loss functions due to different objectives, as given in Section 4.3. The eZF and eMRTC algorithms may not converge if they adopt the sum MSE or sum rate as the loss function, so they cannot guarantee the maximum throughput. the results show that the three algorithms converge to a constant after multiple iterations. In addition, it shows that eMRTC has a faster converging speed, i.e., requires fewer iterations to converge than the other two schemes, and eMMSE has the slowest converging speed. Combined with the complexity of a single iteration given in Section 4.4.4, eMRTC has the lowest time complexity, followed by eZF, and eMMSE has the highest complexity.

In addition to the SI, CCI will also significantly affect the performance of IBFD-MCMU networks. The CCI increases with decreasing cell size and increasing users. To investigate the effects of CCI on IBFD gain, we assume 80 dB of ASIC, 120 dB of SIC, -120 dB of transceiver HWIs factor, and 120 dB of channel uncertainty factor to minimize their effects. Figure 4.6 shows the achievable sum rate of the three beamforming schemes versus the inter-site distance (ISD), which reflects the cell size. Figure 4.6(a) and Figure

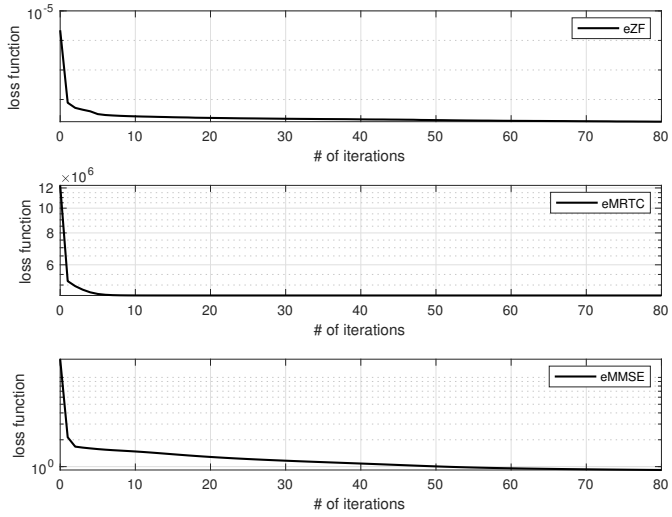
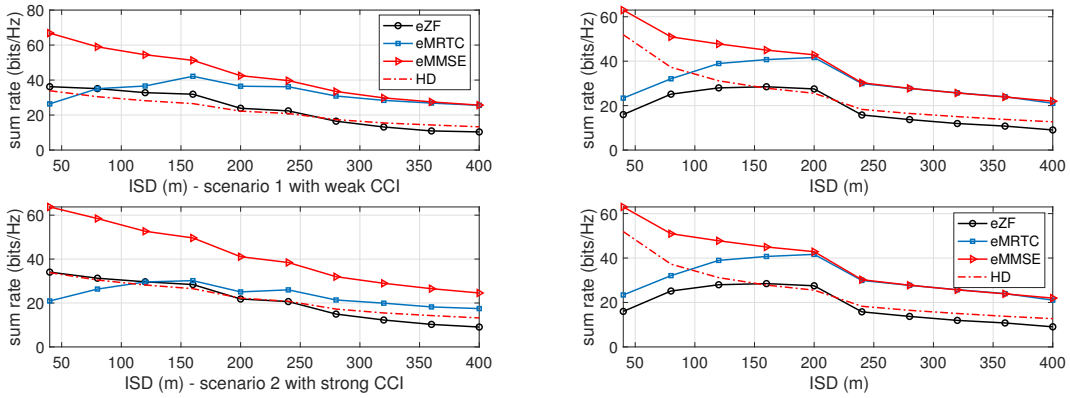


FIGURE 4.5: The convergence behavior of the three linear beamforming schemes.

4.6(b) show the results for 2-cell and 7-cell networks, respectively. We consider Scenario 1 and Scenario 2 in Figure 4.3 for 2-cell networks, which have weak and strong CCI, respectively. For 7-cell networks, we also consider two scenarios: 1) weaker CCI: one downlink user and one uplink user per cell; 1) stronger CCI: two downlink users and two uplink users per cell. It should be noted that both the received signal of interest and interference decreases with increasing ISD due to the long propagation distance. The



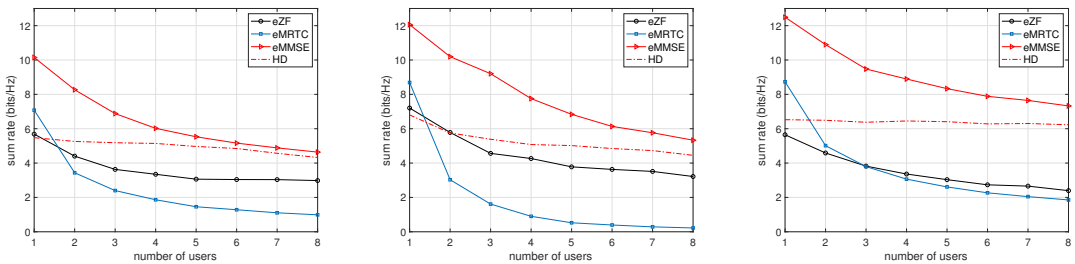
(a) 2-cell networks (b) 7-cell networks

FIGURE 4.6: Achievable sum rate variation with increasing inter-site distance.

results show that eMMSE always achieves the highest sum rate. The achievable sum rate of eMMSE decreases with enlarging cells since the interference could be effectively suppressed even if it is significant (i.e., with short ISD and more cells), so the decreased

signal of interest due to enlarged cells decreases the sum rate. The IBFD gain is compromised by strong CCI (i.e., with multiple micro cells). The eZF beamforming has a similar behavior as eMMSE with 2-cell networks, but it achieves a much lower sum rate than eMMSE due to the lack of signal of interest consideration. With 7 cells, eZF improves the achievable sum rate with enlarging cells at first (i.e., ISD$\leq$ 200 m), then the enlarging cell size decreases the sum rate. The results suggest that eZF cannot effectively suppress complicated interference (e.g., with multiple micro cells). The achievable sum rate of the eMRTC beamforming increases with increasing ISD at first since it improves the received power of the signal of interest to compensate for its increasing pathloss, but the sum rate decreases with enlarging cells. In scenario 1, the eMRTC beamforming approaches the performance of eMMSE when ISD$>$ 350 m, while it cannot achieve this in Scenario 2 due to the stronger CCI in this scenario. The eMRTC beamforming could approach the performance of eMMSE even with 7 cells as long as there is only a single uplink and downlink user within a cell. It can be concluded that with relatively large cell sizes, eMRTC could be employed for IBFD cellular networks from the perspective of energy efficiency since it achieves considerable throughput with much lower computation overheads. For micro cells, eMMSE could achieve multiple times higher throughput than the other two schemes, so it will be the first option even though it has high complexity.

Figure 4.7 shows the average achievable sum rate per user versus the number of downlink and uplink users per cell with different antenna array configurations. We use 2 cells here, and users are randomly located. It can be seen that the IBFD gain decreases



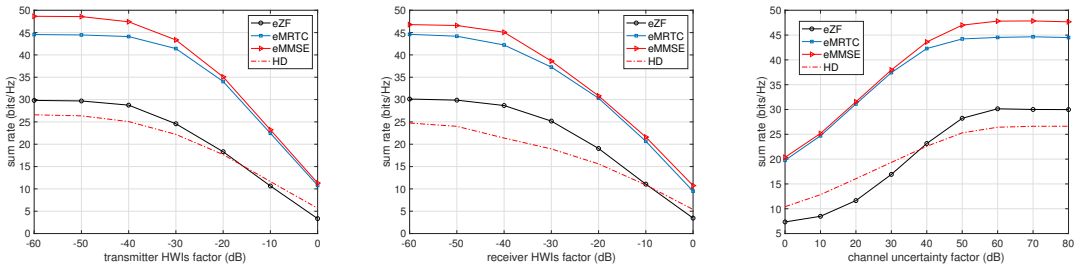
(a) $N_{bs} = M_{bs} = 16$, $N_{ue} = M_{ue} = 2$ (b) $N_{bs} = M_{bs} = 16$, $N_{ue} = M_{ue} = 4$ (c) $N_{bs} = M_{bs} = 64$, $N_{ue} = M_{ue} = 2$

FIGURE 4.7: Averaged achievable sum rate per user versus the number of users per cell with different MIMO configurations.

significantly with an increasing number of users due to the increased CCI. However, this can be compensated by more antennas at BSs or UEs that the sum rate improvement over HD increases from 7.2% with 2 antennas at UEs and 16 antennas at BSs to 19.8%

with 4 antennas at UEs and 16 antennas at BSs and 17.6% with 2 antennas at UEs and 64 antennas at BSs. It also shows that eMRTC outperforms eZF when there are only one downlink user and one uplink user, while eZF outperforms eMRTC with more users. This can be understood, the eZF minimizes interference, so it performs better with stronger interference. Increasing the number of antennas at users and base stations can improve the performance of eMMSE and eZF by providing more spatial DoFs to manage interference. The eMRTC beamforming benefits from more antennas at BSs, but more antennas at users may degrade its performance.

To explore the robustness of the three schemes, Figure 4.8 shows the achievable sum rate variation with increasing transceiver HWIs and channel uncertainty. It can be seen that



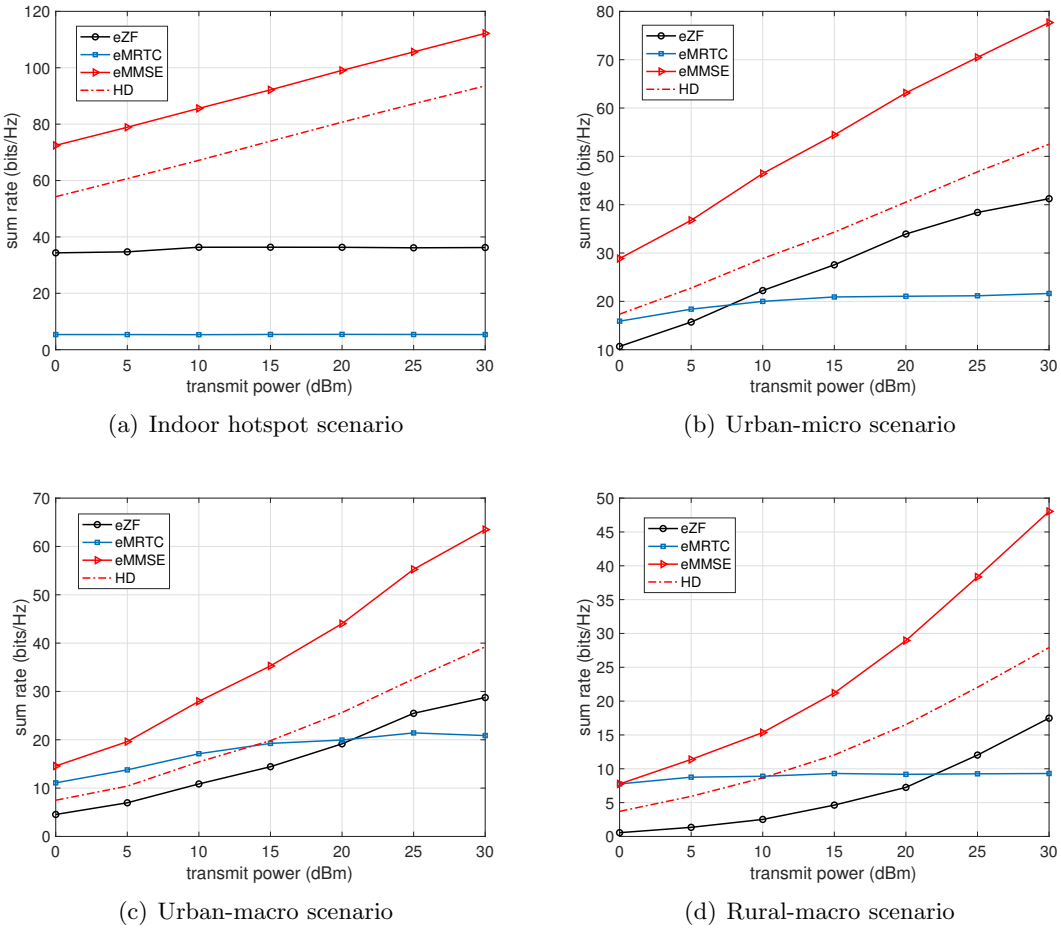
(a) Transmitter HWIs (b) Receiver HWIs (c) Channel uncertainty

FIGURE 4.8: Sum rate variation with increasing transceiver HWIs and channel uncertainty.

the transceiver HWIs and channel uncertainty decrease the achievable sum rate in general. However, they do not decrease the IBFD gain that 98.6%, 99.5%, and 95.2% higher sum rate of the HD radios are achieved by applying eMMSE beamforming with 0 dB of transmitter and receiver HWIs factor and 0 dB of channel uncertainty factor, respectively. The eMRTC beamforming shows stronger robustness to the channel uncertainty than eZF and eMMSE that 0 dB of channel uncertainty factor introduces 47.9% of sum rate degradation while this value increases to 55.1% and 71.1% for eMMSE and eZF, respectively. This is understandable as eMRTC uses much less CSI, and eMMSE includes the effects of channel uncertainty in the interference plus noise covariance matrices.

To provide more useful insights for piratical implementation, we evaluate their performance based on simulations under the four typical scenarios defined in the 3GPP document [67] as listed in Table 4.2. We consider a 2-cell network in simulations, where each base station serves 2 uplink users and 2 downlink users. The performance with more cells and users can be inferred from the above conclusions. Figure 4.9 shows the

achievable sum rate of various beamforming schemes versus the transmission power under the four scenarios. It shows that enhancing the transmit power can increase the



(a) Indoor hotspot scenario

(b) Urban-micro scenario

(c) Urban-macro scenario

(d) Rural-macro scenario

FIGURE 4.9: Performance evaluation of the linear beamformers under four 3GPP-specified scenarios.

system capacity for all four scenarios. However, IBFD gain decreases with increasing transmit power, especially for InH scenarios. For instance, the sum rate improvement of IBFD over HD decreases from 66.3% with 0dBm of transmit power to 47.9% with 30dBm of transmit power under UMi scenarios. As for the eMRTC beamforming scheme, it can achieve considerable performance and outperforms eZF with low transmit power except for indoor hotspot scenarios, but enhancing the transmit power does not have significant improvements. In indoor hotspot scenarios, enhancing the transmit power does not significantly increase the achievable sum rate of eMRTC and eZF beamforming schemes except for eMMSE. eZF has a similar behavior as eMMSE in the other three scenarios, but it achieves a much lower sum rate than eMMSE with reduced computation overheads.

4.6.3 Joint power allocation and beamforming

In this section, we evaluate the performance of the proposed JPAIM algorithm and compare it to the MWSR algorithm proposed in [54], which can maximize the system throughput of IBFD-MCMU networks. Since the joint power allocation and beamforming optimization problem is not jointly convex, the convergence behavior and achievable sum rate depend on the initial points, and the algorithms converge to the local minimum closest to the initial point. Therefore, we ran 1,000 times Monte Carlo simulations and calculated the average value to evaluate their performance to minimize the initial point effects. The simulation variables (e.g., channel matrices and initial points, etc.) are configured to be identical for the two algorithms in each realization for a fair comparison.

Figure 4.10 first compares the converging speed of the two algorithms. It shows that the JPAIM algorithm converges faster than the MWSR algorithm with both insufficient and sufficient ASIC depth realized. The JPAIM algorithm reduces the iteration time to converge by at least half compared to the MWSR Algorithm, which meets the condition analyzed in Section 4.5.3.1 to reduce the overall time complexity. Besides, both algorithms converge faster with insufficient ASIC realized due to the simplified interference.

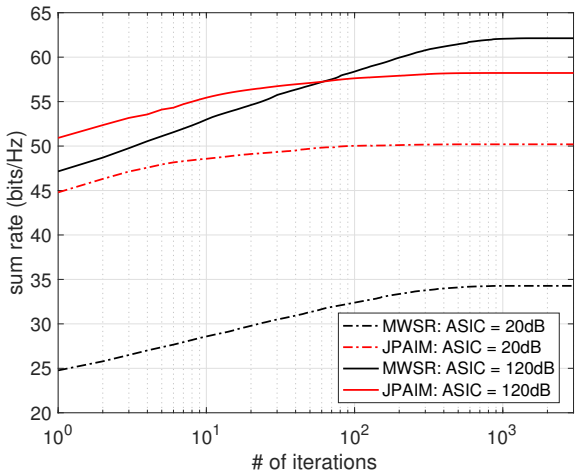


FIGURE 4.10: Convergence behavior (x-axis is in the log scale).

Since our proposed JPAIM algorithm is constructed based on the eMMSE beamforming, i.e., there are constraints on the received RSI power, it is supposed to outperform the existing MWSR algorithm with insufficient ASIC depth. Figure 4.11 shows the achievable sum rate of the JPAIM algorithm versus realized ASIC depths compared to the MWSR

algorithm. The results verify that our proposed JPAIM algorithm achieves a higher sum rate than MWSR when the ASIC depth realized by other techniques is insufficient (i.e., < 60 dB). This is attractive since other ASIC techniques have incredibly high complexity and cost in MIMO systems. It is revealed in the literature that active antenna isolation, which has feasible implementation complexity, can achieve an ASIC depth of around 30 dB [10]. With 30 dB of antenna isolation, the JPAIM algorithm can improve the spectral efficiency by 42.9% in IBFD compared to HD due to its enhanced ASIC capability. In contrast, the MWSR algorithm must require RF cancellers to provide a deeper ASIC depth (i.e., 50 dB) to achieve this level of IBFD gain due to its limited ASIC capability.

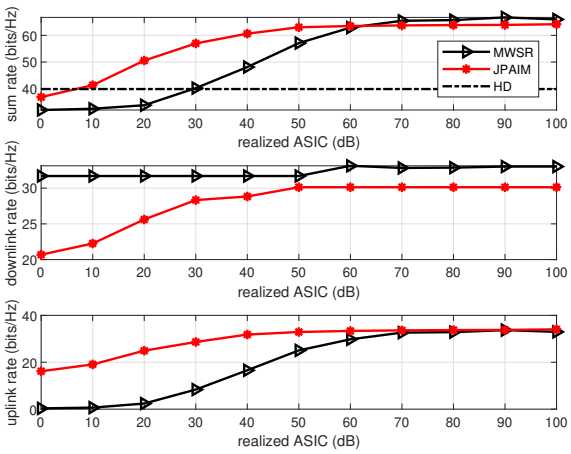
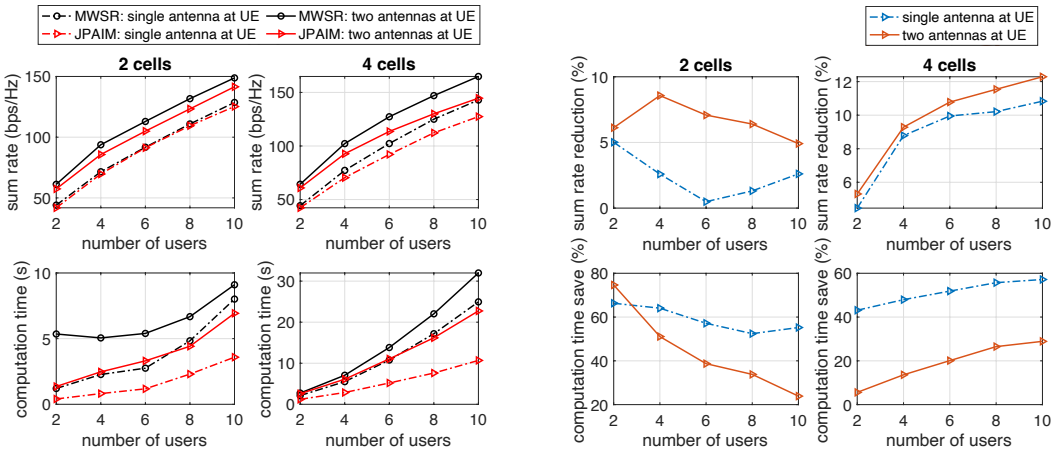


FIGURE 4.11: Achievable spectral efficiency versus realized ASIC depth.

It can be seen from Figure 4.11 that MWSR achieves a higher sum rate (downlink rate, in fact) than JPAIM with sufficient ASIC depths (> 60 dB). The performance degradation comes from two factors. The first one is the downlink precoding errors due to the ASIC enhancement, as explained in the eMMSE section. Another factor lies in the fact that the JPAIM algorithm describes the transmit power via scalars, while the MWSR algorithm describes it in a matrix form in nature, thus allowing a more delicate power allocation (e.g., at the symbol level). In fact, the performance loss mainly stems from the second factor because the Lagrange multiplier associated with the RSI constraints will approach zero when the ASIC depth is sufficient, so the precoding errors are negligible. Figure 4.12 shows the performance difference between the two algorithms in terms of sum rate and average computation time of a realization with different network sizes, i.e., different numbers of cells, users, and antennas at users. The results show that JPAIM achieves a

sum rate close to the one achieved by MWSR, but it takes much less computation time. The sum rate loss generally increases with increasing numbers of cells and users, and it is more significant with multi-antenna users. For single-antenna users, JPAIM saves at least 40% of the computation time at the cost of $< 10\%$ sum rate reduction for most cases. Although the benefits are compromised with two-antenna users, the computation time saved is always higher than the sum rate reduction.



(a) Sum rate and computation time versus the number of users

(b) Sum rate reduction and computation time save percentages

FIGURE 4.12: Performance comparison of JPAIM and MWSR in terms of spectral efficiency and time complexity.

Although the scalar power coefficient causes sum rate degradation, it also brings benefits in terms of time efficiency (i.e., = sum rate/computation time), which is shown in Figure 4.13. The results illustrate that JPAIM achieves higher time efficiency than MWSR under all conditions. This indicates that JPAIM has a lower time complexity in practical implementations than MWSR at the cost of an acceptable sum rate loss. We ran the simulations on the same hardware and platform for a fair comparison. The benefit of JPAIM is significant with single-antenna users, which is meaningful for practical cellular network deployment since many user terminals still use a single antenna.

Finally, the robustness of the proposed algorithm is evaluated. Figure 4.14 shows the achievable sum rate against increasing channel error factors ϱ under various transceiver conditions, i.e., transceiver distortion factors κ and β, which demonstrates the robustness of JPAIM to channel uncertainty. It shows that significant channel uncertainty has less

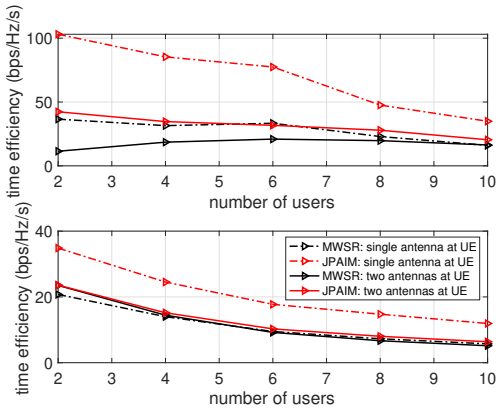


FIGURE 4.13: Time efficiency comparison of the proposed JPAIM and existing MWSR algorithms.

impact on JPAIM than MWSR[8]. The benefits come from the fact that the effects of channel uncertainty are included in our algorithm as detailed in Appendix 4.8. In addition, it is also robust to the transceiver hardware impairments that reducing the dynamic range of transceivers from 70 dB to 50 dB has no noticeable effect on the achievable sum rate. Although a smaller dynamic range (e.g., 30 dB) of transceivers reduces the achievable sum rate, it is acceptable as the reduction is less than 10%.

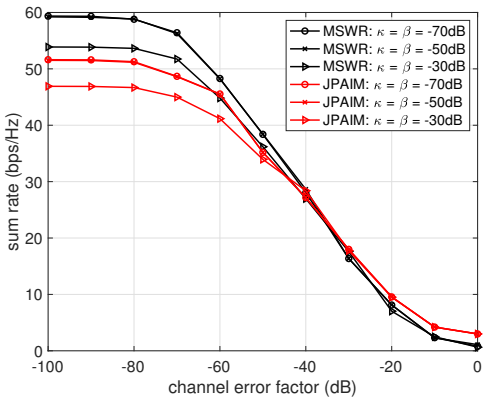


FIGURE 4.14: Robustness of the JPAIM and MWSR algorithm to channel uncertainty and transceiver HWIs.

4.7 Conclusions

In this chapter, we considered a multi-cell multi-user network with multi-antenna users and IBFD base stations, which could be the future cellular network, and studied the

[8]Transceiver HWIs are considered by the MWSR algorithm as in [54], but the channel uncertainty is not.

beamforming schemes accordingly with practical imperfections. We enhanced the conventional ZF and MRTC beamforming schemes to deal with such networks, which could show reasonable performance under specific conditions. The MMSE scheme could maximize the throughput of the MCMU network with MMSE receivers. The eMRTC scheme has considerable performance with trivial interference (e.g., large cells and single-user network) that it could approach the performance of MMSE. In contrast, the eZF scheme outperforms eMRTC with increasing interference (e.g., microcells and more users). The MMSE scheme benefits from large-scale antenna arrays at both base stations and users, while eMRTC benefits only from more antennas at base stations. The increasing transceiver HWIs and channel uncertainty decrease the achievable sum rate of all schemes but do not decrease the IBFD gain. In 3GPP-specified scenarios, the eMMSE scheme could achieve a much higher sum rate than other schemes, and it can be further improved with higher transmit power. In contrast, the performance of the eMRTC scheme does not significantly improve with high transmit powers, especially for indoor hotspot scenarios. Due to the IBFD operation at the base stations, sufficient ASIC depths are the basis of all beamforming schemes to achieve the desired IBFD gain in terms of spectral efficiency. Since conventional RF cancellers have extremely high complexity in MIMO systems, especially in future massive MIMO systems, we exploited beamforming techniques for SIC. For point-to-point systems, forming a precoder and combiner by choosing orthogonal singular vectors of the intended communication channel can mitigate the SI at the IBFD transceiver, and it can prevent receiver saturation by forming the precoder with the singular vectors associated with the smallest singular values. However, this method cannot be extended to multi-user networks where there are multiple communication channels present. In this case, the null-space projection can be employed at the cost of precoding errors, and the tradeoff between the ASIC depth and precoding errors can be made by the projection matrix. However, this method does not crossly design the two objectives, leading to a sub-optimal solution, Therefore, we proposed a joint design by adding the RSI power constraints to the normal MMSE problem, which can enhance the ASIC depth while minimizing the precoding errors. Based on this design, we further considered the power allocation issue and proposed a novel joint power allocation and beamforming design. The formulated problem was converted to an MSE plus RSI minimization problem and then decomposed into two sub-problems to be solved. The closed-form solutions to the sub-problems were derived, and the overall solution was obtained by an iterative algorithm. Benefiting from the

enhanced ASIC depth, our algorithm could achieve 42.9% of the IBFD gain in terms of spectral efficiency in a 3GPP-specified cellular network with feasible antenna isolation only, inspiring a low-cost but efficient IBFD cellular network implementation. In addition to the ASIC enhancement, our algorithm can significantly reduce the computation overheads of the existing method at the cost of acceptable throughput loss, and it is robust to channel uncertainty and transceiver hardware impairments. It saves at least 40% of the computation time at the cost of $< 10\%$ sum rate reduction with single-antenna users.

Appendix 4.8 Effects of channel uncertainty

In this appendix, we give the statistics of the errors caused by the channel uncertainty at the base station and user equipment. For any two specific nodes A and B, we can calculate the covariance of the associated errors as

$$\begin{aligned} Cov\left(\boldsymbol{\Delta}_{B,A}\mathbf{x}_A\right) &= \mathbb{E}\left\{\boldsymbol{\Delta}_{B,A}\mathbf{x}_A\mathbf{x}_A^{\dagger}\boldsymbol{\Delta}_{B,A}^{\dagger}\right\} \\ &= \mathbb{E}_{\boldsymbol{\Delta}_{B,A}}\left\{\boldsymbol{\Delta}_{B,A}\mathbf{T}_A\boldsymbol{\Delta}_{B,A}^{\dagger}\right\} \overset{(a)}{=} \tilde{\sigma}_{B,A}^2 tr\left(\mathbf{T}_A\right)\mathbf{I}, \end{aligned} \tag{4.111}$$

where (a) comes from following derivation

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{\Delta}_{B,A}}\left\{\left[\boldsymbol{\Delta}_{B,A}\mathbf{T}_A\boldsymbol{\Delta}_{B,A}^{\dagger}\right]_{m,n}\right\} \\ &= \mathbb{E}_{\Delta}\left\{\sum_{k=1}^{N_\Delta}\sum_{j=1}^{N_\Delta}\Delta_{mk}T_{kj}\Delta_{nj}^{\dagger}\right\} \\ &= \sum_{k=1}^{N_\Delta}\sum_{j=1}^{N_\Delta}\mathbb{E}_{\Delta}\left\{\Delta_{mk}T_{kj}\Delta_{mk}^{\dagger}\right\}\delta_{mn}\delta_{kj} \\ &\overset{(c)}{=} \sum_{k=1}^{N_\Delta}\mathbb{E}_{\Delta}\left|\Delta_{mk}\right|^2 T_{kk}\delta_{mn} = \tilde{\sigma}_{B,A}^2 tr\left(\mathbf{T}_A\right)\delta_{mn}, \end{aligned} \tag{4.112}$$

where N_Δ denotes the number of columns of matrix $\boldsymbol{\Delta}_{B,A}$, T_{kj} denotes the element at the k^{th} row and j^{th} column of matrix $\mathbf{T}_A$, and Δ_{mk} denotes the element at the m^{th} row and k^{th} column of matrix $\boldsymbol{\Delta}_{B,A}$. δ represents the correlation coefficient, and $\delta_{kj} = 0, \forall\, k \neq j$, $\delta_{kk} = 1, \forall\, k$, $\delta_{mn} = 0, \forall\, m \neq n$, and $\delta_{mm} = 1, \forall\, m$. Thus, the covariance matrices of the

errors are given as

$$Cov\left(\mathbf{e}_g\right) = \underbrace{\sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\tilde{\sigma}_{g,i_j^u}^2 tr\left(\mathbf{T}_{i_j^u}\right) + \sum_{j\neq g}^{G}\tilde{\sigma}_{g,j}^2 tr\left(\mathbf{T}_j\right)}_{\hat{\sigma}_g^2}\mathbf{I}_{M_{bs}}, \tag{4.113}$$

$$Cov\left(\mathbf{e}_{k_g^d}\right) = \underbrace{\sum_{j=1}^{G}\sum_{k=1}^{K_j^d}\tilde{\sigma}_{k_g^d,j}^2 tr\left(\mathbf{T}_{i_j^d}\right) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\tilde{\sigma}_{k_g^d,i_j^u}^2 tr\left(\mathbf{T}_{i_j^u}\right)}_{\hat{\sigma}_{k_g^d}^2}\mathbf{I}_{M_{ue}}. \tag{4.114}$$

Appendix 4.9 Covariance matrices of the received signals

In this appendix, we derive the covariance matrices of the received signals $\mathbf{y}_g$ and $\mathbf{y}_{k_g^d}$, which can be written as

$$\begin{aligned}\mathbf{C}_g = \mathbb{E}\left\{\mathbf{y}_g\mathbf{y}_g^\dagger\right\} &= \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{g,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{g,i_j^u}^\dagger + \sum_{j=1}^{G}\hat{\mathbf{H}}_{g,j}\mathbf{T}_j\hat{\mathbf{H}}_{g,j}^\dagger \\ &+ \beta_{bs}\mathcal{D}\left(\sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{g,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{g,i_j^u}^\dagger + \sum_{j=1}^{G}\hat{\mathbf{H}}_{g,j}\mathbf{T}_j\hat{\mathbf{H}}_{g,j}^\dagger\right) + (\sigma_g^2 + \tilde{\sigma}_g^2)\mathbf{I}_{M_{bs}}.\end{aligned} \tag{4.115}$$

$$\begin{aligned}\mathbf{C}_{k_g^d} = \mathbb{E}\left\{\mathbf{y}_{k_g^d}\mathbf{y}_{k_g^d}^\dagger\right\} &= \sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\hat{\mathbf{H}}_{k_g^d,j}\mathbf{T}_{i_j^d}\hat{\mathbf{H}}_{k_g^d,j}^\dagger + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}^\dagger \\ &+ \beta_{ue}\mathcal{D}\left(\sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\hat{\mathbf{H}}_{k_g^d,j}\mathbf{T}_{i_j^d}\hat{\mathbf{H}}_{k_g^d,j}^\dagger + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}^\dagger\right) + (\sigma_{k_g^d}^2 + \tilde{\sigma}_{k_g^d}^2)\mathbf{I}_{M_{ue}}.\end{aligned} \tag{4.116}$$

Appendix 4.10 MSE expressions

In this appendix, we derive the MSE expressions using the covariance matrices given in Appendix 4.9. The MSE of a single user is a function of the beamforming matrices and allocated power of all other users, i.e., $\alpha_{k_g^d}, \gamma_{k_g^u}, \mathbf{V}_{k_g^d}, \mathbf{V}_{k_g^u}, \mathbf{U}_{k_g^d}$, and $\mathbf{U}_{k_g^u}\ \forall\ k, g$. For

simplicity, we omit the subscripts of these variables and denote the MSE as

$$\begin{aligned}
&\varepsilon_{k_g^d}(\alpha,\gamma,\mathbf{V},\mathbf{U}) = \mathbb{E}\left\{\left\|\mathbf{s}_{k_g^d} - \mathbf{U}_{k_g^d}^{\dagger}\mathbf{y}_{k_g^d}\right\|^2\right\} \\
&= tr\left(\mathbf{U}_{k_g^d}^{\dagger}\left(\sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\hat{\mathbf{H}}_{k_g^d,j}\mathbf{T}_{i_j^d}\hat{\mathbf{H}}_{k_g^d,j}^{\dagger} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}^{\dagger}\right)\mathbf{U}_{k_g^d}\right) \\
&\quad + \beta_{ue}tr\left(\mathbf{U}_{k_g^d}^{\dagger}\mathcal{D}\left(\sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\hat{\mathbf{H}}_{k_g^d,j}\mathbf{T}_{i_j^d}\hat{\mathbf{H}}_{k_g^d,j}^{\dagger} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{k_g^d,i_j^u}^{\dagger}\right)\mathbf{U}_{k_g^d}\right) \\
&\quad + \sigma_{k_g^d}^2 tr\left(\mathbf{U}_{k_g^d}^{\dagger}\mathbf{U}_{k_g^d}\right) - 2\alpha_{k_g^d}\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^d}^{\dagger}\hat{\mathbf{H}}_{k_g^d,g}\mathbf{V}_{k_g^d}\right)\right\} + b_d,
\end{aligned} \tag{4.117}$$

$$\begin{aligned}
&\varepsilon_{k_g^u}(\alpha,\gamma,\mathbf{V},\mathbf{U}) = \mathbb{E}\left\{\left\|\mathbf{s}_{k_g^u} - \mathbf{U}_{k_g^u}^{\dagger}\mathbf{y}_g\right\|^2\right\} \\
&= tr\left(\mathbf{U}_{k_g^u}^{\dagger}\left(\sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{g,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{g,i_j^u}^{\dagger} + \sum_{j=1}^{G}\hat{\mathbf{H}}_{g,j}\mathbf{T}_j\hat{\mathbf{H}}_{g,j}^{\dagger}\right)\mathbf{U}_{k_g^u}\right) \\
&\quad + \beta_{bs}tr\left(\mathbf{U}_{k_g^u}^{\dagger}\mathcal{D}\left(\sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{g,i_j^u}\mathbf{T}_{i_j^u}\hat{\mathbf{H}}_{g,i_j^u}^{\dagger} + \sum_{j=1}^{G}\hat{\mathbf{H}}_{g,j}\mathbf{T}_j\hat{\mathbf{H}}_{g,j}^{\dagger}\right)\mathbf{U}_{k_g^u}\right) \\
&\quad + \sigma_g^2 tr\left(\mathbf{U}_{k_g^u}^{\dagger}\mathbf{U}_{k_g^u}\right) - 2\gamma_{k_g^u}\mathcal{R}\left\{tr\left(\mathbf{U}_{k_g^u}^{\dagger}\hat{\mathbf{H}}_{g,k_g^u}\mathbf{V}_{k_g^u}\right)\right\} + b_u.
\end{aligned} \tag{4.118}$$

Remarking: For fixed full transmit power cases, the MSE expressions can be written with $\alpha_{k_g^d}$ and $\gamma_{k_g^u}$ being replaced by $\sqrt{\frac{P_{bs}}{K_g^d}}$ and $\sqrt{P_{ue}}$, respectively.

Appendix 4.11 Expressions of $\mathbf{\Omega}_g$ and $\mathbf{\Omega}_{k_g^u}$

In this appendix, we give detailed expressions of matrices $\mathbf{\Omega}_g$ and $\mathbf{\Omega}_{k_g^u}$ used in the MMSE precoder solutions.

$$\begin{aligned}
\mathbf{\Omega}_g = &\sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\hat{\mathbf{H}}_{i_j^d,g}^{\dagger}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger} + \mathcal{D}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,g} \\
&+ \kappa_{bs}\mathcal{D}\left(\hat{\mathbf{H}}_{i_j^d,g}^{\dagger}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger} + \mathcal{D}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,g}\right) \\
&+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{j,g}^{\dagger}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger} + \mathcal{D}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,g} \\
&+ \kappa_{bs}\mathcal{D}\left(\hat{\mathbf{H}}_{j,g}^{\dagger}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger} + \mathcal{D}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,g}\right),
\end{aligned} \tag{4.119}$$

$$\begin{aligned}\boldsymbol{\Omega}_{k_g^u} = &\sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\hat{\mathbf{H}}_{i_j^d,k_g^u}^{\dagger}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger}+\mathcal{D}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,k_g^u}\\&+\kappa_{ue}\mathcal{D}\left(\hat{\mathbf{H}}_{i_j^d,k_g^u}^{\dagger}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger}+\mathcal{D}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,k_g^u}\right)\\&+\sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\hat{\mathbf{H}}_{j,k_g^u}^{\dagger}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger}+\mathcal{D}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,k_g^u}\\&+\kappa_{ue}\mathcal{D}\left(\hat{\mathbf{H}}_{j,k_g^u}^{\dagger}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger}+\mathcal{D}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,k_g^u}\right).\end{aligned} \tag{4.120}$$

Appendix 4.12 Proof of Lemma 4.1

The only difference between the two formulations (i.e., minimization problems $(P.1)$ and $(P.2)$) is that the RSI power-related terms appear in different places. The RSI appears in the constraints of problem $(P.1)$, and its included in the objective function of problem $(P.2)$. Since the RSI power is a function of only downlink precoders, this difference does not affect the solution to other variables. The Lagrange functions of the two problems with respect to the downlink precoders $\mathbf{V}_{k_g^d}$ can be given as

$$\mathcal{L}_{P_1} = \boldsymbol{\Sigma}_L\left(\mathbf{V}\right)+\sum_{g=1}^{G}\lambda_g\left(\epsilon_{\mathrm{rsi},g}\left(\mathbf{V}_g\right)-\bar{\epsilon}_{\mathrm{rsi},g}\right), \tag{4.121}$$

$$\mathcal{L}_{P_2} = \boldsymbol{\Sigma}_L\left(\mathbf{V}\right)+\sum_{g=1}^{G}\nu_g\epsilon_{\mathrm{rsi},g}\left(\mathbf{V}_g\right), \tag{4.122}$$

where λ_g is the Lagrange multiplier associated with the constraint $(C.3)$ of problem $(P.1)$, and $\boldsymbol{\Sigma}_L\left(\mathbf{V}\right)$ denotes the sum MSE and power constraint-related terms given as

$$\boldsymbol{\Sigma}_L\left(\mathbf{V}\right)=\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\varepsilon_{k_g^d}\left(\mathbf{V}\right)+\sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\varepsilon_{k_g^u}\left(\mathbf{V}\right)+\sum_{g=1}^{G}\varpi_g\left(\sum_{k=1}^{K_g^d}\alpha_{k_g^d}^2 tr\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)-P_{bs}\right), \tag{4.123}$$

where ϖ_g is the Lagrange multiplier associated with the transmit power constraints of base stations. The optimal solutions to the downlink precoder are obtained by deriving the Lagrange functions with respect to $\mathbf{V}_{k_g^d}$ and set the derivatives to zero, which will be identical if $\nu_g = \lambda_g$. This means optimization problems $(P.1)$ and $(P.2)$ share the same solutions with $\nu_g = \lambda_g$, so they are equivalent. It should be noted that the optimal value of the Lagrange multiplier λ_g is determined by the tolerable RSI power $\bar{\epsilon}_{\mathrm{rsi},g}$. More specifically, λ_g increases with decreasing $\bar{\epsilon}_{\mathrm{rsi},g}$.

Appendix 4.13 Expressions of $\chi_{k_g^d}$ and $\chi_{k_g^u}$

In this appendix, we give detailed expressions of $\chi_{k_g^d}$ and $\chi_{k_g^u}$, which are written as

$$\begin{aligned}\chi_{k_g^d} = &\sum_{j=1}^{G}\sum_{i=1}^{K_j^d} tr\Bigg(\mathbf{U}_{i_j^d}^{\dagger}\hat{\mathbf{H}}_{i_j^d,g}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}+\kappa_{bs}\mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,g}^{\dagger}\mathbf{U}_{i_j^d}\\ &+\beta_{bs}\mathbf{U}_{i_j^d}^{\dagger}\mathcal{D}\left(\hat{\mathbf{H}}_{i_j^d,g}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}+\kappa_{bs}\mathcal{D}\left(\mathbf{V}_{k_g^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,g}^{\dagger}\right)\mathbf{U}_{i_j^d}\Bigg),\\ &+\sum_{j=1}^{G}\sum_{i=1}^{K_j^u} tr\Bigg(\mathbf{U}_{i_j^u}^{\dagger}\hat{\mathbf{H}}_{j,g}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}+\kappa_{bs}\mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,g}^{\dagger}\mathbf{U}_{i_j^u}\\ &+\beta_{ue}\mathbf{U}_{i_j^u}^{\dagger}\mathcal{D}\left(\hat{\mathbf{H}}_{j,g}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}+\kappa_{bs}\mathcal{D}\left(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,g}^{\dagger}\right)\mathbf{U}_{i_j^u}\Bigg),\end{aligned} \tag{4.124}$$

$$\begin{aligned}\chi_{k_g^u} = &\sum_{j=1}^{G}\sum_{i=1}^{K_j^d} tr\Bigg(\mathbf{U}_{i_j^d}^{\dagger}\hat{\mathbf{H}}_{i_j^d,k_g^u}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}+\kappa_{ue}\mathcal{D}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,k_g^u}^{\dagger}\mathbf{U}_{i_j^d}\\ &+\beta_{bs}\mathbf{U}_{i_j^d}^{\dagger}\mathcal{D}\left(\hat{\mathbf{H}}_{i_j^d,k_g^u}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}+\kappa_{ue}\mathcal{D}\left(\mathbf{V}_{k_g^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{i_j^d,k_g^u}^{\dagger}\right)\mathbf{U}_{i_j^u}\Bigg),\\ &+\sum_{j=1}^{G}\sum_{i=1}^{K_j^u} tr\Bigg(\mathbf{U}_{i_j^u}^{\dagger}\hat{\mathbf{H}}_{j,k_g^u}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}+\kappa_{ue}\mathcal{D}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,k_g^u}^{\dagger}\mathbf{U}_{i_j^u}\\ &+\beta_{ue}\mathbf{U}_{i_j^u}^{\dagger}\mathcal{D}\left(\hat{\mathbf{H}}_{j,k_g^u}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}+\kappa_{ue}\mathcal{D}\left(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^{\dagger}\right)\right)\hat{\mathbf{H}}_{j,k_g^u}^{\dagger}\right)\mathbf{U}_{i_j^u}\Bigg).\end{aligned} \tag{4.125}$$

Chapter 5

Physical Layer Based Secret Key Generation

5.1 Introduction

Wireless communication networks are vulnerable to eavesdropping due to the broadcast nature of wireless channels. The mobility and heterogeneity of users and limited resources (e.g., power supply and computing capacity) make it challenging to protect wireless transmission from the physical layer. Thus, the security of wireless networks is conventionally ensured by encryption schemes, where the message is encrypted with a secure key. Classical encryption schemes are applied in the upper layers of the protocol, which achieves computational complexity-based security. These schemes require key distribution by a secure management center and may become ineffective in the future due to the rapidly growing computational capacity. In contrast, physical layer security (PLS) can achieve information-theoretical security without aid from other users or infrastructures. Thus, it has attracted increasing attention from the security community.

Physical layer-based secret key generation (PHY-SKG) schemes utilize the reciprocity and unpredictable randomness of wireless channels to generate the key. Eavesdroppers located more than one half-wavelength away from legitimate users will experience uncorrelated fading. This spatial decorrelation assumption guarantees the security of

Work in this chapter is publishing in IEEE Journal on Selected Areas in Communications Third Quarter 2023 [68], with a preliminary version presented at IEEE ICC May 2023 [69].

the generated key, which is claimed in most related papers [70–72]. In the literature, TDD is usually considered for PHY-SKG schemes, where channel reciprocity is assumed to be held. Various characteristics are exploited to generate the key in the literature, which can be categorized into received signal strength (RSS)-based schemes and CSI-based schemes. The secret key capacity (SKC) of the two methods is compared in [73], which reveals that the RSS-based SKG is seriously penalized compared to the CSI-based scheme. However, CSI-based schemes need to estimate the CSI from the observations and quantize specific parameters of the channel (e.g., channel gain or phase), resulting in high implementation complexity. In [74], a 3D spatial angle-based SKG scheme is proposed for FDD systems based on the assumption that reciprocity exists in terms of the angle of departure and angle of arrival of each dual path.

Many studies have experimentally demonstrated the feasibility of PHY-SKG. In [75], authors carry out a complete RSS-based SKG scheme consisting of several preprocessing techniques on real-world wireless sensor networks and IoT devices, achieving 10%-20% key disagreement rates (KDR) and 1h of updating interval. However, studies reveal that the asymmetric observations due to non-simultaneous measurements in HD systems and the inherent transceiver hardware impairments reduce the PHY-SKG performance [70, 76]. The imperfect channel reciprocity fundamentally limits the secret key capacity (SKC), which is the maximum key generation rate (KGR) at which the secret key can be generated reliably, securely, and uniformly [77]. While a high KGR is necessary for the real-time implementation of PHY-SKG; otherwise, the same key will be used for a long duration, compromising security. These effects are formulated and analyzed by deriving the SKC for the PHY-SKG scheme with HWIs in [76]. It is revealed that the non-reciprocal measurements, which will reduce KGR and increase KDR without proper processing, raise challenges in practical implementation. To combat this challenge, various pre-processing techniques are studied. For instance, a Savitzky–Golay filter is used in [75] to reduce noise while maintaining the shape and height of bandwidth fluctuation caused by the multipath environment of RSS samples. Furthermore, studies also explore neural networks to combat non-reciprocal issues in the ear of artificial intelligence. In [70], two neural networks are trained to generate the key by mapping the non-reciprocal measurements to highly correlated latent vectors at the two legitimate users, yielding high KGR and low KDR with realistic channels and hardware HWIs.

In addition to the probing errors, the coherence time of the channel is also a critical

factor affecting the KGR [78]. For instance, the mobility of users is limited for WLAN, resulting in a relatively stable channel. In this case, a relatively low probing rate is needed to guarantee randomness, yielding a slow KGR. To address this issue, the authors of [79] propose a cooperative SKG scheme to improve the KGR by exploiting the random channels associated with relay nodes in the network as additional random sources, showing a multiplexing gain in the KGR scales linearly with the number of relays. With the recent employment of MIMO systems, studies also investigate their advantages in PLS. The SKC of MIMO-OFDM systems is analyzed in [80], which shows that more randomness is provided by antenna arrays for key generation. In addition to the KGR improvement, studies also reveal the benefits of MIMO systems in security. In [81], a practical PHY-SKG scheme based on precoding matrix indices is proposed, where rotated reference signals are utilized to enhance security. This scheme guarantees the usage of full MIMO gain and can be employed in practice without reconciliation and privacy amplification. Besides, the antenna arrangements have dramatic impacts on the MIMO channel between legitimate users, which poses a greater challenge to the eavesdropper. An optimal beamforming design is proposed in [82] to reduce the pilot overhead of the reciprocal CSI acquisition. However, the realistic correlated MIMO channel may decrease the security due to the correlation between the generated key. To address this issue, a decorrelation vector is utilized in [83] to generate a uniformly-distributed bit sequence with the correlated MIMO channel.

There is always an unavoidable flaw in HD systems, that is, the non-simultaneous probing, which will be more obvious with increasing Doppler shifts. The simultaneous transmission and reception nature of IBFD could provide a solution to this problem, which has been illustrated by theoretical analyses. It is analyzed in [84] that the key rate with IBFD probing is generally higher than its HD counterpart in the high-SNR regime at the cost of SIC. However, the RSI and the overheads of SIC could lead to the IBFD key rate being less than its HD counterpart with low SNR and highly correlated channel observations. Numerical results in [85] illustrate that the KDR drops at least 60%, and the KGR increases to up to 1.8 times with increased entropy of keys by IBFD probing than existing HD-based schemes. Authors in [77] compare the secret key rates with the rate-limited public channel by introducing the reconciliation function in IBFD and HD radios. Simulation results illustrate that IBFD improves the secret key rate and has

negative effects on the eavesdropper's capacity. Although there are many studies exploring IBFD for PHY-SKG, a comprehensive derivation and analysis of the advantages and limits of the IBFD-based SKG are still lacking.

The main contribution of our work is that we give a deep insight into the advantages of IBFD radios on the PLS. We consider an IBFD-MIMOME network which consists of two full-duplex multi-antenna users and an arbitrarily located multi-antenna eavesdropper. The secrecy of such a network is generally analyzed in [86], while we focus on a CFR-based SKG scheme for practical implementation and derive its SKC with practical imperfections described by measurable metrics. In addition to the commonly considered simultaneous probing, we also reveal the advantage of IBFD in reducing errors due to more time-frequency resources. Furthermore, we analyze the properties of collected CFR samples and give the necessary preprocessing methods.

The rest of this chapter is organized as follows. In Section 5.2, we give preliminaries for the SKG scheme, including the attacker model, transceiver imperfections, and key performance indicators. Then, we derive and analyze the SKC by studying the probing errors and considering the eavesdropper's locations in Section 5.3. In Section 5.4, we give a detailed procedure of the SKG protocol by analyzing the sample properties and giving processing accordingly. Both numerical results for SKC evaluation and 3GPP-based SKG simulations and National Institute of Standards and Technology (NIST) test results are given in Section 5.5, followed by conclusions given in Section 5.6.

5.2 Preliminaries

5.2.1 System model

We consider an OFDM-MIMO single-eavesdropper system where two legitimate users (i.e., Alice and Bob) communicate in the presence of a passive eavesdropper (i.e., Eve), as depicted in Figure 5.1. We assume that the communicating nodes have the same number of transmitting and receiving antennas for simplicity, and Alice and Bob are equipped with N_A and N_B transmitting and receiving antennas, respectively. We consider half-wavelength antenna arrays so that the entries of the MIMO channel matrix are

independent of each other [87]. For reconciliation purposes, there is a noiseless public channel between legitimate users to exchange information.

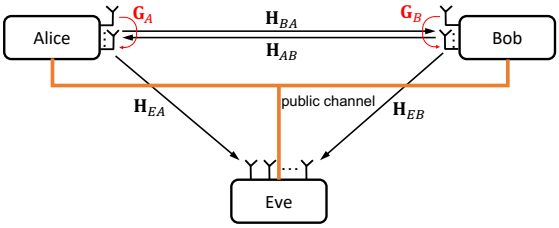


FIGURE 5.1: An OFDM-MIMO single-eavesdropper system.

As for the passive attacker Eve, we usually consider him strong to guarantee the robustness of the SKG scheme. Eve has access to the public channel and knows the communication protocols between legitimate users, so he infers the secret key by eavesdropping on legitimate users' transmissions. In addition, Eve may have high-quality receivers (e.g., with a large dynamic range, high sampling rate, etc.) to capture the probing signal and acquire accurate estimates of the eavesdropping channel. Eve may have powerful storage and computational ability to perform advanced signal processing. Eve could be located anywhere, which means she may be located near legitimate users to experience highly-correlated fading with them. We assume Eve is located near Bob and infer the common randomness for key generation from the probing signal sent by Alice. This assumption does not lose the generality since we can always regard the party that is eavesdropped on as Bob.

5.2.2 Transceiver imperfections

Practical transceivers have limited dynamic range, introducing HWIs to the transmitted and received signals. The limited dynamic range is a natural consequence of imperfect DACs, ADCs, oscillators, and PAs, as introduced earlier. Experimental measurements demonstrate that the transceiver HWIs are independent of the transmitted or received signals, and a circular complex Gaussian model can closely approximate the combined effects of these non-ideal components. Let $\kappa_i \ll 1$ and $\beta_i \ll 1$ characterize the dynamic range of transmitters and receivers of the corresponding user i, respectively, the transceiver distortions can be described by a zero-mean Gaussian model with the variance of κ_i times the power of the transmit signals (or β_i times the power of the received

signals) on that antenna as

$$\begin{aligned} \mathbf{D}_{tx,i} &\sim \mathcal{CN}\left(\mathbf{0}, \kappa_i \mathcal{D}\left(\mathbf{X}_i \mathbf{X}_i^{\dagger}\right)\right), \\ \mathbf{D}_{rx,j} &\sim \mathcal{CN}\left(\mathbf{0}, \beta_j \mathcal{D}\left(\mathbf{Y}_j \mathbf{Y}_j^{\dagger}\right)\right), \end{aligned} \tag{5.1}$$

where $\mathbf{X}_i \in \mathbb{C}^{N_i \times T}$ denote the transmitted symbols on the N_i transmitting antennas and $\mathbf{Y}_j \in \mathbb{C}^{N_j \times T}$ denote the received symbols on the N_j receiving antennas within the duration of T symbols, $i \in \{A, B\}$ and $j \in \{B, A\}$ denote a pair of legitimate users, i.e., $i = A, j = B$ or $i = B, j = A$.

Remarking: The values of κ_i and β_i are related to the measurable EVM of practical transceivers. The hardware impairments model is a verified model based on experiments and has been adopted by many studies in the field of wireless communications [32, 54].

In addition to the distortions, receivers also introduce additive white Gaussian noise (AWGN) to the received signal, which can be described as $\mathbf{W}_i \sim \mathcal{CN}\left(\mathbf{0}, \sigma_{w,i}^2 \mathbf{I}\right)$, where $\sigma_{w,i}^2$ is the thermal noise power of the corresponding receiver. The AWGN power is determined by the thermal noise density (-174 dBm/Hz for typical electronic devices), the noise figure of receivers, and operation bandwidth.

5.2.3 Key performance indicators

The generated key is used to encrypt the signals for secure communications, which have specific randomness, consistency, and refresh rate requirements. The key performance is usually evaluated by the following three metrics.

1) *Key generation rate* is the number of bits generated per second and measures the efficiency of the SKG protocol. A high KGR indicates that the protocol has high efficiency and is desired for the real-time SKG process. KGR can be described as

$$KGR = \frac{N_b}{\Delta_\tau}, \tag{5.2}$$

where Δ_τ is the interval between two probing rounds, and N_b is the averaged number of generated bits per probing round.

2) *Key disagreement rate* is the mismatch rate between the binary key bit sequences generated by Alice and Bob, which describes the robustness of the SKG protocol.

Low KDR indicates that the protocol is robust and requires fewer resources for reconciliation. The KDR has to be within the correction capacity of reconciliation to eliminate the inconsistencies between the two keys. Let $\mathcal{K}_A$ and $\mathcal{K}_B$ denote the key generated by Alice and Bob containing L_K binary bits, and the KDR can be denoted as

$$KDR = \frac{1}{L_K} \sum_{l=1}^{L_K} |\mathcal{K}_A(l) - \mathcal{K}_B(l)| \,. \tag{5.3}$$

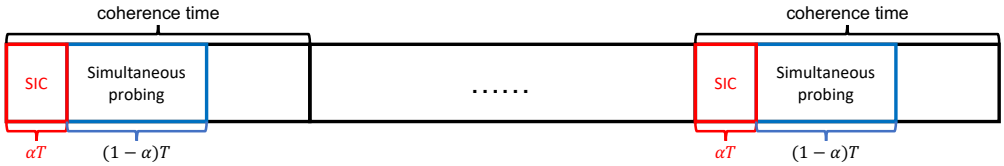
3) *Randomness* of the generated key is the most important metric for a practical SKG protocol since the key should be unpredictable to ensure the security level. Randomness could be verified by the statistical test suite provided by NIST. The test suite consists of 15 statistical tests, which focus on different randomness features. The tests are formulated to test the null hypothesis that the sequence being tested is random. Each test calculates a P-value based on a calculated test statistic value. We refer to [88] for a detailed description of the test suite. The P-value summarizes the strength of the evidence against the null hypothesis that a larger P-value indicates better randomness. 0.01 is usually used as a significance level, which means the null hypothesis is accepted (i.e., the tested sequence is considered random) if P-value$\geq$ 0.01. It is worth noting that the entropy of the key is also used as a performance metric in some studies. However, the Monobit test can evaluate the entropy of a binary bit sequence since they are both determined by the proportion of zeros and ones of the entire sequence. A larger P-value calculated under the Monobit test indicates higher entropy of the key.

5.3 Secret key capacity

SKC is the maximum key generation rate at which the secret key can be generated reliably, securely, and uniformly [77]. As stated in the attacker model, we consider an eavesdropper with high-quality hardware and powerful computation ability. Thus, we assume Eve acquires the eavesdropping channel with trivial errors that can be ignored. Since Eve can be located anywhere, there are two cases: 1) Eve is very close to the legitimate user (we assume it is Bob) so that the eavesdropping and legitimate channels are correlated; 2) Eve is far away from any of the legitimate users and experiences independent fading. We will derive the SKC for both cases in this section.

Figure 5.2 depicts the probing phase in HD and IBFD modes. We assume that a probing spans over T OFDM symbols, which is shorter than the coherence time of the channel, and the probing is done with an interval larger than the coherence time to guarantee randomness. In the conventional HD mode, the two legitimate users equally divide the available probing slots, i.e., Alice sends the probing signal during the first half of the probing period, and Bob sends the probing signal during the last half of the probing period. The non-simultaneous probing causes inconsistent measurements at legitimate users due to the time-varying wireless channel, which could be compensated by the IBFD operation. Due to the simultaneous transmission and reception nature of IBFD radios, Alice and Bob can probe the channel at the same time. However, as illustrated earlier in Chapter 2, effective SIC has to be employed before any IBFD operation, introducing SIC overheads. We will analyze their effects and derive the SKC correspondingly in this section.

(a) Half-duplex mode.



(b) In-band Full-duplex mode.

FIGURE 5.2: Channel probing in IBFD mode and HD mode.

5.3.1 Self-interference cancellation overheads

In IBFD mode, the RF and digital cancellers need to estimate the SI channel and tune their parameters accordingly at the beginning of each coherence time, and they can mitigate the SI effectively for the remaining symbols within the same coherence time. Assume the SIC needs αT symbols of overheads, where $0 \leqslant \alpha \leqslant 1$, and legitimate users work in TDD mode during the SIC phase to reduce the SI channel estimation errors. We employ the typical 3-step SIC method to minimize the effects of SI since it has been experimentally proved to suppress the SI close to the receiver's noise floor

by both studies in the literature and by us in Chapter 3. The 3-step SIC consists of passive antenna isolation, RF cancellation, and digital cancellation, and we derive the corresponding RSI strength with measurable metrics step by step.

The effect of antenna isolation is reflected in pathloss of the direct path component of the SI channel, i.e., the direct path is attenuated by 20dB if 20dB of antenna isolation is provided. We assume ϑ_i dB of cancellation depth is realized by the RF canceller at node i, and we directly reflect their effects on the SI channel strength, i.e., the SI channel is attenuated by ϑ_i dB. We assume that it is frequency-independent, i.e., ϑ_i is identical for all subcarriers k for simplicity, and we do not restrict technologies to achieve it nor the realized suppression depth.

Then, we employ a minimum mean-squared error-based digital canceller to suppress the RSI in the digital frequency domain. After the antenna isolation and RF cancellation, the received RSI signal is digitalized by ADCs with receiver noise and distortions. The digital RSI at the legitimate user i can be written as

$$\begin{aligned}\tilde{\mathbf{Y}}_i^{FD,0}[k] &= \tilde{\mathbf{G}}_i[k]\left(\mathbf{X}_i^{FD,0}[k] + \mathbf{D}_{tx,i}^{FD,0}[k]\right) + \mathbf{D}_i^{FD,0}[k] + \mathbf{W}_i^{FD,0}[k] + \mathbf{D}_{rx,i}^{FD,0}[k], \\ &= \tilde{\mathbf{G}}_i[k]\mathbf{X}_i^{FD,0}[k] + \mathbf{N}_{i,0}^{FD,0}[k],\end{aligned} \tag{5.4}$$

where $\mathbf{X}_i^{FD,0}[k] \in \mathbb{C}^{N_i \times \frac{\alpha T}{2}}$ denote the transmitted signals available for SI channel estimation during the corresponding SIC period of node i[1], $\bar{\mathbf{G}}_i[k] \in \mathbb{C}^{N_i \times N_i}$ denotes the effective SI channel with the passive antenna isolation effects (i.e., the direct path is attenuated), $\tilde{\mathbf{G}}_i[k] = \eta_i \bar{\mathbf{G}}_i[k]$ denotes the effective SI channel after the RF cancellation with $\eta_i = 10^{-\frac{\vartheta_i}{10}}$; $\mathbf{D}_i[k]$ represents the noise and distortions induced by RF cancellers. The noise and distortions of RF cancellers can also be described by the circular complex Gaussian model since it is composed of similar phase noise and nonlinearities as the transceiver hardware impairments. Let $\sigma_{d,i}^2$ describe the power of the canceller noise, then the entries of $\mathbf{D}_i[k]$ are i.i.d. to complex Gaussian distribution with zero mean and variance of $\sigma_{d,i}^2$.

Legitimate users estimate the effective SI channel $\tilde{\mathbf{G}}_i[k]$ form the persevered RSI $\tilde{\mathbf{Y}}_i^{FD,0}[k]$ during the SIC period. The effective SI channel remains unchanged during the same coherence time, so legitimate users are able to generate a replica of the RSI during the

[1]Assume the transmit power is P_i per OFDM symbol, the entries of $\mathbf{X}_i[k]$ are independent and identically distributed (i.i.d.) to complex Gaussian distribution with zero mean and variance of $\frac{P_i}{N_i}$.

IBFD probing phase and cancel the received RSI out. The MMSE channel estimator is utilized to estimate the effective SI channel, and the estimation noise is given as

$$\mathbf{N}_{i,0}^{FD,0}[k] = \tilde{\mathbf{G}}_i[k]\mathbf{D}_{tx,i}^{FD,0}[k] + \mathbf{D}_i^{FD,0}[k] + \mathbf{W}_i^{FD,0}[k] + \mathbf{D}_{rx,i}^{FD,0}[k]. \tag{5.5}$$

Assume the transmitted signals are uncorrelated with the noise, we have the entries of $\mathbf{N}_{i,0}^{FD,0}[k]$ i.i.d. to complex Gaussian distribution with zero mean and variance of

$$\sigma_{n,i,0}^2 = \kappa_i P_i \eta_i \varrho_i + \sigma_{d,i}^2 + \sigma_{w,i}^2 + \beta_i \left((1+\kappa_i) P_i \eta_i \varrho_i + \sigma_{d,i}^2\right). \tag{5.6}$$

According to Appendix 5.7, the estimate of the effective SI channel can be denoted as

$$\begin{aligned}\hat{\mathbf{G}}_i[k] &= \tilde{\mathbf{Y}}_i^{FD,0}[k] \left(\mathbf{X}_i^{FD,0}[k]\right)^{\dagger} \left(\mathbf{R}_{GG,i}\mathbf{X}_i^{FD,0}[k] \left(\mathbf{X}_i^{FD,0}[k]\right)^{\dagger} + N_i \sigma_{n,i,0}^2 \mathbf{I}\right)^{-1} \mathbf{R}_{GG,i} \\ &= \tilde{\mathbf{G}}_i[k] - \boldsymbol{\Delta}_i[k],\end{aligned} \tag{5.7}$$

where $\mathbf{R}_{GG,i} = \mathbb{E}\left\{\left(\tilde{\mathbf{G}}_i[k]\right)^{\dagger} \tilde{\mathbf{G}}_i[k]\right\}$ and $\boldsymbol{\Delta}_i[k]$ has i.i.d zero-mean complex Gaussian elements with variance of

$$\sigma_{\Delta,i,0}^2 = \frac{\sigma_{n,i,0}^2}{\frac{\alpha T P_i}{2N_i} + \frac{\sigma_{n,i,0}^2}{\eta_i \varrho_i}}. \tag{5.8}$$

To have an appropriate estimation, we should have

$$\frac{\alpha T}{2} \geqslant \max\left\{N_A, N_B\right\}, \tag{5.9}$$

as stated in Appendix 5.7.

5.3.2 Channel probing

The key performance of the CFR-based SKG scheme strongly depends on channel estimation accuracy. We assume wireless channel reciprocity is held and effective calibration is implemented at transceivers, which is essential to counteract the hardware imperfections and ensure that the channels observed at both ends are truly reciprocal. Besides, we also assume all time-frequency resources can be used for probing, i.e., ignore the interpolation error, and then we formulate the probing errors as follows.

5.3.2.1 Half-duplex probing

As Figure 5.2 shows, in HD mode, Alice and Bob send orthogonal probing signals during the first and last half of the probing period (i.e., τ_1 and τ_2), respectively. Let $\mathbf{H}_{BA}^{(\tau_1)}[k]$ and $\mathbf{H}_{AB}^{(\tau_2)}[k]$ denote the wireless channels during the corresponding probing period, respectively, the signals received by Alice and Bob are denoted as

$$\begin{aligned}\mathbf{Y}_A^{HD}[k] &= \mathbf{H}_{AB}^{(\tau_2)}[k]\left(\mathbf{X}_B^{HD}[k] + \mathbf{D}_{tx,B}^{HD}[k]\right) + \mathbf{W}_A^{HD}[k] + \mathbf{D}_{rx,A}^{HD}[k] \\ &= \mathbf{H}_{AB}^{(\tau_2)}[k]\mathbf{X}_B^{HD}[k] + \mathbf{N}_A^{HD}[k],\end{aligned} \tag{5.10}$$

$$\begin{aligned}\mathbf{Y}_B^{HD}[k] &= \mathbf{H}_{BA}^{(\tau_1)}[k]\left(\mathbf{X}_A^{HD}[k] + \mathbf{D}_{tx,A}^{HD}[k]\right) + \mathbf{W}_B^{HD}[k] + \mathbf{D}_{rx,B}^{HD}[k] \\ &= \mathbf{H}_{BA}^{(\tau_1)}[k]\mathbf{X}_A^{HD}[k] + \mathbf{N}_B^{HD}[k],\end{aligned} \tag{5.11}$$

where $\mathbf{X}_i^{HD}[k] \in \mathbb{C}^{N_i \times \frac{T}{2}}$ and $\mathbf{N}_i^{HD}[k]$ denotes the noise for channel estimation during HD probing that can be written as

$$\mathbf{N}_A^{HD}[k] = \mathbf{H}_{AB}^{(\tau_2)}[k]\mathbf{D}_{tx,B}^{HD}[k] + \mathbf{W}_A^{HD}[k] + \mathbf{D}_{rx,A}^{HD}[k], \tag{5.12}$$

$$\mathbf{N}_B^{HD}[k] = \mathbf{H}_{BA}^{(\tau_1)}[k]\mathbf{D}_{tx,A}^{HD}[k] + \mathbf{W}_B^{HD}[k] + \mathbf{D}_{rx,B}^{HD}[k]. \tag{5.13}$$

The noise matrices have i.i.d zero-mean complex Gaussian elements with the variance of

$$\sigma_{n,i}^2 = (\beta_i + \kappa_j + \beta_i\kappa_j)P_j\varrho_{ij} + \sigma_{w,i}^2. \tag{5.14}$$

With the known transmitted probing signals, Alice and Bob can acquire the estimate of the wireless channel as

$$\begin{aligned}\hat{\mathbf{H}}_{AB}^{HD}[k] &= \mathbf{Y}_A^{HD}[k]\left(\mathbf{X}_B^{HD}[k]\right)^\dagger\left(\mathbf{R}_{HH,\tau_1}\mathbf{X}_B^{HD}[k]\left(\mathbf{X}_B^{HD}[k]\right)^\dagger + N_A\sigma_{n,A}^2\mathbf{I}\right)^{-1}\mathbf{R}_{HH,\tau_1} \\ &= \mathbf{H}_{AB}^{(\tau_2)}[k] + \mathbf{\Delta}_{AB}^{HD}[k],\end{aligned} \tag{5.15}$$

$$\begin{aligned}\hat{\mathbf{H}}_{BA}^{HD}[k] &= \mathbf{Y}_B^{HD}[k]\left(\mathbf{X}_A^{HD}[k]\right)^\dagger\left(\mathbf{R}_{HH,\tau_2}\mathbf{X}_A^{HD}[k]\left(\mathbf{X}_A^{HD}[k]\right)^\dagger + N_B\sigma_{n,B}^2\mathbf{I}\right)^{-1}\mathbf{R}_{HH,\tau_2} \\ &= \mathbf{H}_{BA}^{(\tau_1)}[k] + \mathbf{\Delta}_{BA}^{HD}[k],\end{aligned} \tag{5.16}$$

where $\mathbf{R}_{HH,\tau_1}$ and $\mathbf{R}_{HH,\tau_2}$ denote the covariance matrices of the legitimate channel during different periods that can be given as

$$\mathbf{R}_{HH,\tau_1} = \mathbb{E}\left\{\left(\mathbf{H}_{BA}^{(\tau_1)}[k]\right)^{\dagger}\mathbf{H}_{BA}^{(\tau_1)}[k]\right\} \tag{5.17}$$

$$\mathbf{R}_{HH,\tau_2} = \mathbb{E}\left\{\left(\mathbf{H}_{AB}^{(\tau_2)}[k]\right)^{\dagger}\mathbf{H}_{AB}^{(\tau_2)}[k]\right\} \tag{5.18}$$

$\mathbf{\Delta}_{AB}^{HD}[k]$ and $\mathbf{\Delta}_{BA}^{HD}[k]$ have i.i.d zero-mean complex Gaussian elements with the variance of

$$\sigma_{\Delta,A}^2 = \frac{\sigma_{n,A}^2}{\frac{TP_B}{2N_B} + \frac{\sigma_{n,A}^2}{\varrho_{AB}}} \text{ and } \sigma_{\Delta,B}^2 = \frac{\sigma_{n,B}^2}{\frac{TP_A}{2N_A} + \frac{\sigma_{n,B}^2}{\varrho_{AB}}}, \tag{5.19}$$

respectively. To have an appropriate estimation, we should have

$$\frac{T}{2} \geqslant \max\{N_A, N_B\}. \tag{5.20}$$

Remarking: Due to the temporal changes of the environment during the transmission interval in HD mode [84], the wireless channel may not be identical during the probing period τ_1 and τ_2, i.e., $\mathbf{H}_{AB}^{(\tau_1)}$ and $\mathbf{H}_{AB}^{(\tau_2)}$ are not identical but highly correlated. However, $\mathbf{H}_{AB}^{(\tau_1)}$ is reciprocal to $\mathbf{H}_{BA}^{(\tau_1)}$ such that $\mathbf{H}_{BA}^{(\tau_1)} = \left(\mathbf{H}_{AB}^{(\tau_1)}\right)^T$ due to the wireless channel reciprocity, so the estimate of Alice-to-Bob channel can be written as

$$\hat{\mathbf{H}}_{BA}^{HD}[k] = \left(\mathbf{H}_{AB}^{(\tau_1)}[k]\right)^T + \mathbf{\Delta}_{BA}^{HD}[k]. \tag{5.21}$$

5.3.2.2 In-band full-duplex probing

During the IBFD probing phase, Alice and Bob simultaneously send orthogonal probing signals during the period consisting of $(1-\alpha)T$ OFDM symbol durations. With the estimate of the effective SI channel $\hat{\mathbf{G}}_i[k]$ obtained from the SIC overheads, legitimate users are able to generate $\hat{\mathbf{G}}_i[k]\mathbf{X}_i^{FD,1}[k]$ to cancel out the received RSI, where $\mathbf{X}_i^{FD,1}[k] \in \mathbb{C}^{N_i \times (1-\alpha)T}$ denotes the corresponding IBFD probing signals. Due to the imperfect estimate of effective SI channel and additional noise from RF cancellers, there will be residual noise caused by SI after digital cancellation denoted as

$$\mathbf{\Psi}_i^{FD,1}[k] = \tilde{\mathbf{G}}_i[k]\mathbf{D}_{tx,i}^{FD,1}[k] + \mathbf{\Delta}_i\mathbf{X}_i^{FD,1}[k] + \mathbf{D}_i^{FD,1}[k]. \tag{5.22}$$

The entries of $\mathbf{\Psi}_i^{FD,1}[k]$ are i.i.d. to complex Gaussian distribution with zero mean and the variance of

$$\sigma_{s,i,0}^2 = \kappa_i P_i \eta_i \varrho_i + \sigma_{\Delta,i,0}^2 P_i + \sigma_{d,i}^2. \tag{5.23}$$

With SIC applied, the received signals at legitimate users during the IBFD probing phase can be denoted as (the superscript for the time period of the channel matrix is omitted here since legitimate users send the probing signals simultaneously)

$$\begin{aligned}\mathbf{Y}_i^{FD,1}[k] &= \mathbf{H}_{ij}[k]\left(\mathbf{X}_j^{FD,1}[k] + \mathbf{D}_{tx,j}^{FD,1}[k]\right) + \mathbf{\Psi}_i^{FD,1}[k] + \mathbf{W}_i^{FD,1}[k] + \mathbf{D}_{rx,i}^{FD,1}[k] \\ &= \mathbf{H}_{ij}[k]\mathbf{X}_j^{FD,1}[k] + \mathbf{N}_i^{FD,1}[k],\end{aligned} \tag{5.24}$$

where $\mathbf{N}_i^{FD,1}[k]$ denotes the noise for channel estimation during the IBFD probing that can be given as

$$\mathbf{N}_i^{FD,1}[k] = \mathbf{H}_{ij}[k]\mathbf{D}_{tx,j}^{FD,1}[k] + \mathbf{\Psi}_i^{FD,1}[k] + \mathbf{W}_i^{FD,1}[k] + \mathbf{D}_{rx,i}^{FD,1}[k], \tag{5.25}$$

which has i.i.d zero-mean complex Gaussian elements with the variance of

$$\sigma_{n,i,1}^2 = (1+\beta_i)\left(\kappa_j P_j \varrho_{ij} + \sigma_{s,i,0}^2\right) + \beta_i P_j \varrho_{ij} + \sigma_{w,i}^2. \tag{5.26}$$

With the known transmitted probing signals, Alice and Bob can estimate the channel using the MMSE channel estimator as given in Appendix 5.7, yielding the estimates with errors as

$$\hat{\mathbf{H}}_{AB}^{FD}[k] = \mathbf{H}_{AB}[k] + \mathbf{\Delta}_{AB}^{FD}[k], \tag{5.27}$$

$$\hat{\mathbf{H}}_{BA}^{FD}[k] = \mathbf{H}_{BA}[k] + \mathbf{\Delta}_{BA}^{FD}[k], \tag{5.28}$$

where the entries of estimator error matrices $\mathbf{\Delta}_{AB}^{FD}[k]$ and $\mathbf{\Delta}_{BA}^{FD}[k]$ are i.i.d to complex Gaussian distribution with zero mean and the variance of

$$\sigma_{\Delta,A,1}^2 = \frac{\sigma_{n,A,1}^2}{\frac{(1-\alpha)TP_B}{N_B} + \frac{\sigma_{n,A,1}^2}{\varrho_{AB}}} \text{ and } \sigma_{\Delta,B,1}^2 = \frac{\sigma_{n,B,1}^2}{\frac{(1-\alpha)TP_A}{N_A} + \frac{\sigma_{n,B,1}^2}{\varrho_{AB}}}, \tag{5.29}$$

respectively. To have an appropriate estimation, we should have

$$(1-\alpha)T \geqslant \max\{N_A, N_B\}. \tag{5.30}$$

Remarking: For IBFD probing, we have $\mathbf{H}_{AB} = (\mathbf{H}_{BA})^T$ due to the wireless channel

reciprocity and simultaneous probing, so the estimated Alice-to-Bob channel can be written as

$$\hat{\mathbf{H}}_{BA}^{FD}[k] = (\mathbf{H}_{AB}[k])^T + \mathbf{\Delta}_{BA}^{FD}[k]. \tag{5.31}$$

5.3.3 Correlated eavesdropping channel

When Eve is very close to Bob, she will experience correlated fading as Bob. We do not consider how Eve processes her observations since this is not our concern, while we assume Eve can extract a channel matrix correlated with $\mathbf{H}_{BA}$ from the correlated observations, which can be modeled as [73][2]

$$\hat{\mathbf{H}}_{EA} = \rho' \mathbf{H}_{BA} + \mathbf{\Lambda}_{BA}, \tag{5.32}$$

where $[\mathbf{\Lambda}_{BA}]_{k,l} \sim \mathcal{CN}\left(0, (1-\rho'^2)\varrho_{AB}\right)\ \forall\, k, l$. The SKC is determined by the conditional mutual information given the column-wise vectorization of the measured CFRs (i.e., $\mathbf{h}_i = \text{vec}\left(\hat{\mathbf{H}}_{ij}[k]\right)$), expressed as

$$C_k = I\left(\mathbf{h}_A; \mathbf{h}_B | \mathbf{h}_E\right). \tag{5.33}$$

Lemma 5.1. *The closed-form expressions of the SKC for HD and IBFD probing are given as*

$$\begin{aligned} C_k^{HD} &= I\left(\mathbf{h}_A^{HD}; \mathbf{h}_B^{HD} \mid \mathbf{h}_E^{HD}\right) \\ &= N_A N_B \log_2 \left(\frac{\varpi_A^{HD}\varpi_B^{HD} - \left(\varpi_A^{HD} + \varpi_B^{HD}\rho^2\right)\rho'^2\varrho_{AB} + \rho'^4\rho^2\varrho_{AB}^2}{\varpi_A^{HD}\varpi_B^{HD} - \left(\varpi_A^{HD} + \varpi_B^{HD}\rho^2\right)\rho'^2\varrho_{AB} + (2\rho'^2 - 1)\rho^2\varrho_{AB}^2} \right), \end{aligned} \tag{5.34}$$

$$\begin{aligned} C_k^{FD} &= I\left(\mathbf{h}_A^{FD}; \mathbf{h}_B^{FD} \mid \mathbf{h}_E^{FD}\right) \\ &= N_A N_B \log_2 \left(\frac{\varpi_A^{FD}\varpi_B^{FD} - (\varpi_A^{FD} + \varpi_B^{FD})\rho'^2\varrho_{AB} + \rho'^4\varrho_{AB}^2}{\varpi_A^{FD}\varpi_B^{FD} - (\varpi_A^{FD} + \varpi_B^{FD})\rho'^2\varrho_{AB} + (2\rho'^2 - 1)\varrho_{AB}^2} \right), \end{aligned} \tag{5.35}$$

where $\varpi_i^{HD} = \varrho_{AB} + \sigma_{\Delta,i}^2$ *and* $\varpi_i^{FD} = \varrho_{AB} + \sigma_{\Delta,i,1}^2$.

Proof. See Appendix 5.8. □

[2]The superscript indicating the probing mode is omitted here because the result is the same for both modes.

The expressions clearly illustrate the gain of MIMO systems that both the transmitting and receiving antenna arrays can increase the SKC proportionally. The SKC is subjected to the probing errors ($\sigma^2_{\Delta,i}$ or $\sigma^2_{\Delta,i,1}$), the correlation between consecutive CFRs (ρ), and the correlation between the eavesdropping and legitimate channels (ρ'). In addition, the SIC overheads (α), probing duration (T), transmit power (P_A and P_B), noise and distortions, and antenna array size (N_A and N_b) will affect the SKC by affecting the probing errors.

5.3.4 Independent eavesdropping channel

Eve will experience independent fading from legitimate users if she is located half a wavelength away from them, which is known as the spatial independence assumption. This assumption is stated and considered in many related studies, such as in [70, 89]. Let $\mathbf{Y}_E[k]$ denote Eve's observation, which is expressed as

$$\begin{aligned}\mathbf{Y}_E[k] = \phi_{EA}\mathbf{H}_{EA}[k]\left(\mathbf{X}_A[k] + \mathbf{D}_{tx,A}[k]\right) \\ + \phi_{EB}\mathbf{H}_{EB}[k]\left(\mathbf{X}_B[k] + \mathbf{D}_{tx,B}[k]\right) + \mathbf{W}_E[k] + \mathbf{D}_{rx,E}[k],\end{aligned} \tag{5.36}$$

where $\phi_{EA}, \phi_{EB} \in \{0, 1\}$ depends on the probing mode and phase, e.g., $\phi_{EA} = 1$ and $\phi_{EB} = 1$ during the IBFD probing phase. Based on the spatial independence assumption that the $\mathbf{H}_{EA}$ and $\mathbf{H}_{EB}$ are independent of $\mathbf{H}_{AB}$ and $\mathbf{H}_{BA}$, Eve cannot extract a correlated CFR from her observations since none of the terms in $\mathbf{Y}_E$ is correlated with the legitimate channel. In this case, the SKC is given as the mutual information given estimated legitimate CFRs.

Lemma 5.2. *The closed-form expressions of the SKC for HD and IBFD probing under spatial independence assumption are given as*

$$C_k^{HD} = I\left(\mathbf{h}_A^{HD}; \mathbf{h}_B^{HD}\right) = N_A N_B \log_2\left(\frac{\varpi_A^{HD}\varpi_B^{HD}}{\varpi_A^{HD}\varpi_B^{HD} - \rho^2\varrho_{AB}^2}\right), \tag{5.37}$$

$$C_k^{FD} = I\left(\mathbf{h}_A^{FD}; \mathbf{h}_B^{FD}\right) = N_A N_B \log_2\left(\frac{\varpi_A^{FD}\varpi_B^{FD}}{\varpi_A^{FD}\varpi_B^{FD} - \varrho_{AB}^2}\right), \tag{5.38}$$

which are equivalent to substituting $\rho' = 0$ to Equations (5.34) and (5.35).

Proof. See Appendix 5.9. □

The expressions suggest that the SKC is mainly limited by the probing error (i.e., $\sigma^2_{\Delta,i}$ for HD and $\sigma^2_{\Delta,i,1}$ for IBFD) and the correlation coefficient (i.e., ρ) in this case. The SKC decreases with increasing probing errors and decreasing correction coefficients. The correlation coefficient is an inherent property of the wireless channel, which depends on the user's moving speed. The probing errors will be affected by many system parameters, as their expressions suggest.

5.3.5 Asymptotic behavior analysis

To investigate the fundamental limits on the performance of PHY-SKG, we derive the asymptotic SKC in the high-SNR regime, which can be obtained by tending the transmit power in Equations (5.37) and (5.38) to infinity. We only consider the case under the spatial independence assumption since we want to focus on its intrinsic limits rather than the eavesdropper's location effect. The asymptotic SKC for HD probing is given as

$$C_k^{HD,asym} = N_A N_B \log_2 \left(\frac{\bar{\varpi}_A^{HD} \bar{\varpi}_B^{HD}}{\bar{\varpi}_A{}^{HD} \bar{\varpi}_B^{HD} - \rho^2 \varrho_{AB}^2} \right), \tag{5.39}$$

where $\bar{\varpi}_i^{HD} = \varrho_{ij} + \bar{\sigma}^2_{\Delta,i}$ with

$$\bar{\sigma}^2_{\Delta,i} = \lim_{P_i \to \infty} \sigma^2_{\Delta,i} = \varrho_{ij} \frac{1}{\frac{T\varrho_{ij}}{2N_j \Sigma_{ij}} + 1}, \tag{5.40}$$

where $\Sigma_{ij} = \kappa_j + \beta_i + \kappa_j \beta_i$. Similarly, we can derive the asymptotic SKC for IBFD probing. We first investigate the IBFD probing error in the high-SNR regime as

$$\begin{aligned} \bar{\sigma}^2_{\Delta,i,1} &= \lim_{P_i, P_j \to \infty} \frac{\sigma^2_{n,i,1}}{\frac{(1-\alpha)TP_j}{N_j} + \frac{\sigma^2_{n,i,1}}{\varrho_{ij}}} \\ &= \lim_{P_i = P_j = P \to \infty} \frac{\frac{\partial \sigma^2_{n,i,1}}{\partial P}}{\frac{(1-\alpha)T}{N_j} + \frac{1}{\varrho_{ij}} \frac{\partial \sigma^2_{n,i,1}}{\partial P}}, \end{aligned} \tag{5.41}$$

where $\sigma^2_{n,i,1}$ is detailed as

$$\begin{aligned} \sigma^2_{n,i,1} = \beta_i P_j \varrho_{ij} + \sigma^2_{w,i} + (1+\beta_i) \Bigg(& \kappa_j P_j \varrho_{ij} + \kappa_i P_i \eta_i \varrho_i \\ & + \frac{\kappa_i P_i \eta_i \varrho_i + \sigma^2_{d,i} + \sigma^2_{w,i} + \sigma^2_r [(1+\kappa_i) P_i \eta_i \varrho_i + \sigma^2_{d,i}]}{\frac{\alpha T P_i}{2N_i} + \frac{\kappa_i P_i \eta_i \varrho_i + \sigma^2_{d,i} + \sigma^2_{w,i} + \beta_i [(1+\kappa_i) P_i \eta_i \varrho_i + \sigma^2_{d,i}]}{\eta_i \varrho_i}} P_i + \sigma^2_{d,i} \Bigg). \end{aligned} \tag{5.42}$$

Thus, we will have

$$\begin{aligned} n_{FD}^{asym} &= \lim_{P\to\infty} \frac{\partial \sigma_{n,i,1}^2}{\partial P} \\ &= \Sigma_{ij}\varrho_{ij} + (1+\beta_i)\left(\kappa_i\eta_i\varrho_i + \frac{2N_i\eta_i\varrho_i\Sigma_{ii}}{\alpha T + 2N_i\Sigma_{ii}}\right), \end{aligned} \tag{5.43}$$

where $\Sigma_{ii} = \kappa_i + \beta_i + \kappa_i\beta_i$. Then, the asymptotic SKC for IBFD probing can be written as

$$C_k^{FD,asym} = N_A N_B \log_2\left(\frac{\bar{\varpi}_A^{FD}\bar{\varpi}_B^{FD}}{\bar{\varpi}_A{}^{FD}\bar{\varpi}_B^{FD} - \varrho_{AB}^2}\right), \tag{5.44}$$

where

$$\bar{\varpi}_i^{FD} = \varrho_{ij} + \varrho_{ij}\frac{1}{\frac{(1-\alpha)T\varrho_{ij}}{N_j n_{FD}^{asym}} + 1}. \tag{5.45}$$

The asymptotic results illustrate that there is an upper bound of the SKC, which is imposed by transceiver HWIs in HD systems. In IBFD systems, the limits are also imposed by the ASIC depth (which is reflected by η_i and ϱ_i) and SIC overheads (α) in addition to the transceiver HWIs.

5.3.6 Conditions for IBFD to gain benefits

To get practical insights, we explore the conditions under which IBFD probing can provide gains over its HD counterpart based on asymptotic behavior. We assume Alice and Bob have identical settings (i.e., identical hardware conditions, transmit power, etc.) for simplicity so that they will have identical probing errors. Let $P_A = P_B = P$, $N_A = N_B = N$, we can denote the probing errors for legitimate users in HD and IBFD mode as

$$\bar{\sigma}_{HD}^2 = \bar{\sigma}_{\Delta,i}^2 = \varrho_{AB}\frac{1}{\frac{T\varrho_{AB}}{2N\Sigma_{ij}} + 1} \;\forall\, i, \tag{5.46}$$

$$\bar{\sigma}_{FD}^2 = \bar{\sigma}_{\Delta,i}^2 = \varrho_{AB}\frac{1}{\frac{T\varrho_{AB}}{2Nn_{FD}^{asym}} + 1} \;\forall\, i. \tag{5.47}$$

By deriving from $C_k^{FD,asym} > C_k^{HD,asym}$, we will have the condition for IBFD to gain benefits over HD as

$$\rho\bar{\sigma}_{FD}^2 < h(1-\rho) + \bar{\sigma}_{HD}^2. \tag{5.48}$$

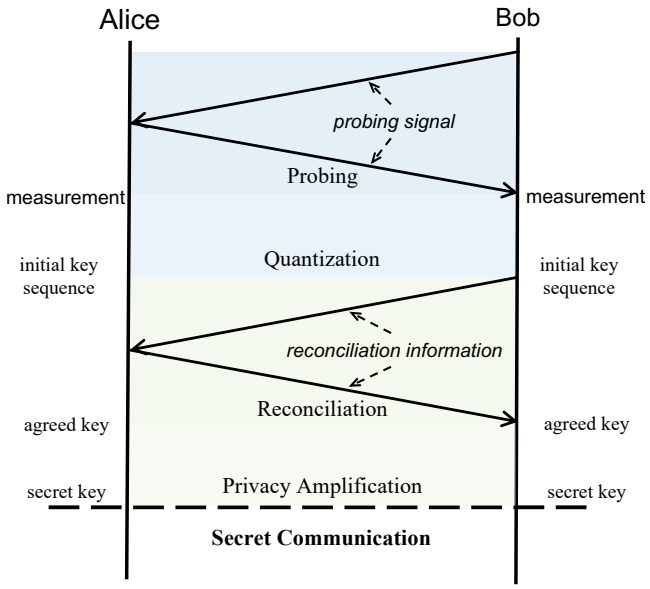


FIGURE 5.3: A typical secret key generation protocol that consists of probing, quantization, reconciliation, and privacy amplification.

This suggests that IBFD could have a larger probing error but still achieve higher SKC than HD with $\rho < 1$. Consider the scenario that $\rho \to 1$, we will have this condition be

$$n_{FD}^{asym} < 2(1-\alpha)\Sigma_{ij}, \tag{5.49}$$

which can be satisfied with appropriate SIC overheads (α) and ASIC depth (which is reflected by η_i and ϱ_i). It reveals that an appropriate SIC scheme (with sufficient ASIC depth and small SIC overheads) is the basis for IBFD to gain benefits. The minimum ASIC depth or the range of SIC overhead that IBFD outperforms HD can be easily calculated based on the condition with specific parameters.

5.4 Secret key generation protocol

A typical secret key generation protocol consists of four phases: channel probing, quantization, information reconciliation, and privacy amplification, as Figure 5.3 depicts. Legitimate users first send probing signals to each other and measure the legitimate channel. The difference between IBFD and HD radios mainly lies in this channel probing phase, as introduced in the previous section. The probing interval is larger than the coherence time of the wireless channel; otherwise, the observations will be highly correlated, compromising the randomness of the key. With the observations of transmitted

probing signals, legitimate users harness the common randomness to generate the secret key by quantizers, e.g., threshold-based quantizers [90], or bidirectional difference quantizers [71]. The bit sequences generated independently by legitimate users may not be identical due to flawed reciprocity caused by practical imperfections such as asymmetric transceiver impairments, half-duplex probing interval, and noise. To eliminate the inconsistencies and achieve key agreement, information reconciliation will be employed by exchanging a message over a public channel. Reconciliation may leak information about the key and compromise security. Thus, privacy amplification is usually utilized to remove the leaked information by leveraging a one-way mapping function, e.g., Hash function [75].

For the CFR-based SKG scheme, one round of channel probing is performed within a coherence time period. At the n^{th} coherence time period, a set of measurements can be obtained at legitimate users, which consist of estimated CFR matrices over all subcarriers as $\mathcal{H}_{i,n} = \left\{\hat{\mathbf{H}}_{ij,n}[k]\right\}_{k=1}^{K}$ [3]. Then, these measurements are appropriately converted into bit sequences independently by Alice and Bob as follows.

5.4.1 Pre-processing

The measured channel matrices on consecutive subcarriers are highly correlated, so directly quantizing the entries of these matrices will decrease the randomness of the generated bit sequences. Therefore, the measurements are pre-processed to enhance the key performance. A feasible solution is intermittently selecting the measured channel matrices for quantization with fixed subcarrier intervals. To fully utilize the channel information on each subcarrier, we alternatively average the measurements within the same subband, which can reduce the effects of estimation errors if the errors are independent on each subcarrier. Assume all the K measurements are divided into M blocks, and the $\lfloor \frac{K}{M} \rfloor$ measurements within the m^{th} block are averaged to obtain a single sample for quantization as (we assume $\frac{K}{M}$ is an integer for simplicity)

$$\hat{\mathbf{H}}_{ij,n,m} = \frac{K}{M} \sum_{k=1}^{\frac{K}{M}} \hat{\mathbf{H}}_{ij,n} \left[\frac{K(m-1)}{M} + k \right]. \tag{5.50}$$

[3] The superscript indicating the probing mode is omitted here since the remaining processing is the same for both modes.

To fully utilize these samples and maximize the KGR, we quantize both the real and imaginary parts of the complex entries of these matrices. Therefore, $\hat{\mathbf{H}}_{ij,n,m} \in \mathbb{C}^{N_i \times N_j}$ is converted into a vector $\mathbf{s}_{i,n,m} \in \mathbb{R}^{2N_A N_B \times 1}$ as

$$\mathbf{s}_{i,n,m} = \begin{bmatrix} \mathcal{R}\left\{\text{vec}\left(\hat{\mathbf{H}}_{ij,n,m}\right)\right\} \\ \mathcal{I}\left\{\text{vec}\left(\hat{\mathbf{H}}_{ij,n,m}\right)\right\} \end{bmatrix}. \tag{5.51}$$

At the end of the n^{th} coherence time period, a total of $L_s = 2MN_AN_B$ of real numbers are collected by Alice and Bob as

$$\mathcal{S}_i = \{\mathbf{s}_{i,n,m}\}_{m=1}^{M}. \tag{5.52}$$

The power of the CFR is identical on all subcarriers in an uncorrelated scattering environment [91], which means the samples are within the same range. Thus, it is not necessary to adjust the range of these samples to generate a uniformly-distributed key. The samples can be normalized by the pathloss ϱ_{AB}, and the scaling does not affect the distribution of samples nor the mutual information between the sample sets collected by legitimate users. Figure 5.4 shows the distribution of processed samples to be quantized at Alice and Bob, which indicates that the samples are subjected to a zero-mean Gaussian distribution.

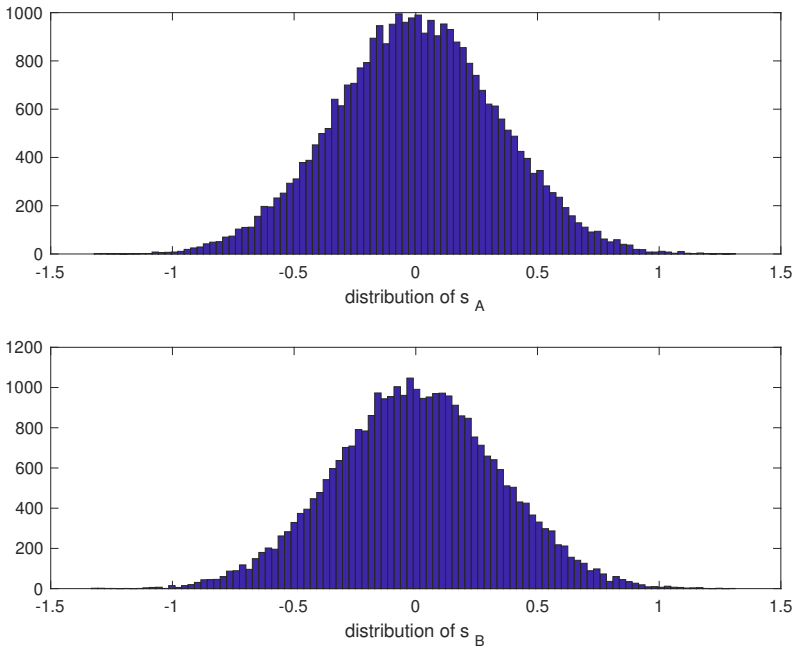


FIGURE 5.4: Distribution of the collected samples.

5.4.2 Segmental quantization

Quantization is employed to convert the processed samples $\mathcal{S}_A$ and $\mathcal{S}_B$ into binary bit sequences. The quantization schemes can be classified into lossy quantizers and lossless quantizers. Lossless quantizers utilize every sample to maximize the KGR, while lossy quantizers set guard strips to strike a tradeoff between the KGR and KDR. The readers are referred to [90] for more information about the quantizer design. We consider a single-bit lossy quantizer to minimize the KDR. To improve the randomness of the generated key bit sequence, we propose segmental quantization to remove the effects of large-scale fading and only quantize the small-scale fading. This is realized by dividing the sample sequence into multiple segments to be independently quantized. The thresholds are calculated according to the samples within different segments. Let $\mathbf{s}_{i,z}$ denote the z^{th} segments consisting of L_b samples at user i, it can be quantized to a binary bit sequence as

$$\mathbf{b}_{i,z} = \begin{cases} 1, \text{ if } \mathbf{s}_{i,z} > q_{+,i,z} \\ 0, \text{ if } \mathbf{s}_{i,z} < q_{-,i,z} \end{cases} \tag{5.53}$$

where $q_{+,i,z} = \mu_{i,z} + ß\sigma_{i,z}$ and $q_{-,i,z} = \mu_{i,z} - \gamma\sigma_{i,z}$ with $0 \leq ß \leq 1$ denote the upper and lower thresholds; $\mu_{i,z}$ and $\sigma^2_{i,z}$ are the mean and variance of the samples in $\mathbf{s}_{i,z}$. The range of $[q_{-,i,z}, q_{+,i,z}]$ is the guard strip that samples fall into the strip will be discarded with their indices recorded into $\mathcal{M}_i$. The key bit sequence can be obtained as $\mathcal{K}_i = \left\{\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, \ldots, \mathbf{b}_{i,L_{k,i}}\right\}$. It is guaranteed to generate a uniformly-distributed key sequence due to the symmetry of the thresholds and the probability distribution function of Gaussian distribution, as Figure 5.4 shows.

5.4.3 Information reconciliation and privacy amplification

Information reconciliation aims at eliminating discrepancies in the initial key sequences after quantization. To this end, Alice and Bob will exchange a message over the public channel, which should minimize the information leakage on the key to guarantee its security of the key. The reconciliation can be implemented with error detection protocols (e.g., Cascade) or error correction codes (e.g., low-density parity-check, Turbo code, and Golay code). The error correction code-based method achieves higher efficiency than the protocol-based one at the cost of higher complexity and more information leakage.

The readers are referred to [92, 93] and references therein for a detailed description of the reconciliation techniques. Since the reconciliation message is transmitted over a public channel, the eavesdropper can hear it, compromising the security of the key. Thus, privacy amplification is leveraged to remove the leaked information, which can be implemented with an extractor or universal hashing functions. The reconciliation and amplification privacy are usually crossly designed.

With the lossy quantizer, we could use the indices as the reconciliation message to perform reconciliation at first. The index-based reconciliation consists of the following steps.

1) Alice and Bob send the indices of the discarded samples $\mathcal{M}_A$ and $\mathcal{M}_B$ to each other through the public channel.

2) Integrate all indices in $\mathcal{M}_A$ and $\mathcal{M}_B$ to from $\mathcal{M}_C$ at Alice and Bob.

3) Alice and Bob discard the bits at corresponding indices of $\mathcal{M}_C$.

This scheme will not cause key information leakage since the shared public message is independent of the key itself. Thus, privacy amplification is not necessary. However, this scheme does not guarantee key agreement, especially when the estimation error is large. For a lossy quantizer, it is efficient to remove a large portion of error bits through the index-based reconciliation method, while more advanced error correction code-based methods could be further employed to achieve secure key agreement along with privacy amplification. Further processing serves as the complement and depends highly on the performance of the initial bit sequence. Similar to other related works [74, 85], simulation results are presented based on initial keys, i.e., the bit sequences after the index-based reconciliation.

5.5 Simulation results

We consider an OFDM system following 3GPP specifications, where the main parameters are given in Table 5.1. The noise of RF cancellers is set to be identical to the thermal noise if not specified. 3GPP TDL models are utilized to construct the wireless MIMO channels based on a Matlab implementation. The “TDL-C” model is employed to construct the channel between legitimate users, and the “TDL-E” model with a K-factor

TABLE 5.1: Parameters for 3GPP-based PHY-SKG simulations.

Parameters	**Values**
Carrier frequency	2.5 GHz
Subcarrier spacing	15 kHz
Bandwidth	4.5 MHz
Modulation	4QAM
Cycle prefix	"normal"
Probing period	1 ms ($T = 14$, if not specified)
Probing interval	25 ms (if not specified)
Pathloss	Table 7.4.1-1 in [3]
Thermal noise density	-174 dBm/Hz
Noise figure	9 dB
ADC/DAC bits	12

of 22 dB is used to construct the SI channel. The MIMO correlation is set to be low to be consistent with the theoretical derivation. UMa scenarios, which are introduced in Table 4.2 in the earlier Chapter 4 are considered for a typical open space and office environment, and the RMS delay spread of 50 ns is utilized for simulations [71]. The distance between legitimate users is 100 m if not specified.

5.5.1 Secret key capacity

In this section, we explore the limiting and affecting factors of the IBFD and HD radios on the SKC based on mathematical derivations. We set $\kappa_i = \beta_i = -50$ dB $\forall\ i$ in this section. Figure 5.5 compares the SKC under an independent eavesdropper, i.e., $\rho' = 0$, with varying transmit power and different channel conditions.

The results illustrate that the reduced correlation between continuous channels (ρ) is the limiting factor for HD probing, which makes it inferior to IBFD probing. The SKC decreases significantly with decreasing ρ for HD probing. IBFD probing with different correlation coefficients ρ is not compared since it does not impose an effect in this mode. For IBFD probing, the most critical factor limiting the key capacity is the ASIC depth. The SI has to be efficiently suppressed in the analog domain to guarantee the effectiveness of digital cancellation, as revealed in Chapter 2. In addition to the SIC depth, another affecting factor is the SIC overheads, which is related to α. SIC overheads reduce the available resources for legitimate channel estimation, increasing the estimation error and reducing the SKC. For $\alpha < 0.5$, more OFDM symbols are available for legitimate channel estimation in IBFD mode than its HD counterpart,

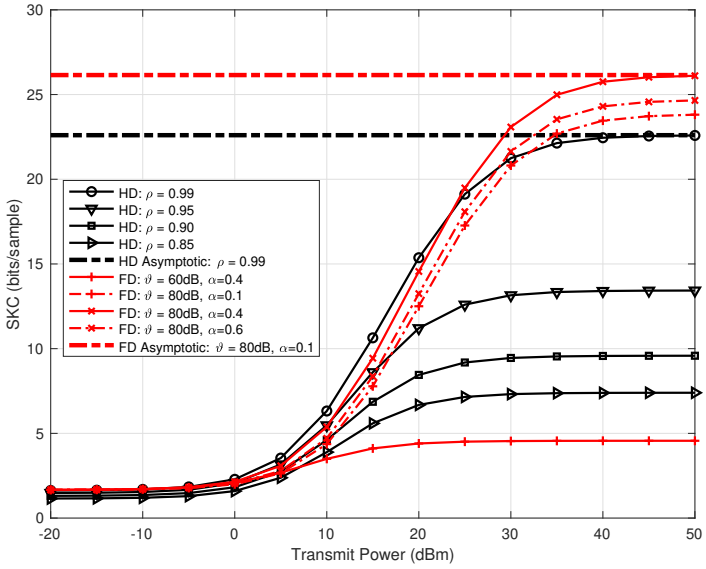


FIGURE 5.5: SKC versus transmit power under various conditions for HD and IBFD probing.

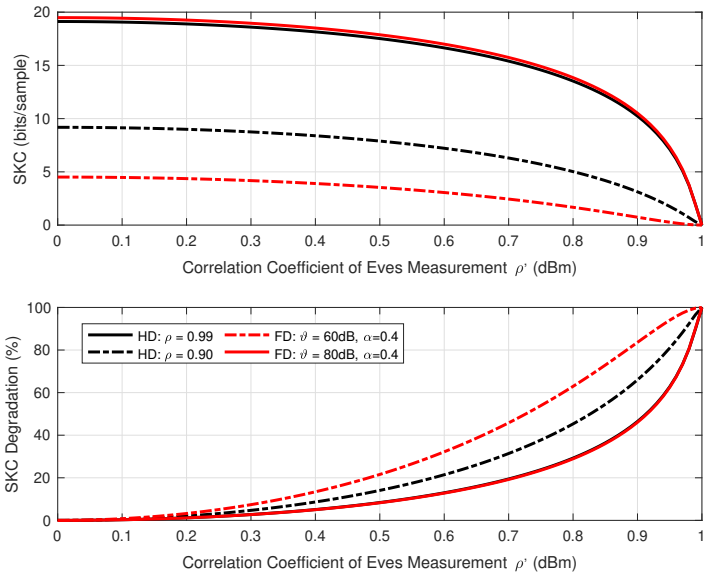


FIGURE 5.6: SKC degradation due to correlated eavesdropping channel.

benefitting the SKC by reducing the estimation errors. However, α has to satisfy the condition that $T \cdot \min\left\{\frac{\alpha}{2}, 1-\alpha\right\} \geqslant \max\{N_A, N_B\}$; otherwise, it will decrease the SKC due to inappropriate digital cancellation, e.g., $\alpha = 0.1$. Too long SIC overheads (e.g., $\alpha = 0.6$) will also decrease the SKC of IBFD probing. In the high-SNR regime, IBFD can achieve a higher SKC than HD with effective ASIC depth and appropriate SIC overheads.

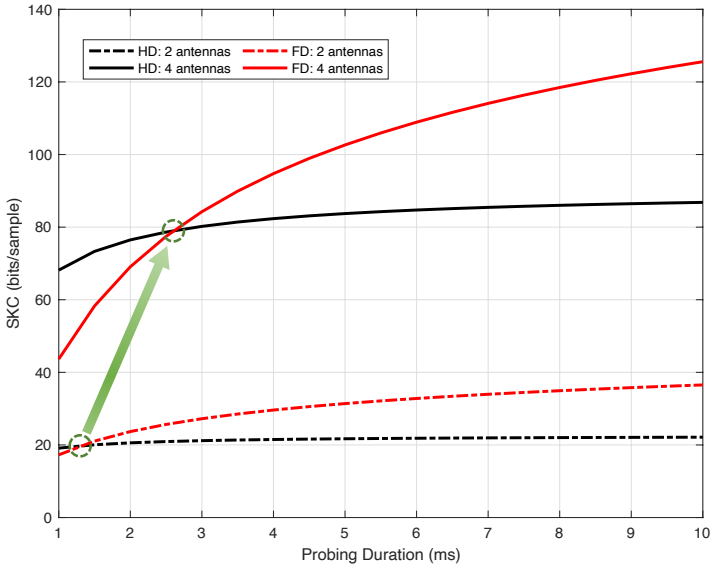


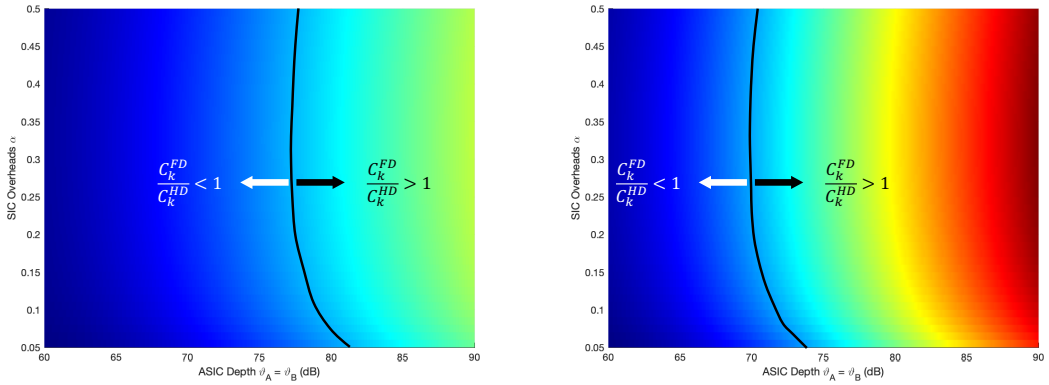
FIGURE 5.7: SKC versus probing duration with different antenna array sizes.

Then, we investigate the effects of Eve's location. Figure 5.6 shows the SKC degradation due to the presence of a close eavesdropper, who can acquire a CFR correlated with the legitimate channel. The results show that if Eve cannot obtain a highly-correlated channel observation, it will not significantly reduce the SKC. The SKC degradation increases with decreasing ρ for HD and poor ASIC configurations (i.e., decreasing ϑ) for IBFD. In addition, IBFD could reduce the SKC degradation with effective ASIC for moving users, which have a low correlation coefficient ρ of continuous channels due to the Doppler shifts, compared to its HD counterpart. However, the SKC degradation for HD and IBFD probing is similar for relatively stationary users.

Although the SKC expressions clearly indicate the gain of antenna arrays, their underlying effects on the probing errors are ambiguous. Figure 5.7 shows the SKC variation with increasing probing duration and enlarged antenna arrays. It can be seen that a larger antenna array can significantly improve the SKC, illustrating the MIMO gain. However, a longer probing duration is required for IBFD to achieve a higher SKC than HD. The reason is that SIC requires longer overheads with enlarging antenna arrays; otherwise, the RSI is increased, increasing the probing errors and reducing the SKC.

As analyzed above, the condition for IBFD to gain benefits over HD depends on the SIC overheads and ASIC depth for a fixed transceiver and channel condition. Thus, we explore the IBFD gain ($= C_k^{FD}/C_k^{HD}$) with varying α and ϑ_A (ϑ_B), as Figure 5.8

shows. The results show that sufficient ASIC is the basis for IBFD to gain benefits, and appropriate SIC overheads can maximize the gain. With rapidly-varying channels (or fast-moving users), i.e., low ρ, IBFD probing has less requirement on the ASIC depth to outperform its HD counterpart.



(a) $\rho = 0.99$ (b) $\rho = 0.95$

FIGURE 5.8: Gain of IBFD probing over HD probing on secret key capacity (i.e., C_k^{FD}/C_k^{HD}) with varying SIC overheads and ASIC depth.

Figure 5.9 shows the effects of transceiver HWIs, which are revealed as a fundamental limit of the SKC for both HD and IBFD probing. The results show that large transceiver

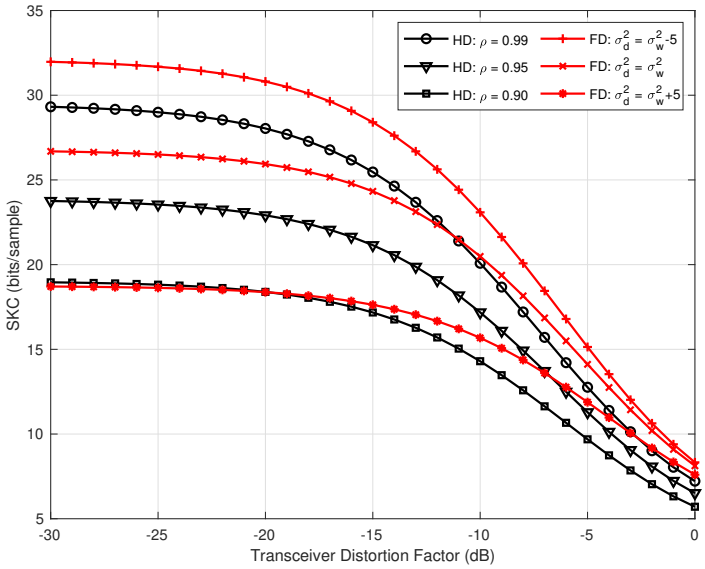


FIGURE 5.9: SKC variation against increasing transceiver HWIs under various conditions ($N_A = N_B = 2$, $P_A = P_B = 25$ dBm).

distortions will reduce the SKC since they increase the inconsistencies between the measurements. For IBFD probing, the additional noise introduced by imperfect hardware of RF cancellers significantly affects the SKC, which is different from its impact on the system capacity. For maximizing the system capacity, as long as the additional noise caused by cancellers is not greater than the thermal noise of receivers, the maximum IBFD gain can be obtained. However, for the SKC, the additional noise needs to be much smaller than the thermal noise to maximize the IBFD gain. With large transceiver HWIs, IBFD probing has obvious advantages over HD probing. The reason is that the penalty of RSI for IBFD becomes trivial when transceiver HWIs are significant.

5.5.2 Secret key generation protocol

In this section, we run Monte Carlo simulations to evaluate the performance of the proposed CFR-based SKG scheme in terms of KGR, KDR, and randomness. We set $\kappa_i = \beta_i = -70$ dB $\forall\ i$, $\alpha = 0.3$ and $T = 14$ for simulations, and we consider 60 dB of ASIC depth, which could be realized by a combination of antenna isolation and RF cancellation. The additional noise caused by RF cancellers is set to be identical to the receivers' thermal noise. Simulation results show that the KDR is decreased to 0 in most cases with the index-based reconciliation, so we show the KDR before the reconciliation. It should be noted that this scheme does not guarantee key agreement but will not cause key information leakage since the shared public message is independent of the key itself. For a lossy quantizer, it is efficient to remove a large portion of error bits through the index-based reconciliation method.

Figure 5.10 shows the KDR and KGR variation against the width of the guard stripe of the lossy quantizers. The samples are collected with IBFD probing. Larger ß yields a wider guard stripe, and the KDR is reduced at the cost of a slow KGR. It also illustrates the benefits of antenna arrays that more antennas benefit the SKG in terms of KGR. Although an enlarged antenna array increases the KDR, it can be reduced by a fairly wide guard strip. For instance, 4-antenna arrays at legitimate users can achieve a KGR of 869 bps and KDR approximate to the order of 10^{-4} with ß $= 0.3$. In contrast, 2-antenna arrays at legitimate users achieve a KGR of 275 bps and KDR approximate to the same order (10^{-4}) with ß $= 0.15$. Thus, an enlarged antenna array improves the key performance in general with appropriate guard stripe and reconciliation.

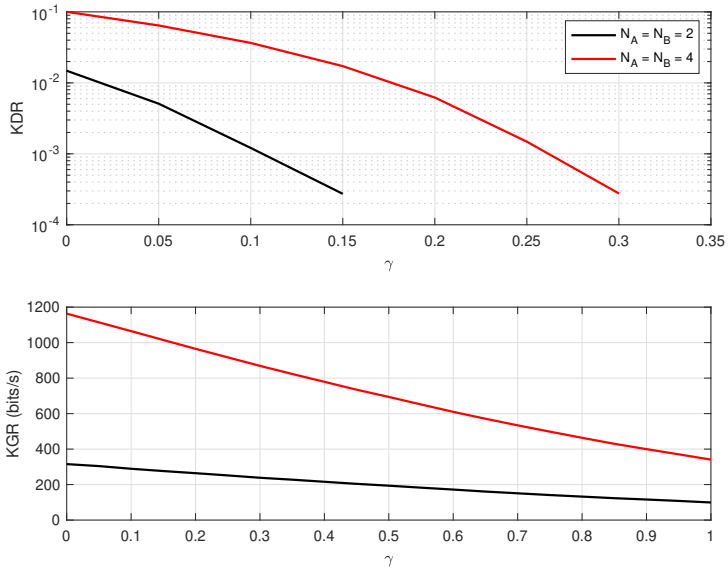
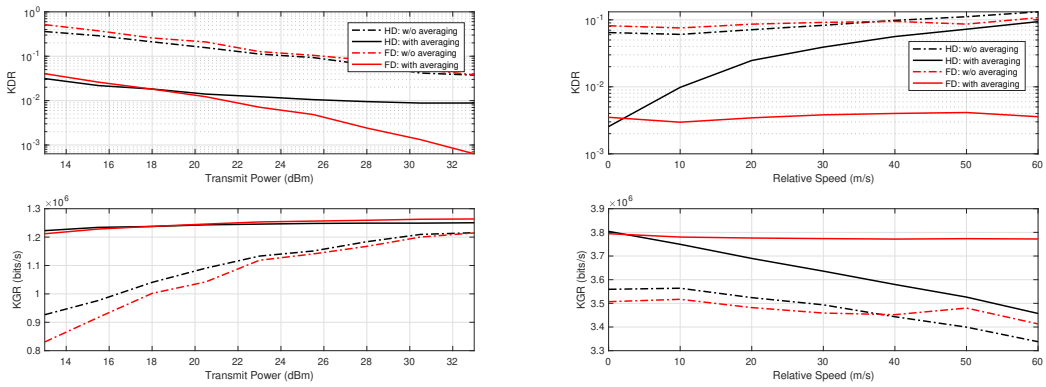


FIGURE 5.10: KDR and KGR variation with widening guard stripes of the lossy quantizers (IBFD probing, $T = 14$).



(a) KDR and KGR versus transmit power (SNR). (b) KDR and KGR versus relative speed between legitimate users.

FIGURE 5.11: Key performance comparison in HD and FD with various transmit power and Doppler shifts ($\Delta_\tau = 25$ ms, $L_b = 40$).

Figure 5.11 compares the key performance of HD and IBFD probing with different SNR (i.e., transmit power) and the relative speed of legitimate users. The quantization segment length is set to 40, i.e., $L_b = 40$ to guarantee the randomness of the generated key. A lossy quantizer with a narrow guard stripe (i.e., ß = 0.01) is utilized here to compare the KDR of different schemes explicitly; otherwise, there will not be error bits due to limited samples. A fast relative speed of legitimate users results in a large Doppler shift of the legitimate channel, reducing the correlation between consecutive CFRs (i.e., ρ). It shows that IBFD probing achieves lower KDR and higher KGR than its HD

counterpart with high transmit power and fast relative speed between legitimate users, which is consistent with the theoretical analysis. Furthermore, it shows the advantages of the CFR averaging operation, which fully utilizes the CFRs within the same band to significantly improve the key performance in terms of both KDR and KGR.

Table 5.2 shows the NIST test results for the generated key with different probing rates and quantization segment lengths, where binary matrix rank test and overlapping template test are not performed due to the limited length of the generated key. The random excursion and random excursion variant tests are performed with $J = 35$ (see [88]) due to the limited length, which may not be reliable. The relative speed of legitimate users is set to be 10 m/s, and the key is generated from IBFD probing samples. Results indicate that there is a tradeoff between the KGR and the randomness of the key. A faster probing rate (i.e., a shorter probing interval of $\Delta_\tau = 5$ ms) can increase the KGR, but the generated key only passes 5 of the 13 tests. Thus, in some ways, the generated key cannot be considered random. In contrast, a low probing rate (i.e., longer probing interval of $\Delta_\tau = 25$ ms) is necessary to guarantee the security of the SKG scheme. The probing rate strongly depends on the environment. In a dynamically changing environment, the probing rate should be adaptively tuned. In addition, it also demonstrates the necessity of segmental quantization. If the samples are not quantized in segments, large-scale fading will cause consecutive 0s or 1s in the generated bit sequence so that it will fail in some tests. With appropriate probing rate and sample segmentation, the generated key passes all 13 tests.

5.6 Conclusions

In this chapter, we investigated the advantages of employing IBFD radios to physical layer-based secret key generation schemes. It has been revealed that RSS-based scheme is seriously penalized compared to the CSI-based scheme, but CSI-based schemes have a higher implementation complexity due to the requirement of CSI acquisition. Thus, for implementation and efficiency considerations, we focused on a CFR-based SKG scheme in OFDM systems since existing 5G and WLAN networks utilize OFDM modulation and provide the CFR estimation resources in the protocol. We derived the secret key capacity of this scheme for both IBFD and HD radios by formulating the difference and

Table 5.2: NIST Test Results (P-values)

	$\Delta_\tau = 5$ ms $L_b = L_k$	$\Delta_\tau = 5$ ms $L_b = 40$	$\Delta_\tau = 25$ ms $L_b = L_k$	$\Delta_\tau = 25$ ms $L_b = 40$
Monobit	0.877	0.6877	0.6496	**0.9692**
Block frequency	0.0	0.9968	0.0	**0.9999**
Runs	0.0002	0.0	0.0049	**0.2456**
Longest runs	0.1421	0.0	0.0312	**0.2689**
DFT (Spectral)	0.0	0.0	**0.9346**	0.8753
Non-overlapping template	0.9461	0.9996	0.9964	**0.9999**
Maurer's universal statistical	0.0	0.0	**0.2311**	0.0141
Linear complexity	0.0	0.0	0.0779	**0.7876**
Serial	0.0	0.0	0.0535	**0.3483**
Approximate entropy	0.0	0.0	0.0676	**0.3431**
Cumulative sums	0.0	0.3859	0.0	**0.6899**
Random excursion	0.0266	0.0045	**0.2782**	0.0444
Random excursion variant	0.3192	0.0449	**0.2787**	0.1088
Number of passed tests	5 (×)	4 (×)	10 (×)	13 (✓)

correlation between the measurements of legitimate users. Two cases of the eavesdropper's location were considered, where the eavesdropper may be very close to legitimate users and experience correlated fading, or the eavesdropper is far away and experience independent fading. Additionally, the RSI effects were described via measurable metrics (e.g., analog cancellation level and noise of RF cancellers). The results illustrate that the fundamental limits of SKC come from the transceiver HWIs, while the non-simultaneous measurements limit the SKC for HD radios and SIC schemes (i.e., ASIC depth and SIC overheads) limit the SKC for IBFD radios. In the high-SNR regime, IBFD could have a larger probing error but achieve a higher SKC than HD with moving users. For stationary users, IBFD probing requires effective ASIC depth to outperform its HD counterpart, while appropriate SIC overheads can maximize the gain. Besides, IBFD probing is more robust to large transceiver HWIs than HD, and it could reduce the SKC degradation due to the presence of a close eavesdropper. MIMO systems can significantly improve the KGR, but longer SIC overheads are required for appropriate SIC. Thus, a longer probing duration is required for IBFD to provide benefits on the SKC over HD. For piratical implementation, the CFRs on subcarriers within the same coherence bandwidth are highly-correlated, so converting all of these CFRs into a bit sequence will decrease its randomness. To make full use of these CFRs, we averaged them

to reduce the effects of estimation errors. In addition, the effects of large-scale fading have to be removed to guarantee randomness of the generated bit sequence, which can be done by segmenting the samples into multiple blocks and independently quantizing them within each block.

Appendix 5.7 MMSE channel estimator

Assume $\mathbf{Y} = \mathbf{HX} + \mathbf{W}$, where $\mathbf{Y} \in \mathbb{C}^{M\times T}$, $\mathbf{H} \in \mathbb{C}^{M\times N}$, $\mathbf{X} \in \mathbb{C}^{N\times T}$, and $\mathbf{W} \in \mathbb{C}^{M\times T}$. The entries of matrices $\mathbf{X}$ and $\mathbf{W}$ are i.i.d. to Gaussian distribution with zero mean and variance of $\frac{P}{N}$ and σ_w^2, respectively. The MMSE estimate of $\mathbf{H}$ is given as

$$\begin{aligned}\hat{\mathbf{H}} &= \mathbf{Y}\left(\mathbb{E}\left\{\mathbf{Y}^\dagger\mathbf{Y}\right\}\right)^{-1}\mathbb{E}\left\{\mathbf{Y}^\dagger\mathbf{H}\right\} \\ &= \mathbf{Y}\left(\mathbf{X}^\dagger\mathbf{R}_{HH}\mathbf{X} + M\sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}^\dagger\mathbf{R}_{HH} \\ &= \mathbf{YX}^\dagger\left(\mathbf{R}_{HH}\mathbf{XX}^\dagger + M\sigma_w^2\mathbf{I}\right)^{-1}\mathbf{R}_{HH},\end{aligned} \tag{5.54}$$

where $\mathbf{R}_{HH} = \mathbb{E}\left\{\mathbf{H}^\dagger\mathbf{H}\right\}$ and $\mathbb{E}\left\{\mathbf{W}^\dagger\mathbf{W}\right\} = M\sigma_w^2\mathbf{I}$. It should have $T \geqslant \max(M, N)$ for an appropriate estimation; otherwise, the matrix in the parentheses of the inverse operation could be rank deficient. The estimation error is given as

$$\boldsymbol{\Delta}_H = \mathbf{H} - \hat{\mathbf{H}}. \tag{5.55}$$

According to the orthogonality principle [94], we have

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{\Delta}_H^\dagger\boldsymbol{\Delta}_H\right\} &= \mathbb{E}\left\{\left(\mathbf{H} - \hat{\mathbf{H}}\right)^\dagger\left(\mathbf{H} - \hat{\mathbf{H}}\right)\right\} \\ &= \mathbf{R}_{HH} - \mathbf{R}_{HH}\mathbf{X}\left(\mathbf{X}^\dagger\mathbf{R}_{HH}\mathbf{X} + M\sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}^\dagger\mathbf{R}_{HH} \\ &= \mathbf{R}_{HH}\left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^\dagger\mathbf{R}_{HH}\mathbf{X} + M\sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}^\dagger\mathbf{R}_{HH}\right] \\ &= \mathbf{R}_{HH}\left[\mathbf{I} - \left(\mathbf{XX}^\dagger\mathbf{R}_{HH} + M\sigma_w^2\mathbf{I}\right)^{-1}\mathbf{XX}^\dagger\mathbf{R}_{HH}\right] \\ &= M\sigma_w^2\mathbf{R}_{HH}\left(\mathbf{XX}^\dagger\mathbf{R}_{HH} + M\sigma_w^2\mathbf{I}\right)^{-1}.\end{aligned} \tag{5.56}$$

Assume each element of the error matrix is independent of each other, then we have $\mathbb{E}\left\{\boldsymbol{\Delta}_H\boldsymbol{\Delta}_H^\dagger\right\} = N\sigma_w^2\mathbf{R}_{HH}\left(\mathbf{XX}^\dagger\mathbf{R}_{HH} + M\sigma_w^2\mathbf{I}\right)^{-1}$. In the case of uncorrelated MIMO channel, i.e., $\mathbf{H}$ has i.i.d zero-mean complex Gaussian elements with variance of ϱ_H such

that $\mathbf{R}_{HH} = M\varrho_H\mathbf{I}$, we can rewrite (5.54) as

$$\hat{\mathbf{H}} = \mathbf{Y}\left(\mathbf{X}^\dagger\mathbf{X} + \frac{\sigma_w^2}{\varrho_H}\mathbf{I}\right)^{-1}\mathbf{X}^\dagger. \tag{5.57}$$

Besides, the estimation error can be described by the circular complex Gaussian model as

$$\boldsymbol{\Delta}_H \sim \mathcal{CN}\left(\mathbf{0}, N\frac{\sigma_w^2}{\frac{PT}{N} + \frac{\sigma_w^2}{\varrho_H}}\mathbf{I}_M\right), \tag{5.58}$$

where the entries of $\boldsymbol{\Delta}_H$ are i.i.d. to zero-mean complex Gaussian distribution with the variance of $\frac{\sigma_w^2}{\frac{PT}{N}+\frac{\sigma_w^2}{\varrho_H}}$.

Appendix 5.8 Proof of Lemma 5.1

The conditional mutual information is computed as

$$I\left(\mathbf{h}_A; \mathbf{h}_B \mid \mathbf{h}_E\right) = \log_2 \frac{|\mathbf{C}_{AE}|\,|\mathbf{C}_{BE}|}{|\mathbf{R}_E|\,|\mathbf{C}_{ABE}|}, \tag{5.59}$$

where $\mathbf{C}_{i,j} = \mathbb{E}\left\{\mathbf{v}_{ij}\mathbf{v}_{ij}^\dagger\right\}$ with $\mathbf{v}_{ij} = \left[(\mathbf{h}_i)^T, (\mathbf{h}_j)^T\right]^T$, $i, j \in \{A, B, E\}$, $\mathbf{C}_{ABE} = \mathbb{E}\left\{\mathbf{v}_{ABE}\mathbf{v}_{ABE}^\dagger\right\}$ with $\mathbf{v}_{ABE} = \left[(\mathbf{h}_A)^T, (\mathbf{h}_B)^T, (\mathbf{h}_E)^T\right]^T$. Thus, we have

$$|\mathbf{C}_{AE}| = \begin{vmatrix} \mathbf{R}_A & \mathbf{R}_{AE} \\ \mathbf{R}_{EA} & \mathbf{R}_E \end{vmatrix} = \mathbf{R}_A\mathbf{R}_E - \mathbf{R}_{AE}\mathbf{R}_{EA}, \tag{5.60}$$

$$|\mathbf{C}_{BE}| = \begin{vmatrix} \mathbf{R}_B & \mathbf{R}_{BE} \\ \mathbf{R}_{EB} & \mathbf{R}_E \end{vmatrix} = \mathbf{R}_B\mathbf{R}_E - \mathbf{R}_{BE}\mathbf{R}_{EB}, \tag{5.61}$$

$$\begin{aligned} |\mathbf{C}_{ABE}| = \begin{vmatrix} \mathbf{R}_A & \mathbf{R}_{AB} & \mathbf{R}_{AE} \\ \mathbf{R}_{BA} & \mathbf{R}_B & \mathbf{R}_{BE} \\ \mathbf{R}_{EA} & \mathbf{R}_{EB} & \mathbf{R}_E \end{vmatrix} &= \mathbf{R}_A\mathbf{R}_B\mathbf{R}_E + \mathbf{R}_{AB}\mathbf{R}_{BE}\mathbf{R}_{EA} + \mathbf{R}_{AE}\mathbf{R}_{BA}\mathbf{R}_{EB} \\ &\quad - \mathbf{R}_B\mathbf{R}_{AE}\mathbf{R}_{EA} - \mathbf{R}_A\mathbf{R}_{BE}\mathbf{R}_{EB} - \mathbf{R}_E\mathbf{R}_{AB}\mathbf{R}_{BA}. \end{aligned} \tag{5.62}$$

The variance matrix of $\mathbf{h}_E$ for both modes is given as

$$\mathbf{R}_E = \mathbb{E}\left\{\mathbf{h}_E \left(\mathbf{h}_E\right)^\dagger\right\} = \left(\rho'^2 \varrho_{AB} + (1-\rho'^2)\varrho_{AB}\right)\mathbf{I}_{N_A N_B} = \varrho_{AB}\mathbf{I}_{N_A N_B}. \tag{5.63}$$

Since the entries of $\mathbf{H}_{ij}$ and $\mathbf{\Delta}_{ij}$ are i.i.d. to zero-mean complex Gaussian distribution, and they are uncorrelated due to the orthogonality principle of MMSE estimators [94], it can be derived that $\mathbf{h}_i$ is distributed to zero-mean Gaussian distribution. The variance matrix of $\mathbf{h}_i$, $\forall\, i \in \{A, B\}$ is different for different probing modes, which is given as

$$\mathbf{R}_i^{HD} = \mathbb{E}\left\{\mathbf{h}_i^{HD} \left(\mathbf{h}_i^{HD}\right)^\dagger\right\} = \left(\varrho_{AB} + \sigma^2_{\Delta,i}\right)\mathbf{I}_{N_A N_B}. \tag{5.64}$$

$$\mathbf{R}_i^{FD} = \mathbb{E}\left\{\mathbf{h}_i^{FD} \left(\mathbf{h}_i^{FD}\right)^\dagger\right\} = \left(\varrho_{AB} + \sigma^2_{\Delta,i,1}\right)\mathbf{I}_{N_A N_B}, \tag{5.65}$$

The covariance matrices for HD probing are given as

$$\begin{aligned}\mathbf{R}_{AB}^{HD} &= \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_2)}\right)\left(\mathrm{vec}\left(\left(\mathbf{H}_{BA}^{(\tau_1)}\right)^T\right)\right)^\dagger\right\} \\ &= \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_2)}\right)\left(\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_1)}\right)\right)^\dagger\right\} = \rho\varrho_{AB}\mathbf{I}_{N_A N_B},\end{aligned} \tag{5.66}$$

$$\begin{aligned}\mathbf{R}_{AE}^{HD} &= \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_2)}\right)\left(\mathrm{vec}\left(\left(\mathbf{H}_{EA}^{(\tau_1)}\right)^T\right)\right)^\dagger\right\} \\ &= \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_2)}\right)\left(\mathrm{vec}\left(\left(\rho'\mathbf{H}_{BA}^{(\tau_1)} + \mathbf{\Lambda}_{BA}^{(\tau_1)}\right)^T\right)\right)^\dagger\right\} = \rho'\rho\varrho_{AB}\mathbf{I}_{N_A N_B},\end{aligned} \tag{5.67}$$

$$\begin{aligned}\mathbf{R}_{BE}^{HD} &= \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_{BA}^{(\tau_1)}\right)\left(\mathrm{vec}\left(\left(\mathbf{H}_{EA}^{(\tau_1)}\right)^T\right)\right)^\dagger\right\} \\ &= \mathbb{E}\left\{\mathrm{vec}\left(\mathbf{H}_{BA}^{(\tau_1)}\right)\left(\mathrm{vec}\left(\left(\rho'\mathbf{H}_{BA}^{(\tau_1)} + \mathbf{\Lambda}_{BA}^{(\tau_1)}\right)^T\right)\right)^\dagger\right\} = \rho'\varrho_{AB}\mathbf{I}_{N_A N_B},\end{aligned} \tag{5.68}$$

where ρ is the correlation coefficients of consecutive CFRs due to temporal changes within different probing slots τ_1 and τ_2 in HD such that [72]

$$\rho = \frac{Cov\left(\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_1)}\right), \mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_2)}\right)\right)}{\sqrt{Var\left(\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_1)}\right)\right)}\sqrt{Var\left(\mathrm{vec}\left(\mathbf{H}_{AB}^{(\tau_2)}\right)\right)}}. \tag{5.69}$$

The covariance matrices for IBFD probing can similarly be given as

$$\mathbf{R}_{AB}^{FD} = \varrho_{AB}\mathbf{I}_{N_A N_B}, \tag{5.70}$$

$$\mathbf{R}_{AE}^{FD} = \rho' \varrho_{AB} \mathbf{I}_{N_A N_B}, \tag{5.71}$$

$$\mathbf{R}_{BE}^{FD} = \rho' \varrho_{AB} \mathbf{I}_{N_A N_B}. \tag{5.72}$$

We will have $\mathbf{R}_{ji}^{\epsilon} = \left(\mathbf{R}_{ij}^{\epsilon}\right)^{\dagger} = \mathbf{R}_{ij}^{\epsilon}$ $\forall\ i, j \in \{A, B, E\}$ since its a real-number identity matrix. Insert these covariance matrices into Equation (5.59), we can obtain the SKC expressions of HD and IBFD modes as Equations (5.34) and (5.35) with $\varpi_i^{HD} = \varrho_{AB} + \sigma_{\Delta,i}^2$ and $\varpi_i^{FD} = \varrho_{AB} + \sigma_{\Delta,i,1}^2$.

Appendix 5.9 Proof of Lemma 5.2

The mutual information of $\mathbf{h}_A$ and $\mathbf{h}_B$ is given as

$$\begin{aligned} I\left(\mathbf{h}_A; \mathbf{h}_B\right) &= H\left(\mathbf{h}_A\right) + H\left(\mathbf{h}_B\right) - H\left(\mathbf{h}_A, \mathbf{h}_B\right) \\ &= \log_2 \frac{\left|\mathbf{R}_A\right| \left|\mathbf{R}_B\right|}{\left|\mathbf{C}_{AB}\right|} = \log_2 \frac{\left|\mathbf{R}_A\right| \left|\mathbf{R}_B\right|}{\left|\mathbf{R}_A\right| \left|\mathbf{R}_B - \mathbf{R}_{AB} \left(\mathbf{R}_A\right)^{-1} \mathbf{R}_{AB}\right|}, \end{aligned} \tag{5.73}$$

where the covariance matrices have been derived in Appendix 5.8. Insert these covariance matrices into Equation (5.73), we can obtain the secret key capacity for HD and IBFD probing under the spatial independence assumption as Equations (5.37) and (5.38), respectively.

Chapter 6

IBFD Integrated Sensing and Communication

6.1 Introduction

As stated earlier in Chapter 1, 6G will transform connected things in 5G to connected intelligence, where the network can have human-like cognition capabilities by enabling many potential services, which can be categorized as 1) high-accuracy localization and tracking; 2) simultaneous imaging, mapping, and localization; 3) augmented human sense; 4) gesture and activity recognition [95]. For this purpose, many applications in 6G require both high-performance sensing and wireless communications. In addition, future wireless communication systems will migrate to higher frequency bands, of which a large portion has been preliminarily assigned to radar systems. Therefore, it is critical to consider the coexistence and cooperation between the two systems, and integrated sensing and communication (ISAC) is regarded as a promising solution. It is envisioned that ISAC will play an important role and enable many emerging applications in future intelligent networks.

By sharing the spectrum and hardware for sensing and communication, ISAC can significantly reduce the implementation cost while enhancing spectral efficiency [96]. Besides, ISAC can improve the performance of both functions, where the physical information

Work in this chapter has been submitted to IEEE Transactions on Vehicular Technology in May 2023, with a preliminary version that is going to be presented at Asilomar Conference on Signals, Systems, and Computers November 2023.

obtained from sensing could help with simplifying the transmission or improving the communication throughput, and the communication signal from intended users can help with perceiving the surrounding environment. To this end, the Tx waveform should be appropriately designed since the two functions have different requirements. OFDM waveforms have been widely studied for ISAC to be compatible with existing communication networks. Studies reveal that OFDM waveforms incorporate frequency diversity and provide discrete phase modulation, offering a joint basis for sensing and communications and fulfilling the requirements for both applications [97, 98].

Although OFDM waveforms could provide high communication throughput and accurate radar sensing, they are not robust to large Doppler shifts, which implies OFDM modulation is not compatible with fast-moving targets (e.g., high-speed vehicles and trains). In contrast, providing reliable service in high-mobility scenarios has been envisioned as an important requirement for future wireless networks to enable many emerging applications (e.g., intelligent automotive systems and moving integrated access and backhaul (IAB)) [99, 100]. To meet this requirement, a new modulation scheme known as OTFS has been recently proposed, which constructs a delay-Doppler (DD) transmission scheme. Studies have demonstrated the advantages of OTFS over OFDM in both communication and sensing functions. OTFS is robust to the inter-channel interference (ICI) caused by large Doppler shifts, showing a much lower bit error ratio (BER) [101] and higher sensing capacity [102] than its OFDM counterpart in high-mobility scenarios. In addition, OTFS modulation also spreads the signal in the entire time-frequency grid to exploit the full diversity gains as OFDM, but it requires less CP than OFDM. Thus, the symbol duration is shortened, enhancing the sensing rate and performance [102] and spectral efficiency [103]. The sparsity and low-variability of the DD-domain channel allow OTFS to fully utilize the time and frequency diversity more efficiently than OFDM and simplify the communication design such as reducing the overhead of channel estimation and equalization [99, 104]. Furthermore, OTFS allows the two functions to corporate more deeply since radar sensing carries out parameter estimations based on the delay, Doppler, and angles, which align perfectly with the DD-based OTFS modulation. For instance, the channel effects can be compensated before communication signal transmission based on the radar-sensing parameters, allowing intended users to bypass channel estimation [105].

The design of OTFS transceivers is of great significance and has received increasing attention recently. Data symbols have a complex two-dimensional (2D) convolution with the DD-domain channel in OTFS systems, making channel estimation and data detection challenging. Conducting the channel estimation in the time-frequency (TF) domain as in OFDM systems results in the high complexity and overhead for OTFS systems. The existing pilot pattern requires large overhead to prevent the pilot from being contaminated under doubly-dispersive channels [106]. A pilot-aided threshold-based channel estimator is proposed for OTFS in [107], where the pilot, guard, and symbols are arranged appropriately in the DD domain to avoid interference under multipath channels with large Doppler shifts. The scheme is compatible with MIMO and multi-user scenarios. In [106], the pilot pattern is further studied for MIMO scenarios, and a low-complexity detector and a low-overhead pilot pattern are proposed for large-scale antenna arrays. A 3 dimension (3D)-structural orthogonal matching pursuit (OMP) algorithm is proposed in [108] to address the challenges for channel estimation in massive MIMO systems due to a large number of antennas, showing accurate estimation with low pilot overhead. In [109], the linear ZF and MMSE equalizers are investigated for OTFS systems, which reduces the complexity without performance degradation. Authors of [100] further extend the work to MIMO cases and consider imperfect CSI cases. However, linear equalizers generally have poor performance [104], so studies usually focus on nonlinear receivers. A message passing (MP) algorithm-based receiver is proposed in [101] by deriving the input-output relation of the OTFS system, which can eliminate ICI and inter-symbol interference (ISI) and mitigate inter-Doppler interference (IDI), showing impressive performance. Furthermore, a reservoir computing-based design is proposed in [103], where a one-shot trained neural network can cope with the channel variations within multiple OTFS frames. It shows a lower BER in the low-SNR regime than conventional methods and a lower time complexity than MP. A joint channel estimation and data detection method are proposed in [110], where the unknown symbols are regarded as virtual pilots to enhance the estimation accuracy, showing improved performance in terms of estimation errors and BER. In addition to these designs, abundant techniques are explored to construct low-complexity OTFS receivers, which are well-summarized in [111].

The effectiveness of the OTFS-ISAC system is verified in [112], where a maximum likelihood (ML) estimator is used for radar parameters estimation, and the MP detector is

used for communication data detection. It demonstrates that OTFS can achieve as accurate radar sensing as state-of-art radar waveforms (e.g., frequency-modulated continuous wave (FMCW)) and conventional OFDM waveforms with satisfactory communication performance. A matched filter algorithm is proposed for range and velocity estimation with OTFS waveforms in [102], showing the inherent advantages of OTFS over OFDM through exploring the structure of the delay-Doppler domain channel. Authors in [113] further consider the OTFS-ISAC with MIMO systems, and a hybrid digital-and-analog beamforming scheme is proposed, achieving the Cramer-Rao lower bound for radar parameter estimation. In [114], a novel ISAC framework based on spatially-spread OTFS is proposed. where the angular domain discretization is enabled by spatial spreading/despreading, simplifying the related estimation and detection problems. In addition, a symbol-wise precoding scheme with only the sensing information (e.g., delay, Doppler shift, and angles obtained from radar sensing) instead of CSI is proposed.

The main contribution of our work is that we propose a novel ISAC framework that is compatible with fast-moving targets and IBFD communications. Although there are many channel estimation, data detection, and sensing schemes proposed in the literature, they still have relatively high complexity, while we significantly reduce the complexity via the MUSIC algorithm and a time domain-based interference cancellation and parameter estimation algorithm. Then, we study the SIC schemes for large Doppler shifts and OTFS symbols to enable IBFD communication. Additionally, we crossly consider the two functions and discuss the efficient transmission scheme.

The rest of this chapter is organized as follows. In Section 6.2, we first describe the ISAC network, followed by wireless channel representations in different domains and an introduction to OTFS modulation. Section 6.3 gives the proposed low-complexity radar sensing design and the cons of existing designs. Then, in Section 6.4, we give a communication design by focusing on the SIC scheme that is compatible with OTFS waveforms and rapidly-changing channels, followed by a discussion of the transmission scheme by crossly conceding the two functions. Finally, simulation results are given in Section 6.5, and conclusions are drawn in Section 6.6.

6.2 Preliminaries

6.2.1 System model

We consider an ISAC network as Figure 6.1 shows, where OTFS modulation is employed at all nodes. The OTFS-ISAC transmitter generates and sends the ISAC signals, while the co-located OTFS-ISAC receiver obtains the backscattered signal and estimates the radar sensing parameters. Meanwhile, the transmitted ISAC signal is sent to the intended OTFS receiver for communication purposes. To improve spectral efficiency and reduce latency, we assume the OTFS-ISAC transceiver operates in IBFD mode so that it simultaneously receives the signal transmitted from the intended OTFS transmitter. Assume all OTFS transceivers are equipped with N_t- and N_r-element transmitting and receiving uniform linear arrays, respectively. We consider a frame with $M\Delta f$ total bandwidth and NT duration, where Δf is the subcarrier spacing, $T = 1/\Delta f$ is the symbol duration, M and N is the number of subcarriers and symbols, respectively.

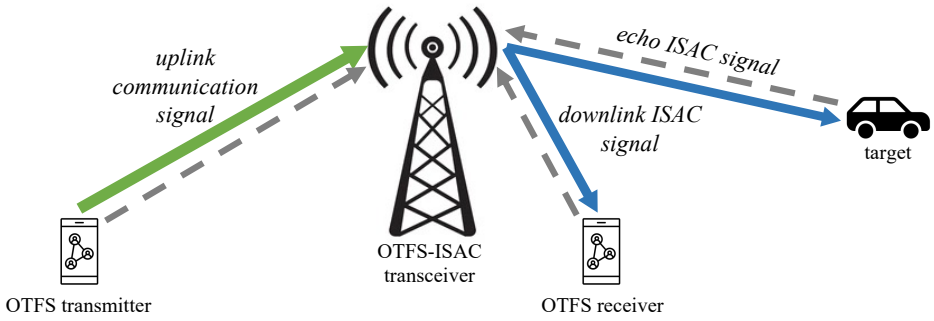


FIGURE 6.1: Integration of sensing with IBFD communication, i.e., IBFD-ISAC, where the ISAC node simultaneously transmits ISAC signals and receives uplink communication signals and echoes from targets.

6.2.2 Communication channel models

6.2.2.1 Delay-Doppler representation

The delay-Doppler domain channel deserves to be investigated due to its slow time-varying property [105]. The channel coefficients in the delay-Doppler domain are limited due to the limited number of objectives in the surrounding environment, yielding a sparse

representation of dimension $N_r \times N_t$ as [115][1]

$$\mathbf{H}(\tau,\nu) = \sum_{l=0}^{L-1} h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l)\delta(\tau-\tau_l)\delta(\nu-\nu_l), \tag{6.1}$$

where L is the number of paths; h_l is the complex channel coefficient; τ_l is the delay; ν_l is the Doppler shift; θ_l and ϕ_l represent the angle of arrival (AoA) and angle of departure (AoD), respectively; and

$$\mathbf{a}_r(\theta_l) = \left[1, e^{-j\pi\sin\theta_l}, \ldots, e^{-j\pi(N_r-1)\sin\theta_l}\right], \tag{6.2}$$

$$\mathbf{a}_t(\phi_l) = \left[1, e^{-j\pi\sin\phi_l}, \ldots, e^{-j\pi(N_t-1)\sin\phi_l}\right]. \tag{6.3}$$

Let ℓ_l and ι_l be the integer delay and Doppler indices in the delay-Doppler grid, which are related to the actual delay and Doppler shift of the l^{th} path as

$$\begin{aligned} \tau_l &= \frac{\ell_l}{M\Delta f} = \frac{\ell_l T}{M} < \tau_{\max} < T, \\ \nu_l &= \frac{\iota_l}{NT} = \frac{\iota_l \Delta f}{N} < \nu_{\max} < \Delta f. \end{aligned} \tag{6.4}$$

Then, we can obtain the discrete delay-Doppler channel representation as

$$\mathbf{H}_{\mathrm{dd}}[m,n] = \begin{cases} h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l), & \text{if } m=\ell_l, n=\iota_l \\ 0, & \text{otherwise} \end{cases}. \tag{6.5}$$

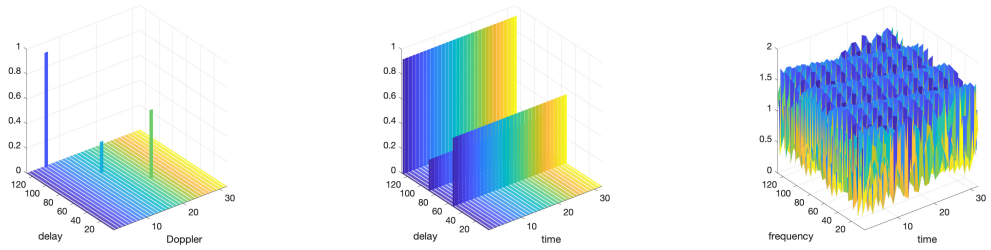
It should be noted that ℓ_l and ι_l are integers, as stated above. They represent the actual delay and Doppler in a discrete grid and are related to the actual delay and Doppler as $\ell_l = \lceil \tau_l M\Delta f \rfloor$ and $\iota_l = \lceil \nu_l NT \rfloor$.

6.2.2.2 Time-delay representation

The time-delay (TD) representation can be given as

$$\begin{aligned} \mathbf{H}(t,\tau) &= \int_{\nu} \mathbf{H}(\tau,\nu) e^{j2\pi\nu(t-\tau)} d\nu \\ &= \sum_{l=0}^{L-1} h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l)\delta(\tau-\tau_l) e^{j2\pi\nu_l(t-\tau)}, \end{aligned} \tag{6.6}$$

[1] We consider an antenna array with $\lambda/2$ spacing, where λ represents the wavelength.



(a) DD channel representation (b) TD channel representation (c) TF channel representation

FIGURE 6.2: The channel representations in different domains.

which is consistent with the TDL models. The discrete representation can be given as

$$\mathbf{H}_{\mathrm{dt}}[m,n] = \begin{cases} h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l) e^{j2\pi \frac{n\kappa_l}{N}}, \ \forall\ n, & \text{if } m = \ell_l \\ 0, & \text{if } m \neq \ell_l \end{cases}. \tag{6.7}$$

6.2.2.3 Time-frequency representation

The time-frequency representation can be given as

$$\begin{aligned} \mathbf{H}(t,f) &= \int_{\tau} \mathbf{H}(t,\tau) e^{-j2\pi f\tau} d\tau \\ &= \sum_{l=0}^{L-1} h_l' \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l) e^{j2\pi\nu_l t} e^{-j2\pi\tau_l f}, \end{aligned} \tag{6.8}$$

where $h_l' = h_l e^{j2\pi\nu_l\tau_l}$. Discretizing the time and frequency axis at nT and $m\Delta f$, we have

$$\mathbf{H}_{\mathrm{tf}} = \sum_{l=0}^{L-1} h_l' \left(\mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l)\right) \odot \left(\mathbf{b}(\tau_l)\mathbf{c}^{\dagger}(\nu_l)\right), \tag{6.9}$$

where $\odot$ denotes element-wise multiplication, and

$$\mathbf{b}(\tau_l) = \left[1, e^{j2\pi\tau_l\Delta f}, \ldots, e^{j2\pi(M-1)\tau_l\Delta f}\right]^T, \tag{6.10}$$

$$\mathbf{c}(\nu_l) = \left[1, e^{j2\pi\nu_l T}, \ldots, e^{j2\pi(N-1)\nu_l T}\right]^T. \tag{6.11}$$

Figure 6.2 shows an instance of the wireless channel in the three domains, where the TF channel is doubly-selective due to the Doppler shifts.

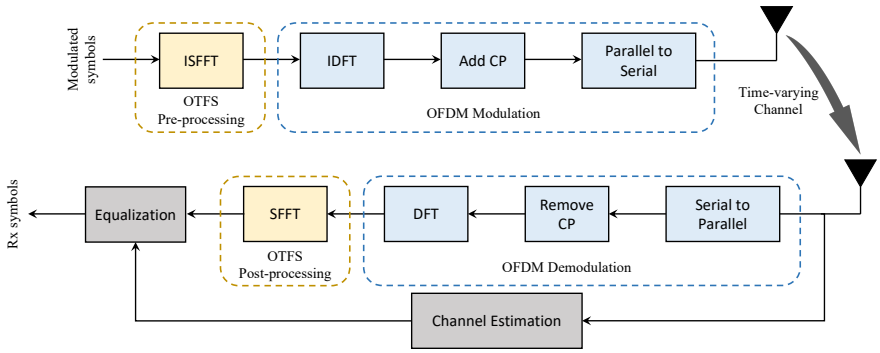


FIGURE 6.3: OTFS architecture based on OFDM transceivers with pre- and post-processing blocks.

6.2.3 Radar channel models

For the radar application, we can assume a mono-static scenario, which means the AoA and AoD are identical for radar channels [113]. The Radar channel can also be described by Equations (6.1), (6.6), and (6.8) with L denoting the number of targets; $\tau_l = \frac{2r_l}{c}$ being the round-trip delay with r_l and c denoting the target range and light speed, respectively; $\nu_l = \frac{2v_l f_c}{c}$ being the round-trip Doppler shift with v_l and f_c denoting the speed and carrier frequency, respectively; and $N_t = N_r$ and $\theta_l = \phi_l$ under the mono-static radar assumption.

6.2.4 OTFS modulation

OTFS can be implemented by adding inverse symplectic fast Fourier transform (ISFFT) and symplectic fast Fourier transform (SFFT) blocks to OFDM systems, as Figure 6.3 shows. Let $\mathbf{X}_{\text{dd}} \in \mathbb{C}^{M\times N}$ denote the symbols in the delay-Doppler grid, which is converted into the time-frequency domain ($\mathbf{X}_{\text{tf}} \in \mathbb{C}^{M\times N}$) as

$$\mathbf{X}_{\text{tf}}[m,n] = \frac{1}{MN}\sum_{p=0}^{M-1}\sum_{q=0}^{N-1}\mathbf{X}_{\text{dd}}[p,q]e^{j2\pi\left(\frac{nq}{N}-\frac{mp}{M}\right)}. \tag{6.12}$$

The rest processing follows OFDM systems to convert it into a continuous-time signal, and it is mapped into the N_t Tx antennas through the precoder $\mathbf{V} \in \mathbb{C}^{N_t\times 1}$ as

$$\mathbf{x}(t) = \mathbf{V}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\mathbf{X}_{\text{tf}}[m,n]g_{\text{tx}}(t-nT)e^{j2\pi m\Delta f(t-nT)}, \tag{6.13}$$

where $g_{\text{tx}}(t)$ is the Tx pulse shaping filter. After the MIMO channel, the receiver combines the received vector by a combiner $\mathbf{U} \in \mathbb{C}^{N_r \times 1}$, yielding the signal as

$$\begin{aligned}\mathbf{y}(t) &= \sum_{l=0}^{L-1} h_l \mathbf{U}^{\dagger} \mathbf{a}_r(\theta_l) \mathbf{a}_t^{\dagger}(\phi_l) \mathbf{x}(t-\tau_l) e^{j2\pi\nu_l t} g_{\text{rx}}(t) \\ &= \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} h_l' \mathbf{X}_{\text{tf}}[m,n] e^{j2\pi\nu_l t} e^{j2\pi m\Delta f(t-\tau_l-nT)} g_{\text{tx}}(t-\tau_l-nT) g_{\text{rx}}(t),\end{aligned} \tag{6.14}$$

where $h_l' = h_l \mathbf{U}^{\dagger} \mathbf{a}_r(\theta_l) \mathbf{a}_t^{\dagger}(\phi_l) \mathbf{V}$, and $g_{\text{rx}}(t)$ is the Rx shaping filter. Then, it follows the OFDM receivers, yielding the time-frequency signal as

$$\begin{aligned}y_{\text{tf}}(t,f) &= \int_{\tau} \mathbf{y}(\tau) g_{\text{rx}}(\tau - t) e^{-j2\pi f\tau} d\tau \\ &= \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} h_l' \mathbf{X}_{\text{tf}}[m,n] e^{-j2\pi m\Delta f \tau_l} \\ &\quad \times \int_{\tau} g_{\text{rx}}(\tau - t)\, g_{\text{tx}}(\tau - \tau_l - nT)\, e^{j2\pi(m\Delta f + \nu_l - f)\tau} d\tau.\end{aligned} \tag{6.15}$$

Discretizing the time and frequency axis at nT and $m\Delta f$, the discrete time-frequency grid can be given as

$$\begin{aligned}\mathbf{Y}_{\text{tf}}[m,n] &= \sum_{l=0}^{L-1} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} h_l' \mathbf{X}_{\text{tf}}[p,q] e^{-j2\pi p\Delta f \tau_l} \\ &\quad \times \int_{\tau} g_{\text{rx}}(\tau - nT)\, g_{\text{tx}}(\tau - \tau_l - qT)\, e^{j2\pi(p\Delta f + \nu_l - m\Delta f)\tau} d\tau.\end{aligned} \tag{6.16}$$

Finally, the symbols can be converted back into the delay-Doppler domain through SFFT as

$$\mathbf{Y}_{\text{dd}}[m,n] = \frac{1}{NM} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \mathbf{Y}_{\text{tf}}[p,q] e^{j2\pi\left(\frac{mp}{M} - \frac{nq}{N}\right)}. \tag{6.17}$$

6.3 Radar sensing design

The ISAC framework consists of two functions: 1) radar sensing-related parameter (e.g., delay, Doppler, and angles information) estimation at the ISAC node; 2) data recovery at the communication receiver. The targets can be tracked with the estimated range, velocity, and angle information [116]. In this section, we study the angle information, range, and speed estimation methods with OTFS waveforms, while the OTFS-based communication design is given later in Section 6.4.

Benefiting from the DD-domain constructive of OTFS waveforms and the fact that the sensing parameters are explicit in the DD domain as Figure 6.2(a) shows, the sensing-related parameters ($\hat{\mathbf{\Gamma}} = \left\{\hat{h}_0, \ldots, \hat{h}_{L-1}, \hat{\tau}_0, \ldots, \hat{\tau}_{L-1}, \hat{\nu}_0, \ldots, \hat{\nu}_{L-1}, \hat{\theta}_0, \ldots, \hat{\theta}_{L-1}\right\}$) can be inferred with the known transmitted DD grid $\mathbf{X}_{\mathrm{dd}}$ and received DD grid $\mathbf{Y}_{\mathrm{dd}}$ via existing estimators, e.g., the maximum likelihood estimator as [112]

$$\hat{\mathbf{\Gamma}} = \arg \min_{\mathbf{\Gamma} \in \mathbb{C}^L \times \mathbb{R}^L \times \mathbb{R}^L \times \mathbb{R}^L} \left\| \mathbf{Y}_{\mathrm{dd}} - \sum_{l=0}^{L-1} \mathbf{G}_{\mathrm{dd}} \mathbf{X}_{\mathrm{dd}} \right\|, \tag{6.18}$$

where $\mathbf{G}_{\mathrm{dd}}$ denotes the delay-Doppler domain channel coefficients. However, these algorithms have very high complexity. Thus, in this section, we will derive a new framework to reduce the complexity.

6.3.1 Angle information estimation

Usually, we will have the covariance matrix of transmitted data symbols as

$$\mathbb{E}\left\{\mathrm{vec}\left(\mathbf{X}_{\mathrm{dd}}\right) \mathrm{vec}\left(\mathbf{X}_{\mathrm{dd}}\right)^{\dagger}\right\} = \sigma_x^2 \mathbf{I}. \tag{6.19}$$

The ISFFT operation that converts the DD grid into the TF grid can be written in the matrix form as

$$\mathbf{X}_{\mathrm{tf}} = \mathbf{F}_M \mathbf{X}_{\mathrm{dd}} \mathbf{F}_N^{\dagger}, \tag{6.20}$$

where $\mathbf{F}_M$ and $\mathbf{F}_N$ denote the DFT and IDFT matrix, respectively, with $\mathbf{F}_M \mathbf{F}_M^{\dagger} = M\mathbf{I}$ and $\mathbf{F}_N^{\dagger} \mathbf{F}_N = N\mathbf{I}$. Thus, we will have

$$\mathbb{E}\left\{\mathrm{vec}\left(\mathbf{X}_{\mathrm{tf}}\right) \mathrm{vec}\left(\mathbf{X}_{\mathrm{tf}}\right)^{\dagger}\right\} = \sigma_x^2 \mathbf{I}, \tag{6.21}$$

which indicates that the spatial covariance matrix of the received data symbols has the same form as in the OFDM systems, which is independent of delays and Dopplers [98]. Thus, we propose to employ the conventional multiple signal classification (MUSIC) algorithms for AoA and AoD estimation for the OTFS-ISAC systems. Under the static-mono radar assumption, we have $\theta_l = \phi_l$, and the spatial covariance matrix of the received signal can be divided into the signal and noise subspaces, yielding the MUSIC spectrum as

$$f\left(\theta\right) = \frac{1}{\mathbf{a}_r(\theta) \mathbf{Q}_n \mathbf{Q}_n^{\dagger} \mathbf{a}_r^{\dagger}(\theta)}, \tag{6.22}$$

where $\mathbf{Q}_n$ contains eigenvectors corresponding to the noise subspace that can be obtained from the eigen-decomposition of the spatial covariance matrix. The MUSIC algorithm has been well studied in the literature, so we do not introduce it in detail, while readers are referred to [5, 98] for the implementation. Assume the Rx antenna arrays have N_r elements, which are half-wavelength spaced, and there are K non-coherent targets with AoAs $\{\theta_k\}_{k=1}^{K}$, then the output of the matched filters can be given as

$$\mathbf{y}(t) = [\mathbf{a}_r(\theta_1), \mathbf{a}_r(\theta_2), \ldots, \mathbf{a}_r(\theta_K)]\, \mathbf{x}(t) + \mathbf{n}(t), \tag{6.23}$$

where $\mathbf{a}_r(\theta_k) = \left[1, e^{-j\pi \sin\theta_k}, \ldots, e^{-j\pi(N_r-1)\sin\theta_k}\right]$ denotes the receive steering vector; α_k represents the amplitude and f_k denotes the Doppler frequency; and $\mathbf{n}(t)$ is the noise vector. With L Rx samples, the covariance matrix of the signal can be estimated as

$$\mathbf{R}_y = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}(l)\mathbf{y}(l)^{\dagger}. \tag{6.24}$$

Performing Eigen decomposition, it can be represented as

$$\mathbf{R}_y = \mathbf{E}_s \mathbf{\Sigma}_s \mathbf{E}_s^{\dagger} + \mathbf{E}_n \mathbf{\Sigma}_n \mathbf{E}_n^{\dagger}, \tag{6.25}$$

where $\mathbf{\Sigma}_s$ consists of the largest K eigenvalues along the diagonal and $\mathbf{\Sigma}_n$ consists of the rest $N_r - K$ eigenvalues; $\mathbf{E}_s$ and $\mathbf{E}_n$ can be regarded as the signal subspace and noise subspace, respectively. The MUSIC spectrum can be thus formulated as

$$S(\theta) = \frac{1}{\mathbf{a}_r^{\dagger}(\theta)\mathbf{E}_n\mathbf{E}_n^{\dagger}\mathbf{a}_r(\theta)}, \tag{6.26}$$

whose K peaks correspond to the AoAs of the K Rx echo signals, i.e., the azimuth of the K targets. One can plot the MUSIC spectrum by continuously changing the value of θ and finding the K peaks to obtain the estimated AoAs. However, the accuracy is limited by the discretization at which the MUSIC spectrum is evaluated, and it requires human interaction to decide on the largest K peaks. Thus, the root-MUSIC algorithm is usually utilized to avoid human interaction. The AoA estimation processing via root-MUSIC algorithm can be summarized as follows.

1) Calculate the covariance matrix of Rx samples as Equation 6.24.

2) Perform Eigen decomposition to the covariance matrix and obtain the eigenvectors corresponding to the $N_r - K$ eigenvalues, which span the noise subspace.

3) Calculate $\mathbf{C} = \mathbf{E}_n \mathbf{E}_n^{\dagger}$ and obtain C_l by summing the l^{th} diagonal of $\mathbf{C}$.

4) Find the zeros of the resulting polynomial $\sum_{l=-(N_r-1)}^{N_r-1} C_l e^{-j\pi l \sin\theta}$.

5) Choose the K roots $(z_k, k = 1, \ldots, K)$ closest to the unit circle from the $N_r - 1$ roots within the unit circle.

6) Calculate the AoA as $\theta_k = \cos^{-1}\left[\frac{-\Im \ln(z_k)}{\pi}\right], \forall\ k$.

6.3.2 Range and velocity estimation

6.3.2.1 A low-complexity TF domain method

There is a low-complexity range and velocity extraction method that is widely used in OFDM systems, which requires only discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) operations in the time-frequency domain. The fact that OTFS modulation is implemented based on OFDM naturally leads to the idea of keeping use of this method for OTFS systems. This method is constructed based on the input-output relationship in the time-frequency domain, which can be obtained from the derivations in Section 6.2.4 as[2]

$$\mathbf{Y}_{\mathrm{tf}}[m,n] = \sum_{l=0}^{L-1}\sum_{p=0}^{M-1}\sum_{q=0}^{N-1} h_l'' \mathbf{X}_{\mathrm{tf}}[p,q] e^{-j2\pi p\Delta f \tau_l} e^{j2\pi\nu_l qT} \sum_{i=0}^{M-1} e^{j2\pi\frac{i(p-m)}{M}} e^{j2\pi\frac{i\nu_l}{M\Delta f}}, \quad (6.27)$$

where the term $e^{j2\pi\frac{i\nu_l}{M\Delta f}}$ denotes the ICI due to Doppler shifts. It can be seen from this relationship that the range and Doppler shift yield orthogonal effects on the time and frequency dimensions if $\frac{i\nu_l}{M\Delta f} \approx 0$. Thus, assume $\nu_{\max} \ll \Delta f$ and ignore the ICI terms, the time-frequency input-output relationship can be approximated as

$$\mathbf{Y}_{\mathrm{tf}} = \sum_{l=0}^{L-1} h_l'' \mathbf{b}(\tau_l)\mathbf{c}^{\dagger}(\nu_l)\mathbf{X}_{\mathrm{tf}}, \quad (6.28)$$

[2]Assume a rectangular function is adopted for the pulse shaping filter, and the guard interval is properly chosen, so that the receiver can cut the same OFDM symbol from its observations, ignoring the time shift of the rect-function.

which is the basis of the low-complexity range and velocity extraction method[5]. Equation (6.28) suggests that the Doppler shift causes an identical phase shift over subcarriers of the same symbol and a linear phase shift between the consecutive modulation symbols on the same subcarrier. In contrast, the range causes an identical phase shift to the consecutive modulation symbols on the same subcarrier and a linear phase shift over subcarriers of the same symbol. Thus, the range of velocity of objects can be detected by processing the transmitted and received time-frequency grids as follows [97].

1) Form the transmitted and received modulation symbol matrices in the 2D time-frequency grid as $\mathbf{Y}_{\text{tf}}$ and $\mathbf{X}_{\text{tf}}$, and then obtain the matrix $\mathbf{D}_{RV} = \mathbf{Y}_{\text{tf}}./\mathbf{x}_{\text{tf}}$, where "./" means divide the corresponding elements of two matrices.

2) *Optional:* Apply windowing filters to the frequency and time axis of $\mathbf{D}_{RV}$. *It can improve the detection performance by reducing the side lobes caused by the subsequent Fourier transforms.*

3) Perform discrete Fourier transform to each row of $\mathbf{D}_{RV}$.

4) Perform inverse discrete Fourier transform to each column of $\mathbf{D}_{RV}$.

After the set of operations, $\mathbf{D}_{RV}$ represents the detection image, where the x-axis is the velocity, and the y-axis is the range. Multiple objects can be detected simultaneously since only linear processing is performed. The maximum detectable range and velocity are determined by the subcarrier spacing and carrier frequency, which are given as

$$r_{\max} = \frac{c}{2\Delta f} \text{ and } v_{\max} = \frac{c\Delta f}{2f_c}, \tag{6.29}$$

respectively. The radar sensing resolution is determined by the resource grid size M and N given as

$$r_{\text{res}} = \frac{c}{2\Delta f N} \text{ and } v_{\text{res}} = \frac{c\Delta f}{2f_c M}. \tag{6.30}$$

The resolution is critical for detecting and separating the object. Therefore, a large number of subcarriers is desired to provide high-resolution range detection, and a large number of symbols is desired to provide high-resolution velocity detection. Figure 6.4 shows the normalized detection image with a different number of subcarriers and symbols under the same condition. A standard 5G NR waveform with 15 kHz subcarrier spacing and 2.5 GHz carrier frequency is utilized, yielding 9.85 km of detectable range and ±337

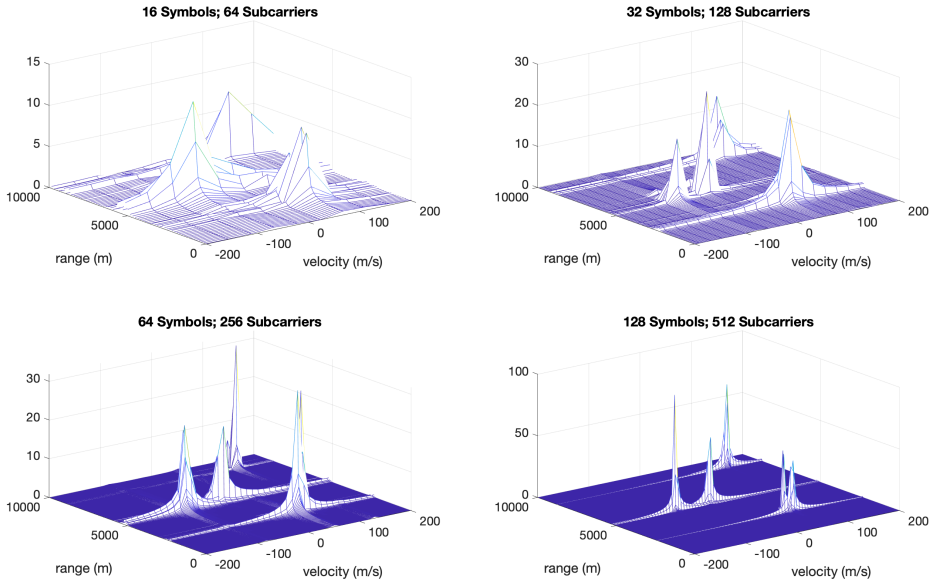


FIGURE 6.4: Normalized detection image calculated from the modulation symbols for five objects with ranges $R \in \{2.30, 2.36, 5.10, 5.23, 8.76\}$ km and velocities $V \in \{63, 48, -80, -10, 130\}$ m/s.

m/s detectable velocity. The velocity resolution of the four resource grids is 42.13 m/s, 21.06 m/s, 10.53 m/s, and 5.27 m/s, respectively. The results show that this method can effectively detect all objects with relatively high resolutions.

6.3.2.2 A low-complexity TD domain method

However, with increasing Doppler shifts approaching Δf, the ICI term cannot be ignored in Equation (6.27) cannot be ignored, which destroys the basis of the time-frequency domain-based method. To achieve accurate radar sensing with large Doppler shifts, we could exploit the delay-Doppler domain input-output relationship of OTFS systems, which can be written as

$$\begin{aligned}\mathbf{Y}_{\mathrm{dd}}[m,n] &= \sum_{m'=0}^{M-1}\sum_{n'=0}^{N-1}\chi(m,m',n,n')\mathbf{X}_{\mathrm{dd}}[m',n'] \\ &= \chi_{m,n}\mathbf{X}_{\mathrm{dd}}[m,n] + \chi(m,m',n,n')\sum_{(m',n')\neq(m,n)}^{M-1}\sum^{N-1}\mathbf{X}_{\mathrm{dd}}[m',n'],\end{aligned} \tag{6.31}$$

where $\chi(m, m', n, n')$ denotes the ISI coefficients and can be approximated assuming a rectangular pulse shaping filter as in [112], denoted as

$$\chi(m,m',n,n') \approx \frac{1}{NM}\sum_{l=0}^{L-1} h_l'' \frac{1-e^{j2\pi(m'-m+\nu_l NT)}}{1-e^{j2\pi\frac{(m'-m+\nu_l NT)}{N}}} \frac{1-e^{j2\pi(n'-n+\tau_l M\Delta f)}}{1-e^{j2\pi\frac{(n'-n+\tau_l M\Delta f)}{N}}} \times e^{-j2\pi\nu_l\frac{m'}{M\Delta f}} \begin{cases} 1 & \text{, if } m' < \frac{\tau_l}{T/M} \\ e^{-j2\pi(\frac{n'}{N}+\nu_l T)} & \text{, otherwise} \end{cases} \tag{6.32}$$

Since both $\mathbf{X}_{\text{dd}}$ and $\mathbf{Y}_{\text{dd}}$ are known by the radar receiver, it can infer the range and velocity parameters $\hat{\mathbf{\Gamma}}_{RV} = \left\{\hat{h}_0, \ldots, \hat{h}_{L-1}, \hat{\tau}_0, \ldots, \hat{\tau}_{L-1}, \hat{\nu}_0, \ldots, \hat{\nu}_{L-1}\right\}$ via existing estimators, e.g., the maximum likelihood estimator as [112]

$$\hat{\mathbf{\Gamma}}_{RV} = \arg \min_{\mathbf{\Gamma}_{RV} \in \mathbb{C}^L \times \mathbb{R}^L \times \mathbb{R}^L} \left\| \mathbf{Y}_{\text{dd}} - \sum_{l=0}^{L-1} \mathbf{G}_{\text{dd}} \mathbf{X}_{\text{dd}} \right\|, \tag{6.33}$$

where $\mathbf{G}_{\text{dd}}$ denotes the delay-Doppler domain channel coefficients that can be derived from Equation (6.31). This estimation problem has a lower complexity than the one in Equation 6.18 due to the reduced number of estimating parameters, i.e., the angle information, which has been obtained via the MUSIC algorithm. However, its complexity is still high due to the complicated input-output relationship and the required iterative processing. Inspiring by the fact that the ISAC receiver has the full knowledge of the transmitted ISAC signal and the time domain input-output relationship is simple as Equation (6.14) suggests, we propose a time domain-based interference cancellation and parameter estimation scheme to extract the range and velocity parameters $\mathbf{\Gamma}_{RV}$. To understand the rationale behind this approach, we formulate the input-output relationship in the delay-time domain as

$$\text{vec}\left(\mathbf{Y}_{\text{td}}\right) = \sum_{l=0}^{L-1} \mathbf{G}_{\text{td}}^l \text{vec}\left(\mathbf{X}_{\text{td}}\right), \tag{6.34}$$

where $\mathbf{G}_{\text{td}}^l$ is the TD channel coefficient matrix of the l^{th} path, which is a matrix of dimension $MN \times MN$ that has l diagonals such that

$$\mathbf{G}_{\text{td}}^l[m,n] = \begin{cases} h_l' e^{j2\pi\nu_l n\frac{T}{M}}, \text{ if } m = \ell_l + i, n = i, \ i >= 0 \\ 0, \text{ otherwise} \end{cases}. \tag{6.35}$$

Figure 6.5 gives an example of the TD channel matrix. The diagonal property of the ma-

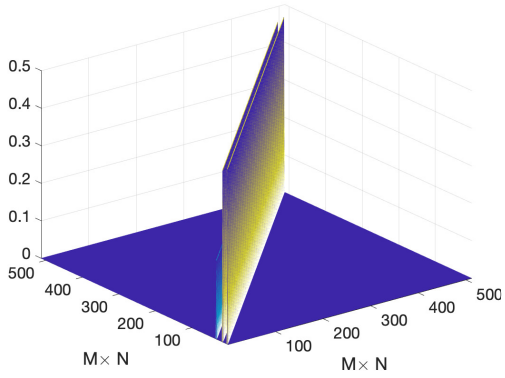


FIGURE 6.5: The delay-time channel matrix representation with $L = 3$ paths and delay indices $\ell = \{2, 15, 32\}$.

trix allows a low-complexity parameter estimation and interference cancellation scheme. For example, assume $L = 2$ with $\{\ell_0 = 1, \ell_1 = 3\}$, we will have

$$\mathbf{Y}_{\mathrm{dt}}[1,0] = h_0' e^{j2\pi\nu_0 \frac{T}{M}} \mathbf{X}_{\mathrm{dt}}[0,0], \tag{6.36}$$

$$\mathbf{Y}_{\mathrm{dt}}[2,0] = h_0' e^{j2\pi\nu_0 2\frac{T}{M}} \mathbf{X}_{\mathrm{dt}}[1,0], \tag{6.37}$$

$$\mathbf{Y}_{\mathrm{dt}}[3,0] = h_0' e^{j2\pi\nu_0 3\frac{T}{M}} \mathbf{X}_{\mathrm{dt}}[2,0] + h_1' e^{j2\pi\nu_1 3\frac{T}{M}} \mathbf{X}_{\mathrm{dt}}[0,0], \tag{6.38}$$

$$\mathbf{Y}_{\mathrm{dt}}[4,0] = h_0' e^{j2\pi\nu_0 4\frac{T}{M}} \mathbf{X}_{\mathrm{dt}}[3,0] + h_1' e^{j2\pi\nu_1 4\frac{T}{M}} \mathbf{X}_{\mathrm{dt}}[1,0]. \tag{6.39}$$

Based on the fact that the OTFS-ISAC transceiver knows all the transmitted data, these expressions indicate that the channel parameters can be estimated without ISI in the time domain in a symbol-by-symbol manner. For instance, τ_0 can be inferred from the indices of $\mathbf{Y}_{\mathrm{td}}$ in Equation (6.36), and ν_0 can be calculated from Equations (6.36) and (6.37). Then, h_0' can be calculated, from which h_0 can be calculated based on the acquired τ_0, ν_0, ϕ_0, and ψ_0. Then, the ISI term caused by the 0^{th} path can be subtracted from Equations (6.38) and (6.39), allowing the ISI-free estimation for the 1^{th} path-related parameters, which follows a similar procedure.

6.4 Communication design

The robustness of OTFS modulation against large Doppler shifts has been demonstrated by many existing studies. For instance, Using the channel estimator and equalizer proposed in [117], the BER against SNR for OTFS and OFDM systems with 16QAM

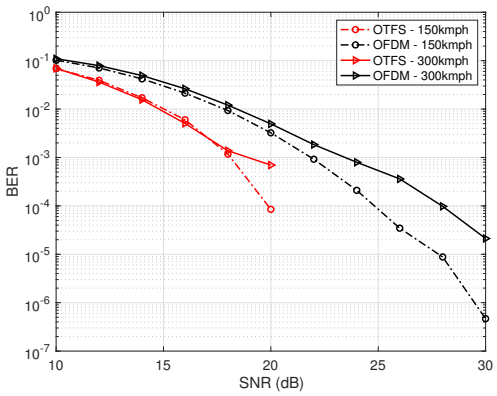


FIGURE 6.6: BER comparison of OFDM and OTFS in high-mobility scenarios.

symbols are plotted in Figure 6.6. We consider moving speeds of 150 km/h for vehicle-to-vehicle (V2V) communication and 300 km/h for high-speed trains. The carrier frequency is 2.5 GHz, and the RMS delay spread is 300 ns. The low-complexity channel estimation and data detection methods have also been extensively studied in the literature, which could be employed to decode the OTFS signal for communications. However, integrating sensing and communication could simplify data decoding by benefiting from radar sensing. In addition, an OTFS-compatible SIC scheme is required to integrate IBFD communications, which is lacking in the literature. In this section, we will develop a feasible SIC scheme and communication protocol for efficient OTFS-ISAC systems.

6.4.1 Self-interference cancellation

Existing SIC studies mainly consider OFDM systems for stationary SI channels, so they may not support the OTFS systems or high-mobility scenarios. Note that although the antennas co-located at the ISAC terminal maintain a relative motion, which suggests the SI channel is still stationary with fast-moving speeds of the terminal, this is only for the direct path. The reflection paths will vary fast due to the moving targets in the surrounding environment, resulting in a rapidly-changing SI channel, which is doubly-selective in the time-frequency domain. Thus, it is necessary to study a feasible SIC scheme compatible with OTFS waveforms and rapidly-changing SI channels to support OTFS-ISAC, providing reliable and efficient uplink channels in high-mobility scenarios. This could also enable many other emerging applications in the future network, such as the moving integrated access and backhaul introduced in the recent 3GPP release [118]. In this section, we study based on the conventional 3-step SIC scheme and analyze the

effects of OTFS modulation and large Doppler shifts on the designs and performance of each step, proposing solutions accordingly.

6.4.1.1 Propagation domain

Propagation domain SIC usually focuses on mitigating the direct path component, which is stationary as analyzed above, and does not rely on the waveforms, so existing methods (e.g., antenna separation, cross-polarization, coupling network, etc.) can be directly employed to the OTFS-ISAC systems. We simply assume the direct path of the SI channel is attenuated by ϑ_{ant} dB to reflect the effects of propagation domain SIC, while we do not restrict technologies to achieve it nor the realized suppression depth. The effective wireless SI channel with its effects can be written as

$$\begin{aligned}\widetilde{\mathbf{H}}_{\text{si}}(\tau,\nu) = 10^{\frac{\vartheta_{\text{ant}}}{10}} \cdot h_0 \mathbf{a}_r(\theta_0)\mathbf{a}_t^{\dagger}(\theta_0)\delta(\tau-0)\delta(\nu-0) \\ + \sum_{l=1}^{L-1} h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\phi_l)\delta(\tau-\tau_l)\delta(\nu-\nu_l),\end{aligned} \tag{6.40}$$

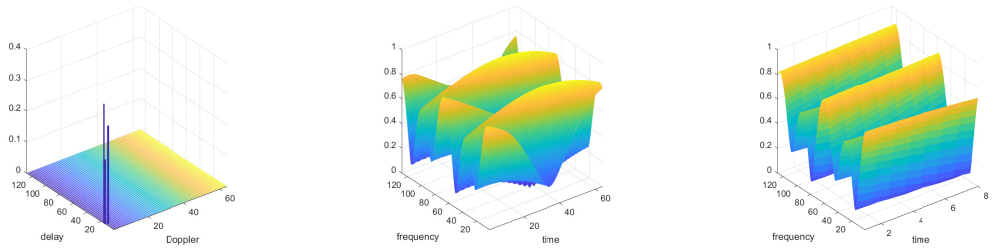
where the trivial propagation delay of the direct path is ignored.

6.4.1.2 Auxiliary RF chain assisted RF cancellation

As shown in Figure 2.7 in Section 2.4, there are two architectures of the RF canceller design: 1) *Stanford* architecture, which builds a finite impulse filter in the RF domain (i.e., a multi-tap canceller) to mimic the wireless SI channel; 2) *Rice* architecture, which regenerates the baseband received SI in the digital domain and converts it to the RF domain via auxiliary RF chains. In this section, we will discuss the two architectures for OTFS-ISAC systems respectively.

The most attractive factor of the *Rice* architecture is that it shares the high flexibility of digital processing. Thus, it can be easily programmed to support the OTFS waveforms, where the relatively constant DD channel could be utilized to combat the large Doppler shifts. According to Equation 6.14, a baseband replica of the received SI can be generated as

$$\hat{\mathbf{y}}_{\text{aux}}(k) = \sum_{l=0}^{L-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \hat{h}_l' \mathbf{X}_{\text{tf}}[m,n] e^{j2\pi\hat{\nu}_l kT} e^{j2\pi m\Delta f((k-n)T-\hat{\tau}_l)}, \tag{6.41}$$



(a) DD domain (b) TF domain (64 time slots) (c) TF domain (8 time slots)

FIGURE 6.7: The stationary SI coupling channel (i.e., with small Doppler shifts) in the DD and TF domains .

where $\hat{h}'_l$, $\hat{\tau}_l$, and $\hat{\nu}_l$ can be obtained by radar sensing, and $\mathbf{X}_{\text{tf}}$ are known by the ISAC terminal. Then, this baseband signal is converted into the RF domain by the auxiliary transmitter chains, which have identical configurations to the ISAC transmitter chains, introducing distortions and noise as

$$\hat{\mathbf{y}}_{\text{aux}}(t) = \sum_{l=0}^{L-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \hat{h}'_l \mathbf{X}_{\text{tf}}[m,n] e^{j2\pi\hat{\nu}_l t} e^{j2\pi m\Delta f(t-\hat{\tau}_l-nT)} + \mathbf{d}_{\text{aux}}(t) + \mathbf{n}_{\text{aux}}(t), \quad (6.42)$$

where $\mathbf{d}_{\text{aux}}(t)$ and $\mathbf{n}_{\text{aux}}(t)$ denote the distortions and AWGN caused by the auxiliary chains, respectively. It should be noted that Equation 6.14 only demonstrates the OTFS modulation processing, so it ignores the transceiver imperfections, while the real received SI should be written as

$$\mathbf{y}_{\text{si}}(t) = \sum_{l=0}^{L-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} h'_l \mathbf{X}_{\text{tf}}[m,n] e^{j2\pi\nu_l t} e^{j2\pi m\Delta f(t-\tau_l-nT)} + \mathbf{d}_{\text{isac}}(t) + \mathbf{n}_{\text{isac}}(t), \quad (6.43)$$

where $\mathbf{d}_{\text{isac}}(t)$ and $\mathbf{n}_{\text{isac}}(t)$ denote the distortions and AWGN by the ISAC transmitter, respectively. Subtract the generated $\hat{\mathbf{y}}_{\text{aux}}(t)$ from the received SI for cancellation, there will leave RSI due to estimation errors and transmitter distortions and noise from all RF chains. Thus, the performance will be limited by the transceiver conditions.

6.4.1.3 Multi-tap canceller

In a fairly stationary environment, which means there are no fast-moving objects around, the SI channel could remain unchanged for a few time slots, and it changes with a relatively long time interval, as Figure 6.7 shows. Under this condition, the conventional multi-tap canceller can be directly utilized for RF cancellation with OTFS waveforms

since the cancellation rationale does not change. The coherence time of the SI channel is fairly long so that the canceller can be tuned based on the pilots within the duration of the first symbols, and it works for the rest duration of the same coherence time period. However, the RF waveform is OTFS modulated here, so the tuning method will be different from existing studies.

A canceller is inserted between each Tx/Rx antenna pair. The feedback loops tap the received SI from the ISAC receiver input, and the ADCs sample it at $t = kT$, obtaining received samples as

$$\begin{aligned}\mathbf{y}_{\mathrm{si}}(k) = 10^{\frac{\vartheta_{\mathrm{ant}}}{10}} \cdot h_0 \mathbf{a}_r(\theta_0)\mathbf{a}_t^{\dagger}(\theta_0) \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \mathbf{X}_{\mathrm{tf}}[m,n] e^{j2\pi m\Delta f(k-n)T} \\ + \sum_{l=1}^{L-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\theta_l)\mathbf{X}_{\mathrm{tf}}[m,n] e^{j2\pi\nu_l kT} e^{j2\pi m\Delta f((k-n)T-\tau_l)}.\end{aligned} \tag{6.44}$$

There are MN samples collected by each receiver chain within a frame, and we use the first M samples to estimate the channel frequency response of the corresponding SI coupling to tune the weights and phases of the taps, which can be implemented as follows. We first convert the M time samples into the frequency domain as

$$\mathbf{r}_{\mathrm{si},i} = \mathbf{F}_M \left(\mathbf{y}_{\mathrm{si}}^i(1:M)\right)^T, \tag{6.45}$$

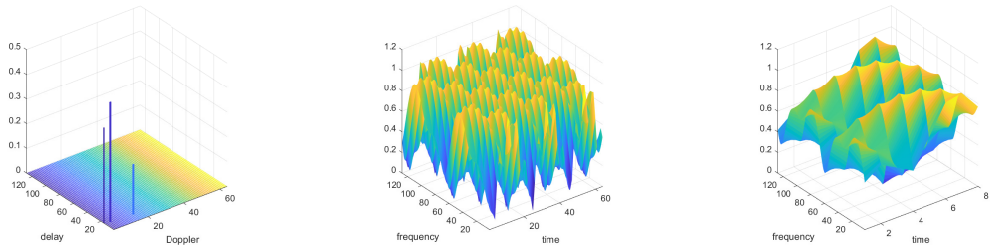
where $\mathbf{y}_{\mathrm{si}}^i$ denotes the i^{th} row of $\mathbf{y}_{\mathrm{si}}(k)$. $\Upsilon_{\mathrm{si},i}$ is actually the first column of the received signal in the time-frequency grid, so its associated transmitted signal is $\mathbf{X}_{\mathrm{tf}}[:,1]$. Then, we can form a diagonal matrix with the corresponding transmitted symbols on its diagonal as

$$\mathbf{T}_{\mathrm{si},i} = \mathcal{D}\left(\mathbf{X}_{\mathrm{tf}}[:,1]\right), \tag{6.46}$$

Thus, the CFR of the SI coupling channel can be obtained through the MMSE estimator as

$$\mathbf{G}_{\mathrm{si},ij} = \left(\mathbf{T}_{\mathrm{si},j}^{\dagger}\mathbf{T}_{\mathrm{si},j} + \sigma^2\mathbf{I}_M\right)^{-1}\mathbf{T}_{\mathrm{si},j}^{\dagger}\mathbf{r}_{\mathrm{si},i}. \tag{6.47}$$

Then, the multi-tap canceller can be tuned by minimizing the difference between the canceller's CFR and $\mathbf{G}_{\mathrm{si},ij}$ within the operation band, which has been detailed in Sections 2.4.2 and 3.2.2 for the two architectures. The tuning process is implemented for each frame, which means the tuned canceller remains unchanged for the remaining $(N-1)T$ duration of the frame. Thus, the cancellation performance may degrade with time if



(a) DD domain (b) TF domain (64 time slots) (c) TF domain (8 time slots)

FIGURE 6.8: The rapidly-changing SI coupling channel (i.e., with large Doppler shifts) in the DD and TF domains .

the SI coupling channel varies during the remaining frame duration, which happens in high-mobility scenarios. To deal with the rapidly-changing channel, one solution is to increase the tuning frequency. However, this yields tuning speed concerns and spectral efficiency reduction. For high-speed V2V and train communications, the SI coupling channel could change at each discrete time period, as Figure 6.8 shows, which means the canceller needs to be tuned accordingly with an extremely short interval, e.g., T. Otherwise, the cancellation performance will be severely compromised due to the varied SI channel. However, the response speed of practical electronic components and limited computation capacity may not support such a high tuning frequency. In addition, it may make half of the time-frequency grid unavailable, which reduces the spectral efficiency by half.

Thus, a new canceller architecture, which can adaptively adapt to the rapidly-changing coupling channel and works with a relatively long tuning interval, is required. Inspired by the relatively constant DD channel, it is expected to construct the canceller based on the delay-Doppler channel representation, which writes an input-output relationship between the s^{th} Tx antenna and the r^{th} Rx antenna in the time domain as

$$\begin{aligned} y_{r,s}(t) &= \int_0^{\tau_{\max}} \int_\nu \left(\sum_{l=1}^{L} h''_{l,rs} \delta(\tau - \tau_l)\delta(\nu - \nu_l) \right) e^{j2\pi\nu(t-\tau)} x_s(t-\tau) d\nu d\tau \\ &= \sum_{l=1}^{L} \left(\int_\nu \int_0^{\tau_{\max}} h''_{l,rs} \delta(\tau - \tau_l)\delta(\nu - \nu_l) e^{j2\pi\nu(t-\tau)} x_s(t-\tau) d\nu d\tau \right) \\ &= \sum_{l=1}^{L} \left(\int_\nu h''_{l,rs} \delta(\nu - \nu_l) e^{j2\pi\nu(t-\tau_l)} x_s(t-\tau_l) d\nu \right) \\ &= \sum_{l=1}^{L} \underbrace{h''_{l,rs} e^{-j2\pi\nu_l\tau_l}}_{\alpha_{l,rs} e^{j2\pi\phi_l}} e^{j2\pi\nu_l t} x_s(t-\tau_l), \end{aligned} \tag{6.48}$$

where $h''_{l,rs} = \left[h_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\dagger}(\theta_l)\mathbf{V}\right]_{r,s}$. The expression suggests that the time-varying effect is caused by the Doppler shifts, which introduce time-dependent phase shifting $e^{j2\pi\nu_l t}$. Thus, a finite impulse response (FIR) filter with time-dependent phase shifters can be leveraged to characterize the time-varying coupling channel. This architecture can be implemented based on the typical time-domain multi-tap canceller, but an additional tunable time-dependent phase shifter is added on each tap to include the time-varying effect caused by the Doppler shifts, i.e., $e^{j2\pi\nu_l t}$, as Figure 6.9 shows.

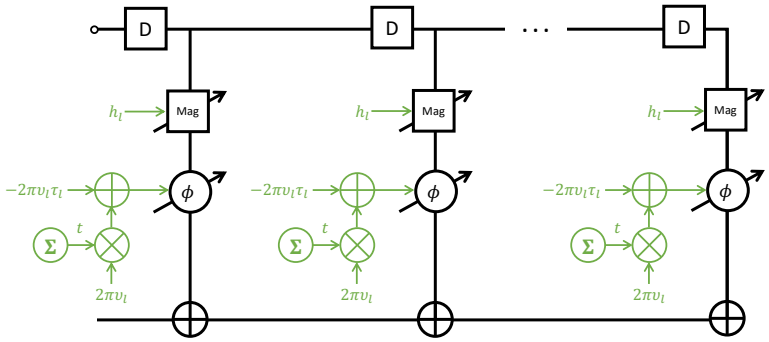


FIGURE 6.9: The architecture of a variant of multi-tap cancellers compatible with large Doppler shifts.

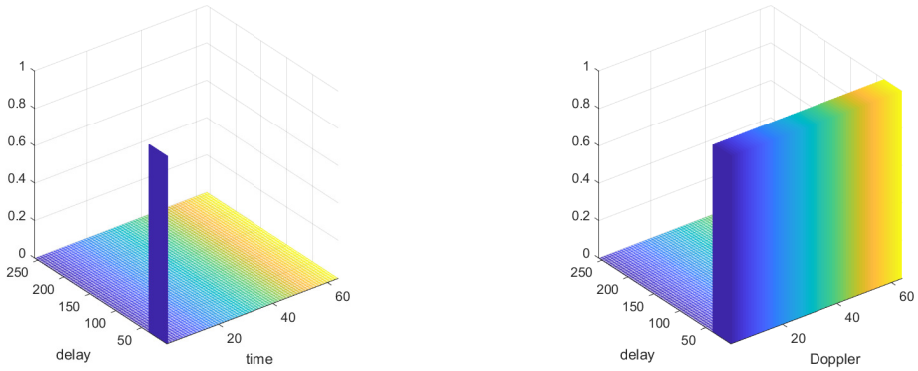
Remarking: The canceller can be easily tuned with the help of radar sensing, where the tuning parameters $h_{l,rs}$, τ_l and ν_l have been obtained. In addition, these parameters can be obtained within the first T duration of the frame. However, the architecture is proposed based on mathematical derivation, while its effectiveness needs to be verified by hardware simulations and experiments in the future.

6.4.1.4 Digital cancellation

Digital SIC has high flexibility, so it can easily regenerate the RSI in any domain (e.g., time-delay, time-frequency, delay-Doppler domains) and subtract it. The non-linear digital canceller introduced earlier in Section 2.5 can also be employed for the OTFS systems since it is a time sample-wise operation regardless of the modulation method.

6.4.2 Pilot-based transmission scheme

With the existing channel estimator, data detector, and our proposed SIC scheme, it is possible to enable efficient IBFD communications in the OTFS-ISAC systems. Then,



(a) delay-time domain (b) delay-Doppler domain

FIGURE 6.10: The proposed pilot pattern in the DT and DD domains .

we study the waveform design, i.e., transmission scheme, by crossly considering the two functions of ISAC. The low-complexity channel parameter estimation method introduced in Section 6.3.2.2 could be leveraged for channel estimation of the communication function. To this end, the pilot symbols are mapped in the delay-time domain, occupying the first $\ell_{\max}$ resource blocks of the 0^{th} time slot. However, it should be noted that these symbols in the delay-time domain will spread over the whole Doppler axis in the delay-Doppler domain, as Figure 6.10 shows. Therefore, the first $\ell_{\max}$ rows need to be reserved for the delay-time pilot symbols for this low-complexity interference-free estimator[3]. The delay-time domain pilot pattern has another advantage is that it will not spread the pilot symbols over time, so the estimation process can start as long as the first M time samples (of the 0^{th} time slot) have been collected.

In contrast, the conventional delay-Doppler domain-based pilot pattern requires a total of $4\ell_{\max}\kappa_{\max}$ resource blocks are required for a single pilot plus surrounding guard intervals [119]. In addition, it spreads the pilot symbols over time, so the estimation process starts after the whole frame (i.e., MN time samples) is collected, increasing the latency.

Remarking: In addition to the latency reduction, the proposed transmission and estimation scheme has advantages in terms of complexity and could have a higher spectral efficiency if $N < 4\kappa_{\max}$, which happens if there are distant objectives.

[3]This pilot requirement is for communication function only since the receiver does not have the full knowledge of the transmitted signal, while it is not necessary for the radar sensing since the radar receiver has the full knowledge

6.4.3 Bypassing channel estimation transmission scheme

The OTFS-ISAC node can determine the location and speed of all other nodes (both OTFS transmitters and receivers) within its range according to the delay, Doppler, and angle information acquired from radar sensing. Based on the estimated location and speed information, the other nodes can be tracked, i.e., their motion in the following time slots can be predicted. Then, the channel effects (e.g., delay, Doppler shifts, and pathloss) can be predicted, allowing the OTFS-ISAC receiver to bypass channel estimation while detecting the uplink data. Thus, the OTFS transmitter does not need to insert pilots and can use all resource blocks for data transmission, which improves spectral efficiency. In addition, the delay-Doppler domain-based pilot pattern spreads the pilot over time and frequency. Thus, bypassing channel estimation with the help of radar sensing can eliminate the pilot overhead, significantly reducing the latency. Downlink transmission can also benefit from radar sensing by compensating for the channel effects at the OTFS-ISAC transmitter based on the constructed topology of the surrounding environment. The readers are referred to [105] for details of constructing the dynamic topology and compensating methods. By doing this, the OTFS receiver will receive the signal without channel effects, so it can directly perform OTFS demodulation and data detection to recover the transmitted data.

6.5 Simulation results

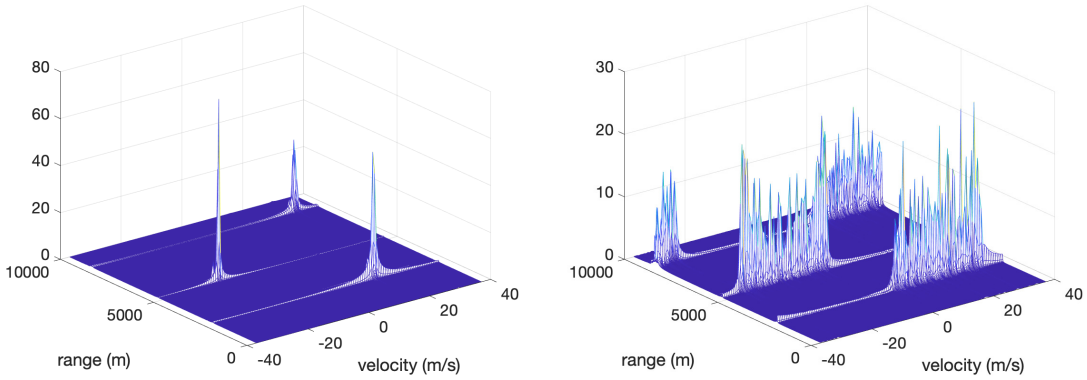
Our simulations follow 3GPP specifications, where the simulation parameters are given in Table 6.1. We run Mont-Carlo simulations for 500 frames. The Doppler shift at the l^{th} path is generated as $\iota_l = \iota_{\max}\cos(\psi_l)$ with $\psi_l \sim \mathcal{U}(0, \pi)$. The channel tap positions are assumed to be known at the receiver, and the normalized Doppler shift for each path is assumed to be an integer.

6.5.1 Radar sensing

We first illustrate the flaws of the low-complexity time-frequency domain range and velocity detection method introduced in Section 6.3.2.1. Figure 6.11 shows the normalized detection image obtained by this method, where the received signal is generated based

TABLE 6.1: Simulation parameters for OTFS-ISAC

Parameters	**Values**
Number of subcarriers	$M = 1024$
Number of time slots	$N = 8$
OFDM cyclic prefix	64
Subcarrier spacing	$\Delta f = 15$ kHz
Carrier frequency	$f_c = 4$ GHz
Transmit power (OTFS-ISAC)	23 dBm
Transmit power (OTFS transmitter)	0 dBm
Modulation	QPSK
Channel	EVA model in [120]
Coding	Turbo code (1/3)
Number of transmitting antennas	2
Number of receiving antenna	16



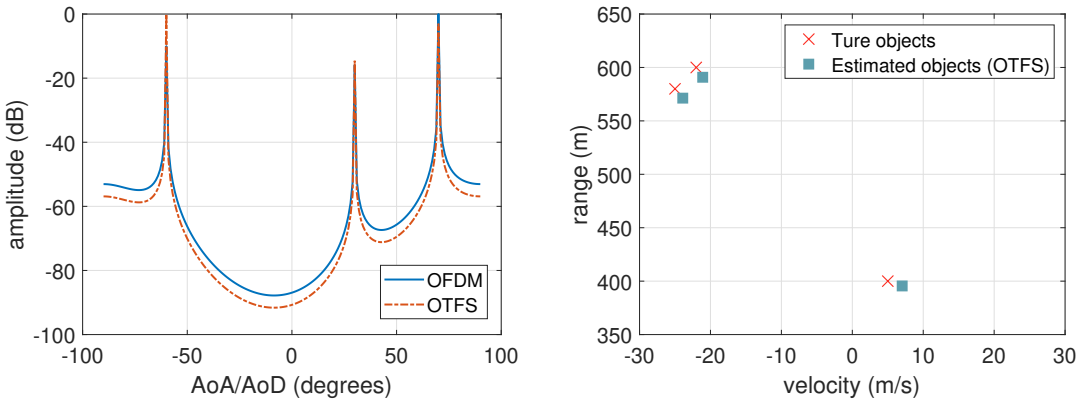
(a) based on Equation (6.28)

(b) based on Equation (6.27)

FIGURE 6.11: Normalized detection image obtained through the TF domain method for three objects with ranges $r \in \{2300, 5230, 8760\}$ m and velocities $v \in \{16, -17, 30\}$ m/s. .

on Equations (6.27) and (6.28), respectively. It can be seen that this method only works under the assumption that the ICI can be ignored. However, with the real received waveforms with ICI, it cannot clearly identify the target velocity.

Figure 6.12 shows the radar sensing results of the MUSCI algorithm and the proposed delay-time domain-based estimation methods. The results verified the effectiveness of the MUSIC algorithm for OTFS waveforms, that it can achieve identical performance as in its OFDM counterpart. Our proposed range and velocity estimation method can accurately detect all objects within the range. The minor error comes from the limited detection resolution due to the limited size of the resource grid. In contrast, accurate velocity estimation for fast-moving objects is challenging for OFDM waveforms.



(a) MUSIC spectrum (b) Range and velocity estimation

FIGURE 6.12: Radar sensing performance evaluation.

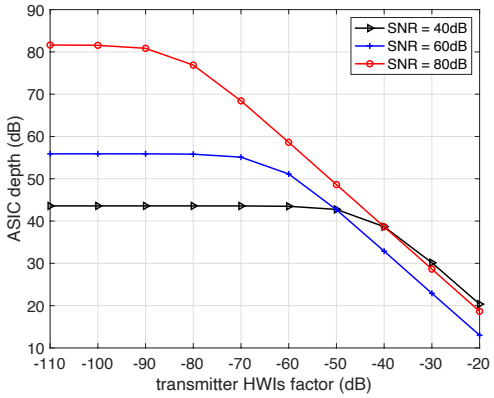


FIGURE 6.13: Achievable ASIC depth of the auxiliary chain-based canceller with various transmitter distortion factors and SNR for channel estimation.

6.5.2 Communication performance

We first demonstrate the SIC performance to guarantee effective IBFD communications. Figure 6.13 shows the achievable ASIC depth against the dynamic range of the transmitter and auxiliary chain with different SNRs of the received SI. The results indicate that the performance of the auxiliary chain-based canceler is mainly subjected to the accuracy of SI channel estimation and the dynamic range (i.e., EVM) of the transmitter and the auxiliary chain. The SI channel can usually be estimated with high SNR due to the proximity of the transmitter and receiver. However, high SNR results in significant performance degradation with poor hardware conditions (i.e., large HWI factors). The reason is that high SNR indicates a high transmit power, which results in significant transceiver HWIs to the digitalized signal. In addition, the achievable ASIC depth reduces severely with decreased quality of transmitter and auxiliary chain hardware.

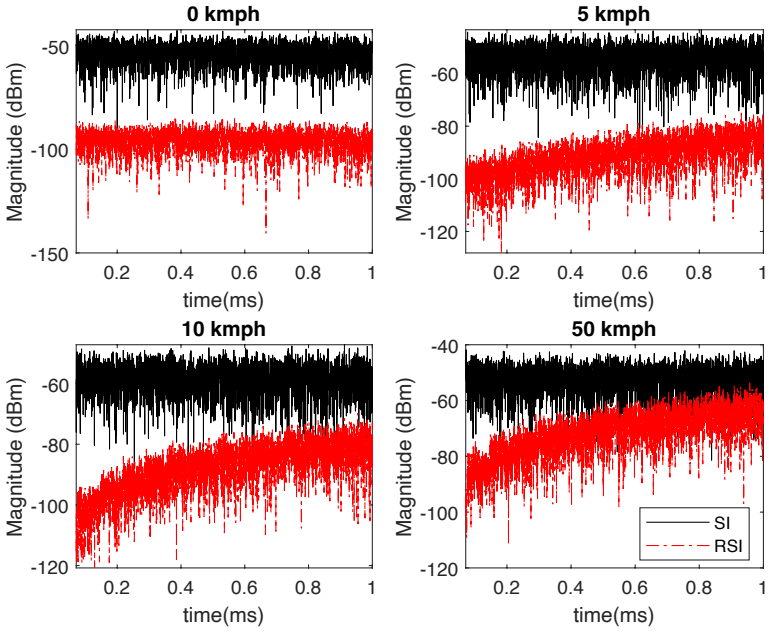


FIGURE 6.14: Received SI samples in the time domain versus the RSI after multi-tap RF cancellers with different speeds (0, 5, 10, 50 kmph).

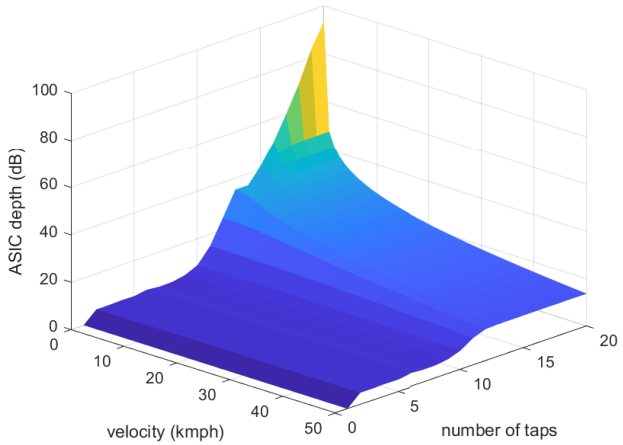
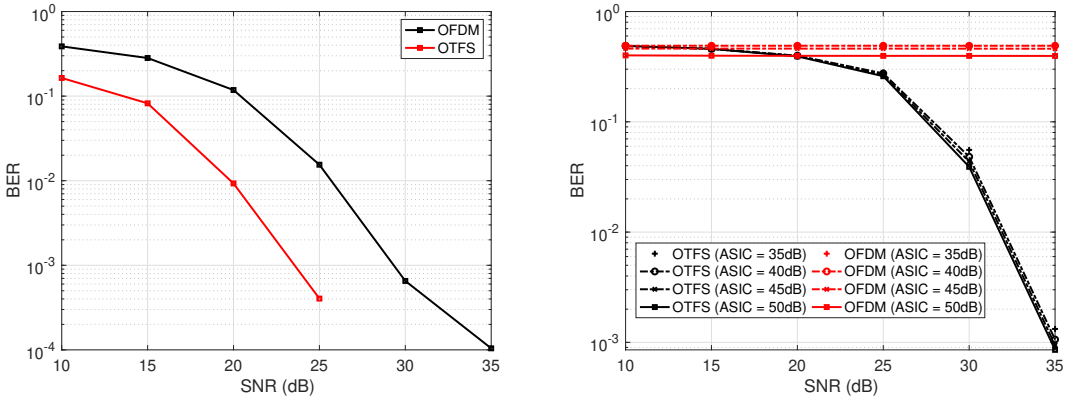


FIGURE 6.15: Achievable ASIC depth of the typical multi-tap canceller versus the number of taps and velocity.

Figure 6.14 shows the magnitude of received SI and the RSI samples after a 15-tap canceller. It can be seen that with high Doppler shift (i.e., fast velocity), the magnitude of RSI increases with time due to the varying SI channel, compromising the effectiveness of the tuned canceller. Figure 6.15 shows the achievable ASIC depth with a different number of taps of cancellers and Doppler shifts. It suggests that increasing the number of taps can generally improve the ASIC capacity of cancellers. However, with a high Doppler shift, the achievable ASIC depth will be limited, and more taps may not further



(a) downlink (b) uplink

FIGURE 6.16: Communication performance evaluation of the OTFS-ISAC system in terms of BER .

improve the ASIC depth. For instance, with a velocity of 50 kmph, the ASIC is not further increased with more than 12 taps. In contrast, $>$ 80 dB of ASIC can be achieved with a 20-tap canceller with a stationary environment (i.e., with a velocity of 0 kmph).

With the proposed SIC scheme applied, Figure 6.16 shows the communication performance of the proposed framework compared to the conventional OFDM-based design. It can be seen that OTFS is superior to OFDM in high-mobility scenarios. For downlink transmission, enhancing the transmit power can compensate for the BER performance loss of OFDM systems. However, this is not the case for uplink transmission due to the presence of SI. As revealed above, existing SIC schemes usually ignore the ICI, invalidating the cancellers with fast-moving objects. In contrast, our proposed radar sensing-based SIC scheme effectively suppresses SI with large Doppler shifts, providing reliable uplink transmission.

6.6 Conclusions

In this chapter, we studied a novel framework for ISAC networks. To enable emerging applications in future networks, we introduced OTFS modulation to support high-mobility scenarios (e.g., intelligent automotive systems, high-speed trains, etc.) and integrated IBFD communication to enhance spectral efficiency. The performance of conventional OFDM waveforms will be severely compromised in terms of both radar sensing and communication by the ICI due to the lack of guard interval in the frequency domain,

while OTFS waveforms could be free of ICI effects with appropriate designs. In addition to the robustness of Doppler shifts, OTFS-ISAC allows the two functions to cooperate more deeply than their OFDM counterpart. For instance, the channel effects can be compensated before communication signal transmission based on the radar-sensing parameters, and the new transmission scheme allows intended users to bypass channel estimation. In addition, the SIC canceller can be tuned based on the parameters obtained from radar sensing. However, to deal with the rapidly-changing SI coupling channel, a new multi-tap canceller is required since the conventional one faces the tuning speed problem. By exploiting the relatively constant DD channel, a variant of the multi-tap canceller, which adds time-dependent phase shifters, was proposed. We also proposed a low-complexity radar sensing scheme, where the angle information is obtained via the conventional MUSIC algorithm, and the DD channel parameters are acquired via a delay-time domain-based interference cancellation and estimation scheme. The proposed method has advantages in terms of pilot overhead and complexity over existing methods and can achieve higher spectral efficiency in the presence of distant targets. Numerical results demonstrate that the proposed ISAC waveform and associated estimation algorithm can provide both reliable communications and accurate radar sensing with reduced latency, improved spectral efficiency, etc.

Chapter 7

Conclusions

7.1 Summary of contributions

The demand for the wireless network's capacity is increasing during the evolution, motivating new techniques to enhance resource utilization efficiency. IBFD is considered as a promising candidate in future networks due to its potential to double spectral efficiency, reduce latency, enhance emerging applications, etc. In this paper, we study the feasibility of IBFD by investigating effective SIC schemes for wideband operations, massive MIMO, and cellular networks. Then, we explore the advantages of IBFD in cellular networks and emerging applications over its HD counterpart. Our contributions can be summarized as follows.

- *Novel SIC schemes:* Starting with a single antenna transceiver, we studied the SIC techniques in Chapter 2. We reviewed existing designs to provide Tx-Rx isolation for both shared and separate antenna interfaces. These methods usually provide only direct path mitigation, while they cannot handle reflection components. We innovatively built upon existing SIC techniques, addressing the limitations of direct path mitigation by exploring RF domain cancellation with a unique focus on a time domain-based multi-tap canceller. Our distinctive approach, combining frequency domain tuning with fixed delays and nonlinear digital cancellation, challenges conventional RF canceller designs. Notably, we proposed an optimal ASIC depth, dictated by the transceiver quality, and provided thorough validation through both 3GPP-specified simulations and real-world testing in the later

Chapter 3. Additionally, we formulated the final RSI power (i.e., after the digital canceller) and illustrated that there exists an optimal ASIC depth that can minimize the SI effects at IBFD transceivers, which strongly depends on the transceiver quality.

- *Pioneering PAC design:* To break the limitation on the canceller's flexibility of electrical delay lines, we studied PAC designs in Chapter 3. We gave the general model of the PAC, where four branches are used to implement the tunable phase since it is challenging to tune the phase of RF signals through the modulated optical signal. We also proposed an alternative low-complexity tuning algorithm for this architecture. Then, we designed a fiber-array canceller and verified its effectiveness through both hardware simulations and experiments. Additionally, we discussed its ability to construct more taps based on experimental measurements.

- *Feasible SIC schemes for MIMO systems:* Expanding on our novel SIC approach, Chapter 2 unveils our methodology for MIMO system integration. Utilizing hybrid beamforming architecture, we offer insights into the impacts on the RMS delay spread and the SIC performance. Inspired by the spatial DoF provided by MIMO systems, we exploited beamforming techniques for SIC since it requires only digital processing without additional hardware. We revealed that digital precoding could suppress the SI in the propagation domain before the receiver, fulfilling the function of the complex and expensive RF cancellation. Then, we studied beamforming cancellation in Chapter 4 and proposed the eMMSE beamforming scheme to enhance the ASIC depth while minimizing the sum rate loss, which can achieve considerable IBFD gain in MCMU networks with low implementation complexity.

- *CCI reduction via resource allocation:* Addressing a fundamental challenge with IBFD radios, we introduced a game theoretic user allocation algorithm in Chapter 2. This unique approach provides a universally adaptable solution for diverse communication scenarios, ensuring global optimum allocation. We considered a heterogeneous environment for generality so it can be employed for IIoT scenarios, V2V communications, cellular networks, etc., by changing the system parameters. The algorithm has feasible complexity, and its performance does not rely on the initial point, guaranteeing to achieve the global optimum.

- *Enhanced beamforming schemes for IBFD cellular networks:* Chapter 4 pioneers an enhanced approach to beamforming for future IBFD cellular networks. Our work redefines traditional linear beamforming schemes, offering a groundbreaking JPABF scheme that synergizes power allocation and beamforming. The algorithm we propose significantly boosts efficiency by minimizing computational overhead. We evaluated their performance in terms of complexity and spectral efficiency under various 3GPP-specified scenarios to provide practical insights.

- *IBFD-Powered Security Boost:* Chapter 5 delves into our groundbreaking research on the advantages of IBFD radios for PLS. We focused on a CFR-based SKG scheme to investigate the advantages of IBFD radios on the PLS. By characterizing the practical imperfections via measurable metrics, we formulated the probing errors and RSI strengths after a 3-step SIC. Then, we derived the SKC accordingly in the presence of a passive eavesdropper. We compared the SKC of HD and IBFD probing and revealed that hardware impairments, SIC overheads, and ASIC depth are the three limiting factors of the CFR-based SKG with IBFD probing. Then, we ran 3GPP-specified simulations to verify the theoretical analysis and verified the effectiveness of the generated key in terms of randomness via the NIST test suite.

- *A novel OTFS-ISAC framework:* Envisioning the demands of future networks, Chapter 6 introduces our avant-garde OTFS-ISAC framework. By seamlessly integrating OTFS modulation, we tackle the challenges of ICI in high-mobility scenarios. We illustrated the performance degradation of both radar sensing and communication with OFDM waveforms in high-mobility scenarios. Then, we developed a low-complexity radar sensing scheme for OTFS waveforms, where the angle information is obtained via the MUSIC algorithm and the DD channel parameters, which include the range and velocity information of targets, are obtained via a low-complexity delay-time domain-based interference cancellation and estimation scheme. Then, we studied the IBFD communication design and proposed two RF cancellation techniques that are compatible with large Doppler shifts. In addition, we also discussed digital cancellation for OTFS systems. Simulation results showed that the proposed framework could provide accurate radar sensing and reliable IBFD communication.

7.2 Further work

7.2.1 Self-interference cancellation

- The experimental demo in Chapter 3 only carries out a 2-tap RF canceller with low transmit power for the feasibility demonstration purpose. Its scalability is only speculated based on the measurements of the two taps. In addition, the canceller itself also has the defects of phase instability and slow tuning speed. Therefore, future work needs to solve the defects of the canceller and try to build more taps to verify its scalability, which is desired to be implemented on a photonic chip using the integration processes from the CMOS industry.

- Reservoir computing has recently been developed for processing time-dependent information, where the reservoir is capable of universal computation due to its unique structure. Furthermore, photonics reservoir computing, where a photonic chip is trained for certain tasks, attracts lots of research interest due to its hardware advantages [121]. It is experimentally demonstrated in [122] that a trained photonic chip could effectively compensate for the fiber nonlinearity and outperform the electronic tapped delay line filter. Both the structure and task of the photonics reservoir in [122] are highly similar to the PAC for SIC introduced in Chapter 3, which motivates us to further develop the canceller based on the idea of reservoir computing to further improve the performance and efficiency.

- With the recent technique explosion of artificial intelligence in terms of computing power and algorithms (i.e., models), it is possible to characterize a complex relationship via neural networks on real-world processors with low time complexity. Thus, it inspires us to train a single neural network for both analog and digital SIC and develop a variant of the auxiliary chain-based canceller. Let $\mathcal{F}_{\mathrm{rx}}(\cdot)$, $\mathcal{F}_{\mathrm{aux}}(\cdot)$, and $\mathcal{F}_{\mathrm{nn}}(\cdot)$ describe the input-output relationship of the receiver chain, auxiliary transmitter chain, and neural network, respectively, the canceller can perform both analog and digital SIC by training the neural network to reduce the loss as

$$\mathcal{L}_{\mathrm{nn}} = \mathcal{F}_{\mathrm{rx}}\left(g_{\mathrm{si}}(t) - \mathcal{F}_{\mathrm{aux}}\left(\mathcal{F}_{\mathrm{nn}}\left(s(n)\right)\right)\right), \tag{7.1}$$

where $s(n)$ is the baseband transmit symbol that is used as the input of the neural network. The feasibility and effectiveness of this idea need to be verified by developing an appropriate network architecture in future work.

7.2.2 Beamforming studies

- As future networks move to higher frequency bands, massive MIMO will be widely employed by the user side instead of base stations only. Therefore, the hybrid beamforming scheme introduced in Section 2.6 should be considered while designing the beamforming scheme for cellular networks. Future work can decompose the proposed all-digital linear beamformers into hybrid digital-RF beamformers for performance evaluation. In addition, the RF beamformer introduces both new concerns and techniques, such as the codebook design due to the limited resolution and beam management. Thus, future work can study the optimal design with the new formulation (i.e., formulate the transmitted and received signals with RF and digital beamforming matrices) of the convex optimization problem.

- The proposed algorithm in Chapter 4 uses a single scalar to reflect the allocated transmit power, so all transmitting antennas are equally scaled to satisfy the transmit power constraint. To enable a finer power allocation, a diagonal matrix could be leveraged to reflect the power allocation policies, yielding the transmitted signal of user i as

$$\mathbf{x}_i = \mathbf{\Lambda}_i \mathbf{V}_i \mathbf{s}_i + \mathbf{d}_{tx,i}, \tag{7.2}$$

  where $\mathbf{\Lambda}_i = \mathcal{D}\left(\alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,N_t}\right)$ consists of the power coefficients for the N_t transmitting antennas at user i on its diagonal. Then, the representations of the received signal, MSE, covariance matrices, problem formulation in Equation (4.66), etc., can be modified accordingly, and obtain a finer JPABF scheme through solving this problem to further improve the performance.

7.2.3 Physical layer security

- In IBFD radios, the PHY-SKG scheme benefits from simultaneous probing and more time-frequency resources, legitimate users can acquire estimates that are closer to the reciprocal wireless channel than its HD counterpart but suffers from

more severe HWIs due to SI effects, as illustrated in Chapter 5. Let $\mathbf{H}$ denote the reciprocal wireless legitimate channel, ρ denote the correlation coefficient between the estimated and actual channels, $\mathbf{D}_A$ and $\mathbf{D}_B$ denote the hardware HWIs at the two legitimate users, respectively. Then, the inconsistency between the legitimate users' measurements can be written as

$$\mathbf{\Delta} = 2(1-\rho)\mathbf{H} + \mathbf{D}_A + \mathbf{D}_B. \tag{7.3}$$

IBFD radios yield a higher ρ but also larger $\mathbf{D}_A$ and $\mathbf{D}_B$. It is challenging to compensate for the reduced channel correlation via digital processing, limiting the HD SKC. In contrast, many studies have illustrated that machine learning can effectively compensate for hardware HWIs, including reservoir computing [123], encoders and decoders [70], deep neural networks [124], convolutional neural networks [125], etc. Thus, future work could explore effective machine learning approaches to compensate for the increased hardware HWIs in IBFD radios to enhance the upper bound of SKC.

- With the development of ISAC, the eavesdropper might be able to reconstruct the full wireless environments, compromising the security. The SI channel, which is unique in IBFD radios, could be exploited as a randomness source for the key-based PLS designs, e.g., the rotated matrix at legitimate users as in [81], private keys for asymmetric encryption [126], further enhancing the security level. It should be noted that as long as the two legitimate users are far away, which is a reasonable assumption since they do not need wireless communication if they are that close, the eavesdropper cannot acquire both of their SI channels at the same time, guaranteeing the security. Thus, future work can explore the use of SI to enhance wireless security.

- In addition to encryption-based schemes, beamforming can enable a keyless PLS scheme, which is more energy efficient and is suitable for low-cost devices. Thus, it is considered as a promising technique for PLS designs. By steering the transmitted beams in the desired direction, the downlink rate of unauthorized devices is severely degraded to prevent them from obtaining sensitive data. The future network is very likely to adopt IBFD radios and cell-free massive MIMO and intelligent reflecting surface (IRS), so it is very meaningful to study the PLS beamforming design in such

a network for future wireless security, which is lacking in existing studies [127] and could be a good research direction for future work. Further work can consider the joint power allocation and beamforming scheme to enhance the secrecy capacity.

7.2.4 Integrated sensing and communication

- The variant of the multi-tap canceller in Figure 6.9 is theoretically proposed based on the mathematical derivation of the time-domain input-output relationship. The first step of future work is to verify its feasibility and effectiveness through experiments, which could be done through a small number of taps constructed by discrete components first. Similar to the limits of electrical components as stated in Chapter 2, it is motivated to implement such cancellers through photonics and further photonics chips also.

- As stated in Section 6.4, the ISAC node takes advantage of the fact it has full knowledge of its own transmitted signal, it can perform the radar sensing parameters in any domain to inspire a very low-complexity algorithm, such as our proposed time domain-based algorithm. In contrast, the receiver does not have this advantage, and the DD symbols will spread over the time and frequency domains such that the payload will contaminate the pilots in the TF domain. Therefore, DD domain-based estimation seems as the best choice for OTFS receivers, yielding relatively high complexity even though lots of studies have efforted to reduce it [100, 106–109]. Thus, future work could study deeper cooperation of the two functions to enable low-complexity OTFS communications with the help of radar sensing, such as the bypassing channel estimation scheme.

- The beamforming scheme has a significant effect on the communication and backscatter links, affecting the communication and radar sensing performance. However, there is a contradiction in the beamforming designs for ISAC devices. For radar sensing, the transmitted should cover a wide angular sector to scan possible targets. In contrast, a narrow beam towards the intended user is desired to enhance the communication quality. Therefore, it is worth developing an appropriate beamforming design for ISAC, including the optimal beamformer for the two functions and the switch between the two functions.

List of Publications

- H. Luo, A. Bishnu, and T. Ratnarajah, "In-Band Full-Duplex Radios in 6G Networks: Implementation and Applications", *Foundations and Trends® in Networking*, vol. 13, no. 1, pp. 1-105, 2023.

- H. Luo, N. Garg, and T. Ratnarajah, "A Channel Frequency Response-Based Secret Key Generation Scheme in In-band Full-duplex MIMO-OFDM Systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2951-2965, Sep. 2023.

- H. Luo, N. Garg, and T. Ratnarajah, "Beamforming Design for Multi-cell Multi-user In-band Full-duplex MIMO Networks with Global and Local CSI," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10218-10233, Aug. 2023.

- H. Luo, A. Bishnu, and T. Ratnarajah, "Design and Analysis of In-Band Full-Duplex Private 5G Networks Using FR2 Band," *IEEE Access*, vol. 9, pp.166886-166905, 2021.

- H. Luo, M. Holm, and T. Ratnarajah, "On the performance of active analog self-interference cancellation techniques for beyond 5G systems," *China Commun.*, vol. 18, no. 10, pp.158-168, 2021.

- H. Luo, A. Bishnu, and T. Ratnarajah, "On the Feasibility and Performance of Self-interference Cancellation in FR2 Band," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 907-912, Seoul, South Korea, May 16-20, 2022.

- H. Luo, N. Garg, and T. Ratnarajah, "A Channel Frequency Response-Based Secret Key Generation Scheme in In-band Full-duplex MIMO Systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Rome, Italy, May 28-June 1, 2023.

- H. Luo, M. Holm, and T. Ratnarajah, "Design and Analysis of Wide-band Self-interference Cancellation for Full-duplex Wireless Networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 2697-2702, Austin, TX, USA, April 10-13, 2022.

- H. Luo, M. Holm, and T. Ratnarajah, "Wideband active analog self-interference cancellation for 5G and beyond full-duplex systems," in *Proc. 54th Asilomar Conference on Signals, Systems, and Computers*, pp. 868-872, Pacific Grove, California, Nov. 1-4, 2020.

- H. Luo, A. Bishnu, and T. Ratnarajah, "In-Band Full-Duplex Integrated Sensing and Communication Using OTFS Waveforms," in *Proc. 57th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, Oct. 29-Nov. 1, 2023.

- H. Luo, N. Garg, M. Holm, and T. Ratnarajah, "Joint Power Allocation and Beamforming for In-band Full-duplex Multi-cell Multi-user Networks," submitted.

- H. Luo, A. Bishnu, M. Sellathurai, and T. Ratnarajah, "Transceiver Design for In-band Full-duplex MIMO Systems in Highly Time-Varying Channel," submitted.

- J. Zhang, H. Luo, N. Garg, A. Bishnu, M. Holm, and T. Ratnarajah, "Design and analysis of wideband in-band-full-duplex FR2-IAB networks," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 6, pp. 4183-4196, 2022.

- J. Zhang, H. Luo, N. Garg, M. Holm, and T. Ratnarajah, "Design and Analysis of mmWave Full-Duplex Integrated Access and Backhaul Networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1-6, Montreal, Canada, June 14-23, 2021.

Bibliography

[1] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6g wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, 2019.

[2] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6g wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, 2020.

[3] 3GPP, "Study on channel model for frequencies from 0.5 to 100 ghz," tech. rep., 3rd Generation Partnership Project (3GPP), 2020.

[4] K. E. Kolodziej, S. Yegnanarayanan, and B. T. Perry, "Fiber bragg grating delay lines for wideband self-interference cancellation," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4005–4014, 2019.

[5] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[6] K. David and H. Berndt, "6g vision and requirements: Is there any need for beyond 5g?," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 72–80, 2018.

[7] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, *et al.*, "On the road to 6g: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surv. Tutor.*, 2023.

[8] H. Luo, M. Holm, and T. Ratnarajah, "On the performance of active analog self-interference cancellation techniques for beyond 5g systems," *China Commun.*, vol. 18, no. 10, pp. 158–168, 2021.

[9] H. Luo, A. Bishnu, and T. Ratnarajah, "Design and analysis of in-band full-duplex private 5g networks using fr2 band," *IEEE Access*, vol. 9, pp. 166886–166905, 2021.

[10] H. Luo, M. Holm, and T. Ratnarajah, "Design and analysis of wideband self-interference cancellation for full-duplex wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 2697–2702, IEEE, 2022.

[11] H. Luo, A. Bishnu, and T. Ratnarajah, "On the feasibility and performance of self-interference cancellation in fr2 band," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 907–912, IEEE, 2022.

[12] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM Conf.*, pp. 375–386, 2013.

[13] M. Jain, J. I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. ACM Mobicom Annu. Int. Conf. Mobile Comput. Netw.*, pp. 301–312, 2011.

[14] X. Su, X. Han, S. Fu, S. Wang, C. Li, Q. Tan, G. Zhu, C. Wang, Z. Wu, Y. Gu, *et al.*, "Optical multipath rf self-interference cancellation based on phase modulation for full-duplex communication," *IEEE Photon. J.*, vol. 12, no. 4, pp. 1–14, 2020.

[15] C. Motz, T. Paireder, H. Pretl, and M. Huemer, "A survey on self-interference cancellation in mobile lte-a/5g fdd transceivers," *IEEE Trans. Circuits Syst. II: Express Br.*, vol. 68, no. 3, pp. 823–829, 2021.

[16] K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3025–3041, 2019.

[17] T. Chen, S. Garikapati, A. Nagulu, A. Gaonkar, M. Kohli, I. Kadota, H. Krishnaswamy, and G. Zussman, "A survey and quantitative evaluation of integrated circuit-based antenna interfaces and self-interference cancellers for full-duplex," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1753–1776, 2021.

[18] 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, p. 1699, 2022.

[19] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 2, pp. 680–694, 2014.

[20] A. Balatsoukas-Stimming, P. Belanovic, K. Alexandris, and A. Burg, "On self-interference suppression methods for low-complexity full-duplex mimo," in *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 992–997, IEEE, 2013.

[21] R. Askar, T. Kaiser, B. Schubert, T. Haustein, and W. Keusgen, "Active self-interference cancellation mechanism for full-duplex wireless transceivers," in *Proc. International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, pp. 539–544, IEEE, 2014.

[22] 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. Wirel. Commun.*, 2022.

[23] 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.

[24] V. Panse, T. K. Jain, P. K. Sharma, and A. Kothari, "Digital self-interference cancellation in the era of machine learning: A comprehensive review," *Phys. Commun.*, vol. 50, p. 101526, 2022.

[25] A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, pp. 1–5, IEEE, 2018.

[26] 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, 2020.

[27] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6291–6307, 2022.

[28] D. Korpi, M. Heino, C. Icheln, K. Haneda, and M. Valkama, "Compact inband full-duplex relays with beyond 100 db self-interference suppression: Enabling techniques and field measurements," *IEEE Trans. Antennas Propag.*, vol. 65, no. 2, pp. 960–965, 2016.

[29] A. Nadh, J. Samuel, A. Sharma, S. Aniruddhan, and R. K. Ganti, "A taylor series approximation of self-interference channel in full-duplex radios," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 7, pp. 4304–4316, 2017.

[30] 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.

[31] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex mimo relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541–1553, 2012.

[32] A. Bishnu, M. Holm, and T. Ratnarajah, "Performance evaluation of full-duplex iab multi-cell and multi-user network for fr2 band," *IEEE Access*, vol. 9, pp. 72269–72283, 2021.

[33] N. Reiskarimian, J. Zhou, T.-H. Chuang, and H. Krishnaswamy, "Analysis and design of two-port n-path bandpass filters with embedded phase shifting," *IEEE Trans. Circuits Syst. II: Express Br.*, vol. 63, no. 8, pp. 728–732, 2016.

[34] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in mimo full-duplex transceivers under crosstalk," *EURASIP J. Wirel. Commun. Netw.*, vol. 2017, pp. 1–15, 2017.

[35] J. Zhang, N. Garg, M. Holm, and T. Ratnarajah, "Design of full duplex millimeter-wave integrated access and backhaul networks," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 60–67, 2021.

[36] J. Zhang, H. Luo, N. Garg, M. Holm, and T. Ratnarajah, "Design and analysis of mmwave full-duplex integrated access and backhaul networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, IEEE, 2021.

[37] J. Zhang, H. Luo, N. Garg, A. Bishnu, M. Holm, and T. Ratnarajah, "Design and analysis of wideband in-band-full-duplex fr2-iab networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4183–4196, 2021.

[38] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5g urllc: Design challenges and system concepts," in *Proc. 15th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6, IEEE, 2018.

[39] H. Luo, M. Holm, and T. Ratnarajah, "Wideband active analog self-interference cancellation for 5g and beyond full-duplex systems," in *Proc. 54th Asilomar Conference on Signals, Systems, and Computers*, pp. 868–872, IEEE, 2020.

[40] X. Yu, J. Ye, L. Yan, T. Zhou, X. Zou, and W. Pan, "Photonic-assisted multipath self-interference cancellation for wideband mimo radio-over-fiber transmission," *J. Light. Technol.*, vol. 40, no. 2, pp. 462–469, 2022.

[41] Z. Zhu, C. Gao, S. Zhao, T. Zhou, G. Wang, H. Li, and Q. Tan, "Photonics-assisted ultrawideband rf self-interference cancellation with signal of interest recovery and fiber transmission," *J. Light. Technol.*, vol. 40, no. 3, pp. 655–663, 2022.

[42] Y. Xing, S. Li, X. Xue, and X. Zheng, "Photonic-assisted rf self-interference cancellation based on optical spectrum processing," *J. Light. Technol.*, vol. 40, no. 7, pp. 2015–2022, 2022.

[43] M. Han, T. Shi, and Y. Chen, "Digital-assisted photonic analog wideband multipath self-interference cancellation," *IEEE Photon. Technol. Lett.*, vol. 34, no. 5, pp. 299–302, 2022.

[44] T. Shi and Y. Chen, "Photonics-assisted analog wideband self-interference cancellation for in-band full-duplex mimo systems with adaptive digital amplitude and delay pre-matching," *J. Light. Technol.*, 2023.

[45] A. Pandey, K. Van Gasse, and D. Van Thourhout, "Integrated photonics approach to radio-frequency self-interference cancellation," *Optics Continuum*, vol. 1, no. 8, pp. 1668–1675, 2022.

[46] H. Luo, N. Garg, and T. Ratnarajah, "Beamforming design for in-band full-duplex multi-cell multi-user mimo networks with global and local csi," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10218–10233, 2023.

[47] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, 2010.

[48] M. Mohammadi, H. A. Suraweera, Y. Cao, I. Krikidis, and C. Tellambura, "Full-duplex radio for uplink/downlink wireless access with spatially random nodes," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5250–5266, 2015.

[49] M. Mohammadi, H. A. Suraweera, and C. Tellambura, "Uplink/downlink rate analysis and impact of power allocation for full-duplex cloud-rans," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 9, pp. 5774–5788, 2018.

[50] R. Shafin and L. Liu, "Multi-cell multi-user massive fd-mimo: Downlink precoding and throughput analysis," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 1, pp. 487–502, 2018.

[51] J. Ye, A. Kammoun, M.-S. Alouini, *et al.*, "Sum-rate analysis of a multi-cell multi-user miso system under double scattering channels," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 332–349, 2021.

[52] A. C. Cirik, S. Biswas, S. Vuppala, and T. Ratnarajah, "Beamforming design for full-duplex mimo interference channels–qos and energy-efficiency considerations," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4635–4651, 2016.

[53] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell mimo networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590–1608, 2017.

[54] E. Balti and N. Mensi, "Zero-forcing max-power beamforming for hybrid mmwave full-duplex mimo systems," in *Proc. lnt. Conf. Adv. Syst. Emerg. Technol. (IC_ASET)*, pp. 344–349, IEEE, 2020.

[55] A. Koc and T. Le-Ngoc, "Intelligent non-orthogonal beamforming with large self-interference cancellation capability for full-duplex multiuser massive mimo systems," *IEEE Access*, vol. 10, pp. 51771–51791, 2022.

[56] I. P. Roberts, J. G. Andrews, and S. Vishwanath, "Hybrid beamforming for millimeter wave full-duplex under limited receive dynamic range," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 12, pp. 7758–7772, 2021.

[57] M. M. Fadoul and C. Y. Leow, "Joint nullspace projection-based interference mitigation for full-duplex relay-assisted multicell networks," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2392–2399, 2020.

[58] X. Huang, A. T. Le, and Y. J. Guo, "Transmit beamforming for communication and self-interference cancellation in full duplex mimo systems: A trade-off analysis," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 6, pp. 3760–3769, 2021.

[59] 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. Wirel. Commun.*, 2022.

[60] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X.-G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (noma) in 5g millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, 2018.

[61] X. Yu, T. Teng, X. Dang, S.-H. Leung, and F. Xu, "Joint power allocation and beamforming for energy-efficient design in multiuser distributed mimo systems," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4128–4143, 2021.

[62] J. Choi, J. Park, and N. Lee, "Energy efficiency maximization precoding for quantized massive mimo systems," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 6803–6817, 2022.

[63] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical csi," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, 2019.

[64] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Exploiting amplitude control in intelligent reflecting surface aided wireless communication with imperfect csi," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4216–4231, 2021.

[65] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[66] 3GPP, "5g nr; physical channels and modulation," tech. rep., 3rd Generation Partnership Project (3GPP), 2018.

[67] H. Luo, N. Garg, and T. Ratnarajah, "A channel frequency response-based secret key generation scheme in in-band full-duplex mimo-ofdm systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2951–2965, 2023.

[68] H. Luo, N. Garg, and T. Ratnarajah, "A channel frequency response-based secret key generation scheme in in-band full-duplex mimo systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, IEEE, 2023.

[69] X. Wei and D. Saha, "Knew: Key generation using neural networks from wireless channels," in *Proc. ACM Workshop on Wireless Security and Machine Learning*, pp. 45–50, 2022.

[70] D. Guo, K. Cao, J. Xiong, D. Ma, and H. Zhao, "A lightweight key generation scheme for the internet of things," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12137–12149, 2021.

[71] G. Li, A. Hu, J. Zhang, L. Peng, C. Sun, and D. Cao, "High-agreement uncorrelated secret key generation based on principal component analysis preprocessing," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3022–3034, 2018.

[72] F. Rottenberg, T.-H. Nguyen, J.-M. Dricot, F. Horlin, and J. Louveaux, "Csi-based versus rss-based secret-key generation under correlated eavesdropping," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1868–1881, 2020.

[73] K. Lin, Z. Ji, Y. Zhang, G. Chen, P. L. Yeoh, and Z. He, "Secret key generation based on 3d spatial angles for uav communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 1–6, IEEE, 2021.

[74] A. K. Junejo, F. Benkhelifa, B. Wong, and J. A. McCann, "Lora-lisk: a lightweight shared secret key generation scheme for lora networks," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4110–4124, 2021.

[75] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, "Hardware-impaired phy secret key generation with man-in-the-middle adversaries," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 856–860, 2022.

[76] H. Vogt, Z. H. Awan, and A. Sezgin, "Secret-key generation: Full-duplex versus half-duplex probing," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 639–652, 2018.

[77] N. Aldaghri and H. Mahdavifar, "Physical layer secret key generation in static environments," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2692–2705, 2020.

[78] L. Lai, Y. Liang, and W. Du, "Cooperative key generation in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1578–1588, 2012.

[79] K. Anjana and R. Ramanathan, "Impact of channel estimation errors on secret key capacity in mimo-ofdm systems," in *Proc. Int. Conf. Commun. Inf. Comput. Technol. (ICCICT)*, pp. 1–6, IEEE, 2021.

[80] C.-Y. Wu, P.-C. Lan, P.-C. Yeh, C.-H. Lee, and C.-M. Cheng, "Practical physical layer security schemes for mimo-ofdm systems using precoding matrix indices," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1687–1700, 2013.

[81] G. Li, C. Sun, E. A. Jorswieck, J. Zhang, A. Hu, and Y. Chen, "Sum secret key rate maximization for tdd multi-user massive mimo wireless networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 968–982, 2020.

[82] C. Chen and M. A. Jensen, "Secret key establishment using temporally and spatially correlated wireless channel coefficients," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 205–215, 2010.

[83] A. Sadeghi, M. Zorzi, and F. Lahouti, "Analysis of key generation rate from wireless channel in in-band full-duplex communications," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, pp. 104–109, IEEE, 2016.

[84] Z. Zhuang, S. Jiang, Y. Xu, X. Luo, and X. Cheng, "A physical layer key generation scheme based on full-duplex mode in wireless networks without fixed infrastructure," in *Proc. International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1–5, IEEE, 2019.

[85] R. Sohrabi, Q. Zhu, and Y. Hua, "Secrecy analyses of a full-duplex mimome network," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5968–5982, 2019.

[86] M. Wang, F. Gao, S. Jin, and H. Lin, "An overview of enhanced massive mimo with array signal processing techniques," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 886–901, 2019.

[87] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, and E. Barker, "A statistical test suite for random and pseudorandom number generators for cryptographic applications," tech. rep., Booz-allen and hamilton inc mclean va, 2001.

[88] G. Li, C. Sun, W. Xu, M. Di Renzo, and A. Hu, "On maximizing the sum secret key rate for reconfigurable intelligent surface-assisted multiuser systems," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 211–225, 2021.

[89] M. Adil, S. Wyne, and S. J. Nawaz, "On quantization for secret key generation from wireless channel samples," *IEEE Access*, vol. 9, pp. 21653–21668, 2021.

[90] Y. Liu, S. C. Draper, and A. M. Sayeed, "Exploiting channel diversity in secret key generation from multipath fading randomness," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 5, pp. 1484–1497, 2012.

[91] G. Li, Z. Zhang, Y. Yu, and A. Hu, "A hybrid information reconciliation method for physical layer key generation," *Entropy*, vol. 21, no. 7, p. 688, 2019.

[92] K. Moara-Nkwe, Q. Shi, G. M. Lee, and M. H. Eiza, "A novel physical layer secure key generation and refreshment scheme for wireless sensor networks," *IEEE Access*, vol. 6, pp. 11374–11387, 2018.

[93] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[94] D. K. P. Tan, J. He, Y. Li, A. Bayesteh, Y. Chen, P. Zhu, and W. Tong, "Integrated sensing and communication in 6g: Motivations, use cases, requirements, challenges and future directions," in *Proc. IEEE International Online Symposium on Joint Communications and Sensing (JC&S)*, pp. 1–6, IEEE, 2021.

[95] P. Kumari, S. A. Vorobyov, and R. W. Heath, "Adaptive virtual waveform design for millimeter-wave joint communication–radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 715–730, 2019.

[96] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proc. IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.

[97] M. F. Keskin, H. Wymeersch, and V. Koivunen, "Mimo-ofdm joint radar-communications: Is ici friend or foe?," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1393–1408, 2021.

[98] H. Qu, G. Liu, L. Zhang, S. Wen, and M. A. Imran, "Low-complexity symbol detection and interference cancellation for otfs system," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1524–1537, 2020.

[99] P. Singh, A. Gupta, H. B. Mishra, and R. Budhiraja, "Low-complexity zf/mmse mimo-otfs receivers for high-speed vehicular communication," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 209–227, 2022.

[100] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 10, pp. 6501–6515, 2018.

[101] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Orthogonal time frequency space (otfs) modulation based radar system," in *Proc. of IEEE Radar Conf. (RadarConf)*, pp. 1–6, IEEE, 2019.

[102] Z. Zhou, L. Liu, J. Xu, and R. Calderbank, "Learning to equalize otfs," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 9, pp. 7723–7736, 2022.

[103] C. Jin, Z. Bie, X. Lin, W. Xu, and H. Gao, "A simple two-stage equalizer for otfs with rectangular windows," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1158–1162, 2020.

[104] W. Yuan, Z. Wei, S. Li, J. Yuan, and D. W. K. Ng, "Integrated sensing and communication-assisted orthogonal time frequency space transmission for vehicular networks," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 6, pp. 1515–1528, 2021.

[105] Y. Shan and F. Wang, "Low-complexity and low-overhead receiver for otfs via large-scale antenna array," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5703–5718, 2021.

[106] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for otfs in delay–doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, 2019.

[107] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, "Channel estimation for orthogonal time frequency space (otfs) massive mimo," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204–4217, 2019.

[108] J. Cheng, H. Gao, W. Xu, Z. Bie, and Y. Lu, "Low-complexity linear equalizers for otfs exploiting two-dimensional fast fourier transform," *arXiv preprint arXiv:1909.00524*, 2019.

[109] X. Wang, W. Shen, C. Xing, J. An, and L. Hanzo, "Joint bayesian channel estimation and data detection for otfs systems in leo satellite communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4386–4399, 2022.

[110] Z. ZHANG, H. LIU, Q. WANG, and P. FAN, "A survey on low complexity detectors for otfs systems," *ZTE Communications*, vol. 19, no. 4, pp. 3–15, 2022.

[111] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of otfs for joint radar parameter estimation and communication," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 9, pp. 5951–5965, 2020.

[112] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "Hybrid digital-analog beamforming and mimo radar with otfs modulation," *arXiv preprint arXiv:2009.08785*, 2020.

[113] S. Li, W. Yuan, C. Liu, Z. Wei, J. Yuan, B. Bai, and D. W. K. Ng, "A novel isac transmission framework based on spatially-spread orthogonal time frequency space modulation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1854–1872, 2022.

[114] T. Thaj and E. Viterbo, "Low complexity iterative rake decision feedback equalizer for zero-padded otfs systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15606–15622, 2020.

[115] J. Zhang, N. Garg, and T. Ratnarajah, "In-band-full-duplex integrated sensing and communications for iab networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 12782–12796, 2022.

[116] S. P. Lavery and T. Ratnarajah, "Airborne phased array ofdm joint radar-communications system," in *Proc. IEEE Radar Conf. (RadarConf)*, pp. 1–6, IEEE, 2022.

[117] N. Hashimoto, N. Osawa, K. Yamazaki, and S. Ibi, "Channel estimation and equalization for cp-ofdm-based otfs in fractional doppler channels," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, pp. 1–7, IEEE, 2021.

[118] 3GPP, "Nr; study on integrated access and backhaul," tech. rep., 3rd Generation Partnership Project (3GPP), 2018.

[119] L. Gaudio, G. Colavolpe, and G. Caire, "Otfs vs. ofdm in the presence of sparsity: A fair comparison," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 6, pp. 4410–4423, 2021.

[120] T. ETSI, "Lte evolved universal terrestrial radio access (e-utra), base station (bs) radio transmission and reception," *LTE, Evolved Universal Terrestrial Radio Access (E-UTRA) Base Station (BS) Radio Transmission and Reception*, 2010.

[121] G. Van der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017.

[122] S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation," *Opt. Express.*, vol. 29, no. 20, pp. 30991–30997, 2021.

[123] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Express*, vol. 20, no. 20, pp. 22783–22795, 2012.

[124] S. Liu, T. Wang, and S. Wang, "Joint compensation of cfo and iq imbalance in ofdm receiver: A deep learning based approach," in *Proc. IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 793–798, IEEE, 2021.

[125] A. Singh, S. Sharma, K. Deka, and V. Bhatia, "Dl-based otfs signal detection in presence of hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 12, no. 09, pp. 1533–1537, 2023.

[126] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.

[127] H. Sharma, N. Kumar, and R. Tekchandani, "Physical layer security using beamforming techniques for 5g and beyond networks: A systematic review," *Phys. Commun.*, p. 101791, 2022.