SPECTRUM AND ENERGY EFFICIENT DESIGNS FOR NEXT GENERATION

WIRELESS COMMUNICATION SYSTEMS

by

Yahia Ramadan Ahmed Mohamed Ramadan

APPROVED BY SUPERVISORY COMMITTEE:

_____

Dr. Hlaing Minn, Chair

_____

Dr. Naofal Al-Dhahir

_____

Dr. John P. Fonseka

_____

Dr. Murat Torlak

Dedicated to my family and my fiancée, I would not be anything without your support.

SPECTRUM AND ENERGY EFFICIENT DESIGNS FOR NEXT GENERATION

WIRELESS COMMUNICATION SYSTEMS

by

YAHIA RAMADAN AHMED MOHAMED RAMADAN, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2018

ACKNOWLEDGMENTS

SPECTRUM AND ENERGY EFFICIENT DESIGNS FOR NEXT GENERATION WIRELESS COMMUNICATION SYSTEMS

Yahia Ramadan Ahmed Mohamed Ramadan, PhD
The University of Texas at Dallas, 2018

Supervising Professor: Dr. Hlaing Minn, Chair

Millimeter wave (mmWave) and Tera-Hertz (THz) communications are promising technologies for next generation wireless communication systems. However, their transceivers suffer high cost and high power consumption. On the other hand, the coexistence between wireless communication systems and wireless passive services (i.e., radio astronomy systems (RASs)) is an issue. This dissertation develops spectrum and energy efficient designs for mmWave/THz transceivers and proposes a new coexistence paradigm between wireless communication systems and RAS.

The first part of the dissertation proposes hybrid analog-digital precoding designs for multiuser millimeter wave systems with different hardware complexities and different types of channel knowledge to enhance physical layer security and maximize average network sum-rate. The proposed hybrid precoders achieve performance comparable to that of the fully digital precoding with much hardware-complexity reduction.

The second part of the dissertation studies nonlinear THz communication systems by incorporating the nonlinearity aspects of the low-cost THz devices and the inphase and quadrature (I/Q) imbalance effect into the signal model. The proposed precompensation schemes overcome the prominent problems experienced in the existing THz systems, namely severe

nonlinear distortions of the modulation symbols as well as spectral spreading and/or large spectrum sidelobes, and mitigate the I/Q imbalance effect.

The third part of the dissertation proposes a new spectrum sharing paradigm between cellular wireless communications (CWC) and radio astronomy systems (RAS) which enables geographical and spectral coexistence between CWC and RAS. The proposed paradigm offers: 1) certain guaranteed spectrum access to RAS, which is impossible in the existing paradigm, 2) capability to handle higher peak and mean traffics to CWC under spectrum restructuring of both CWC and RAS bands, and 3) overall improved spectrum utilization. Furthermore, a shared spectrum access strategy for RAS and WiFi systems is also developed by modifying the distributed medium access protocol.

CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Next generation wireless communication systems demand an exponential increase in data rate. The spectrum available in the microwave band is too scarce to answer such data rate need. This leads to a potential use of the underutilized millimeter-wave (mmWave) band (ranging from 30 GHz to 300 GHz [109]) and Tera-Hertz (THz) band (ranging from 0.1 THz to 10 THz [119]). On the other hand, radio astronomy systems (RASs) make astronomical observations in all available atmospheric windows ranging from 2 MHz to 1000 GHz and above [55]. In this dissertation, on one hand we develop energy-efficient designs for mmWave/THz transceivers for next generation wireless communication systems, and on the other hand we propose a new coexistence paradigm between wireless communication systems and RAS.

Millimeter-wave communications can support multiple Gbps data rates, but since the carrier frequencies are so high, mmWave links suffer higher propagation path loss. Antenna arrays can be used to compensate such losses [109]. Tens of antennas can be packed into a small area in mmWave transceivers due to the tiny wavelength. However, implementing a separate radio-frequency (RF) chain for each antenna is impractical due to the high cost and power of mixed-signal devices. An efficient solution to reduce the hardware complexity and the power consumption is the hybrid analog-digital precoding, where the antenna array with N_{T} elements is connected via an analog RF precoder to N_{RF} RF chains ($N_{\mathrm{RF}} < N_{\mathrm{T}}$) which process the digitally-precoded transmitted streams [12]. The hardware constraints of the RF precoder and the coupling between the baseband and RF precoders make the design of the hybrid precoders challenging, and we address this issue in this dissertation.

The THz band is one of the least explored areas. It exhibits unique features unavailable in the frequency bands of current and near-future communication systems. It can

accommodate detection of specific types of gaseous molecules in targeted environments, thus enabling various applications including breath analysis for non-invasive medical diagnosis and indoor/industrial air quality control [118, 151, 117, 88, 52, 114, 110, 115, 125]. The THz band offers substantially larger bandwidth and data rates not feasible in the current and 5G communication systems, thus the IEEE 802.15 Task Group 3d is developing a THz communication standard for applications such as wireless kiosks for multimedia and software download, small-cells wireless fronthaul and backhaul, and for data centers. The small form-factor of THz circuits also enables chip-to-chip or within-device communications and enhanced beamforming for physical layer security of wireless communications. These advantages are very attractive but for broader consumer applications the major barrier is the cost of THz devices. Signal generation circuitries for low-cost THz devices are different from the conventional ones of the lower frequency bands. This is due to unavailability of THz oscillators and THz CMOS power amplifiers which is commonly known as "the THz Gap" [102]. Thus, low-cost THz transceivers rely on nonlinear devices in contrast to the linear devices of the lower frequency bands and they suffer from nonlinear distortions. In this dissertation, we develop solutions to this problem.

Spectrum sharing has been viewed as an efficient approach for enhanced spectrum utilization [103, 38]. Some progresses have been made for releasing and relocation of some of the federal radio spectrum, e.g., [103], and some rule making for shared spectrum access and management, e.g., [38]. The obvious trend is that the active usage of the radio spectrum will rise substantially across time, frequency and space, which is much desirable from the spectrum utilization perspective but it could cause harmful radio frequency interference (RFI) to passive services such as radio astronomy and earth exploration remote sensing. These passive services provide economically and scientifically important observations of Earth's environment, our solar system and the cosmos (e.g., weather/climate forecasting, fresh water resources, discoveries of pulsars, cosmic microwave background radiation,

gigantic molecular clouds where new stars are being born, active galactic nuclei where giant black holes reside, or solar flares affecting infrastructures and lives on Earth) [91]. Due to their benefits to society, several of the ITU-R recommendations stipulate for protection of them [97, 72, 55, 59, 56, 61, 57, 126]. Thus, it is very crucial to develop new spectrum sharing paradigms between the active wireless communications and the passive remote sensing systems that answer the needs of both types of systems. We also address this issue in this dissertation.

1.2   Outline and Contributions

In the first part of the dissertation (chapters 2, 3 and 4), we propose hybrid analog-digital precoding designs for multiuser millimeter wave systems with different hardware complexities and different types of channel knowledge to enhance physical layer security and maximize average network sum-rate. The proposed hybrid precoders achieve performance comparable to that of the fully digital precoding with much hardware-complexity reduction.

In the second part of the dissertation (chapter 5), we study nonlinear THz communication systems by incorporating the nonlinearity aspects of the low-cost THz devices and the inphase and quadrature (I/Q) imbalance effect into the signal model. The proposed precompensation schemes overcome the prominent problems experienced in the existing THz systems, namely severe nonlinear distortions of the modulation symbols as well as spectral spreading and/or large spectrum sidelobes, and mitigate the I/Q imbalance effect.

In the third part of the dissertation (chapters 6 and 7), we propose a new spectrum sharing paradigm between cellular wireless communications (CWC) and radio astronomy systems (RAS) which enables geographical and spectral coexistence between CWC and RAS. The proposed paradigm offers: 1) certain guaranteed spectrum access to RAS, which is impossible in the existing paradigm, 2) capability to handle higher peak and mean traffics to CWC under spectrum restructuring of both CWC and RAS bands, and 3) overall improved

spectrum utilization. Furthermore, a shared spectrum access strategy for RAS and WiFi systems is also developed by modifying the distributed medium access protocol.

1.3 Notations

The following notations are used throughout this dissertation: $\mathbf{A}$ is a matrix, and $\|\mathbf{A}\|_{\mathrm{F}}$ is its Frobenius norm. $\mathbf{a}$ is a vector, and $\|\mathbf{a}\|$ is its $l2$-norm. a is a scalar. $(\cdot)^T$ and $(\cdot)^H$ are the transpose and conjugate transpose operators respectively. $\mathbf{I}_N$ is the identity matrix of order N. $\mathrm{Tr}\,[\mathbf{A}]$ denotes the trace of $\mathbf{A}$. $\mathcal{E}_{\max}[\mathbf{A}]$ is the principal eigenvector of $\mathbf{A}$, and $\lambda_{\max}[\mathbf{A}]$ is the corresponding maximum eigenvalue, while $\mathcal{E}_{1:\mathrm{N}}[\mathbf{A}]$ is the first N principal eigenvectors of $\mathbf{A}$. $\mathcal{N}[\mathbf{A}]$ returns the orthonormal basis of the null space of $\mathbf{A}$. $\mathrm{diag}\,(a_1, a_2, \ldots, a_N)$ returns the diagonal concatenation of elements $a_1, a_2, \ldots, a_N$, while $\mathrm{blkdiag}\,(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N)$ returns the diagonal concatenation of block matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N$. $\gcd\,(a_1, a_2)$ is the greatest common divisor of a_1 and a_2. $\mathbb{P}\,(x)$ and $\mathbb{E}\,(x)$ denote the probability and expectation of x. $\mathbf{x} \sim \mathcal{CN}\,(0, \Sigma)$ means that $\mathbf{x}$ is a circularly-symmetric complex Gaussian random vector with zero mean and covariance matrix Σ. $\mathfrak{R}\,\{x\}$ is the real part of x. $\hat{x}$ denotes the estimate of x. We use $*$ to denote the convolution operation. We use MATLAB notations, where $\mathbf{a}\,(i:j)$ consists of the ith to the jth elements of $\mathbf{a}$, $\mathbf{A}\,(i, j)$ denotes the $(i, j)^{\mathrm{th}}$ element of $\mathbf{A}$, $\mathbf{A}\,(i:j, m)$ consists of the ith to the jth elements of the mth column of $\mathbf{A}$, $\mathbf{A}\,(i, :)$ consists of the ith row of $\mathbf{A}$, and $\mathbf{A}\,(i:j, :)$ and $\mathbf{A}\,(:, i:j)$ consist of the ith to the jth rows and columns of $\mathbf{A}$ respectively.

PART I

HYBRID ANALOG-DIGITAL PRECODING DESIGNS FOR MILLIMETER

WAVE MIMO SYSTEMS

CHAPTER 2

SECURE HYBRID PRECODING FOR MMWAVE MISO-OFDM SYSTEMS [1]

2.1 Introduction

Due to the broadcast nature of wireless links, wireless communication is susceptible to eavesdropping. As a result, physical layer security has recently gained a lot of interest in the literature especially for multiple-antenna systems. The spatial degrees of freedom can be exploited to enhance the main channel and degrade the channels to eavesdroppers (Eves). This enhances the secrecy rate which is defined as the minimum difference between the achievable rate of the main channel and each achievable rate of the channels to Eves [116]. This chapter develops hybrid precoding designs for physical layer security in mmWave systems.

With full channel knowledge at the transmitter (Alice), the beamforming strategy was proven to achieve the secrecy capacity when the intended receiver (Bob) has a single antenna [75]. A semidefinite programming (SDP) framework was developed in [76] to maximize the secrecy rate with perfect or imperfect channel knowledge and multiple eavesdroppers with multiple antennas. The generation of artificial noise (AN) on the null space of the main channel was also introduced to degrade only the channels to Eve. If Alice does not have any knowledge of the channels to Eves, AN is uniformly spread on the null space of the main channel (isotropic AN) [92]. With partial channel knowledge at Alice, spatially-selective AN on the null space of the main channel is generated to effectively degrade the channels to Eves [42, 78]. We note that the works in [75, 76, 92, 42, 78] were restricted to secure baseband precoding with full RF chains.

As for secure RF precoding, previous works have focused on directional modulation (DM). In [28, 29], the RF precoder is designed such that the transmitted symbol is correctly mod-

---

ulated along the desired direction while the signal constellation is distorted along the other directions. The antenna subset modulation (ASM) proposed in [130] adopts the same idea besides choosing a different subset of antennas at each symbol. Using this approach, an additional randomness in the constellation along the other directions is introduced. However, previous works in [28, 29, 130] consider only the case that Alice has a single RF chain and the channels are line-of-sight (LOS). The first work to consider secure RF precoding for frequency-selective channels was in [107], where the RF precoder is designed to maximize the secrecy rate with full or partial channel knowledge at Alice. However, the work in [107] considers only the case that Alice has a single RF chain, and lacks designing the hybrid precoder if Alice has multiple RF chains.

With partial channel knowledge at Alice, the sparse structure of mmWave channels was exploited in [66] to generate spatially-selective AN to minimize the secrecy outage probability. However, the work in [66] was also restricted to secure baseband precoding with full RF chains. The first work to consider secure hybrid precoding was in [153], where the baseband preocder is designed to generate isotropic AN, and the RF precoder is designed to enhance the main channel. However, the work in [153] considers only the case that Alice does not have any knowledge of the channels to Eves and the channels are flat fading. The security aspect for two-way relaying was considered in [48], where the source nodes have multiple antennas connected to a single RF chain while the relay has multiple antennas connected to two RF chains, and the secure RF precoders are designed to maximize the secrecy rate. The works in [66, 153, 48] were restricted to flat fading channels, but mmWave channels are likely to be frequency-selective due to the large transmission bandwidth [96]. To the best of our knowledge, secure hybrid precoding for frequency-selective mmWave channels with full or partial channel knowledge has not been developed in the literature. For flat fading channels, the hybrid precoder can be desgined per symbol. However, for frequency-selective channels, the RF precoder has to be fixed across the subcarriers of orthogonal frequency

division multiplexing (OFDM) symbol as it is applied in the time domain, while the baseband precoder is designed per subcarrier [6], thus creating a different design problem. Since we assume that Alice has some knowledge of the channels to Eves, we consider secure hybrid precoding using the beamforming strategy. Incorporation of AN in our framework requires a different development than the existing AN works, and hence we introduce it in chapter 2.

In this chapter, we investigate the hybrid precoder design in mmWave multiple-input single-output (MISO) systems for physical layer security. We consider two types of channel knowledge at the transmitter. With full channel knowledge at Alice, we design the hybrid precoder to maximize the secrecy rate. With partial channel knowledge at Alice, we design the hybrid precoder to maximize the secrecy throughput. Our main contributions are summarized as follows:

- With full channel knowledge at Alice, we provide sufficient conditions on the minimum number of RF chains needed to realize the performance of the fully digital precoding.

- By maximizing the average projection between the fully digital precoder and the hybrid precoder, we propose a low-complexity closed-form hybrid precoder design which achieves exactly the same performance of the alternating minimization algorithms in [146].

- We extend the conventional projected maximum ratio transmissions (P-MRT) scheme to realize the hybrid precoder. We propose two P-MRT schemes. The first P-MRT scheme nulls the channels to Eves at time domain (TD-PMRT), while the second P-MRT scheme nulls the channels to Eves at frequency domain (FD-P-MRT). The two schemes TD-P-MRT and FD-P-MRT have different regions of feasibility. We define our P-MRT as an adaptive scheme that applies the one with higher secrecy rate between TD-P-MRT and FD-P-MRT for each channel realization.

8

- We propose an iterative hybrid precoder design to maximize the secrecy rate. The optimal baseband preocder is obtained as a function of the RF precoder. As a result, we write the secrecy rate as a function of the RF precoder only. Then, we propose a simple gradient ascent algorithm to design the RF precoder.

- With partial channel knowledge, where Alice has full knowledge of the channel to Bob but has knowledge only of the angles of departure (AoDs) of the propagation paths to Eves, we derive a secrecy outage probability upper bound. We convert the secrecy throughput maximization problem into a sequence of secrecy outage probability minimization problems, each is solved for a fixed target secrecy rate. Then, we propose an alternating minimization algorithm, based on gradient descent, to minimize the secrecy outage probability.

- We present extensive simulation results to show that the proposed hybrid precoding designs achieve performance close to that of the fully digital precoding at low and moderate signal-to-noise ratios (SNRs), and sometimes at high SNRs depending on the system parameters.

The rest of this chapter is organized as follows. In section II, we describe the system and the channel models. In section III, with full channel knowledge at Alice, the hybrid precoder is designed to maximize the secrecy rate. In section IV, with partial channel knowledge at Alice, the hybrid precoder is designed for secrecy throughput maximization. In section V, we present the numerical results. Finally, section VI concludes this chapter.

2.2 System and Channel Models

2.2.1 System Model

We consider a secrecy mmWave MISO-OFDM system with K single-antenna receivers as shown in Fig. 2.1. The transmitter (Alice) sends a confidential message to the first re-
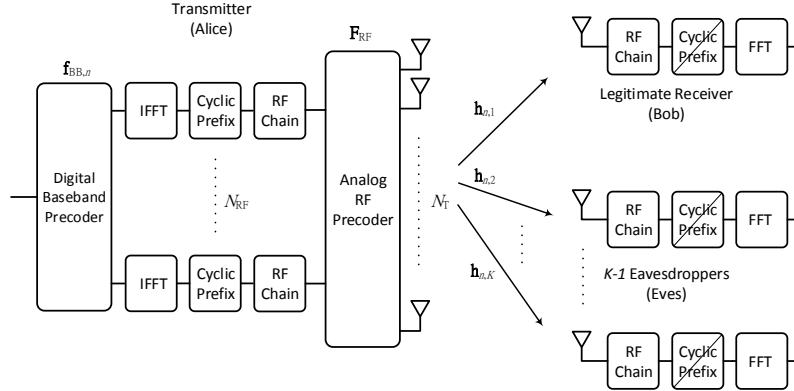
Figure 2.1. Secrecy mmWave massive MISO-OFDM system with $K - 1$ eavesdroppers.

ceiver (Bob), while the rest $K - 1$ receivers are eavesdroppers (Eves). We assume that the transmitter is equipped with a uniform linear array (ULA) with N_T antennas. The spacing between antennas is half the wavelength. To reduce the hardware complexity and the power consumption, the uniform linear antenna array is connected via an analog RF precoder to N_{RF} RF chains ($N_{RF} < N_T$) which process the digitally-precoded transmitted stream.

Due to the large transmission bandwidth of mmWave communications, mmWave channels are likely to be frequency-selective. Hence, OFDM is one of the most appropriate modulation techniques as it can convert the frequency-selective fading channel into a number of parallel flat fading subchannels. The received signal $y_{n,k}$ at the nth subcarrier and k^{th} receiver is given by

$$y_{n,k} = \mathbf{h}_{n,k} \mathbf{F}_{RF} \mathbf{f}_{BB,n} s_n + z_{n,k}, \tag{2.1}$$

where $\mathbf{h}_{n,k} \in \mathbb{C}^{1 \times N_T}$ is the mmWave frequency domain channel at the nth subcarrier to the kth receiver, $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_{RF}}$ is the analog RF precoder, $\mathbf{f}_{BB,n} \in \mathbb{C}^{N_{RF} \times 1}$ is the digital baseband precoder at the nth subcarrier, s_n is the transmitted modulated symbol with $\mathbb{E}\left[|s_n|^2\right] = P_T$, and $z_{n,k}$ is the zero-mean additive white complex Gaussian noise with variance σ^2 at the nth subcarrier and the kth receiver. Since the analog RF precoder $\mathbf{F}_{RF}$ is applied in time-domain, it is fixed across the subcarriers of OFDM symbol. On the other hand, the digital baseband precoder $\mathbf{f}_{BB,n}$ is designed per subcarrier since it is applied in frequency-domain.

10

The RF precoder $\mathbf{F}_{\mathrm{RF}}$ and baseband precoder $\mathbf{f}_{\mathrm{BB},n}$ have to be designed jointly due to the coupled power constraint, $\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 = 1 \ \forall n \in \{1, 2, \ldots N_{\mathrm{C}}\}$, where N_{C} is the number of subcarriers.

2.2.2 Channel Model

Millimeter-wave channels are expected to have limited scattering [145, 141]. Therefore, we adopt the sparse geometric multipath channel model in [113]. Under this model, the discrete-time channel vector $\tilde{\mathbf{h}}_{t,k} \in \mathbb{C}^{1 \times N_{\mathrm{T}}}$ at time instant t to the kth receiver is given by

$$\tilde{\mathbf{h}}_{t,k} = \sqrt{N_{\mathrm{T}}} \sum_{l=1}^{L} \alpha_{l,k} \mathbf{a}_{l,k}^{H} \delta\left(t - \tau_{l,k}\right), \tag{2.2}$$

where L is the number of propagation paths, $\alpha_{l,k}$ and $\tau_{l,k}$ are the channel gain and delay (in samples) of the the lth path to the kth receiver, $\{\alpha_{l,k}\}$ are independent complex Gaussian random variables with zero-mean and power delay profile $\{\rho_1, \rho_2, \ldots, \rho_L\}$, $\mathbf{a}_{l,k}$ is the transmit steering vectors of the lth path to the kth receiver with azimuth angle of departure (AoD) of $\varphi_{l,k}$,

$$\mathbf{a}_{l,k} = \frac{1}{\sqrt{N_{\mathrm{T}}}} \left[1, e^{-j\pi \cos(\varphi_{l,k})}, \ldots, e^{-j\pi(N_{\mathrm{T}}-1)\cos(\varphi_{l,k})}\right]^{T}, \tag{2.3}$$

and $\delta\left(t\right)$ is the Dirac delta function. Assuming perfect synchronization, the frequency domain channel vector $\mathbf{h}_{n,k}$ at the nth subcarrier to the kth receiver is given by

$$\mathbf{h}_{n,k} = \sqrt{N_{\mathrm{T}}} \sum_{l=1}^{L} \alpha_{l,k} \mathbf{a}_{l,k}^{H} \omega_{n,l,k}, \tag{2.4}$$

where $\omega_{n,l,k}$ is defined as $\omega_{n,l,k} = \exp\left(\frac{-j2\pi(n-1)\tau_{l,k}}{N_{\mathrm{c}}}\right)$. The frequency domain channel vector $\mathbf{h}_{n,k}$ can also be written in a compact form as

$$\mathbf{h}_{n,k} = \mathbf{w}_{n,k} \mathbf{D}_k \mathbf{A}_{\mathrm{T},k}^{H}, \tag{2.5}$$

where $\mathbf{w}_{n,k} = [\omega_{n,1,k}, \omega_{n,2,k}, \ldots, \omega_{n,L,k}] \in \mathbb{C}^{1 \times L}$, $\mathbf{D}_k = \mathrm{diag}\left(\alpha_{1,k}, \alpha_{2,k}, \ldots, \alpha_{L,k}\right) \in \mathbb{C}^{L \times L}$, and $\mathbf{A}_{\mathrm{T},k} \in \mathbb{C}^{N_{\mathrm{T}} \times L}$ is the transmit array response matrix to the kth receiver given by

$$\mathbf{A}_{\mathrm{T},k} = \left[\mathbf{a}_{1,k}, \mathbf{a}_{2,k}, \ldots, \mathbf{a}_{L,k}\right]. \tag{2.6}$$
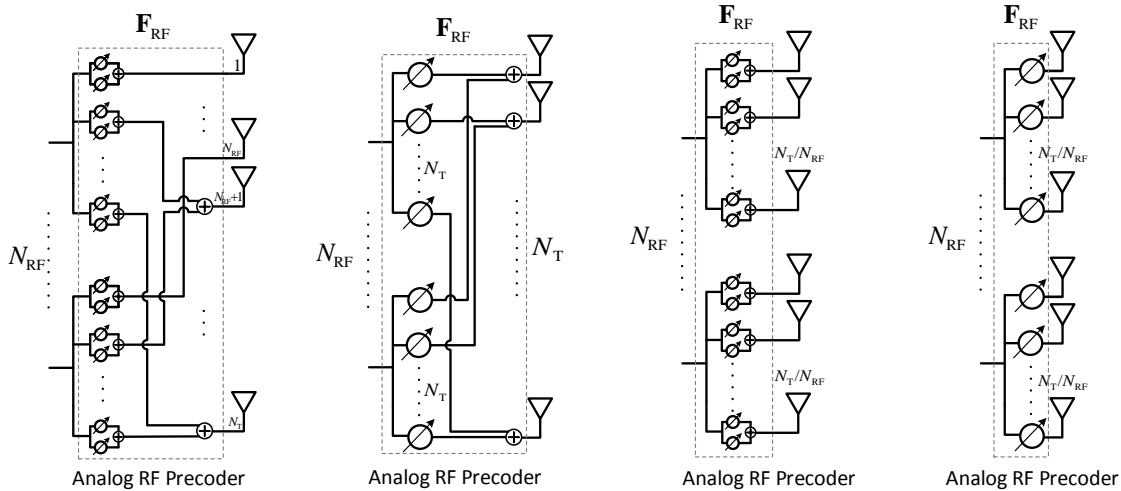
(a) Fully-connected struc-(b) Fully-connected struc-(c) Subarray structure S1(d) Subarray structure
ture F1   ture F2   S2

Figure 2.2. Analog RF precoding structures.

2.2.3 Analog RF Precoder Structures

The analog RF preocder $\mathbf{F}_{\mathrm{RF}}$ is usually implemented using analog phase shifters and analog combiners. Four structures for the analog RF precoder are shown in Fig. 2.2 The fully-connected structure F1 requires $2N_{\mathrm{RF}}(N_{\mathrm{T}} - N_{\mathrm{RF}} + 1)$ analog phase shifters and $N_{\mathrm{RF}}(N_{\mathrm{T}} - N_{\mathrm{RF}}) + N_{\mathrm{T}}$ analog combiners [19], while the fully-connected structure F2 requires $N_{\mathrm{T}}N_{\mathrm{RF}}$ analog phase shifters and N_{T} analog combiners. The subarray structure S1 requires $2N_{\mathrm{T}}$ analog phase shifters and N_{T} analog combiners, while the subarray structure S2 requires only N_{T} analog phase shifters.

In [19], it was shown that the fully-connected structure F1 has no constraints on the entries of $\mathbf{F}_{\mathrm{RF}}$. Each non-zero entry of $\mathbf{F}_{\mathrm{RF}}$ can be expressed as a sum of two analog phase shifters. The other three structures have constraints on the entries of $\mathbf{F}_{\mathrm{RF}}$. For the fully-connected structure F2, we have $|\mathbf{F}_{\mathrm{RF}}(l, m)| = 1 \ \forall l, m$. For the subarray structure S1, $\mathbf{F}_{\mathrm{RF}}$

has to be expressed as

$$
\mathbf{F}_{\mathrm{RF}} = \begin{bmatrix} \mathbf{f}_{\mathrm{RF},1} & 0 & \cdots & 0 \\ 0 & \mathbf{f}_{\mathrm{RF},2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{f}_{\mathrm{RF},N_{\mathrm{RF}}} \end{bmatrix}, \tag{2.7}
$$

where $\mathbf{f}_{\mathrm{RF},r} \in \mathbb{C}^{\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}} \times 1} \; \forall r \in \{1, 2, \ldots, N_{\mathrm{RF}}\}$. Similarly, each entry of $\mathbf{f}_{\mathrm{RF},r}$ can be expressed as a sum of two analog phase shifters. The subarray structure S2 has the same constraint in (2.7) and an additional constraint that $|\mathbf{f}_{\mathrm{RF},r}(l)| = 1 \; \forall l$. Let us denote by $\mathcal{F}_{\mathrm{RF}}^{\mathrm{F1}}$ the set of all $N_{\mathrm{T}} \times N_{\mathrm{RF}}$ complex matrices and by $\mathcal{F}_{\mathrm{RF}}^{\mathrm{F2}}$ the set of analog RF precoders satisfying the constraint of fully-connected structure F2, while by $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S1}}$ and $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S2}}$ the sets of analog RF precoders satisfying the constraints of subarray structure S1 and S2 respectively. Note that $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S2}} \subset \mathcal{F}_{\mathrm{RF}}^{\mathrm{S1}} \subset \mathcal{F}_{\mathrm{RF}}^{\mathrm{F1}}$ and $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S2}} \subset \mathcal{F}_{\mathrm{RF}}^{\mathrm{F2}} \subset \mathcal{F}_{\mathrm{RF}}^{\mathrm{F1}}$.

2.2.4 Hybrid Precoding Design Problems

The achievable rate R_k of the kth receiver is given by

$$
R_k = \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(1 + \gamma \left|\mathbf{h}_{n,k} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\right|^2\right), \tag{2.8}
$$

where $\gamma = P_{\mathrm{T}}/\sigma^2$ is the transmit SNR per subcarrier. Throughout the chapter, we assume that the secure coding is applied jointly across all subchannels (coding across sub-messages [14]). With full knowledge of all channels at the transmitter, maximizing the secrecy rate R_{sec} given by [116]

$$
R_{\mathrm{sec}} = \min_k \left\{R_1 - R_k\right\}_{k=2}^{K} \tag{2.9}
$$

is preferred. With full knowledge of the channel to Bob and partial knowledge of the channels to Eves at the transmitter, maximizing the secrecy throughput η_{sec} given by [143, 138]

$$
\eta_{\mathrm{sec}} = R_{\mathrm{sec}} \left(1 - \epsilon_{\mathrm{sec}}\right) \tag{2.10}
$$

13

is preferred, where ϵ_{sec} is the secrecy outage probability.

The subsequent parts of the chapter focus on designing the secure hybrid precoder for the the aforementioned two types of channel knowledge at the transmitter:

1. With full knowledge of all channels at the transmitter, section III focuses on designing the hybrid precoder to maximize the secrecy rate R_{sec},

$$
\begin{aligned}
[\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}] \;&=\; \underset{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}}{\arg\max} \; R_{\mathrm{sec}}, \\
&\text{s.t.} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \; \|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 = 1 \, \forall n. \quad (2.11)
\end{aligned}
$$

2. With full knowledge of the channel to Bob and partial knowledge of the channels to Eves at the transmitter, section IV focuses on designing the hybrid precoder to maximize the secrecy throughput η_{sec},

$$
\begin{aligned}
[\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}] \;&=\; \underset{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}}{\arg\max} \; \eta_{\mathrm{sec}}, \\
&\text{s.t.} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \; \|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 = 1 \, \forall n. \quad (2.12)
\end{aligned}
$$

Note that for the two hybrid precoder design problems in (2.11) and (2.12), $\mathcal{F}_{\mathrm{RF}}$ can be $\mathcal{F}_{\mathrm{RF}}^{\mathrm{F1}}$, $\mathcal{F}_{\mathrm{RF}}^{\mathrm{F2}}$, $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S1}}$, or $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S2}}$ according to the used structure. We consider all the four structures in our solutions.

2.3 Hybrid Precoder Design for Secrecy Rate Maximization with Full Channel Knowledge

In this section, the hybrid precoder is designed to maximize the secrecy rate R_{sec} for a given transmit SNR per subcarrier γ. We assume that Alice has full knowledge of the channels to Bob and to Eves. Bob and Eves have full knowledge of their channels to Alice. These assumptions become realistic if Eves are active nodes which have communicated with Alice [152]. Furthermore, we assume that Eves do not cooperate.

2.3.1 Fully Digital Precoding Design and the Minimum Number of RF Chains to Realize It

When $N_{\mathrm{RF}} = N_{\mathrm{T}}$, Alice applies fully digital precoding to maximize the secrecy rate R_{sec}. The optimal baseband precoder can be obtained by solving the optimization problem in (2.11) using semidefinite programming (SDP) as in [76]. As a closed-form near-optimal solution, we assume that Eves cooperate with each other. Therefore, the $K-1$ Eves are considered as one Eve with $K-1$ antennas. Note that this approximation gives a secrecy rate lower bound $\tilde{R}_{\mathrm{sec}}$, which can be written as

$$\tilde{R}_{\mathrm{sec}} = \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(\frac{1 + \gamma \left| \mathbf{h}_{n,1} \mathbf{f}_{\mathrm{opt},n} \right|^2}{1 + \gamma \left\| \mathbf{H}_n \mathbf{f}_{\mathrm{opt},n} \right\|^2} \right), \qquad (2.13)$$

where $\mathbf{H}_n = \left[\mathbf{h}_{n,2}^T, \mathbf{h}_{n,3}^T, \ldots, \mathbf{h}_{n,K}^T \right]^T \in \mathbb{C}^{(K-1) \times N_{\mathrm{T}}}$ and $\mathbf{f}_{\mathrm{opt},n} \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ is the fully digital precoder. Satisfying the power constraint $\left\| \mathbf{f}_{\mathrm{opt},n} \right\|^2 = 1$, $\mathbf{f}_{\mathrm{opt},n}$ maximizing (2.13) is obtained using generalized eigenvector decomposition as [116]

$$\mathbf{f}_{\mathrm{opt},n} = \boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{H}_n^H \mathbf{H}_n \right)^{-1} \left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \right) \right], \qquad (2.14)$$

and the corresponding $\tilde{R}_{\mathrm{sec}}$ is given by

$$\tilde{R}_{\mathrm{sec}} = \frac{1}{N_{\mathrm{c}}} \sum_{n=1}^{N_{\mathrm{c}}} \log_2 \left(\lambda_{\max} \left[\left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{H}_n^H \mathbf{H}_n \right)^{-1} \left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \right) \right] \right). \qquad (2.15)$$

Let us define $\mathbf{F}_{\mathrm{opt}} = [\mathbf{f}_{\mathrm{opt},1}, \mathbf{f}_{\mathrm{opt},2}, \ldots, \mathbf{f}_{\mathrm{opt},N_{\mathrm{C}}}] \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{C}}}$. Next, we provide sufficient conditions on the number of RF chains needed for the hybrid precoder to realize the performance of fully digital precoding (i.e., expressing $\mathbf{F}_{\mathrm{opt}}$ as $\mathbf{F}_{\mathrm{opt}} = \mathbf{F}_{\mathrm{RF}} [\mathbf{f}_{\mathrm{BB},1}, \mathbf{f}_{\mathrm{BB},2}, \ldots, \mathbf{f}_{\mathrm{BB},N_{\mathrm{C}}}]$).

Proposition 1. *To realize the performance of fully digital precoding, it is sufficient for the hybrid precoding utilizing the fully-connected structure F1 that $N_{\mathrm{RF}} \geq KL$. For the hybrid precoding utilizing the fully-connected structure F2, the sufficient condition becomes $N_{\mathrm{RF}} \geq 2KL$, and it reduces to $N_{\mathrm{RF}} \geq KL$ only if all the channels follow the mmWave channel model in (2.2).*

Proof. See Appendix A. □

Proposition 2. *For the subarray structures, there is no sufficient condition depending only on the number of RF chains to realize the performance of fully digital precoding.*

Proof. See Appendix B. □

For practical system parameters, the above sufficient conditions are not likely to be satisfied. Next, we provide different hybrid precoder designs to maximize the secrecy rate.

2.3.2 Low-Complexity Secrecy Hybrid Precoding Desgins

Approximating The Fully Digital Precoding (App-FD)

Traditionally, the hybrid precoder is designed to approximate the fully digital precoder by minimizing the average Euclidean distance between the fully digital precoder and the hybrid precoder [142, 94, 146]. Different from the average Euclidean distance criterion, we design the hybrid precoder to approximate the fully digital precoder by maximizing the average projection between the fully digital precoder and the hybrid precoder. Interestingly, the two criteria are related to each other, and they have similarity in the design of baseband precoder (see Appendix C).

In the following, we obtain closed-form solutions for the hybrid precoder maximizing the average projection between the fully digital precoder and the hybrid precoder. In other words, the hybrid precoder is designed as

$$
[\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}] = \underset{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}}{\arg\max} \sum_{n=1}^{N_{\mathrm{C}}} \left\| \mathbf{f}_{\mathrm{opt},n}^{H} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right\|^{2},
$$
$$
\text{s.t.} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \ \left\| \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right\|^{2} = 1 \ \forall n. \tag{2.16}
$$

After applying the power constraint $\left\| \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right\|^{2} = 1$ into the objective function and dropping the constraint on the entries of the RF precoder, we have

$$
[\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}] = \underset{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}}{\arg\max} \sum_{n=1}^{N_{\mathrm{C}}} \frac{\mathbf{f}_{\mathrm{BB},n}^{H} \left(\mathbf{F}_{\mathrm{RF}}^{H} \mathbf{f}_{\mathrm{opt},n} \mathbf{f}_{\mathrm{opt},n}^{H} \mathbf{F}_{\mathrm{RF}} \right) \mathbf{f}_{\mathrm{BB},n}}{\mathbf{f}_{\mathrm{BB},n}^{H} \left(\mathbf{F}_{\mathrm{RF}}^{H} \mathbf{F}_{\mathrm{RF}} \right) \mathbf{f}_{\mathrm{BB},n}}. \tag{2.17}
$$

16

Using generalized eigenvector decomposition, $\mathbf{f}_{\text{BB},n}$ maximizing (2.17) is given by

$$\mathbf{f}_{\text{BB},n} = \kappa_n \boldsymbol{\mathcal{E}}_{\max}\left[\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-1}\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}\mathbf{f}_{\text{opt},n}^H\mathbf{F}_{\text{RF}}\right] = \frac{\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-1}\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}}{\left\|\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-\frac{1}{2}}\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}\right\|}, \qquad (2.18)$$

where κ_n is a scaling factor to have $\|\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},n}\|^2 = 1$, and the second equality holds since $\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-1}\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}\mathbf{f}_{\text{opt},n}^H\mathbf{F}_{\text{RF}}$ is a rank-one matrix. Substituting (2.18) into (2.17), we get

$$\begin{aligned}
\mathbf{F}_{\text{RF}} &= \underset{\mathbf{F}_{\text{RF}}}{\arg\max}\ \sum_{n=1}^{N_{\text{C}}}\lambda_{\max}\left[\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-1}\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}\mathbf{f}_{\text{opt},n}^H\mathbf{F}_{\text{RF}}\right] \\
&= \underset{\mathbf{F}_{\text{RF}}}{\arg\max}\ \text{trace}\left[\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-\frac{1}{2}}\mathbf{F}_{\text{RF}}^H\left(\sum_{n=1}^{N_{\text{C}}}\mathbf{f}_{\text{opt},n}\mathbf{f}_{\text{opt},n}^H\right)\mathbf{F}_{\text{RF}}\left(\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\right)^{-\frac{1}{2}}\right] \\
&= \boldsymbol{\mathcal{E}}_{1:N_{\text{RF}}}\left[\sum_{n=1}^{N_{\text{C}}}\mathbf{f}_{\text{opt},n}\mathbf{f}_{\text{opt},n}^H\right].
\end{aligned} \qquad (2.19)$$

For the fully-connected structure F1, as there are no constraints on the entries of $\mathbf{F}_{\text{RF}}$, the RF precoder is simply given by (2.19). Since $\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}} = \mathbf{I}_{N_{\text{RF}}}$ due to (2.19), $\mathbf{f}_{\text{BB},n}$ in (2.18) is simplified to

$$\mathbf{f}_{\text{BB},n} = \frac{\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}}{\|\mathbf{F}_{\text{RF}}^H\mathbf{f}_{\text{opt},n}\|}. \qquad (2.20)$$

For the fully-connected structure F2, we obtain $\mathbf{F}_{\text{RF}}$ as

$$\mathbf{F}_{\text{RF}} = \exp\left(j\angle\boldsymbol{\mathcal{E}}_{1:N_{\text{RF}}}\left[\sum_{n=1}^{N_{\text{C}}}\mathbf{f}_{\text{opt},n}\mathbf{f}_{\text{opt},n}^H\right]\right), \qquad (2.21)$$

which satisfies the unit modulus constraint and is a good approximation to (2.19). Then, we obtain $\mathbf{f}_{\text{BB},n}$ as in (2.18) since $\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}} \neq \mathbf{I}_{N_{\text{RF}}}$.

For the subarray structures, $\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}} = \text{diag}\left(\|\mathbf{f}_{\text{RF},1}\|^2, \|\mathbf{f}_{\text{RF},2}\|^2, \ldots, \|\mathbf{f}_{\text{RF},N_{\text{RF}}}\|^2\right)$, and hence (2.19) can be solved for each $\mathbf{f}_{\text{RF},r}$ as

$$\mathbf{f}_{\text{RF},r} = \underset{\mathbf{f}_{\text{RF},r}}{\arg\max}\ \frac{\mathbf{f}_{\text{RF},r}^H\left(\sum_{n=1}^{N_{\text{C}}}\mathbf{f}_{\text{opt},n,r}\mathbf{f}_{\text{opt},n,r}^H\right)\mathbf{f}_{\text{RF},r}}{\mathbf{f}_{\text{RF},r}^H\mathbf{f}_{\text{RF},r}}, \qquad (2.22)$$

where $\mathbf{f}_{\text{opt},n,r} = \mathbf{f}_{\text{opt},n}\left((r-1)\frac{N_{\text{T}}}{N_{\text{RF}}}+1 : r\frac{N_{\text{T}}}{N_{\text{RF}}}\right) \in \mathbb{C}^{\frac{N_{\text{T}}}{N_{\text{RF}}}\times 1}\ \forall r \in \{1,2,\ldots,N_{\text{RF}}\}$. Therefore, $\mathbf{f}_{\text{RF},r}$ is obtained in a closed-form for the subarray structure S1 as

$$\mathbf{f}_{\text{RF},r} = \boldsymbol{\mathcal{E}}_{\max}\left[\sum_{n=1}^{N_{\text{C}}}\mathbf{f}_{\text{opt},n,r}\mathbf{f}_{\text{opt},n,r}^H\right], \qquad (2.23)$$

while $\mathbf{f}_{\mathrm{RF},r}$ for the subarray structure S2 is obtained as

$$\mathbf{f}_{\mathrm{RF},r} = \exp\left(j\angle\boldsymbol{\mathcal{E}}_{\max}\left[\sum_{n=1}^{N_{\mathrm{C}}}\mathbf{f}_{\mathrm{opt},n,r}\mathbf{f}_{\mathrm{opt},n,r}^{H}\right]\right). \tag{2.24}$$

Since $\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{RF}} = \mathbf{I}_{N_{\mathrm{RF}}}$ for the subarray structures due to (2.23) and (2.24), $\mathbf{f}_{\mathrm{BB},n}$ in (2.18) is simplified to

$$\mathbf{f}_{\mathrm{BB},n} = \frac{\left[\mathbf{f}_{\mathrm{RF},1}^{H}\mathbf{f}_{\mathrm{opt},n,1}, \mathbf{f}_{\mathrm{RF},2}^{H}\mathbf{f}_{\mathrm{opt},n,2}, \ldots, \mathbf{f}_{\mathrm{RF},N_{\mathrm{RF}}}^{H}\mathbf{f}_{\mathrm{opt},n,N_{\mathrm{RF}}}\right]^{T}}{\sqrt{\sum_{r=1}^{N_{\mathrm{RF}}}\left|\mathbf{f}_{\mathrm{RF},r}^{H}\mathbf{f}_{\mathrm{opt},n,r}\right|^{2}}}. \tag{2.25}$$

Although the average projection criterion and the average Euclidean distance criterion are slightly different (as shown in Appendix C), we observe in our numerical results that the proposed closed-form hybrid precoder achieves exactly the same performance of the hybrid precoder obtained in [146] which applies two nested iterative algorithms to design the hybrid precoder. The computational complexity of App-FD is $\mathcal{O}\left(N_{\mathrm{T}}^{3}N_{\mathrm{C}} + N_{\mathrm{T}}^{2}N_{\mathrm{C}}K\right)$.

Projected Maximum Ratio Transmission (P-MRT)

The main idea of P-MRT is to maximize the average SNR of Bob $\left(\overline{\mathrm{SNR}}_{\mathrm{B}}\right)$ in the null space of channels to Eves [76]. Generally, P-MRT is suboptimal at low and moderate SNRs but optimal at high SNRs. We will show how to design P-MRT using the hybrid precoder. Nulling the channels to Eves can be done using the analog RF precoder (at time domain TD-P-MRT) or using the digital baseband precoder (at frequency domain FD-P-MRT). The two schemes TD-P-MRT and FD-P-MRT have different regions of feasibility (as will be shown). As a result, our P-MRT adaptively selects the better scheme from TD-P-MRT and FD-P-MRT depending on the system parameters and channel realizations. This will yield higher secrecy rate.

TD-P-MRT   First, we consider the fully-connected structures F1 and F2. For both structures, it is necessary that $N_{\mathrm{T}} > (K-1)L$ to apply TD-P-MRT. We express $\mathbf{F}_{\mathrm{RF}}$ as

$$\mathbf{F}_{\mathrm{RF}} = \mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}, \tag{2.26}$$

where $\mathbf{U}_{\mathrm{RF}} = \mathcal{N}\left[\mathbf{A}_{\mathrm{T,Eves}}\right] \in \mathbb{C}^{N_{\mathrm{T}} \times (N_{\mathrm{T}} - (K-1)L)}$ is a unitary matrix in the null space of the channels to Eves, where $\mathbf{A}_{\mathrm{T,Eves}} \in \mathbb{C}^{N_{\mathrm{T}} \times (K-1)L}$ is given by

$$\mathbf{A}_{\mathrm{T,Eves}} = [\mathbf{A}_{\mathrm{T},2}, \mathbf{A}_{\mathrm{T},3}, \ldots, \mathbf{A}_{\mathrm{T},K}]. \tag{2.27}$$

We design $\tilde{\mathbf{F}}_{\mathrm{RF}}$ and $\mathbf{f}_{\mathrm{BB},n}$ to maximize $\overline{\mathrm{SNR}}_{\mathrm{B}}$ given by

$$\overline{\mathrm{SNR}}_{\mathrm{B}} = \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \frac{\left|\mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\right\|^2} = \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \frac{\mathbf{f}_{\mathrm{BB},n}^H \left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}}\right) \mathbf{f}_{\mathrm{BB},n}}{\mathbf{f}_{\mathrm{BB},n}^H \left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}}\right) \mathbf{f}_{\mathrm{BB},n}}, \tag{2.28}$$

where $\mathbf{U}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}} = \mathbf{I}_{N_{\mathrm{T}} - (K-1)L}$. Using generalized eigenvector decomposition, $\mathbf{f}_{\mathrm{BB},n}$ maximizing (2.28) is given by

$$\mathbf{f}_{\mathrm{BB},n} = \kappa_n \boldsymbol{\mathcal{E}}_{\max}\left[\left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}}\right)^{-1} \tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}}\right] = \frac{\left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}}\right)^{-1} \tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H}{\left\|\left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H\right\|}, \tag{2.29}$$

where κ_n is a scaling factor to have $\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\|^2 = 1$, the second equality holds since $\left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}}\right)^{-1} \tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}}$ is a rank-one matrix, and the corresponding $\overline{\mathrm{SNR}}_{\mathrm{B}}$ is expressed as

$$\overline{\mathrm{SNR}}_{\mathrm{B}} = \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \lambda_{\max}\left[\left(\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}}\right)^{-1} \tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}}\right]$$

$$= \frac{\gamma}{N_{\mathrm{C}}} \mathrm{trace}\left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \tilde{\mathbf{F}}_{\mathrm{RF}}^H \left(\sum_{n=1}^{N_{\mathrm{C}}} \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}}\right) \tilde{\mathbf{F}}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}\right]. \tag{2.30}$$

For the fully-connected structure F1, $\tilde{\mathbf{F}}_{\mathrm{RF}}$ maximizing (2.30) is obtained as

$$\tilde{\mathbf{F}}_{\mathrm{RF}} = \boldsymbol{\mathcal{E}}_{1:N_{\mathrm{RF}}}\left[\sum_{n=1}^{N_{\mathrm{C}}} \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}}\right]. \tag{2.31}$$

Since $\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}} = \mathbf{I}_{N_{\mathrm{RF}}}$ due to (2.31), $\mathbf{f}_{\mathrm{BB},n}$ in (2.29) is simplified to

$$\mathbf{f}_{\mathrm{BB},n} = \frac{\tilde{\mathbf{F}}_{\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H}{\left\|\mathbf{h}_{n,1} \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}}\right\|}, \tag{2.32}$$

19

which is the well-known MRT precoder for the equivalent channel $\mathbf{h}_{n,1}\mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}$. For the fully-connected structure F2, we need $\mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}$ to satisfy the unit modulus constraint. Consequently, we obtain $\tilde{\mathbf{F}}_{\mathrm{RF}}$ using only half of the number of RF chains as

$$\tilde{\mathbf{F}}_{\mathrm{RF}} = \mathcal{E}_{1:\frac{N_{\mathrm{RF}}}{2}}\left[\sum_{n=1}^{N_{\mathrm{C}}}\mathbf{U}_{\mathrm{RF}}^H\mathbf{h}_{n,1}^H\mathbf{h}_{n,1}\mathbf{U}_{\mathrm{RF}}\right], \tag{2.33}$$

and $\tilde{\mathbf{f}}_{\mathrm{BB},n} = \frac{\tilde{\mathbf{F}}_{\mathrm{RF}}^H\mathbf{U}_{\mathrm{RF}}^H\mathbf{h}_{n,1}^H}{\|\mathbf{h}_{n,1}\mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}\|} \in \mathbb{C}^{\frac{N_{\mathrm{RF}}}{2}\times 1}$. Then, $\mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}$ is decomposed (as described in proof of Proposition 1) as $\mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}} = \mathbf{Q}_{\mathrm{RF}}\mathbf{R}_{\mathrm{BB}}$ where $\mathbf{Q}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}}\times N_{\mathrm{RF}}}$ is with unit modulus entries and $\mathbf{R}_{\mathrm{BB}} \in \mathbb{R}^{N_{\mathrm{RF}}\times\frac{N_{\mathrm{RF}}}{2}}$. Finally, we set $\mathbf{F}_{\mathrm{RF}} = \mathbf{Q}_{\mathrm{RF}}$ and $\mathbf{f}_{\mathrm{BB},n} = \mathbf{R}_{\mathrm{BB}}\tilde{\mathbf{f}}_{\mathrm{BB},n} \in \mathbb{C}^{N_{\mathrm{RF}}\times 1}$.

For the subarray structure S1, each $\mathbf{f}_{\mathrm{RF},r}$ has to null the channels to Eves. As a result, we need $N_{\mathrm{T}} > N_{\mathrm{RF}}(K-1)L$. However, we may have $N_{\mathrm{T}} > (K-1)L$ but $N_{\mathrm{T}} \leq N_{\mathrm{RF}}(K-1)L$. Therefore, we divide the RF chains into $\tilde{N}_{\mathrm{RF}}$ distinct groups, each group has $\frac{N_{\mathrm{RF}}}{\tilde{N}_{\mathrm{RF}}}$ RF chains which process the same digitally-preocded symbols, where $\tilde{N}_{\mathrm{RF}} = \gcd\left(\min(\lfloor\frac{N_{\mathrm{T}}-1}{(K-1)L}\rfloor, N_{\mathrm{RF}}), N_{\mathrm{RF}}\right)$. Therefore, we should have $N_{\mathrm{T}} > \tilde{N}_{\mathrm{RF}}(K-1)L$ provided that $N_{\mathrm{T}} > (K-1)L$. Equivalently, we proceed assuming that we have $\tilde{N}_{\mathrm{RF}}$ chains, each is connected to $\frac{N_{\mathrm{T}}}{\tilde{N}_{\mathrm{RF}}}$ antennas such that $N_{\mathrm{T}} > \tilde{N}_{\mathrm{RF}}(K-1)L$. We express $\mathbf{f}_{\mathrm{RF},r}$ as

$$\mathbf{f}_{\mathrm{RF},r} = \mathbf{U}_{\mathrm{RF},r}\tilde{\mathbf{f}}_{\mathrm{RF},r}, \tag{2.34}$$

where $\mathbf{U}_{\mathrm{RF},r} = \mathcal{N}[\mathbf{A}_{\mathrm{T,Eves},r}] \in \mathbb{C}^{\frac{N_{\mathrm{T}}}{\tilde{N}_{\mathrm{RF}}}\times\left(\frac{N_{\mathrm{T}}}{\tilde{N}_{\mathrm{RF}}}-(K-1)L\right)}$ is a unitary matrix in the null space of the channels to Eves seen by the rth RF chain group, $\mathbf{A}_{\mathrm{T,Eves},r} = \mathbf{A}_{\mathrm{T,Eves}}\left((r-1)\frac{N_{\mathrm{T}}}{\tilde{N}_{\mathrm{RF}}}+1:r\frac{N_{\mathrm{T}}}{\tilde{N}_{\mathrm{RF}}},:\right)$, and $\tilde{\mathbf{f}}_{\mathrm{RF},r} \in \mathbb{C}^{\left(\frac{N_{\mathrm{T}}}{\tilde{N}_{\mathrm{RF}}}-(K-1)L\right)\times 1}$. Let $\mathbf{U}_{\mathrm{RF}} = \mathrm{blkdiag}\left(\mathbf{U}_{\mathrm{RF},1},\mathbf{U}_{\mathrm{RF},2},\ldots,\mathbf{U}_{\mathrm{RF},\tilde{N}_{\mathrm{RF}}}\right) \in \mathbb{C}^{N_{\mathrm{T}}\times\left(N_{\mathrm{T}}-\tilde{N}_{\mathrm{RF}}(K-1)L\right)}$ and $\tilde{\mathbf{F}}_{\mathrm{RF}} = \mathrm{blkdiag}\left(\tilde{\mathbf{f}}_{\mathrm{RF},1},\tilde{\mathbf{f}}_{\mathrm{RF},2},\ldots,\tilde{\mathbf{f}}_{\mathrm{RF},N_{\mathrm{RF}}}\right) \in \mathbb{C}^{\left(N_{\mathrm{T}}-\tilde{N}_{\mathrm{RF}}(K-1)L\right)\times N_{\mathrm{RF}}}$, then we have $\mathbf{F}_{\mathrm{RF}} = \mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}$, and $\mathbf{U}_{\mathrm{RF}}^H\mathbf{U}_{\mathrm{RF}} = \mathbf{I}_{N_{\mathrm{T}}-\tilde{N}_{\mathrm{RF}}(K-1)L}$. Therefore, $\overline{\mathrm{SNR}}_{\mathrm{B}}$ and $\mathbf{f}_{\mathrm{BB},n}$ can also be given by (2.28) and (2.29) respectively. Furthermore, $\overline{\mathrm{SNR}}_{\mathrm{B}}$ can be simplified to

$$\overline{\mathrm{SNR}}_{\mathrm{B}} = \frac{\gamma}{N_{\mathrm{C}}}\sum_{r=1}^{\tilde{N}_{\mathrm{RF}}}\frac{\tilde{\mathbf{f}}_{\mathrm{RF},r}^H\left(\sum_{n=1}^{N_{\mathrm{C}}}\mathbf{U}_{\mathrm{RF},r}^H\mathbf{h}_{n,1,r}^H\mathbf{h}_{n,1,r}\mathbf{U}_{\mathrm{RF},r}\right)\tilde{\mathbf{f}}_{\mathrm{RF},r}}{\tilde{\mathbf{f}}_{\mathrm{RF},r}^H\tilde{\mathbf{f}}_{\mathrm{RF},r}}, \tag{2.35}$$

where $\mathbf{h}_{n,1,r} = \mathbf{h}_{n,1}\left((r-1)\frac{N_\mathrm{T}}{\tilde{N}_\mathrm{RF}} + 1 : r\frac{N_\mathrm{T}}{\tilde{N}_\mathrm{RF}}\right) \ \forall r \in \left\{1, 2, \ldots, \tilde{N}_\mathrm{RF}\right\}$. Therefore, we obtain $\tilde{\mathbf{f}}_{\mathrm{RF},r}$ which maximizes (2.35) as

$$\tilde{\mathbf{f}}_{\mathrm{RF},r} = \mathcal{E}_\mathrm{max}\left[\sum_{n=1}^{N_\mathrm{C}} \mathbf{U}_{\mathrm{RF},r}^H \mathbf{h}_{n,1,r}^H \mathbf{h}_{n,1,r} \mathbf{U}_{\mathrm{RF},r}\right]. \tag{2.36}$$

Since $\tilde{\mathbf{F}}_{\mathrm{RF}}^H \tilde{\mathbf{F}}_{\mathrm{RF}} = \mathbf{I}_{\tilde{N}_\mathrm{RF}}$ due to (2.36), $\mathbf{f}_{\mathrm{BB},n}$ is simplified to

$$\mathbf{f}_{\mathrm{BB},n} = \frac{\left[\tilde{\mathbf{f}}_{\mathrm{RF},1}^H \mathbf{U}_{\mathrm{RF},1}^H \mathbf{h}_{n,1,1}^H, \tilde{\mathbf{f}}_{\mathrm{RF},2}^H \mathbf{U}_{\mathrm{RF},2}^H \mathbf{h}_{n,1,2}^H, \ldots, \tilde{\mathbf{f}}_{\mathrm{RF},N_\mathrm{RF}}^H \mathbf{U}_{\mathrm{RF},N_\mathrm{RF}}^H \mathbf{h}_{n,1,N_\mathrm{RF}}^H\right]^T}{\sqrt{\sum_{r=1}^{N_\mathrm{RF}} \left|\mathbf{h}_{n,1,r} \mathbf{U}_{\mathrm{RF},r} \tilde{\mathbf{f}}_{\mathrm{RF},r}\right|^2}}, \tag{2.37}$$

which is the well-known MRT precoder for the equivalent channel $\mathbf{h}_{n,1}\mathbf{U}_\mathrm{RF}\tilde{\mathbf{F}}_\mathrm{RF}$. Note that entries of $\mathbf{U}_{\mathrm{RF},r}\tilde{\mathbf{f}}_{\mathrm{RF},r}$ are not likely to have unit modulus. Moreover, approximating $\mathbf{U}_{\mathrm{RF},r}\tilde{\mathbf{f}}_{\mathrm{RF},r}$ by

$\exp\left(j\angle\left(\mathbf{U}_{\mathrm{RF},r}\tilde{\mathbf{f}}_{\mathrm{RF},r}\right)\right)$ results in losing the null space property. Therefore, TD-P-MRT is not applicable for the subarray structure S2. The computational complexity of TD-P-MRT is $\mathcal{O}\left(N_\mathrm{T}^3 + N_\mathrm{T}N_\mathrm{RF}N_\mathrm{C}\right)$.

FD-P-MRT For the four structures, it is necessary that $N_\mathrm{RF} \geq K$ to apply FD-P-MRT. We express the baseband precoder $\mathbf{f}_{\mathrm{BB},n}$ as

$$\mathbf{f}_{\mathrm{BB},n} = \mathbf{U}_{\mathrm{BB},n}\tilde{\mathbf{f}}_{\mathrm{BB},n}, \tag{2.38}$$

where $\mathbf{U}_{\mathrm{BB},n} = \mathcal{N}\left[\mathbf{H}_n\mathbf{F}_\mathrm{RF}\right] \in \mathbb{C}^{N_\mathrm{RF} \times (N_\mathrm{RF}-(K-1))}$ is a unitary matrix in the null space of the equivalent frequency domain channels to Eves $\mathbf{H}_n\mathbf{F}_\mathrm{RF}$. We design $\mathbf{F}_\mathrm{RF}$ and $\tilde{\mathbf{f}}_{\mathrm{BB},n}$ to maximize $\overline{\mathrm{SNR}}_\mathrm{B}$ given by

$$\overline{\mathrm{SNR}}_\mathrm{B} = \frac{\gamma}{N_\mathrm{C}}\sum_{n=1}^{N_\mathrm{C}} \frac{\left|\mathbf{h}_{n,1}\mathbf{F}_\mathrm{RF}\mathbf{U}_{\mathrm{BB},n}\tilde{\mathbf{f}}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_\mathrm{RF}\mathbf{U}_{\mathrm{BB},n}\tilde{\mathbf{f}}_{\mathrm{BB},n}\right\|^2} = \frac{\gamma}{N_\mathrm{C}}\sum_{n=1}^{N_\mathrm{C}} \frac{\tilde{\mathbf{f}}_{\mathrm{BB},n}^H\left(\mathbf{U}_{\mathrm{BB},n}^H\mathbf{F}_\mathrm{RF}^H\mathbf{h}_{n,1}^H\mathbf{h}_{n,1}\mathbf{F}_\mathrm{RF}\mathbf{U}_{\mathrm{BB},n}\right)\tilde{\mathbf{f}}_{\mathrm{BB},n}}{\tilde{\mathbf{f}}_{\mathrm{BB},n}^H\left(\mathbf{U}_{\mathrm{BB},n}^H\mathbf{F}_\mathrm{RF}^H\mathbf{F}_\mathrm{RF}\mathbf{U}_{\mathrm{BB},n}\right)\tilde{\mathbf{f}}_{\mathrm{BB},n}}. \tag{2.39}$$

Using generalized eigenvector decomposition, $\tilde{\mathbf{f}}_{\mathrm{BB},n}$ maximizing (2.39) is given by

$$
\begin{aligned}
\tilde{\mathbf{f}}_{\mathrm{BB},n} &= \kappa_n \boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right)^{-1} \mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right] \\
&= \frac{\left(\mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right)^{-1} \mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H}{\left\| \left(\mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right)^{-\frac{1}{2}} \mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \right\|},
\end{aligned}
\tag{2.40}
$$

where κ_n is a scaling factor to have $\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\|^2 = 1$, the second equality holds since $\left(\mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right)^{-1} \mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n}$ is a rank-one matrix, and the corresponding $\overline{\mathrm{SNR}}_{\mathrm{B}}$ is expressed as

$$
\begin{aligned}
\overline{\mathrm{SNR}}_{\mathrm{B}} &= \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \lambda_{\max} \left[\left(\mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right)^{-1} \mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right] \\
&= \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \left(\mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} \mathbf{U}_{\mathrm{BB},n} \right)^{-1} \mathbf{U}_{\mathrm{BB},n}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H.
\end{aligned}
\tag{2.41}
$$

Note that $\mathbf{U}_{\mathrm{BB},n}$ is a function of $\mathbf{F}_{\mathrm{RF}}$, and $\mathbf{F}_{\mathrm{RF}}$ is fixed across subcarriers while $\mathbf{U}_{\mathrm{BB},n}$ is not. Even if we fix $\mathbf{U}_{\mathrm{BB},n}$, we cannot get $\mathbf{F}_{\mathrm{RF}}$ in a closed-form. Consequently, we cannot get $\mathbf{F}_{\mathrm{RF}}$ in a closed-form or by an alternately optimizing algorithm. As suboptimal solutions, we obtain $\mathbf{F}_{\mathrm{RF}}$ for the fully-connected structures F1 and F2 as in (2.19) and (2.21) respectively, and for the subarray structures S1 and S2 as in (2.23) and (2.24) respectively. Then, $\mathbf{f}_{\mathrm{BB},n}$ is obtained as in (2.38). The computational complexity of FD-P-MRT is $\mathcal{O}\left(N_{\mathrm{T}}^3 N_{\mathrm{C}} + N_{\mathrm{RF}}^3 N_{\mathrm{C}} \right)$.

2.3.3   Iterative Secrecy Hybrid Precoding Design

The drawback of the aforementioned solutions is that they do not directly consider the original problem in (2.11). In this subsection, we propose an iterative hybrid precoding design to maximize the secrecy rate lower bound $\tilde{R}_{\mathrm{sec}}$, which gives a near-optimal solution to the problem in (2.11). The reason why we choose $\tilde{R}_{\mathrm{sec}}$ to maximize is that we can write $\tilde{R}_{\mathrm{sec}}$ as a function of $\mathbf{F}_{\mathrm{RF}}$ only (as will be shown).

Similar to (2.13), $\tilde{R}_{\mathrm{sec}}$ can be written as a function of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{f}_{\mathrm{BB},n}$ as

$$
\tilde{R}_{\mathrm{sec}} = \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(\frac{1 + \gamma \left| \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right|^2}{1 + \gamma \left\| \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right\|^2} \right).
\tag{2.42}
$$

Applying the power constraint $\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 = 1$ into (2.42), we get

$$\tilde{R}_{\mathrm{sec}} = \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(\frac{\mathbf{f}_{\mathrm{BB},n}^H \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \mathbf{f}_{\mathrm{BB},n}}{\mathbf{f}_{\mathrm{BB},n}^H \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right) \mathbf{f}_{\mathrm{BB},n}} \right). \tag{2.43}$$

Using generalized eigenvector decomposition, $\mathbf{f}_{\mathrm{BB},n}$ maximizing (2.43) is given by

$$\begin{aligned}
\mathbf{f}_{\mathrm{BB},n} &= \kappa_n \boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right)^{-1} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \right] \\
&= \frac{\boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right)^{-1} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \right]}{\left\| \mathbf{F}_{\mathrm{RF}} \, \boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right)^{-1} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \right] \right\|},
\end{aligned} \tag{2.44}$$

where κ_n is a scaling factor to have $\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 = 1$, and the corresponding $\tilde{R}_{\mathrm{sec}}$ is expressed as

$$\tilde{R}_{\mathrm{sec}} = \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(\lambda_{\max} \left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right)^{-1} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \right] \right). \tag{2.45}$$

Now, we can write the hybrid precoding design problem as a function of $\mathbf{F}_{\mathrm{RF}}$ only as

$$\begin{aligned}
\mathbf{F}_{\mathrm{RF}} = \ & \arg\max_{\mathbf{F}_{\mathrm{RF}}} \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(\lambda_{\max} \left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right)^{-1} \right. \right. \\
& \qquad\qquad\qquad \left. \left. \times \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \right] \right), \\
& \text{s.t.} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}.
\end{aligned} \tag{2.46}$$

It is straightforward to show that $\tilde{R}_{\mathrm{sec}}$ is a non-convex function of $\mathbf{F}_{\mathrm{RF}}$. Moreover, the constraint is non-convex (except for $\mathcal{F}_{\mathrm{RF}} = \mathcal{F}_{\mathrm{RF}}^{\mathrm{F1}}$). We propose a suboptimal gradient ascent algorithm to design $\mathbf{F}_{\mathrm{RF}}$. However, the maximum eigenvalue function $\lambda_{\max}[\mathbf{X}]$ is non-differentiable. Since $\lambda_{\max}[\mathbf{X}] \geq \frac{1}{\mathrm{rank}[\mathbf{X}]} \mathrm{Tr}[\mathbf{X}]$, we have

$$\tilde{R}_{\mathrm{sec}} \geq \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \log_2 \left(\frac{1}{N_{\mathrm{RF}}} \mathrm{Tr} \left[\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}_n^H \mathbf{H}_n \mathbf{F}_{\mathrm{RF}} \right)^{-1} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} + \gamma \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \right) \right] \right). \tag{2.47}$$

The gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}$ of the lower bound in (2.47) is given by [101]

$$\nabla_{\mathbf{F}_{\mathrm{RF}}} = \frac{2}{N_{\mathrm{RF}}N_{\mathrm{C}}\ln 2} \sum_{n=1}^{N_{\mathrm{C}}} \frac{\left(\mathbf{I}_{N_{\mathrm{T}}} - \mathbf{C}_n\mathbf{F}_{\mathrm{RF}}\left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{C}_n\mathbf{F}_{\mathrm{RF}}\right)^{-1}\mathbf{F}_{\mathrm{RF}}^H\right)\mathbf{B}_n\mathbf{F}_{\mathrm{RF}}\left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{C}_n\mathbf{F}_{\mathrm{RF}}\right)^{-1}}{\log_2\left(\frac{1}{N_{\mathrm{RF}}}\mathrm{Tr}\left[\left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{C}_n\mathbf{F}_{\mathrm{RF}}\right)^{-1}\left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{B}_n\mathbf{F}_{\mathrm{RF}}\right)\right]\right)},$$

(2.48)

where $\mathbf{C}_n = \mathbf{I}_{N_{\mathrm{T}}} + \gamma\mathbf{H}_n^H\mathbf{H}_n$ and $\mathbf{B}_n = \mathbf{I}_{N_{\mathrm{T}}} + \gamma\mathbf{h}_{n,1}^H\mathbf{h}_{n,1}$. Using the gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}$ in (2.48), we obtain the RF precoder $\mathbf{F}_{\mathrm{RF}}$ by Algorithm 1, where P is the number of iterations, $\mathcal{M}(\mathbf{X}) =$ blkdiag $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_{\mathrm{RF}}}) \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{RF}}}$ and $\mathbf{x}_r = \mathbf{X}\left((r-1)\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}} + 1 : r\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}, r\right) \in \mathbb{C}^{\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}} \times 1}$ $\forall r \in \{1, 2, \ldots, N_{\mathrm{RF}}\}$. The step size α is obtained by a backtracking line search [21].

It is well-known that gradient ascent algorithm is highly affected by the initial solution since it is a local solver. As an initialization, we propose to use the hybrid precoder of App-FD or P-MRT, where we choose the scheme that achieves higher secrecy rate. The computational complexity of Algorithm 1 is $\mathcal{O}\left(N_{\mathrm{T}}^3 N_{\mathrm{C}} + N_{\mathrm{T}} N_{\mathrm{RF}} N_{\mathrm{C}} + \left(N_{\mathrm{RF}}^3 + N_{\mathrm{T}} N_{\mathrm{RF}}^2\right) N_{\mathrm{C}} P\right)$.

2.4   Hybrid Precoder Design for Secrecy Throughput Maximization with Partial Channel Knowledge

In this section, we design the hybrid precoder to maximize the secrecy throughput defined as $\eta_{\mathrm{sec}} = R_{\mathrm{sec}}(1 - \epsilon_{\mathrm{sec}})$ [143, 138], where ϵ_{sec} is the secrecy outage probability. We assume that Alice has full knowledge of the channel to Bob but has partial knowledge of the channels to Eves. Since the variation of the AoDs is slower than the variation of the channel gains [136, 105], Alice uses the previous estimate of the AoDs of the propagation paths to Eves as a partial channel knowledge for the next multiple OFDM symbols. Bob and Eves have full knowledge of their channels to Alice. Furthermore, we assume that Eves do not cooperate.

In the following, we derive a secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$, and hence we get a secrecy throughput lower bound $\tilde{\eta}_{\mathrm{sec}} = R_{\mathrm{sec}}(1 - \tilde{\epsilon}_{\mathrm{sec}})$. Then, we design the hybrid precoder to maximize the secrecy throughput lower bound $\tilde{\eta}_{\mathrm{sec}}$.

Algorithm 1 Iterative Hybrid precoder design for secrecy rate maximization

---

Initialization: Obtain $\mathbf{F}_{\mathrm{RF}}^{(0)}$ as in App-FD or P-MRT (choose the scheme that achieves higher secrecy rate), $p = 0$.

while $p \leq P$ (or any other appropriate stopping criterion)

    Calculate the gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}$ using (2.48).

    Updating rule:

        For fully-connected structure F1,

$$\alpha = \arg\max_{\alpha} \tilde{R}_{\mathrm{sec}} \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}.$$

        For fully-connected structure F2,

$$\alpha = \arg\max_{\alpha} \tilde{R}_{\mathrm{sec}} \left(\exp\left(j\angle \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right) \right) \right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \exp\left(j\angle \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right) \right).$$

        For subarray structure S1,

$$\alpha = \arg\max_{\alpha} \tilde{R}_{\mathrm{sec}} \left(\boldsymbol{\mathcal{M}} \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right) \right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \boldsymbol{\mathcal{M}} \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right).$$

        For subarray structure S2,

$$\alpha = \arg\max_{\alpha} \tilde{R}_{\mathrm{sec}} \left(\boldsymbol{\mathcal{M}} \left(\exp\left(j\angle \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right) \right) \right) \right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \boldsymbol{\mathcal{M}} \left(\exp\left(j\angle \left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)} \right) \right) \right).$$

    $p = p + 1$.

end while

Obtain the baseband precoders $\{\mathbf{f}_{\mathrm{BB},n}\}$ using (2.44).

2.4.1 Secrecy Outage Probability

The secrecy outage probability ϵ_{sec} is defined as [16, 14]

$$\epsilon_{\mathrm{sec}} = \mathbb{P} \left(\min_{k} \{R_1 - R_k\}_{k=2}^{K} < R_{\mathrm{sec}} \right). \tag{2.49}$$

It can be rewritten as

$$\epsilon_{\mathrm{sec}} = 1 - \mathbb{P} \left(\max_{k} \{R_k\}_{k=2}^{K} < R_{\mathrm{o}} \right) = 1 - \prod_{k=2}^{K} \mathbb{P} \left(R_k < R_{\mathrm{o}} \right), \tag{2.50}$$

where $R_{\mathrm{o}} = R_1 - R_{\mathrm{sec}}$. We derive a tight secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$. Applying the power constraint $\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 = 1$ into (2.8) and then Jensen's inequality, we get

$$R_k \leq \log_2 \left(1 + \gamma \frac{1}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \frac{|\mathbf{h}_{n,k}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}|^2}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2} \right). \tag{2.51}$$

Using the triangle inequality, we have

$$\frac{\left|\mathbf{h}_{n,k}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2} = \frac{\left|\sqrt{N_{\mathrm{T}}}\sum_{l=1}^{L}\alpha_{l,k}\mathbf{a}_{l,k}^H\omega_{n,l,k}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2} \le N_{\mathrm{T}}\sum_{l=1}^{L}\left|\alpha_{l,k}\right|^2\frac{\left|\mathbf{a}_{l,k}^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2}. \tag{2.52}$$

Applying (2.52) and (2.51) into (2.50), we get

$$\epsilon_{\mathrm{sec}} \le 1 - \prod_{k=2}^{K}\mathbb{P}\left(\sum_{l=1}^{L}\left|\alpha_{l,k}\right|^2\sum_{n=1}^{N_{\mathrm{C}}}\frac{\left|\mathbf{a}_{l,k}^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2} < \delta\right), \tag{2.53}$$

where $\delta = \frac{\left(2^{R_{\mathrm{o}}}-1\right)N_{\mathrm{C}}}{\gamma\,N_{\mathrm{T}}}$. The term $\sum_{l=1}^{L}\left|\alpha_{l,k}\right|^2\sum_{n=1}^{N_{\mathrm{C}}}\frac{\left|\mathbf{a}_{l,k}^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2}$ represents a sum of indepen-

dent exponentially-distributed random variables with mean values of $\left\{\rho_l\sum_{n=1}^{N_{\mathrm{C}}}\frac{\left|\mathbf{a}_{l,k}^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2}\right\}$,

which results in an L-stage hypo-exponentially distributed random variable with a cumula-

tive distribution function (CDF) given by [122]

$$\mathbb{P}\left(\sum_{l=1}^{L}\left|\alpha_{l,k}\right|^2\sum_{n=1}^{N_{\mathrm{C}}}\frac{\left|\mathbf{a}_{l,k}^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2} < \delta\right) = 1 - g_k\left(\mathbf{F}_{\mathrm{RF}},\left\{\mathbf{f}_{\mathrm{BB},n}\right\},R_{\mathrm{sec}}\right), \tag{2.54}$$

where $g_k\left(\mathbf{F}_{\mathrm{RF}},\left\{\mathbf{f}_{\mathrm{BB},n}\right\},R_{\mathrm{sec}}\right)$ is given by

$$g_k\left(\mathbf{F}_{\mathrm{RF}},\left\{\mathbf{f}_{\mathrm{BB},n}\right\},R_{\mathrm{sec}}\right) = \sum_{l=1}^{L}e^{-\frac{\delta}{r_{l,k}}}\prod_{n\neq l}\frac{r_{l,k}}{r_{l,k}-r_{n,k}}, \tag{2.55}$$

and $r_{l,k} = \rho_l\sum_{n=1}^{N_c}\frac{\left|\mathbf{a}_{l,k}^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2}$. Therefore, the secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$

is written as

$$\tilde{\epsilon}_{\mathrm{sec}}\left(\mathbf{F}_{\mathrm{RF}},\left\{\mathbf{f}_{\mathrm{BB},n}\right\},R_{\mathrm{sec}}\right) = 1 - \prod_{k=2}^{K}\left(1 - g_k\left(\mathbf{F}_{\mathrm{RF}},\left\{\mathbf{f}_{\mathrm{BB},n}\right\},R_{\mathrm{sec}}\right)\right), \tag{2.56}$$

and hence we get a secrecy throughput lower bound $\tilde{\eta}_{\mathrm{sec}} = R_{\mathrm{sec}}\left(1 - \tilde{\epsilon}_{\mathrm{sec}}\right)$.

2.4.2 Low-Complexity Secrecy Hybrid Precoding Designs

In this subsection, we investigate if the low-complexity secrecy hybrid precoding strategies

in subsection 2.3.2 are applicable in case of partial channel knowledge. Examining the

secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$ in (2.56) with $N_{\mathrm{RF}} = N_{\mathrm{T}}$ ($\mathbf{F}_{\mathrm{RF}} = \mathbf{I}_{N_{\mathrm{T}}}$), we

know that the fully digital precoders $\{\mathbf{f}_{\text{BB},n}\}$ have to be designed jointly and do not have closed-form expressions, in contrast to maximizing the secrecy rate in subsection 2.3.1 where the fully digital precoders are designed separately and have closed-form expressions. As a result, we cannot have sufficient conditions on the number of RF chains needed to realize the performance of the fully digital precoding in case of partial channel knowledge. Furthermore, it is not tractable to design the hybrid precoder by approximating the fully digital precoders in case of partial channel knowledge.

Since TD-P-MRT described in subsection 2.3.2 requires only the knowledge of the AoDs to Eves which is available at Alice in case of partial channel knowledge, TD-P-MRT can be applied also in case of partial channel knowledge with the same constraints on N_{RF}. On the contrary, FD-P-MRT cannot be applied since it requires the knowledge of frequency domain channels to Eves which is not available at Alice in case of partial channel knowledge.

2.4.3 Secrecy Hybrid Precoding Design

Following the derived secrecy outage probability upper bound $\tilde{\epsilon}_{\text{sec}}$ in (2.56), the hybrid precoder is designed to maximize the secrecy throughput lower bound $\tilde{\eta}_{\text{sec}}$ as

$$
\begin{aligned}
[\mathbf{F}_{\text{RF}}, \{\mathbf{f}_{\text{BB},n}\}, R_{\text{sec}}] &= \underset{\mathbf{F}_{\text{RF}}, \{\mathbf{f}_{\text{BB},n}\}, R_{\text{sec}}}{\arg\max} \quad R_{\text{sec}} \left(1 - \tilde{\epsilon}_{\text{sec}} \left(\mathbf{F}_{\text{RF}}, \{\mathbf{f}_{\text{BB},n}\}, R_{\text{sec}}\right)\right), \\
&\text{s.t.} \quad 0 \le R_{\text{sec}} \le R_{\text{MRT}}, \ \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}},
\end{aligned}
\tag{2.57}
$$

where R_{MRT} is the maximum rate of Bob achieved by the maximum ratio transmission (MRT) scheme in [99] where the hybrid precoder is designed to maximize the rate of Bob ignoring Eves, and the power constraint is removed since it is applied into the objective function. The optimization problem in (2.57) is non-convex, and cannot be solved directly. In the following, we will focus on the secrecy outage probability minimization problem written as

$$
\begin{aligned}
[\mathbf{F}_{\text{RF}}, \{\mathbf{f}_{\text{BB},n}\}] &= \underset{\mathbf{F}_{\text{RF}}, \{\mathbf{f}_{\text{BB},n}\}}{\arg\min} \quad \tilde{\epsilon}_{\text{sec}} \left(\mathbf{F}_{\text{RF}}, \{\mathbf{f}_{\text{BB},n}\}, R_{\text{sec}}\right), \\
&\text{s.t.} \quad \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}}.
\end{aligned}
\tag{2.58}
$$

Then, we will show how to use the secrecy outage probability minimization problem to maximize the secrecy throughput. It is straightforward to show that $\tilde{\epsilon}_{\mathrm{sec}}$ is a non-convex function of $\mathbf{F}_{\mathrm{RF}}$ and $\{\mathbf{f}_{\mathrm{BB},n}\}$. Next, we propose an alternating algorithm which designs the RF precoder and the baseband precoders iteratively to minimize the secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$.

RF Precoding Design

First, we fix the baseband precoders $\{\mathbf{f}_{\mathrm{BB},n}\}$ and optimize over the RF precoder $\mathbf{F}_{\mathrm{RF}}$ to minimize the secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$. We propose a suboptimal gradient descent algorithm to design the RF precoder. The gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}$ of the the secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$ with respect to the RF precoder is obtained (after mathematical manipulations) as

$$\nabla_{\mathbf{F}_{\mathrm{RF}}} = \sum_{k=2}^{K} \frac{1}{1-g_k} \sum_{l=1}^{L} \left(\frac{r_{l,k} \nabla_{\mathbf{F}_{\mathrm{RF}}}(\delta) - \delta \nabla_{\mathbf{F}_{\mathrm{RF}}}(r_{l,k})}{r_{l,k}^2} + \sum_{n \neq l} \frac{r_{n,k} \nabla_{\mathbf{F}_{\mathrm{RF}}}(r_{l,k}) - r_{l,k} \nabla_{\mathbf{F}_{\mathrm{RF}}}(r_{n,k})}{r_{l,k}^2 - r_{l,k} r_{n,k}} \right)$$
$$\times e^{-\frac{\delta}{r_{l,k}}} \prod_{n \neq l} \frac{r_{l,k}}{r_{l,k} - r_{n,k}}, \tag{2.59}$$

where

$$\nabla_{\mathbf{F}_{\mathrm{RF}}}(r_{l,k}) = \frac{\rho_l \left(\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 \mathbf{a}_{l,k}\mathbf{a}_{l,k}^H - \left|\mathbf{a}_{l,k}^H\mathbf{f}_{\mathrm{BB},n}\right|^2 \mathbf{I}_{N_{\mathrm{T}}} \right) \mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\mathbf{f}_{\mathrm{BB},n}^H}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^4}, \tag{2.60}$$

$$\nabla_{\mathbf{F}_{\mathrm{RF}}}(\delta) = \frac{2^{R_\circ} \left(\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^2 \mathbf{h}_{n,1}^H\mathbf{h}_{n,1} - \left|\mathbf{h}_{n,1}\mathbf{f}_{\mathrm{BB},n}\right|^2 \mathbf{I}_{N_{\mathrm{T}}} \right) \mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\mathbf{f}_{\mathrm{BB},n}^H}{N_{\mathrm{T}} \left(1 + \gamma \left|\mathbf{h}_{n,1}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2 \right) \|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|^4}. \tag{2.61}$$

Using the gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}$ in (2.59), we obtain the RF precoder $\mathbf{F}_{\mathrm{RF}}$ by Algorithm 2, where P_{RF} is the number of iterations, and the step size α is obtained by a backtracking line search.

Baseband Precoding Design

Now, we fix the RF precoder $\mathbf{F}_{\mathrm{RF}}$ and optimize over the baseband precoders $\{\mathbf{f}_{\mathrm{BB},n}\}$ to minimize the secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$. We propose a suboptimal gradient

Algorithm 2 RF precoding design for secrecy outage probability minimization

---

Input: $\mathbf{F}_{\mathrm{RF}}^{(0)}$, $\{\mathbf{f}_{\mathrm{BB},n}\}$, R_{sec}.

$p = 0$.

while $p \leq P_{\mathrm{RF}}$ (or any other appropriate stopping criterion)

    Calculate the gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}$ using (2.59)–(2.61).

    Updating rule:

        For fully-connected structure F1,

$$\alpha = \arg\min_{\alpha} \tilde{\epsilon}_{\mathrm{sec}}\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}, \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}\right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}.$$

        For fully-connected structure F2,

$$\alpha = \arg\min_{\alpha} \tilde{\epsilon}_{\mathrm{sec}}\left(\exp\left(j\angle\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right)\right), \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}\right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \exp\left(j\angle\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right)\right).$$

        For subarray structure S1,

$$\alpha = \arg\min_{\alpha} \tilde{\epsilon}_{\mathrm{sec}}\left(\mathcal{M}\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right), \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}\right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \mathcal{M}\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right).$$

        For subarray structure S2,

$$\alpha = \arg\min_{\alpha} \tilde{\epsilon}_{\mathrm{sec}}\left(\mathcal{M}\left(\exp\left(j\angle\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right)\right)\right), \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}\right),$$

$$\mathbf{F}_{\mathrm{RF}}^{(p+1)} = \mathcal{M}\left(\exp\left(j\angle\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha \nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right)\right)\right).$$

    $p = p + 1$.

end while

Output: $\mathbf{F}_{\mathrm{RF}}^{(P_{\mathrm{RF}}+1)}$

---

descent algorithm to design the baseband precoders. The gradient $\nabla_{\mathbf{f}_{\mathrm{BB},n}}$ of the the secrecy outage probability upper bound $\tilde{\epsilon}_{\mathrm{sec}}$ with respect to the baseband precoder is obtained (after mathematical manipulations) as

$$\nabla_{\mathbf{f}_{\mathrm{BB},n}} = \sum_{k=2}^{K} \frac{1}{1-g_k} \sum_{l=1}^{L} \left(\frac{r_{l,k} \nabla_{\mathbf{f}_{\mathrm{BB},n}}(\delta) - \delta \nabla_{\mathbf{f}_{\mathrm{BB},n}}(r_{l,k})}{r_{l,k}^2} + \sum_{n \neq l} \frac{r_{n,k} \nabla_{\mathbf{f}_{\mathrm{BB},n}}(r_{l,k}) - r_{l,k} \nabla_{\mathbf{f}_{\mathrm{BB},n}}(r_{n,k})}{r_{l,k}^2 - r_{l,k} r_{n,k}} \right)$$

$$\times e^{-\frac{\delta}{r_{l,k}}} \prod_{n \neq l} \frac{r_{l,k}}{r_{l,k} - r_{n,k}}, \tag{2.62}$$

where

$$\nabla_{\mathbf{f}_{\mathrm{BB},n}}(r_{l,k}) = \frac{\rho_l \left(\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\|^2 \mathbf{F}_{\mathrm{RF}}^H \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H - \left| \mathbf{a}_{l,k}^H \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right|^2 \mathbf{F}_{\mathrm{RF}}^H \right) \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}}{\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\|^4}, \tag{2.63}$$

$$\nabla_{\mathbf{f}_{\mathrm{BB},n}}(\delta) = \frac{2^{R_{\mathrm{o}}} \left(\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\|^2 \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H \mathbf{h}_{n,1} - \left| \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right|^2 \mathbf{F}_{\mathrm{RF}}^H \right) \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}}{N_{\mathrm{T}} \left(1 + \gamma \left| \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n} \right|^2 \right) \|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\|^4}. \tag{2.64}$$

Algorithm 3 Baseband precoding design for secrecy outage probability minimization

---

Input: $\mathbf{F}_{\mathrm{RF}}, \left\{\mathbf{f}_{\mathrm{BB},n}^{(0)}\right\}, R_{\mathrm{sec}}$.

$p = 0$.

while $p \leq P_{\mathrm{BB}}$ (or any other appropriate stopping criterion)

    Calculate the gradient $\left\{\nabla_{\mathbf{f}_{\mathrm{BB},n}}^{(p)}\right\}$ using (2.62)–(2.64).

    Updating rule:

$$\alpha = \arg\min_{\alpha} \tilde{\epsilon}_{\mathrm{sec}} \left(\mathbf{F}_{\mathrm{RF}}, \left\{\mathbf{f}_{\mathrm{BB},n}^{(p)} + \alpha \nabla_{\mathbf{f}_{\mathrm{BB},n}}^{(p)}, \right\}, R_{\mathrm{sec}}\right),$$

$$\left\{\mathbf{f}_{\mathrm{BB},n}^{(p+1)}\right\} = \left\{\mathbf{f}_{\mathrm{BB},n}^{(p)} + \alpha \nabla_{\mathbf{f}_{\mathrm{BB},n}}^{(p)}\right\}.$$

    $p = p + 1$.

end while

Output: $\left\{\mathbf{f}_{\mathrm{BB},n}^{(P_{\mathrm{BB}}+1)}\right\}$

Using the gradient $\nabla_{\mathbf{f}_{\mathrm{BB},n}}$ in (2.62), we obtain the baseband precoders $\{\mathbf{f}_{\mathrm{BB},n}\}$ by Algorithm 3, where P_{BB} is the number of iterations, and the step size α is obtained by a backtracking line search.

Initial Hybrid Precoder

Generally, the average SNR of Bob $\overline{\mathrm{SNR}}_{\mathrm{B}}$ is expressed as $\overline{\mathrm{SNR}}_{\mathrm{B}} = \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \frac{\left|\mathbf{h}_{n,1}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right|^2}{\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\right\|^2}$. For any RF precoder $\mathbf{F}_{\mathrm{RF}}$, the baseband precoder $\mathbf{f}_{\mathrm{BB},n}$ which maximizes $\overline{\mathrm{SNR}}_{\mathrm{B}}$ is obtained as

$$\mathbf{f}_{\mathrm{BB},n} = \frac{\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-1} \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H}{\left\|\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H\right\|}, \tag{2.65}$$

and the corresponding $\overline{\mathrm{SNR}}_{\mathrm{B}} = \frac{\gamma}{N_{\mathrm{C}}} \sum_{n=1}^{N_{\mathrm{C}}} \mathbf{h}_{n,1} \mathbf{F}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-1} \mathbf{F}_{\mathrm{RF}}^H \mathbf{h}_{n,1}^H$. As an initial solution, we design the RF precoder $\mathbf{F}_{\mathrm{RF}}$ to maximize the average received SNR of Bob in the null space of the $K-1$ expected strongest directions to Eves.

For the fully-connected structure F1, we express $\mathbf{F}_{\mathrm{RF}}$ as

$$\mathbf{F}_{\mathrm{RF}} = \mathbf{U}_{\mathrm{RF}} \tilde{\mathbf{F}}_{\mathrm{RF}}, \tag{2.66}$$

where $\mathbf{U}_{\mathrm{RF}} = \mathcal{N}\left[\boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L} \rho_l \mathbf{a}_{l,2} \mathbf{a}_{l,2}^H\right], \boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L} \rho_l \mathbf{a}_{l,3} \mathbf{a}_{l,3}^H\right], \ldots, \boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L} \rho_l \mathbf{a}_{l,K} \mathbf{a}_{l,K}^H\right]\right]$ $\in \mathbb{C}^{N_{\mathrm{T}} \times (N_{\mathrm{T}}-(K-1))}$ is a unitary matrix in the null space of the $K-1$ expected strongest

directions to Eves. Similar to (2.31), $\tilde{\mathbf{F}}_{\mathrm{RF}}$ which maximizes $\overline{\mathrm{SNR}}_{\mathrm{B}}$ is obtained as $\tilde{\mathbf{F}}_{\mathrm{RF}} = \boldsymbol{\mathcal{E}}_{1:\mathrm{N_{RF}}}\left[\sum_{n=1}^{N_{\mathrm{C}}} \mathbf{U}_{\mathrm{RF}}^{H}\mathbf{h}_{n,1}^{H}\mathbf{h}_{n,1}\mathbf{U}_{\mathrm{RF}}\right]$. For the fully-connected structure F2, we obtain $\mathbf{F}_{\mathrm{RF}}$ as

$$\mathbf{F}_{\mathrm{RF}} = \exp\left(j\angle\mathbf{U}_{\mathrm{RF}}\tilde{\mathbf{F}}_{\mathrm{RF}}\right), \tag{2.67}$$

which satisfies the unit modulus constraint and is a good approximation to (2.66).

For the subarray structure S1, we express $\mathbf{f}_{\mathrm{RF},r}$ as

$$\mathbf{f}_{\mathrm{RF},r} = \mathbf{U}_{\mathrm{RF},r}\tilde{\mathbf{f}}_{\mathrm{RF},r}, \tag{2.68}$$

where $\mathbf{U}_{\mathrm{RF},r} = \mathcal{N}\left[\boldsymbol{\mathcal{E}}\left[\sum_{l=1}^{L}\rho_l\mathbf{a}_{l,2,r}\mathbf{a}_{l,2,r}^{H}\right], \boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L}\rho_l\mathbf{a}_{l,3,r}\mathbf{a}_{l,3,r}^{H}\right], \ldots, \boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L}\rho_l\mathbf{a}_{l,K,r}\mathbf{a}_{l,K,r}^{H}\right]\right]$ $\in \mathbb{C}^{\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}\times\left(\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}-(K-1)\right)}$ is a unitary matrix in the null space of the $K-1$ expected strongest directions to Eves seen by the rth RF chain, and $\mathbf{a}_{l,k,r} = \mathbf{a}_{l,k}\left((r-1)\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}+1:r\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}\right)$. Similar to (2.36), $\tilde{\mathbf{f}}_{\mathrm{RF},r}$ which maximizes $\overline{\mathrm{SNR}}_{\mathrm{B}}$ is obtained as $\tilde{\mathbf{f}}_{\mathrm{RF},r} = \boldsymbol{\mathcal{E}}_{\max}\left[\sum_{n=1}^{N_{\mathrm{C}}}\mathbf{U}_{\mathrm{RF},r}^{H}\mathbf{h}_{n,1,r}^{H}\mathbf{h}_{n,1,r}\mathbf{U}_{\mathrm{RF},r}\right]$. For the subarray structure S2, we obtain $\mathbf{f}_{\mathrm{RF},r}$ as

$$\mathbf{f}_{\mathrm{RF},r} = \exp\left(j\angle\mathbf{U}_{\mathrm{RF},r}\tilde{\mathbf{f}}_{\mathrm{RF},r}\right), \tag{2.69}$$

which satisfies the unit modulus constraint and is a good approximation to (2.68). The initial RF precoders in (2.66) and (2.67) require $N_{\mathrm{T}} \geq K$, while the ones in (2.68) and (2.69) require $N_{\mathrm{T}} \geq KN_{\mathrm{RF}}$. These two conditions are likely to be satisfied for large-scale mmWave systems.

To solve the secrecy throughput maximization problem in (2.57), we convert it into a sequence of secrecy outage probability minimization problems, each one is solved (as illustrated above) for a fixed target secrecy rate R_{sec}. The secrecy rate R_{sec} which maximizes the secrecy throughput is obtained by one dimensional search. The detailed algorithm is shown in Algorithm 4, where P is the number of iterations. Note that for fully digital precoding, the throughput maximization problem can also be solved using Algorithm 4 after excluding the RF precoder design step. The computational complexity of Algorithm 4 is $\mathcal{O}\left(N_{\mathrm{T}}^{3} + N_{\mathrm{T}}N_{\mathrm{RF}}N_{\mathrm{C}} + \max\left(N_{\mathrm{C}}, KL\right)\left(N_{\mathrm{RF}}N_{\mathrm{T}}P_{\mathrm{BB}} + N_{\mathrm{T}}^{2}P_{\mathrm{RF}}\right)P\right)$.

Algorithm 4 Hybrid precoding design for secrecy throughput maximization

Initialization: Obtain $\mathbf{F}_{\mathrm{RF}}^{(0)}$ as in (2.66), (2.67), (2.68), or (2.69) according to the used structure, then $\left\{\mathbf{f}_{\mathrm{BB},n}^{(0)}\right\}$ as in (2.65), $p = 0$.

$R_{\mathrm{sec}}^{(0)} = \underset{R_{\mathrm{sec}}}{\arg\max} \, R_{\mathrm{sec}} \left(1 - \tilde{\epsilon}_{\mathrm{sec}} \left(\mathbf{F}_{\mathrm{RF}}^{(0)}, \left\{\mathbf{f}_{\mathrm{BB},n}^{(0)}\right\}, R_{\mathrm{sec}}\right)\right)$, s.t. $0 \le R_{\mathrm{sec}} \le R_{\mathrm{MRT}}$.

while $p \le P$ (or any other appropriate stopping criterion)

$\quad \mathbf{F}_{\mathrm{RF}}^{(p+1)} = \underset{\mathbf{F}_{\mathrm{RF}}}{\arg\min} \, \tilde{\epsilon}_{\mathrm{sec}} \left(\mathbf{F}_{\mathrm{RF}}, \left\{\mathbf{f}_{\mathrm{BB},n}^{(p)}\right\}, R_{\mathrm{sec}}^{(p)}\right)$, using Algorithm 2 with $\mathbf{F}_{\mathrm{RF}}^{(p)}$ as an initial solution.

$\quad \left\{\mathbf{f}_{\mathrm{BB},n}^{(p+1)}\right\} = \underset{\{\mathbf{f}_{\mathrm{BB},n}\}}{\arg\min} \, \tilde{\epsilon}_{\mathrm{sec}} \left(\mathbf{F}_{\mathrm{RF}}^{(p+1)}, \{\mathbf{f}_{\mathrm{BB},n}\}, R_{\mathrm{sec}}^{(p)}\right)$, using Algorithm 3 with $\left\{\mathbf{f}_{\mathrm{BB},n}^{(p)}\right\}$ as an initial solution.

$\quad R_{\mathrm{sec}}^{(p+1)} = \underset{R_{\mathrm{sec}}}{\arg\max} \, R_{\mathrm{sec}} \left(1 - \tilde{\epsilon}_{\mathrm{sec}} \left(\mathbf{F}_{\mathrm{RF}}^{(p+1)}, \left\{\mathbf{f}_{\mathrm{BB},n}^{(p+1)}\right\}, R_{\mathrm{sec}}\right)\right)$, s.t. $R_{\mathrm{sec}}^{(p)} \le R_{\mathrm{sec}} \le R_{\mathrm{MRT}}$.

$\quad p = p + 1$.

end while

$\{\mathbf{f}_{\mathrm{BB},n}\} = \left\{\dfrac{\mathbf{f}_{\mathrm{BB},n}}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|}\right\}$

2.5 Simulation Results

In this section, we evaluate the performance of the proposed hybrid precoding designs and compare them with the performance of the fully digital precoding by means of Monte-Carlo simulations. Regarding the simulation setup, we assume that Alice has 64 antennas ($N_{\mathrm{T}} = 64$). The number of RF chains N_{RF} will be an adjustable parameter through the numerical results. All channels follow the mmWave channel model described in subsection 2.2.2 with 12 propagation paths ($L = 12$), where the channel gains $\{\alpha_{l,k}\}$ are complex Gaussian random variables with zero mean and exponential power delay profile defined as $\left\{\rho_l = \frac{q^{l-1}(1-q)}{(1-q^L)}\right\}$ where $q = 0.36$, and the angles of departure (AoDs) $\{\varphi_{l,k}\}$ are uniformly-distributed within $[0 \; 2\pi)$. The number of subcarriers N_{C} is 256.

If P-MRT is feasible, the optimal precoding strategy at high SNRs is to maximize the rate of Bob in the null space of channels to Eves, and hence we have $\lim_{\gamma \to \infty} \frac{R_{\mathrm{sec}}(\gamma)}{\log_2(\gamma)} = \lim_{\gamma \to \infty} \frac{R_1(\gamma)}{\log_2(\gamma)} = 1$ since Bob has a single antenna [128]. This result means that the secrecy rate should have a unit slope at high SNRs if P-MRT is feasible. Similarly, the secrecy throughput should have a unit slope if P-MRT is feasible. Table 2.1 summarizes the neces-

Table 2.1. Necessary conditions to apply P-MRT

Channel knwoledge	Fully Digital Precoding $(N_{\mathrm{RF}} = N_{\mathrm{T}})$	Hybrid Analog-Digital Precoding $(N_{\mathrm{RF}} < N_{\mathrm{T}})$	
		Structures F1, F2, and S1	Subarray structure S2
Full knowledge	$N_{\mathrm{T}} \geq K$ (FD-P-MRT)	$N_{\mathrm{T}} > (K-1)L$ (TD-P-MRT) or $N_{\mathrm{RF}} \geq K$ (FD-P-MRT)	$N_{\mathrm{RF}} \geq K$ (FD-P-MRT)
Partial knowledge	$N_{\mathrm{T}} > (K-1)L$ (TD-P-MRT)	$N_{\mathrm{T}} > (K-1)L$ (TD-P-MRT)	infeasible

sary conditions (mentioned in subsections 2.3.2 and 2.4.2) to apply P-MRT for fully digital precoding and hybrid precoding with full or partial channel knowledge. If P-MRT is infeasible, the secrecy rate and secrecy throughput will not have a unit slope at high SNRs.

2.5.1 Achievable Secrecy Rate with Full Channel Knowledge

Fig. 2.3 shows the achievable secrecy rate as a function of the transmit SNR with $N_{\mathrm{RF}} = 4$ and different numbers of Eves $(K-1)$. With two Eves (Fig. 2.3a), we observe that Algorithm 1 achieves a secrecy rate very close to that achieved by the fully digital precoding for the whole SNR range. The four RF precoder structures achieve slightly different secrecy rates due to the different hardware complexities. Algorithm1 approaches P-MRT at high SNRs since P-MRT is optimal at high SNRs. Note that FD-P-MRT is feasible for all structures since $N_{\mathrm{RF}} > K$. With four Eves (Fig. 2.3b), the secrecy rate gap between the fully digital precoding and Algorithm 1 increases as SNR increases. Algorithm 1 (F1, F2, and S1) approaches the corresponding P-MRT with unit secrecy slope at high SNRs, while Algorithm 1 (S2) does not achieve the unit secrecy slope. The reason is that FD-P-MRT is infeasible for all structures since $N_{\mathrm{RF}} < K$, while TD-P-MRT is feasible only for the structures F1, F2, and S1 since $N_{\mathrm{T}} > (K-1)L$.

For both Fig. 2.3a and Fig. 2.3b, we observe that App-FD achieve approximately the same secrecy rate as Algorithm 1 at low SNRs. However, App-FD does not achieve the unit secrecy slope at moderate and high SNRs, and the secrecy rate gap between Algorithm 1 and App-FD increases as SNR increases. The proposed adaptive hybrid precoder max(App-
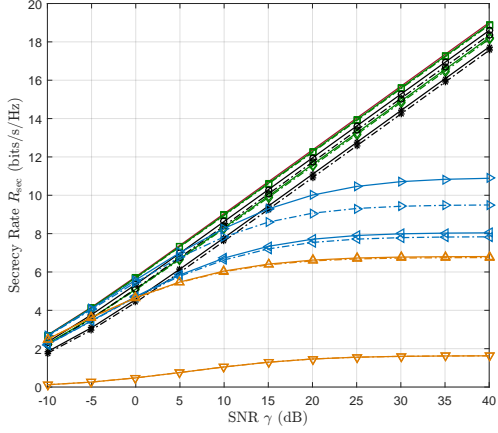
(a) Secrecy rate with 2 Eves ($K = 3$)	(b) Secrecy rate with 4 Eves ($K = 5$)
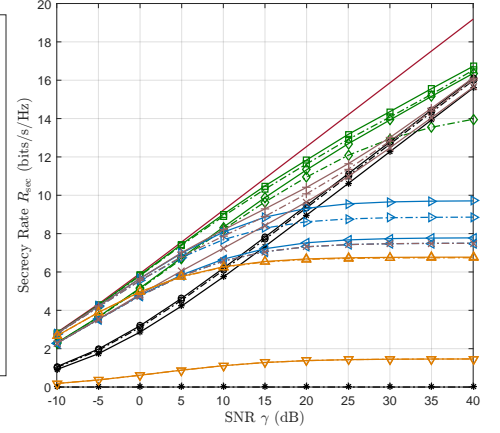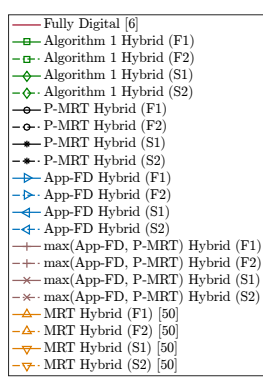
Figure 2.3. Achievable secrecy rate as a function of transmit SNR γ with $N_{\mathrm{RF}} = 4$ and different numbers of Eves ($K - 1$).

FD, P-MRT) in Fig. 2.3b combines the low-moderate SNR advantage of App-FD and the high SNR advantage of P-MRT, and yields better secrecy rate performance than App-FD and P-MRT with low computational complexity (both App-FD and P-MRT are obtained in closed forms. Note that max(App-FD, P-MRT) reduces to App-FD if P-MRT is infeasible. On the contrary, MRT of [99] achieves the worst secrecy rate at moderate and high SNRs due to ignoring Eves.

With different number of RF chains N_{RF}, Fig. 2.4 shows the achievable secrecy rate as a function of the transmit SNR with six Eves ($K = 7$). As expected, Algorithm 1 achieves the highest secrecy rate among the hybrid precoding designs. With $N_{\mathrm{RF}} = 4$ (Fig. 2.4a), the secrecy rate gap between the fully digital precoding and all hybrid precoding schemes increases as SNR increases. Algorithm 1 does not achieve the unit secrecy slope at high SNRs since P-MRT is infeasible for all structures. On the other hand, with $N_{\mathrm{RF}} = 8$ (Fig. 2.4b), FD-P-MRT is feasible for all structures. As a result, Algorithm 1 achieves a secrecy rate very close to that achieved by the fully digital precoding. From Fig. 2.4a and Fig. 2.4b, we can observe the significant effect of the number of RF chains on the achievable secrecy rate.

34
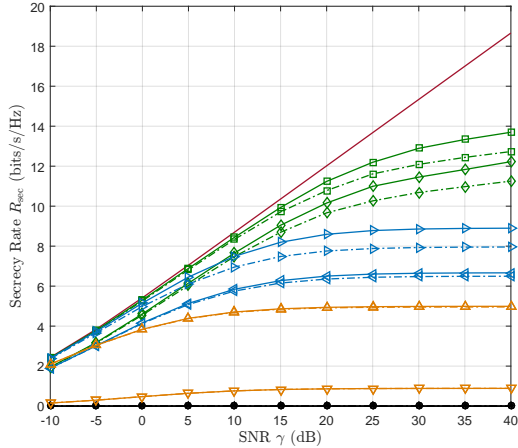
(a) Secrecy rate with $N_{\mathrm{RF}} = 4$ (b) Secrecy rate with $N_{\mathrm{RF}} = 8$
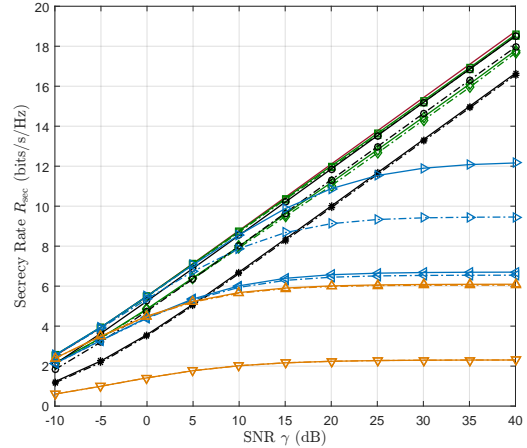
Figure 2.4. Achievable secrecy rate as a function of transmit SNR γ with 6 Eves ($K = 7$) and different numbers of RF chains N_{RF}.

2.5.2 Achievable Secrecy Throughput with Partial Channel Knowledge

Fig. 2.5 shows the achievable secrecy throughput as a function of the transmit SNR with $N_{\mathrm{RF}} = 4$ and different numbers of Eves ($K-1$). With four Eves (Fig. 2.5a), we observe that the hybrid precoding with Algorithm 4 (F1, F2, and S1) achieves a secrecy throughput very close to that of the fully digital precoding with Algorithm 4, and they have a unit secrecy slope at high SNRs since TD-PMRT is feasible for the fully digital precoding and the hybrid precoding structures F1, F2, and S1 due to $N_{\mathrm{T}} > (K-1)L$. On the contrary, Algorithm 4 (S2) does not achieve the unit secrecy slope since TD-P-MRT is not feasible for the subarray structure S2. With six Eves (Fig. 2.5b), the secrecy throughput gap between the fully digital precoding with Algorithm 4 and the hybrid precoding with Algorithm 4 increases as SNR increases. However, the hybrid precoding with Algorithm 4 achieves good performance at low and moderate SNRs. All precoding schemes (including the fully digital precoding) do not achieve the unit secrecy slope since TD-P-MRT is infeasible. Similarly, MRT of [99] achieves the worst secrecy throughput due to ignoring Eves. Note that increasing the number of RF chains (even if $N_{\mathrm{RF}} = N_T$) will not achieve the unit secrecy slope at high SNRs since

35
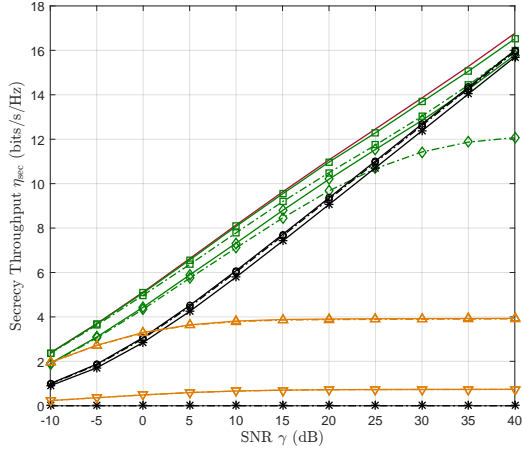
(a) Secrecy throughput with 4 Eves ($K = 5$)   (b) Secrecy throughput with 6 Eves ($K = 7$)
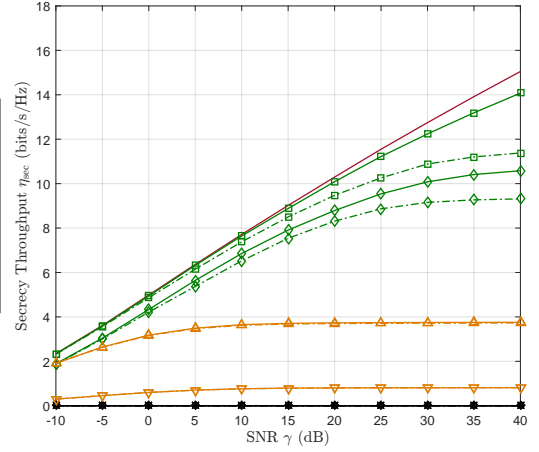
Figure 2.5. Achievable secrecy throughput as a function of transmit SNR γ with $N_{\mathrm{RF}} = 4$ and different numbers of Eves ($K - 1$).

FD-P-MRT is infeasible in case of partial channel knowledge, and TD-P-MRT is infeasible due to $N_{\mathrm{T}} < (K - 1)L$.

2.5.3 Tightness of Secrecy Rate and Throughput Lower Bounds

Fig. 2.6 shows an example for the tightness of the secrecy rate lower bound $\tilde{R}_{\mathrm{sec}}$ and the secrecy throughput lower bound $\tilde{\eta}_{\mathrm{sec}}$ with $N_{\mathrm{RF}} = 4$ and 4 Eves ($K = 5$). The sold lines are for the exact values while the dotted lines are for the lower bound values. We observe from Fig. 2.6a that the secrecy rate lower bound is tight for the structures F1, F2, and S1, while the secrecy rate lower bound predicts the performance behavior of the structure S2 efficiently. From Fig. 2.6b, we observe that the proposed secrecy throughput lower bound is very tight for all structures. The difference between the exact values and the lower bound values over the whole SNR range and all schemes is less than 0.0798 bits/s/Hz.

The reason of the tightness can be explained as follows. When the hybrid precoder is well designed (as the proposed schemes do) with $N_{\mathrm{T}} \gg K$ (large-scale mmWave systems) and small L (limited scattering mmWave channels), the average receive SNR of Eves will be
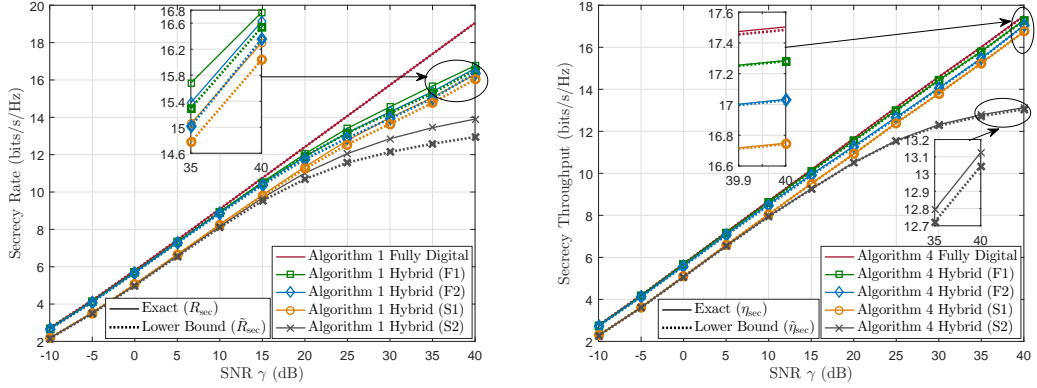
(a) Tightness of secrecy rate lower bound $\tilde{R}_{\text{sec}}$ (b) Tightness of secrecy throughput lower bound $\tilde{\eta}_{\text{sec}}$

Figure 2.6. Tightness of secrecy rate and throughput lower bounds with $N_{\text{RF}} = 4$, and 4 Eves ($K = 5$).

very small compared to the average receive SNR of Bob [144] (due to the capability of generating very sharp beams avoiding, as much as possible, the directions to Eves). Therefore, considering Eves as one Eve with $K - 1$ antennas (the approximation of section 3) will not decrease the secrecy rate significantly. Similarly, applying the inequalities of section 4 will not increase the rates of Eves significantly.

2.5.4 Convergence of Algorithms 1 and 4 and Effect of Finite Resolution Phase Shifters

Fig. 2.7 shows an example for the convergence of Algorithm 1 and Algorithm 4 with $N_{\text{RF}} = 4$, 4 Eves ($K = 5$), and transmit SNR $\gamma = 15$ dB. The secrecy rate and throughput achieved by Algorithms 1 and 4 after each iteration are plotted in Fig. 2.7a and Fig. 2.7b respectively. We observe that both Algorithms 1 and 4 converge in a small number of iterations, where 10 iterations are sufficient for all RF precoder structures.

With finite resolution phase shifters, Fig. 2.8 shows the secrecy rate and throughput achieved by Algorithms 1 and 4 versus different numbers of quantization bits for the phase shifters with $N_{\text{RF}} = 4$, 4 Eves ($K = 5$), and transmit SNR $\gamma = 15$ dB. After designing the hybrid precoder using Algorithm 1 or Algorithm 4, the phases of the RF precoder are
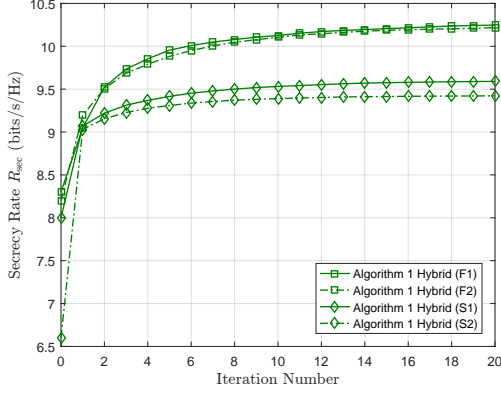
(a) Convergence of Algorithm 1
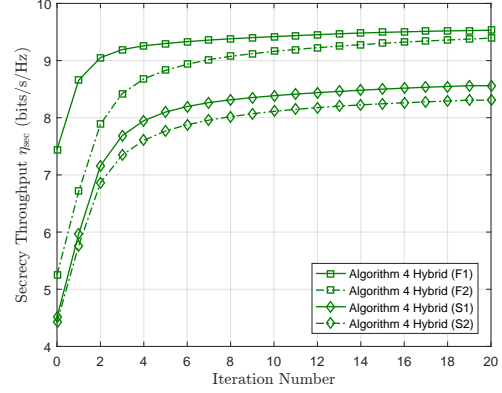
(b) Convergence of Algorithm 4

Figure 2.7. Convergence of Algorithm 1 and Algorithm 4 with $N_{\mathrm{RF}} = 4$, 4 Eves ($K = 5$), and transmit SNR $\gamma = 15$ dB.
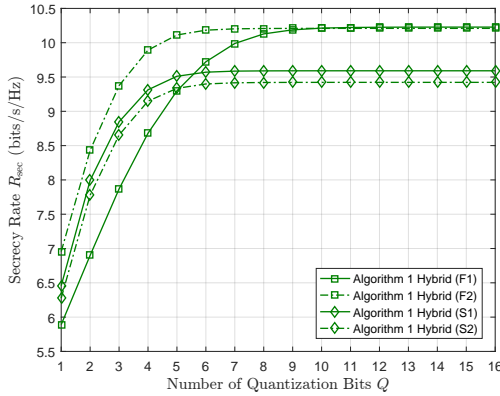
(a) Secrecy rate achieved by Algorithm 1 versus number of quantization bits
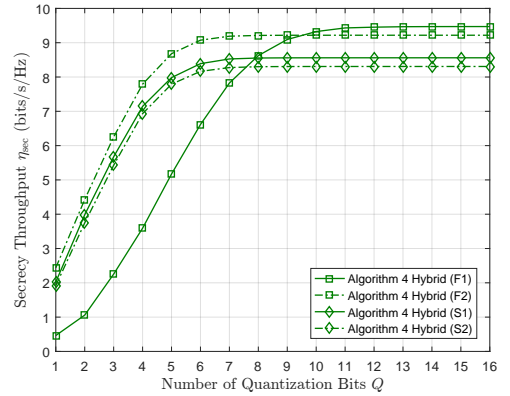
(b) Secrecy throughput achieved by Algorithm 4 versus number of quantization bits

Figure 2.8. Effect of finite resolution phase shifters on the secrecy rate and throughput achieved by Algorithm 1 and Algorithm 4 with $N_{\mathrm{RF}} = 4$, 4 Eves ($K = 5$), and transmit SNR $\gamma = 15$ dB.

quantized into Q bits such that $\angle \mathbf{F}_{\mathrm{RF}}(i,j) \in \left\{ 0, \frac{2\pi}{2^Q}, \ldots, \frac{2\pi 2^{Q-1}}{2^Q} \right\} \, \forall \mathbf{F}_{\mathrm{RF}}(i,j) \neq 0$. We observe that 6 quantization bits are sufficient for the RF precoder structures F2, S1, and S2 with secrecy rate/throughput loss less than 0.2 bits/s/Hz. On the other hand, the RF precoder structure F1 requires at least 10 quantization bits to outperform the structure F2. The reason is that the structure F1 uses approximately twice the number of phase shifters of the structure F2. As a result, the quantization error of the structure F1 is larger than that of the remaining structures.

2.6 Conclusion

In this chapter, we have designed the hybrid analog-digital precoder for physical layer security. With full channel knowledge at the transmitter, we provided sufficient conditions for the hybrid precoder to realize the performance of the fully digital precoding. If the sufficient conditions are not satisfied, we design the hybrid precoder to maximize the secrecy rate. By maximizing the average projection between the fully digital precoder and the hybrid precoder, we proposed a low-complexity closed-form hybrid precoder design. The conventional P-MRT scheme is extended to realize the hybrid precoder. Two P-MRT schemes were presented. TD-P-MRT nulls the channels to Eves at time domain, and FD-P-MRT nulls the channels to Eves at frequency domain. Moreover, we proposed an iterative hybrid precoder design, based on gradient ascent, which converges in a small number of iterations and achieves secrecy rate close to that achieved by the fully digital precoding.

With partial channel knowledge at the transmitter, we derived a secrecy outage probability upper bound. The secrecy throughput maximization problem is converted into a sequence of secrecy outage probability minimization problems. Then, the hybrid precoder is designed to minimize the secrecy outage probability by an iterative hybrid precoder design, based on gradient descent, which converges in a small number of iterations and achieves secrecy throughput close to that achieved by the fully digital precoding. With finite resolution phase shifters, we showed that 6 quantization bits are sufficient for the structures F2, S1, and S2. On the contrary, 10 quantization bits are needed for the structure F1 to outperform the structure F2.

Appendix A - Proof of Proposition 1

As shown in (2.14), the optimal fully digital precoder $\mathbf{f}_{\mathrm{opt},n}$ is the principal generalized eigen-vector corresponding to the maximum eigenvalue of the pencil $\left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{H}_n^H \mathbf{H}_n, \; \mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{h}_{n,1}^H \mathbf{h}_{n,1}\right)$. Among the N_{T} generalized eigenvalues, $(N_{\mathrm{T}} - K)$ of them are equal to 1 and obtained using any vector that is orthogonal to the space spanned by $\left[\mathbf{h}_{n,1}^H, \mathbf{H}_n^H\right]$. The other K generalized eigenvectors corresponding to the other K eigenvalues (including the maximum eigenvalue) lie completely in the space spanned by $\left[\mathbf{h}_{n,1}^H, \mathbf{H}_n^H\right]$. Therefore, we can write $\mathbf{f}_{\mathrm{opt},n}$ as

$$\mathbf{f}_{\mathrm{opt},n} = \beta_n \frac{\mathbf{\Pi}_n \mathbf{h}_{n,1}^H}{\left\|\mathbf{\Pi}_n \mathbf{h}_{n,1}^H\right\|} + \sqrt{1 - \beta_n^2} \frac{\mathbf{\Pi}_n^\perp \mathbf{h}_{n,1}^H}{\left\|\mathbf{\Pi}_n^\perp \mathbf{h}_{n,1}^H\right\|}, \tag{2.70}$$

where $0 \leq \beta_n \leq 1$, $\beta_n = \frac{\left|\mathbf{h}_{n,1} \mathbf{\Pi}_n \boldsymbol{\mathcal{E}}_{\max}\left[\left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{H}_n^H \mathbf{H}_n\right)^{-1}\left(\mathbf{I}_{N_{\mathrm{T}}} + \gamma \mathbf{h}_{n,1}^H \mathbf{h}_{n,1}\right)\right]\right|}{\left\|\mathbf{\Pi}_n \mathbf{h}_{n,1}^H\right\|}$, $\mathbf{\Pi}_n = \mathbf{H}_n^H \left(\mathbf{H}_n \mathbf{H}_n^H\right)^{-1} \mathbf{H}_n \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ denotes the orthogonal projection onto the space spanned by $\mathbf{H}_n$, and $\mathbf{\Pi}_n^\perp = \mathbf{I}_{N_{\mathrm{T}}} - \mathbf{\Pi}_n$ denotes the projection onto its orthogonal complement. Equation (2.70) can be rewritten as

$$\mathbf{f}_{\mathrm{opt},n} = \mathbf{H}_{\mathrm{TD}} \mathbf{W}_n \mathbf{p}_n, \tag{2.71}$$

where $\mathbf{H}_{\mathrm{TD}} \in \mathbb{C}^{N_{\mathrm{T}} \times KL}$ is the time domain channel matrix to the K receivers given by

$$\mathbf{H}_{\mathrm{TD}} = \left[\tilde{\mathbf{h}}_{\tau_{1,1,1}}^H, \tilde{\mathbf{h}}_{\tau_{2,1,1}}^H, \ldots, \tilde{\mathbf{h}}_{\tau_{L,1,1}}^H, \ldots, \tilde{\mathbf{h}}_{\tau_{1,K,K}}^H, \tilde{\mathbf{h}}_{\tau_{2,K,K}}^H, \ldots, \tilde{\mathbf{h}}_{\tau_{L,K,K}}^H\right], \tag{2.72}$$

$\mathbf{W}_n = \mathrm{blkdiag}\left(\mathbf{w}_{n,1}^H, \mathbf{w}_{n,2}^H, \ldots, \mathbf{w}_{n,K}^H\right) \in \mathbb{C}^{KL \times K}$, and $\mathbf{p}_n = \left[\mu_n, \nu_n \mathbf{h}_{n,1} \mathbf{H}_n^H \left(\mathbf{H}_n \mathbf{H}_n^H\right)^{-1}\right]^H \mathbb{C}^{K \times 1}$, where $\mu_n = \frac{\sqrt{1 - \beta_n^2}}{\left\|\mathbf{\Pi}_n^\perp \mathbf{h}_{n,1}^H\right\|}$ and $\nu_n = \frac{\beta_n}{\left\|\mathbf{\Pi}_n \mathbf{h}_{n,1}^H\right\|} - \frac{\sqrt{1 - \beta_n^2}}{\left\|\mathbf{\Pi}_n^\perp \mathbf{h}_{n,1}^H\right\|}$. The performance of fully digital pre-coding can be realized by setting $\mathbf{F}_{\mathrm{RF}}\left(:, 1:KL\right) = \mathbf{H}_{\mathrm{TD}}$ and $\mathbf{f}_{\mathrm{BB},n}\left(1:KL\right) = \mathbf{W}_n \mathbf{p}_n$ if $N_{\mathrm{RF}} \geq KL$. Thus, to realize the performance of fully digital precoding, it is sufficient for the hybrid precoding utilizing the fully-connected structure F1 that $N_{\mathrm{RF}} \geq KL$. This completes the proof of the first statement.

For the fully-connected structure F2, we have to satisfy the unit modulus constraint. In [149], it was shown that any vector $\mathbf{x} \in \mathbb{C}^{N \times 1}$ can be expressed as $\mathbf{x} = \tilde{\mathbf{X}}\tilde{\mathbf{x}}$, where $\tilde{\mathbf{X}} \in \mathbb{C}^{N \times 2}$

is with unit modulus entries and $\tilde{\mathbf{x}} \in \mathbb{R}^{2 \times 1}$. Following this decomposition, $\mathbf{H}_{\mathrm{TD}}$ can be expressed as $\mathbf{H}_{\mathrm{TD}} = \mathbf{Q}_{\mathrm{RF}} \mathbf{R}_{\mathrm{BB}}$, where $\mathbf{Q}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}} \times 2KL}$ is with unit modulus entries obtained as

$$\mathbf{Q}_{\mathrm{RF}} (l, 2m-1) = \exp \left(j \left(\angle \mathbf{H}_{\mathrm{TD}} (l, m) - \cos^{-1} \left(\frac{|\mathbf{H}_{\mathrm{TD}} (l, m)|^2 + b_{\mathrm{max},m} b_{\mathrm{min},m}}{|\mathbf{H}_{\mathrm{TD}} (l, m)|^2 (b_{\mathrm{max},m} + b_{\mathrm{min},m})} \right) \right) \right),$$

(2.73)

$$\mathbf{Q}_{\mathrm{RF}} (l, 2m) = \exp \left(j \left(\angle \mathbf{H}_{\mathrm{TD}} (l, m) + \cos^{-1} \left(\frac{|\mathbf{H}_{\mathrm{TD}} (l, m)|^2 - b_{\mathrm{max},m} b_{\mathrm{min},m}}{|\mathbf{H}_{\mathrm{TD}} (l, m)|^2 (b_{\mathrm{max},m} - b_{\mathrm{min},m})} \right) \right) \right),$$

(2.74)

$\forall l \in \{1, 2, \ldots N_{\mathrm{T}}\}$ and $\forall m \in \{1, 2, \ldots KL\}$ where $b_{\mathrm{max},m} = \max_l |\mathbf{H}_{\mathrm{TD}} (l, m)|$, $b_{\mathrm{min},m} = \min_l |\mathbf{H}_{\mathrm{TD}} (l, m)|$, and $\mathbf{R}_{\mathrm{BB}} \in \mathbb{R}^{2KL \times KL}$ is obtained as $\mathbf{R}_{\mathrm{BB}} (2m-1, m) = (b_{\mathrm{max},m} + b_{\mathrm{min},m})/2$, $\mathbf{R}_{\mathrm{BB}} (2m, m) = (b_{\mathrm{max},m} - b_{\mathrm{min},m})/2 \, \forall m \in \{1, 2, \ldots KL\}$. Therefore, we can set $\mathbf{F}_{\mathrm{RF}} (:, 1 : 2KL) = \mathbf{Q}_{\mathrm{RF}}$ and $\mathbf{f}_{\mathrm{BB},n} (1 : 2KL) = \mathbf{R}_{\mathrm{BB}} \mathbf{W}_n \mathbf{p}_n$ if $N_{\mathrm{RF}} \geq 2KL$. As a result, it is sufficient that $N_{\mathrm{RF}} \geq 2KL$ for the hybrid precoding utilizing the fully-connected structure F2 to realize the performance of fully digital precoding. By assuming that all the channels follow the mmWave channel model in (2.2), (2.71) can be written as $\mathbf{f}_{\mathrm{opt},n} = \mathbf{A}_{\mathrm{T}} \mathbf{D}_{\mathrm{T}} \mathbf{W}_n \mathbf{p}_n$, where $\mathbf{A}_{\mathrm{T}} = [\mathbf{A}_{\mathrm{T},1}, \mathbf{A}_{\mathrm{T},2}, \ldots, \mathbf{A}_{\mathrm{T},K}] \in \mathbb{C}^{N_{\mathrm{T}} \times KL}$ and $\mathbf{D}_{\mathrm{T}} = \mathrm{blkdiag} \left(\mathbf{D}_1^H, \mathbf{D}_2^H, \ldots, \mathbf{D}_K^H \right) \in \mathbb{C}^{KL \times KL}$. Since $\mathbf{A}_{\mathrm{T}}$ is with unit modulus entries, the sufficient condition for the hybrid precoding utilizing the fully-connected structure F2 reduces to $N_{\mathrm{RF}} \geq KL$ by setting $\mathbf{F}_{\mathrm{RF}} (:, 1 : KL) = \mathbf{A}_{\mathrm{T}}$ and $\mathbf{f}_{\mathrm{BB},n} (1 : KL) = \mathbf{D}_{\mathrm{T}} \mathbf{W}_n \mathbf{p}_n$. This completes the proof of the second statement. Note that the obtained sufficient conditions vanish if $N_{\mathrm{T}} \leq KL$ since $N_{RF} < N_{\mathrm{T}}$.

Appendix B - Proof of Proposition 2

In Proposition 1, it was shown that if $N_{\mathrm{T}} \leq KL$, there is no sufficient condition depending only on the number of RF chains to realize the performance of fully digital precoding since $N_{RF} < N_{\mathrm{T}}$. Therefore, we consider the case that $KL \leq N_{\mathrm{RF}} < N_{\mathrm{T}}$. To realize the

performance of the fully digital precoding, $\mathbf{f}_{\mathrm{BB},n}$ has to be expressed as $\mathbf{f}_{\mathrm{BB},n} = \mathbf{B}_{\mathrm{BB}}\mathbf{W}_n\mathbf{p}_n$, where $\mathbf{B}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times KL}$ is a mapping matrix. We have to design $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{B}_{\mathrm{BB}}$ such that $\mathbf{H}_{\mathrm{TD}} = \mathbf{F}_{\mathrm{RF}}\mathbf{B}_{\mathrm{BB}}$. Let $N_{\mathrm{RF}} = N_{\mathrm{T}} - 1 \geq KL$ which means that each RF chain is connected to one antenna except only one RF chain that is connected to two antennas. Without loss of generality, let the first RF chain be the only RF chain that is connected to two antennas, we should have

$$\mathbf{H}_{\mathrm{TD}}(1,l) = \mathbf{f}_{\mathrm{RF},1}(1)\,\mathbf{B}_{\mathrm{BB}}(1,l), \tag{2.75}$$

$$\mathbf{H}_{\mathrm{TD}}(2,l) = \mathbf{f}_{\mathrm{RF},1}(2)\,\mathbf{B}_{\mathrm{BB}}(1,l), \tag{2.76}$$

where $l \in \{1, 2, \ldots, KL\}$. To satisfy (2.75) and (2.76), it is necessary that $\mathbf{H}_{\mathrm{TD}}(2,:) = c\,\mathbf{H}_{\mathrm{TD}}(1,:)$ where c is a constant, which is not guaranteed and depends on the channel realizations. As a result, the hybrid precoding utilizing the subarray structure S1 cannot generally realize the fully digital precoding for any $N_{\mathrm{RF}} \leq N_{\mathrm{T}} - 1$. Since $\mathcal{F}_{\mathrm{RF}}^{\mathrm{S2}} \subset \mathcal{F}_{\mathrm{RF}}^{\mathrm{S1}}$, we arrive at the same conclusion for the subarray structure S2. This completes the proof of Proposition 2.

Appendix C - Approximating The Fully Digital Precoding

For the average Euclidean distance criterion, the hybrid precoder is designed to approximate the fully digital precoding as [142, 94, 146]

$$[\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}] = \underset{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{f}_{\mathrm{BB},n}\}}{\arg\min} \sum_{n=1}^{N_{\mathrm{C}}} \left\| \mathbf{f}_{\mathrm{opt},n} - \frac{\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},n}\|} \right\|^2. \tag{2.77}$$

It is straightforward to show that for any $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{f}_{\mathrm{BB},n}$ is the least squares solution which can be expressed (after appropriate normalization) as

$$\mathbf{f}_{\mathrm{BB},n} = \frac{\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-1} \mathbf{F}_{\mathrm{RF}}^H \mathbf{f}_{\mathrm{opt},n}}{\left\| \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \mathbf{F}_{\mathrm{RF}}^H \mathbf{f}_{\mathrm{opt},n} \right\|}, \tag{2.78}$$

which is exactly the same as (2.18), and the corresponding $\sum_{n=1}^{N_C} \left\| \mathbf{f}_{\text{opt},n} - \frac{\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},n}}{\|\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},n}\|} \right\|^2 = 2N_C - 2\sum_{n=1}^{N_C} \sqrt{\mathbf{f}_{\text{opt},n}^H \mathbf{F}_{\text{RF}} \left(\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}}\right)^{-1} \mathbf{F}_{\text{RF}}^H \mathbf{f}_{\text{opt},n}}$. Thus, (2.77) can be written as a function of $\mathbf{F}_{\text{RF}}$ only as

$$\mathbf{F}_{\text{RF}} = \arg\max_{\mathbf{F}_{\text{RF}}} \sum_{n=1}^{N_C} \sqrt{\lambda_{\max}\left[\left(\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}}\right)^{-1} \mathbf{F}_{\text{RF}}^H \mathbf{f}_{\text{opt},n} \mathbf{f}_{\text{opt},n}^H \mathbf{F}_{\text{RF}}\right]}, \tag{2.79}$$

which is similar to the first equality of (2.19) except for the square root. For (2.19), we have a closed-

form solution of $\mathbf{F}_{\text{RF}} = \boldsymbol{\mathcal{E}}_{1:\text{N}_{\text{RF}}}\left[\sum_{n=1}^{N_C} \mathbf{f}_{\text{opt},n}^H \mathbf{f}_{\text{opt},n}\right]$, while there is no a closed-form solution for (2.79).

CHAPTER 3

SECURE AN-AIDED HYBRID PRECODING FOR MMWAVE MISO SYSTEMS WITH PARTIAL CHANNEL KNOWLEDGE [1]

3.1 Introduction

Physical layer security is typically introduced by precoding schemes at the transmitter (Alice). For millimeter-wave (mmWave) communications [109], some of existing works consider radio-frequency (RF) precoding with a single RF chain in line-of-sight (LOS) channels [130, 137, 154] or non-line-of-sight (NLOS) channels [107]. The works in [67, 135, 133] apply secure baseband precoding with full RF chains. For mmWave systems, hybrid analog-digital precoders are preferred due to the hardware complexity and power consumption concern [12]. Secure hybrid precoding schemes are developed in [108, 144, 153, 36]. The work in [108] relies on the beamforming strategy with no artificial noise (AN) which results in secrecy performance degradation at moderate and high signal-to-noise ratios (SNRs). The works in [144, 153, 36] develop an AN-aided hybrid precoder. However, the optimal power allocation between the confidential signal and AN was not considered in [144], but addressed in [153] when the number of RF chains is greater than the number of antennas of Eve for Rayleigh fading channels and in [36] for mmWave LOS channels. Moreover, [144] and [153] consider only the case that Alice does not have any knowledge of the channels to Eves. For cellular systems, base station (Alice) may know the directional information of other active users (Eves) and exploiting such partial channel knowledge to enhance secure communication is much desirable. To the best of our knowledge, secure AN-aided hybrid precoding for mmWave NLOS systems with partial channel knowledge has not been developed in the literature.

In this chapter, we propose an AN-aided hybrid precoder design to maximize the average secrecy rate with partial channel knowledge. We derive a closed-form expression for the average signal-to-interference-and-noise ratio (SINR) of Eve as a function of the hybrid precoder. Using the average SINRs of Eves, we obtain a secrecy rate lower bound. Since the hybrid precoder design problem is non-convex, we propose a suboptimal solution. Numerical results show that the proposed AN-aided hybrid precoder achieves comparable performance to that of the fully digital precoder, with much lower hardware complexity. Moreover, the proposed AN-aided hybrid precoder outperforms the maximum-ratio-transmission (MRT) hybrid precoder of [99].

3.2 System and Channel Models

3.2.1 System Model

We consider a secrecy mmWave MISO system with K single-antenna receivers. The transmitter (Alice) sends a confidential message to the first receiver (Bob), while the rest $K - 1$ receivers are eavesdroppers (Eves). We assume that the transmitter is equipped with a large-scale uniform linear array (ULA) with N_T antennas ($N_\mathrm{T} \gg K$). The spacing between antennas is half the wavelength. To reduce the hardware complexity and the power consumption, the antenna array is connected via an analog RF precoder to N_RF RF chains ($N_\mathrm{RF} < N_\mathrm{T}$) which process the digitally-precoded streams.

We consider a narrow-band transmission, where the received signal y_k at the kth receiver is given by

$$y_k = \mathbf{h}_k \mathbf{x} + n_k, \tag{3.1}$$

where $\mathbf{h}_k \in \mathbb{C}^{1 \times N_\mathrm{T}}$ is the mmWave channel to the kth receiver, $n_k \sim \mathcal{CN}\left(0, \sigma^2\right)$ is the additive white complex Gaussian noise with zero mean and variance σ^2 at the kth receiver, and $\mathbf{x} \in \mathbb{C}^{N_\mathrm{T} \times 1}$ is the transmit signal vector given by

$$\mathbf{x} = \sqrt{\phi}\, \mathbf{F}_\mathrm{RF} \mathbf{f}_\mathrm{BB} s + \sqrt{1 - \phi}\, \mathbf{F}_\mathrm{RF} \mathbf{U}_\mathrm{BB} \mathbf{z}, \tag{3.2}$$

where ϕ is the power fraction allocated to the confidential signal, $(1 - \phi)$ is the power fraction allocated to AN, $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{RF}}}$ is the analog RF precoder, $s \sim \mathcal{CN}(0, P)$ is the coded confidential signal, $\mathbf{f}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times 1}$ is the signal digital baseband precoder, $\mathbf{z} \sim \mathcal{CN}(0, P\mathbf{I}_{N_{\mathrm{RF}}})$ is the noise vector artificially generated by Alice, and $\mathbf{U}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{RF}}}$ is the AN digital baseband precoder. To maintain the power constraint $\mathbb{E}\{\mathbf{x}^H \mathbf{x}\} = P$, where P is the transmit power, we should have $\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}\|^2 = \|\mathbf{F}_{\mathrm{RF}}\mathbf{U}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = 1$.

Using (3.1) and (3.2), the achievable rate R_k of the kth receiver is given by

$$R_k = \log_2\left(1 + \frac{\phi\gamma\,|\mathbf{h}_k\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}|^2}{(1 - \phi)\gamma\,\|\mathbf{h}_k\mathbf{F}_{\mathrm{RF}}\mathbf{U}_{\mathrm{BB}}\|^2 + 1}\right), \tag{3.3}$$

where $\gamma = P/\sigma^2$ is the transmit signal-to-noise ratio (SNR). The average secrecy rate $\bar{R}_{\mathrm{sec}}$ is given by [116]

$$\bar{R}_{\mathrm{sec}} = \mathbb{E}\left\{R_1 - \max_k\{R_k\}_{k=2}^K\right\}, \tag{3.4}$$

where the expectation is performed over the channels to Bob and Eves.

The analog RF precoder $\mathbf{F}_{\mathrm{RF}}$ is usually implemented using analog phase shifters. We consider the subarray structure [99], where each RF chain is connected to $\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}$ antennas via analog phase shifters. Therefore, $\mathbf{F}_{\mathrm{RF}}$ has to be expressed as $\mathbf{F}_{\mathrm{RF}} = \mathrm{blkdiag}\left(\mathbf{f}_{\mathrm{RF},1}, \mathbf{f}_{\mathrm{RF},2}, \ldots, \mathbf{f}_{\mathrm{RF},N_{\mathrm{RF}}}\right)$, where $\mathbf{f}_{\mathrm{RF},r} \in \mathbb{C}^{\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}} \times 1} \ \forall r \in \{1, 2, \ldots, N_{\mathrm{RF}}\}$ and $\left|[\mathbf{f}_{\mathrm{RF},r}]_m\right| = \frac{1}{\sqrt{N_{\mathrm{T}}/N_{\mathrm{RF}}}} \ \forall r, m$ [99]. Therefore, we have $\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}} = \mathbf{I}_{N_{\mathrm{RF}}}$. Moreover, the power constraint is reduced to $\|\mathbf{f}_{\mathrm{BB}}\|^2 = \|\mathbf{U}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = 1$. Let $\mathcal{F}_{\mathrm{RF}}$ be the set of analog RF precoders satisfying the constraints of subarray structure, then we should have $\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}$.

3.2.2 Channel Model

Millimeter-wave channels are expected to have limited scattering [145, 141]. Therefore, we adopt a sparse geometric multipath channel model, where the channel vector $\mathbf{h}_k$ to the kth user is given by

$$\mathbf{h}_k = \sqrt{\frac{N_{\mathrm{T}}}{L}} \sum_{l=1}^L \alpha_{l,k}\mathbf{a}_{l,k}^H, \tag{3.5}$$

where L is the number of propagation paths, $\alpha_{l,k}$ is the complex channel gain of the lth path to the kth user, $\mathbf{a}_{l,k}$ is the transmit steering vector of the lth path to the kth user with azimuth angle of departure (AoD) of $\varphi_{l,k}$, and

$$\mathbf{a}_{l,k} = \frac{1}{\sqrt{N_\text{T}}} \left[1, e^{-j\pi \cos(\varphi_{l,k})}, \dots, e^{-j\pi(N_\text{T}-1)\cos(\varphi_{l,k})} \right]^T. \tag{3.6}$$

Denoted by $\mathbf{A}_{\text{T},k} \in \mathbb{C}^{N_\text{T} \times L}$, the transmit array response matrix to the kth receiver is given by $\mathbf{A}_{\text{T},k} = [\mathbf{a}_{1,k}, \mathbf{a}_{2,k}, \dots, \mathbf{a}_{L,k}]$. Since each resolvable path consists of several paths, similar to [67, Section II-B] and [1, Section III-E], the channel gains $\{\alpha_{l,k}\}$ are assumed to be independent complex Gaussian random variables with zero-mean and unit variance.

3.3   AN-aided Hybrid Precoding Design

We design the hybrid precoder to maximize the average secrecy rate $\bar{R}_\text{sec}$ for a given transmit SNR γ. We assume that Alice has full knowledge of the channel to Bob but has partial knowledge of the channels to Eves. Similar to [107, 67, 108], Alice has knowledge only of the AoDs of the paths to Eves. Bob and Eves have full knowledge of their channels to Alice. These assumptions become realistic if Eves are active nodes which have communicated with Alice [152]. We assume also that Eves do not cooperate.

We illustrate the partial channel knowledge as follows. Since we assume that Eves are active nodes in the system (Bob and Eves play interchangeable roles), the channel knowledge of Bob and Eves may not be up-to-date. Alice re-estimates the channel to Bob. Since the coherence time of AoDs is much longer (tens or hundreds times as reported in [79]) than that of the channel gains [107, 67, 136, 105, 134], Alice uses the estimates of the AoDs of the paths to Eves as partial channel knowledge (assuming that the AoDs remain almost unchanged), and do not re-estimate the channel gains of Eves.

With full knowledge of the channel to Bob and partial knowledge of the channels to Eves, maximizing the secrecy rate R_sec given by

$$R_\text{sec} = R_1 - \mathbb{E}\left\{ \max_k \{R_k\}_{k=2}^K \right\}, \tag{3.7}$$

where the expectation is performed over the unknown channel gains of Eves, is equivalent to maximizing the average secrecy rate $\bar{R}_{\text{sec}}$ [43]. Since it is difficult to get R_{sec} in a closed-form, we derive a secrecy rate lower bound $\tilde{R}_{\text{sec}}$ as

$$
\begin{aligned}
R_{\text{sec}} &= R_1 - \mathbb{E}\left\{\log_2\left(1 + \max_k \{\text{SINR}_k\}_{k=2}^{K}\right)\right\} \\
&\geq R_1 - \log_2\left(1 + \mathbb{E}\left\{\max_k \{\text{SINR}_k\}_{k=2}^{K}\right\}\right) \\
&\geq R_1 - \log_2\left(1 + \sum_{k=2}^{K} \mathbb{E}\{\text{SINR}_k\}\right) \triangleq \tilde{R}_{\text{sec}},
\end{aligned}
\tag{3.8}
$$

where $\mathbb{E}\{\text{SINR}_k\} = \mathbb{E}\left\{\frac{\gamma\phi|\mathbf{h}_k\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}|^2}{\gamma(1-\phi)\|\mathbf{h}_k\mathbf{F}_{\text{RF}}\mathbf{U}_{\text{BB}}\|^2+1}\right\}$ is the average SINR of the kth receiver, the first inequality holds due to Jensen's inequality, and the second inequality holds since $\max_k \{x_k\}_{k=2}^{K} \leq \sum_{k=2}^{K} x_k$. We need to evaluate $\mathbb{E}\{\text{SINR}_k\}$ to get $\tilde{R}_{\text{sec}}$. It can be rewritten as

$$
\mathbb{E}\{\text{SINR}_k\} = \mathbb{E}\left[\frac{\phi\boldsymbol{\alpha}_k^H \mathbf{A}_k \boldsymbol{\alpha}_k}{(1-\phi)\boldsymbol{\alpha}_k^H \mathbf{B}_k \boldsymbol{\alpha}_k + \delta}\right],
\tag{3.9}
$$

where $\boldsymbol{\alpha}_k = [\alpha_{1,k}, \alpha_{2,k}, \ldots, \alpha_{L,k}]^T$, $\mathbf{A}_k = \mathbf{A}_{\text{T},k}^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB}} \mathbf{f}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{A}_{\text{T},k}$, $\mathbf{B}_k = \mathbf{A}_{\text{T},k}^H \mathbf{F}_{\text{RF}} \mathbf{U}_{\text{BB}} \mathbf{U}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{A}_{\text{T},k}$, and $\delta = \frac{L}{\gamma N_{\text{T}}}$. We can notice that $\mathbb{E}\{\text{SINR}_k\}$ in (3.9) is the expected value of a ratio of quadratic forms of $\boldsymbol{\alpha}_k$. Moreover, $\boldsymbol{\alpha}_k$ is a circularly-symmetric complex Gaussian random vector with a probability density function (PDF) $f_{\boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k)$ given by [120]

$$
f_{\boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k) = \frac{e^{-\boldsymbol{\alpha}_k^H \boldsymbol{\alpha}_k}}{\pi^L}.
\tag{3.10}
$$

Let w_1 and w_2 be two random variables such that $\mathbb{P}(w_2 > 0) = 1$. In [83], it was shown that

$$
\mathbb{E}\left[\frac{w_1}{w_2}\right] = \int_0^\infty \left[\frac{\partial}{\partial s} M_{w_1,w_2}(s,-r)\right]_{s=0} dr,
\tag{3.11}
$$

where $M_{w_1,w_2}(s,r)$ is the joint moment generating function of w_1 and w_2. Define $w_1 = \phi\boldsymbol{\alpha}_k^H \mathbf{A}_k \boldsymbol{\alpha}_k$ and $w_2 = (1-\phi)\boldsymbol{\alpha}_k^H \mathbf{B}_k \boldsymbol{\alpha}_k + \delta$. Then, we obtain $M_{w_1,w_2}(s,-r)$ as

$$
M_{w_1,w_2}(s,-r) = \int_{-\infty}^\infty \pi^{-L} e^{-\boldsymbol{\alpha}_k^H \boldsymbol{\alpha}_k} e^{\phi\boldsymbol{\alpha}_k^H \mathbf{A}_k \boldsymbol{\alpha}_k s - (1-\phi)\boldsymbol{\alpha}_k^H \mathbf{B}_k \boldsymbol{\alpha}_k r - \delta r} d\boldsymbol{\alpha}_k
$$

$$= e^{-\delta r} \left| \mathbf{I}_L - \phi \mathbf{A}_k s + (1 - \phi) \mathbf{B}_k r \right|^{-1}$$

$$\times \underbrace{\int_{-\infty}^{\infty} \pi^{-L} \left| \mathbf{I}_L - \phi \mathbf{A}_k s + (1 - \phi) \mathbf{B}_k r \right| e^{-\boldsymbol{\alpha}_k^H (\mathbf{I}_L - \phi \mathbf{A}_k s + (1-\phi) \mathbf{B}_k r) \boldsymbol{\alpha}_k} d\boldsymbol{\alpha}_k}_{=1 \text{ (area under PDF)}}$$

$$= e^{-\delta r} \left| \mathbf{I}_L - \phi \mathbf{A}_k s + (1 - \phi) \mathbf{B}_k r \right|^{-1}. \tag{3.12}$$

Therefore, we get

$$\mathbb{E} \left\{ \mathrm{SINR}_k \right\} = \int_0^{\infty} \left[\frac{\partial}{\partial s} M_{w_1, w_2}(s, -r) \right]_{s=0} dr$$

$$= \int_0^{\infty} e^{-\delta r} \left| \mathbf{I}_L + (1 - \phi) \mathbf{B}_k r \right|^{-1} \mathrm{Tr} \left[(\mathbf{I}_L + (1 - \phi) \mathbf{B}_k r)^{-1} \phi \mathbf{A}_k \right] dr, \tag{3.13}$$

which can be simplified to

$$\mathbb{E} \left\{ \mathrm{SINR}_k \right\} = \mathbf{f}_{\mathrm{BB}}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{A}_{\mathrm{T},k} \mathbf{U}_k \mathrm{diag} \left(\left\{ \underbrace{\int_0^{\infty} \frac{\phi e^{-\delta r} dr}{(1 + (1 - \phi) \lambda_{k,l} r) \prod_{m=1}^L (1 + (1 - \phi) \lambda_{k,m} r)}}_{=I_{k,l} \text{ (obtained in a closed-form in Appendix)}} \right\}_{l=1}^L \right)$$

$$\times \mathbf{U}_k^H \mathbf{A}_{\mathrm{T},k}^H \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB}}, \tag{3.14}$$

where $\{\lambda_{k,l}\}_{l=1}^L$ and $\mathbf{U}_k \in \mathbb{C}^{L \times L}$ are the eigenvalues and the eigenvectors matrix of $\mathbf{B}_k$ respectively, and the integral $I_{k,l}$ in (3.14) is obtained in a closed-form in Appendix. Therefore, we can write $\tilde{R}_{\mathrm{sec}}$ in a closed-form.

Using the secrecy rate lower bound $\tilde{R}_{\mathrm{sec}}$, the AN-aided hybrid precoder design problem is expressed as

$$\max_{\phi, \mathbf{F}_{\mathrm{RF}}, \mathbf{f}_{\mathrm{BB}}, \mathbf{U}_{\mathrm{BB}}} \tilde{R}_{\mathrm{sec}},$$

$$\text{s.t. } \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \ 0 \leq \phi \leq 1, \ \|\mathbf{f}_{\mathrm{BB}}\|^2 = 1, \ \|\mathbf{U}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = 1. \tag{3.15}$$

The optimization problem in (3.15) is non-convex since the objective function and the RF precoder constraint are non-convex. Therefore, we propose a suboptimal AN-aided hybrid precoder design to maximize the secrecy rate lower bound $\tilde{R}_{\mathrm{sec}}$. To get an efficient solution

and decouple the hybrid precoder and the power fraction ϕ designs, we put $\phi = 1$ (no AN) and design the hybrid precoder. Then, we get ϕ maximizing $\tilde{R}_{\text{sec}}$ by any efficient one-dimensional search.

With $\phi = 1$ (no AN), $\mathbb{E}\left\{\text{SINR}_k\right\}$ in (3.14) is simplified to

$$\mathbb{E}\left\{\text{SINR}_k\right\} = \frac{\gamma N_{\text{T}}}{L}\left\|\mathbf{A}_{\text{T},k}^H \mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}\right\|^2. \tag{3.16}$$

Therefore, $\tilde{R}_{\text{sec}}$ in (3.8) can be written while applying the power constraint $\|\mathbf{f}_{\text{BB}}\|^2 = 1$ as

$$\tilde{R}_{\text{sec}} = \log_2\left(\frac{\mathbf{f}_{\text{BB}}^H\left(\mathbf{I}_{N_{\text{RF}}} + \gamma \mathbf{F}_{\text{RF}}^H \mathbf{h}_1^H \mathbf{h}_1 \mathbf{F}_{\text{RF}}\right)\mathbf{f}_{\text{BB}}}{\mathbf{f}_{\text{BB}}^H\left(\mathbf{I}_{N_{\text{RF}}} + \frac{\gamma N_{\text{T}}}{L}\sum_{k=2}^K \mathbf{F}_{\text{RF}}^H \mathbf{A}_{\text{T},k}\mathbf{A}_{\text{T},k}^H \mathbf{F}_{\text{RF}}\right)\mathbf{f}_{\text{BB}}}\right). \tag{3.17}$$

Using the generalized eigenvector decomposition, we obtain $\mathbf{f}_{\text{BB}}$ maximizing (3.17) as a function of $\mathbf{F}_{\text{RF}}$ as

$$\mathbf{f}_{\text{BB}} = \boldsymbol{\mathcal{E}}_{\max}\left[\left(\mathbf{I}_{N_{\text{RF}}} + \frac{\gamma N_{\text{T}}}{L}\sum_{k=2}^K \mathbf{F}_{\text{RF}}^H \mathbf{A}_{\text{T},k}\mathbf{A}_{\text{T},k}^H \mathbf{F}_{\text{RF}}\right)^{-1}\left(\mathbf{I}_{N_{\text{RF}}} + \gamma \mathbf{F}_{\text{RF}}^H \mathbf{h}_1^H \mathbf{h}_1 \mathbf{F}_{\text{RF}}\right)\right]. \tag{3.18}$$

If a fully digital precoder $\mathbf{f}_{\text{FD}}$ (i.e., $\mathbf{F}_{\text{RF}} = \mathbf{I}_{N_{\text{T}}}$) was used, (3.18) would be written as

$$\mathbf{f}_{\text{FD}} = \boldsymbol{\mathcal{E}}_{\max}\left[\left(\mathbf{I}_{N_{\text{T}}} + \frac{\gamma N_{\text{T}}}{L}\sum_{k=2}^K \mathbf{A}_{\text{T},k}\mathbf{A}_{\text{T},k}^H\right)^{-1}\left(\mathbf{I}_{N_{\text{T}}} + \gamma \mathbf{h}_1^H \mathbf{h}_1\right)\right]. \tag{3.19}$$

We propose to obtain the RF precoder $\mathbf{F}_{\text{RF}}$ as

$$\mathbf{f}_{\text{RF},r} = \frac{1}{\sqrt{N_{\text{T}}/N_{\text{RF}}}}\exp\left(j\angle\left[\mathbf{f}_{\text{FD}}\right]_{(r-1)\frac{N_{\text{T}}}{N_{\text{RF}}}+1:r\frac{N_{\text{T}}}{N_{\text{RF}}}}\right)\ \forall r, \tag{3.20}$$

which satisfies the constraints of the subarray structure and is a good approximation to $\mathbf{f}_{\text{FD}}$.

Now, we design the AN baseband precoder $\mathbf{U}_{\text{BB}}$ to be in the null space of the equivalent channel to Bob $\hat{\mathbf{h}}_1 = \mathbf{h}_1 \mathbf{F}_{\text{RF}}$ and directed to Eves as

$$\mathbf{U}_{\text{BB}} = \frac{\Pi_{\hat{\mathbf{h}}_1}\mathbf{V}}{\left\|\Pi_{\hat{\mathbf{h}}_1}\mathbf{V}\right\|_{\text{F}}}, \tag{3.21}$$

where $\Pi_{\hat{\mathbf{h}}_1} = \left(\mathbf{I}_{N_{\text{RF}}} - \hat{\mathbf{h}}_1^H\left(\hat{\mathbf{h}}_1 \hat{\mathbf{h}}_1^H\right)^{-1}\hat{\mathbf{h}}_1\right) \in \mathbb{C}^{N_{\text{RF}}\times N_{\text{RF}}}$ is the orthogonal complement projector of $\hat{\mathbf{h}}_1$, and $\mathbf{V} = \boldsymbol{\mathcal{E}}_{1:\min(N_{\text{RF}},(K-1)L)}\left[\Pi_{\hat{\mathbf{h}}_1}^H \mathbf{F}_{\text{RF}}^H \mathbf{A}_{\text{T,Eves}}\mathbf{A}_{\text{T,Eves}}^H \mathbf{F}_{\text{RF}}\Pi_{\hat{\mathbf{h}}_1}\right]$, where $\mathbf{A}_{\text{T,Eves}} = $

$[\mathbf{A}_{\mathrm{T},2}, \mathbf{A}_{\mathrm{T},3}, \ldots, \mathbf{A}_{\mathrm{T},K}] \in \mathbb{C}^{N_{\mathrm{T}} \times (K-1)L}$. Note that $\mathbf{U}_{\mathrm{BB}}$ is normalized to satisfy the power constraint $\|\mathbf{U}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = 1$. Finally, we obtain the power fraction ϕ by any efficient one-dimensional search (e.g., golden section search) as

$$\phi = \arg\max_{\phi} \tilde{R}_{\mathrm{sec}}, \text{ s.t. } 0 \le \phi \le 1, \tag{3.22}$$

where $\tilde{R}_{\mathrm{sec}}$ is given by (3.8) using $\mathbb{E}\{\mathrm{SINR}_k\}$ in (3.14). The computational complexity of the proposed AN-aided hybrid precoder is $\mathcal{O}(N_{\mathrm{T}}^3 + N_{\mathrm{T}}^2 KL + N_{\mathrm{RF}}^3 + KL^3)$.

3.4   Numerical Results

We evaluate the performance of the proposed AN-aided hybrid precoder by means of Monte-Carlo simulations. We assume that Alice has 32 antennas and 4 RF chains. All channels follow the mmWave channel model described in Subsection 3.2.2 with 6 propagation paths, and the angles of departure $\{\varphi_{l,k}\}$ are uniformly-distributed within $[0\ 2\pi)$.

We consider four baseline algorithms. The first algorithm is the MRT hybrid precoder of [99], which designs the hybrid precoder to maximize the rate of Bob while ignoring Eves. The second baseline algorithm (denoted by "Reduced Fully Digital Precoder") applies random antenna selection for each RF chain from its own antenna subset and performs the proposed AN-aided digital precoding (i.e., $N_{\mathrm{T}} = N_{\mathrm{RF}} = 4$). The third baseline algorithm is the no-AN hybrid precoder which is based on the proposed hybrid precoder with no AN. The fourth baseline algorithm (denoted by "Fully Digital Precoder") is the fully digital precoder which applies the proposed AN-aided precoding with full RF chains (i.e., $N_{\mathrm{T}} = N_{\mathrm{RF}} = 32$). Note that the first, second, and third baseline algorithms are expected to be performance lower bounds, while the fourth baseline algorithm is expected to be a performance upper bound.

Fig. 3.1 shows the achievable average secrecy rate (left y-axis) and the optimal power fraction ϕ (right y-axis) as a function of the transmit SNR γ with 4 Eves ($K = 5$). As expected, the AN-aided fully digital precoding achieves the highest average secrecy rate due to the use
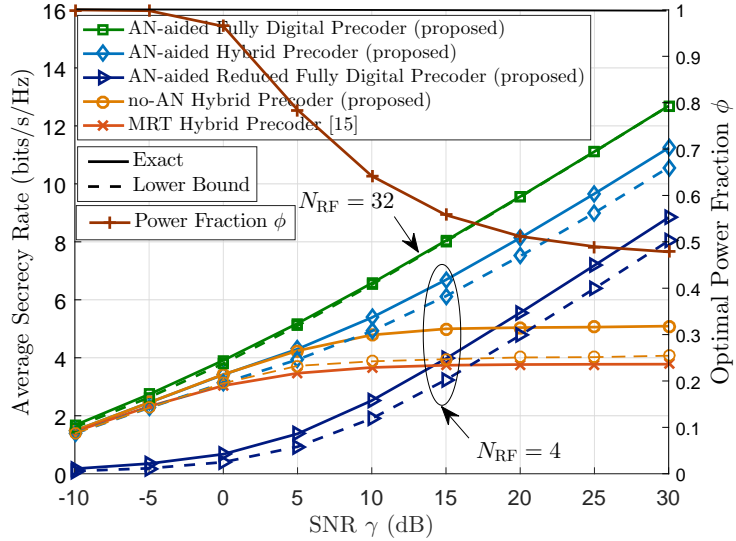
Figure 3.1. Achievable average secrecy rate $\bar{R}_{\text{sec}}$ and optimal power fraction ϕ with 4 Eves ($K = 5$).

of one RF chain per antenna (very large hardware complexity). The proposed AN-aided hybrid precoder achieves comparable performance to that of the AN-aided fully digital precoder, with much lower hardware complexity (only 4 RF chains are used). The performance loss is due to the modulus constraint and the limited number of RF chains. The proposed AN-aided hybrid precoder outperforms the AN-aided reduced fully digital precoder, which verifies the effectiveness of the proposed RF precoder design. The SNR gap between the proposed AN-aided hybrid precoder and the AN-aided reduced fully digital precoder is about 7.5 dB at high transmit SNRs. We also observe that the derived secrecy rate lower bound predicts the performance behavior efficiently. The MRT hybrid precoder achieves the worst average secrecy rate at moderate and high transmit SNRs due to ignoring Eves. At low transmit SNRs, the optimal power fraction ϕ is approximately 1 (no AN) that is why the proposed AN-aided hybrid precoder and the no-AN hybrid precoder have approximately the same performance at low transmit SNRs. As transmit SNR increases, the optimal power fraction ϕ decreases which means allocating more power to the AN. That is why the performance of no-AN hybrid precoder degrades significantly at moderate and high transmit SNRs. On the contrary,
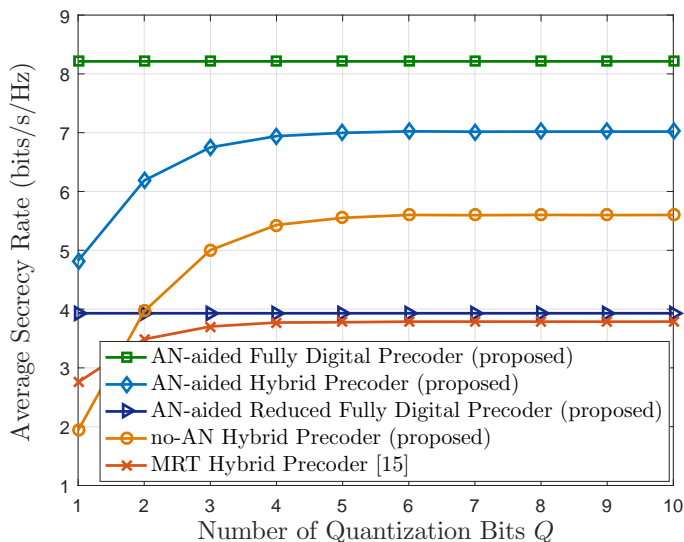
Figure 3.2. Effect of finite resolution phase shifters on the average secrecy rate $\bar{R}_{\text{sec}}$ with 4 Eves ($K = 5$) and transmit SNR $\gamma = 15$ dB.

the secrecy rate achieved by the proposed AN-aided hybrid precoder linearly increases with transmit SNR, thanks to the optimal power allocation to AN. The same above performance behavior is observed when increasing the number of Eves.

With finite resolution phase shifters, Fig. 3.2 shows the achievable average secrecy rate versus different numbers of quantization bits for the phase shifters with 4 Eves ($K = 5$) and transmit SNR $\gamma = 15$ dB. After designing the RF precoder, the phases of the RF precoder are quantized into Q bits such that $\angle \left[\mathbf{f}_{\text{RF},r}\right]_m \in \left\{0, \frac{2\pi}{2^Q}, \ldots, \frac{2\pi 2^{Q-1}}{2^Q}\right\} \forall r, m$. We observe that 4 quantization bits are sufficient for the proposed AN-aided hybrid precoder with secrecy rate loss less than 0.01 bits/s/Hz. With the same number of quantization bits, the proposed AN-aided hybrid precoder outperforms the other hybrid precoders.

3.5   Conclusion

We proposed an AN-aided hybrid precoder design to enhance the physical layer security. The proposed AN-aided hybrid precoder achieves comparable secrecy rate to that of the fully digital precoder, with much lower hardware complexity. Moreover, the proposed AN-aided

hybrid precoder outperforms the conventional MRT hybrid precoder. With finite resolution phase shifters, we showed that 4 quantization bits are sufficient for the proposed AN-aided hybrid precoder with secrecy rate loss less than 0.01 bits/s/Hz.

Appendix A - Evaluating the Integral $I_{k,l}$ in (3.14)

Defining $\mu_{k,l} = \frac{1}{(1-\phi)\lambda_{k,l}}$, the integral $I_{k,l}$ in (3.14) can be written as $I_{k,l} = \phi\mu_{k,l}\left(\prod_{m=1}^{L}\mu_{k,m}\right)\tilde{I}_{k,l}$, where

$$
\begin{aligned}
\tilde{I}_{k,l} &= \int_0^\infty \frac{e^{-\delta r}dr}{(r+\mu_{k,l})\prod_{m=1}^L(r+\mu_{k,m})} \\
&= \int_0^\infty \left(\frac{a_{k,l}e^{-\delta r}}{(r+\mu_{k,l})^2} + \frac{b_{k,l}e^{-\delta r}}{(r+\mu_{k,l})} + \sum_{m=1,\,m\neq l}^L \frac{c_{k,m}e^{-\delta r}}{(r+\mu_{k,m})}\right)dr \\
&= a_{k,l}\left(\delta e^{\mu_{k,l}\delta}\mathrm{Ei}\left(-\mu_{k,l}\delta\right) + \frac{1}{\mu_{k,l}}\right) - b_{k,l}e^{\mu_{k,l}\delta}\mathrm{Ei}\left(-\mu_{k,l}\delta\right) - \sum_{m=1,\,m\neq l}^L c_{k,m}e^{\mu_{k,m}\delta}\mathrm{Ei}\left(-\mu_{k,m}\delta\right),
\end{aligned}
$$

$$(3.23)$$

where $\mathrm{Ei}(x) = -\int_{-x}^\infty \frac{e^{-t}}{t}dt$ is the exponential integral, and the partial fraction coefficients are obtained as $a_{k,l} = \frac{1}{\prod_{n=1,\,n\neq l}^L(\mu_{k,n}-\mu_{k,l})}$, $b_{k,l} = \left[\frac{\partial}{\partial r}\frac{1}{\prod_{n=1,\,n\neq l}^L(r+\mu_{k,n})}\right]_{r=-\mu_{k,l}}$, and $c_{k,m} = \frac{1}{(\mu_{k,l}-\mu_{k,m})\prod_{n=1,\,n\neq m}^L(\mu_{k,n}-\mu_{k,m})}$.

HYBRID PRECODING FOR MMWAVE MULTIUSER SYSTEMS WITH

PARTIAL CHANNEL KNOWLEDGE [1]

4.1 Introduction

The existing hybrid precoder designs for mmWave systems can be divided into two categories. In the first category, it is assumed that the transmitter has full channel knowledge before designing the hybrid precoder. For a single user, it was shown that by minimizing the average Euclidean distance between the fully digital precoder and the hybrid precoder, the hybrid precoder can achieve performance very close to that of the fully digital precoding [142, 94, 146]. For multiple users, a two-stage hybrid precoder design was developed in [7, 9]. At the first stage, the transmitter and the users jointly select (using a feedback from the users) the best pair of RF precoder and RF combiner to maximize the channel gain. Then, the baseband precoder is designed as a zero-forcing (ZF) precoder for the equivalent channels. In [93], the hybrid precoder design was enhanced by minimizing the mean-squared error (MSE) of the transmitted data streams. In the second category, it is assumed that the transmitter has knowledge of the second-order channel statistics [5, 8, 27, 100]. Using the second-order channel statistics, the transmitter designs the RF precoder. Then, the transmitter estimates the equivalent channel based on which the baseband precoder is designed as a ZF precoder.

We note that most of previous works assume full channel knowledge at the transmitter either before designing the hybrid precoder (the first category) or after designing the RF precoder (the second category). On the contrary, the assumption of partial channel knowledge at the transmitter, where the transmitter has knowledge of the angles of departure (AoDs) of the propagation paths only, is more practical. Since the AoDs are invariant with frequency

---

in time-division duplex (TDD) or frequency-division duplex (FDD) operation mode, the transmitter does not need any feedback [136, 5, 105]. Moreover, the variation of the AoDs is slower than the variation of the channel gains that makes it possible to use the same hybrid precoder for multiple symbols. This chapter designs the hybrid precoder to maximize the average sum rate of mmWave multiuser downlink systems with partial channel knowledge in contrast to the full channel knowledge assumption of the existing approaches.

Our contributions can be summarized as follows. 1) We design the ZF hybrid precoder with partial channel knowledge if $N_{\mathrm{T}} > (K-1)L$ and $N_{\mathrm{RF}} \geq 2K$, where L is the number of propagation paths per user. 2) If these conditions are not satisfied, we propose another hybrid precoder design that is based on signal-to-interference-and-noise ratio (SINR) and signal-to-leakage-and-noise ratio (SLNR). We derive closed-form expressions for the average SINR and SLNR. Using a lower bound on SLNR, we obtain the baseband precoder as a function of the RF precoder. Then, we propose a simple gradient ascent algorithm which designs the RF precoder using the gradient of the closed-form SLNR. To ensure the convergence, we update the RF precoder to maximize the average sum rate obtained using the closed-form SINR. 3) Since the hybrid precoder design with partial channel knowledge for multiple users has not been discussed in the literature, we develop the closed-form Eigenvector-SLNR (EV-SLNR) hybrid precoder (based on previous works with some modifications) which we consider as our benchmark design. 4) We present sum-rate characteristics of the proposed hybrid precoders with partial channel knowledge under different values of N_{RF} and SNR.

The rest of this chapter is organized as follows. In section II, we describe the system and the channel models. In section III, we present the proposed ZF hybrid precoder. In section IV, we introduce the proposed SINR-SLNR hybrid precoder. In section V, we develop the EV-SLNR hybrid precoder. In section VI, we present the numerical results. Finally, section VII concludes the chapter.
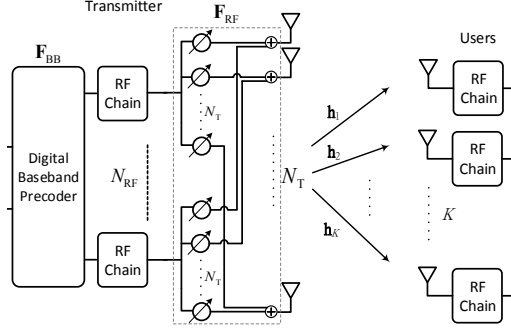
Figure 4.1. mmWave multiuser system with K single-antenna users.

4.2 System and Channel Models

4.2.1 System Model

We consider a mmWave multiuser downlink system with K single-antenna users as shown in Fig. 4.1. The transmitter is equipped with a uniform linear array (ULA) with N_{T} antennas. The spacing between antennas is half the wavelength. To reduce the hardware complexity and the power consumption, the uniform linear antenna array is connected via an analog RF precoder to N_{RF} RF chains ($K \leq N_{\mathrm{RF}} < N_{\mathrm{T}}$) which process the digitally-precoded streams.

We consider a narrow-band transmission, where the received signal y_k at the kth user is given by

$$y_k = \mathbf{h}_k \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},k} s_k + \sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{h}_k \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},i} s_i + z_k, \tag{4.1}$$

where $\mathbf{h}_k \in \mathbb{C}^{1 \times N_{\mathrm{T}}}$ is the mmWave channel to the kth user, $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{RF}}}$ is the analog RF precoder, $\mathbf{f}_{\mathrm{BB},k} \in \mathbb{C}^{N_{\mathrm{RF}} \times 1}$ is the digital baseband precoder for the kth user, s_k is the transmitted modulated symbol to the kth user with $\mathbb{E}\left[|s_k|^2\right] = P$, and z_k is the zero-mean additive white complex Gaussian noise with variance σ^2 at the kth user. The RF precoder $\mathbf{F}_{\mathrm{RF}}$ and the baseband precoder $\mathbf{F}_{\mathrm{BB}} = [\mathbf{f}_{\mathrm{BB},1}, \mathbf{f}_{\mathrm{BB},2}, \dots, \mathbf{f}_{\mathrm{BB},K}] \in \mathbb{C}^{N_{\mathrm{RF}} \times K}$ have to be designed jointly due to the coupled power constraint $\|\mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}\|_{\mathrm{F}}^2 = K$. The analog RF preocder $\mathbf{F}_{\mathrm{RF}}$ is usually implemented using analog phase shifters and analog combiners. We consider

the fully-connected structure which requires $N_{\rm T} N_{\rm RF}$ analog phase shifters and $N_{\rm T}$ analog combiners. Therefore, we have the constraint that $|\mathbf{F}_{\rm RF}(m,n)| = 1 \; \forall m,n$.

4.2.2 Channel Model

Millimeter-wave channels are expected to have limited scattering [145, 141]. Therefore, we adopt a sparse geometric multipath channel model, where the channel vector $\mathbf{h}_k$ to the $k^{\rm th}$ user is given by

$$\mathbf{h}_k = \sqrt{\frac{N_{\rm T}}{L}} \sum_{l=1}^{L} \alpha_{l,k} \mathbf{a}_{l,k}^H, \tag{4.2}$$

where L is the number of propagation paths, $\alpha_{l,k}$ is the channel gain of the $l^{\rm th}$ path to the $k^{\rm th}$ user, $\mathbf{a}_{l,k}$ is the transmit steering vectors of the $l^{\rm th}$ path to the $k^{\rm th}$ user with azimuth angle of departure (AoD) of $\varphi_{l,k}$, and

$$\mathbf{a}_{l,k} = \frac{1}{\sqrt{N_{\rm T}}} \left[1, e^{-j\pi \cos(\varphi_{l,k})}, \ldots, e^{-j\pi(N_{\rm T}-1)\cos(\varphi_{l,k})} \right]^T. \tag{4.3}$$

Let $\mathbf{A}_{{\rm T},k} \in \mathbb{C}^{N_{\rm T} \times L}$ be the transmit array response matrix to the $k^{\rm th}$ user given by

$$\mathbf{A}_{{\rm T},k} = [\mathbf{a}_{1,k}, \mathbf{a}_{2,k}, \ldots, \mathbf{a}_{L,k}]. \tag{4.4}$$

Since each resolvable path consists of several paths, similar to [1, Section III-E], the channel gains $\{\alpha_{l,k}\}$ are assumed to be independent complex Gaussian random variables with zero-mean and unit variance.

Throughout the chapter, we assume partial channel knowledge at the transmitter, where the transmitter has knowledge of only the AoDs of the propagation paths. Since the AoDs are invariant with frequency in TDD or FDD operation mode, the transmitter does not need any feedback [136, 5, 105]. Moreover, the variation of the AoDs is slower than the variation of the channel gains that makes it possible to use the same hybrid precoder for multiple symbols.

4.2.3 Hybrid Precoding Design Problem

The achievable rate R_k of the kth user is given by

$$R_k = \log_2 \left(1 + \frac{\left| \mathbf{h}_k \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \right|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} \left| \mathbf{h}_k \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},i} \right|^2 + \delta} \right), \tag{4.5}$$

where $\delta = 1/\gamma$ and $\gamma = P/\sigma^2$ is the transmit SNR per user. Our aim is to maximize the average sum rate R_{sum} given by

$$R_{\text{sum}} = \sum_{k=1}^{K} \mathbb{E} \left[\log_2 \left(1 + \frac{\left| \mathbf{h}_k \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \right|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} \left| \mathbf{h}_k \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},i} \right|^2 + \delta} \right) \right], \tag{4.6}$$

where the expectation is performed only over the unknown channel gains $\{\alpha_{l,k}\}$. Using Jensen's inequality, we have

$$R_{\text{sum}} \leq \sum_{k=1}^{K} \log_2 \left(1 + \overline{\text{SINR}}_k \right) \triangleq R_{\text{sum}}^{\text{UB}}, \tag{4.7}$$

where $\overline{\text{SINR}}_k = \mathbb{E} \left[\frac{\left| \mathbf{h}_k \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \right|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} \left| \mathbf{h}_k \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},i} \right|^2 + \delta} \right]$ is the average receive SINR of the kth user. Alternatively, we design the hybrid precoder to maximize the average sum rate upper bound $R_{\text{sum}}^{\text{UB}}$,

$$
\begin{aligned}
[\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}] \quad &= \quad \arg\max_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} R_{\text{sum}}^{\text{UB}}, \\
\text{s.t.} \quad &\left| \mathbf{F}_{\text{RF}} \left(m, n \right) \right| = 1 \,\forall m, n, \\
&\left\| \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \right\|_{\text{F}}^2 = K.
\end{aligned}
\tag{4.8}
$$

Next, we propose different hybrid precoding designs to maximize the average sum rate upper bound $R_{\text{sum}}^{\text{UB}}$.

4.3 Proposed Zero-Forcing Hybrid Precoder

The zero-forcing (ZF) hybrid precoder is designed to null the interference to each user. With full channel knowledge, the baseband precoder is used to null the interference to each user

[7, 9]. As a result, it is only necessary that $N_{\mathrm{RF}} \geq K$ to apply the ZF precoder with full channel knowledge. On the other hand, with partial channel knowledge, using the baseband precoder to null the interference to each user requires that $N_{\mathrm{RF}} > (K-1)L$ since it is necessary to null L directions for each one of the $K-1$ interfering users. However, the condition $N_{\mathrm{RF}} > (K-1)L$ is not likely to be satisfied in mmWave systems due to the high cost and power of RF chains. Alternatively, we use the RF precoder to null the interference to each user. Therefore, we obtain a simpler condition on N_{RF} (as will be shown).

Dropping the unit modulus constraint, let $\mathbf{f}_k$ be the precoder for the kth user. We express $\mathbf{f}_k$ as

$$\mathbf{f}_k = \mathbf{U}_k \tilde{\mathbf{f}}_k, \tag{4.9}$$

where $\mathbf{U}_k = \mathcal{N}\left[\mathbf{A}_{\mathrm{T},1}, \ldots, \mathbf{A}_{\mathrm{T},k-1}, \mathbf{A}_{\mathrm{T},k+1}, \ldots, \mathbf{A}_{\mathrm{T},K}\right] \in \mathbb{C}^{N_{\mathrm{T}} \times (N_{\mathrm{T}}-(K-1)L)}$ is a semi-unitary matrix in the null space of the directions to the $K-1$ interfering users, and $\tilde{\mathbf{f}}_k \in \mathbb{C}^{(N_{\mathrm{T}}-(K-1)L)\times 1}$. Therefore, it is necessary that $N_{\mathrm{T}} > (K-1)L$. We design $\tilde{\mathbf{f}}_k$ to maximize the expected receive SNR of the kth user $(\overline{\mathrm{SNR}}_k)$ given by

$$\begin{aligned}
\overline{\mathrm{SNR}}_k &= \mathbb{E}\left[\gamma \left|\mathbf{h}_k \mathbf{U}_k \tilde{\mathbf{f}}_k\right|^2\right] \\
&= \frac{\gamma N_{\mathrm{T}}}{L} \tilde{\mathbf{f}}_k^H \mathbf{U}_k^H \left(\sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbb{E}\left[|\alpha_{l,k}|^2\right] \mathbf{a}_{l,k}^H\right) \mathbf{U}_k \tilde{\mathbf{f}}_k \\
&= \frac{\gamma N_{\mathrm{T}}}{L} \tilde{\mathbf{f}}_k^H \mathbf{U}_k^H \left(\sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H\right) \mathbf{U}_k \tilde{\mathbf{f}}_k.
\end{aligned} \tag{4.10}$$

By maximizing $\overline{\mathrm{SNR}}_k$ with the constraint $\|\mathbf{f}_k\|^2 = 1$, we obtain $\tilde{\mathbf{f}}_k$ as

$$\tilde{\mathbf{f}}_k = \mathcal{E}_{\max}\left[\mathbf{U}_k^H \left(\sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H\right) \mathbf{U}_k\right], \tag{4.11}$$

and the corresponding $R_{\mathrm{sum,ZF}}^{\mathrm{UB}}$ is expressed as

$$R_{\mathrm{sum,ZF}}^{\mathrm{UB}} = \sum_{k=1}^{K} \log_2\left(1 + \frac{\gamma N_{\mathrm{T}}}{L} \lambda_{\max}\left[\mathbf{U}_k^H \left(\sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H\right) \mathbf{U}_k\right]\right). \tag{4.12}$$

Therefore, the precoder matrix $\mathbf{F}$ is expressed as $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_K] \in \mathbb{C}^{N_\mathrm{T} \times K}$. However, the precoder matrix $\mathbf{F}$ is not likely to satisfy the unit modulus constraint. In [149], it was shown that any vector $\mathbf{x} \in \mathbb{C}^{N \times 1}$ can be expressed as $\mathbf{x} = \tilde{\mathbf{X}}\tilde{\mathbf{x}}$, where $\tilde{\mathbf{X}} \in \mathbb{C}^{N \times 2}$ is with unit modulus entries and $\tilde{\mathbf{x}} \in \mathbb{R}^{2 \times 1}$. Following this decomposition, we decompose $\mathbf{F}$ as $\mathbf{F} = \mathbf{Q}_\mathrm{RF}\mathbf{R}_\mathrm{BB}$, where $\mathbf{Q}_\mathrm{RF} \in \mathbb{C}^{N_\mathrm{T} \times 2K}$ is with unit modulus entries, and $\mathbf{R}_\mathrm{BB} \in \mathbb{R}^{2K \times K}$. Therefore, it is also necessary that $N_\mathrm{RF} \geq 2K$ to apply the ZF hybrid precoder by setting $\mathbf{F}_\mathrm{RF}(:, 1:2K) = \mathbf{Q}_\mathrm{RF}$ and $\mathbf{F}_\mathrm{BB}(1:2K, :) = \mathbf{R}_\mathrm{BB}$. In conclusion, with partial channel knowledge, it is necessary that $N_\mathrm{T} > (K-1)L$ and $N_\mathrm{RF} \geq 2K$ to apply the ZF hybrid precoder.

4.4 Proposed SINR-SLNR Hybrid Precoder

If the conditions $N_\mathrm{T} > (K-1)L$ and/ or $N_\mathrm{RF} \geq 2K$ are not satisfied, the zero-forcing hybrid precoder is infeasible. Therefore, we propose another hybrid precoder design that is based on SINR and SLNR. Let $\overline{\mathrm{SINR}}_k$ and $\overline{\mathrm{SLNR}}_k$ be the expected receive SINR and SLNR of the kth user respectively. Next, we drive closed-form expressions for $\overline{\mathrm{SINR}}_k$ and $\overline{\mathrm{SLNR}}_k$. Then, we use both $\overline{\mathrm{SINR}}_k$ and $\overline{\mathrm{SLNR}}_k$ to maximize $R_\mathrm{sum}^\mathrm{UB}$.

4.4.1 A Closed-Form Expression for SINR

The average receive SINR of the kth user $\overline{\mathrm{SINR}}_k$ is given by

$$\overline{\mathrm{SINR}}_k = \mathbb{E}\left[\frac{|\mathbf{h}_k \mathbf{F}_\mathrm{RF}\mathbf{f}_{\mathrm{BB},k}|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} |\mathbf{h}_k \mathbf{F}_\mathrm{RF}\mathbf{f}_{\mathrm{BB},i}|^2 + \delta}\right]. \tag{4.13}$$

It can be rewritten as

$$\overline{\mathrm{SINR}}_k = \mathbb{E}\left[\frac{\boldsymbol{\alpha}_k^H \mathbf{A}_k \boldsymbol{\alpha}_k}{\boldsymbol{\alpha}_k^H \mathbf{B}_k \boldsymbol{\alpha}_k + \tilde{\delta}}\right], \tag{4.14}$$

where $\boldsymbol{\alpha}_k = [\alpha_{1,k}, \alpha_{2,k}, \ldots, \alpha_{L,k}]^T$, $\mathbf{A}_k = \mathbf{A}_{\mathrm{T},k}^H \mathbf{F}_\mathrm{RF}\mathbf{f}_{\mathrm{BB},k}\mathbf{f}_{\mathrm{BB},k}^H \mathbf{F}_\mathrm{RF}^H \mathbf{A}_{\mathrm{T},k}$, $\mathbf{B}_k = \mathbf{A}_{\mathrm{T},k}^H \mathbf{F}_\mathrm{RF}\left(\sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{f}_{\mathrm{BB},i}\mathbf{f}_{\mathrm{BB},i}^H\right) \mathbf{F}_\mathrm{RF}^H \mathbf{A}_{\mathrm{T},k}$, and $\tilde{\delta} = \frac{L}{\gamma N_\mathrm{T}}$. We can notice that $\overline{\mathrm{SINR}}_k$ in (4.14) is the expected value of a ratio of quadratic forms of $\boldsymbol{\alpha}_k$. Moreover, $\boldsymbol{\alpha}_k$ is

61

a circularly-symmetric complex Gaussian random vector with a probability density function (PDF) $f_{\boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k)$ given by [120]

$$f_{\boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k) = \frac{e^{-\boldsymbol{\alpha}_k^H \boldsymbol{\alpha}_k}}{\pi^L}. \tag{4.15}$$

Let w_1 and w_2 be two random variables such that $\mathbb{P}(w_2 > 0) = 1$. In [83], it was shown that

$$\mathbb{E}\left[\frac{w_1}{w_2}\right] = \int_0^\infty \left[\frac{\partial}{\partial s} M_{w_1, w_2}(s, -r)\right]_{s=0} dr, \tag{4.16}$$

where $M_{w_1, w_2}(s, r)$ is the joint moment generating function of w_1 and w_2. Define $w_1 = \boldsymbol{\alpha}_k^H \mathbf{A}_k \boldsymbol{\alpha}_k$ and $w_2 = \boldsymbol{\alpha}_k^H \mathbf{B}_k \boldsymbol{\alpha}_k + \tilde{\delta}$, we obtain $M_{w_1, w_2}(s, -r)$ as

$$M_{w_1, w_2}(s, -r) = \int_{-\infty}^\infty \pi^{-L} e^{-\boldsymbol{\alpha}_k^H \boldsymbol{\alpha}_k} e^{\boldsymbol{\alpha}_k^H \mathbf{A}_k \boldsymbol{\alpha}_k s - \boldsymbol{\alpha}_k^H \mathbf{B}_k \boldsymbol{\alpha}_k r - \tilde{\delta} r} d\boldsymbol{\alpha}_k$$

$$= e^{-\tilde{\delta} r} |\mathbf{I}_L - \mathbf{A}_k s + \mathbf{B}_k r|^{-1}. \tag{4.17}$$

Therefore, we can evaluate $\overline{\mathrm{SINR}}_k$ as

$$\overline{\mathrm{SINR}}_k = \int_0^\infty \left[\frac{\partial}{\partial s} e^{-\tilde{\delta} r} |\mathbf{I}_L - \mathbf{A}_k s + \mathbf{B}_k r|^{-1}\right]_{s=0} dr$$

$$= \int_0^\infty e^{-\tilde{\delta} r} |\mathbf{I}_L + \mathbf{B}_k r|^{-1} \mathrm{Tr}\left[(\mathbf{I}_L + \mathbf{B}_k r)^{-1} \mathbf{A}_k\right] dr, \tag{4.18}$$

which can be simplified to

$$\overline{\mathrm{SINR}}_k = \mathbf{f}_{\mathrm{BB},k}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{A}_{\mathrm{T},k} \mathbf{V}_k \mathrm{diag}\left(\left\{\underbrace{\int_0^\infty \frac{e^{-\tilde{\delta} r} dr}{(1 + \lambda_{k,l} r) \prod_{m=1}^M (1 + \lambda_{k,m} r)}}_{I_{k,l}}\right\}_{l=1}^M\right) \mathbf{V}_k^H \mathbf{A}_{\mathrm{T},k}^H \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},k},$$

$$\tag{4.19}$$

where $M = \min(K - 1, L)$, and $\{\lambda_{k,l}\}_{l=1}^M$ and $\mathbf{V}_k \in \mathbb{C}^{L \times M}$ are the eigenvalues and the eigenvectors matrix of $\mathbf{B}_k$ respectively. The integral $I_{k,l}$ in (4.19) is evaluated in Appendix. As a result, we can also write $R_{\mathrm{sum}}^{\mathrm{UB}}$ in a closed-form. However, $\overline{\mathrm{SINR}}_k$ in (4.19) is a non-convex and non-differentiable function of the hybrid precoder, which makes the optimization problem in (4.8) difficult to solve.

4.4.2 A Closed-Form Expression for SLNR

An alternative approach to design the hybrid precoder is to consider SLNR. We derive a closed-form expression for $\overline{\text{SLNR}}_k$. Interestingly, the derived $\overline{\text{SLNR}}_k$ is a differentiable function of the hybrid precoder (as will be shown).

The average receive SLNR of the kth user $\overline{\text{SLNR}}_k$ is given by

$$\overline{\text{SLNR}}_k = \mathbb{E}\left[\frac{\frac{L}{N_{\text{T}}} \left|\mathbf{h}_k \mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2}{\frac{L}{N_{\text{T}}} \sum_{\substack{i=1 \\ i\neq k}}^{K} \left|\mathbf{h}_i \mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2 + \tilde{\delta}} \right]. \tag{4.20}$$

We know that $\sqrt{\frac{L}{N_{\text{T}}}}\mathbf{h}_k\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k} = \sum_{l=1}^{L} \alpha_{l,k}\mathbf{a}_{l,k}^{H}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}$ which is a complex Gaussian random variable with zero-mean and variance of $\sum_{l=1}^{L} \left|\mathbf{a}_{l,k}^{H}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2$. Therefore, $v_1 = \frac{L}{N_{\text{T}}}\left|\mathbf{h}_k\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2$ is an exponential random variable with mean value $\mu_{k,k} = \sum_{l=1}^{L} \left|\mathbf{a}_{l,k}^{H}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2$, and its PDF is given by [120]

$$f_{v_1}(v_1) = \frac{e^{-\frac{v_1}{\mu_{k,k}}}}{\mu_{k,k}}, \ v_1 > 0. \tag{4.21}$$

In addition, $v_2 = \frac{L}{N_{\text{T}}} \sum_{\substack{i=1 \\ i\neq k}}^{K} \left|\mathbf{h}_i\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2$ is a sum of independent exponential random variables with mean values of $\left\{\mu_{i,k} = \sum_{l=1}^{L} \left|\mathbf{a}_{l,i}^{H}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\right|^2\right\}_{\substack{i=1 \\ i\neq k}}^{K}$, which results in a hypo-exponential random variable with a PDF given by [120]

$$f_{v_2}(v_2) = \sum_{\substack{i=1 \\ i\neq k}}^{K} \frac{t_i e^{-\frac{v_2}{\mu_{i,k}}}}{\mu_{i,k}}, \ v_2 > 0, \tag{4.22}$$

where $t_i = \prod_{\substack{m=1 \\ m\neq i, m\neq k}}^{K} \frac{\mu_{i,k}}{\mu_{i,k}-\mu_{m,k}}$. Therefore, we can obtain the joint moment generating function $M_{v_1,v_2}(s,r)$ of v_1 and v_2 as

$$M_{v_1,v_2}(s,r) = e^{-\tilde{\delta}r} \int_0^\infty \frac{e^{-v_1\left(\frac{1}{\mu_{k,k}}-s\right)}}{\mu_{k,k}} dv_1$$
$$\times \sum_{\substack{i=1 \\ i\neq k}}^{K} \int_0^\infty \frac{t_i e^{-v_2\left(\frac{1}{\mu_{i,k}}-r\right)}}{\mu_{i,k}} dv_2$$

$$= e^{-\tilde{\delta}r} \frac{1}{1-\mu_{k,k}s} \sum_{\substack{i=1 \\ i \neq k}}^{K} \frac{t_i}{1-\mu_{i,k}r}. \tag{4.23}$$

Using (4.16), we can evaluate $\overline{\text{SLNR}}_k$ as

$$\overline{\text{SLNR}}_k = \int_0^\infty \left[\frac{\partial}{\partial s} e^{-\tilde{\delta}r} \frac{1}{1-\mu_{k,k}s} \sum_{\substack{i=1 \\ i \neq k}}^{K} \frac{t_i}{1+\mu_{i,k}r} \right]_{s=0} dr$$

$$= \mu_{k,k} \sum_{\substack{i=1 \\ i \neq k}}^{K} t_i \int_0^\infty \frac{e^{-\tilde{\delta}r}}{1+\mu_{i,k}r} dr$$

$$= -\mu_{k,k} \sum_{\substack{i=1 \\ i \neq k}}^{K} \frac{t_i}{\mu_{i,k}} e^{\frac{\tilde{\delta}}{\mu_{i,k}}} \text{Ei}\left(-\frac{\tilde{\delta}}{\mu_{i,k}}\right), \tag{4.24}$$

where $\text{Ei}(x) = -\int_{-x}^\infty \frac{e^{-t}}{t} dt$ is the exponential integral. We can also get a lower bound on $\overline{\text{SLNR}}_k$ as

$$\overline{\text{SLNR}}_k \geq \frac{\mathbb{E}\left[\frac{L}{N_\text{T}} |\mathbf{h}_k \mathbf{F}_\text{RF} \mathbf{f}_{\text{BB},k}|^2\right]}{\mathbb{E}\left[\frac{L}{N_\text{T}} \sum_{i \neq k} |\mathbf{h}_i \mathbf{F}_\text{RF} \mathbf{f}_{\text{BB},k}|^2 + \tilde{\delta}\right]}$$

$$= \frac{\sum_{l=1}^{L} \left|\mathbf{a}_{l,k}^H \mathbf{F}_\text{RF} \mathbf{f}_{\text{BB},k}\right|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \left|\mathbf{a}_{l,i}^H \mathbf{F}_\text{RF} \mathbf{f}_{\text{BB},k}\right|^2 + \tilde{\delta}} \triangleq \overline{\text{SLNR}}_k^{\text{LB}}, \tag{4.25}$$

where the inequality holds since the numerator and denominator are independent and $\mathbb{E}\left[\frac{1}{x}\right] \geq \frac{1}{\mathbb{E}[x]}$ by Jensen's inequality.

4.4.3 Optimizing Algorithm

Using $\overline{\text{SLNR}}_k^{\text{LB}}$, we obtain the baseband precoder as a function of the RF precoder. To decouple the design of the baseband precoders $\{\mathbf{f}_{\text{BB},k}\}$, we relax the power constraint to $\|\mathbf{F}_\text{RF} \mathbf{f}_{\text{BB},k}\|^2 = 1$, which also satisfies $\|\mathbf{F}_\text{RF} \mathbf{F}_\text{BB}\|_\text{F}^2 = K$. Applying the new power constraint into $\overline{\text{SLNR}}_k^{\text{LB}}$, we get

$$\overline{\text{SLNR}}_k^{\text{LB}} = \frac{\sum_{l=1}^{L} \left| \mathbf{a}_{l,k}^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \right|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \left| \mathbf{a}_{l,i}^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \right|^2 + \tilde{\delta} \left| \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \right|^2}$$

$$= \frac{\mathbf{f}_{\text{BB},k}^H \left(\mathbf{F}_{\text{RF}}^H \sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H \mathbf{F}_{\text{RF}} \right) \mathbf{f}_{\text{BB},k}}{\mathbf{f}_{\text{BB},k}^H \left(\mathbf{F}_{\text{RF}}^H \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \mathbf{a}_{l,i} \mathbf{a}_{l,i}^H \mathbf{F}_{\text{RF}} + \tilde{\delta} \mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} \right) \mathbf{f}_{\text{BB},k}}. \quad (4.26)$$

Using generalized eigenvector decomposition, we obtain $\mathbf{f}_{\text{BB},k}$ maximizing $\overline{\text{SLNR}}_k^{\text{LB}}$ as

$$\mathbf{f}_{\text{BB},k} = \frac{\boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{F}_{\text{RF}}^H \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \mathbf{a}_{l,i} \mathbf{a}_{l,i}^H \mathbf{F}_{\text{RF}} + \tilde{\delta} \mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} \right)^{-1} \mathbf{F}_{\text{RF}}^H \sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H \mathbf{F}_{\text{RF}} \right]}{\left\| \mathbf{F}_{\text{RF}} \boldsymbol{\mathcal{E}}_{\max} \left[\left(\mathbf{F}_{\text{RF}}^H \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{l=1}^{L} \mathbf{a}_{l,i} \mathbf{a}_{l,i}^H \mathbf{F}_{\text{RF}} + \tilde{\delta} \mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} \right)^{-1} \mathbf{F}_{\text{RF}}^H \sum_{l=1}^{L} \mathbf{a}_{l,k} \mathbf{a}_{l,k}^H \mathbf{F}_{\text{RF}} \right] \right\|}. \quad (4.27)$$

Now, we propose a suboptimal gradient ascent algorithm to design the RF precoder $\mathbf{F}_{\text{RF}}$. Note that $\overline{\text{SLNR}}_k$ in (4.24) is differentiable while $\overline{\text{SINR}}_k$ in (4.19) is not. Therefore, we use the gradient of $\sum_{k=1}^{K} \overline{\text{SLNR}}_k$ to maximize $R_{\text{sum}}^{\text{UB}}$. The gradient $\nabla_{\mathbf{F}_{\text{RF}}}$ of $\sum_{k=1}^{K} \overline{\text{SLNR}}_k$ with respect to the RF precoder $\mathbf{F}_{\text{RF}}$ is obtained (after mathematical manipulations) as

$$\nabla_{\mathbf{F}_{\text{RF}}} = \sum_{k=1}^{K} \sum_{\substack{i=1 \\ i \neq k}}^{K} \Bigg[\left(\frac{\mu_{k,k} \nabla_{\mathbf{F}_{\text{RF}}} (\mu_{i,k}) - \mu_{i,k} \nabla_{\mathbf{F}_{\text{RF}}} (\mu_{k,k})}{\mu_{i,k}^2} t_i - \frac{\mu_{k,k}}{\mu_{i,k}} \nabla_{\mathbf{F}_{\text{RF}}} (t_i) \right) e^{\frac{\tilde{\delta}}{\mu_{i,k}}} \text{Ei} \left(-\frac{\tilde{\delta}}{\mu_{i,k}} \right)$$

$$+ t_i \frac{\mu_{k,k}}{\mu_{i,k}} \frac{\tilde{\delta} e^{\frac{\tilde{\delta}}{\mu_{i,k}}} \text{Ei} \left(-\frac{\tilde{\delta}}{\mu_{i,k}} \right) + \mu_{i,k}}{\mu_{i,k}^2} \nabla_{\mathbf{F}_{\text{RF}}} (\mu_{i,k}) \Bigg], \quad (4.28)$$

$$\nabla_{\mathbf{F}_{\text{RF}}} (\mu_{i,k}) = \sum_{l=1}^{L} \mathbf{a}_{l,i} \mathbf{a}_{l,i}^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k} \mathbf{f}_{\text{BB},k}^H, \quad (4.29)$$

$$\nabla_{\mathbf{F}_{\text{RF}}} (t_i) = t_i \sum_{\substack{m=1 \\ m \neq i, m \neq k}}^{K} \frac{\mu_{i,k} \nabla_{\mathbf{F}_{\text{RF}}} (\mu_{m,k}) - \mu_{m,k} \nabla_{\mathbf{F}_{\text{RF}}} (\mu_{i,k})}{\mu_{i,k}^2 - \mu_{i,k} \mu_{m,k}}. \quad (4.30)$$

Using the gradient $\nabla_{\mathbf{F}_{\text{RF}}}$ in (4.28), we obtain the RF precoder $\mathbf{F}_{\text{RF}}$ by Algorithm 1, where P_{RF} is the number of iterations, and the updating rule is solved by a backtracking line search

Algorithm 1 Hybrid precoder design for average sum rate maximization

---

Initialization: Obtain $\mathbf{F}_{\mathrm{RF}}^{(0)}$ as the proposed ZF RF precoder if it is feasible. If not, obtain $\mathbf{F}_{\mathrm{RF}}^{(0)}$ randomly while satisfying the unit modulus constraint.
Then, obtain $\mathbf{F}_{\mathrm{BB}}^{(0)}$ using (4.27).
while $p \leq P$ (or any other appropriate stopping criterion)
　　Calculate the gradient $\nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}$ using (4.28)–(4.30) with $\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{BB}}^{(p)}$.
　　Updating rule:

$$
\begin{aligned}
\left[\mathbf{F}_{\mathrm{RF}}^{(p+1)}, \mathbf{F}_{\mathrm{BB}}^{(p+1)}, \alpha\right] \quad &= \quad \underset{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \alpha}{\arg\max} \; R_{\mathrm{sum}}^{\mathrm{UB}}\left(\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}\right), \\
\text{s.t.} \quad &\mathbf{F}_{\mathrm{RF}} = e^{j\angle\left(\mathbf{F}_{\mathrm{RF}}^{(p)} + \alpha\nabla_{\mathbf{F}_{\mathrm{RF}}}^{(p)}\right)}, \\
&\mathbf{F}_{\mathrm{BB}} \text{ as in (4.27).}
\end{aligned}
$$

　　$p = p + 1$.
end while
Output: $\mathbf{F}_{\mathrm{RF}}^{(P+1)}, \mathbf{F}_{\mathrm{BB}}^{(P+1)}$.

---

over the step size α [21]. We initialize the RF precoder as the proposed ZF RF precoder if it is feasible ($N_{\mathrm{T}} > (K-1)L$ and $N_{\mathrm{RF}} \geq 2K$). If not, we initialize the RF precoder randomly while satisfying the unit modulus constraint $|\mathbf{F}_{\mathrm{RF}}(m,n)| = 1 \; \forall m, n$. Note that the updating rule in Algorithm 1 is done based on $R_{\mathrm{sum}}^{\mathrm{UB}}$, which ensures an increase in $R_{\mathrm{sum}}^{\mathrm{UB}}$ in each iteration, and hence it ensures the convergence of the algorithm. Note that we have a closed-form for $R_{\mathrm{sum}}^{\mathrm{UB}}$ thanks to the closed-form SINR in (4.19).

4.5　Eigenvector-SLNR Hybrid Precoder

In this section, we present the Eigenvector-SLNR (EV-SLNR) hybrid precoder, which we consider as our benchmark design. Specifically, we develop EV-SLNR hybrid precoder based on [80, 8, 27] with some modifications to satisfy the unit modulus constraint and the power constraint.

As proposed in [80, 8], the RF precoder $\mathbf{F}_{\mathrm{RF}}$ is obtained using the principal eigenvectors of $\left\{\sum_{l=1}^{L} \mathbf{a}_{l,k}\mathbf{a}_{l,k}^{H}\right\}_{k=1}^{K}$ as

$$
\mathbf{F}_{\mathrm{RF}} = \left[e^{j\angle\boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L}\mathbf{a}_{l,1}\mathbf{a}_{l,1}^{H}\right]}, \ldots, e^{j\angle\boldsymbol{\mathcal{E}}_{\max}\left[\sum_{l=1}^{L}\mathbf{a}_{l,K}\mathbf{a}_{l,K}^{H}\right]}\right], \tag{4.31}
$$

where we use only the phases of the principal eigenvectors to satisfy the unit modulus constraint. As proposed in [27], the baseband precoder $\mathbf{f}_{\text{BB},k}$ is designed to maximize $\overline{\text{SLNR}}_k^{\text{LB}}$. In [27], the power constraint $\|\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},k}\|^2 = 1$ is applied after obtaining $\mathbf{f}_{\text{BB},k}$. On the contrary, we apply the power constraint into $\overline{\text{SLNR}}_k^{\text{LB}}$ before obtaining $\mathbf{f}_{\text{BB},k}$. Therefore, our closed-form baseband precoder in (4.27) is more accurate than that in [27]. The main advantage of EV-SLNR hybrid precoder is that the RF precoder and the baseband precoder can be written in closed-forms as in (4.27) and (4.31). However, EV-SLNR hybrid precoder utilizes only K RF chains out of the available N_{RF} RF chains, which results in some performance loss.

4.6 Numerical Results

We evaluate the performance of the proposed hybrid precoder designs and compare them with the performance of EV-SLNR hybrid precoder by means of Monte-Carlo simulations. Note that the proposed SINR-SLNR hybrid precoder is different from EV-SLNR hybrid precoder in the design of the RF precoder. EV-SLNR hybrid precoder utilizes only K RF chains out of the available N_{RF} RF chains. On the contrary, the proposed SINR-SLNR hybrid precoder utilizes the available N_{RF} RF chains.

Regarding the simulation setup, we assume that the transmitter has 20 antennas ($N_{\text{T}} = 20$), and we have 4 users ($K = 4$). The number of RF chains N_{RF} will be an adjustable parameter. All channels follow the mmWave channel model described in subsection 4.2.2 with 6 propagation paths ($L = 6$), where the channel gains $\{\alpha_{l,k}\}$ are zero-mean and unit-variance complex Gaussian random variables, and the angles of departure $\{\varphi_{l,k}\}$ are uniformly-distributed within $[0 \ 2\pi)$.

Fig. 4.2 shows the average SINR ($\frac{1}{K}\sum_{k=1}^{K}\overline{\text{SINR}}_k$) and the average SLNR ($\frac{1}{K}\sum_{k=1}^{K}\overline{\text{SINR}}_k$) versus transmit SNR γ with $N_{\text{RF}} = 4$. We observe that the closed-form expressions in (4.19) and (4.24) exactly match the corresponding values obtained by simulation, which verifies our
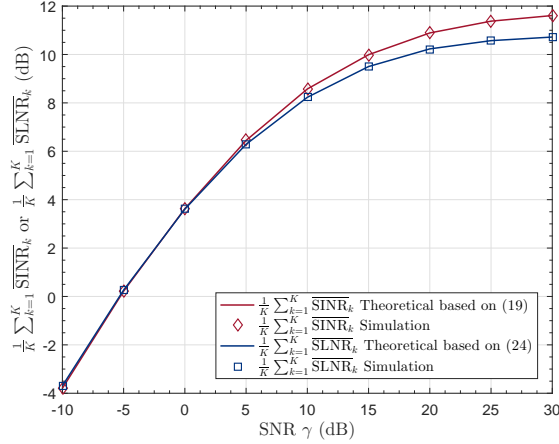
Figure 4.2. Verification of our closed-form expressions of SINR in (4.19) and SLNR in (4.24) with values obtained by simulation with $N_{\mathrm{RF}} = 4$.

closed-form expressions. We also observe that the average SLNR gives a good approximation to the average SINR, which enables us to use the gradient in (4.28) of the average SLNR to maximize the sum rate (as used in Algorithm 1).

Fig. 4.3 shows the achievable sum rate R_{sum} as a function of the transmit SNR γ with different numbers of RF chains. With $N_{\mathrm{RF}} = 4$, the proposed ZF hybrid precoder is infeasible. On the other hand, the proposed SINR-SLNR hybrid precoder and EV-SLNR hybrid precoder are feasible. The proposed SINR-SLNR hybrid precoder outperforms EV-SLNR hybrid precoder specifically at moderate and high SNRs. The sum-rate gap between the proposed SINR-SLNR hybrid precoder and EV-SLNR hybrid precoder increases as we increase N_{RF} to 6 and 8. The reason is that EV-SLNR hybrid precoder utilizes only K RF chains out of the available N_{RF} RF chain, while the proposed SINR-SLNR hybrid precoder utilizes the available N_{RF} RF chains. With $N_{\mathrm{RF}} = 8$, the proposed ZF hybrid precoder is feasible, and it outperforms EV-SLNR at high SNRs since nulling the interference is the optimal strategy at high SNRs. However, the proposed ZF hybrid precoder achieves the lowest sum rate at low SNRs since the system is power-limited and nulling the interference is not the optimal strategy at low SNRs. We also observe that with the same number of
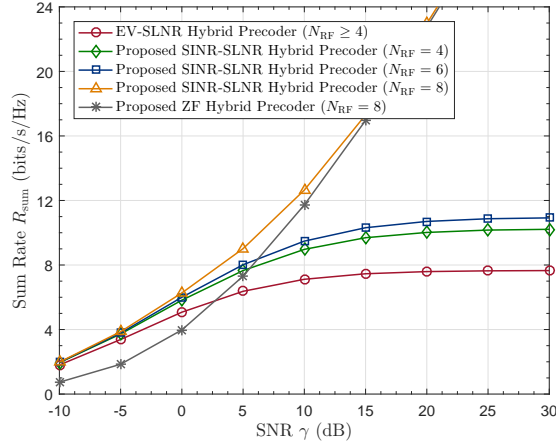
Figure 4.3. Achievable sum rate R_{sum} as a function of transmit SNR γ with $K = 4$ and different numbers of N_{RF}.

RF chains the proposed SINR-SLNR hybrid precoder outperforms the proposed ZF hybrid precoder at low and moderate SNRs on the expense of higher computational complexity.

4.7 Conclusion

For mmWave multiuser downlink systems, the use of hybrid precoders and partial channel knowledge is more practical than that of fully digital precoders and/or full channel knowledge. We developed three hybrid precoders (ZF, SINR-SLNR, and EV-SLNR) based on partial channel knowledge in contrast to the full channel knowledge required in the existing approaches. The interference nulling of ZF hybrid precoder is feasible with partial channel knowledge if necessary conditions on the numbers of transmit antennas and RF chains are satisfied, which is more demanding than that with full channel knowledge. For the scenarios not satisfying the necessary conditions of ZF, we proposed SINR-SLNR hybrid precoder. We also developed EV-SLNR hybrid precoder (a modification of the existing approaches to fit to the partial channel knowledge scenario) as a benchmark. When ZF is not feasible, there will be interference and hence the sum-rate performance of SINR-SLNR and EV-SLNR hybrid precoders would not linearly grow with SNR at high SNRs. However, the SINR-SLNR hybrid precoder outperforms the EV-SLNR hybrid precoder and its sum-rate performance is also

enhanced by increasing the number of RF chains while it is not the case for the EV-SLNR hybrid precoder. When ZF is feasible, the sum-rate performance of the ZF hybrid precoder is substantially better at high SNRs, but worse at low SNRs than those of the other schemes. With the same number of RF chains, the proposed SINR-SLNR hybrid precoder yields better sum-rate performance than all the other schemes on the expense of higher computational complexity.

Appendix A - Evaluating the Integral $I_{k,l}$ in (4.19)

Defining $\mu_{k,l} = \frac{1}{\lambda_{k,l}}$, the integral $I_{k,l}$ in (4.19) can be written as $I_{k,l} = \mu_{k,l} \left(\prod_{m=1}^{M} \mu_{k,m} \right) \tilde{I}_{k,l}$, where

$$
\begin{aligned}
\tilde{I}_{k,l} &= \int_0^\infty \frac{e^{-\tilde{\delta}r} dr}{(r + \mu_{k,l}) \prod_{m=1}^{M} (r + \mu_{k,m})} \\
&= \int_0^\infty \left(\frac{a_{k,l} e^{-\tilde{\delta}r}}{(r + \mu_{k,l})^2} + \frac{b_{k,l} e^{-\tilde{\delta}r}}{(r + \mu_{k,l})} + \sum_{m=1, m \neq l}^{M} \frac{c_{k,m} e^{-\tilde{\delta}r}}{(r + \mu_{k,m})} \right) dr \\
&= a_{k,l} \left(\tilde{\delta} e^{\mu_{k,l}\tilde{\delta}} \mathrm{Ei} \left(-\mu_{k,l}\tilde{\delta} \right) + \frac{1}{\mu_{k,l}} \right) - b_{k,l} e^{\mu_{k,l}\tilde{\delta}} \mathrm{Ei} \left(-\mu_{k,l}\tilde{\delta} \right) \\
&\quad - \sum_{m=1, m \neq l}^{M} c_{k,m} e^{\mu_{k,m}\tilde{\delta}} \mathrm{Ei} \left(-\mu_{k,m}\tilde{\delta} \right),
\end{aligned}
\tag{4.32}
$$

where $\mathrm{Ei}(x) = -\int_{-x}^\infty \frac{e^{-t}}{t} dt$ is the exponential integral, and the partial fraction coefficients are obtained as $a_{k,l} = \frac{1}{\prod_{n=1, n\neq l}^{M} (\mu_{k,n} - \mu_{k,l})}$, $b_{k,l} = \left[\frac{\partial}{\partial r} \frac{1}{\prod_{n=1, n\neq l}^{M} (r + \mu_{k,n})} \right]_{r=-\mu_{k,l}}$, and $c_{k,m} = \frac{1}{(\mu_{k,l} - \mu_{k,m}) \prod_{n=1, n\neq m}^{M} (\mu_{k,n} - \mu_{k,m})}$.

PART II

LOW-COST NONLINEAR TERA-HERTZ TRANSMITTERS

CHAPTER 5

PRECOMPENSATION AND SYSTEM PARAMETERS ESTIMATION FOR LOW-COST NONLINEAR TERA-HERTZ TRANSMITTERS IN THE PRESENCE OF I/Q IMBALANCE [1]

5.1 Introduction

Signal generation circuitries for low-cost THz devices are different from the conventional ones of the lower frequency bands. This is due to unavailability of THz oscillators and THz CMOS power amplifiers which is commonly known as "the THz Gap" [102]. Thus, low-cost THz transceivers rely on nonlinear devices in contrast to the linear devices of the lower frequency bands. Another crucial limitation of low-cost THz devices is the output power constraint which has a direct effect on the sensing/communication range and performance.

There are a few low-cost state-of-the-art CMOS THz transmitter architectures in the recent literature. The two most promising ones are the so-called frequency-multiplier-last architecture from the UC-Berkeley [69, 68, 98] and the cubic mixer architecture from Japan [70, 71]. The former [69, 68] has an advantage of 14.5 dB higher output power and 8 dB lower DC power consumption than the latter. However, the frequency-multiplier-last architecture [69, 68, 98] is not capable of transmitting quadrature amplitude modulation (QAM) schemes while the mixer architecture [70, 71] is. The research group of the latter also developed a doubler mixer architecture in [124] which yields a higher output power due to the use of doubler rather than tripler. This chapter focuses on the frequency-multiplier-last architecture since it has substantial advantages in terms of the transmitter output power and DC power consumption which are much needed to address the propagation range limitation and energy

---

efficiency. The major limitations/challenges of the frequency-multiplier-last architecture reported in the recent literature are its incapability with QAM transmission and its undesired spectrum spreading, both due to nonlinear distortions. These nonlinear distortions impose new fundamental challenges for designing reliable and efficient communications systems in the nonlinear regime.

If compared to the existing communication systems, THz communication introduces several challenges including communication range limitation, distance and frequency dependent channel characteristics, and more difficult synchronization, equalization and distortion compensation [2]. When low-cost constraint (i.e., CMOS device) is imposed, the most energy-efficient transmitter architecture in the literature [69, 68] causes an additional challenge of nonlinear distortion to the message signals which prevents reliable transmission of QAM signals. Another source of distortion in low-cost THz transmitters is the inphase and quadrature (I/Q) imbalance. Ideally, the inphase (I) and quadrature (Q) branches of the mixers should have equal amplitude and 90° phase difference. However, this is rarely the case in practice, resulting in I/Q imbalance. Communication theory in such nonlinear systems in the presence of I/Q imbalance has not been investigated in the literature and we address it in this chapter.

In [22], a precompensation scheme was proposed for nonlinear power amplifiers (PAs) in the absence of I/Q imbalance. In [10, 148, 74] and the references therein, closed-feedback polynomial-based precompensation schemes were developed for nonlinear PAs in the presence of I/Q imbalance. Due to the closed-feedback nature, these schemes do not require the knowledge of the nonlinear device (NLD) parameters and the I/Q imbalance parameters. Unlike the case of nonlinear PAs, the closed-feedback polynomial-based precompensation is not applicable to nonlinear frequency multipliers in the presence of I/Q imbalance. This fact is due to the third-order and fifth-order powers relationship (for frequency tripler) between the input and the output of the frequency multiplier. Therefore, the NLD parameters and

the I/Q imbalance parameters should be estimated first, and then be used to precompensate the transmitted signal, but this problem has not been addressed in the literature.

Motivated by the challenges mentioned above, we study such nonlinear THz communication system in the presence of I/Q imbalance. Our main contributions are summarized as follows:

- An accurate signal model is derived incorporating both the nonlinearity aspects of the low-cost THz devices and the I/Q imbalance effect.

- We show that without precompensation and with pulse-shaping filter spanning more than one symbol, the transmitter output constellation experiences significant distortions in the absence or presence of I/Q imbalance. In addition, the existing frequency-multiplier-last architecture is not capable of transmitting QAM.

- A precompensation scheme is proposed to compensate the nonlinearity and I/Q imbalance effects and enable low-cost THz QAM transmission. The proposed precompensation scheme requires the knowledge of the NLD parameters and the I/Q imbalance parameters.

- Due to the nonlinearity of the frequency multiplier, we show that the I/Q imbalance parameters cannot be estimated unless the following two conditions are satisfied. First, the pilot symbols are inter-symbol-interference-free. Second, the first half of the pilot sequence is real only while the second half is imaginary only, or the first half of the pilot sequence is imaginary only while the second half is real only.

- Using a separate measurement circuitry for testing/calibrating the THz transmitters, we propose a maximum-likelihood (ML) estimator and its practical implementation to estimate the NLD parameters and the I/Q imbalance parameters based on two phases of measurements.

- We derive closed-form expressions for the Cramér–Rao lower bounds (CRLBs) of the system parameters estimates as benchmark metrics to evaluate the performance of the proposed ML estimator.

- A pilot sequence design is developed for testing/calibrating the THz transmitters to enhance the performance of the proposed ML estimator.

- We present performance characteristics of the proposed precompensation and parameters estimation schemes which show efficient handling of severe nonlinear distortions and spectrum spreading issues of the low-cost frequency-multiplier-last THz transmitter, and mitigation of the I/Q imbalance effect.

The rest of the chapter is organized as follows. In section II, we describe the system model and derive an accurate signal model. In section III, we show the effects of the pulse-shaping filter and the I/Q imbalance on the signal spectrum and constellation. In section IV, the proposed precompensation scheme is introduced. In section V, we describe the proposed system parameters estimation method. In section VI, we derive the CRLBs for system parameters estimation, and design the pilot sequence used to estimate the system parameters. Numerical results are presented in section VII. Finally, section VIII concludes the chapter.

5.2   System and Signal Model

To investigate the effect of nonlinear THz transmitters in the presence of I/Q imbalance, we consider a single-antenna THz communication system. The extension to multi-antenna systems can be straightly done. The transmitter utilizes the frequency-multiplier-last architecture [69, 68] for an in-phase (I) and quadrature (Q) transmission as shown in Fig. 5.1.
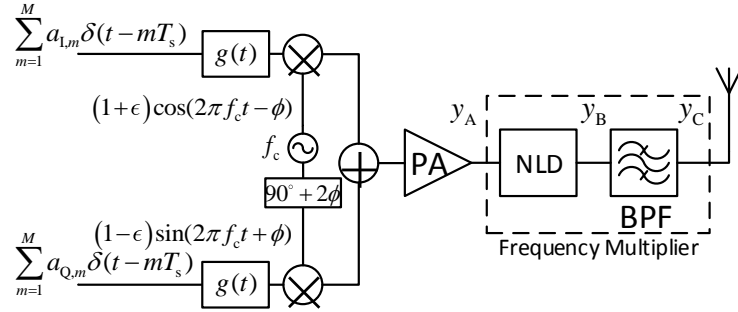
Figure 5.1. Frequency-multiplier-last transmitter architecture for I/Q transmission [69, 68] in the presence of I/Q imbalance. (BPF = bandpass filter, PA = power amplifier, NLD = nonlinear device)

The frequency multiplier block is implemented by a nonlinear device (NLD) followed by a bandpass filter (BPF), and it is the source of nonlinear distortion. Ideally, the I and Q branches of the mixers should have equal amplitude and 90° phase difference. However, this is rarely the case in practice, resulting in I/Q imbalance.

For a block of M modulation symbols, the I and Q branches can be given by $s_{\mathrm{I}}(t) = \sum_{m=1}^{M} a_{\mathrm{I},m}\delta(t - mT_{\mathrm{s}})$ and $s_{\mathrm{Q}}(t) = \sum_{m=1}^{M} a_{\mathrm{Q},m}\delta(t - mT_{\mathrm{s}})$ where $a_m \triangleq a_{\mathrm{I},m} + ja_{\mathrm{Q},m}$ is a modulation symbol[2], T_{s} is the symbol duration, and $R_{\mathrm{s}} = 1/T_{\mathrm{s}}$ is the symbol rate. With a pulse-shaping filter $g(t)$, the baseband signal $\tilde{s}(t) = \tilde{s}_{\mathrm{I}}(t) + j\tilde{s}_{\mathrm{Q}}(t) = s(t) * g(t)$ is given by

$$\tilde{s}(t) = \sum_{m=1}^{M} a_m g(t - mT_{\mathrm{s}}). \tag{5.1}$$

Then, the output of the non-ideal mixer (in the presence of I/Q imbalance) with frequency f_{c} is given by [81]

$$y_{\mathrm{A}}(t) = (1 + \epsilon)\,\tilde{s}_{\mathrm{I}}(t)\cos(2\pi f_{\mathrm{c}}t - \phi)$$

$$- (1 - \epsilon)\,\tilde{s}_{\mathrm{Q}}(t)\sin(2\pi f_{\mathrm{c}}t + \phi), \tag{5.2}$$

---

[2]For multicarrier modulation, $\{a_m\}$ should be the complex time-domain output samples of the inverse discrete Fourier transform (IDFT) block, and M should be the IDFT size.

where ϵ and ϕ are the amplitude imbalance and phase imbalance respectively between I and Q branches. After some manipulations, (5.2) can be rewritten as

$$y_{\mathrm{A}}(t) = \bar{s}_{\mathrm{I}}(t) \cos(2\pi f_{\mathrm{c}} t) - \bar{s}_{\mathrm{Q}}(t) \sin(2\pi f_{\mathrm{c}} t), \tag{5.3}$$

where $\bar{s}(t) = \bar{s}_{\mathrm{I}}(t) + j\bar{s}_{\mathrm{Q}}(t)$, and $\bar{s}(t)$ is related to $\tilde{s}(t)$ as [129]

$$\bar{s}(t) = \mu\tilde{s}(t) + \nu\tilde{s}^*(t), \tag{5.4}$$

where

$$\mu = \cos(\phi) - j\epsilon\sin(\phi), \tag{5.5}$$

$$\nu = \epsilon\cos(\phi) - j\sin(\phi). \tag{5.6}$$

We consider a memoryless polynomial model for the NLD (i.e., for an input x, the output is $\sum_{b=1}^{B} A_b x^b$). [3] The larger order terms typically have smaller coefficients and their spectrum spreadings are wider. Furthermore, larger order terms may introduce some floor of distortions. The NLD used in the THz transmitter typically has a differential output, and hence even order terms will be canceled. Therefore, the nonlinear device output signal $y_{\mathrm{B}}(t)$ is given by[4]

$$\begin{aligned}
y_{\mathrm{B}}(t) = & A_1 \left(\bar{s}_{\mathrm{I}}(t) \cos(2\pi f_{\mathrm{c}} t) - \bar{s}_{\mathrm{Q}}(t) \sin(2\pi f_{\mathrm{c}} t)\right) \\
& + A_3 \left(\bar{s}_{\mathrm{I}}(t) \cos(2\pi f_{\mathrm{c}} t) - \bar{s}_{\mathrm{Q}}(t) \sin(2\pi f_{\mathrm{c}} t)\right)^3 \\
& + A_5 \left(\bar{s}_{\mathrm{I}}(t) \cos(2\pi f_{\mathrm{c}} t) - \bar{s}_{\mathrm{Q}}(t) \sin(2\pi f_{\mathrm{c}} t)\right)^5 + \dots.
\end{aligned} \tag{5.7}$$

---

[3]The hardware implementation of the frequency-multiplier-last transmitter in [69] implements wideband amplifiers, frequency multiplier, and antennas. Therefore, the power amplifier and the frequency multiplier have approximately flat responses. The sensitivity of the proposed solutions to frequency-selective NLDs and frequency-dependent I/Q imbalance is investigated in section VII.

[4]With approximately flat responses for power amplifier and frequency multiplier, if the level of the input signal to the power amplifier is not within the linear region of the power amplifier, the power amplifier will introduce nonlinear distortion. However, since the two nonlinear blocks (the power amplifier and the frequency multiplier) are cascaded, the nonlinearity effect of the two nonlinear blocks can be modeled by a single nonlinear polynomial model as in (5.7).

If we consider the desired signal to be centered around a THz carrier $f_\mathrm{T} = 3f_\mathrm{c}$, it will be typically contributed mainly by third-order power $(\cdot)^3$ and partially by fifth-order power $(\cdot)^5$. The last BPF allows the desired signal centered around the THz carrier f_T and suppresses other terms, yielding $y_\mathrm{C}(t)$ as

$$y_\mathrm{C}(t) = \underbrace{\left[\frac{A_3}{4} \left(\bar{s}_\mathrm{I}^3(t) - 3\bar{s}_\mathrm{I}(t)\bar{s}_\mathrm{Q}^2(t) \right) + \frac{5A_5}{16} \left(\bar{s}_\mathrm{I}^5(t) - 3\bar{s}_\mathrm{I}(t)\bar{s}_\mathrm{Q}^4(t) - 2\bar{s}_\mathrm{I}^3(t)\bar{s}_\mathrm{Q}^2(t) \right) \right]}_{x_\mathrm{BB,I}(t)} \cos(2\pi f_\mathrm{T}t)$$

$$- \underbrace{\left[\frac{A_3}{4} \left(3\bar{s}_\mathrm{I}^2(t)\bar{s}_\mathrm{Q}(t) - \bar{s}_\mathrm{Q}^3(t) \right) + \frac{5A_5}{16} \left(2\bar{s}_\mathrm{I}^2(t)\bar{s}_\mathrm{Q}^3(t) + 3\bar{s}_\mathrm{I}^4(t)\bar{s}_\mathrm{Q}(t) - \bar{s}_\mathrm{Q}^5(t) \right) \right]}_{x_\mathrm{BB,Q}(t)} \sin(2\pi f_\mathrm{T}t). \quad (5.8)$$

The equivalent baseband signal of $y_\mathrm{C}(t)$ is $x_\mathrm{BB}(t) = x_\mathrm{BB,I}(t) + jx_\mathrm{BB,Q}(t)$. By defining $\tilde{A}_3 = A_3/4$ and $\tilde{A}_5 = 5A_3/16$ and grouping the terms in (5.8) carefully, $x_\mathrm{BB}(t)$ is simplified into

$$x_\mathrm{BB}(t) = \tilde{A}_3\bar{s}^3(t) + \tilde{A}_5\bar{s}^3(t)\left|\bar{s}(t)\right|^2,$$

$$= \tilde{A}_3 \left(\mu\tilde{s}(t) + \nu\tilde{s}^*(t)\right)^3 + \tilde{A}_5\left(\mu\tilde{s}(t) + \nu\tilde{s}^*(t)\right)^3$$

$$\times \left|\mu\tilde{s}(t) + \nu\tilde{s}^*(t)\right|^2. \quad (5.9)$$

From (5.9), we can observe that I/Q imbalance and nonlinearity of the frequency multiplier cause a specific form of nonlinear distortion to the modulation symbols as well as spectral spreading and larger spectrum sidelobes. The I/Q imbalance modifies the baseband signal $\tilde{s}(t)$ into $\bar{s}(t)$, and then the nonlinearity modifies the baseband signal $\bar{s}(t)$ into $x_\mathrm{BB}(t)$. We can also write the output of the frequency multiplier $y_\mathrm{C}(t) = \Re\left\{x_\mathrm{BB}(t)e^{j2\pi f_\mathrm{T}t}\right\}$ as

$$y_\mathrm{C}(t) = \Re\left\{ \tilde{A}_3\left(\sum_{m=1}^{M} (\mu a_m + \nu a_m^*)g(t - mT_\mathrm{s}) \right)^3 \right.$$

$$+ \tilde{A}_5\left(\sum_{m=1}^{M} (\mu a_m + \nu a_m^*)g(t - mT_\mathrm{s}) \right)^3$$

$$\left. \times \left| \sum_{m=1}^{M} (\mu a_m + \nu a_m^*)g(t - mT_\mathrm{s}) \right|^2 \right\} e^{j2\pi f_\mathrm{T}t} \right\}. \quad (5.10)$$

78

Equation (5.10) shows exactly how the I/Q imbalance parameters (μ and ν), the frequency multiplier nonlinearity parameters ($\tilde{A}_3$ and $\tilde{A}_5$), and the pulse-shaping filter $g(t)$ affect the modulation symbols $\{a_m\}$ and the THz transmitted signal.

5.3 Pulse Shaping and I/Q Imbalance Effects

An important entity which influences the nonlinear distortions of the frequency-multiplier architecture is the transmit baseband pulse-shaping filter before the frequency multiplier. For low-cost THz transmitters, the filtering stage after the frequency multiplier circuit is implemented by means of on-chip connection line and on-chip antenna, and hence the filtering performance at that stage is rather loose. Thus, the role of the baseband pulse-shaping filter is more prominent for controlling output power spectrum. Differences from the existing lower band systems are complications due to nonlinearity of the frequency multiplier and the loose filtering after the frequency multiplier.

Recent literature demonstrated feasibility of quadrature phase-shift keying (QPSK) in such nonlinear systems [69, 68] where the message points in the QPSK constellation just experience a predefined permutation. However, these results are valid only in the absence of I/Q imbalance and if the baseband pulse-shaping filter impulse response is limited within one symbol interval (T_s) which causes high levels of spectrum sidelobes. To keep adjacent channel interferences at an acceptable level, the spectrum sidelobes need to be substantially lowered which requires the use of a larger span of the pulse-shaping filter.

To illustrate this, for QPSK modulation with normalized input energy $E_\mathrm{norm} = 20$ dBm to the NLD, Fig. 5.2 shows the power spectrum density (PSD) of the signals before and after the frequency multiplier and the transmitter output constellation in the absence (Fig. 5.2a)/presence (Fig. 5.2b) of I/Q imbalance and for two different pulse-shaping filters: 1) rectangular pulse-shaping filter spanning one symbol and 2) root-raised cosine (RRC) pulse-shaping filter spanning 12 symbols. Here, the NLD contains both the third-order and

fifth-order power terms. On the other hand, we present the results for only the third-order power term of the NLD in the absence of I/Q imbalance in [106]. From the results of Fig. 5.2a and those of [106], we observe that a larger filter span yields a lower spectrum sidelobe but a severe degradation of the transmitter output constellation due to nonlinear distortion. With the rectangular pulse-shaping filter, the third-order power of the NLD results in only a predefined permutation to the QPSK constellation points. On the other hand, with the RRC pulse-shaping filter, the third-order power of the NLD results in a severe distortion to the QPSK constellation. The fifth-order power of the NLD results in a slight distortion to the transmitter output constellation for both pulse-shaping filters.

In the presence of I/Q imbalance (Fig. 5.2b), we observe that the I/Q imbalance adds additional distortions to the transmitter output constellation for both pulse-shaping filters. These numerical results show that new communication strategies are needed for both QPSK and QAM signals for energy and spectrum efficient low-cost THz systems. In the next section, we propose a precompensation scheme to mitigate the distortions caused by the I/Q imbalance and the nonlinearity of the frequency multiplier.

5.4   Proposed Precompensation Scheme

Since the I/Q imbalance and nonlinear distortion are caused by two different blocks (i.e., I/Q imbalance is caused by the non-ideal mixer and nonlinear distortion is caused by the frequency multiplier), we propose to precompensate each block separately. However, this method requires the knowledge of I/Q imbalance parameters (ϵ and ϕ) and the NLD parameters ($\tilde{A}_3$ and $\tilde{A}_5$). In this section, we propose a precompensation scheme assuming that these system parameters have been already estimated. In the next section, we will describe the estimation method of these parameters.

PSD before NLD PSD after NLD Output constellation after NLD

Rectangular pulse (T_s)

RRC pulse ($12T_s$, $\beta_s = 0.332$)

(a) In the absence of I/Q imbalance

PSD before NLD PSD after NLD Output constellation after NLD

Rectangular pulse (T_s)

RRC pulse ($12T_s$, $\beta_s = 0.332$)

(b) In the presence of I/Q imbalance
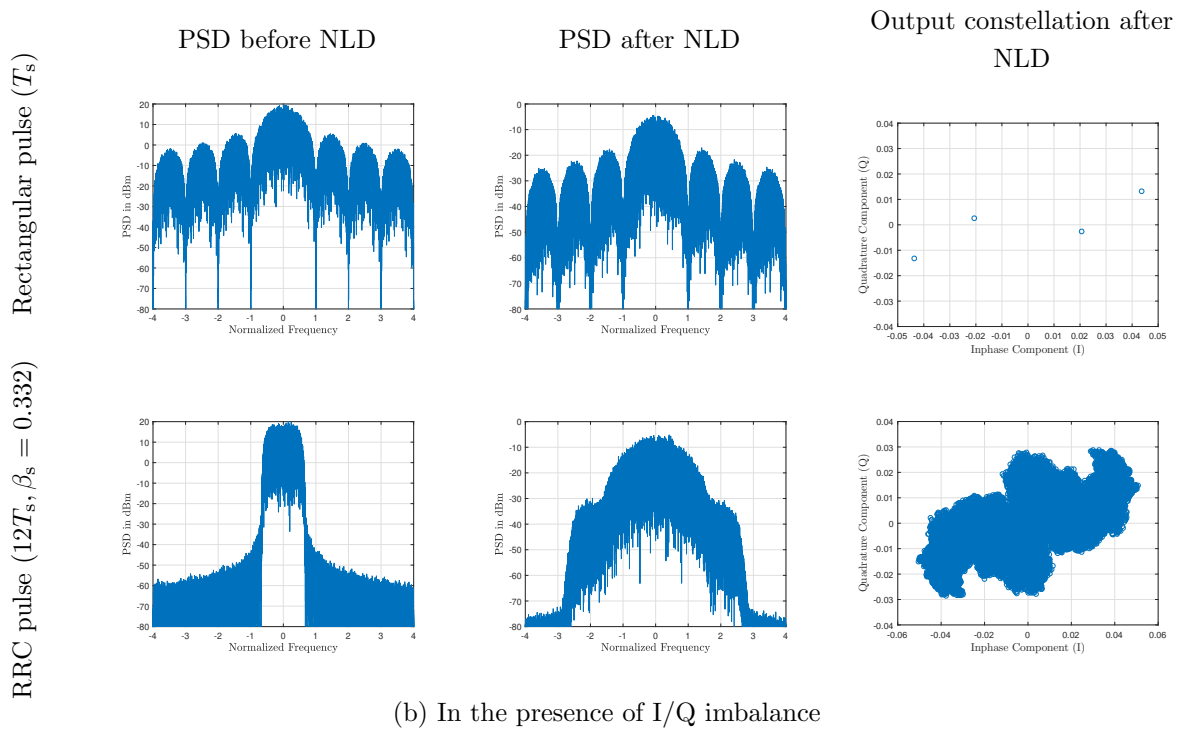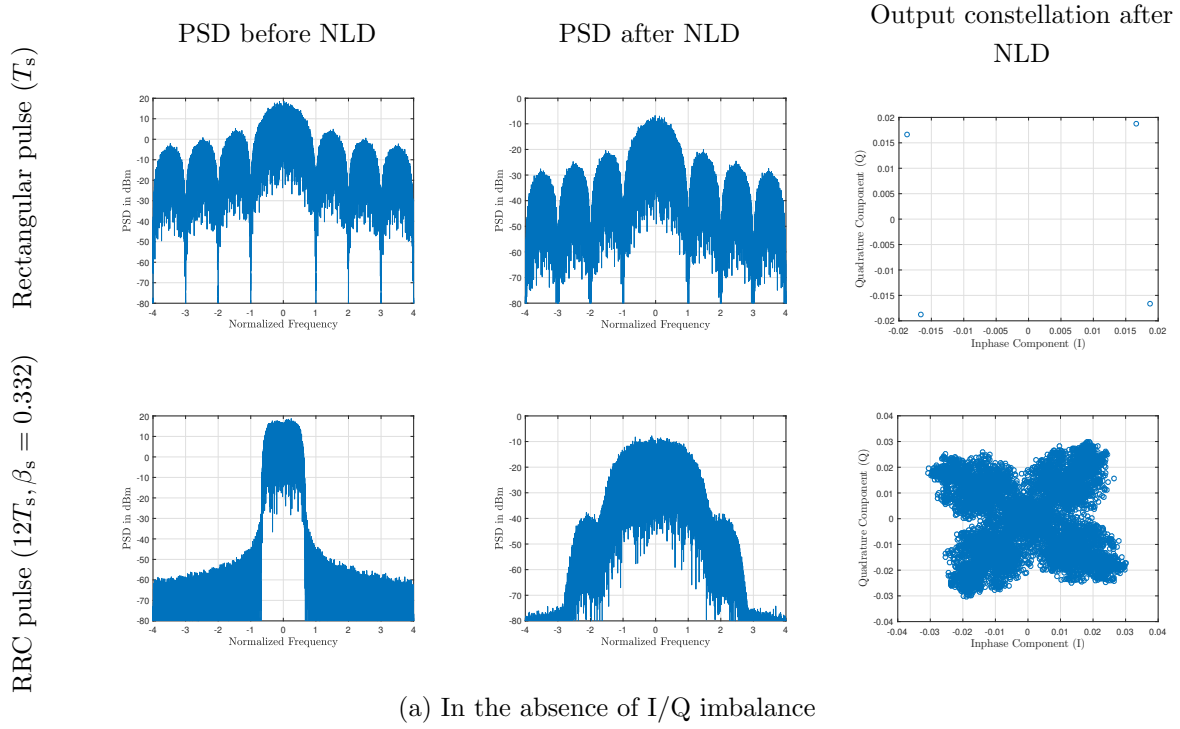
Figure 5.2. Combined effects of the baseband pulse-shaping filter and the NLD on QPSK (the existing scheme in [69, 68]) in the absence/presence of I/Q imbalance. (β_s = roll-off factor, $E_{\mathrm{norm}} = 20$ dBm, $A_3 = 1.26 - 0.081j$, $A_5 = -0.96 + 0.29j$, $\epsilon = 20\%$, $\phi = 10°$)

5.4.1 Precompensator Setup

For the desired baseband signal $\tilde{s}(t)$ of (5.1), our design seeks to make the output of the frequency multiplier as close as possible to $\sqrt{\gamma}\tilde{s}(t)$ where γ is a scaling factor to be determined to satisfy a normalized energy requirement E_{norm} at the output of the power amplifier. First, we start with the precompensation of the frequency multiplier in the absence of the I/Q imbalance. Then, we modify the precompensated signal to compensate the I/Q Imbalance in subsection 5.4.4.

Suppose that $x_{\mathrm{pre}}(t)$ is the baseband signal after precompensating the nonlinearity effect of the frequency multiplier (i.e., in Fig. 5.1, $y_{\mathrm{A}} = x_{\mathrm{pre}}(t)$). Then, we design precompensation such that its output signal $x_{\mathrm{pre}}(t)$ is given by

$$\underset{x_{\mathrm{pre}}(t)}{\arg\min} \int_{T_{\mathrm{B}}} \left| \sqrt{\gamma}\tilde{s}(t) - \tilde{A}_3 x_{\mathrm{pre}}^3(t) - \tilde{A}_5 x_{\mathrm{pre}}^3(t) \left| x_{\mathrm{pre}}(t) \right|^2 \right|^2 dt, \tag{5.11}$$

where T_{B} is the time span of the signal block. Using a highly enough sampling rate R_{c} which is L times the symbol rate R_{s}, the discrete-time implementation of (5.11) can be given as

$$\underset{\{c_k\}}{\arg\min} \sum_{k=1}^{K} \left| \tilde{s}_k - \tilde{A}_3 c_k^3 - \tilde{A}_5 c_k^3 |c_k|^2 \right|^2, \tag{5.12}$$

where $\tilde{s}_k = \sqrt{\gamma}\tilde{s}(kT_{\mathrm{c}})$, $c_k = x_{\mathrm{pre}}(kT_{\mathrm{c}})$, and $K = LM$. Then, the solution samples $\{c_k\}$ can be used to construct $x_{\mathrm{pre}}(t)$ as

$$x_{\mathrm{pre}}(t) = \sum_{k=1}^{K} c_k q(t - kT_{\mathrm{c}}), \tag{5.13}$$

where $q(t)$ is the interpolation (construction) filter. In order to maintain zero error of construction (zero integrand) at time instants $\{kT_{\mathrm{c}}\}$ in (5.11), $q(t)$ should satisfy Nyquist's zero inter-symbol interference criterion. Furthermore, $q(t)$ should preserve $\{c_k\}$ and also be independent of $\{c_k\}$. Thus, we propose to obtain $q(t)$ as a truncated raised cosine (RC) filter given by [4]

$$q(t) = \frac{\mathrm{sinc}\left(\frac{t}{T_{\mathrm{c}}}\right)\cos\left(\frac{\pi\beta_{\mathrm{c}}t}{T_{\mathrm{c}}}\right)}{1 - \left(\frac{2\beta_{\mathrm{c}}t}{T_{\mathrm{c}}}\right)^2}, \quad -\frac{QT_{\mathrm{c}}}{2} \le t \le \frac{QT_{\mathrm{c}}}{2}, \tag{5.14}$$

where $\text{sinc}(x) = \sin(\pi x)/\pi x$, β_c is the roll-off factor to be determined, and QT_c is the time span.

In summary, our approach based on (5.12) will generate K samples $\{c_k\}$ which will be converted to the continuous-time signal $x_\text{pre}(t)$ as in (5.13). However, the optimization problem in (5.12) is nonconvex and does not have a closed-form solution. Thus, we first determine the scaling factor γ in subsection 5.4.2. Then, we obtain the samples $\{c_k\}$ that solve (5.12) in subsection 5.4.3.

5.4.2 Determining the Scaling Factor γ

The normalized energy of the precompensated signal $x_\text{pre}(t)$ is always higher than that of the desired output signal $\tilde{s}(t)$ due to the nonlinear relationship between the input and the output of the frequency multiplier. Therefore, to keep the normalized energy of the precompensated signal $x_\text{pre}(t)$ at E_norm, the desired signal $\tilde{s}(t)$ should be scaled properly as $\sqrt{\gamma}\tilde{s}(t)$ such that its precompensated signal has approximately a normalized energy E_norm. Note that scaling the precompensated signal (after it is obtained) distorts the transmit output constellation. This fact is due to the nonlinear relationship between the input and output of the frequency multiplier. We propose to ignore the contribution of the fifth-order power of the NLD and scale the desired signal $\tilde{s}(t)$ such that its precompensated signal has a normalized energy E_norm.

By ignoring the contribution of the fifth-order power of the NLD, the virtual sample c_k is given by $c_k = (\sqrt{\gamma}\tilde{s}(kT_\text{c})/\tilde{A}_3)^{1/3}$. The normalized energy $E_{\text{norm},x_\text{pre}}$ of $x_\text{pre}(t)$ can be approximated as

$$E_{\text{norm},x_\text{pre}} = \frac{1}{T_\text{B}} \int_{T_\text{B}} |\sum_{k=1}^{K} c_k q(t - kT_\text{c})|^2 dt$$

$$\approx (\frac{1}{K} \sum_{k=1}^{K} |c_k|^2) \frac{1}{T_\text{c}} \int_{T_\text{B}} |q(t)|^2 dt$$

83

$$\approx \frac{\gamma^{1/3}}{|\tilde{A}_3|^{2/3}} \Big(\frac{1}{M} \sum_{m=1}^{M} |a_m|^2 \Big)^{1/3} \frac{1}{T_c} \int_{T_B} |q(t)|^2 dt$$

$$= \Big(\frac{\gamma}{|\tilde{A}_3|^2} \Big)^{1/3} E_{\text{QAM}}^{1/3} \frac{1}{1 - \frac{\beta_c}{4}}, \tag{5.15}$$

where $(1/T_c) \int_{T_B} q^2(t) dt = 1/(1 - \beta_c/4)$ for the raised-cosine construction filter $q(t)$ of (5.14) if $T_B \gg T_c$ [65], and $E_{\text{QAM}} = \sum_m |a_m|^2/M$ is the average symbol energy of the QAM symbols. From (5.15), we note that γ should be set as

$$\gamma = \Big(1 - \frac{\beta_c}{4} \Big)^3 \frac{E_{\text{norm}}^3 |\tilde{A}_3|^2}{E_{\text{QAM}}}, \tag{5.16}$$

to have $E_{\text{norm},x_{\text{pre}}} \approx E_{\text{norm}}$. Using γ as in (5.16), we observe from our numerical results that we always have $E_{\text{norm},x_{\text{pre}}} \lesssim E_{\text{norm}}$.

5.4.3 Nonlinearity Precompensation Algorithms

First, we propose a suboptimal gradient-descent based algorithm to get the virtual samples $\{c_k\}$ that solve the nonconvex problem in (5.12). Then, we modify the virtual samples $\{c_k\}$ to another virtual samples $\{u_k\}$ to reduce the large spectrum sidelobes of the output of the frequency multiplier.

Let W_k be the sample-wise square-error objective function of (5.12). The complex gradient $\nabla_{c_k} W_k$ of W_k with respect to c_k is given by (see Appendix A)

$$\nabla_{c_k} W_k = 3 \left| \tilde{A}_3 \right|^2 |c_k|^4 c_k + 8 \Re \left\{ \tilde{A}_3^* \tilde{A}_5 \right\} |c_k|^6 c_k + 5 \left| \tilde{A}_5 \right|^2 |c_k|^8 c_k$$

$$- 3 \tilde{A}_3^* (c_k^*)^2 \tilde{s}_k - 4 \tilde{A}_5^* |c_k|^2 (c_k^*)^2 \tilde{s}_k - \tilde{A}_5 c_k^4 \tilde{s}_k^*. \tag{5.17}$$

The gradient-descent based algorithm is illustrated in Algorithm 1, where $\mathcal{P}_{\text{comp}}$ is the number of iterations. As a good initial solution $c_k^{(0)}$, we propose to obtain $c_k^{(0)}$ as $c_k^{(0)} = \left| \tilde{s}_k/\tilde{A}_3 \right|^{1/3} e^{j(\angle(\tilde{s}_k/\tilde{A}_3))/3}$ which is the solution of (5.12) when we ignore the contribution of the fifth-order power of the NLD. The step size $\mathfrak{s}$ of the updating rule is obtained by back-tracking line search [21]. Since in each iteration the step size $\mathfrak{s}$ is obtained such that the

sample-wise objective function W_k is decreased, it is guaranteed that Algorithm 1 converges to a local minimum.

Algorithm 1 Step 1 of proposed sequence generation

Inputs: $\tilde{A}_3$, $\tilde{A}_5$, $\{a_m\}$, $g(t)$, T_s, T_c, E_{norm}
Obtain γ as in (5.16)
for $k = 1$ to K
    Obtain $\tilde{s}_k$ as: $\tilde{s}_k = \sqrt{\gamma}\tilde{s}(kT_c)$
                   $= \sqrt{\gamma}\sum_{m=1}^{M} a_m g(kT_c - mT_s)$
    Obtain $c_k^{(0)}$ as: $c_k^{(0)} = \left| \tilde{s}_k / \tilde{A}_3 \right|^{1/3} e^{j\left(\angle\left(\tilde{s}_k/\tilde{A}_3\right)\right)/3}$
    for $p = 1$ to $\mathcal{P}_{comp}$ (or any other stopping criterion)
        Calculate the gradient $\nabla_{c_k}^{(p)} W_k$ using (5.17)
        Updating rule:
            $\mathfrak{s} = \arg\min_{\mathfrak{s}} W_k(c_k^{(p)} - \mathfrak{s}\nabla_{c_k}^{(p)} W_k)$
        $c_k^{(p+1)} = c_k^{(p)} - \mathfrak{s}\nabla_{c_k}^{(p)} W_k$
    end for
end for
Output: $\left\{ c_k^{(\mathcal{P}_{comp}+1)} \right\}$

Due to the third-order power relationship between the input and output of the frequency multiplier, for each c_k obtained by Algorithm 1, there are three virtual samples $\{\tilde{u}_{k,l}\}_{l=1}^{3}$ related to c_k as

$$\tilde{u}_{k,l} = |c_k| \, e^{j\left(\angle c_k + \frac{2\pi(l-2)}{3}\right)}, \text{ for } l = 1, 2, 3; \, \forall k, \tag{5.18}$$

which have the same square-error value of (5.12), i.e., $W_k(c_k) = W_k(\tilde{u}_{k,l})$, for $l = 1, 2, 3$. Therefore, we have 3^K possible virtual sequences for $\{\tilde{u}_{k,l}\}$ with the same square-error value of (5.12). To reduce the large spectrum sidelobes of the frequency multiplier output, we propose to choose the virtual sequence having the smallest PSD spread (slowest time variation) since the samples $\{\tilde{s}_k\}$ are correlated due to the higher sampling rate R_c than the symbol rate R_s of $\tilde{s}(t)$. However, the exhaustive search over 3^K virtual sequences to find the optimal solution is not practically affordable for large values of K. Hence, we propose a low-complexity suboptimal algorithm to obtain the virtual sequence $\{u_k\}$ as described in Algorithm 2. The

main idea of Algorithm 2 is to find the virtual samples sequentially by minimizing the Euclidean distance between each two successive virtual samples.

Algorithm 2 Step 2 of proposed virtual sequence generation

Input: $\{c_k\}$
$u_1 = c_1$
for $k = 2$ to K
 Calculate the three samples $\{\tilde{u}_{k,l}\}_{l=1}^3$ as in (5.18)
 Calculate the Euclidean distances between each of
 $\{\tilde{u}_{k,l}\}_{l=1}^3$ and u_{k-1}: $\left\{ e_l = |\tilde{u}_{k,l} - u_{k-1}|^2 \right\}$
 Find the $\tilde{u}_{k,l}$ with minimum Euclidean distance:
  $\tilde{l} = \arg \min_{l} \{e_l\}$
 Obtain u_k as: $u_k = \tilde{u}_{k,\tilde{l}}$
end for
Output: $\{u_k\}$

5.4.4   I/Q Imbalance Precompensation

Suppose that we have already obtained the virtual sequence $\{u_k\}$ using Algorithms 1 and 2. From (5.4), we can precompensate the I/Q imbalance effect by modifying the virtual samples $\{u_k\}$ to $\{d_k\}$ as

$$d_k = \frac{\mu^* u_k - \nu u_k^*}{|\mu|^2 - |\nu|^2}, \ \forall k, \tag{5.19}$$

which precompensates the effect of I/Q imbalance completely (i.e., $\mu d_k + \nu d_k^* = u_k$, $\forall k$). Note that the average sample energies of $\{c_k\}$ and $\{u_k\}$ are the same. Therefore, the normalized energy constraint E_{norm} is still satisfied.

 The proposed precompensation scheme can be summarized as in Fig. 5.3, where $\{a_{\text{I},m} + ja_{\text{Q},m}\}$ are QAM symbols with the symbol rate R_s. Algorithms 1 generates $\{c_{\text{I},k}, c_{\text{Q},k}\}$ at the rate of R_c. Algorithms 2 generates $\{u_{\text{I},k}, u_{\text{Q},k}\}$ at the rate of R_c. Then $\{d_{\text{I},k}, d_{\text{Q},k}\}$ are generated at the rate of R_c. Note that the construction filter $q(t)$ in Fig. 5.3 has a wider bandwidth than that of $g(t)$ in Fig. 5.1. The proposed precompensation scheme requires the knowledge of the NLD parameters ($\tilde{A}_3$ and $\tilde{A}_5$) and the I/Q imbalance parameters (ϵ and
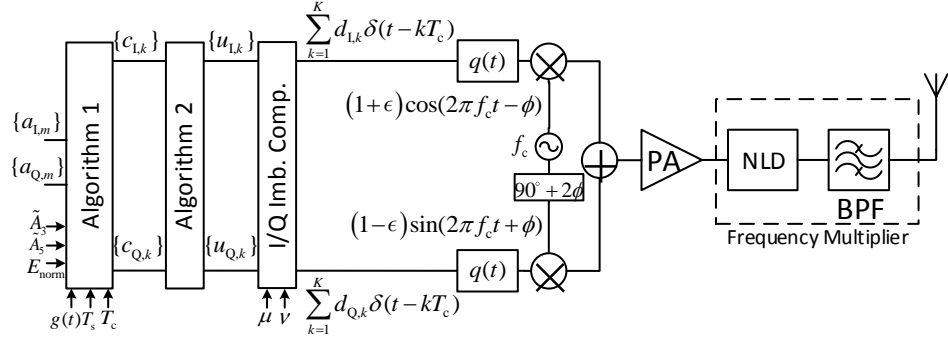
Figure 5.3. Proposed precompensation scheme based on the frequency-multiplier-last transmitter architecture for THz I/Q transmission.
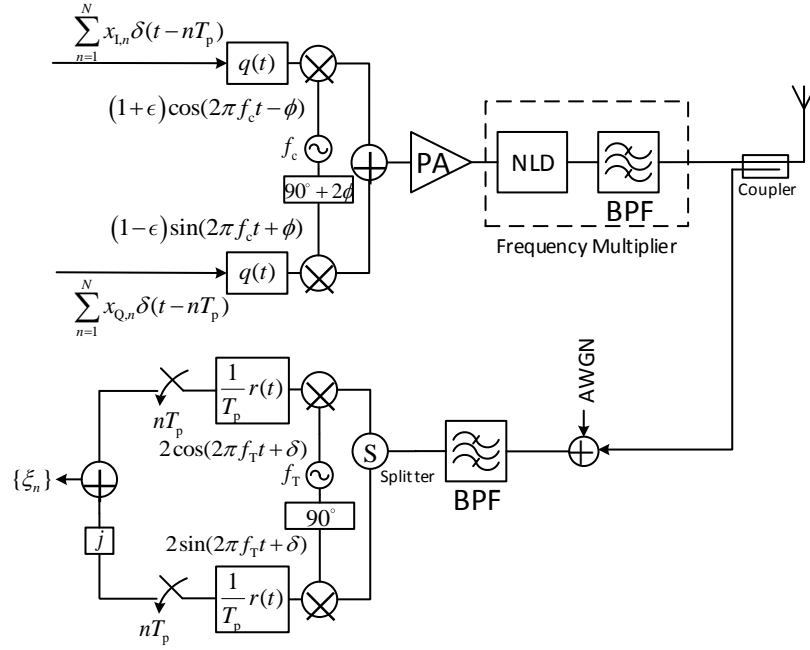
Figure 5.4. Measurement setup for parameters estimation during THz device testing/calibration phase.

ϕ). In the next section, we will describe the estimation method of these parameters. The

computational complexity of the proposed precompensation scheme is $\mathcal{O}\left(ML(\mathcal{P}_{\mathrm{comp}}+3)\right)$,

which is a linear function of the modulation block size M.

5.5 Proposed System Parameters Estimation

It is essential to estimate the NLD parameters ($\tilde{A}_3$ and $\tilde{A}_5$) and the I/Q imbalance parameters (ϵ and ϕ) to apply the proposed precompensation scheme in the previous section. For practicality, in the estimation stage we consider the same proposed transmitter architecture of section 5.4. Note that Algorithm 1, Algorithm 2, I/Q imbalance compensation are deactivated since the system parameters are still unknown.

5.5.1 Estimator Setup

We assume that the parameters estimation is performed during the THz device testing/calibration phase. The measurement setup for (offline) parameters estimation is shown in Fig. 5.4, where the lower branch (the feedback RF chain) is a separate measurement circuitry for testing/calibrating the THz transmitters. This measurement circuitry can be built with higher quality devices and hence we assume it does not introduce any type of distortions. For simplicity, we just show functional blocks for the measurement circuitry without elaborating details (e.g., the generation of THz frequency at the feedback RF chain, which can be done as in [127]). Note that since the parameters of our system are static, deterministic estimation methods can be applied [73]. For dynamic system identification, the reader is referred to [33, 32, 31, 30] for new-type parameter estimation methods.

Examining the nonlinear signal model of section II, we note that due to the nonlinearity of the frequency multiplier, the I/Q imbalance parameters cannot be estimated unless the following two conditions are satisfied. First, the pilot symbols are inter-symbol-interference-free (ISI-free). Second, the first half of the pilot sequence is real only while the second half is imaginary only, or the first half of the pilot sequence is imaginary only while the second half is real only. To illustrate this, suppose $\{x_n = x_{\mathrm{I},n} + jx_{\mathrm{Q},n}\}_{n=1}^{N}$ is the pilot sequence to be transmitted with rate $R_{\mathrm{p}} = 1/T_{\mathrm{p}}$ to estimate the system parameters as shown in Fig. 5.4. Based on (5.10) and the transmitter architecture of Fig. 5.4, we illustrate these two

$$y_C(t) = \Re\left\{ \left(\tilde{A}_3 \Big(\sum_{n=1}^{N} (\mu x_n + \nu x_n^*) q(t - nT_p) \Big)^3 + \tilde{A}_5 \Big(\sum_{n=1}^{N} (\mu x_n + \nu x_n^*) q(t - nT_p) \Big)^3 \right. \right.$$

$$\left. \left. \times \Big| \sum_{n=1}^{N} (\mu x_n + \nu x_n^*) q(t - nT_p) \Big|^2 \right) e^{j2\pi f_T t} \right\} \tag{5.20}$$

$$\overset{(a)}{=} \Re\left\{ \sum_{n=1}^{N} \left(\tilde{A}_3 (\mu x_n + \nu x_n^*)^3 q^3(t - nT_p) + \tilde{A}_5 (\mu x_n + \nu x_n^*)^3 \big| \mu x_n + \nu x_n^* \big|^2 q^5(t - nT_p) \right) e^{j2\pi f_T t} \right\} \tag{5.21}$$

$$\overset{(b)}{=} \begin{cases} \Re\left\{ \sum_{n=1}^{N} \left(\tilde{A}_3 x_{I,n}^3 q^3(t - nT_p)(\mu + \nu)^3 + \tilde{A}_5 x_{I,n}^5 q^5(t - nT_p)(\mu + \nu)^3 \big| \mu + \nu \big|^2 \right) e^{j2\pi f_T t} \right\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } x_{Q,n} = 0,\ \forall n \\[2em] \Re\left\{ -j \sum_{n=1}^{N} \left(\tilde{A}_3 x_{Q,n}^3 q^3(t - nT_p)(\mu - \nu)^3 + \tilde{A}_5 x_{Q,n}^5 q^5(t - nT_p)(\mu - \nu)^3 \big| \mu - \nu \big|^2 \right) e^{j2\pi f_T t} \right\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } x_{I,n} = 0,\ \forall n \end{cases} \tag{5.22}$$

conditions in (5.20)$-$(5.22), where $\overset{(a)}{=}$ is obtained assuming that the pilot symbols are ISI-free, and then $\overset{(b)}{=}$ is obtained assuming that the pilot symbols are either real or imaginary. With satisfying these two conditions, we observe from (5.22) that the pilot symbols and the I/Q imbalance parameters are decoupled.

To accommodate the first condition, as shown in Fig. 5.4, the pilot symbols $\{x_{p,n}\}$ are generated every $T_p = QT_c$, where QT_c is the time span of the construction filter $q(t)$ to avoid the inter-symbol-interference. Similarly, the output of the receive filter $r(t)$ of the feedback RF chain is sampled every T_p. To accommodate the second condition, we perform two phases of measurements. Suppose $\chi_p = \{x_{p,1}, x_{p,2}, \ldots, x_{p,N}\}$ be a length-N real-valued pilot sequence (to be designed in details in the next section). In the first phase, the real-valued pilot symbols are transmitted only on the inphase branch (i.e., $x_{I,n} = x_{p,n}$, $x_{Q,n} = 0$, $\forall n$). In the second phase, the real-valued pilot symbols are transmitted only on the quadrature branch (i.e., $x_{Q,n} = x_{p,n}$, $x_{I,n} = 0$, $\forall n$). Based on (5.22) and the transmitter architecture in

Fig. 5.4, the observation symbols $\{\xi_{n,1}\}_{n=1}^{N}$ and $\left\{\tilde{\xi}_{n,2}\right\}_{n=1}^{N}$ (sampled every T_{p} at the output of the receive filter $r(t)$ of the feedback chain) of the first and second measurement phases respectively can be written as

$$\xi_{1,n} = e^{j\delta}(x_{\mathrm{p},n}^3 r_3 \tilde{A}_3(\mu+\nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5(\mu+\nu)^3|\mu+\nu|^2) + \eta_{1,n} \tag{5.23}$$

$$\tilde{\xi}_{2,n} = -je^{j\delta}(x_{\mathrm{p},n}^3 r_3 \tilde{A}_3(\mu-\nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5(\mu-\nu)^3|\mu-\nu|^2) + \tilde{\eta}_{2,n}, \tag{5.24}$$

where δ is the oscillator phase shift of the feedback RF chain, $r_3 = (1/T_{\mathrm{p}})(q^3(t)*r(t))_{t=T_{\mathrm{p}}}$, $r_5 = (1/T_{\mathrm{p}})(q^5(t)*r(t))_{t=T_{\mathrm{p}}}$, and $\eta_{1,n}$ and $\tilde{\eta}_{2,n}$ are the zero-mean additive white complex Gaussian noise (AWGN) with variance σ_η^2 of the first and second measurement phases respectively. For convenience, we consider $\{\xi_{2,n}\}$ instead of $\left\{\tilde{\xi}_{2,n}\right\}$, where $\xi_{2,n} = j\tilde{\xi}_{2,n}$, $\forall n$ which is given by

$$\xi_{2,n} = e^{j\delta}(x_{\mathrm{p},n}^3 r_3 \tilde{A}_3(\mu-\nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5(\mu-\nu)^3|\mu-\nu|^2) + \eta_{2,n}, \tag{5.25}$$

where $\eta_{2,n} = j\tilde{\eta}_{2,n}$ is also zero-mean additive white complex Gaussian noise with variance σ_η^2.

From (5.23) and (5.25), we notice that the phase shift δ can be absorbed into the estimates of the NLD coefficients $\tilde{A}_3$ and $\tilde{A}_5$ resulting in a fixed rotation in the receive constellation which can be considered as a part of the channel to be compensated at the receiver. Therefore, without loss of generality, we consider $\delta = 0$ in the rest of the chapter. From (5.23) and (5.25), we also notice that the observation symbols $\{\xi_{1,n}\}$ and $\{\xi_{2,n}\}$ are affected by r_3 and r_5 which depend on the design of the receive filter $r(t)$. Note that $r(t)$ is normalized such that $(1/T_{\mathrm{p}}) \int_{-T_{\mathrm{p}}/2}^{T_{\mathrm{p}}/2} r^2(t)\, dt = 1$. We consider three choices for $r(t)$. The first choice $r_1(t)$ is a normalized version of $q(t)$ given by

$$r_1(t) = q(t) \Big/ \sqrt{\frac{1}{T_{\mathrm{p}}} \int_{-T_{\mathrm{p}}/2}^{T_{\mathrm{p}}/2} q^2(t)\, dt}. \tag{5.26}$$

90
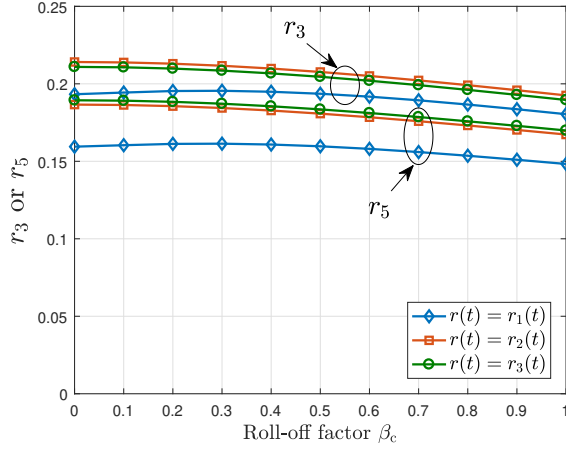
Figure 5.5. Values of r_3 and r_5 for different choices of $r(t)$ versus the roll-off factor β_c with a time span of $Q = 12$.

The second choice $r_2(t)$ is given by

$$r_2(t) = q^3(t)/\sqrt{\frac{1}{T_p} \int_{-T_p/2}^{T_p/2} q^6(t)\, dt}, \tag{5.27}$$

which maximizes r_3 (since it is the matched filter of $q^3(t)$). The third choice $r_3(t)$ is given by

$$r_3(t) = q^5(t)/\sqrt{\frac{1}{T_p} \int_{-T_p/2}^{T_p/2} q^{10}(t)\, dt}, \tag{5.28}$$

which maximizes r_5 (since it is the matched filter of $q^5(t)$). Fig. 5.5 shows the values of r_3 and r_5 for the three choices of $r(t)$ versus the roll-off factor β_c with a time span of $Q = 12$. We observe that the first choice $r_1(t)$ yields the smallest values of r_3 and r_5, while there is no significant difference between the second and third choices of $r(t)$. Therefore, in the rest of the chapter, we consider the second choice of $r(t)$, i.e., $r(t) = r_2(t)$ as in (5.27).

5.5.2   Estimation Algorithm

Using the observation symbols $\{\xi_{1,n}\}$ and $\{\xi_{2,n}\}$ of the two measurement phases, we propose a maximum-likelihood (ML) estimator. The log-likelihood function (LLF) $\mathcal{L}(\{\xi_{1,n}\}, \{\xi_{2,n}\}; \tilde{A}_3, \tilde{A}_5, \epsilon, \phi)$ of the observation symbols is given by [73]

$$\mathcal{L}(\{\xi_{1,n}\}, \{\xi_{2,n}\}; \tilde{A}_3, \tilde{A}_5, \epsilon, \phi) = -2N \ln \pi \sigma_\eta^2 - \frac{1}{\sigma_\eta^2}$$

$$\times \left(\sum_{n=1}^{N} |\xi_{1,n} - x_{\mathrm{p},n}^3 r_3 \tilde{A}_3 (\mu + \nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5 (\mu + \nu)^3 |\mu + \nu|^2|^2 \right.$$

$$\left. + \sum_{n=1}^{N} |\xi_{2,n} - x_{\mathrm{p},n}^3 r_3 \tilde{A}_3 (\mu - \nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5 (\mu - \nu)^3 |\mu - \nu|^2|^2 \right), \quad (5.29)$$

where μ and ν are related to ϵ and ϕ as in (5.5) and (5.6). Therefore, the estimation of the system parameters can be formulated as an ML estimator given by

$$[\widehat{\tilde{A}_3}, \widehat{\tilde{A}_5}, \hat{\epsilon}, \hat{\phi}] = \arg\min_{\tilde{A}_3, \tilde{A}_5, \epsilon, \phi} \tilde{\mathcal{L}}(\{\xi_{1,n}\}, \{\xi_{2,n}\}; \tilde{A}_3, \tilde{A}_5, \epsilon, \phi), \quad (5.30)$$

where $\tilde{\mathcal{L}}(\{\xi_{1,n}\}, \{\xi_{2,n}\}; \tilde{A}_3, \tilde{A}_5, \epsilon, \phi)$ (or $\tilde{\mathcal{L}}$ for simplicity) is given by

$$\tilde{\mathcal{L}} = \sum_{n=1}^{N} |\xi_{1,n} - x_{\mathrm{p},n}^3 r_3 \tilde{A}_3 (\mu + \nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5 (\mu + \nu)^3 |\mu + \nu|^2|^2$$

$$+ \sum_{n=1}^{N} |\xi_{2,n} - x_{\mathrm{p},n}^3 r_3 \tilde{A}_3 (\mu - \nu)^3 + x_{\mathrm{p},n}^5 r_5 \tilde{A}_5 (\mu - \nu)^3 |\mu - \nu|^2|^2. \quad (5.31)$$

The optimization problem in (5.30) is nonconvex and cannot be solved directly. Using good initial estimates for the I/Q imbalance parameters (will be obtained later), we propose an alternating algorithm which estimates the NLD parameters and the I/Q imbalance parameters alternately.

NLD Parameters Estimation

We fix the estimation of the I/Q imbalance parameters and estimate the NLD parameters. Define $\boldsymbol{\zeta}_1 = [\xi_{1,1}, \xi_{1,2}, \ldots, \xi_{1,N}]^T$, $\boldsymbol{\zeta}_2 = [\xi_{2,1}, \xi_{2,2}, \ldots, \xi_{2,N}]^T$, $\boldsymbol{\zeta} = [\boldsymbol{\zeta}_1^T, \boldsymbol{\zeta}_2^T]^T$, $\mathbf{x}_{\mathrm{p},3} = [x_{\mathrm{p},1}^3, x_{\mathrm{p},2}^3, \ldots, x_{\mathrm{p},N}^3]^T$, $\mathbf{x}_{\mathrm{p},5} = [x_{\mathrm{p},1}^5, x_{\mathrm{p},2}^5, \ldots, x_{\mathrm{p},N}^5]^T$, $\mathbf{X}_{\mathrm{p},1} = [r_3(\mu + \nu)^3 \mathbf{x}_{\mathrm{p},3}, r_5(\mu + \nu)^3 |\mu + \nu|^2 \mathbf{x}_{\mathrm{p},5}]$, $\mathbf{X}_{\mathrm{p},2} = [r_3(\mu - \nu)^3 \mathbf{x}_{\mathrm{p},3}, r_5(\mu - \nu)^3 |\mu - \nu|^2 \mathbf{x}_{\mathrm{p},5}]$, and $\mathbf{X}_{\mathrm{p}} = [\mathbf{X}_{\mathrm{p},1}^T, \mathbf{X}_{\mathrm{p},2}^T]^T$. Then, the optimization problem in (5.30) can be written (while fixing ϵ and ϕ) in a matrix form as

$$[\widehat{\tilde{A}_3}, \widehat{\tilde{A}_5}] = \arg\min_{\tilde{A}_3, \tilde{A}_5} ||\boldsymbol{\zeta} - \mathbf{X}_{\mathrm{p}}[\tilde{A}_3, \tilde{A}_5]^T||^2, \quad (5.32)$$

which is a linear estimation problem, and its solution is given by

$$[\widehat{\tilde{A}_3}, \widehat{\tilde{A}_5}]^T = (\mathbf{X}_{\mathrm{p}}^H \mathbf{X}_{\mathrm{p}})^{-1} \mathbf{X}_{\mathrm{p}}^H \boldsymbol{\zeta}. \tag{5.33}$$

I/Q Imbalance Parameters Estimation

Now, we fix the estimation of the NLD parameters and enhance the estimation of the I/Q imbalance parameters. However, the estimation of the I/Q imbalance parameters of (5.30) (while fixing $\tilde{A}_3$ and $\tilde{A}_5$) is still a nonconvex optimization problem due to the nonlinearity of the NLD. We propose a suboptimal gradient-descent based estimator for the I/Q imbalance parameters. Define $\mathbf{e} = [\epsilon, \phi]^T$, $\alpha_1 = \mu + \nu$, and $\alpha_2 = \mu - \nu$. Then, the gradient $\nabla_{\mathbf{e}}\tilde{\mathcal{L}}$ of $\tilde{\mathcal{L}}$ with respect to $\mathbf{e}$ is given by $\nabla_{\mathbf{e}}\tilde{\mathcal{L}} = [\partial\tilde{\mathcal{L}}/\partial\epsilon, \partial\tilde{\mathcal{L}}/\partial\phi]^T$, where $\partial\tilde{\mathcal{L}}/\partial\epsilon$ and $\partial\tilde{\mathcal{L}}/\partial\phi$ are obtained as (see Appendix B)

$$\frac{\partial\tilde{\mathcal{L}}}{\partial\epsilon} = 2\sum_{n=1}^{N}\mathfrak{R}\{\lambda_{1,n}e^{j\phi} - \lambda_{2,n}e^{-j\phi}\}, \tag{5.34}$$

$$\frac{\partial\tilde{\mathcal{L}}}{\partial\phi} = 2\sum_{n=1}^{N}\mathfrak{R}\{\lambda_{1,n}j\alpha_1^* - \lambda_{2,n}j\alpha_2^*\}, \tag{5.35}$$

where $\lambda_{i,n}$ is given by

$$\lambda_{i,n} = 3\left|\tilde{A}_3\right|^2 |\alpha_i|^4 \alpha_i + 8\mathfrak{R}\left\{\tilde{A}_3^*\tilde{A}_5\right\}|\alpha_i|^6 \alpha_i + 5\left|\tilde{A}_5\right|^2 |\alpha_i|^8 \alpha_i$$
$$-3\tilde{A}_3^*(\alpha_i^*)^2 \zeta_{i,n} - 4\tilde{A}_5^* |\alpha_i|^2 (\alpha_i^*)^2 \zeta_{i,n} - \tilde{A}_5\alpha_i^4\zeta_{i,n}^*, , \forall i, n. \tag{5.36}$$

The gradient-descent based estimator of the I/Q imbalance parameters is illustrated in Algorithm 3, where $\mathcal{P}_{\mathrm{IQ}}$ is the number of iterations. The step size $\mathfrak{s}$ of the updating rule is obtained by back-tracking line search [21]. Since in each iteration the step size $\mathfrak{s}$ is obtained such that the objective function $\tilde{\mathcal{L}}$ is decreased, it is guaranteed that Algorithm 3 converges to a local minimum.

Algorithm 3 ML Estimator of I/Q Imbalance Parameters

Inputs: $\tilde{A}_3$, $\tilde{A}_5$, $\epsilon^{(0)}$, $\phi^{(0)}$, $\mathcal{P}_{\mathrm{IQ}}$

for $p = 1$ to $\mathcal{P}_{\mathrm{IQ}}$ (or any other stopping criterion)

    Calculate the gradient $\nabla_{\mathbf{e}}^{(p)}\mathcal{L}$ using $(5.34)-(5.36)$

    Updating rule:

$$\mathfrak{s} = \arg\min_{\mathfrak{s}} \tilde{\mathcal{L}}\left(\tilde{A}_3, \tilde{A}_5, \epsilon^{(p)} - \mathfrak{s}\frac{\partial\tilde{\mathcal{L}}}{\partial\epsilon}^{(p)}, \phi^{(p)} - \mathfrak{s}\frac{\partial\tilde{\mathcal{L}}}{\partial\phi}^{(p)}\right)$$

$$\epsilon^{(p+1)} = \epsilon^{(p)} - \mathfrak{s}\frac{\partial\tilde{\mathcal{L}}}{\partial\epsilon}^{(p)},$$

$$\phi^{(p+1)} = \phi^{(p)} - \mathfrak{s}\frac{\partial\tilde{\mathcal{L}}}{\partial\phi}^{(p)},$$

end for

Output: $\epsilon^{(\mathcal{P}_{\mathrm{IQ}}+1)}$, $\phi^{(\mathcal{P}_{\mathrm{IQ}}+1)}$

Initial Estimates of I/Q Imbalance Parameters

An important step in the proposed estimator is to get good initial estimates for the I/Q imbalance parameters. To do this, we define $\rho_1 = \tilde{A}_3(\mu + \nu)^3$, $\rho_2 = \tilde{A}_3(\mu - \nu)^3$, $\rho_3 = \tilde{A}_5(\mu + \nu)^3|\mu + \nu|^2$ and $\rho_4 = \tilde{A}_5(\mu - \nu)^3|\mu - \nu|^2$. Then, we estimate ρ_1 from the observation symbols $\{\xi_{1,n}\}$ of the first measurement phase and estimate ρ_2 from the observation symbols $\{\xi_{2,n}\}$ of the second measurement phase, as follows. Let us define

$$\tilde{\mathbf{X}}_{\mathrm{p}} = [r_3\mathbf{x}_{\mathrm{p},3},\ r_5\mathbf{x}_{\mathrm{p},5}]. \tag{5.37}$$

Similar to (5.32), we can write the ML estimators of ρ_1 and ρ_2 as

$$\hat{\rho}_1 = \arg\min_{\rho_1} ||\boldsymbol{\zeta}_1 - \tilde{\mathbf{X}}_{\mathrm{p}}[\rho_1, \rho_3]^T||^2, \tag{5.38}$$

$$\hat{\rho}_2 = \arg\min_{\rho_2} ||\boldsymbol{\zeta}_2 - \tilde{\mathbf{X}}_{\mathrm{p}}[\rho_2, \rho_4]^T||^2, \tag{5.39}$$

which are two independent linear estimation problems, and their solutions are given by

$$\hat{\rho}_1 = [(\tilde{\mathbf{X}}_{\mathrm{p}}^H\tilde{\mathbf{X}}_{\mathrm{p}})^{-1}\tilde{\mathbf{X}}_{\mathrm{p}}^H\boldsymbol{\zeta}_1]_{(1)}, \tag{5.40}$$

$$\hat{\rho}_2 = [(\tilde{\mathbf{X}}_{\mathrm{p}}^H\tilde{\mathbf{X}}_{\mathrm{p}})^{-1}\tilde{\mathbf{X}}_{\mathrm{p}}^H\boldsymbol{\zeta}_2]_{(1)}. \tag{5.41}$$

Ignoring the noise effect, we can write $(\hat{\rho}_1/\hat{\rho}_2)^{1/3}$ (using the definitions of ρ_1, ρ_2, μ, and ν) as

$$
\begin{aligned}
(\frac{\hat{\rho}_1}{\hat{\rho}_2})^{1/3} &= \frac{\mu + \nu}{\mu - \nu} \\
&= \frac{(1 + \epsilon)(\cos\phi - j\sin\phi)}{(1 - \epsilon)(\cos\phi + j\sin\phi)} \\
&= \frac{(1 + \epsilon)}{(1 - \epsilon)}(\cos 2\phi - j\sin 2\phi) \triangleq \kappa_R + j\kappa_I.
\end{aligned} \tag{5.42}
$$

However, we have three roots for $(\hat{\rho}_1/\hat{\rho}_2)^{1/3}$, and hence we need to pick the correct root. Examining (5.42), we know that the correct root has two properties. For any ϵ and $|\phi| < 45°$ ($|\phi|$ is much smaller than $45°$ in practice), $\kappa_R > 0$ and $\kappa_R > |\kappa_I|$. Therefore, we pick the root that satisfies $\kappa_R > 0$ and $\kappa_R > |\kappa_I|$. Due to noise, if there are more than one root satisfying these two properties, we pick any one of them. Using κ_R and κ_I, we obtain the initial estimates of ϵ and ϕ as

$$
\hat{\epsilon} = \frac{2\sqrt{\kappa_R^2 + \kappa_I^2}}{1 + \sqrt{\kappa_R^2 + \kappa_I^2}} - 1, \tag{5.43}
$$

$$
\hat{\phi} = -0.5\tan^{-1}(\frac{\kappa_I}{\kappa_R}). \tag{5.44}
$$

The proposed ML estimator of the system parameters is summarized in Algorithm 4, where $\mathcal{P}_{ML}$ is the number of iterations. The computational complexity of the proposed ML estimator is $\mathcal{O}\left(((24 + \mathcal{P}_{IQ})N + 24)(\mathcal{P}_{ML} + 1)\right)$, which is a linear function of the length N of the pilot sequence. Note that while fixing the I/Q imbalance parameters, the estimation of the NLD parameters is a linear estimation problem, and the ML estimator of the NLD parameters is unbiased. While fixing the NLD parameters, the estimation of the I/Q imbalance parameters is a nonlinear estimation problem, and the ML estimator of the I/Q imbalance parameters is biased. Therefore, the ML algorithm is, in general, a biased estimator. However, as the length of the pilot sequence increases, the bias converges to zero (asymptotically unbiased) [73].

Algorithm 4 ML Estimator of System Parameters

Inputs: $\{\xi_{1,n}\}$, $\{\xi_{2,n}\}$, P_{ML}

Obtain $\hat{\rho}_1$ and $\hat{\rho}_2$ using (5.40) and (5.41)

Obtain κ_{R} and κ_{I} as $(\hat{\rho}_1/\hat{\rho}_2)^{1/3} = \kappa_{\text{R}} + j\kappa_{\text{I}}$ while
  choosing the root that satisfies $\kappa_{\text{R}} > 0$ and $\kappa_{\text{R}} > |\kappa_{\text{I}}|$.
  If there are more than one root satisfying these two
  properties, choose any one of them.

Obtain $\hat{\epsilon}^{(0)}$ and $\hat{\phi}^{(0)}$ using (5.43) and (5.44)

Obtain $\widehat{\tilde{A}_3}^{(0)}$ and $\widehat{\tilde{A}_5}^{(0)}$ using (5.33) while fixing
$\epsilon = \hat{\epsilon}^{(0)}$ and $\phi = \hat{\phi}^{(0)}$

for $p = 1$ to $\mathcal{P}_{\text{ML}}$ (or any other stopping criterion)
    Obtain $\hat{\epsilon}^{(p)}$ and $\hat{\phi}^{(p)}$ using Algorithm 3 with $\hat{\epsilon}^{(p-1)}$

    and $\hat{\phi}^{(p-1)}$ as initial solutions while fixing $\tilde{A}_3 = \widehat{\tilde{A}_3}^{(p-1)}$
    and $\tilde{A}_5 = \widehat{\tilde{A}_5}^{(p-1)}$

    Obtain $\widehat{\tilde{A}_3}^{(p)}$ and $\widehat{\tilde{A}_5}^{(p)}$ using (5.33) while fixing $\epsilon = \hat{\epsilon}^{(p)}$
    and $\phi = \hat{\phi}^{(p)}$

end for

Output: $\widehat{\tilde{A}_3}^{(\mathcal{P}_{\text{ML}})}$, $\widehat{\tilde{A}_5}^{(\mathcal{P}_{\text{ML}})}$, $\hat{\epsilon}^{(\mathcal{P}_{\text{ML}})}$, $\hat{\phi}^{(\mathcal{P}_{\text{ML}})}$

5.6 Cramér–Rao Lower Bounds for System Parameters Estimation and Pilot Sequence Design

5.6.1 CRLBs for System Parameters Estimation

The Cramer-Rao lower bounds (CRLBs) provide limits on the estimate variances for a set of deterministic parameters. In other words, if $\hat{x}$ is the unbiased estimate of x, then the mean square error (MSE) of $\hat{x}$ is $\text{MSE}(\hat{x}) = \mathbb{E}\{|\hat{x} - x|^2\} \geq \text{CRLB}(x)$. We examine the CRLB as a benchmark metric to evaluate the performance of the proposed ML estimator. Define the vector system parameter $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = [\tilde{A}_3, \tilde{A}_5, \epsilon, \phi]^T$ which represents the system parameters

that we estimate. The Fisher information matrix (FIM) $\mathbf{I}(\boldsymbol{\theta})$ is given by

$$
\mathbf{I}(\boldsymbol{\theta}) =
\begin{bmatrix}
\mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})^*(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})^*(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})^*(\frac{\partial \mathcal{L}}{\partial \epsilon})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})^*(\frac{\partial \mathcal{L}}{\partial \phi})\} \\
\mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})^*(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})^*(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})^*(\frac{\partial \mathcal{L}}{\partial \epsilon})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})^*(\frac{\partial \mathcal{L}}{\partial \phi})\} \\
\mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \epsilon})(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \epsilon})(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \epsilon})^2\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \epsilon})(\frac{\partial \mathcal{L}}{\partial \phi})\} \\
\mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \phi})(\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \phi})(\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \phi})(\frac{\partial \mathcal{L}}{\partial \epsilon})\} & \mathbb{E}\{(\frac{\partial \mathcal{L}}{\partial \phi})^2\}
\end{bmatrix}, \quad (5.45)
$$

where $\mathcal{L}$ is the LLF in (5.29) [73]. Then, $\mathrm{CRLB}(\tilde{A}_3) = [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{(1,1)}$, $\mathrm{CRLB}(\tilde{A}_5) = [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{(2,2)}$, $\mathrm{CRLB}(\epsilon) = [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{(3,3)}$, and $\mathrm{CRLB}(\phi) = [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{(4,4)}$ [73].

We provide $\mathbf{I}(\boldsymbol{\theta})$ in closed-form in the next proposition. Note that the upper triangle entries of $\mathbf{I}(\boldsymbol{\theta})$ are sufficient to construct $\mathbf{I}(\boldsymbol{\theta})$ since $\mathbf{I}(\boldsymbol{\theta})$ is a Hermitian matrix. For simplicity, similar to the previous section, we set $\alpha_1 = \mu + \nu$ and $\alpha_2 = \mu - \nu$.

Proposition 3. *The upper triangle entries of the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ of the vector system parameter $\boldsymbol{\theta}$ can be written in closed-form expressions as in* (5.65)$-$(5.74) *shown in Appendix C.*

Proof. Please refer to Appendix C. □

5.6.2 Proposed Pilot Sequence Design

We design the length-N real-valued pilot sequence $\chi_{\mathrm{p}} = \{x_{\mathrm{p},1}, x_{\mathrm{p},2}, \ldots, x_{\mathrm{p},N}\}$ to enhance the performance of the proposed ML estimator of the previous subsection. As shown in Algorithm 4, the first step in the proposed ML estimator is to estimate $\rho_1 = \tilde{A}_3(\mu + \nu)^3$ and $\rho_2 = \tilde{A}_3(\mu - \nu)^3$ to get initial estimates for the I/Q imbalance parameters. Therefore, we propose to design the pilot sequence χ_{p} to minimize the CRLBs of ρ_1 and ρ_2.

Since the same pilot sequence χ_{p} is transmitted in the two measurement phases (on the I branch in the first phase and on the Q branch in the second phase), we have $\mathrm{CRLB}(\rho_1) = \mathrm{CRLB}(\rho_2)$. Define $\boldsymbol{\psi} = [\rho_1, \rho_3]$. To derive $\mathrm{CRLB}(\rho_1)$, first we need to obtain the Fisher

information matrix $\mathbf{I}(\boldsymbol{\psi})$ of $\boldsymbol{\psi}$ which is given by

$$\mathbf{I}(\boldsymbol{\psi}) = \begin{bmatrix} \mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_1^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_1^*})\} & \mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_1^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_3^*})\} \\ \mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_3^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_1^*})\} & \mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_3^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_3^*})\} \end{bmatrix}. \tag{5.46}$$

Similar to $(5.65)-(5.67)$ in Appendix C, we can obtain $\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_1^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_1^*})\}$, $\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_1^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_3^*})\}$, and $\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_3^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_3^*})\}$ as

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_1^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_1^*})\} = \frac{1}{\sigma_\eta^2}r_3^2\sum_{n=1}^{N}x_{\mathrm{p},n}^6, \tag{5.47}$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_3^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_3^*})\} = \frac{1}{\sigma_\eta^2}r_5^2\sum_{n=1}^{N}x_{\mathrm{p},n}^{10}, \tag{5.48}$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\rho_1^*})^*(\frac{\partial\mathcal{L}}{\partial\rho_3^*})\} = \frac{1}{\sigma_\eta^2}r_3 r_5\sum_{n=1}^{N}x_{\mathrm{p},n}^8. \tag{5.49}$$

Therefore, we can write $\mathrm{CRLB}(\rho_1) = [\mathbf{I}^{-1}(\boldsymbol{\psi})]_{(1,1)}$ as

$$\begin{aligned} \mathrm{CRLB}(\rho_1) &= \frac{\sigma_\eta^2}{r_3^2}\frac{\sum_{n=1}^{N}x_{\mathrm{p},n}^{10}}{(\sum_{n=1}^{N}x_{\mathrm{p},n}^6)(\sum_{n=1}^{N}x_{\mathrm{p},n}^{10}) - (\sum_{n=1}^{N}x_{\mathrm{p},n}^8)^2} \\ &= \frac{\sigma_\eta^2}{r_3^2}\frac{\sum_{n_1=1}^{N}x_{\mathrm{p},n_1}^{10}}{\sum_{n_2=1}^{N}\sum_{\substack{n_1=1 \\ n_1\neq n_2}}^{N}(x_{\mathrm{p},n_1}^6 x_{\mathrm{p},n_2}^{10} - x_{\mathrm{p},n_1}^8 x_{\mathrm{p},n_2}^8)} \end{aligned} \tag{5.50}$$

We note that it is only the pilot-energy allocation among the pilot symbols that has an effect on $\mathrm{CRLB}(\rho_1)$. We also note that the order of the pilot symbols does not affect $\mathrm{CRLB}(\rho_1)$. To estimate ρ_1, we know that the rank of the observation matrix $\tilde{\mathbf{X}}_{\mathrm{p}}$ in (5.37) should be two or more since we have two unknown parameters ρ_1 and ρ_3. If we consider an equal-energy allocation among the pilot symbols, the rows of $\tilde{\mathbf{X}}_{\mathrm{p}}$ in (5.37) becomes the same, and hence $\tilde{\mathbf{X}}_{\mathrm{p}}$ becomes a rank-one matrix. Therefore, the equal-energy pilot sequence design has $\mathrm{CRLB}(\rho_1) = \infty$, and hence it cannot be used to estimate ρ_1. This fact can also be viewed from (5.50), where the denominator becomes zeros ($\mathrm{CRLB}(\rho_1) = \infty$) with the equal-energy allocation among the pilot symbols.

The normalized energy $E_{\text{norm,pilot}}$ of the pilot signal $x_{\text{pilot}}(t) = \sum_{n=1}^{N} x_{\text{p},n} q(t - nT_{\text{p}})$ is given by

$$
\begin{aligned}
E_{\text{norm,pilot}} &= \frac{1}{NT_{\text{p}}} \int_{NT_{\text{p}}} \sum_{n=1}^{N} x_{\text{p},n}^2 q^2(t - nT_{\text{p}}) dt \\
&= \frac{1}{N} \sum_{n=1}^{N} x_{\text{p},n}^2 \frac{1}{QT_{\text{c}}} \int_{T_{\text{p}}} q^2(t) dt \\
&= \frac{1}{NQ(1 - \frac{\beta_{\text{c}}}{4})} \sum_{n=1}^{N} x_{\text{p},n}^2,
\end{aligned}
\tag{5.51}
$$

where $(1/T_{\text{c}}) \int_{T_{\text{p}}} q^2(t) dt = 1/(1 - \beta_{\text{c}}/4)$ for the raised-cosine construction filter $q(t)$ of (5.14) if $T_{\text{p}} \gg T_{\text{c}}$ [65]. Then, the pilot sequence which minimizes $\text{CRLB}(\rho_1)$ is designed as

$$
\min_{\{x_{\text{p},n}\}} \frac{\sum_{n_1=1}^{N} x_{\text{p},n_1}^{10}}{\sum_{n_2=1}^{N} \sum_{\substack{n_1=1 \\ n_1 \neq n_2}}^{N} \left(x_{\text{p},n_1}^6 x_{\text{p},n_2}^{10} - x_{\text{p},n_1}^8 x_{\text{p},n_2}^8 \right)},
$$

$$
\text{s.t. } x_{\text{p},n_1}^2 \leq E_{\text{max,pilot}}, \ \forall n_1 \in \{1, 2, \ldots, N\}
$$

$$
\sum_{n_1=1}^{N} x_{\text{p},n_1}^2 \leq E_{\text{tot,pilot}},
\tag{5.52}
$$

where $E_{\text{tot,pilot}}$ is the total pilot-signal energy obtained from (5.51) as $E_{\text{tot,pilot}} = NQ(1 - \beta_{\text{c}}/4)E_{\text{norm,pilot}}$, and $E_{\text{max,pilot}}$ is a design value representing the allowable peak pilot-signal energy. By using the transformation $y_{\text{p},n} = \ln(x_{\text{p},n}^2/E_{\text{tot,pilot}}) \ \forall n$ and $\mathfrak{y}_{\text{max}} = E_{\text{max,pilot}}/E_{\text{tot,pilot}}$, the optimization problem in (5.52) is equivalent to

$$
\min_{\{y_{\text{p},n}\}} \frac{\sum_{n_1=1}^{N} e^{5y_{\text{p},n_1}}}{\sum_{n_2=1}^{N} \sum_{\substack{n_1=1 \\ n_1 \neq n_2}}^{N} \left(e^{5y_{\text{p},n_1} + 3y_{\text{p},n_2}} - e^{4y_{\text{p},n_1} + 4y_{\text{p},n_2}} \right)},
$$

$$
\text{s.t. } e^{y_{\text{p},n_1}} \leq \mathfrak{y}_{\text{max}}, \ \forall n_1 \in \{1, 2, \ldots, N\},
$$

$$
\sum_{n_1=1}^{N} e^{y_{\text{p},n_1}} \leq 1.
\tag{5.53}
$$

We note that the numerator of the objective function of (5.53) is a convex function, while the denominator is a nonconvecx function since it is a difference of two convex functions

(DCFs). Therefore, the objective function of (5.53) is nonconvex. We also note that the two constraints of (5.53) are convex. Therefore, the optimization problem in (5.53) is nonconvex due to the nonconvexity of the objective function.

To make the optimization problem in (5.53) tractable, we propose to divide it into two optimization problems P1 and P2 as follows:

$$\text{P1: } \min_{\tau} \frac{\tau}{-h(\tau)}$$

$$\text{s.t. } 0 < \tau \leq \tau_{\max}, \tag{5.54}$$

$$\text{P2: } h(\tau) = \min_{\{y_{\text{p},n}\}} \sum_{n_2=1}^{N} \sum_{\substack{n_1=1 \\ n_1 \neq n_2}}^{N} \left(e^{4y_{\text{p},n_1}+4y_{\text{p},n_2}} - e^{5y_{\text{p},n_1}+3y_{\text{p},n_2}} \right),$$

$$\text{s.t. } \sum_{n_1=1}^{N} e^{5y_{\text{p},n_1}} = \tau,$$

$$e^{y_{\text{p},n_1}} \leq \mathfrak{y}_{\max}, \ \forall n_1 \in \{1, 2, \ldots, N\},$$

$$\sum_{n_1=1}^{N} e^{y_{\text{p},n_1}} \leq 1, \tag{5.55}$$

where $\tau_{\max}$ is obtained as

$$\tau_{\max} = \max_{\{y_{\text{p},n}\}} \sum_{n=1}^{N} e^{5y_{\text{p},n}},$$

$$\text{s.t. } e^{y_{\text{p},n}} \leq \mathfrak{y}_{\max}, \ \forall n \in \{1, 2, \ldots, N\},$$

$$\sum_{n_1=1}^{N} e^{y_{\text{p},n}} \leq 1,$$

$$= \lfloor \mathfrak{y}_{\max}^{-1} \rfloor \mathfrak{y}_{\max}^{5} + (1 - \lfloor \mathfrak{y}_{\max}^{-1} \rfloor \mathfrak{y}_{\max})^{5}. \tag{5.56}$$

The idea is to fix the numerator of the objective function of (5.53), and that leads to the first constraint in (5.55). The proof of the solution equivalence of problems (5.53) and (5.54) can be easily obtained by following the argument in [77]. To solve P1, we propose

100

to solve sequences of P2 with different values of τ efficiently spanning the interval $(0, \tau_{\max}]$. Then, we pick $\{y_{\mathrm{p},n}\}$ associated to τ that minimizes $-\tau/h(\tau)$. Finally, we obtain $\{x_{\mathrm{p},n}\}$ as $x_{\mathrm{p},n} = e^{(y_{\mathrm{p},n}+\ln E_{\mathrm{tot,pilot}})/2}$, $\forall n$. Therefore, in the following, we focus on solving P2. Then, we summarize the whole proposed solution.

As discussed above, the objective function of P2 is DCFs which is nonconvex. We propose a suboptimal successive convex approximation approach to solve P2. Approximating the negative exponential functions of the objective function of P2 by their first order Taylor expansions around $\{y_{\mathrm{p},n}^{(0)}\}$, we obtain a convex optimization problem $\tilde{\mathrm{P}}2$ as

$$
\tilde{\mathrm{P}}2: \tilde{h}(\tau) = \min_{\{y_{\mathrm{p},n}\}} \sum_{n_2=1}^{N} \sum_{\substack{n_1=1 \\ n_1 \neq n_2}}^{N} \left(e^{4y_{\mathrm{p},n_1}+4y_{\mathrm{p},n_2}} - e^{5y_{\mathrm{p},n_1}^{(0)}+3y_{\mathrm{p},n_2}^{(0)}} \right.
$$

$$
\left. \times \left(5y_{\mathrm{p},n_1} + 3y_{\mathrm{p},n_2} - (5y_{\mathrm{p},n_1}^{(0)} + 3y_{\mathrm{p},n_2}^{(0)}) + 1 \right) \right),
$$

$$
\text{s.t.} \sum_{n_1=1}^{N} e^{5y_{\mathrm{p},n_1}} = \tau,
$$

$$
e^{y_{\mathrm{p},n_1}} \leq \mathfrak{y}_{\max}, \ \forall n_1 \in \{1, 2, \ldots, N\},
$$

$$
\sum_{n_1=1}^{N} e^{y_{\mathrm{p},n_1}} \leq 1, \tag{5.57}
$$

where $\{y_{\mathrm{p},n}^{(0)}\}$ is an initial solution. Using this initial solution, the convex optimization problem $\tilde{\mathrm{P}}2$ can be solved using any convex optimization solver (e.g., CVX [50, 49]) to get the optimum value $\{y_{\mathrm{p},n}\}$. Then, we set $y_{\mathrm{p},n}^{(0)} = y_{\mathrm{p},n}$, $\forall n$ and solve the optimization problem $\tilde{\mathrm{P}}2$ again. The new solution is then used as an initial solution for another iteration, and so on. Eventually, $\{y_{\mathrm{p},n}\}$ will converge to a local solution of the original nonconvex problem P2 [21].

The proposed pilot sequence design is summarized in Algorithm 5, where $d\tau$ is the step size that we use to linearly search for the optimal τ in the interval $(0, \tau_{\max}]$, and $\mathcal{P}_{\mathrm{pilot}}$ defines how many times we successively solve $\tilde{\mathrm{P}}2$ to get the solution of P2. As an initial solution, we generate uniformly-distributed random pilot-symbol energies (using a uniform

random generator, e.g., "unifrnd" of MATLAB) while satisfying the allowable peak pilot-symbol energy. We properly scale this initial solution to satisfy the total pilot-symbol energy constraint and the first constraint of $\tilde{\text{P}}2$. Then, we solve sequences of P2 with different values of τ. Then, we pick $\{y_{\text{p},n}\}$ associated to τ that minimizes $-\tau/h\left(\tau\right)$. Finally, we obtain the pilot sequence $\{x_{\text{p},n}\}$ as $x_{\text{p},n} = e^{(y_{\text{p},n}+\ln E_{\text{tot,pilot}})/2}$, $\forall n$. The computational complexity of the proposed pilot sequence design is $\mathcal{O}\left((2N + \lfloor\tau_{\max}/d\tau\rfloor\,\mathcal{P}_{\text{pilot}}(N^6 + N^2))\right)$ [21]. Note that the proposed pilot sequence design is done once and offline.

Algorithm 5 Proposed Pilot Sequence Design

Inputs: $E_{\text{norm,pilot}}$, $E_{\text{max,pilot}}$, N, Q, β_{c}, $d\tau$, $\mathcal{P}_{\text{pilot}}$
$E_{\text{tot,pilot}} = NQ(1 - \beta_{\text{c}}/4)E_{\text{norm,pilot}}$
$\mathfrak{y}_{\max} = E_{\text{max,pilot}}/E_{\text{tot,pilot}}$
$\tau_{\max} = \lfloor\mathfrak{y}_{\max}^{-1}\rfloor\,\mathfrak{y}_{\max}^5 + (1 - \lfloor\mathfrak{y}_{\max}^{-1}\rfloor\,\mathfrak{y}_{\max})^5$
$y_{\text{p},n}^{(0)} = \ln(\text{unifrnd}\,(0,\mathfrak{y}_{\max}))$, $\forall n$
if $\sum_{n=1}^{N} e^{y_{\text{p},n}^{(0)}} > 1$
    $y_{\text{p},n}^{(0)} := \ln(e^{y_{\text{p},n}^{(0)}}/\sum_{n=1}^{N} e^{y_{\text{p},n}^{(0)}})$, $\forall n$
end if
if $\sum_{n=1}^{N} e^{5y_{\text{p},n}^{(0)}} > d\tau$
    $y_{\text{p},n}^{(0)} := \ln(e^{y_{\text{p},n}^{(0)}}/(\sum_{n=1}^{N} e^{5y_{\text{p},n}^{(0)}})^{1/5})$, $\forall n$
end if
for $w = 1$ to $\lfloor\tau_{\max}/d\tau\rfloor$
    $\tau^{(w)} = w\,d\tau$
    $y_{\text{p},n}^{(0)} = y_{\text{p},n}^{(w-1)}$, $\forall n$
    for $p = 1$ to $\mathcal{P}_{\text{pilot}}$ (or any other stopping criterion)
        Obtain $\tilde{h}_w = \tilde{h}\left(\tau^{(w)}\right)$ and $\{y_{\text{p},n}\}$ by solving $\tilde{\text{P}}2$ using
        CVX solver with $\tau = \tau^{(w)}$ and $\{y_{\text{p},n}^{(0)}\}$
        $y_{\text{p},n}^{(0)} = y_{\text{p},n}$, $\forall n$
    end for
    $y_{\text{p},n}^{(w)} = y_{\text{p},n}$, $\forall n$
end for
$w_{\text{opt}} = \arg\min_{w} \{-\tau^{(w)}/\tilde{h}_w\}$
$x_{\text{p},n} = e^{(y_{\text{p},n}^{(w_{\text{opt}})}+\ln E_{\text{tot,pilot}})/2}$, $\forall n$
Output: $\{x_{\text{p},n}\}$

5.7 Simulation Results

In this section, we evaluate the performance of our proposed system parameters estimation and precompensation for low-cost nonlinear Tera-Hertz transmitters in the presence of I/Q imbalance. With the proposed pilot sequence design and using the proposed ML system parameters estimator, we compare the estimate variances of the system parameters with the derived CRLBs. Then, we evaluate the performance of the proposed precompensation algorithms.

We consider a desired transmitted signal $\tilde{s}(t)$ constructed from a block of $M = 128$ 16-QAM modulation symbols convolved with a desired pulse-shaping filter $g(t)$. Unless mentioned otherwise, we consider $g(t)$ as an RRC pulse-shaping filter with time span of $12T_{\mathrm{s}}$ and roll-off factor $\beta_{\mathrm{s}} = 0.332$. We generate our numerical results based on 10^4 16-QAM blocks. The NLD parameters are $A_3 = 1.26 - 0.081j$ and $A_5 = -0.96 + 0.29j$, and the I/Q imbalance parameters are $\epsilon = 20\%$ and $\phi = 10°$. The PSD of the AWGN at the feedback RF chain is -174 dBm/Hz. The symbol data rate R_{s} is 20×10^6 symbols/s. The proposed compensation scheme utilizes an RC construction filter $q(t)$ as in (5.14) with time span of $12T_{\mathrm{c}}$ and roll-off factor $\beta_{\mathrm{c}} = 0.332$. We set $\mathcal{P}_{\mathrm{comp}} = \mathcal{P}_{\mathrm{IQ}} = \mathcal{P}_{\mathrm{pilot}} = 5$. Unless mentioned otherwise, the upsampling factor L is 2.

5.7.1 System Parameters Estimation

We consider a pilot sequence with length $N = 100$ and peak pilot-symbol energy $E_{\mathrm{max,pilot}}$ of 6 dB above the normalized pilot-signal energy $E_{\mathrm{norm,pilot}}$. Fig. 5.6 shows $\mathrm{CRLB}(\rho_1)$ at various normalized pilot-signal energies $E_{\mathrm{norm,pilot}}$ with two pilot sequence designs as follows: 1) Random-energy pilot sequence design, where the pilot-symbol energies are generated randomly while satisfying the allowable peak pilot-symbol energy and the total pilot-symbol energy constraints. 2) Proposed pilot sequence design, where the pilot-symbol energies are optimized as in Algorithm 5. We observe that $\mathrm{CRLB}(\rho_1)$ with the proposed pilot sequence
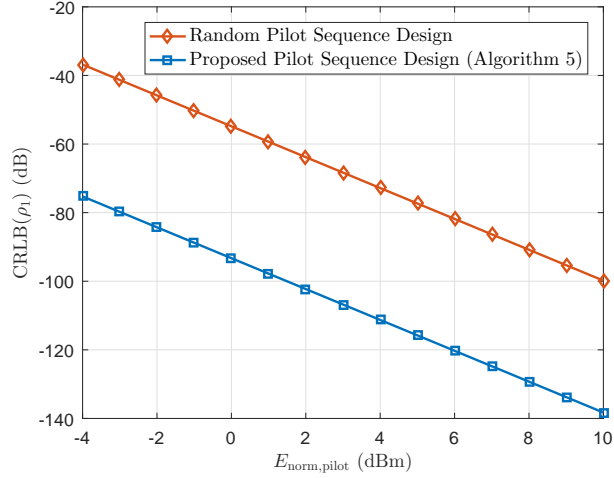
Figure 5.6. CRLB(ρ_1) at various normalized pilot-signal energies $E_{\text{norm,pilot}}$ with two pilot sequence designs.

design is about 40 dB smaller than that with the random-energy pilot sequence design, which verifies the effectiveness of the proposed pilot sequence design. This result is expected since the proposed pilot sequence is designed to minimize CRLB(ρ_1), while the random-energy pilot sequence is not.

Fig. 5.7 compares the MSEs of the proposed ML estimator with the derived CRLBs of the system parameters at various normalized pilot-signal energies $E_{\text{norm,pilot}}$. From Fig. 5.7, we observe that the initial estimates of the I/Q imbalance parameters yield MSEs of 10 dB above the CRLBs, while the initial estimates of the NLD parameters yield MSEs very close to the CRLBs. Using the proposed ML estimator (with 5 iterations), we observe that all the MSEs of the estimates of the system parameters are very close to the CRLBs, which verifies the excellent performance of the proposed ML system parameters estimator.

5.7.2 Effects of the Pulse-shaping Filter and I/Q Imbalance without Precompensation

For illustration, we consider two types of pulse-shaping filters for $g(t)$: 1) Rectangular pulse-shaping filter with time span of T_{s} and 2) RRC pulse-shaping filter with time span of $12T_{\text{s}}$ and

104
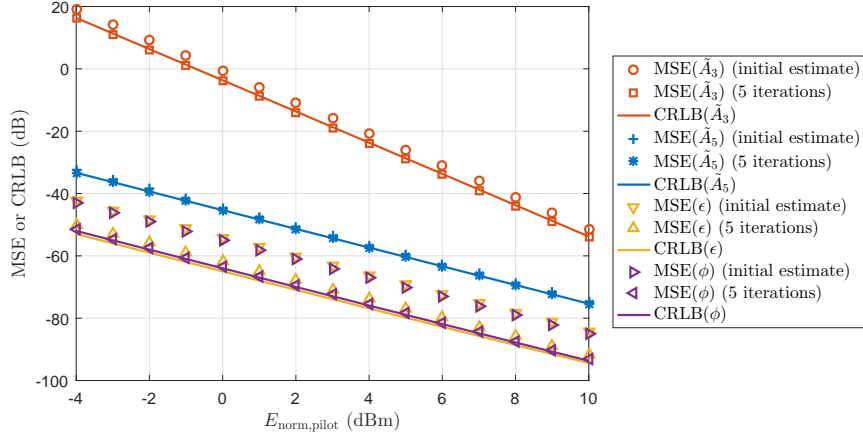
Figure 5.7. MSEs of the proposed estimator and CRLBs for the system parameters estimation at various normalized pilot-signal energies $E_{\mathrm{norm,pilot}}$.

roll-off factor $\beta_{\mathrm{s}} = 0.332$. For the existing scheme in [69, 68] (i.e., without precompensation) with $E_{\mathrm{norm}} = 20$ dBm, Fig. 5.8 shows the PSD of the signals before and after the NLD and the transmitter output constellation. In the absence of I/Q imbalance, unlike QPSK modulation with rectangular pulse-shaping which just experiences predefined permutation due to the third-order power of the NLD and slight distortion due to the fifth-order power of the NLD as shown in Fig. 5.2, 16-QAM modulation with rectangular pulse-shaping experiences serious distortions in the transmitter output constellation. Comparing the results between the two pulse-shaping filters in the absence of I/Q imbalance, one can observe that a larger filter span yields a lower spectrum sidelobe but a severe degradation of the transmitter output constellation due to nonlinear distortion. In the presence of I/Q imbalance, we observe that the I/Q imbalance adds additional distortions to the 16-QAM constellation for both pulse-shaping filters.

5.7.3 Effects of the Proposed Precompensation Schemes

Without Algorithm 2, we name the proposed precompensation scheme of section 5.4 as "proposed basic precompensation scheme", while with Algorithm 2 we name it as "proposed improved precompensation scheme". Based on the estimates of the system parameters with
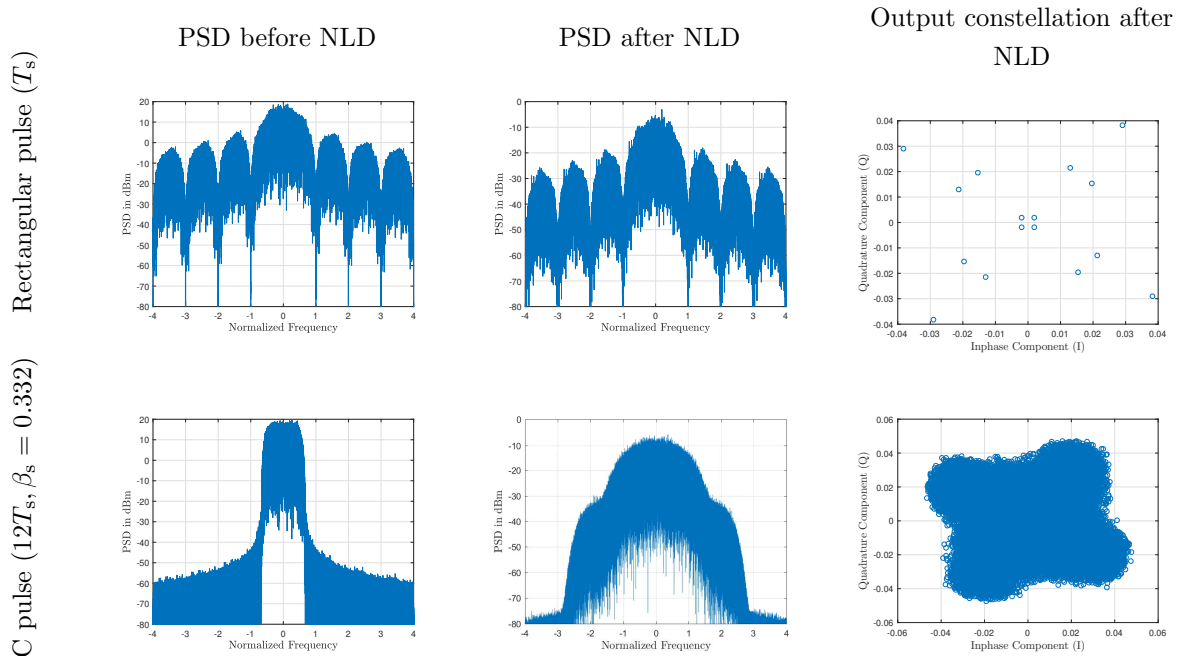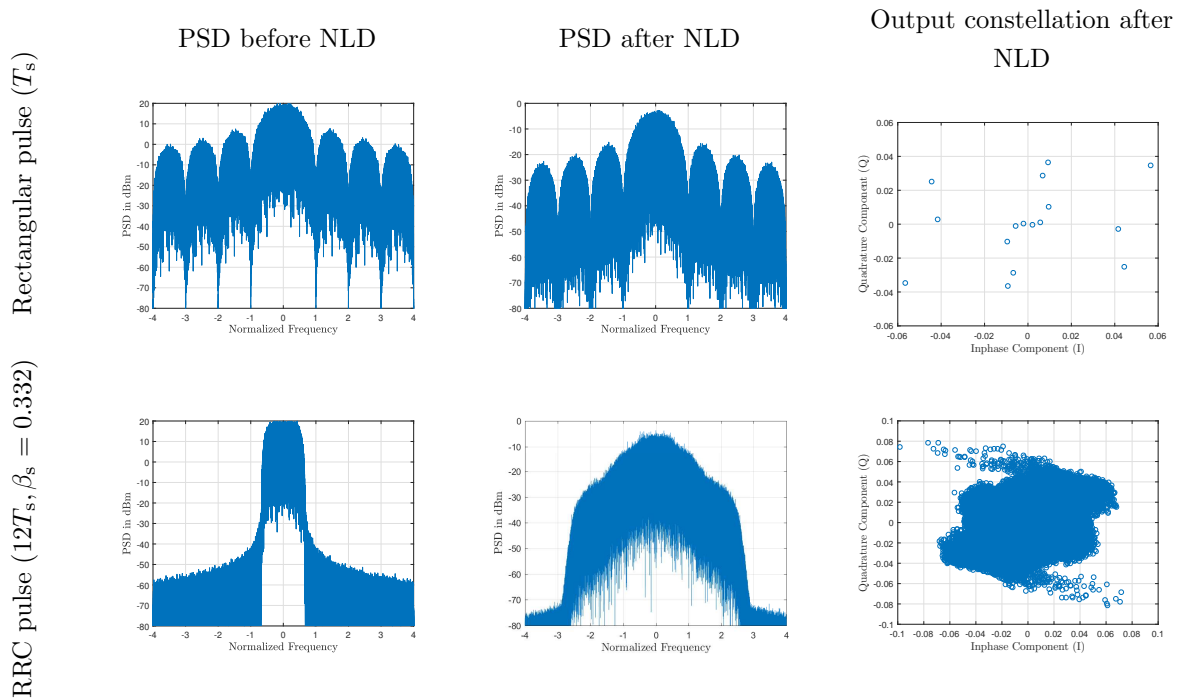
Figure 5.8. Combined effects of the baseband pulse-shaping filter and the NLD on 16-QAM without precompensation (the existing scheme in [69, 68]) in the absence/presence of I/Q imbalance.

Figure 5.9. Combined effects of the baseband pulse-shaping filter and the NLD on 16-QAM with the proposed precompensation schemes with the upsampling factor $L = 2$ in the presence of I/Q imbalance.

$E_{\mathrm{norm}} = 20$ dBm, Fig. 5.9 shows the performances of the proposed basic precompensation scheme and the proposed improved precompensation scheme with the upsampling factor $L = 2$. Fig. 5.10 extends the results of Fig. 5.9 to the case of the upsampling factor $L = 4$. Comparing the signal constellation results of the larger filter span between the existing scheme in Fig. 5.8 and the proposed schemes in Fig. 5.9 and Fig. 5.10, we can clearly see performance advantages of the proposed precompensation schemes in suppressing the nonlinear distortions and the I/Q imbalance distortion. Compared to the proposed basic precompensation scheme, we observe that the proposed improved precompensation scheme has better PSDs (lower sidelobe levels and no spikes) before and after the NLD and less constellation distortion. This performance improvement comes from the fact that the proposed improved precompensation scheme (with Algorithm 2) modifies the virtual samples to have slower time variations, which results in a more-spectrally-constrained construction of $x_{\mathrm{pre}}(t)$
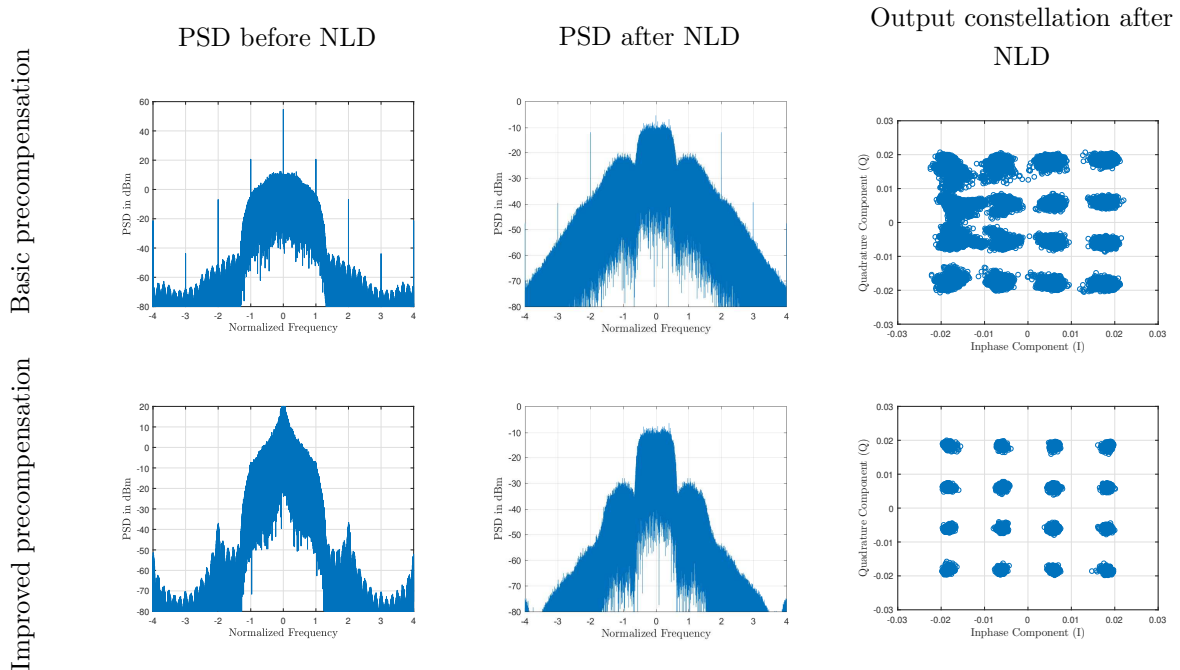
Figure 5.10. Combined effects of the baseband pulse-shaping filter and the NLD on 16-QAM with the proposed precompensation schemes with the upsampling factor $L = 4$ in the presence of I/Q imbalance.

from the virtual samples. The proposed precompensation schemes with $L = 4$ yield better PSDs and less distortions than those with $L = 2$ due to its higher sampling rate LR_s which results in smaller aliasing issue and better construction for $x_{\text{pre}}(t)$ from the virtual samples. The cost for a larger L is the increased computational complexity in the precompensation stage.

5.7.4 Effects of the Roll-off Factor of the RC Construction Filter

With the proposed improved precompensation scheme with $E_{\text{norm}} = 20$ dBm, Fig. 5.11 shows the normalized error vector magnitude (NEVM) of transmitter output constellation[5] (left y-axis) and normalized (with respect to the main lobe peak) maximum sidelobe level (NMSL)

---

[5]NEVM is defined as NEVM $= \sqrt{\sum_{m=1}^{M} |a_m - \tilde{a}_m|^2 / \sum_{m=1}^{M} |a_m|^2}$, where $\tilde{a}_m$ is the measured modulation symbol [40].
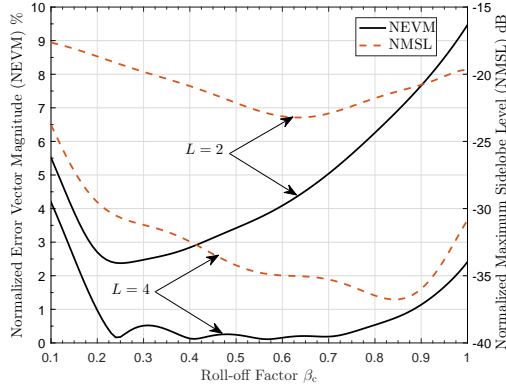
Figure 5.11. NEVM and NMSL versus roll-off factor β_c of the RC construction filter $q(t)$. of transmitter output (right y-axis) versus the roll-off factor β_c of the RC construction filter $q(t)$ with different values of the upsampling factor L. We observe that the minimum NEVM is not obtained along with the minimum NMSL. In other words, we have a trade off between NEVM and NMSL. Therefore, appropriate values for the roll-off factor β_c of the construction filter $q(t)$ and the upsampling factor L have to be determined according to the system requirements. For example, if the system requires the sidelobe level to be at most -20 dB, we can select $L = 2$ for lower complexity and $\beta_c = 0.32$ which yields an NEVM of about 2.5%. If the system needs to limit its sidelobe level to be below -35 dB, then we can select $L = 4$ and $\beta_c = 0.7$ which yields an NEVM of about 0.19%.

5.7.5 Effects of Frequency-Dependent I/Q Imbalance and Frequency-Selective NLDs

The frequency-dependent I/Q imbalance is modeled by a frequency-independent I/Q imbalance followed by imbalance filters $\varsigma_I(t)$ and $\varsigma_Q(t)$ for the I and Q branches respectively [89]. Using the Wiener model [46], the frequency selectivity of the NLDs (the power amplifier and the frequency multiplier) can be modeled by a filter $\varrho(t)$ followed by the memoryless polynomial model in (7). Therefore, $\varsigma_I(t)$ and $\varrho(t)$ can be combined as $\tilde{\varsigma}_I(t) = \varsigma_I(t) * \varrho(t)$, and $\varsigma_Q(t)$ and $\varrho(t)$ can be combined as $\tilde{\varsigma}_Q(t) = \varsigma_Q(t) * \varrho(t)$.
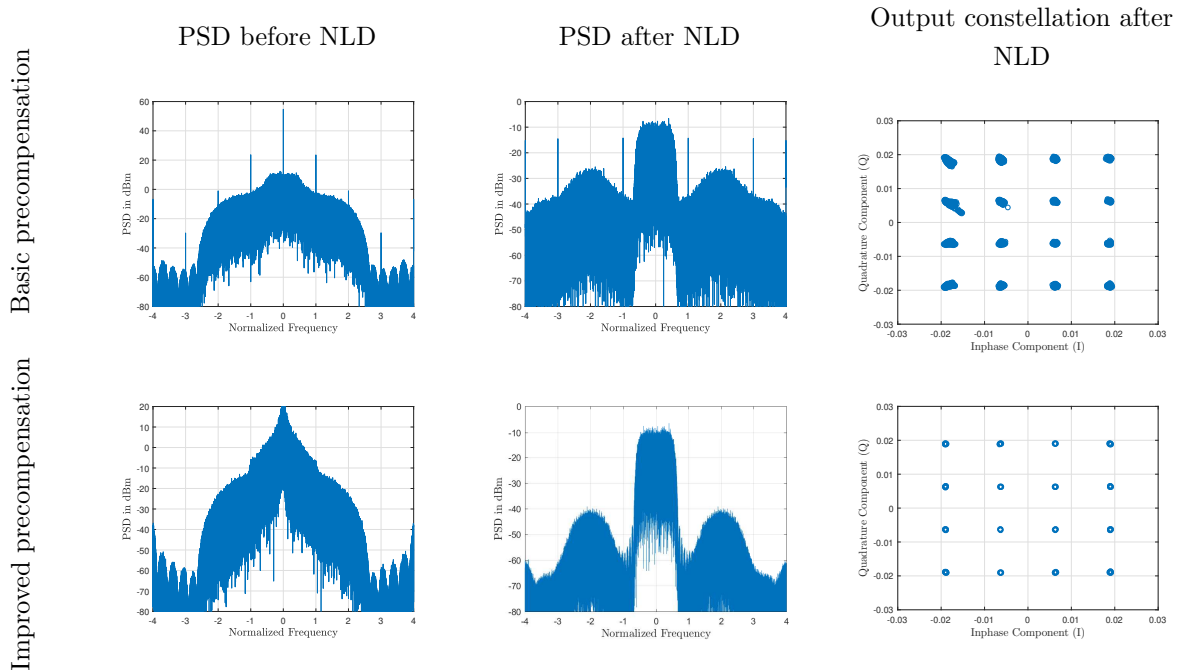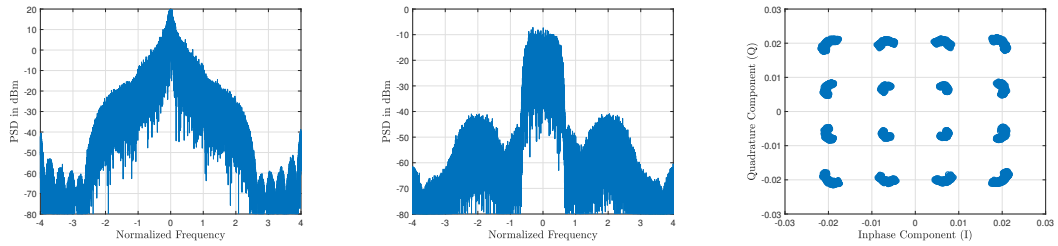
Figure 5.12. Combined effects of the baseband pulse-shaping filter and the NLD on 16-QAM with the proposed improved precompensation scheme with the upsampling factor $L = 4$ in the presence of frequency-dependent I/Q imbalance.

With $E_{\mathrm{norm}} = 20$ dBm, Fig. 5.12 shows the performances of the proposed improved precompensation scheme with the upsampling factor $L = 4$ in the presence of frequency-dependent I/Q imbalance and frequency-selective NLDs, where $\tilde{\varsigma}_{\mathrm{I}}(t) = 0.015\delta(t) + \delta(t - T_{\mathrm{s}}) + 0.02\delta(t - 2T_{\mathrm{s}}) + 0.01\delta(t - 3T_{\mathrm{s}})$ and $\tilde{\varsigma}_{\mathrm{Q}}(t) = 0.01\delta(t) + \delta(t - T_{\mathrm{s}}) + 0.03\delta(t - 2T_{\mathrm{s}}) + 0.02\delta(t - 3T_{\mathrm{s}})$. From Fig. 5.10 and Fig. 5.12, we observe that the frequency-dependent I/Q imbalance and frequency-selective NLDs add additional distortions to the signal constellation. This distortion comes from the ISI introduced by the filters of the I/Q imbalance and the NLDs. The NEVM of the signal constellation with frequency-dependent I/Q imbalance and frequency-selective NLDs is 8.74% which still meets the requirements of the 3rd generation partnership project (3GPP) standard ($< 12.5\%$) [37]. Therefore, the proposed precompensation and estimation schemes are still applicable.

5.8   Conclusion

Low cost and energy efficiency are crucial factors for broader deployment of THz applications. The recently proposed frequency-multiplier-last THz transmitter architecture is the most promising one in this regard, but it suffers from nonlinear distortions and is incapable of supporting QAM transmission. In this chapter, we developed solutions to overcome limitations/drawbacks of this architecture. By incorporating nonlinearity aspects of the low-cost THz devices and the I/Q imbalance into the signal model, our study shows that when the

pulse-shaping filter spans more than one symbol, the existing result that QPSK modulation does not experience distortions except a fixed permutation of the constellation points no longer holds. For out-of-band spectrum control, the pulse-shaping filter span of several symbol intervals is needed and in this case both QPSK and QAM suffer severe nonlinear distortions.

We developed a precompensation scheme which can suppress nonlinear distortions, mitigate the I/Q imbalance effect, and enable QAM transmission capability. The proposed precompensation scheme offers very attractive output power spectral density control and nonlinear distortion suppression. Using a separate measurement circuitry for testing/calibrating the THz transmitters, we proposed a maximum-likelihood (ML) estimator to estimate the NLD parameters and the I/Q imbalance parameters. We derived closed-form expressions for the Cramér–Rao lower bounds (CRLBs) of the system parameters estimates as benchmark metrics to evaluate the performance of the proposed ML estimator. We designed the pilot sequence used in testing/calibrating the THz transmitters to enhance the performance of the proposed ML estimator. We also presented how to trade off between complexity and performance of the proposed scheme by means of the upsampling factor and the construction filter roll-off factor. The effects of frequency-dependent I/Q imbalance and frequency-selective NLDs are investigated, and the results show that the proposed scheme can keep the total distortion within the 3GPP requirement.

Appendix A - Steps to Obtain (5.17)

For a real-valued objective function $f(x)$ of a complex variable x, it is shown in [73] that the complex gradient $\nabla_x (f)$ is given by $\nabla_x (f) = (\partial f/\partial \Re\{x\} + \partial f/\partial \Im\{x\})/2 = \partial f/\partial x^*$. Therefore, we write the sample-wise square-error objective function W_k as a function of c_k and c_k^* as

$$W_k = \left|\tilde{A}_3\right|^2 (c_k^*)^3 c_k^3 + 2\Re\left\{\tilde{A}_3^* \tilde{A}_5\right\} (c_k^*)^4 c_k^4 + \left|\tilde{A}_5\right|^2 (c_k^*)^5 c_k^5$$

$$-\tilde{A}_3^* \left(c_k^*\right)^3 \tilde{s}_k - \tilde{A}_5^* \left(c_k^*\right)^4 c_k \tilde{s}_k - \tilde{A}_5 c_k^* c_k^4 \tilde{s}_k^* - \tilde{A}_3 c_k^3 \tilde{s}_k^* + |\tilde{s}_k|^2. \tag{5.58}$$

Then, we obtain $\nabla_{c_k} W_k$ by taking the partial derivative $\partial W_k / \partial c_k^*$ of W_k with respect to c_k^* while treating c_k as a constant, which straightly yields (5.17).

Appendix B - Steps to Obtain $(5.34)-(5.36)$

Using the chain rule via the intermediate complex variables α_1 and α_2, we have

$$
\begin{aligned}
\frac{\partial \tilde{\mathcal{L}}}{\partial \epsilon} &= \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_1^*} \frac{\partial \alpha_1^*}{\partial \epsilon} + \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_1} \frac{\partial \alpha_1}{\partial \epsilon} + \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_2^*} \frac{\partial \alpha_2^*}{\partial \epsilon} + \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_2} \frac{\partial \alpha_2}{\partial \epsilon} \\
&= 2\Re\{ \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_1^*} \frac{\partial \alpha_1^*}{\partial \epsilon} + \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_2^*} \frac{\partial \alpha_2^*}{\partial \epsilon} \}.
\end{aligned}
\tag{5.59}
$$

Similarly, we have

$$\frac{\partial \tilde{\mathcal{L}}}{\partial \phi} = 2\Re\{ \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_1^*} \frac{\partial \alpha_1^*}{\partial \phi} + \frac{\partial \tilde{\mathcal{L}}}{\partial \alpha_2^*} \frac{\partial \alpha_2^*}{\partial \phi} \}. \tag{5.60}$$

Next, we obtain $\partial \tilde{\mathcal{L}} / \partial \alpha_i^*$ as $\partial \tilde{\mathcal{L}} / \partial \alpha_i^* = \sum_{n=1}^N \lambda_{i,n}$, and by applying the same procedure as in Appendix A, we obtain $\lambda_{i,n}$ as in (5.36). Finally, we simply obtain $\partial \alpha_1^* / \partial \epsilon = \cos\phi + j\sin\phi = e^{j\phi}$, $\partial \alpha_2^* / \partial \epsilon = -\cos\phi + j\sin\phi = -e^{-j\phi}$, $\partial \alpha_1^* / \partial \phi = j(1 + \epsilon)(\cos\phi + j\sin\phi) = j\alpha_1^*$, and $\partial \alpha_2^* / \partial \phi = -j(1 - \epsilon)(\cos\phi - j\sin\phi) = -j\alpha_2^*$. By arranging these results, we can obtain $\partial \tilde{\mathcal{L}} / \partial \epsilon$ and $\partial \tilde{\mathcal{L}} / \partial \phi$ as in $(5.34)-(5.36)$.

Appendix C - Proof of Proposition 2

Here, we provide a sketch for the proof since the full proof is based on lengthy algebraic manipulations. First, with the help of Appendices A and B, we obtain $\partial \mathcal{L} / \partial \tilde{A}_3^*$, $\partial \mathcal{L} / \partial \tilde{A}_5^*$, $\partial \mathcal{L} / \partial \epsilon$, and $\partial \mathcal{L} / \partial \phi$. Then, we use the definitions of $\xi_{1,n}$ and $\xi_{2,n}$ in (5.23) and (5.25) to simplify these derivatives as in $(5.61)-(5.64)$. Then, the upper triangle entries of the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ can be obtained as in $(5.65)-(5.74)$ by using the facts that $\mathbb{E}\{\eta_{i,n}\eta_{i,n}^*\} = \sigma_\eta^2 \, \forall i, n$, $\mathbb{E}\{\eta_{i,n}\eta_{i,m}^*\} = 0 \, \forall i, n \neq m$, $\mathbb{E}\{\eta_{i,n}\eta_{j,m}^*\} = 0 \, \forall n, m, i \neq j$, and $\mathbb{E}\{(\eta_{i,n})^2\} = \mathbb{E}\{(\eta_{i,n}^*)^2\} = 0 \, \forall i, n$.

112

$$\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*} = \frac{r_3}{\sigma_\eta^2} \sum_{n=1}^{N} \left\{ (\alpha_1^*)^3 x_{\mathrm{p},n}^3 \eta_{1,n} + (\alpha_2^*)^3 x_{\mathrm{p},n}^3 \eta_{2,n} \right\}, \tag{5.61}$$

$$\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*} = \frac{r_5}{\sigma_\eta^2} \sum_{n=1}^{N} \left\{ (\alpha_1^*)^3 |\alpha_1|^2 x_{\mathrm{p},n}^5 \eta_{1,n} + (\alpha_2^*)^3 |\alpha_2|^2 x_{\mathrm{p},n}^5 \eta_{2,n} \right\}, \tag{5.62}$$

$$\frac{\partial \mathcal{L}}{\partial \epsilon} = \frac{1}{\sigma_\eta^2} \sum_{n=1}^{N} \left\{ 2\Re\{ (3r_3\tilde{A}_3^*(\alpha_1^*)^2 x_{\mathrm{p},n}^3 \eta_{1,n} + 4r_5\tilde{A}_5^*(\alpha_1^*)^3 \alpha_1 x_{\mathrm{p},n}^5 \eta_{1,n} + r_5\tilde{A}_5(\alpha_1)^4 x_{\mathrm{p},n}^5 \eta_{1,n}^*) e^{j\phi} \} \right.$$
$$\left. - 2\Re\{ (3r_3\tilde{A}_3^*(\alpha_2^*)^2 x_{\mathrm{p},n}^3 \eta_{2,n} + 4r_5\tilde{A}_5^*(\alpha_2^*)^3 \alpha_2 x_{\mathrm{p},n}^5 \eta_{2,n} + r_5\tilde{A}_5(\alpha_2)^4 x_{\mathrm{p},n}^5 \eta_{2,n}^*) e^{-j\phi} \} \right\}, \tag{5.63}$$

$$\frac{\partial \mathcal{L}}{\partial \phi} = \frac{1}{\sigma_\eta^2} \sum_{n=1}^{N} \left\{ 2\Re\{ (3r_3\tilde{A}_3^*(\alpha_1^*)^2 x_{\mathrm{p},n}^3 \eta_{1,n} + 4r_5\tilde{A}_5^*(\alpha_1^*)^3 \alpha_1 x_{\mathrm{p},n}^5 \eta_{1,n} + r_5\tilde{A}_5(\alpha_1)^4 x_{\mathrm{p},n}^5 \eta_{1,n}^*)(j\alpha_1^*) \} \right.$$
$$\left. - 2\Re\{ (3r_3\tilde{A}_3^*(\alpha_2^*)^2 x_{\mathrm{p},n}^3 \eta_{2,n} + 4r_5\tilde{A}_5^*(\alpha_2^*)^3 \alpha_2 x_{\mathrm{p},n}^5 \eta_{2,n} + r_5\tilde{A}_5(\alpha_2)^4 x_{\mathrm{p},n}^5 \eta_{2,n}^*)(j\alpha_2^*) \} \right\}. \tag{5.64}$$

$$\mathbb{E}\{ (\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})^* (\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*}) \} = \frac{1}{\sigma_\eta^2} r_3^2 (|\alpha_1|^6 + |\alpha_2|^6) \sum_{n=1}^{N} x_{\mathrm{p},n}^6 \tag{5.65}$$

$$\mathbb{E}\{ (\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*})^* (\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*}) \} = \frac{1}{\sigma_\eta^2} r_5^2 (|\alpha_1|^{10} + |\alpha_2|^{10}) \sum_{n=1}^{N} x_{\mathrm{p},n}^{10} \tag{5.66}$$

$$\mathbb{E}\{ (\frac{\partial \mathcal{L}}{\partial \tilde{A}_3^*})^* (\frac{\partial \mathcal{L}}{\partial \tilde{A}_5^*}) \} = \frac{1}{\sigma_\eta^2} r_3 r_5 (|\alpha_1|^8 + |\alpha_2|^8) \sum_{n=1}^{N} x_{\mathrm{p},n}^8 \tag{5.67}$$

$$\mathbb{E}\{ (\frac{\partial \mathcal{L}}{\partial \epsilon})^2 \} = \frac{1}{\sigma_\eta^2} \sum_{n=1}^{N} \left\{ 2\Re\{ (6r_3 r_5 \tilde{A}_3^* \tilde{A}_5 \alpha_1^2 |\alpha_1|^4 x_{\mathrm{p},n}^8 + 8r_5^2 |\tilde{A}_5|^2 \alpha_1^2 |\alpha_1|^6 x_{\mathrm{p},n}^{10}) e^{j2\phi} \} \right.$$
$$+ 2\Re\{ (6r_3 r_5 \tilde{A}_3^* \tilde{A}_5 \alpha_2^2 |\alpha_2|^4 x_{\mathrm{p},n}^8 + 8r_5^2 |\tilde{A}_5|^2 \alpha_2^2 |\alpha_2|^6 x_{\mathrm{p},n}^{10}) e^{-j2\phi} \}$$
$$+ 18r_3^2 |\tilde{A}_3|^2 (|\alpha_1|^4 + |\alpha_2|^4) x_{\mathrm{p},n}^6 + 34r_5^2 |\tilde{A}_5|^2 (|\alpha_1|^8 + |\alpha_2|^8) x_{\mathrm{p},n}^{10}$$
$$\left. + 2\Re\{ 24r_3 r_5 \tilde{A}_3^* \tilde{A}_5 (|\alpha_1|^6 + |\alpha_2|^6) x_{\mathrm{p},n}^8 \} \right\} \tag{5.68}$$

$$\mathbb{E}\{ (\frac{\partial \mathcal{L}}{\partial \phi})^2 \} = \frac{1}{\sigma_\eta^2} \sum_{n=1}^{N} \left\{ -2\Re\{ (6r_3 r_5 \tilde{A}_3^* \tilde{A}_5 \alpha_1^2 |\alpha_1|^4 x_{\mathrm{p},n}^8 + 8r_5^2 |\tilde{A}_5|^2 \alpha_1^2 |\alpha_1|^6 x_{\mathrm{p},n}^{10})(\alpha_1^*)^2 \} \right.$$
$$- 2\Re\{ (6r_3 r_5 \tilde{A}_3^* \tilde{A}_5 \alpha_2^2 |\alpha_2|^4 x_{\mathrm{p},n}^8 + 8r_5^2 |\tilde{A}_5|^2 \alpha_2^2 |\alpha_2|^6 x_{\mathrm{p},n}^{10})(\alpha_2^*)^2 \}$$
$$+ (18r_3^2 |\tilde{A}_3|^2 |\alpha_1|^4 x_{\mathrm{p},n}^6 + 34r_5^2 |\tilde{A}_5|^2 |\alpha_1|^8 x_{\mathrm{p},n}^{10} + 2\Re\{ 24r_3 r_5 \tilde{A}_3^* \tilde{A}_5 |\alpha_1|^6 x_{\mathrm{p},n}^8 \}) |\alpha_1|^2$$

$$+ (18r_3^2|\tilde{A}_3|^2|\alpha_2|^4 x_{\mathrm{p},n}^6 + 34r_5^2|\tilde{A}_5|^2|\alpha_2|^8 x_{\mathrm{p},n}^{10} + 2\Re\{24r_3r_5\tilde{A}_3^*\tilde{A}_5|\alpha_2|^6 x_{\mathrm{p},n}^8\})|\alpha_2|^2\Big\} \qquad (5.69)$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\epsilon})(\frac{\partial\mathcal{L}}{\partial\phi})\} = \frac{1}{\sigma_\eta^2}\sum_{n=1}^{N}\Big\{2\Re\{(6r_3r_5\tilde{A}_3^*\tilde{A}_5\alpha_1^2|\alpha_1|^4 x_{\mathrm{p},n}^8 + 8r_5^2|\tilde{A}_5|^2\alpha_1^2|\alpha_1|^6 x_{\mathrm{p},n}^{10})e^{j\phi}(j\alpha_1^*)\}$$

$$+ 2\Re\{(6r_3r_5\tilde{A}_3^*\tilde{A}_5\alpha_2^2|\alpha_2|^4 x_{\mathrm{p},n}^8 + 8r_5^2|\tilde{A}_5|^2\alpha_2^2|\alpha_2|^6 x_{\mathrm{p},n}^{10})e^{-j\phi}(j\alpha_2^*)\}$$

$$- \Re\{(18r_3^2|\tilde{A}_3|^2|\alpha_1|^4 x_{\mathrm{p},n}^6 + 34r_5^2|\tilde{A}_5|^2|\alpha_1|^8 x_{\mathrm{p},n}^{10} + 2\Re\{24r_3r_5\tilde{A}_3^*\tilde{A}_5|\alpha_1|^6 x_{\mathrm{p},n}^8\})e^{j\phi}(j\alpha_1)\}$$

$$- \Re\{(18r_3^2|\tilde{A}_3|^2|\alpha_2|^4 x_{\mathrm{p},n}^6 + 34r_5^2|\tilde{A}_5|^2|\alpha_2|^8 x_{\mathrm{p},n}^{10} + 2\Re\{24r_3r_5\tilde{A}_3^*\tilde{A}_5|\alpha_2|^6 x_{\mathrm{p},n}^8\})e^{j\phi}(j\alpha_2)\}\Big\}$$

$$(5.70)$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\tilde{A}_3^*})^*(\frac{\partial\mathcal{L}}{\partial\epsilon})\} = \frac{1}{\sigma_\eta^2}\sum_{n=1}^{N}\Big\{(3r_3^2\tilde{A}_3^*\alpha_1|\alpha_1|^4 x_{\mathrm{p},n}^6 + 4r_3r_5\tilde{A}_5^*\alpha_1|\alpha_1|^6 x_{\mathrm{p},n}^8)e^{j\phi} + r_3r_5\tilde{A}_5^*\alpha_1^*|\alpha_1|^6 e^{-j\phi}$$

$$- (3r_3^2\tilde{A}_3^*\alpha_2|\alpha_2|^4 x_{\mathrm{p},n}^6 + 4r_3r_5\tilde{A}_5^*\alpha_2|\alpha_2|^6 x_{\mathrm{p},n}^8)e^{-j\phi} - r_3r_5\tilde{A}_5^*\alpha_2^*|\alpha_2|^6 e^{j\phi}\Big\} \qquad (5.71)$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\tilde{A}_3^*})^*(\frac{\partial\mathcal{L}}{\partial\phi})\} = \frac{1}{\sigma_\eta^2}\sum_{n=1}^{N}\Big\{(3r_3^2\tilde{A}_3^*\alpha_1|\alpha_1|^4 x_{\mathrm{p},n}^6 + 4r_3r_5\tilde{A}_5^*\alpha_1|\alpha_1|^6 x_{\mathrm{p},n}^8)(j\alpha_1^*) - r_3r_5\tilde{A}_5^*\alpha_1^*|\alpha_1|^6(j\alpha_1)$$

$$- (3r_3^2\tilde{A}_3^*\alpha_2|\alpha_2|^4 x_{\mathrm{p},n}^6 + 4r_3r_5\tilde{A}_5^*\alpha_2|\alpha_2|^6 x_{\mathrm{p},n}^8)(j\alpha_2^*) + r_3r_5\tilde{A}_5^*\alpha_2^*|\alpha_2|^6 e^{-j\phi}(j\alpha_2)\Big\} \quad (5.72)$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\tilde{A}_5^*})^*(\frac{\partial\mathcal{L}}{\partial\epsilon})\} = \frac{1}{\sigma_\eta^2}\sum_{n=1}^{N}\Big\{(3r_3r_5\tilde{A}_3^*\alpha_1|\alpha_1|^6 x_{\mathrm{p},n}^8 + 4r_5^2\tilde{A}_5^*\alpha_1|\alpha_1|^8 x_{\mathrm{p},n}^{10})e^{j\phi} + r_5^2\tilde{A}_5^*\alpha_1^*|\alpha_1|^8 e^{-j\phi}$$

$$- (3r_3r_5\tilde{A}_3^*\alpha_2|\alpha_2|^6 x_{\mathrm{p},n}^8 + 4r_5^2\tilde{A}_5^*\alpha_2|\alpha_2|^8 x_{\mathrm{p},n}^{10})e^{-j\phi} - r_5^2\tilde{A}_5^*\alpha_2^*|\alpha_2|^8 e^{j\phi}\Big\} \qquad (5.73)$$

$$\mathbb{E}\{(\frac{\partial\mathcal{L}}{\partial\tilde{A}_5^*})^*(\frac{\partial\mathcal{L}}{\partial\phi})\} = \frac{1}{\sigma_\eta^2}\sum_{n=1}^{N}\Big\{(3r_3r_5\tilde{A}_3^*\alpha_1|\alpha_1|^6 x_{\mathrm{p},n}^8 + 4r_5^2\tilde{A}_5^*\alpha_1|\alpha_1|^8 x_{\mathrm{p},n}^{10})(j\alpha_1^*) - r_5^2\tilde{A}_5^*\alpha_1^*|\alpha_1|^8(j\alpha_1)$$

$$- (3r_3r_5\tilde{A}_3^*\alpha_2|\alpha_2|^6 x_{\mathrm{p},n}^8 + 4r_5^2\tilde{A}_5^*\alpha_2|\alpha_2|^8 x_{\mathrm{p},n}^{10})(j\alpha_2^*) + r_5^2\tilde{A}_5^*\alpha_2^*|\alpha_2|^8(j\alpha_2)\Big\} \qquad (5.74)$$

PART III

SPECTRUM SHARING BETWEEN WIRELESS COMMUNICATIONS AND

RADIO ASTRONOMY SYSTEMS

CHAPTER 6

A NEW PARADIGM FOR SPECTRUM SHARING BETWEEN CELLULAR WIRELESS COMMUNICATIONS AND RADIO ASTRONOMY SYSTEMS [1]

6.1 Introduction

Radio astronomy systems (RASs) are allocated with some dedicated primary bands and some secondary bands, but their spectral requirements continue to increase and some observations in the bands allocated to the active wireless services are essential for RAS [91, 44]. In the past, sporadic spectrum use of active services allows RAS to observe in the active service bands during the unused time intervals. But such opportunities are diminishing quickly as new wireless systems with dynamic spectrum access proliferate. Furthermore, RFI detection (important task for RAS [60, 86, 51, 95, 24, 35, 20, 39]) will become more challenging as commonly used non-Gaussianity test would become invalid for such emerging scenarios. These imply smaller spectrum access opportunities and corrupted or less reliable scientific observations for RAS, thus hindering several benefits and scientific advancements that RAS can offer.

To protect from RFI, RAS receivers are typically set up at remote locations and a radio quiet zone centered around each RAS receiver (additionally surrounded by a radio coordination zone) is imposed by regulation. In view of population growths, ever-expanding active wireless services, and increased interests in RAS observations and in developing new RAS sites, the existing geographical isolation approach of RAS receivers in remote locations together with radio protection zones [90, 131] would not be able to answer the future needs of the society. New spectrum sharing strategies are in great need, which we address in this chapter.

Our contributions are summarized below. We develop a new paradigm of spectrum sharing between CWC and RAS based on creating a shared spectrum access zone (SSAZ) and three-phase spectrum access which enables geographical and spectral coexistence between CWC and RAS. Solutions for different cases of single and multiple RAS receivers as well as single and multiple shared bands are presented. We also develop advanced resource allocation strategies by exploiting CWC's traffic statistics. Furthermore, built-in fine tuning feature is also described to address design mismatches and service evolutions of CWC and RAS. Simulation study illustrates that the proposed paradigm offers i) certain guaranteed spectrum access to RAS which is impossible in the existing paradigm, ii) capability to handle higher peak and mean traffics to CWC under spectrum restructuring of both CWC and RAS bands, and iii) overall improved spectrum utilization.

The rest of the chapter is organized as follows. Section 6.2 discusses the perspectives of the proposed work with respective to the existing approaches. Section 6.3 introduces key approaches of the proposed paradigm, while their detailed designs are described in Sections 6.4 and 6.5. Section 6.6 presents performance evaluation results, and Section 6.7 concludes the chapter.

6.2 Related Works

There exist several works on spectrum sharing for cognitive radio (CR) or dynamic spectrum access (DSA), e.g., [3, 150, 84, 45, 147, 53, 104, 26, 121, 140, 41]. Although the spectrum sharing theme is common, system characteristics, requirements, and specific technical solutions are different between the considered spectrum sharing paradigm and the existing CR/DSA. Here we briefly describe those differences which distinguish the proposed approach from the existing ones.

First of all, the systems involved in CR/DSA are active systems while the proposed approach is between active and passive systems. With active systems, one can sense the

signal strength of the other and can make necessary adjustment which is not the case when passive systems are involved. Second, RAS single-dish (interferometric arrays) receivers could be 50-60 dB (10-20 dB) more sensitive than receivers in active systems [39, 56]. This means negligible interference for active systems could cause disruptive interference for passive systems. For example, [17, 18] reported that underlay spectrum sharing between CWC and RAS is not applicable. The spectrum sensing methods considered in CR/DAS [3, 150, 84, 45, 147] are based on the characteristics of the active systems and they do not have sufficiently high sensitivity to detect very low level interferences which are still detrimental to RAS.

Third, the scales of signal receiving time interval (frame length for CR/DSA and measurement duration for RAS) are quite different. For CR/DSA, the time scale is on the order of millisecond to seconds. But for RAS, a measurement could require several hours to a few months. The time windows to watch out for interference are quite different between CR/DSA and RAS. Erroneous signal detection could be easily overcome in CR/DSA by error control techniques within a short time interval but it could be a significant loss of time and resources and potentially a complete loss of scientific information for some time-critical observations for RAS.

Fourth, opportunistic spectrum access as used in CR/DSA is not suitable for the considered scenario. Imposing RAS to apply secondary opportunistic spectrum access will not work well due to several reasons- i) unreliable RFI detection of RAS and hence unreliable scientific data, ii) no guarantee of required spectrum access time for RAS thus ruling out some time-critical observations or time-scheduled measurements, iii) unfair spectrum allocation between two different types of systems. On the other hand, CWC cannot play the part of secondary opportunistic users as in CR/DSA since CWC cannot sense when RAS is performing its measurement. Furthermore, due to different time scales, latency requirement of CWC would not be satisfied. The above points clearly illustrate that the technical solutions developed for CR/DSA would not be directly applicable for the considered spectrum sharing scenario.

As for spectrum sharing between active and passive systems, [139, 123] considered spectrum sharing between CWC and fixed satellite service earth station (FSS-ES). CWC can work only outside a protection region. The idea is similar to the radio quiet zone. In [63], a transmit-nulling space division multiple access spectrum sharing is considered to effectively reduce the protection region size. Since the interference threshold of FSS-ES is quite larger than that of RAS, the protection region is quite smaller for FSS-ES. Furthermore, these works assume no spectrum sharing inside the protection region. In [13], an interweave spectrum sharing approach is developed to protect RAS from satellite interferences. It considers a radio quite zone around RAS and lacks in sharing the spectrum with CWC.

6.3 Proposed Shared Spectrum Access

The proposed shared spectrum access is based on the new paradigm of geographical coexistence between CWC and RAS systems with some form of guaranteed RFI-free spectrum access for RAS systems. We propose the following novel approaches to enable such coexistence paradigm. Their detailed design aspects will be described in the next section.

1. Three-Phase Time Division Spectrum Access: We develop a shared spectrum access strategy based on the time division approach with three phases. Across time, each frame contains three phases, namely, CWC only phase, CWC+RAS phase, and RAS only phase. The three phases are repeated but the durations of the three phases can be time-dependent and different from repetition to repetition. They are predefined well in advance and known to both CWC and RAS.

During the CWC only phase, CWC transmits but RAS does not collect its data as it is heavily corrupted by RFI from CWC. During the CWC+RAS phase, each CWC base station (BS) or its users can transmit during a phase-2 transmit interval (P2TI) which is BS-dependent and predefined. Fine tuning of the shared spectrum access can also be done during this phase. During the RAS only phase, there is no CWC transmission within the

SSAZ, thus RAS is free from CWC-induced RFI. RAS conducts reliable data collection during this phase.

2. Three-Phase Time-Frequency Division Spectrum Access: When multiple bands are shared, we devise a spectrum access strategy which follows the same three phase principle in each band but positions the frames of different bands in a way to minimize the physical layer latency of CWC. The design would also depend on the frequency spacing of the shared bands and the CWC's out-of-band spectrum characteristics.

3. Adaptive Shared Spectrum Access: As CWC traffic loads vary across time, we also propose to enhance the above three phase spectrum access strategies by using different modes of resource allocation between CWC and RAS at different hours of each day. Such adaptation modes and schedules are predefined well in advance based on the average traffic statistics across time and/or the planned off-times of RAS. They can be adjusted over a larger time scale in response to service evolution of CWC and RAS under fair resource sharing between CWC and RAS.

4. Built-in Fine Tuning for the Shared Spectrum Access: We propose that each RAS site has additional processing units to perform fine tuning for the shared spectrum access. During the CWC+RAS phase, RAS receiver does not collect scientific data and RAS can do RFI measurement, fine timing synchronization with CWC BSs, and fine tuning of SSAZ. This provides a built-in fine tuning mechanism for addressing design mismatches of the shared spectrum access or for accommodating service evolutions of CWC and RAS. Details will be discussed in Section 6.5.

6.4 Design Steps

6.4.1 Shared Spectrum Access Zone (SSAZ)

SSAZ with 1 RAS receiver: An important step in developing the proposed shared spectrum access is to define a SSAZ within which CWC and RAS follow the three phase spectrum

access while CWC systems outside SSAZ can access their spectrum freely but their transmissions do not cause harmful RFI to RAS. For this, we adopt the interference power threshold as defined in ITU-R RA.769-2 as the maximum RFI level the RAS receiver can tolerate and denote it by I_{th}. We consider downlink of CWC with full load to determine the RFI level at RAS. This corresponds to a worst case RFI scenario (safer design for RFI protection) for RAS. We consider a path loss model $L(d, f_{\text{c}})$ at distance r and CWC carrier frequency f_{c} as

$$L(r, f_{\text{c}}) = \sigma f_{\text{c}}^2 r^{-n} 10^{-\beta r/10}, \tag{6.1}$$

where $\sigma = \left(\frac{4\pi}{3 \times 10^8}\right)^2$, n is the path loss exponent[2], and β is the atmospheric absorption loss in dB scale at f_{c}. Suppose the SSAZ is with radius R_{SSAZ} meters centered at the RAS receiver. We consider two models for the distribution of CWC cells surrounding the RAS receiver.

Hexagon CWC Cells

We consider a deterministic model, where hexagon CWC cells, each with radius R_{CWC} meters, surround the RAS receiver as shown in Fig. 6.1. The ith tier cell ring has K_i hexagon cells with the same distance $r_i = \sqrt{3}\sqrt{a^2 + a(b+1) + b^2 + \frac{1}{3}} R_{\text{CWC}}$ between BSs of the ith tier and RAS receiver (this result is obtained after applying translation of axis to the result obtained in [47, Eq. 15.2]), where K_i is the number of all possible ways to get the same distance r_i by choosing any integer values for a and b. Then, the SSAZ design based on S outer tiers can be given by the radius R_{SSAZ} of the SSAZ as

$$R_{\text{SSAZ}} = \min_{i_0} r_{i_0}, \quad \text{s.t. } P_{\text{BS}} G_{\text{CWC}} G_{\text{RAS}} \sum_{i=i_0+1}^{i_0+S} K_i L(r_i) < I_{\text{th}}, \tag{6.2}$$

---

[2]We consider a simple distance-dependent path loss exponent defined as $n = 2$ if $d \leq d_{\text{LOS}}$ and $n = n_{\text{NLOS}}$ if $d > d_{\text{LOS}}$, where $n = 2$ and $n = n_{\text{NLOS}}$ are the line-of-sight (LOS) and non-line-of-sight (NLOS) path loss exponents respectively. For a safer design, we use the horizon distance d_{Horizon} in km as our d_{LOS}. We obtain d_{Horizon} as $d_{\text{Horizon}} < 3.57(\sqrt{h_{\text{RAS}}} + \sqrt{h_{\text{CWC}}})$, where h_{RAS} and h_{CWC} are the heights of the RAS receiver and CWC base station in meters respectively.
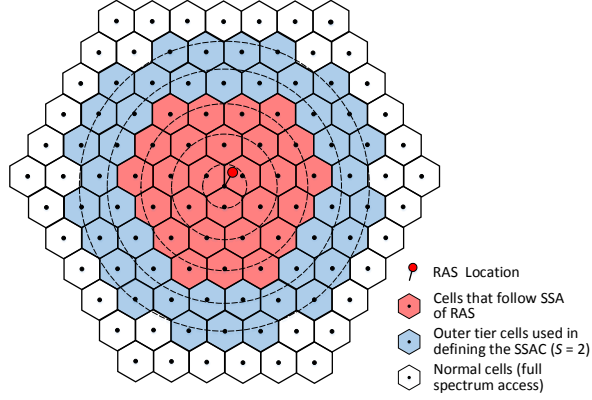
Figure 6.1. Tiers of CWC cell rings around an RAS receiver

where P_{BS} is the transmit power of CWC BSs, G_{CWC} is the transmit antenna gain of CWC BSs, and G_{RAS} is the antenna gain of RAS receiver for CWC interferences. Note that due to terrain constraints, associated path-loss and antenna down-tilting, using a finite S is justified. Fig. 6.1 illustrates the tiers of CWC hexagon cell rings around an RAS receiver, as well as the outer tiers ($S = 2$ for presentation convenience; those in the blue color) used in defining the SSAZ. In this illustration, the SSAZ is composed of 3 tiers (for presentation convenience) of CWC cells in the red color and these cells will follow the shared spectrum access approach. The CWC cells outside the SSAZ, i.e., those in the blue and white color, have full spectrum access and do not follow the shared spectrum access approach.

Poisson Point Process (PPP) Model

We also consider a stochastic geometry model, where CWC cells are distrusted as a homogeneous Poisson point process (PPP) with intensity ρ_{BS}. Then, the SSAZ design based on the mean interference and outer distance R_{out} can be given by the radius R_{SSAZ} of the SSAZ as

$$R_{\mathrm{SSAZ}} = \min_{r_0} r_0, \quad \mathrm{s.t.} \ 2\pi P_{\mathrm{BS}}\rho_{\mathrm{BS}}G_{\mathrm{CWC}}G_{\mathrm{RAS}} \int_{r_0}^{R_{\mathrm{out}}} rL(r)\,\mathrm{d}r < I_{\mathrm{th}}. \tag{6.3}$$

Let $\tilde{\beta} = \frac{\beta \ln 10}{10}$. Using the path loss model in (6.1), we have

$$\int_{r_0}^{R_{\mathrm{out}}} rL(r)\,\mathrm{d}r = \sigma f_{\mathrm{c}}^2 \tilde{\beta}^{n-2}\left(\Gamma(2-n, \tilde{\beta}r_0) - \Gamma(2-n, \tilde{\beta}R_{\mathrm{out}})\right) \tag{6.4}$$

122

where $\Gamma(m, x_0) = \int_{x_0}^{\infty} x^{m-1} \exp(-x)\, dx$ is the incomplete gamma function. Therefore, R_{SSAZ} can be obtained by solving the non-linear equation $\Gamma(2 - n, \tilde{\beta} R_{\mathrm{SSAZ}}) = \Gamma(2 - n, \tilde{\beta} R_{\mathrm{out}}) + I_{\mathrm{th}}/(2\pi P_{\mathrm{BS}} \rho_{\mathrm{BS}}\ G_{\mathrm{CWC}} G_{\mathrm{RAS}} \sigma f_{\mathrm{c}}^2 \tilde{\beta}^{n-2})$. It is worth mentioning that by setting $R_{\mathrm{out}} = r_{i_0 + S}$ and $\rho_{\mathrm{BS}} = \left(3\sqrt{3} R_{\mathrm{CWC}}^2/2\right)^{-1}$, R_{SSAZ} obtained assuming the PPP model is approximately the same as R_{SSAZ} obtained assuming hexagon CWC cells.

SSAZ with M RAS receivers: When there are M RAS receivers at different locations within a region of a potential SSAZ, the above design can be modified as follows. First, for each RAS receiver, we apply the above design as if there were only one RAS receiver. For RAS receiver m, we obtain the zone $\widetilde{\mathrm{SSAZ}}_m$ within which there are K_{in}^m BSs with the corresponding BS-to-RAS distances of $\{d_{i,m}\}$. Then, the final SSAZ(s) is (are) given by the union of all M zones as $\{\mathrm{SSAZ}_n\} = \cup_{m=1}^{M} \widetilde{\mathrm{SSAZ}}_m$. An illustration of a scenario with 3 RAS receivers is shown in Fig. 6.2 where $\widetilde{\mathrm{SSAZ}}_1$ and $\widetilde{\mathrm{SSAZ}}_2$ are combined into SSAZ_1 while $\widetilde{\mathrm{SSAZ}}_3$ stands as SSAZ_2. In other words, there are two RAS receivers with different locations within SSAZ_1 and one RAS receiver within SSAZ_2. Note that there are three groups of cells denoted by different colors within SSAZ_1 due to how their P2TIs are computed (described in the next section).

6.4.2 Location-Dependent Three-Phase Frame Structure

After determining the SSAZ, the next step is to design the frame structure for the shared spectrum access. Let $T_{\mathrm{CWC},k}$, $T_{\mathrm{CWC+RAS}}$ and $T_{\mathrm{RAS},k}$ denote the durations of the three phases in each frame at hour k, and $T_{\mathrm{f}} = T_{\mathrm{CWC},k} + T_{\mathrm{CWC+RAS}} + T_{\mathrm{RAS},k}$ denotes the frame duration. We address $T_{\mathrm{CWC+RAS}}$ first. This phase absorbs propagation delays of different CWC BSs/users within the SSAZ so that those CWC signals do not arrive at the RAS receiver during the RAS only phase. The maximum propagation delay difference of CWC BSs/users is $R/(3 \times 10^8)$ (propagation time from the edge of SSAZ to the RAS receiver). This imposes a lower limit on $T_{\mathrm{CWC+RAS}}$ and we can set $T_{\mathrm{CWC+RAS}} > \kappa R/(3 \times 10^8)$ where we use $\kappa > 1$, e.g., $\kappa = 2$,
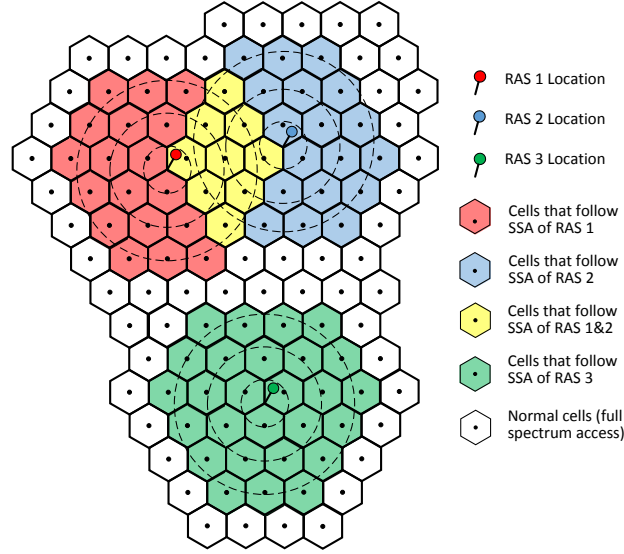
Figure 6.2. A scenario with 3 RAS sites which results in a SSAZ with 2 RAS receivers and another SSAZ with one RAS receiver

to accommodate propagation model mismatches and RFI measurement/testing during the CWC+RAS phase.

To limit overhead due to $T_{\text{CWC+RAS}}$, T_{f} should be substantially larger than $T_{\text{CWC+RAS}}$. On the other hand, $T_{\text{CWC+RAS}} + T_{\text{RAS},k}$ causes a physical layer off duration for CWC which affects CWC's latency performance. To limit the latency, we can introduce a constraint of maximum physical layer off duration $\tau_{\text{off,max}}$ such that $T_{\text{CWC+RAS}} + T_{\text{RAS},k} < \tau_{\text{off,max}}$. Such constraint can be relaxed if there are other CWC-only resources or shared resources to support the required latency performance. The resource allocation between $T_{\text{CWC},k}$ and $T_{\text{RAS},k}$ could be set based on an agreed spectrum sharing policy and an advanced approach is described in Section 6.4.4.

We allow CWC BSs/users to transmit during their P2TIs provided their transmissions do not cause RFI during the RAS only phase. Let $J_{\widetilde{\text{SSAZ}},i}$ denote the index set of $\widetilde{\text{SSAZ}}$s such that BS i is located within each of these SSAZs and $d_{i,m}$ represent the distance of BS i to $\widetilde{\text{SSAZ}}_m$, $m \in J_{\widetilde{\text{SSAZ}},i}$. We will use orthogonal frequency division multiplexing (OFDM) since with a proper precoding it can also represent a single-carrier system. Let T_{sym} represent the OFDM symbol duration including the cyclic prefix interval. Then, we can design P2TI
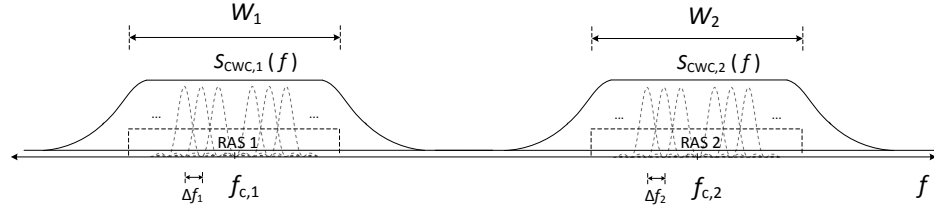
Figure 6.3. Power spectral density of CWC signals appearing at the two observation bands of RAS

during the CWC+RAS phase for BS i, denoted by $T_{\text{ext},i}$ as

$$T_{\text{ext},i} = \min_{m \in J_{\widehat{\text{SSAZ}},i}} \left\lfloor \frac{T_{\text{CWC+RAS}} - d_{i,m}/(3 \times 10^8)}{T_{\text{sym}}} \right\rfloor T_{\text{sym}} \tag{6.5}$$

which represents an extension of the CWC transmission duration for BS i. The minimization in the above design avoids causing RFI to RAS when a BS is located within an overlapped region of two or more individual SSAZs. For example, within SSAZ$_1$ in Fig. 6.2, P2TIs are computed based on RAS site 1 for the BSs in the red group, on RAS site 2 for those in the blue group; but P2TI of a BS in the yellow group is selected as the minimum of the two extension durations due to RAS site 1 and site 2. By this, the yellow group avoids RFI to both RAS sites. Note that different BSs in the same SSAZ have the same durations of the three phases but their P2TIs can be different depending on their locations.

6.4.3 Design for Multiple Shared Bands

The design presented in the previous sections can be extended for multiple shared bands. As an example, we consider a scenario with two shared bands. For simplicity of exposure, we assume in each band that the CWC carrier frequency and the RAS observation band center frequency are the same. Furthermore, the bandwidths of CWC and RAS are the same in each band but are different from band to band. Fig. 6.3 shows the power spectral density (PSD) of CWC signals appearing at the two observation bands of RAS, where $f_{\text{c},i}$ and W_i are the center frequency and the bandwidth of the ith band of CWC/RAS. We need to determine the SSAZ such that RFI to the RAS's data collection in any band is less than the

acceptable interference threshold. The OFDM system at the uth band has an avergae BS transmit power of $P_{\mathrm{BS},u}$, a subcarrier spacing of Δf_u, an OFDM symbol duration of $1/\Delta f'_u$ (including cyclic prefix; thus $\Delta f'_u < \Delta f_u$), $N_{\mathrm{c},u}$ used subcarriers with their subcarrier index set $\mathcal{I}_u$, and the carrier frequency $f_{\mathrm{c},u}$. Then the PSD of the CWC signal at the uth band, denoted by $S_{\mathrm{CWC},u}(f)$, is given by

$$S_{\mathrm{CWC},u}(f) = \sum_{i \in \mathcal{I}_u} \frac{P_{\mathrm{BS},u}}{N_{\mathrm{c},u}\Delta f'_u} \mathrm{sinc}^2\left(\frac{f - f_{\mathrm{c},u} - i\Delta f_u}{\Delta f'_u}\right), \tag{6.6}$$

where $\mathrm{sinc}(x) \triangleq \sin(\pi x)/(\pi x)$. The normalized interfering CWC signal power centered at $f_{\mathrm{c},u}$ with bandwidth of W_u within the RAS observation band centered at $f_{\mathrm{c},v}$ with bandwidth of W_v, denoted by $L_{\mathrm{s}}(f_{\mathrm{c},u}, W_u, f_{\mathrm{c},v}, W_v)$, is calculated as

$$L_{\mathrm{s}}(f_{\mathrm{c},u}, W_u, f_{\mathrm{c},v}, W_v) = \int_{f_{\mathrm{c},v} - \frac{W_v}{2}}^{f_{\mathrm{c},v} + \frac{W_v}{2}} \frac{S_{\mathrm{CWC},u}(f)}{P_{\mathrm{BS},u}} df < \sum_{i \in \mathcal{I}_u} \int_{f_{\mathrm{c},v} - \frac{W_v}{2}}^{f_{\mathrm{c},v} + \frac{W_v}{2}} \frac{\Delta f'_u/(N_{\mathrm{c},u}\pi^2)}{(f - f_{\mathrm{c},u} - i\Delta f)^2} df \tag{6.7}$$

$$= \sum_{i \in \mathcal{I}_u} \frac{\Delta f'_u\, W_v/(N_{\mathrm{u}}\pi^2)}{\left(f_{\mathrm{c},u} + i\Delta f - f_{\mathrm{c},v} + \frac{W_v}{2}\right)\left(f_{\mathrm{c},u} + i\Delta f - f_{\mathrm{c},v} - \frac{W_v}{2}\right)}.$$

Therefore, the SSAZ is defined by the radius R_{SSAZ} given by

$$R_{\mathrm{SSAZ}} = \max\left(R_{\mathrm{SSAZ},1}, R_{\mathrm{SSAZ},2}\right), \tag{6.8}$$

$$R_{\mathrm{SSAZ},1} = \min_{i_0} r_{i_0}$$

$$\text{s.t. } P_{\mathrm{BS},1} \sum_{i=i_0+1}^{i_0+S} K_i L(r_i, f_{\mathrm{c},1}) + P_{\mathrm{BS},2} \sum_{i=1}^{i_0+S} K_i L(r_i, f_{\mathrm{c},1}) L_{\mathrm{s}}(f_{\mathrm{c},2}, W_2, f_{\mathrm{c},1}, W_1) < \frac{I_{\mathrm{th},1}}{G_{\mathrm{CWC}}G_{\mathrm{RAS}}},$$

$$R_{\mathrm{SSAZ},2} = \min_{i_0} r_{i_0}$$

$$\text{s.t. } P_{\mathrm{BS},2} \sum_{i=i_0+1}^{i_0+S} K_i L(r_i, f_{\mathrm{c},2}) + P_{\mathrm{BS},1} \sum_{i=1}^{i_0+S} K_i L(r_i, f_{\mathrm{c},2}) L_{\mathrm{s}}(f_{\mathrm{c},1}, W_1, f_{\mathrm{c},2}, W_2) < \frac{I_{\mathrm{th},2}}{G_{\mathrm{CWC}}G_{\mathrm{RAS}}},$$

where $I_{\mathrm{th},1}$ and $I_{\mathrm{th},2}$ are the interference thresholds of the two bands. Similarly, the PPP model described in subsection 6.4.1 can be used to define the SSAZ.

Now, we introduce a frame time offset between the frames of the two bands such that the physical layer off duration (hence latency) of CWC is minimized. This results in a
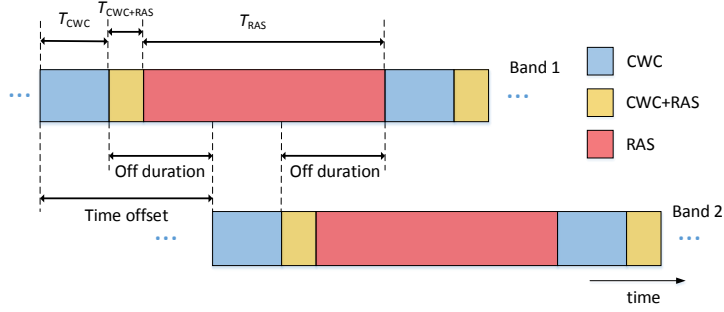
Figure 6.4. Time-frequency division spectrum access for two shared bands

time-frequency division spectrum access for CWC and RAS. Fig. 6.4 illustrates the frame structures for the two shared bands. In this case, with a time offset of $T_\mathrm{f}/2$, the physical layer off duration is reduced from $T_\mathrm{RAS} + T_\mathrm{CWC+RAS}$ to $\max([T_\mathrm{RAS} + T_\mathrm{CWC+RAS} - T_\mathrm{CWC}]/2, 0)$.

6.4.4   Resource Adaptation based on CWC Traffic Statistics

CWC wireless traffics show specific temporal usage characteristics and an example of such statistics within a day is shown in Fig. 6.5 [25]. Utilizing such traffic statistics, we propose to enhance the resource allocation through adaptation across time. We consider two resource allocation problems under the constraint that CWC and RAS must have predefined minimum resource amounts in each frame and the predefined ratio of the total resource amounts per day between CWC and RAS is approximately maintained. In the first problem, we optimize the resource allocation across time to maximize CWC traffic support. In the second, we maximize the allocated resources to RAS across time during each day while limiting the CWC blockage probability per hour.

Single Shared Band

The frame structure at hour k can be defined by $T_\mathrm{CWC,k}$, $T_\mathrm{CWC+RAS}$, and $T_\mathrm{RAS,k}$. Suppose one subframe duration T_sf is the minimum time unit that can be allocated among the three phases, and there are C resource blocks within T_sf. One frame duration T_f has n_f subframes. The numbers of subframes within $T_\mathrm{CWC,k}$, $T_\mathrm{CWC+RAS}$, and $T_\mathrm{RAS,k}$ are $n_\mathrm{CWC,k}$, $n_\mathrm{CWC+RAS}$, and
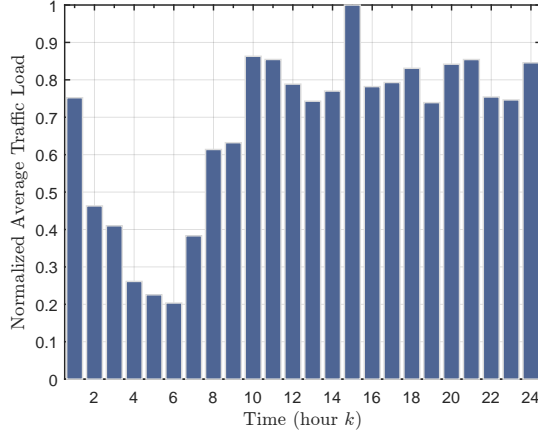
127

Figure 6.5. An example of average (normalized) traffic load within a day

$n_{\mathrm{RAS},k}$ respectively, where $n_{\mathrm{CWC},k} + n_{\mathrm{CWC+RAS}} + n_{\mathrm{RAS},k} = n_{\mathrm{f}}$ is the same across all hours. Note that the actual traffic load fluctuates and hence it can sometimes be higher than the average traffic load during the considered hour. Suppose the CWC system is designed to support a maximum of $n_{\mathrm{max,CWC}} C$ resource blocks per frame which is α times the maximum average traffic load, e.g., $\alpha = 1.2$. First, the average traffic loads per frame across 24 hours are scaled in units of resource blocks, e.g., λ_k arrivals per frame, each requiring a resource block, during hour k. Then, with $\lambda_{\max} = \max_k(\lambda_k)$, we have $n_{\mathrm{max,CWC}} C = \alpha \lambda_{\max}$. For illustration, the actual traffic per frame during hour k is modeled as a Poisson process with mean λ_k. Let $F_{k,n} = \lambda_k^n \exp(-\lambda_k)/n!$ denote the probability that n resource blocks/frame are requested during hour k. Then, the CWC throughput per hour for hour k, denoted by $\rho_k(n_{\mathrm{CWC},k})$ or simply ρ_k, is defined as

$$\rho_k = \sum_{n=0}^{\infty} F_{k,n} \min\left(n, n_{\mathrm{CWC},k} C\right). \tag{6.9}$$

Suppose CWC and RAS require a minimum of $n_{\mathrm{min,CWC}}$ and $n_{\mathrm{min,RAS}}$ subframes respectively in each frame, and the ratio of resource amounts per day between CWC and RAS during the CWC only and RAS only phases is γ. Then our first resource allocation problem becomes designing the time-dependent frame structure to maximize the average CWC

throughput as

$$P1 : \underset{\{n_{\mathrm{CWC},k}\}}{\arg\max} \sum_{k=1}^{24} \rho_k(n_{\mathrm{CWC},k}), \tag{6.10}$$

$$\text{s.t. } C1 : \sum_{k=1}^{24} n_{\mathrm{CWC}k} \leq N_{\mathrm{CWC}},$$

$$C2 : n_{\mathrm{min,CWC}} \leq n_{\mathrm{CWC},k} \leq n_{\mathrm{max,CWC}} \quad \forall k.$$

The constraint C1 is to satisfy the ratio γ of resource amount per day between CWC and RAS, where N_{CWC} is the maximum total number of subframes that can be allocated to CWC during a day which is a function of γ, and $N_{\mathrm{CWC}} = 24\,(n_{\mathrm{f}} - n_{\mathrm{CWC+RAS}})\,\gamma/\,(1+\gamma)$. The constraint C2 is to make sure that CWC and RAS have at least a minimum number of subframes per each hour, and $n_{\mathrm{max,CWC}} = n_{\mathrm{f}} - n_{\mathrm{CWC+RAS}} - n_{\mathrm{min,RAS}}$. Note that γ has to be less than or equal γ_{max}, where $\gamma_{\mathrm{max}} = n_{\mathrm{max,CWC}}/n_{\mathrm{min,RAS}}$ is the maximum achievable ratio of resource amounts, otherwise the problem P1 will be infeasible.

The optimization P1 is a mixed integer convex programming (MICP) problem, and it has two difficulties. First, $n_{\mathrm{CWC},k}$ has a non-convex integer constraint. Second, although ρ_k in (6.9) is a concave function, it does not have a closed-form expression, and hence we cannot use the existing MICP solvers (e.g. MOSEK [87]). As a solution to the second difficulty, we truncate the infinite sum in (6.9) to a finite number of terms, and then we solve the optimization problem with the truncated expression by any existing MICP solvers. We keep increasing the number of terms and solving the optimization problem until the resulting throughput effectively does not change. However, the number of needed terms increases as λ_k increases. Alternatively, we can avoid the second difficulty by applying an accurate approximation to (6.9) as follows. Using the Gaussian approximation of Poisson distribution, which is accurate for $\lambda_k \geq 10$ [120], we approximate ρ_k by $\tilde{\rho}_k$ expressed as (see Appendix A)

$$\tilde{\rho}_k = n_{\text{CWC},k}\, C - (n_{\text{CWC},k}\, C - \lambda_k)\, \Phi\left(\frac{n_{\text{CWC},k}\, C - \lambda_k}{\sqrt{\lambda_k}}\right) - \sqrt{\frac{\lambda_k}{2\pi}}\, \exp(-\frac{(n_{\text{CWC},k}\, C - \lambda_k)^2}{2\lambda_k}),$$

(6.11)

where $\Phi\left(x\right)$ is the cumulative distribution function (CDF) of the standard normal distribution.[3] It can be shown (see Appendix B) that

$$\frac{\partial \tilde{\rho}_k}{\partial n_{\text{CWC},k}} = C\left(1 - \Phi\left(\frac{n_{\text{CWC},k}\, C - \lambda_k}{\sqrt{\lambda_k}}\right)\right),$$

(6.12)

which is a monotonically decreasing function. Therefore, $\tilde{\rho}_k$ is a concave function. Now, we can use any MICP solver for P1 but with $\tilde{\rho}_k$. We also apply another method which solves an integer-relaxed problem and then applies an appropriate rounding [21]. Both methods converge to the same solution. We adopt the second method (which has lower computational complexity) to solve the problem P1. Let $\tilde{\text{P}}1$ be the integer-relaxed problem of P1 but with $\tilde{\rho}_k$. Since $\tilde{\rho}_k$ is a concave function and the constraints C1 and C2 are convex, the problem $\tilde{\text{P}}1$ is a convex optimization problem. Therefore, solving the dual problem of $\tilde{\text{P}}1$, denoted by D-$\tilde{\text{P}}1$, gives the same solution of the problem $\tilde{\text{P}}1$. Define the Lagrangian to be

$$\mathcal{L}(\{n_{\text{CWC},k}\}, \mu) = \sum_{k=1}^{24} \tilde{\rho}_k - \mu\left(\sum_{k=1}^{24} n_{\text{CWC},k} - N_{\text{CWC}}\right),$$

(6.13)

where μ is the Lagrange multiplier associated with the constraint C1. The problem D-$\tilde{\text{P}}1$ is given by

$$\text{D-}\tilde{\text{P}}1 : \underset{\{n_{\text{CWC},k}\}, \mu}{\arg\max}\ \mathcal{L}(\{n_{\text{CWC},k}\}, \mu),$$

(6.14)

$$\text{s.t. C2}: n_{\min,\text{CWC}} \le n_{\text{CWC},k} \le n_{\max,\text{CWC}} \quad \forall k.$$

We propose a successive resource allocation method solving the problem D-$\tilde{\text{P}}1$. Dropping C2, the Karush-Kuhn-Tucker (KKT) conditions are

$$\frac{\partial \mathcal{L}}{\partial n_{\text{CWC},k}} = C\left(1 - \Phi\left(\frac{n_{\text{CWC},k}\, C - \lambda_k}{\sqrt{\lambda_k}}\right)\right) - \mu = 0,$$

(6.15)

[3]For practical system parameters, λ_k's are quite greater than 10 which yield an accurate approximation for ρ_k. We observe in our evaluation that solving P1 with $\tilde{\rho}_k$ gives the same optimal solution obtained by solving P1 with ρ_k.

$$\mu \left(\sum_{k=1}^{24} n_{\text{CWC},k} - N_{\text{CWC}} \right) = 0, \tag{6.16}$$

where $\mu \geq 0$. From (6.15), we get

$$\frac{n_{\text{CWC},k}\, C - \lambda_k}{\sqrt{\lambda_k}} = \frac{n_{\text{CWC},l}\, C - \lambda_l}{\sqrt{\lambda_l}}, \tag{6.17}$$

for $k, l \in \{1, 2, \ldots, 24\}$. Using (6.16) and (6.17), we can solve the resource allocation problem D-$\tilde{\text{P}}$1 in a successive manner to satisfy C2 as illustrated in Algorithm 1 which is described in the next page. Let $f_{\text{asc},\lambda}(\cdot)$ represent a function which orders $\{\lambda_k\}$ in an ascending order and $f_{\text{asc},\lambda}^{-1}(\cdot)$ be the corresponding inverse ordering function. Then, for $\{\tilde{\lambda}_k\} = f_{\text{asc},\lambda}(\{\lambda_k\})$, we have $\{\lambda_k\} = f_{\text{asc},\lambda}^{-1}(\{\tilde{\lambda}_k\})$. Generally, elements of $\{\tilde{n}_{\text{CWC},k}\}_{k=J}^{M}$ obtained by Algorithm 1 are not integers. Thus, we proceed as follows. First, the quantity $\Delta = N_{\text{CWC}} - \sum_{k=1}^{24} \lfloor \tilde{n}_{\text{CWC},k} \rfloor$ and the set $\mathcal{D} = \{ \lceil \tilde{n}_{\text{CWC},k} \rceil - \tilde{n}_{\text{CWC},k} \}_{k=J}^{M}$ are calculated. Next, we set $\tilde{n}_{\text{CWC},k} = \lfloor \tilde{n}_{\text{CWC},k} \rfloor$ for $k = J, J+1, \ldots, M$. Then, we distribute Δ over as many elements of $\{\tilde{n}_{\text{CWC},k}\}_{k=J}^{M}$ as possible by 1-unit increase according to the ascending order of $\mathcal{D}$. Next, we obtain $\{n_{\text{CWC},k}\} = f_{\text{asc},\lambda}^{-1}(\{\tilde{n}_{\text{CWC},k}\})$ and then $n_{\text{RAS},k} = n_{\text{f}} - n_{\text{CWC+RAS}} - n_{\text{CWC},k}\ \forall k \in \{1, 2, \ldots, 24\}$.

The second resource allocation problem is to maximize the allocated resources to RAS provided that constraints on the CWC blockage probabilities per hour denoted by $\{B_k\}$ are satisfied. For hour k, B_k, defined as the ratio of the average blocked traffic requests to the average total traffic requests, is given by (see Appendix C)

$$\begin{aligned}
B_k &= \frac{1}{\lambda_k} \sum_{n=n_{\text{CWC},k}\, C+1}^{\infty} (n - n_{\text{CWC},k}\, C)\, F_{k,n} \\
&= 1 - Q\left(n_{\text{CWC},k}\, C - 1, \lambda_k \right) - \frac{n_{\text{CWC},k}\, C}{\lambda_k} \left(1 - Q\left(n_{\text{CWC},k}\, C, \lambda_k \right) \right), \tag{6.18}
\end{aligned}$$

where $Q(x, \lambda) = \sum_{n=0}^{x} \frac{\exp(-\lambda)\lambda^x}{x!}$ is the regularized incomplete gamma function.

The resource allocation problem P2 reads as

$$\text{P2} : \underset{n_{\text{CWC},k}}{\arg\min}\ n_{\text{CWC},k}, \tag{6.19}$$

131

Algorithm 1 Resource Allocation P1

Input: $\{\lambda_k\}$, C, $n_{\text{min,CWC}}$, $n_{\text{max,CWC}}$, N_{CWC}

Initial: $J = 1$, $M = 24$, flag $= 1$

1: **Obtain** $\{\tilde{\lambda}_k\}$ and $f_{\text{asc},\lambda}^{-1}(\cdot)$ s.t. $\{\tilde{\lambda}_k\} = f_{\text{asc},\lambda}(\{\lambda_k\})$

2: **while** flag **do**

3: $\quad \tilde{n}_{\text{CWC},J} = \dfrac{N_{\text{CWC}} - \sum_{k=J}^{M} \frac{1}{C}\left(\tilde{\lambda}_k - \sqrt{\tilde{\lambda}_k \tilde{\lambda}_J}\right)}{\sum_{k=J}^{M} \sqrt{\frac{\tilde{\lambda}_k}{\tilde{\lambda}_J}}}$

4: $\quad$ **for** $k = J + 1$ to M **do**

5: $\quad\quad \tilde{n}_{\text{CWC},k} = \sqrt{\frac{\tilde{\lambda}_k}{\tilde{\lambda}_J}}\tilde{n}_{\text{CWC},J} + \frac{1}{C}\left(\tilde{\lambda}_k - \sqrt{\tilde{\lambda}_k \tilde{\lambda}_J}\right)$

6: $\quad$ **end for**

7: $\quad$ flag $= 0$

8: $\quad$ **while** $\tilde{n}_{\text{CWC},J} < n_{\text{min,CWC}}$ **do**

9: $\quad\quad \tilde{n}_{\text{CWC},J} = n_{\text{min,CWC}}$

10: $\quad\quad N_{\text{CWC}} = N_{\text{CWC}} - n_{\text{min,CWC}}$

11: $\quad\quad J = J + 1$

12: $\quad\quad$ flag $= 1$

13: $\quad$ **end while**

14: $\quad$ **if not** flag **then**

15: $\quad\quad$ **while** $\tilde{n}_{\text{CWC},M} > n_{\text{max,CWC}}$ **do**

16: $\quad\quad\quad \tilde{n}_{\text{CWC},M} = n_{\text{max,CWC}}$

17: $\quad\quad\quad N_{\text{CWC}} = N_{\text{CWC}} - n_{\text{max,CWC}}$

18: $\quad\quad\quad M = M - 1$

19: $\quad\quad\quad$ flag $= 1$

20: $\quad\quad$ **end while**

21: $\quad$ **end if**

22: **end while**

Output: $\{\tilde{n}_{\text{CWC},k}\}$, J, M, $f_{\text{asc},\lambda}^{-1}(\cdot)$

$$\text{s.t. } C1 : B_k \leq \epsilon_k,$$

$$C2 : n_{\text{CWC},k} \geq n_{\text{min,CWC}},$$

where ϵ_k is the maximum CWC blockage probability at hour k. Note that ϵ_k has to be greater than $B_{\text{min},k}$, where $B_{\text{min},k}$ is the minimum CWC blockage probability at hour k when $n_{\text{CWC},k} = n_{\text{f}} - n_{\text{CWC+RAS}} - 1$ (only one subframe is allocated to RAS), otherwise no subframes will be allocated to RAS or the problem P2 will be infeasible. The constraint C2 is to guarantee at least a minimum number of subframes for CWC per each hour or it can

be viewed as a maximum CWC blockage probability $B_{\max,k} = B_k\left(n_{\min,\text{CWC}}\right)$. The solution of the resource allocation problem P2 is to set $n_{\text{CWC},k} = n_{\min,\text{CWC}}$ if $\epsilon_k \geq B_{\max,k}$. If $B_{\min,k} < \epsilon_k < B_{\max,k}$, $n_{\text{CWC},k}$ can be obtained by bisection search (or any other one-dimensional search method) within $[N_l, N_u]$, where $N_l = n_{\min,\text{CWC}}$ and $N_u = n_{\text{f}} - n_{\text{CWC+RAS}} - 1$.

Two Shared Bands

Now, we consider the case of two shared bands (Band 1 and Band 2) between CWC and RAS. The two bands may have different ratios of resource amounts between CWC and RAS given by γ_1 and γ_2. In addition, they may have different bandwidths, but we assume that they are in the same frequency range. Therefore, they can have the same frame length T_{f}, but with different numbers of resource blocks per subframe given by C_1 and C_2. Thus, the CWC throughput maximization problem P3 can be written as

$$\text{P3}: \underset{\{n_{\text{CWC}v,k}\}}{\arg\max} \sum_{k=1}^{24} \rho_k \left(C_1 n_{\text{CWC1},k} + C_2 n_{\text{CWC2},k}\right), \tag{6.20}$$

$$\text{s.t. C1}: \sum_{k=1}^{24} n_{\text{CWC}v,k} \leq N_{\text{CWC}v} \quad \forall v,$$

$$\text{C2}: n_{\min,\text{CWC}} \leq n_{\text{CWC}v,k} \leq n_{\max,\text{CWC}} \quad \forall v, k,$$

where $v \in \{1, 2\}$ referring to the band number, and $N_{\text{CWC}v} = 24\left(n_{\text{f}} - n_{\text{CWC+RAS}}\right)\gamma_v / \left(1 + \gamma_v\right)$.

Since CWC throughput is a function of the sum of total available resource blocks $r_{\text{CWC},k} = C_1 n_{\text{CWC1},k} + C_2 n_{\text{CWC2},k}$, we divide the problem P3 into two sub-problems P3.1 and P3.2 or P3.3 as will be discussed next. The first sub-problem is given as

$$\text{P3.1}: \underset{\{r_{\text{CWC},k}\}}{\arg\max} \sum_{k=1}^{24} \rho_k \left(r_{\text{CWC},k}\right), \tag{6.21}$$

$$\text{s.t. C1}: \sum_{k=1}^{24} r_{\text{CWC},k} \leq C_1 N_{\text{CWC1}} + C_2 N_{\text{CWC2}},$$

$$\text{C2}: \left(C_1 + C_2\right) n_{\min,\text{CWC}} \leq r_{\text{CWC},k} \leq \left(C_1 + C_2\right) n_{\max,\text{CWC}} \quad \forall k,$$

where in P3.1 we combine the constraints of the two bands to be similar to that of the single band case.

We can notice that P3.1 has the same structure of P1, and hence it can be solved using Algorithm 1 to obtain $\{\tilde{r}_{\text{CWC},k}\}$ where $\{\tilde{r}_{\text{CWC},k}\} = f_{\text{asc},\lambda}(\{r_{\text{CWC},k}\})$. Then, $\{\tilde{r}_{\text{CWC},k}\}$ are distributed among the two bands satisfying the constraints of the original problem P3 as follows. For $\{\tilde{r}_{\text{CWC},k}\}_{k=1}^{J-1}$ and $\{\tilde{r}_{\text{CWC},k}\}_{k=M+1}^{24}$, the corresponding CWC resource amounts $\{\tilde{n}_{\text{CWC}v,k}\}_{k=1}^{J-1}$ and $\{\tilde{n}_{\text{CWC}v,k}\}_{k=M+1}^{24}$ in each band will be the same as $n_{\text{min,CWC}}$ and $n_{\text{max,CWC}}$ respectively. For $\{\tilde{n}_{\text{CWC}v,k}\}_{k=J}^{M}$, there are many ways of distribution among the two bands having the same $\{\tilde{r}_{\text{CWC},k}\}_{k=J}^{M}$. For the case of two different bandwidths $(C_1 \neq C_2)$, we propose to choose the solution maximizing the minimum sum of numbers of CWC subframes per hour in order to have a minimum latency. In other words, the second sub-problem can be given by

$$\text{P3.2}: \quad \underset{\{\tilde{n}_{\text{CWC}v,k}\}_{k=J}^{M}}{\arg\max} \quad \min\{\tilde{n}_{\text{CWC1},k} + \tilde{n}_{\text{CWC2},k}\}_{k=J}^{M} \tag{6.22}$$

$$\text{s.t. C1}: \tilde{n}_{\text{CWC1},k}C_1 + \tilde{n}_{\text{CWC2},k}C_2 = \tilde{r}_{\text{CWC},k} \quad \forall k \in \{J, J+1, \ldots, M\},$$

$$\text{C2}: \sum_{k=1}^{24} \tilde{n}_{\text{CWC}v,k} \leq N_{\text{CWC}v} \quad \forall v,$$

$$\text{C3}: n_{\text{min,CWC}} \leq \tilde{n}_{\text{CWC}v,k} \leq n_{\text{max,CWC}} \quad \forall v, \forall k \in \{J, J+1, \ldots, M\},$$

where the constraint C1 represents the output of sub-problem P3.1. For the case of same bandwidth $(C_1 = C_2 = C)$, $\{\tilde{n}_{\text{CWC1},k} + \tilde{n}_{\text{CWC2},k}\}_{k=J}^{M}$ is already determined by $\{\tilde{r}_{\text{CWC},k}/C\}_{k=J}^{M}$. Hence, one way of allocation is to distribute the resource equally, as much as possible, among the two bands. In other words, it can be given by

$$\text{P3.3}: \quad \underset{\{\tilde{n}_{\text{CWC}v,k}\}_{k=J}^{M}}{\arg\min} \quad \max\{|\tilde{n}_{\text{CWC1},k} - \tilde{n}_{\text{CWC2},k}|\}_{k=J}^{M} \tag{6.23}$$

$$\text{s.t. C1}: \tilde{n}_{\text{CWC1},k} + \tilde{n}_{\text{CWC2},k} = \tilde{r}_{\text{CWC},k}/C \quad \forall k \in \{J, J+1, \ldots, M\},$$

$$\text{C2}: \sum_{k=1}^{24} \tilde{n}_{\text{CWC}v,k} \leq N_{\text{CWC}v}, \quad \forall v,$$

$$C3 : n_{\min,\text{CWC}} \leq \tilde{n}_{\text{CWC}v,k} \leq n_{\max,\text{CWC}} \qquad \forall v, \forall k \in \{J, J+1, \ldots, M\},$$

where the constraint C1 represents the output of sub-problem P3.1. The problems P3.2 and P3.3 are convex optimization problems since the constraints of the two sub-problems are convex and the objective functions are concave and convex respectively. Therefore, they can be solved by convex programming techniques such as interior point methods [21]. Then, elements of $\{\tilde{n}_{\text{CWC}v,k}\}_{k=J}^{M}$ are rounded by a similar way to the one used for P1. Next, we obtain $\{n_{\text{CWC}v,k}\} = f_{\text{asc},\lambda}^{-1}(\{\tilde{n}_{\text{CWC}v,k}\})$ and then $n_{\text{RAS}v,k} = n_{\text{f}} - n_{\text{CWC+RAS}} - n_{\text{CWC}v,k}$ $\forall k \in \{1, 2, \ldots, 24\}, \forall v.$

The second resource allocation problem P2 can be extended to the case of two shared bands as shown below as P4,

$$\text{P4} : \underset{n_{\text{CWC1},k},\, n_{\text{CWC2},k}}{\arg\min} \quad \max\{n_{\text{CWC1},k},\, n_{\text{CWC2},k}\}, \tag{6.24}$$

$$\text{s.t. C1} : B_k(n_{\text{CWC1},k}C_1 + n_{\text{CWC2},k}C_2) \leq \epsilon_k,$$

$$\text{C2} : n_{\text{CWC}v,k} \geq n_{\min,\text{CWC}} \qquad \forall v.$$

Similarly, P4 is divided into two sub-problems, P4.1 and P4.2, in terms of $r_{\text{CWC},k}$ as follows,

$$\text{P4.1} : \underset{r_{\text{CWC},k}}{\arg\min}\, r_{\text{CWC},k}, \tag{6.25}$$

$$\text{s.t. C1} : B_k(r_{\text{CWC},k}) \leq \epsilon_k,$$

$$\text{C2} : r_{\text{CWC},k} \geq (C_1 + C_2)\, n_{\min,\text{CWC}},$$

where P4.1 is solved by the same way as in P2. Then, $r_{\text{CWC},k}$ is distributed among the two bands, if with different bandwidths, as in P4.2,

$$\text{P4.2} : \underset{n_{\text{CWC1},k},\, n_{\text{CWC2},k}}{\arg\min} \quad \max\{n_{\text{CWC1},k},\, n_{\text{CWC2},k}\}, \tag{6.26}$$

$$\text{s.t. C1} : n_{\text{CWC1},k}C_1 + n_{\text{CWC2},k}C_2 = r_{\text{CWC},k},$$

$$\text{C2} : n_{\text{CWC}v,k} \geq n_{\min,\text{CWC}} \qquad \forall v,$$

where the constraint C1 represents the output of sub-problem P4.1. The sub-problem P4.2 is a convex optimization problem since the objective function and the constraints are convex. Therefore, it can be solved by convex programming techniques such as interior point methods [21]. If the two bands have the same bandwidth, then $n_{\mathrm{CWC1},k} = n_{\mathrm{CWC2},k} = r_{\mathrm{CWC},k}/(2C)$. Finally, $n_{\mathrm{CWC1},k}$ and $n_{\mathrm{CWC2},k}$ are rounded up.

In our resource adaptation scheme presented above, the resources are allocated for each hour. The time interval of resource allocation (which we will call resource allocation time unit) can be set to other value as needed. For example, we can set the resource allocation time unit to have a duration of 15 minutes instead of an hour, and in this case the total number of time units per day N_{time} will be 96 instead of 24.

6.4.5 Resource Adaptation based on Spectrum Access Characteristics of CWC and RAS

In practice, RAS sometimes has down-times due to maintenance, equipment update, calibrations, testing, etc. Except for RAS calibrations and testing down-times, the other down-times can be incorporated in the resource adaptation. Let Ω be the set containing the resource allocation time units of RAS down-times (except for RAS calibrations and testing down-times) for a considered day, and H be the size of Ω. If Ω is not time-sensitive within that day and can be designed, the proposed strategy is to set them to be the H resource allocation time units centered at the time unit with the highest traffic load. This setting is logical since the time units CWC wants the most with all spectrum resources are during its highest traffic load. The resource allocation algorithm first assigns all spectrum resources to CWC for the time units defined in Ω. Next, we modify the previous resource allocation problems by excluding the resource allocation time units of Ω and adjusting N_{CWC} to be $N_{\mathrm{CWC}} = \left((N_{\mathrm{time}} - H)\left(n_{\mathrm{f}} - n_{\mathrm{CWC+RAS}}\right)\gamma - Hn_{\mathrm{f}}\right)/(1+\gamma)$. Then, solving the modified problems in the same way as before gives the resource allocation for the time units outside Ω.

6.5 Built-in Fine Tuning

RAS requires to have spectrum access without harmful RFI while CWC desires to have more spectrum access opportunities. The proposed approach facilitates them by means of the SSAZ and BS-dependent guard time intervals during the CWC+RAS phase. However, the channel conditions, the cell structures, and the wireless traffic statistics in practical environments may be different from those used in the design. Thus, meeting the system design goal (e.g., avoiding harmful RFI for RAS) is uncertain. For this critical practical issue, we propose a fine tuning mechanism which is a built-in characteristic of the proposed three phase spectrum access. During initial deployment stage, the CWC+RAS phase is designed conservatively to avoid causing RFI to RAS in the RAS only phase. We also equip each RAS site with additional processing units to measure RFI from CWC and propagation delays of in-zone CWC BSs. CWC and RAS coordinate for automatic fine tuning.

In the first phase of fine tuning, we adjust the SSAZ. RAS main receiver is used to measure RFI from the CWC cells outside SSAZ during the CWC+RAS phase[4] (e.g., by subtracting the noise power estimate from the received power estimate). If RFI level is below the acceptable threshold, increasing the SSAZ is not required. Otherwise, SSAZ should be increased by including the bordering outside tier of cells. This process is repeated until the RFI level falls below the acceptable threshold. If the RFI level is substantially lower than the threshold, we can remove CWC cells in the bordering inside tier from the SSAZ, one at a time, as long as the RFI level is below the threshold.

In the second phase of fine tuning, we adjust P2TIs for cells within SSAZ. BSs within SSAZ, one at a frame according to the predefined order, transmit predefined training signals during a few beginning symbols (duration less than initial $T_{\text{ext},i}$) of the CWC+RAS phase. RAS measures power levels of the received training signals either in the same way as the RFI

---

[4]Auxiliary receivers could also be incorporated in performing these tasks as in [15, 23, 62]

measurement mentioned above or based on the power of the channel estimates computed based on the training signals, e.g., [85]. RAS also measures the propagation delays of those training signals (e.g., based on the delay in the channel estimates [85]). Then $\{T_{\text{ext},i}\}$ are adjusted to be the largest integer multiple of T_{sym} such that CWC transmission during $T_{\text{ext},i}$ at the beginning of the CWC+RAS phase does not cause RFI to the RAS only phase. RAS control station can compute the updated $T_{\text{ext},i}$ and feed it back to BS i.

The above two phases of fine tuning just need to be done once at the initial deployment of the shared spectrum access. During that time, both CWC and RAS can operate in their respective CWC only phase and RAS only phase, thus, without any interruption of their normal operation. If needed, the CWC+RAS phase duration can also be adjusted. Due to service evolution, new CWC cell sites may need to be added or the existing cells may need to be re-structured (e.g., with different coverage zones). Similarly, new RAS sites may be introduced or a different level of RFI threshold may be imposed. For those scenarios, the fine tuning task can be invoked on a demand basis during the CWC+RAS phase. This provides a flexible coexistence between CWC and RAS while accommodating service evolutions of CWC and RAS.

Note that backbone communication between CWC and RAS is required for coordination but it would be infrequent and with low data rate. Furthermore, timing synchronization is required between CWC BSs and RAS receivers. For this, one can adopt an existing inter-BS synchronization scheme as used in TDD systems [132] to avoid inter-cell interference between uplink and downlink. In fact, the fine tuning task also provides fine timing synchronization. As mentioned before, each BS can transmit known training signal during the fine tuning stage from which RAS can obtain the combined timing offset and propagation delay (e.g., [85]). Then RAS can provide feedback to CWC BSs through the backbone link to adjust their transmission time. Since this feedback is not time-critical, the feedback link can be based on a dedicated backbone link or even an internet-based connection between CWC

Table 6.1. SSAZ radius for different CWC cell types and frequency bands

CWC Cell Type	System Parameters	SSAZ Radius (R_{SSAZ})	
		Hexagon Cells	PPP Model
Macro Cells	$f_c = 2.695$ GHz, $R_{\text{CWC}} = 1$ km $P_{\text{BS}} = 43$ dBm, $G_{\text{CWC}} = 0$ dB $n_{\text{NLOS}} = 4$ [47], $\beta = 0.006$ dB/km [58] $G_{\text{RAS}} = -13$ dB [59][5], $I_{\text{th}} = -207$ dBW [56] $h_{\text{CWC}} = h_{\text{RAS}} = 24$ m	139.45 km	139.59 km
Small Cells	$f_c = 24$ GHz, $R_{\text{CWC}} = 200$ m $P_{\text{BS}} = 30$ dBm, $G_{\text{CWC}} = 0$ dB $n_{\text{NLOS}} = 3.4$ [112], $\beta = 0.23$ dB/km [58] $G_{\text{RAS}} = -13$ dB [59], $I_{\text{th}} = -195$ dBW [56] $h_{\text{CWC}} = h_{\text{RAS}} = 24$ m	40.06 km	40.06 km

and RAS inter-system control/management units. Such fine timing synchronization can be implemented on a regular basis (e.g., once per day) or a need-basis.

6.6   Performance Evaluation

6.6.1   Shared Spectrum Access Zone (SSAZ)

The first step in our design is to determine the shared spectrum access zone. We evaluate the radius of the SSAZ for two different types of CWC cells and frequency ranges, namely, macro cells operating at the microwave frequency band and small cells operating at/near the millimeter wave (simply called mmWave) band. The systems parameters and the corresponding SSAZ radii are listed in Table 6.1. We observe that R_{SSAZ} with small cells is quite smaller than that with macro cells due to the higher propagation path loss and the smaller BS transmit power of mmWave signals. In addition, we observe that R_{SSAZ} obtained with the assumption of hexagon CWC cells is approximately the same as that obtained with the assumption of PPP model with $\rho_{\text{BS}} = \left(3\sqrt{3}R_{\text{CWC}}^2/2\right)^{-1}$. Next, we perform our remaining

---

[5]The radio astronomy station receives CWC interference through the antenna side lobe.

Table 6.2. Simulation parameters for mmWave shared bands

Parameter	Value
Carrier frequencies $(f_{c,1},\ f_{c,2})$	24, 26 GHz
Bandwidth (W)	500 MHz
Subcarrier spacing (Δf)	720 kHz
Index set of used subcarriers $(\mathcal{I})$	$\{-345, \ldots, -1, 1, \ldots, 345\}$
Symbol duration (T_{sym})	1.604 μs
Cyclic prefix (CP)	215 ns
Subframe duration (T_{sf})	38.5 μs
Frame duration (T_{f})	1.694 ms
No. subframes per frame (n_{f})	44
No. of resource blocks per subframe (C)	115
No. of subcarriers per resource block	6
Traffic normalization constant (α)	1.2
Ratio of resource amounts per day between CWC and RAS (γ)	4/3
Min. RAS duration per frame $(T_{\mathrm{min,RAS}})$	0.4235 ms $(n_{\mathrm{min,RAS}} = 11)$

performance evaluations using mmWave small cells since it is more appealing to incorporate the proposed paradigm into the future 5G networks so as to avoid disruption to the operation of current systems.

6.6.2 System Design for mmWave Shard Bands

The simulation parameters used in our next design are listed in Table 6.2 in addition to the mmWave system parameters in Table 6.1. A subcarrier spacing of 720 kHz is used to have robustness against phase noise. The typical RMS delay spread for mmWave channels is in the order of 10 ns-100 ns [109]. Thus, a cyclic prefix (CP) of 215 ns is appropriate to mitigate the inter symbol interference (ISI). The CP overhead is about 13.5% of the symbol duration. A time division duplex (TDD) mode is considered in which the uplink and downlink subframes alternate. A guard period (GP) of 0.7 μs between the downlink and uplink subframes is used to absorb the propagation delay.

As two bands (24 GHz and 26 GHz) are used, (6.8) yields an SSAZ with radius of 45.5 km which is larger than the single band result in Table 6.1 due to the RFI induced by the adjacent band. Note that this SSAZ size requires a time duration of 151.69 μs to absorb the propagation delays of different CWC BSs within the SSAZ. Therefore, we set $T_{\text{CWC+RAS}} = 154$ μs, i.e., $n_{\text{CWC+RAS}} = 4$ subframes.

We evaluate the existing paradigm where the spectrum is allocated solely to CWC and the proposed paradigm where the spectrum is shared between CWC and RAS according to the proposed shared spectrum access. We consider one band and two bands. For the two bands case, for simplicity of exposure, we consider that each of the two bands has the same bandwidth as in the one band case, i.e., $C_1 = C_2 = C$, and the ratio of the average spectrum requirement between CWC and RAS is the same for the two bands, i.e., $\gamma_1 = \gamma_2 = \gamma$. Similarly, the traffic loads are doubled for the two bands case if compared to the one band case, and hence we set λ_{max} for the two bands case twice that for the one band case. Due to these settings, the resource allocations of the two shared bands will be the same.

6.6.3 Radio Latency

We consider the round trip time (RTT) as the minimum radio latency. The RTT is defined as the time from the start of data transmission to the reception of the acknowledgment (ACK). A simple RTT model is considered in which one subframe is used for data transmission, data decoding takes one subframe interval, and then ACK is transmitted in the next reverse link subframe. Therefore, the minimum RTT is about three subframes duration of 115.5 μs. However, due to the RAS observation duration, the ACK of the last uplink/downlink subframe is received in the next frame. In addition, if a service is required during the RAS period, it has to wait until the next CWC period. Thus, the minimum number of CWC subframes per frame $n_{\text{min,CWC}}$ affects the radio latency requirement. On the other hand, for most of the existing CWC services, an average radio latency of 1 ms is more than sufficient.
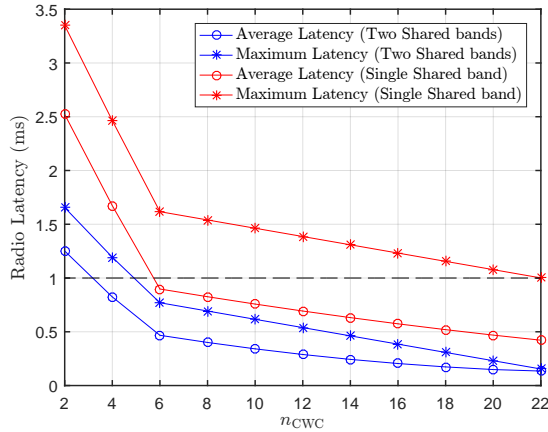
Figure 6.6. Maximum and average radio latency for CWC

Fig. 6.6 shows the average and maximum radio latency as a function of $n_{\min,\text{CWC}}$ for the proposed paradigm. The results show that in order to satisfy a maximum radio latency of 1 ms, $n_{\min,\text{CWC}}$ has to be at least 22 subframes for the single shared band case. On the other hand, $n_{\min,\text{CWC}}$ is only 6 subframes for the two shared bands case assuming that TDD is applied in both bands. If the constraint is an average radio latency of 1 ms instead of the maximum radio latency of 1 ms, $n_{\min,\text{CWC}}$ decreases for both one band and two bands cases, which is expected. Note that a smaller value of $n_{\min,\text{CWC}}$ (for the two bands case if compared to the one band case) could reduce unused spectrum for CWC during very low traffic load, thus it could enhance spectrum utilization efficiency.

Note that defining $n_{\min,\text{CWC}}$ puts a minimum value $\gamma_{\min}$ for the ratio of the total resource amounts per day between CWC and RAS that can be achieved as $\gamma_{\min} = n_{\min,\text{CWC}}/(n_{\text{f}} - n_{\min,\text{CWC}} - n_{\text{CWC+RAS}})$. As a result, τ_{latency}, $n_{\min,\text{CWC}}$, and $\gamma_{\min}$ are related to each other, and sometimes there may be a trade off between τ_{latency} and γ. For example, for the single shared band case and a maximum radio latency of 1 ms ($n_{\min,\text{CWC}} = 22$), the equal resource allocation between CWC and RAS per day ($\gamma = 1$) is not achievable since $\gamma_{\min} = 1.22$. However, these requirements are achievable for the two shared bands case since $\gamma_{\min} = 0.1765$ with $n_{\min,\text{CWC}} = 6$. Based on the results in Fig. 6.6, for the next performance evaluation, we

142
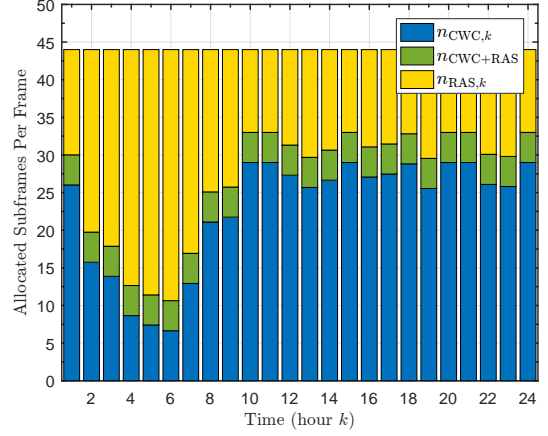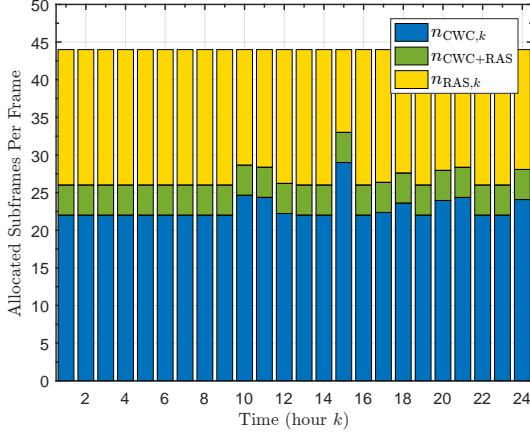
(a) Resource allocation for single shared band ($n_{\min,\mathrm{CWC}} = 22$)

(b) Resource allocation for two shared bands ($n_{\min,\mathrm{CWC}} = 6$)

Figure 6.7. Adaptive resource allocation for maximizing average throughput

set $n_{\min,\mathrm{CWC}}$ to 22 and 6 subframes for the single shared band and two shared bands cases, respectively.

6.6.4 CWC Throughput

Next, we evaluate CWC throughput performance for the one band and two bands cases under the CWC-only spectrum allocation paradigm and the proposed shared spectrum access paradigm. We apply the solution for maximizing the average CWC throughput (i.e., for P1 and P3). Fig. 6.7a and Fig. 6.7b show the resource allocation in each band during each hour for the cases of single shared band (P1) and two shared bands (P3), respectively. The corresponding CWC throughput performances are shown in Fig. 6.8 where the average throughput is normalized by the maximum average traffic load of each of the one band and two bands cases for the convenience of presentation. Thus, for the CWC-only spectrum allocation, the normalized average throughputs are the same for the one band and two bands cases, and they are shown as one entity in the plot. For relative (un-normalized) average throughput comparison between the one band and two bands cases, the values for the two bands cases in the figure should be doubled.

The results in Fig. 6.7a and Fig. 6.7b show that for the two shared bands case, the resource amounts allocated to CWC approximately follow CWC's average traffic loads shown in Fig. 6.5 in a better way than that for the single shared band case. At the hours with low traffic load, the resource amount needed to maintain the CWC latency requirement is larger than the average traffic demand for the evaluated system. Thus, it is observed that the resource amounts allocated to CWC at the low-traffic hours are set to $n_{\min,\text{CWC}}$ to satisfy the latency requirement. This illustrates a tradeoff scenario between spectrum utilization and latency support for CWC. At the hours with high traffic load, the resource amounts allocated to CWC are limited by $n_{\max,\text{CWC}}$. Consequently, as shown in Fig. 6.8, CWC experiences reduced throughput at the hours with high traffic load which is expected due to allocation of a minimum of $n_{\min,\text{RAS}} = 11$ subframes per frame to RAS. We also observe in Fig. 6.8 that the two band sharing gives some throughput advantage over the one band sharing. The reason can be explained as follows. Due to the frame time offset between the two bands, the resource amount needed to maintain latency requirement is smaller for the two band scenario than the one band scenario. This translates that a smaller resource amount is allocated at low-traffic hours (thus a slightly smaller normalized throughput) but a larger resource amount at hours with medium or close-to-high traffic loads (thus a larger normalized throughput) for the two band case than the single band case. Furthermore, we observe that except at peak traffic load, the proposed paradigm (especially with two shared bands) yields CWC throughput similar to the CWC-only spectrum allocation.

6.6.5 CWC Blockage Probability

Next, we evaluate CWC traffic blockage probability of the proposed paradigm under the resource allocation problems P2 (for one band) and P4 (for two bands). Fig. 6.9 shows the allocated subframes in each band to RAS and Fig. 6.10 presents the corresponding CWC blockage probability during each hour. To satisfy CWC's maximum latency constraint of 1
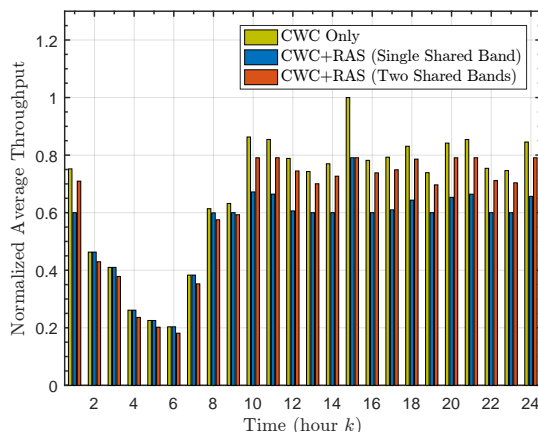
Figure 6.8. CWC average throughput comparison between the CWC only allocation and the proposed shared allocation

ms, the maximum number of RAS subframes per frame $n_{\mathrm{max,RAS}}$ is 18 and 34 for the single shared band and two shared bands cases respectively. Therefore, as observed in Fig. 6.9, for the two shared band case, more resources can be allocated to RAS at the hours with low traffic load than that for the single shared band case while satisfying the CWC blockage probability constraint of 0.1. The results in Fig. 6.9 yield the resource ratio of $\gamma = 1.7586$ for the single band and $\gamma = 1.2482$ for the two bands case. On the other hand, Fig. 6.10 shows no blockage at the hours with low traffic load for the single shared band case since the minimum CWC subframes $n_{\mathrm{min,CWC}}$ to satisfy the latency requirement is notably greater than the minimum CWC subframes to satisfy the blockage probability constraint. Overall, the proposed allocations for both one band and two bands cases satisfy latency and blockage probability constraints and the two bands case gives more spectrum access to RAS while the one band case yields smaller CWC blockage probability.

6.6.6 Average Spectrum Utilization

As spectrum is a precious resource, spectrum utilization is an important performance indicator for spectrum allocation and access policies/technologies. Fig. 6.11 shows the average spectrum utilization per hour defined as $\eta_k = (\rho_k + n_{\mathrm{RAS},k}C)/(n_{\mathrm{f}}C)$ for the single shared
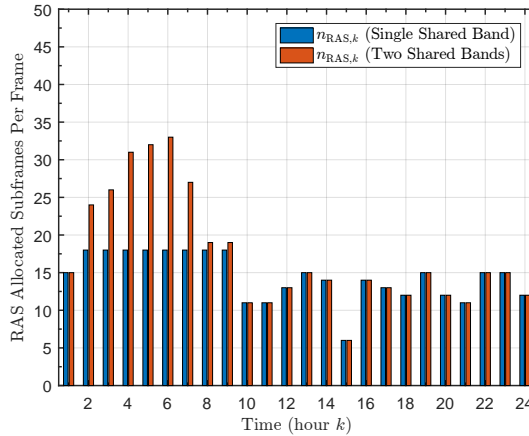
Figure 6.9. Allocated subframes per band to RAS during each hour under P2 and P4
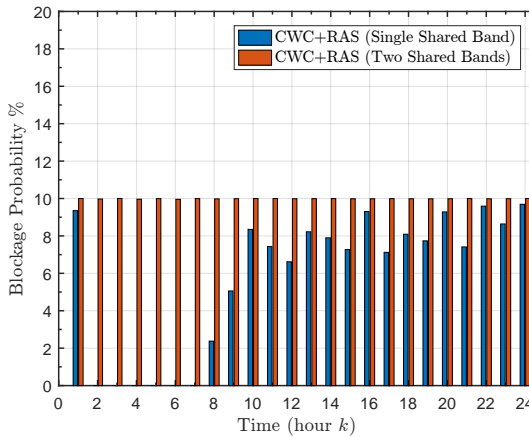
Figure 6.10. CWC blockage probability during each hour under P2 and P4

band (P1) and $\eta_k = (\rho_k + n_{\mathrm{RAS1},k}C_1 + n_{\mathrm{RAS2},k}C_2)/(n_{\mathrm{f}}(C_1 + C_2))$ for the two shared bands (P3). The results under P2 and P4 show similar behavior and hence the plot is omitted. As expected, the proposed shared spectrum access yields average spectrum utilization of about 91% which is much larger than if the spectrum is allocated to CWC only. Under the proposed paradigm, the two shared bands case yields higher spectrum utilization efficiency than the single shared band case. The reason is that for the single shared band case, the average spectrum utilization decreases at the hours with low traffic load since $n_{\mathrm{min,CWC}} C$ imposed by the latency constraint for this case is greater than the average required resource amounts.

146

Figure 6.11. Spectrum utilization comparison between the CWC only allocation and the shared allocation

6.6.7 Performance for Different Deployment Scenarios

The performance results presented in the previous sections could be related to several deployment scenarios. If the scenario is that new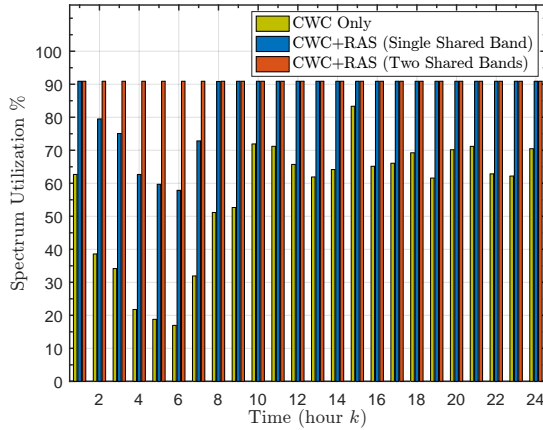 band(s) is (are) allocated for shared spectrum access between CWC and RAS and CWC has no other band, then the results of the proposed shared access represent an example performance of a fair and guaranteed shared spectrum access. In this case, both CWC and RAS benefit from the new band(s).

Suppose the scenario is that CWC also has another non-shared band in addition to the new shared band(s). Then the CWC's non-shared band can be used to handle its latency requirement and the constraint on $n_{\text{min,CWC}}$ can be removed or reduced in the resource allocation optimization. The result would be a throughput improvement for CWC, and an example result can be seen in Fig. 6.8 by comparing the normalized throughput between the case with $n_{\text{min,CWC}} = 22$ (single band) and that with $n_{\text{min,CWC}} = 6$ (two bands).

Another scenario is that band 1 was previously allocated to CWC and band 2 was to RAS, but now the two bands are restructured as shared bands for CWC and RAS. In this scenario, the advantage of the spectrum restructuring can be observed as follows. First, RAS now has certain guaranteed spectrum access in both bands but in the previous allocation RAS does not have guaranteed spectrum access in band 1. Regarding CWC throughput

147

performance, by comparing the normalized average throughput for the CWC only case with two times the normalized average throughput of the proposed two bands sharing case, we can infer that the spectrum restructuring provides capability of serving much higher peak traffic as well as mean traffic. Overall, the proposed shared spectrum access paradigm offers a win-win outcome for both CWC and RAS.

6.7   Conclusions

We have proposed a new paradigm of shared spectrum access between CWC and RAS by means of the three-phase spectrum access strategy. Shared spectrum access zones (SSAZs) are established around RAS sites. CWC systems within those zones follow the three-phase access while those outside SSAZ have full spectrum access. Both time-division and time-frequency division three-phase access schemes are presented. Interaction between CWC latency, minimum resource amount and resource ratio between CWC and RAS is discussed. A built-in fine tuning mechanism is described for addressing design mismatches and service evolutions. Average traffic pattern based resource allocation is further developed. The simulation results show that CWC systems within SSAZ experience slight throughput reduction at peak traffic hours (if compared to the case where spectrum is solely allocated to CWC only) but RAS achieves substantial RFI-free spectrum access which were infeasible in the existing paradigm. When a new shared band is allocated or the existing CWC and RAS bands are restructured as shared bands, the proposed approach offers benefits to both CWC and RAS. The overall spectrum utilization is also substantially enhanced, thus illustrating high potentials of the proposed paradigm.

Appendix A - Derivation of (6.11)

Using the Gaussian approximation of Poisson distribution, we approximate ρ_k by $\tilde{\rho}_k$ as

$$
\tilde{\rho}_k = \int_{-\infty}^{\infty} \min(n, n_{\mathrm{CWC},k}C) \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n = \int_{-\infty}^{n_{\mathrm{CWC},k}C} n \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n
$$

$$
+ n_{\mathrm{CWC},k}C \int_{n_{\mathrm{CWC},k}C}^{\infty} \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n
$$

$$
= \int_{-\infty}^{n_{\mathrm{CWC},k}C} (n-\lambda_k) \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n + \lambda_k \int_{-\infty}^{n_{\mathrm{CWC},k}C} \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n
$$

$$
+ n_{\mathrm{CWC},k}C \left(1 - \int_{-\infty}^{n_{\mathrm{CWC},k}C} \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n \right)
$$

$$
= n_{\mathrm{CWC},k}C - (n_{\mathrm{CWC},k}C - \lambda_k)\,\Phi\left(\frac{n_{\mathrm{CWC},k}C - \lambda_k}{\sqrt{\lambda_k}}\right) - \sqrt{\frac{\lambda_k}{2\pi}} \exp\left(\frac{-(n_{\mathrm{CWC},k}C - \lambda_k)^2}{2\lambda_k}\right), \quad (6.27)
$$

where $\Phi(x)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Appendix B - Derivation of (6.12)

Using (6.27), we can write $\frac{\partial \tilde{\rho}_k}{\partial n_{\mathrm{CWC},k}}$ as

$$
\frac{\partial \tilde{\rho}_k}{\partial n_{\mathrm{CWC},k}} = \frac{\partial}{\partial n_{\mathrm{CWC},k}} \left[n_{\mathrm{CWC},k}C - \int_{-\infty}^{n_{\mathrm{CWC},k}C} \underbrace{(n_{\mathrm{CWC},k}C - n) \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}}_{g\left(n, n_{\mathrm{CWC},k}C\right)}\, \mathrm{d}n \right]. \quad (6.28)
$$

Applying Leibniz integral rule [120, Eq. A.2-1]), we get

$$
\frac{\partial \tilde{\rho}_k}{\partial n_{\mathrm{CWC},k}} = C - \left(\int_{-\infty}^{n_{\mathrm{CWC},k}C} \frac{\partial g\left(n, n_{\mathrm{CWC},k}C\right)}{\partial n_{\mathrm{CWC},k}}\, \mathrm{d}n \right) = C - \left(C \int_{-\infty}^{n_{\mathrm{CWC},k}C} \frac{\exp\left(\frac{-(n-\lambda_k)^2}{2\lambda_k}\right)}{\sqrt{2\pi\lambda_k}}\, \mathrm{d}n \right)
$$

$$
= C \left(1 - \Phi\left(\frac{n_{\mathrm{CWC},k}C - \lambda_k}{\sqrt{\lambda_k}}\right) \right). \quad (6.29)
$$

Appendix C - Derivation of (6.18)

For hour k, the CWC blockage probability, B_k, in (6.18) is derived as follows.

$$
\begin{aligned}
B_k &= \frac{1}{\lambda_k} \sum_{n=n_{\text{CWC},k}C+1}^{\infty} \left(n - n_{\text{CWC},k}C\right) \frac{\lambda_k^n \exp\left(-\lambda_k\right)}{n!} \\
&= \sum_{n-1=n_{\text{CWC},k}C}^{\infty} \frac{\lambda_k^{n-1} \exp\left(-\lambda_k\right)}{(n-1)!} - \frac{n_{\text{CWC},k}C}{\lambda_k} \sum_{n=n_{\text{CWC},k}C+1}^{\infty} \frac{\lambda_k^n \exp\left(-\lambda_k\right)}{n!} \\
&= 1 - Q\left(n_{\text{CWC},k}C - 1, \lambda_k\right) - \frac{n_{\text{CWC},k}C}{\lambda_k}\left(1 - Q\left(n_{\text{CWC},k}C, \lambda_k\right)\right),
\end{aligned}
\qquad (6.30)
$$

where $Q\left(x, \lambda\right) = \sum_{n=0}^{x} \frac{\lambda^x \exp(-\lambda)}{x!}$ is the regularized incomplete gamma function.

SPECTRUM SHARING BETWEEN WIFI AND RADIO ASTRONOMY [1]

7.1 Introduction

One of the most widely used active wireless systems is WiFi but its deployment has caused 5-6 GHz band close to unusable for RAS (a feedback from a radio astronomer at Arecibo radio observatory) over the past 10 years or so. There are no protected RAS bands there but there are some exciting spectral lines to observe in that band. Thus, this chapter develops an approach for spectrum coexistence between WiFi and RAS. Note that WiFi uses distributed MAC. Our work on the shared spectrum access between cellular wireless communications with centralized MAC and RAS is presented in chapter 6.

Our main contribution in this chapter is a new paradigm of spectrum coexistence between WiFi and RAS based on time-division-embedded distributed WiFi MAC protocols which enable geographical and spectral coexistence between WiFi and RAS. We present detailed design aspects, proposed MAC protocols, and advanced resource allocation based on WiFi's traffic statistics. Simulation study illustrates that the proposed approach can address the needs of WiFi and RAS and provide substantially enhanced spectrum utilization.

7.2 Coexistence Access Zone (CAZ)

An important step in developing the proposed coexistence access paradigm is to define a coexistence access zone (CAZ) within which WiFi and RAS follow the proposed coexistence spectrum access protocol while WiFi systems outside CAZ can access their spectrum freely but their transmissions do not cause harmful RFI to RAS. For this, we adopt the interference

---

power level threshold as defined in ITU-R RA.769-2 as the maximum RFI level the RAS receiver can tolerate and denote it by I_{th}. We consider downlink of WiFi with full load to determine the RFI level at RAS. Suppose the CAZ is with radius R_{CAZ} kilometers centered at the RAS receiver, and the WiFi hexagon cells, each with radius r kilometers, surround the RAS receiver as shown in Fig. 7.1. In calculating interference power experienced at RAS, we approximate each hexagonal cell by its inscribed circular cell with radius $\sqrt{3}r/2$. This yields closer WiFi access points (APs) to RAS than the hexagon case, thus giving an upper bound on the actual interference power and hence a safer design for RFI protection of RAS. The ith tier cell ring has $K_i \triangleq \lfloor \pi(2i-1) \rfloor$ cells with the same distance $d_i = (2i-1)\sqrt{3}r/2$ between AP and RAS receiver. Denote the AP of cell j at the ith tier cell ring by $\text{AP}_{i,j}$. Then, the CAZ design can be given by the radius R_{CAZ} of the CAZ which is computed based on I_{out} outer tiers of WiFi cells as

$$R_{\text{CAZ}} = \min_{i_0}\{(2i_0 - 1)\sqrt{3}r/2\}, \tag{7.1}$$

$$\text{s.t.} \sum_{i=i_0+1}^{i_0+I_{\text{out}}} \frac{K_i P_{\text{AP}}}{L_{\text{tot}}(d_i)} < I_{\text{th}},$$

where P_{AP} is the transmitted power of WiFi APs, and $L_{\text{tot}}(d_i)$ is the total propagation loss at the ith tier. Note that due to terrain constraints, associated path-loss and antenna down-tilting, using a finite I_{out} is justified. Fig. 7.1 illustrates the tiers of WiFi hexagon cell rings around an RAS receiver, as well as the outer tiers ($I_{\text{out}} = 2$ for presentation convenience) used in defining the CAZ. In this illustration, the CAZ is composed of 3 tiers (for presentation convenience) of WiFi cells, and WiFi transmissions in these cells follow the proposed coexistence access protocol. WiFi transmissions in other cells follow their original access protocol.

The total propagation loss is given by

$$L_{\text{tot}} = \frac{L_{\text{p}}}{G_{\text{WiFi}}G_{\text{RAS}}L_{\text{s}}}, \tag{7.2}$$
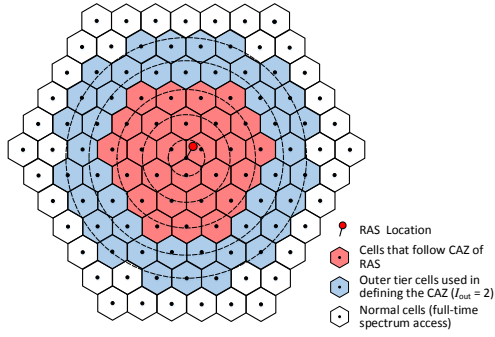
152

Figure 7.1. Tiers of WiFi cell rings around an RAS receiver.

where G_{WiFi} is the transmit antenna gain of WiFi APs, G_{RAS} is the antenna gain of RAS receivers (for WiFi interferences), that appears in the band of RAS, and L_{p} is the propagation path loss. WiFi adopts orthogonal frequency division multiplexing (OFDM) technology. Suppose the OFDM system has a subcarrier spacing of Δf, an OFDM symbol duration of $1/\Delta f'$ (including cyclic prefix interval; thus $\Delta f' < \Delta f$), N_{u} used subcarriers with their subcarrier index set $\mathcal{I}$, and the carrier frequency of $f_{\text{c}}^{\text{WiFi}}$. Then the power spectral density (PSD) of the WiFi OFDM signal is given by

$$S_{\text{WiFi}}(f) = \sum_{i \in \mathcal{I}} \frac{P_{\text{AP}}}{N_{\text{u}}\Delta f'} \text{sinc}^2 \left(\frac{f - f_{\text{c}}^{\text{WiFi}} - i\Delta f}{\Delta f'} \right) \tag{7.3}$$

where $\text{sinc}(x) \triangleq \sin(\pi x)/(\pi x)$. Note that the existing PSD expression (e.g., in [64]) uses $\Delta f' = \Delta f$ in the above equation which is inaccurate due to the cyclic prefix. Fig. 7.2 shows how the PSD of WiFi signal appears at the observation band of RAS since both systems operate at frequency bands close to each other, where $f_{\text{c}}^{\text{RAS}}$ is the RAS observation band center frequency, W^{RAS} is the RAS observation bandwidth, and W^{WiFi} is the bandwidth of WiFi. The normalized interfering WiFi signal power within the RAS bandwidth, denoted by L_{s}, is calculated as

$$L_{\text{s}} = \int_{f_{\text{l}}}^{f_{\text{u}}} \frac{S_{\text{WiFi}}(f)}{P_{\text{AP}}} df < \sum_{i \in \mathcal{I}} \int_{f_{\text{l}}}^{f_{\text{u}}} \frac{\Delta f'/(N_{\text{u}}\pi^2)}{(f - f_{\text{c}}^{\text{WiFi}} - i\Delta f)^2} df$$

$$= \sum_{i \in \mathcal{I}} \frac{\Delta f' \, W^{\text{RAS}}/(N_{\text{u}}\pi^2)}{(f_{\text{c}}^{\text{WiFi}} + i\Delta f - f_{\text{l}})(f_{\text{c}}^{\text{WiFi}} + i\Delta f - f_{\text{u}})} \tag{7.4}$$
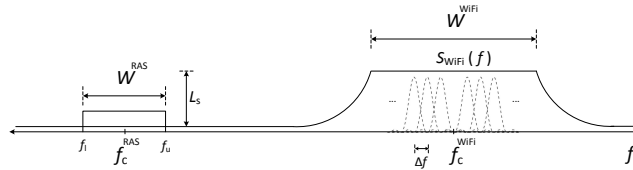
153

Figure 7.2. PSD of WiFi appears at the observation band of RAS.

where $f_l = f_c^{\text{RAS}} - \frac{W^{\text{RAS}}}{2}$ and $f_u = f_c^{\text{RAS}} + \frac{W^{\text{RAS}}}{2}$.

Note that L_s is evaluated in [64] after applying Maclaurin expansion of the function $\text{sinc}^2(x)$. Therefore, the expression of L_s in [64] is accurate only for any small frequency window inside the WiFi band. Since the two systems are separated in frequency, using the expression of L_s in [64] is not valid. That is why we derive an upper bound which is valid for any frequency separation.

As for the propagation path loss L_p, we adopt the empirical propagation model in [82]. This model is shown to have good fit to actual measurements for large distances in order of tens kilometers [11].

7.3 Coexistence MAC Protocols

After determining the CSZ, the next step is to show how WiFi MAC protocol can be modified (for WiFi devices within the CAZ) to enable the coexistence between the two systems. The idea is to develop a time-division approach such that all WiFi devices within the CAZ are silent during the time intervals allocated to RAS. Since WiFi systems use distributed medium access control (MAC) protocol, to maintain compatibility with the existing WiFI MAC protocol principle, we propose to embed the time division spectrum access between WiFi and RAS within the distributed MAC framework. We make a brief review on the WiFi access and sensing modes. Then, we present the proposed coexistence schemes.

### 7.3.1 An overview on WiFi access and sensing modes

The MAC layer of WiFi networks is based on carrier sense multiple access with collision avoidance (CSMA/CA). The fundamental mechanism is called distributed coordination function

(DCF) which is a random access scheme based on CSMA/CA. DCF has two access modes for packet transmission. The first one is the basic two-way hand shaking mode called basic access. If the channel is sensed idle for a time duration equal to a distributed interframe space (DIFS), the WiFi node can transmit. If the channel is sensed busy, the node monitors the channel until it is measured idle for a DIFS, and then it generates a random backoff interval before transmission. The successful transmission is identified by the reception of an immediate acknowledgment after a time duration called short interframe space (SIFS). Note that the basic access mode applies physical carrier sensing. The second access mode is the four way hand shaking mode called request-to-send/clear-to-send (RTS/CTS). After the channel is sensed idle or after the random backoff time, the sender WiFi STA reserves the channel by sending an RTS frame. The receiver sends a CTS frame to acknowledge the reception of RTS. The RTS and CTS frames include information about the length of the packet to be transmitted and hence the time duration needed for transmission. Other WiFi nodes apply virtual carrier sensing by decoding the CTS frame and setting an internal counter called network allocation vector (NAV) equal to the time duration needed for transmission. During this duration, they do not sense or access the medium. Similarly, the successful transmission is identified by the reception of an immediate acknowledgment after a SIFS.

7.3.2   Proposed coexistence MAC protocols

We consider beacon transmission based WiFi association [111]. WiFi AP transmits a beacon frame periodically, and WiFi STAs connect to that WiFi AP if the beacon frame is received properly. The beacon frame is transmitted every target beacon transmission time (TBTT), denoted by T_{TBTT}, after a time duration defined by the point coordination function interframe space (PIFS) to ensure the medium is free.

The proposed coexistence schemes are shown in Fig. 7.3. Across time, we divide every TBTT into three phases, namely, WiFi only phase, Pre-RAS phase, and RAS only phase
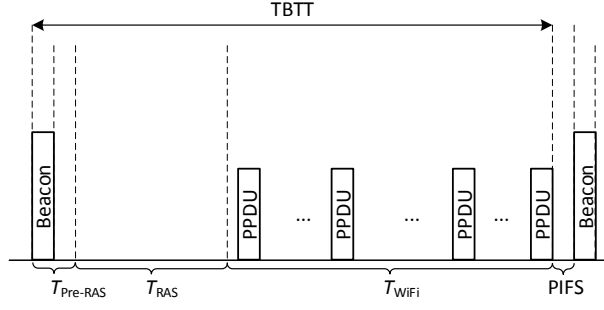
(a) Beacon based coexistence MAC protocol
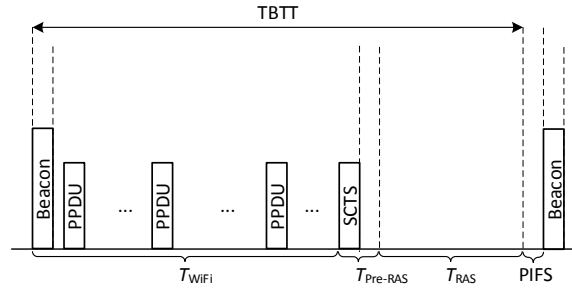
(b) CTS based coexistence MAC protocol

Figure 7.3. Proposed coexistence MAC protocols for beacon transmission based WiFi association.

with time durations of T_{WiFi}, $T_{\text{Pre}-\text{RAS}}$, and T_{RAS} respectively. The duration of the three phases can be time-dependent and different from TBTT to another TBTT. During the WiFi only phase, WiFi nodes can transmit physical protocol data units (PPDUs) but RAS does not collect data as it is heavily corrupted by RFI from WiFi devices. In fact, during such time intervals, RAS switches off its external scientific data measurement but it can measure its reference noise power level (as used in VLA telescopes of NRAO). The Pre-RAS phase consists of either a beacon frame or a special clear-to-send (SCTS) frame (depending on the WiFi access mode) followed by a guard time T_{G} to absorb the propagation delays of different WiFi devices. T_{G} can be designed as

$$T_{\text{G}} = \kappa\sigma \left\lceil \frac{R_{\text{CAZ}}/\left(3 \times 10^8\right)}{\sigma} \right\rceil \tag{7.5}$$

156

where R_{CAZ} is in meters, σ is the WiFi time-slot duration in seconds, and $\kappa \geq 1$ (e.g., $\kappa = 2$) is to accommodate propagation model mismatches. During the RAS only phase, there is no WiFi transmission within the CAZ. Thus, RAS is free from WiFi-induced RFI. RAS conducts reliable data collection during this phase.

Time-division-embedded beacon-based distributed MAC: The proposed beacon based coexistence protocol is based on a modified basic WiFi MAC access mode as shown in Fig. 7.3a. The beacon transmission based WiFi association is modified to act as follows. After the reception of the beacon frame, WiFi nodes keep silent for a time duration equal to $T_{\mathrm{G}} + T_{\mathrm{RAS}}$. Note that $T_{\mathrm{Pre-RAS}} = T_{\mathrm{B}} + T_{\mathrm{G}}$, where T_B is the time duration of the beacon frame. The remaining time of T_{TBTT} is left for WiFi only phase. For practical deployment, this proposed protocol needs to be included in the WiFi access modes as a new WiFi access mode for coexistence. Moreover, the RAS observation time should be known and updated for all WiFi nodes within the CAZ.

Time-division-embedded CTS-based distributed MAC: Fig. 7.3b shows the proposed CTS-based coexistence scheme, where we assume that DCF applies RTS/CTS access mode in order to exploit the virtual carrier sensing. During the Pre-RAS phase, WiFi APs broadcast a special CTS (SCTS) frame having a value in the packet length field long enough for the RAS observation time which is equal to the remaining time in that TBTT. After receiving this SCTS frame, WiFi STAs keep silent until the next TBTT. The SCTS frame is followed by the guard time T_{G}. Note that $T_{\mathrm{Pre-RAS}} = T_{\mathrm{CTS}} + T_{\mathrm{G}}$, where T_{CTS} is the time duration of the CTS frame. It is worth mentioning that one SCTS can make WiFi devices silent for a time duration between 1ms and approximately 32ms with 1ms step. A typical value for T_{TBTT} is 100ms. Thus if the required RAS observation time is more than 32ms per each TBTT, multiple SCST are sent to satisfy this requirement. Note that for the proposed CTS based coexistence scheme, WiFi APs should know the required RAS observation time, while WiFi STAs do not need this information. Therefore, any new WiFi STA with CTS

access mode entering the CAZ will follow automatically the proposed CTS based coexistence scheme. For practical deployment, only WiFi APs will need a modified protocol to transmit SCTS frames for reserving RAS spectrum access times.

Note that synchronization is required between WiFi backbone network and RAS for the proposed coexistence schemes.

7.4 Resource Adaptation

WiFi wireless traffics show specific temporal usage characteristics. Utilizing such traffic statistics, we propose to enhance the resource allocation through adaptation across time. Under the constraint that RAS must have a minimum resource amount in each frame and the ratio of the total resource amounts per day between WiFi and RAS is maintained, we optimize the resource allocation across time to maximize WiFi traffic support.

The average per-user WiFi throughput at hour k (denoted by ρ_k) is given by

$$\rho_k = R_{\text{PPDU}} \tilde{\rho}_k \frac{T_{\text{WiFi},k}}{T_{\text{TBTT}}} \tag{7.6}$$

where R_{PPDU} is the physical data rate and $\tilde{\rho}_k$ is the normalized WiFi throughput at hour k provided that WiFi utilizes the whole TBTT [34]. $\tilde{\rho}_k$ is a function of the packet arrival rate λ_k at hour k, and hence it can be different for a different k.[2]

Suppose WiFi and RAS require minimum time durations $T_{\text{WiFi,min}}$ and $T_{\text{RAS,min}}$ in each TBTT respectively, and the ratio of resource amounts per day between WiFi and RAS during the WiFi only and RAS only phases is γ. Then our resource allocation problem becomes designing the time-dependent TBTT structure to maximize the WiFi average throughput as

$$\{T_{\text{WiFi},k}\} = \underset{\{T_{\text{WiFi},k}\}}{\arg\max} \left(\frac{1}{24} \sum_{k=1}^{24} \rho_k^v \right)^{\frac{1}{v}}, \tag{7.7}$$

---

[2]Detail of the calculation of $\tilde{\rho}_k$ is referred to [34].

$$\text{s.t.} \quad \text{C1} : \sum_{k=1}^{24} T_{\text{WiFi},k} \leq T_{\text{WiFi}},$$

$$\text{C2} : T_{\text{WiFi,min}} \leq T_{\text{WiFi},k} \leq T_{\text{WiFi,max}} \qquad \forall k,$$

where υ is the generalized mean exponent $(-\infty < \upsilon \leq 1)$ representing the fairness between different ρ_k (e.g., $\upsilon = 1$ represents the arithmetic mean), $T_{\text{WiFi}} = 24(T_{\text{TBTT}} - T_{\text{Pre-RAS}})\gamma/(1 + \gamma)$ and $T_{\text{WiFi,max}} = T_{\text{TBTT}} - T_{\text{Pre-RAS}} - T_{\text{RAS,min}}$. Note that $T_{\text{RAS},k} = T_{\text{TBTT}} - T_{\text{Pre-RAS}} - T_{\text{WiFi},k}$. Moreover, we should have $\gamma_{\text{min}} \leq \gamma \leq \gamma_{\text{max}}$, where $\gamma_{\text{max}} = T_{\text{WiFi,max}}/T_{\text{RAS,min}}$ and $\gamma_{\text{min}} = T_{\text{WiFi,min}}/(T_{\text{TBTT}} - T_{\text{Pre-RAS}} - T_{\text{WiFi,min}})$. In other words, γ_{max} and γ_{min} are the maximum and minimum achievable ratios of resource amounts respectively, otherwise the optimization problem will be infeasible. The optimization problem in (7.7) is convex since the objective function and the constraints are concave and convex respectively. Therefore, it can be solved by convex programming techniques such as interior point methods [21]. For $\upsilon = 1$ (arithmetic average throughput), it can be easily shown that the optimization problem becomes a linear programming problem. Hence, it can be solved as follows. First, we set $T_{\text{WiFi},k} = T_{\text{WiFi,min}} \, \forall k$. T_{WiFi} is updated as $T_{\text{WiFi}} = T_{\text{WiFi}} - 24T_{\text{WiFi,min}}$. Then, T_{WiFi} is distributed over as many elements of $\{T_{\text{WiFi},k}\}$ according to the descending order of $\{\tilde{\rho}_k\}$ such that $T_{\text{WiFi},k} \leq T_{\text{WiFi,max}} \, \forall k$.

7.5 Performance Evaluation

In this section, we provide numerical results to analyze the proposed coexistence access paradigm between WiFi and RAS. The simulation parameters are listed in Table 7.1. The WiFi parameters are based on IEEE 802.11a standard [54] while the RAS parameters are based on ITU-R RA.769-2. Wireless traffic load statistics within a day is shown in Fig. 7.4 [25] which will be used in our performance evaluation.

Fig. 7.5 shows the received interference power at RAS. As an upper bound for the interference power experienced at the RAS receiver and hence a safer design, I_{out} is set to ∞

Table 7.1. Simulation parameters for 5GHz band

Parameter	Value
WiFi carrier frequency $\left(f_{\mathrm{c}}^{\mathrm{WiFi}}\right)$	5.18 GHz
WiFi bandwidth (W^{WiFi})	20 MHz
Subcarrier spacing (Δf)	312.5 kHz
Index set of used subcarriers $(\mathcal{I})$	$\{-26,\ldots,-1,1,\ldots,26\}$
AP transmit power (P_{AP})	40 dBm
AP antenna gain (G_{WiFi})	6 dB
WiFi cell radius (r)	100m
AP antenna height (h_1)	10m
RAS observing center frequency $\left(f_{\mathrm{c}}^{\mathrm{RAS}}\right)$	4.995 GHz
RAS observing bandwidth (W^{RAS})	10 MHz
RAS receive antenna gain for WiFi (G_{RAS})	-12 dB
RAS antenna height (h_2)	30m
Outer tiers used to define CAZ (I_{out})	∞
RAS interference threshold (I_{th})	-207 dBW
RAS minimum time per TBTT $(T_{\mathrm{min,RAS}})$	10ms
WiFi minimum time per TBTT $(T_{\mathrm{min,WiFi}})$	30ms
WiFi / RAS resource amounts ratio (γ)	3
Generalized mean exponent (υ)	0.5
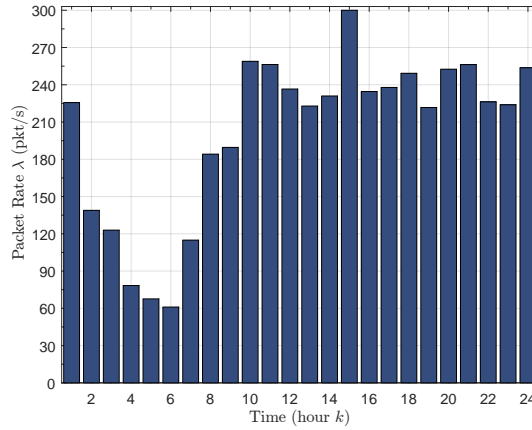
Figure 7.4. Typical average traffic load within a day.

in calculating (7.1). According to Fig. 7.5, the coexistence access zone is defined by 47Km surrounding the RAS receiver. This distance requires a time duration of 162μs out of TBTT (0.162% overhead) to absorb the propagation delays of different WiFi APs within the CAZ. Note that WiFi cells outside the CAZ access the spectrum freely.

As for maximizing the average WiFi throughput, Fig. 7.6 shows the resource allocation during each hour. The time resource amounts allocated to WiFi approximately follow WiFi average traffic load. At the hours with high traffic load, the time resource amounts allocated

Figure 7.5. Received interference power at RAS.

to WiFi are limited by $T_{\text{WiFi,max}}$ to satisfy the RAS minimum observation time $T_{\text{RAS,min}}$. On the other hand, at the hours wit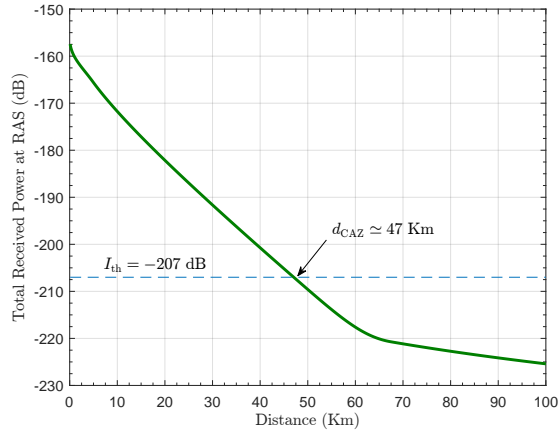h low traffic load, the time resource amounts allocated to WiFi are set to $T_{\text{WiFi,min}}$. Consequently, as shown in Fig. 7.7, WiFi experiences reduced throughput due to the portion of the time allocated to RAS.

The resource allocation in Fig. 7.6 and the WiFi throughputs in Fig. 7.7 can be adjusted through different settings of the minimum time per TBTT for RAS and WiFi and the resource amount ratio between WiFi and RAS. Also note that during the time intervals allocated to WiFi, RAS could do other internal processing which is not connected to the antenna output, for example, measuring noise power from an internal noise source as used in the old VLA of NRAO.

7.6 Conclusion

We have proposed a new paradigm for the coexistence between WiFi and RAS by means of the time-division coexistence access strategy. Coexistence access zone (CAZ) is established around each RAS site. WiFi systems within the zones follow either one of the two proposed time-division-embedded distributed MAC protocols while those outside CAZ have full spectrum access. Average traffic pattern based resource allocation is further developed. The simulation results show that WiFi systems within CAZ experience slight throughput
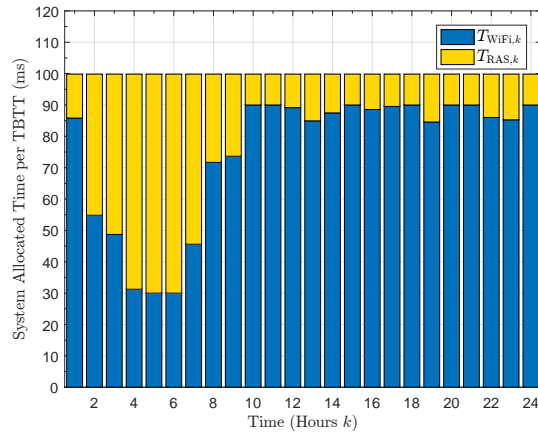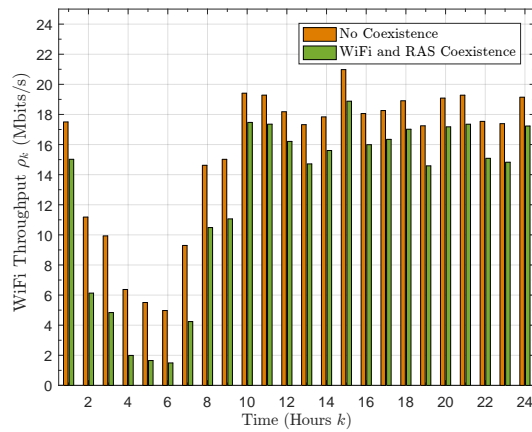
Figure 7.6. Adaptive resource allocation

Figure 7.7. WiFi average throughput comparison between the WiFi only (no coexistence) and the coexistence schemes.

reduction but RAS achieves substantial RFI-free spectrum access which were infeasible in the existing paradigm, thus illustrating high potentials of the proposed paradigm.

CHAPTER 8

CONCLUSION

Millimeter-wave large-scale antenna systems typically apply hybrid analog-digital precoders to reduce hardware complexity and power consumption. This dissertation proposed hybrid precoders for millimeter-wave systems to enhance the physical layer security under two types of channel knowledge. Based on the beamforming strategy only, the proposed hybrid precoders achieve performance close to that of the fully digital precoding at low and moderate SNRs, and sometimes at high SNRs depending on the system parameters. With the aid of artificial noise, the proposed AN-aided hybrid precoder achieves higher secrecy rate. To maximize average network sum-rate, this dissertation developed three hybrid precoders (ZF, SINR-SLNR, and EV-SLNR) based on partial channel knowledge in contrast to the full channel knowledge required in the existing approaches. This part resulted in the following publications:

- Yahia R. Ramadan, Hlaing Minn, and Ahmed S. Ibrahim, "Hybrid Analog-Digital Precoding Design for Secrecy mmWave MISO-OFDM Systems," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5009-5026, Nov. 2017.

- Yahia R. Ramadan and Hlaing Minn, "Artificial Noise Aided Hybrid Precoding Design for Secure mmWave MISO Systems with Partial Channel Knowledge," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1729-1733, Nov. 2017.

- Yahia R. Ramadan and Hlaing Minn, "Novel Hybrid Precoding Designs for mmWave Multiuser Systems with Partial Channel Knowledge," in *IEEE Global Communications Conference (Globecom) 2017*, Singapore, December 2017.

The most promising low-cost THz transmitter architecture in the literature is the so-called frequency-multiplier-last architecture. However, it is incapable of transmitting QAM due to

the architecture's inherent nonlinear distortions. This dissertation developed a precompensation scheme which can suppress nonlinear distortions, mitigate the I/Q imbalance effect, and enable QAM transmission capability. The proposed precompensation scheme offers very attractive output power spectral density control and nonlinear distortion suppression. Using a separate measurement circuitry for testing/calibrating the THz transmitters, we proposed an ML estimator to estimate the NLD parameters and the I/Q imbalance parameters. We derived closed-form expressions for the CRLBs of the system parameters estimates as benchmark metrics to evaluate the performance of the proposed ML estimator. We designed the pilot sequence used in testing/calibrating the THz transmitters to enhance the performance of the proposed ML estimator. We also presented how to trade off between complexity and performance of the proposed scheme by means of the upsampling factor and the construction filter roll-off factor. The effects of frequency-dependent I/Q imbalance and frequency selective NLDs were investigated, and the results showed that the proposed scheme can keep the total distortion within the 3GPP requirement. This part resulted in the following publications:

- Yahia R. Ramadan, Hlaing Minn, and Mahmoud E. Abdelgelil, "Precompensation and System Parameters Estimation for Low-Cost Nonlinear Tera-Hertz Transmitters in the Presence of I/Q Imbalance," accepted in *IEEE Access*, 2018.

- Yahia R. Ramadan, Mahmoud E. Abdelgelil, and Hlaing Minn, "Novel Pre-compensation Schemes for Low-Cost Nonlinear Tera-Hertz Transmitters," in *IEEE International Conference on Communications (ICC) 2018*, Kansas City, May 2018.

In contrast to the existing paradigm of geographical and spectral isolation between CWC and RAS, this dissertation developed a three phase spectrum access which enables geographical and spectral coexistence between CWC and RAS. Shared spectrum access zone (SSAZ) is created around the RAS site and CWC cells within the SSAZ follow the three phase

spectrum access scheme while those outside the SSAZ have full spectrum access. Additionally, system characteristics based improved spectrum sharing is developed. Furthermore, a built-in fine tuning mechanism is presented for addressing mismatches between design and practical environments as well as for facilitating service evolutions. Performance evaluation results demonstrate that the proposed paradigm offers i) certain guaranteed spectrum access to RAS which is impossible in the existing paradigm, ii) capability to handle higher peak and mean traffics to CWC under spectrum restructuring of both CWC and RAS bands, and iii) overall improved spectrum utilization. Furthermore, a shared spectrum access strategy for RAS and WiFi systems was also developed by modifying the distributed medium access protocol. This part resulted in the following publications:

- Yahia R. Ramadan, Hlaing Minn, and Yucheng Dai, "A New Paradigm for Spectrum Sharing between Cellular Wireless Communications and Radio Astronomy Systems," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3985-3999, Sept. 2017.

- Hlaing Minn, Yahia R. Ramadan, and Yucheng Dai, "A New Shared Spectrum Access Paradigm between Cellular Wireless Communications and Radio Astronomy," in *IEEE Global Communications Conference (Globecom) 2016*, Washington DC, USA, December 2016.

- Yahia R. Ramadan, Yucheng Dai, Hlaing Minn, and Fabiano S. Rodrigues "Spectrum Sharing Between WiFi and Radio Astronomy," in *RFI 2016 Conference*, Socorro, NM, USA, October 2016.

BIBLIOGRAPHY

[1] Akdeniz, M. R., Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip (2014, Jun.). Millimeter wave channel modeling and cellular capacity evaluation. *IEEE J. Sel. Areas Commun. 32*(6), 1164–1179.

[2] Akyildiz, I. F., J. M. Jornet, and C. Han (2014). Terahertz band: Next frontier for wireless communications. *Physical Communication 12*, 16–32.

[3] Akyildiz, I. F., W.-Y. Lee, M. C. Vuran, and S. Mohanty (2006). Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Comput. Netw. 50*(13), 2127 – 2159.

[4] Alagha, N. S. and P. Kabal (1999, Jul.). Generalized raised-cosine filters. *IEEE Trans. on Commun. 47*(7), 989–997.

[5] Alkhateeb, A., O. El Ayach, G. Leus, and R. Heath (2013, February). Hybrid precoding for millimeter wave cellular systems with partial channel knowledge. In *Information Theory and Applications Workshop (ITA), 2013*, pp. 1–5.

[6] Alkhateeb, A. and R. W. Heath (2016, May). Frequency selective hybrid precoding for limited feedback millimeter wave systems. *IEEE Transactions on Communications 64*(5), 1801–1818.

[7] Alkhateeb, A., R. W. Heath, and G. Leus (2015, Jun.). Achievable rates of multi-user millimeter wave systems with hybrid precoding. In *Proc. IEEE ICCW 2015*, pp. 1232–1237.

[8] Alkhateeb, A., G. Leus, and R. W. Heath (2014, Nov.). Multi-layer precoding for full-dimensional massive MIMO systems. In *Proc. IEEE Asilomar Conf. Signal, Syst. Comput. 2014*, pp. 815–819.

[9] Alkhateeb, A., G. Leus, and R. W. Heath (2015, Nov.). Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE Trans. Wireless Commun. 14*(11), 6481–6494.

[10] Anttila, L., P. Handel, and M. Valkama (2010, Apr.). Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters. *IEEE Trans. Microw. Theory Techn. 58*(4), 730–739.

[11] Asplund, H., J. Medbo, and J. E. Berg (2011, Sep.). Measurements of beyond horizon propagation loss. In *IEEE-APS APWC 2011*, pp. 159–162.

[12] Ayach, O., R. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi (2012, June). Low complexity precoding for large millimeter wave MIMO systems. In *Communications (ICC), 2012 IEEE International Conference on*, pp. 3724–3729.

[13] Ayodele, P. and F. Olabisi (2015, June). Interference protection of radio astronomy services using cognitive radio spectrum sharing models. In *European Conf. Networks and Communications (EuCNC)*, pp. 86–90.

[14] Baldi, M., F. Chiaraluce, N. Laurenti, S. Tomasin, and F. Renna (2014, November). Secrecy transmission on parallel channels: Theoretical limits and performance of practical codes. *IEEE Transactions on Information Forensics and Security 9*(11), 1765–1779.

[15] Barnbaum, C. and R. F. Bradley (1998). A new approach to interference excision in radio astronomy: real-time adaptive cancellation. *The Astronomical Journal 116*(5), 2598.

[16] Barros, J. and M. Rodrigues (2006, July). Secrecy capacity of wireless channels. In *Information Theory, 2006 IEEE International Symposium on*, pp. 356–360.

[17] Bentum, M., A. Boonstra, and W. Baan (2009). The coexistence of cognitive radio and radio astronomy. In *16th Annual Symp. IEEE/CVT 2009*, Louvain-la-Neuve, Belgium.

[18] Bentum, M., A. Boonstra, and W. Baan (2010, Mar.). Impact of cognitive radio on radio astronomy. In *Proceedings of the RFI Mitigation Workshop, RFI 2010*. Proc. Science (POS).

[19] Bogale, T. E., L. B. Le, A. Haghighat, and L. Vandendorpe (2016, May). On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming. *IEEE Transactions on Wireless Communications 15*(5), 3311–3326.

[20] Boonstra, A.-J. *Radio frequency interference mitigation in radio astronomy*. Ph.D. dissertation, Technical Univ. Delft, ASTRON, Delft, The Netherlands, 2005.

[21] Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge, UK: Cambridge University Press.

[22] Braithwaite, R. N. (2017, Oct.). Digital predistortion of an RF power amplifier using a reduced volterra series model with a memory polynomial estimator. *IEEE Trans. Microw. Theory Techn. 65*(10), 3613–3623.

[23] Briggs, F., J. Bell, and M. Kesteven (2000). Removing radio interference from contaminated astronomical spectra using an independent reference signal and closure relations. *The Astronomical Journal 120*(6), 3351.

[24] Briggs, F. and J. Kocz (2005). Overview of technical approaches to radio frequency interference mitigation. *Radio Science 40*(5).

[25] Chen, X., Y. Jin, S. Qiang, W. Hu, and K. Jiang (2015). Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale. In *IEEE Intl. Conf. Commun. (ICC)*.

[26] Chowdhury, K. and I. Akyildiz (2008, Jan.). Cognitive wireless mesh networks with dynamic spectrum access. *IEEE J. Sel. Areas Commun. 26*(1), 168 –181.

[27] Dai, M. and B. Clerckx (2017). Multiuser millimeter wave beamforming strategies with quantized and statistical CSIT. *IEEE Trans. Wireless Commun.*. doi:`10.1109/TWC.2017.2737009`.

[28] Daly, M. and J. Bernhard (2009, September). Directional modulation technique for phased arrays. *Antennas and Propagation, IEEE Transactions on 57*(9), 2633–2640.

[29] Daly, M., E. Daly, and J. Bernhard (2010, May). Demonstration of directional modulation using a phased array. *Antennas and Propagation, IEEE Transactions on 58*(5), 1545–1550.

[30] Ding, F. (2013, Jan.). Coupled-least-squares identification for multivariable systems. *IET Control Theory Applications 7*(1), 68–79.

[31] Ding, F., G. Liu, and X. P. Liu (2010, Aug.). Partially coupled stochastic gradient identification methods for non-uniformly sampled systems. *IEEE Trans. Autom. Control 55*(8), 1976–1981.

[32] Ding, F., X. Liu, and Y. Gu (2016). An auxiliary model based least squares algorithm for a dual-rate state space system with time-delay using the data filtering. *Journal of the Franklin Institute 353*(2), 398–408.

[33] Ding, F., F. Wang, L. Xu, T. Hayat, and A. Alsaedi (2017). Parameter estimation for pseudo-linear systems using the auxiliary model and the decomposition technique. *IET Control Theory Applications 11*(3), 390–400.

[34] Duffy, K., D. Malone, and D. J. Leith (2005, Aug.). Modeling the 802.11 distributed coordination function in non-saturated conditions. *IEEE Commun. Lett. 9*(8), 715–717.

[35] Ellingson, S. W. (2004). RFI mitigation and the SKA. *Experimental Astronomy 17*(1-3), 261–267.

[36] Eltayeb, M., T. A.-N. J. Choi, and R. W. Heath (2017). Enhancing secrecy with multi-antenna transmission in millimeter wave vehicular communication systems. *IEEE Trans. Veh. Technol.*. doi:`10.1109/TVT.2017.2681965`.

[37] ETSI (2012, Nov.). TS 136 104 v9. 13.0, LTE; evolved universal terrestrial radio access (E-UTRA); base station (BS) radio transmission and reception (3GPP TS 36.104 version 9.13. 0 release 9).

[38] FCC (2014, Apr.). Further notice of proposed rulemaking. Technical Report FCC 14-49.

[39] Ford, J. and K. Buch (2014, July). RFI mitigation techniques in radio astronomy. In *IEEE Intl. Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 231–234.

[40] Forestier, S., P. Bouysse, R. Quere, A. Mallet, J. M. Nebus, and L. Lapierre (2004, Apr.). Joint optimization of the power-added efficiency and the error-vector measurement of 20-GHz phemt amplifier through a new dynamic bias-control method. *IEEE Trans. Microw. Theory Techn. 52*(4), 1132–1141.

[41] Gao, L., X. Wang, Y. Xu, and Q. Zhang (2011, Apr.). Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach. *IEEE J. Sel. Areas Commun. 29*(4), 843 –855.

[42] Gerbracht, S., C. Scheunert, and E. A. Jorswieck (2012, April). Secrecy outage in MISO systems with partial channel information. *IEEE Transactions on Information Forensics and Security 7*(2), 704–716.

[43] Gerbracht, S., A. Wolf, and E. A. Jorswieck (2010, Feb). Beamforming for fading wiretap channels with partial channel information. In *2010 International ITG Workshop on Smart Antennas (WSA)*, pp. 394–401.

[44] Gergely, T. (2014, Mar.). Spectrum access for the passive services: The past and the future (invited paper). *IEEE Proc. 102*(3), 393–398.

[45] Ghasemi, A. and E. Sousa (2008). Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. *IEEE Commun. Mag. 46*(4), 32–39.

[46] Gilabert, P., G. Montoro, and E. Bertran (2005, Dec). On the Wiener and Hammerstein models for power amplifier predistortion. In *Proc. Asia-Pacific Microwave Conf.*, Volume 2, pp. 1–4.

[47] Goldsmith, A. (2005). *Wireless Communications*. Cambridge University Press.

[48] Gong, S., C. Xing, Z. Fei, and S. Ma (2016, June). Millimeter-wave secrecy beamforming designs for two-way amplify-and-forward MIMO relaying networks. *IEEE Transactions on Vehicular Technology*. doi:`10.1109/TVT.2016.2578943`.

[49] Grant, M. and S. Boyd (2008). Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura (Eds.), *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited.

[50] Grant, M. and S. Boyd (2014). CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`.

[51] Guner, B., N. Niamsuwan, and J. Johnson (2010, July). Performance study of a cross-frequency detection algorithm for pulsed sinusoidal RFI in microwave radiometry. *IEEE Trans. Geosci. Remote Sens. 48*(7), 2899–2908.

[52] He, Y., E. P. J. Parrott, and E. Pickwell-MacPherson (2017, Mar.). Adaptive sampling for terahertz time-domain spectroscopy and imaging. *IEEE Trans. THz Sci. Technol. 7*(2), 118–123.

[53] Huang, S., X. Liu, and Z. Ding (2011, Feb.). Decentralized cognitive radio control based on inference from primary link control information. *IEEE J. Sel. Areas Commun. 29*(2), 394 –406.

[54] IEEE (1999, Dec.). IEEE standard for telecommunications and information exchange between systems - LAN/MAN specific requirements - part 11: Wireless medium access control (MAC) and physical layer (PHY) specifications: High speed physical layer in the 5 GHz band. *IEEE Std 802.11a-1999*, 1–102.

[55] ITU-R (2003a, Jun.). Preferred frequency bands for radio astronomical measurements. Recommendation RA.314-10, ITU.

[56] ITU-R (2003b, Mar.). Protection creteria used for radio astronomical measurements. Recommendation RA.769-2, ITU.

[57] ITU-R (2007, Jun.). Protection of the radio astronomy service in frequency bands shared with other services. Recommendation RA.1031-2, ITU.

[58] ITU-R (2013a, Sep.). Attenuation by atmospheric gases. Recommendation P.676-10, ITU.

[59] ITU-R (2013b, Dec.). Space research earth station and radio astronomy reference antenna radiation pattern for use in interference calculations, including coordination procedures, for frequencies less than 30 GHz. Technical Report SA.509-3, ITU.

[60] ITU-R (2013c, Nov.). Techniques for mitigation of radio frequency interference in radio astronomy. Technical Report RA.2126-1, ITU.

[61] ITU-R (2015, Mar.). Levels of data loss to radio astronomy observations and percentage-of-time criteria resulting from degradation by interference for frequency bands allocated to the radio astronomy service on a primary basis. Recommendation RA.1513-2, ITU.

[62] Jeffs, B. D., L. Li, and K. F. Warnick (2005). Auxiliary antenna-assisted interference mitigation for radio astronomy arrays. *IEEE Trans. Signal Process. 53*(2), 439–451.

[63] Jo, H.-S. (2015). A multi-antenna based spectrum sharing for IMT-Advanced SDMA systems. *Intl. J. Multimedia and Ubiquitous Eng. 10*(1), 265–276.

[64] Jo, H. S., H. G. Yoon, J. Lim, W. G. Chung, J. G. Yook, and H. K. Park (2006, Jun.). The coexistence of ofdm-based systems beyond 3g with fixed service microwave systems. *Journal of Commun. and Netw. 8*(2), 187–193.

[65] Joost, M. (2010). Theory of root-raised cosine filter. *Research and Development 47829*.

[66] Ju, Y., H. M. Wang, T. X. Zheng, and Q. Yin (2016, April). Secure transmission with artificial noise in millimeter wave systems. In *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–6.

[67] Ju, Y., H. M. Wang, T. X. Zheng, and Q. Yin (2017, May). Secure transmissions in millimeter wave systems. *IEEE Transactions on Communications 65*(5), 2114–2127.

[68] Kang, S., S. V. Thyagarajan, and A. M. Niknejad (2014, Jun.). A 240GHz wideband QPSK transmitter in 65nm CMOS. In *Proc. IEEE Radio Frequency Integrated Circuits Symposium (RFIC) 2014*, pp. 353–356.

[69] Kang, S., S. V. Thyagarajan, and A. M. Niknejad (2015, Oct.). A 240 GHz fully integrated wideband QPSK transmitter in 65 nm CMOS. *IEEE Journal of Solid-State Circuits 50*(10), 2256–2267.

[70] Katayama, K., K. Takano, S. Amakawa, S. Hara, A. Kasamatsu, K. Mizuno, K. Takahashi, T. Yoshida, and M. Fujishima (2016a, Jan.). 20.1 a 300GHz 40nm CMOS transmitter with 32-QAM 17.5Gb/s/ch capability over 6 channels. In *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 342–343.

[71] Katayama, K., K. Takano, S. Amakawa, S. Hara, A. Kasamatsu, K. Mizuno, K. Takahashi, T. Yoshida, and M. Fujishima (2016b, Dec.). A 300 GHz CMOS transmitter with 32-QAM 17.5 Gb/s/ch capability over six channels. *IEEE J. Solid-State Circuits 51*(12), 3037–3048.

[72] Katzenstein, W. E., R. P. Moore, and H. G. Kimball (1981, Aug.). Spectrum allocations above 40 GHz. *IEEE Trans. Commun. COM-29*(8), 1136–1141.

[73] Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

[74] Khan, Z. A., E. Zenteno, P. Händel, and M. Isaksson (2017, Jan.). Digital predistortion for joint mitigation of I/Q imbalance and MIMO power amplifier distortion. *IEEE Trans. Microw. Theory Techn. 65*(1), 322–333.

[75] Khisti, A. and G. W. Wornell (2010, July). Secure transmission with multiple antennas I: The MISOME wiretap channel. *IEEE Transactions on Information Theory 56*(7), 3088–3104.

[76] Li, Q. and W. K. Ma (2011a, Augest). Optimal and robust transmit designs for MISO channel secrecy by semidefinite programming. *IEEE Transactions on Signal Processing 59*(8), 3799–3812.

[77] Li, Q. and W. K. Ma (2011b, Aug.). Optimal and robust transmit designs for MISO channel secrecy by semidefinite programming. *IEEE Trans. Signal Process. 59*(8), 3799–3812.

[78] Li, Q. and W. K. Ma (2013, May). Spatially selective artificial-noise aided transmit optimization for MISO multi-eves secrecy rate maximization. *IEEE Transactions on Signal Processing 61*(10), 2704–2717.

[79] Li, Z., S. Han, and A. F. Molisch (2017, Jul.). Optimizing channel-statistics-based analog beamforming for millimeter-wave multi-user massive mimo downlink. *IEEE Trans. Wireless Commun. 16*(7), 4288–4303.

[80] Liang, L., Y. Dai, W. Xu, and X. Dong (2014, Oct.). How to approach zero-forcing under RF chain limitations in large mmwave multiuser systems? In *IEEE/CIC ICCC 2014*, pp. 518–522.

[81] Liu, C.-L. (1998, Aug.). Impacts of I/Q imbalance on QPSK-OFDM-QAM detection. *IEEE Trans. Consum. Electron. 44*(3), 984–989.

[82] Lustgarten, M. N. and J. A. Madison (1977, Aug.). An empirical propagation model (epm-73). *IEEE Trans. Electromagn. Compat EMC-19*(3), 301–309.

[83] Magnus, J. (1986). The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Economie et de Statistique 4*, 95–109.

[84] Meng, J., W. Yin, H. Li, E. Hossain, and Z. Han (2011, Feb.). Collaborative spectrum sensing from sparse observations in cognitive radio networks. *IEEE J. Sel. Areas Commun. 29*(2), 327 –337.

[85] Minn, H., V. K. Bhargava, and K. B. Letaief (2003, Jul.). A robust timing and frequency synchronization for OFDM systems. *IEEE Trans. Wireless Commun. 2*(4), 822–839.

[86] Misra, S. and C. Ruf (2012, May). Analysis of radio frequency interference detection algorithms in the angular domain for SMOS. *IEEE Trans. Geosci. Remote Sens. 50*(5), 1448–1457.

[87] MOSEK ApS (2015). *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*

[88] Nahata, A., A. S. Weling, and T. F. Heinz (1996). A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling. *Applied Physics Letters 69*(16), 2321–2323.

[89] Narasimhan, B., D. Wang, S. Narayanan, H. Minn, and N. Al-Dhahir (2009, June). Digital compensation of frequency-dependent joint Tx/Rx I/Q imbalance in OFDM systems under high mobility. *IEEE J. Sel. Topics Signal Process. 3*(3), 405–417.

[90] National Academies of Sciences, Engineering, and Medicine (2007). *Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses.* NRC of the National Academies, The National Academies Press.

[91] National Research Council (2010). *Spectrum Management for Science in the 21st Century.* Committee on Scientific Use of the Radio Spectrum, Committee on Radio Frequencies, and National Research Council, The National Academies Press.

[92] Negi, R. and S. Goel (2005, September). Secret communication using artificial noise. In *Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd*, Volume 3, pp. 1906–1910.

[93] Nguyen, D. H. N., L. B. Le, and T. Le-Ngoc (2016, May). Hybrid MMSE precoding for mmWave multiuser MIMO systems. In *IEEE ICC 2016*, pp. 1–6.

[94] Ni, W., X. Dong, and W. Lu (2015). Near-optimal hybrid processing for massive MIMO systems via matrix decomposition.

[95] Njoku, E., P. Ashcroft, T. Chan, and L. Li (2005, May). Global survey and statistics of radio-frequency interference in AMSR-E land observations. *IEEE Trans. Geosci. Remote Sens. 43*(5), 938–947.

[96] Nsenga, J., A. Bourdoux, and F. Horlin (2010, May). Mixed analog/digital beamforming for 60 GHz MIMO frequency selective channels. In *Communications (ICC), 2010 IEEE International Conference on*, pp. 1–6.

[97] Pankonin, V. and R. Proce (1981, Aug.). Radio astronomy and spectrum management: The impact of WARC-79. *IEEE Trans. Commun. COM-29*(8), 1228–1237.

[98] Park, J. D., S. Kang, S. V. Thyagarajan, E. Alon, and A. M. Niknejad (2012, Jun.). A 260 GHz fully integrated CMOS transceiver for wireless chip-to-chip communication. In *Proc. Symposium on VLSI Circuits (VLSIC)*, pp. 48–49.

[99] Park, S., A. Alkhateeb, and R. W. H. Jr. (2016). Dynamic subarrays for hybrid precoding in wideband mmwave MIMO systems.

[100] Park, S., J. Park, A. Yazdan, and R. W. Heath (2017, July). Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems. *IEEE Trans. Signal Process. 65*(14), 3818–3832.

[101] Petersen, K. and M. Pedersen (2007, September). *The Matrix Cookbook*.

[102] Petrov, V., A. Pyattaev, D. Moltchanov, and Y. Koucheryavy (2016, Oct.). Terahertz band communications: Applications, research challenges, and standardization activities. In *Proc. IEEE International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) 2016*, pp. 183–190.

[103] Pritzker, P. and L. E. Strickling (2014, June). Fourth interim progress report on the ten-year plan and timetable and plan for quantitative assessments of spectrum usage. Technical report, U.S. Department of Commerce.

[104] Qiu, W., B. Xie, H. Minn, and C.-C. Chong (2012, Jan.). Interference-controlled transmission schemes for cognitive radio in frequency-selective time-varying fading channels. *IEEE Trans. Wireless Commun. 11*(1), 142–153.

[105] Ramadan, Y., A. Ibrahim, and M. Khairy (2015, March). Minimum outage RF beamforming for millimeter wave MISO-OFDM systems. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pp. 557–561.

[106] Ramadan, Y. R., M. E. Abdelgelil, and H. Minn (2018, May). Novel pre-compensation schemes for low-cost nonlinear Tera-Hertz transmitters. accepted in *IEEE International Conference on Communications (ICC) 2018*.

[107] Ramadan, Y. R., A. S. Ibrahim, and M. M. Khairy (2016, May). RF beamforming for secrecy millimeter wave MISO-OFDM systems. In *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6.

[108] Ramadan, Y. R., H. Minn, and A. S. Ibrahim (2017). Hybrid analog-digital precoding design for secrecy mmWave MISO-OFDM systems. *IEEE Trans. Commun.*. doi:`10.1109/TCOMM.2017.2734666`.

[109] Rappaport, T., S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez (2013). Millimeter wave mobile communications for 5G cellular: It will work! *Access, IEEE 1*, 335–349.

[110] Rothbart, N., K. Schmalz, J. Borngraber, D. Kissinger, and H. W. Hubers (2016, Dec.). Gas spectroscopy by voltage-frequency tuning of a 245 GHz SiGe transmitter and receiver. *IEEE Sensors J. 16*(24), 8863–8864.

[111] Rupasinghe, N. and İ. Güvenç (2015, Mar.). Reinforcement learning for licensed-assisted access of lte in the unlicensed spectrum. In *IEEE WCNC 2015*, pp. 1279–1284.

[112] Samimi, M. K., T. S. Rappaport, and G. R. MacCartney (2015). Probabilistic omnidirectional path loss models for millimeter-wave outdoor communications. *IEEE Wireless Commun. Lett. 4*(4), 357–360.

[113] Sayeed, A. (2006). Sparse multipath wireless channels: Modeling and implications. In *Proceedings of ASAP*.

[114] Schmalz, K., J. Borngraber, P. Neumaier, N. Rothbart, D. Kissinger, and H. W. Hubers (2016, Jun.). Gas spectroscopy system at 245 and 500 GHz using transmitters and receivers in SiGe BiCMOS. In *Proc. Global Symposium on Millimeter Waves (GSMM) ESA Workshop on Millimetre-Wave Technology and Applications 2016*, pp. 1–4.

[115] Schmalz, K., N. Rothbart, P. F. X. Neumaier, J. Borngraber, H. W. Hubers, and D. Kissinger (2017, May). Gas spectroscopy system for breath analysis at mm-wave/THz using SiGe BiCMOS circuits. *IEEE Trans. Microw. Theory Techn. 65*(5), 1807–1818.

[116] Shafiee, S. and S. Ulukus (2007, June). Achievable rates in Gaussian MISO channels with secrecy constraints. In *2007 IEEE International Symposium on Information Theory*, pp. 2466–2470.

[117] Sharma, N., J. Zhang, Q. Zhong, W. Choi, J. P. McMillan, C. F. Neese, F. C. D. Lucia, and K. O. Kenneth (2014, Sept.). 85-to-127 GHz CMOS transmitter for rotational spectroscopy. In *Proc. IEEE Custom Integrated Circuits Conference 2014*, pp. 1–4.

[118] Sharma, N., Q. Zhong, Z. Chen, W. Choi, J. P. McMillan, C. F. Neese, R. Schueler, I. Medvedev, F. D. Lucia, and K. O (2016, Jun.). 200-280 GHz CMOS RF front-end of transmitter for rotational spectroscopy. In *Proc. IEEE Symposium on VLSI Technology 2016*, pp. 1–2.

[119] Siegel, P. H. (2002, Mar). Terahertz technology. *IEEE Trans. Microw. Theory Techn. 50*(3), 910–928.

[120] Stark, H. and J. Woods (2012). *Probability, statistics, and random processes for engineers.* Prentice Hall.

[121] Stevenson, C. R., G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell (2009, Jan.). IEEE 802.22: The first cognitive radio wireless regional area network standard. *IEEE Commun. Mag. 47*(1), 130 –138.

[122] Stewart, W. J. (2009). *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling.* Princeton University Press.

[123] Sun, Q. and S. Nan (2012, Oct). Coexistence studies between LTE-hotspot indoor and earth station of fixed satellite service in the band 3400-3600 MHz. In *IEEE ICSP 2012*, Volume 3, pp. 2275–2278.

[124] Takano, K., S. Amakawa, K. Katayama, S. Hara, R. Dong, A. Kasamatsu, I. Hosako, K. Mizuno, K. Takahashi, T. Yoshida, and M. Fujishima (2017, Feb.). 17.9 a 105Gb/s 300GHz CMOS transmitter. In *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 308–309.

[125] Tang, A., B. Drouin, Y. Kim, G. Virbila, and M. C. F. Chang (2017, May). 95-105 GHz 352 mW all-silicon cavity-coupled pulsed echo rotational spectroscopy system in 65 nm CMOS. *IEEE Trans. THz Sci. Technol. 7*(3), 244–249.

[126] Thompson, A. R. (2004). ITU-R recommendations of particular importance to radio astronomy. In *Spectrum Management for Radio Astronomy*, Volume 1, pp. 121.

[127] Thyagarajan, S. V., S. Kang, and A. M. Niknejad (2015, Oct.). A 240 GHz fully integrated wideband qpsk receiver in 65 nm CMOS. *IEEE J. Solid-State Circuits 50*(10), 2268–2280.

[128] Tse, D. and P. Viswanath (2005). *Fundamentals of wireless communication.* Cambridge university press.

[129] Tubbax, J., B. Come, L. V. der Perre, L. Deneire, S. Donnay, and M. Engels (2003, May). Compensation of IQ imbalance in OFDM systems. In *Proc. IEEE International Conference on Communications (ICC) 2003*, Volume 5, pp. 3403–3407.

[130] Valliappan, N., A. Lozano, and R. W. Heath (2013, August). Antenna subset modulation for secure millimeter-wave wireless communication. *IEEE Transactions on Communications 61*(8), 3231–3245.

[131] van Driel, W. (2009). Radio quiet, please!- protecting radio astronomy from interference. In *IAU Symposium: The Role of Astronomy in Society and Culture*, Volume 260.

[132] Venmani, D. P., O. L. Moult, F. Deletre, Y. Lagadec, and Y. Morlon (2016, Sep.). On the role of network synchronization for future cellular networks: An operator's perspective. *IEEE Commun. Mag. 54*(9), 58–64.

[133] Vuppala, S., S. Biswas, and T. Ratnarajah (2016, Aug.). An analysis on secure communication in millimeter/micro-wave hybrid networks. *IEEE Trans. Commun. 64*(8), 3507–3519.

[134] Wallace, J. W. and M. A. Jensen (2006, Nov). Time-varying MIMO channels: Measurement, analysis, and modeling. *IEEE Transactions on Antennas and Propagation 54*(11), 3265–3273.

[135] Wang, C. and H. M. Wang (2016, Aug.). Physical layer security in millimeter wave cellular networks. *IEEE Trans. on Wireless Commun. 15*(8), 5569–5585.

[136] Wang, J., P. Ding, M. Zoltowski, and D. Love (2005, December). Space-time coding and beamforming with partial channel state information. In *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, Volume 5, pp. 3149–3153.

[137] Wang, L., M. Elkashlan, T. Q. Duong, and R. W. Heath (2014, June). Secure communication in cellular networks: The benefits of millimeter wave mobile broadband. In *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 115–119.

[138] Wang, W., K. C. Teh, and K. H. Li (2017, March). Secrecy throughput maximization for MISO multi-eavesdropper wiretap channels. *IEEE Transactions on Information Forensics and Security 12*(3), 505–515.

[139] Wang, W., F. Zhou, W. Huang, B. Wang, and Y. Zhang (2010, Oct). Coexistence studies between LTE system and earth station of fixed satellite service in the 3400-3600 MHz frequency bands in China. In *IEEE IC-BNMT*, pp. 1125–1130.

[140] Weiss, T. and F. Jondral (2004, Mar.). Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency. *IEEE Commun. Mag. 42*(3), S8–14.

[141] Xu, H., V. Kukshya, and T. S. Rappaport (2002, April). Spatial and temporal characteristics of 60-GHz indoor channels. *IEEE Journal on Selected Areas in Communications 20*(3), 620–630.

[142] Xu, Z., S. Han, Z. Pan, and C. L. I (2015, June). Alternating beamforming methods for hybrid analog and digital MIMO transmission. In *2015 IEEE International Conference on Communications (ICC)*, pp. 1595–1600.

[143] Yan, S., G. Geraci, N. Yang, R. Malaney, and J. Yuan (2014, June). On the target secrecy rate for SISOME wiretap channels. In *2014 IEEE International Conference on Communications (ICC)*, pp. 987–992.

[144] Yang, N., L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo (2015, April). Safeguarding 5G wireless communication networks using physical layer security. *IEEE Communications Magazine 53*(4), 20–27.

[145] Yong, S.-K., P. Xia, and A. Valdes-Garcia (2010). *60 GHz technology for Gbps WLAN and WPAN: From theory to practice.* Wiley.

[146] Yu, X., J. C. Shen, J. Zhang, and K. B. Letaief (2016, April). Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems. *IEEE Journal of Selected Topics in Signal Processing 10*(3), 485–500.

[147] Yucek, T. and H. Arslan (2009, Mar.). A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE, Communications Surveys Tutorials 11*(1), 116–130.

[148] Zhan, P., K. Qin, and S. Cai (2011, Dec.). Joint compensation model for memory power amplifier and frequency-dependent nonlinear IQ impairments. *Electronics Letters 47*(25), 1382–1384.

[149] Zhang, X., A. F. Molisch, and S.-Y. Kung (2005, November). Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection. *IEEE Transactions on Signal Processing 53*(11), 4091–4103.

[150] Zhao, G., G. Li, and C. Yang (2009, Sep.). Proactive detection of spectrum opportunities in primary systems with power control. *IEEE Trans. Wireless Commun. 8*(9), 4815 –4823.

[151] Zhong, Q., W. Choi, N. Sharma, Z. Ahmad, J. P. McMillan, C. F. Neese, F. C. D. Lucia, and K. K. O (2016, May). 225-280 GHz receiver for rotational spectroscopy. In *Proc. IEEE Radio Frequency Integrated Circuits Symposium (RFIC) 2016*, pp. 298–301.

[152] Zhou, X., L. Song, and Y. Zhang (2013). *Physical Layer Security in Wireless Communications*. CRC Press.

[153] Zhu, J., W. Xu, and N. Wang (2016, October). Secure massive MIMO systems with limited RF chains. *IEEE Transactions on Vehicular Technology*. doi:`10.1109/TVT.2016.2615885`.

[154] Zhu, Y., L. Wang, K. K. Wong, and R. W. Heath (2017, May). Secure communications in millimeter wave Ad Hoc networks. *IEEE Transactions on Wireless Communications 16*(5), 3205–3217.

BIOGRAPHICAL SKETCH

Yahia Ramadan received the B.S. (with highest honors) and M.S. degrees in Electronics and Communications Engineering from Cairo University, Cairo, Egypt, in 2011 and 2015, respectively. He received his Ph.D. degree in Electrical Engineering from The University of Texas at Dallas, Richardson, TX, USA in 2018. His research work is in the areas of millimeter-wave communications, Tera-Hertz communications, device-to-device communications, massive MIMO, physical layer security, nonlinear distortion and RF impairments compensations, estimation theory, and spectrum access and sharing.

He was a 5G RF standrads intern at Intel Corporation, Hillsboro, OR, USA, from June 2018 to August 2018. He was a research/teaching assistant at The University of Texas at Dallas, Richardson, TX, USA, from 2016 to 2018, and a research/teaching assistant at Cairo University, Cairo, Egypt, from 2012 to 2015. He received the Mary and Richard Templeton Graduate Fellowship in 2017, and Award of Excellence from the Faculty of Engineering, Cairo University, Egypt from 2006 to 2011.

Yahia Ramadan Ahmed Mohamed Ramadan

August 2018

Contact Information:

Department of Electrical Engineering
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021, U.S.A.

Voice: +1 (213) 709-7479
Email: yahia.ramadan@utdallas.edu
      yahia.r.ramadan@ieee.org

Educational History:

B.S., Electronics and Communications Engineering, Cairo University, Cairo, Egypt, 2011
M.S., Electronics and Communications Engineering, Cairo University, Cairo, Egypt, 2015
Ph.D., Electrical Engineering, The University of Texas at Dallas, Richardson, TX, USA, 2018

Employment History:

5G RF standrads intern, Intel Corporation, June 2018 – August 2018
Research/Teaching Assistant, University of Texas at Dallas, TX, USA, January 2016 – May 2018
Research/Teaching Assistant, Cairo University, Cairo, Egypt, December 2012 – August 2015

Professional Recognitions and Honors:

Mary and Richard Templeton Graduate Fellowship, The University of Texas at Dallas, TX, USA, 2017
Award of Excellence from the Faculty of Engineering, Cairo University, Cairo, Egypt, 2006 – 2011

Professional Memberships:

Institute of Electrical and Electronics Engineers (IEEE), 2014–present
IEEE Communications Society (IEEE ComSoc), 2014–present

PUBLICATIONS

Journals

1. **Yahia R. Ramadan**, Hlaing Minn, and Mahmoud E. Abdelgelil, "Precompensation and System Parameters Estimation for Low-Cost Nonlinear Tera-Hertz Transmitters in the Presence of I/Q Imbalance," accepted in *IEEE Access*, 2018.

2. **Yahia R. Ramadan** and Hlaing Minn, "Artificial Noise Aided Hybrid Precoding Design for Secure mmWave MISO Systems with Partial Channel Knowledge," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1729-1733, Nov. 2017.

3. **Yahia R. Ramadan**, Hlaing Minn, and Ahmed S. Ibrahim, "Hybrid Analog-Digital Precoding Design for Secrecy mmWave MISO-OFDM Systems," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5009-5026, Nov. 2017.

4. **Yahia R. Ramadan**, Hlaing Minn, and Yucheng Dai, "A New Paradigm for Spectrum Sharing between Cellular Wireless Communications and Radio Astronomy Systems," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3985-3999, Sept. 2017.

5. **Yahia R. Ramadan**, Ahmed S. Ibrahim, and Mohamed M. Khairy, "Robust RF Beamforming for Millimeter Wave MIMO-OFDM Systems," in *International Journal of Wireless Information Networks: Special Issue on Toward 5G Emerging Technologies*, September 2015.

Conferences

1. **Yahia R. Ramadan**, Mahmoud E. Abdelgelil, and Hlaing Minn, "Novel Pre-compensation Schemes for Low-Cost Nonlinear Tera-Hertz Transmitters," in *IEEE International Conference on Communications (ICC) 2018*, Kansas City, May 2018.

2. Yucheng Dai, **Yahia R. Ramadan**, Hlaing Minn, Jiu Xiong, Jin Liu, and Alan Gatherer "A New Receiver Architecture for MIMO Beam-Forming Applications," in *IEEE Global Communications Conference (Globecom) 2017*, Singapore, December 2017.

3. **Yahia R. Ramadan** and Hlaing Minn, "Novel Hybrid Precoding Designs for mmWave Multiuser Systems with Partial Channel Knowledge," in *IEEE Global Communications Conference (Globecom) 2017*, Singapore, December 2017.

4. **Yahia R. Ramadan**, Yucheng Dai, Hlaing Minn, and Fabiano S. Rodrigues "Spectrum Sharing Between WiFi and Radio Astronomy," in *RFI 2016 Conference*, Socorro, NM, USA, October 2016.

5. Hlaing Minn, **Yahia R. Ramadan**, and Yucheng Dai, "A New Shared Spectrum Access Paradigm between Cellular Wireless Communications and Radio Astronomy," in *IEEE Global Communications Conference (Globecom) 2016*, Washington DC, USA, December 2016.

6. **Yahia R. Ramadan**, Ahmed S. Ibrahim, and Mohamed M. Khairy, "RF Beamforming for Secrecy Millimeter Wave MISO-OFDM Systems," in *IEEE International Conference on Communications (ICC) 2016*, Kuala Lumpur, Malaysia, May 2016.

7. Belal S. Amin, **Yahia R. Ramadan**, Ahmed S. Ibrahim, and Mahmoud H. Ismail, "Power Allocation for Device-to-Device Communication Underlaying Massive MIMO Multicasting Networks," in *IEEE Wireless Communications and Networking Conference (WCNC) 2015*, New Orleans, USA, March 2015.

8. **Yahia R. Ramadan**, Ahmed S. Ibrahim, and Mohamed M. Khairy, "Minimum Outage RF Beamforming for Millimeter Wave MISO-OFDM Systems," in *IEEE Wireless Communications and Networking Conference (WCNC) 2015*, New Orleans, USA, March 2015.

9. **Yahia R. Ramadan**, Ahmed S. Ibrahim, and Mohamed M. Khairy, "Reliable RF Beamforming for Millimeter Wave MIMO-OFDM Systems," in *IEEE 25th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2014*, Washington DC, USA, September 2014.