Energy Efficient and Low Complexity Techniques for the Next Generation Millimeter Wave Hybrid MIMO Systems

Aryan Kaushik

Supervised By: John Thompson

THE UNIVERSITY
of EDINBURGH

Thesis submitted for the degree of
Doctor of Philosophy to the
College of Science and Engineering at
The University of Edinburgh
January 2020

Abstract

The fifth generation (and beyond) wireless communication systems require increased capacity, high data rates, improved coverage and reduced energy consumption. This can be potentially provided by unused available spectrum such as the Millimeter Wave (MmWave) frequency spectrum above 30 GHz. The high bandwidths for mmWave communication compared to sub-6 GHz microwave frequency bands must be traded off against increased path loss, which can be compensated using large-scale antenna arrays such as the Multiple-Input Multiple-Output (MIMO) systems. The analog/digital Hybrid Beamforming (HBF) architectures for mmWave MIMO systems reduce the hardware complexity and power consumption using fewer Radio Frequency (RF) chains and support multi-stream communication with high Spectral Efficiency (SE). Such systems can also be optimized to achieve high Energy Efficiency (EE) gains with low complexity but this has not been widely studied in the literature. This PhD project focussed on designing energy efficient and low complexity communication techniques for next generation mmWave hybrid MIMO systems.

Firstly, a novel architecture with a framework that dynamically activates the optimal number of RF chains was designed. Fractional programming was used to solve an EE maximization problem and the Dinkelbach Method (DM) based framework was exploited to optimize the number of active RF chains and the data streams. The DM is an iterative and parametric algorithm where a sequence of easier problems converge to the global solution. The HBF matrices were designed using a codebook-based fast approximation solution called gradient pursuit which was introduced as a cost-effective and fast approximation algorithm. This work maximizes EE by exploiting the structure of RF chains with full resolution sampling unlike existing baseline approaches that use fixed RF chains and aim only for high SE.

Secondly, an efficient sparse mmWave channel estimation algorithm was developed with low resolution Analog-to-Digital Converters (ADCs) at the receiver. The sparsity of the mmWave channel was exploited and the estimation problem was tackled using compressed sensing through the Stein's unbiased risk estimate based parametric denoiser. The Expectation-maximization density estimation was used to avoid the need to specify the channel statistics. Furthermore, an energy efficient mmWave hybrid MIMO system was developed with Digital-to-Analog Converters (DACs) at the transmitter where the best subset of the active RF chains and the DAC resolution were selected. A novel technique based on the DM and subset selection optimization was implemented for EE maximization. This work exploits the low resolution sampling at the converting units and provides more efficient solutions in terms of EE and channel estimation than existing baselines in the literature.

Thirdly, the DAC and ADC bit resolutions and the HBF matrices were jointly optimized for EE maximization. The flexibility in choosing the bit resolution for each DAC and ADC was considered and they were optimized on a frame-by-frame basis unlike the existing approaches, based on the fixed resolution sampling. A novel decomposition of the HBF matrices to three parts was introduced to represent the analog beamformer matrix, the DAC/ADC bit resolution matrix and the baseband beamformer matrix. The alternating direction method of multipliers was used to solve this matrix factorization problem as it has been successfully applied to other non-convex matrix factorization problems in the literature. This work considers EE maximization with low resolution sampling at both the DACs and the ADCs simultaneously, and jointly optimizes the HBF and DAC/ADC bit resolution matrices, unlike the existing baselines that use fixed bit resolution or otherwise optimize either DAC/ADC bit resolution or HBF matrices.

Lay Summary

In this modern digital age of 21$^{\mathrm{st}}$ century, mobile users demand better communication technology which should be mainly cost-efficient, with less complex hardware and high speed. The microwave frequency spectrum that we currently use for mobile broadband is limited to a very crowded frequency range. There is an enhanced demand for an unused and available spectrum which can be resolved by the use of millimeter wave frequency spectrum. The larger bandwidth channels means higher data rates and we can further benefit by using multiple antenna systems at millimeter wave. The use of a hybrid architecture, which involves both digital and analog units used in conventional technologies, reduces the hardware complexity and power consumption for such systems while still supporting communication with multiple streams. In the existing literature, the millimeter wave multiple antenna systems are designed for high data rates but designing such systems for high energy efficiency with low complexity solutions and keeping high data rates, has not been widely studied. So this thesis focuses on designing energy efficient and low complexity techniques for the next generation millimeter wave multiple antenna systems. We provide energy efficient solutions by exploiting the structure of complex and power hungry components such as the radio frequency chains and associated conversion units. We also provide an efficient and low complexity solution to estimate the millimeter wave channel and consider the impact of resolution sampling associated with the conversion unit.

Declaration of Originality

I hereby declare that the research conducted in this thesis and the thesis itself has been composed solely by myself at the Institute for Digital Communications in the School of Engineering, The University of Edinburgh. The thesis has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own. The included publications are my own work, except where indicated in the thesis.

Aryan Kaushik

29 January 2020

Acknowledgements

First and foremost, I will always be short of words to thank my PhD research supervisor Prof. John Thompson without whom this research would have not been possible. He has always been there for me, providing unending support and motivation. He believed in my potentials and provided constructive suggestions, a positive atmosphere and a lot of encouragement. He trained me in building efficient research skills which led me to finish my PhD research effectively and in time. I would always be grateful to him for providing me the opportunity to conduct my PhD research on such exciting problems and being an extraordinary support throughout the study.

A very special thanks to my parents Gopal Dass Kaushik and Savitri Kaushik for providing an incredible support and encouragement to follow my dreams. I will always be grateful to them for supporting me personally and financially. I can not thank enough my girlfriend Gina for her continuous love and support. She always stood by me no matter where we were, held my hand firmly in all the life experiences and gave me strength to keep going.

I would like to thank all my colleagues at IDCOM and a special mention to Evangelos Vlachos who being a friend supported me in my research with a helpful attitude and we had fruitful discussions about my research work. I would like to thank Alessandro Perelli who being a friend became a continuous support in my

research. I would also like to acknowledge Mehrdad Yaghoobi who helped to kick-start my research and provided continuous advice.

I would also like to acknowledge the colleagues at the University of Luxembourg, Luxembourg, firstly, Symeon Chatzinotas for hosting my research visit, and my colleague and friend Christos Tsinos for being a good support in conducting the collaborative research successfully. I would like to acknowledge Rongke Liu for hosting my research visits at Beihang University, China, and Bruno Clerckx for hosting my research visit at the Imperial College London, UK.

Lastly I would like to thank all my friends from different countries and cultures for long lasting friendships and enjoyable experiences during my stay in Edinburgh.

Contents

List of Figures

List of Tables

List of Algorithms

Notations

a	Scalar
$\|a\|$	Determinant of a
$\mathbf{a}$	Vector
$\\|\mathbf{a}\\|_p$	p-norm of $\mathbf{a}$
$\mathbf{A}$	Matrix
$\|\mathbf{A}\|$	Determinant of $\mathbf{A}$
$\mathbf{A}^T$	Transpose of $\mathbf{A}$
$\mathbf{A}^H$	Complex conjugate transpose of $\mathbf{A}$
$[\mathbf{A}]_k$	k-th column of matrix $\mathbf{A}$
$[\mathbf{A}]_{kl}$	Matrix entry at the k-th row and l-th column
$\mathbf{A}^{(i)}$	i-th column of $\mathbf{A}$
$\mathbf{A}^{\dagger}$	Pseudo inverse of $\mathbf{A}$
$\\|\mathbf{A}\\|_F$	Frobenius norm of $\mathbf{A}$
$\mathbf{A}\|_{\Gamma}$	A matrix consisting of rows of matrix $\mathbf{A}$ with indices from Γ set
$[\mathbf{A}\|\mathbf{B}]$	Horizontal concatenation
$\mathbf{0}_{X \times Y}$	$X \times Y$ all-zeros matrix
$\mathbb{1}_{\mathcal{S}}\{\mathbf{A}\}$	An indicator function of a set $\mathcal{S}$ that acts over a matrix $\mathbf{A}$; defined as $0 \; \forall \; \mathbf{A} \in \mathcal{S}$ and $\infty \; \forall \; \mathbf{A} \notin \mathcal{S}$
$\mathcal{CN}(\mathbf{a}; \mathbf{A})$	Complex Gaussian vector; mean $\mathbf{a}$, covariance $\mathbf{A}$
$\mathbb{C}$	Set of complex numbers
$\mathbb{C}^{A \times B}$	Set of $A \times B$ matrices with complex entries
$\mathrm{diag}(\mathbf{A})$	Vector generated by the diagonal elements of $\mathbf{A}$
$\mathbb{E}\{\cdot\}$	Expectation operator
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{+}$	Set of positive real numbers
$\mathbb{R}\{\cdot\}$	Real part of a complex variable
$\mathbb{R}^{A \times B}$	Set of $A \times B$ matrices with real entries
$\mathrm{tr}(\mathbf{A})$	Trace of $\mathbf{A}$
$\mathrm{vec}(\mathbf{A})$	Vector with column entries of $\mathbf{A}$
$x \cup y$	Union of x and y union disjoint sets
$\mathbf{X} \in \mathbb{C}^{A \times B}$	$A \times B$ size $\mathbf{X}$ matrix with complex entries
$\mathbf{X} \in \mathbb{R}^{A \times B}$	$A \times B$ size $\mathbf{X}$ matrix with real entries
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$

List of Symbols

α_{il}	Gain of l-th ray in i-th cluster
α	Scalar penalty parameter
b^r	Bit resolution at ADC
b^t	Bit resolution at DAC
δ	Multiplicative distortion parameter
$\mathcal{D}$	Set of diagonal sparse matrices
$\mathcal{D}_{\mathrm{RX}}$	Set representing the finite states of the quantizer at the RX
$\mathcal{D}_{\mathrm{TX}}$	Set representing the finite states of the quantizer at the TX
EE	Energy efficiency
$\mathcal{F}$	Set of possible phase shifts in $\mathbf{F}_{\mathrm{RF}}$
γ_{R}	Trade-off parameter between rate and power consumption at the RX
γ_{T}	Trade-off parameter between rate and power consumption at the TX
L_{R}	Number of RX RF chains
L_{T}	Number of TX RF chains
m	Minimum value of bit resolution range
M	Maximum value of bit resolution range
N_{cl}	Number of multi-path clusters
N_{ray}	Number of multi-path rays
N_{R}	Number of RX antennas
N_{s}	Number of streams
N_{T}	Number of TX antennas
ϕ^t_{il}	Azimuth angles of departure
ϕ^r_{il}	Azimuth angles of arrival
P	Total consumed power
P_{ADC}	Power consumed per bit in the ADC
P_{CR}	Power required by all circuit components at the RX
P_{CT}	Power required by all circuit components at the TX
P_{DAC}	Power consumed per bit in the DAC
P_{DR}	Power for total quantization operation at the RX
P_{DT}	Power for total quantization operation at the TX
$P_{\max}$	Maximum allocated power
P_{PR}	Power per phase shifter at the RX
P_{PT}	Power per phase shifter at the TX
P_{R}	Power per antenna at the RX
P_{RX}	Power consumption at the RX
P_{T}	Power per antenna at the TX
P_{TX}	Power consumption at the TX

$Q(x)$	Uniform scalar quantizer with scalar complex input x applied to both real and imaginary parts
R	Information rate
$\mathcal{W}$	Set of possible phase shifts in $\mathbf{W}_{\mathrm{RF}}$
$\mathbf{a}_{\mathrm{R}}$	Normalized receive array response vector
$\mathbf{a}_{\mathrm{T}}$	Normalized transmit array response vector
$\boldsymbol{\epsilon}_{\mathrm{RX}}$	Additive quantization noise for ADC
$\boldsymbol{\epsilon}_{\mathrm{TX}}$	Additive quantization noise for DAC
$\boldsymbol{\eta}$	Combined effect of the additive white Gaussian RX noise and quantization noise
$\mathbf{n}$	Complex Gaussian noise with i.i.d. entries
$\mathbf{t}$	transmitted signal with dimensions $N_{\mathrm{T}} \times 1$
$\mathbf{r}$	RX output signal with dimensions $N_{\mathrm{s}} \times 1$
$\mathbf{x}/\mathbf{s}$	TX signal vector with dimensions $N_{\mathrm{s}} \times 1$
$\mathbf{y}$	Received signal with dimensions $N_{\mathrm{R}} \times 1$
$\mathbf{A}$	Arbitrary matrix for ADMM approach
$\mathbf{A}_{\mathcal{F}}$	Projection of matrix $\mathbf{A}$ onto set $\mathcal{F}$
$\mathbf{C}_{\epsilon\mathrm{R}}$	Diagonal covariance matrix for ADCs
$\mathbf{C}_{\epsilon\mathrm{T}}$	Diagonal covariance matrix for DACs
$\boldsymbol{\Delta}_{\mathrm{RX}}$	Diagonal matrix with values depending on the bit resolution of each ADC
$\boldsymbol{\Delta}_{\mathrm{TX}}$	Diagonal matrix with values depending on the bit resolution of each DAC
$\mathbf{F}_{\mathrm{BB}}$	Baseband precoder with dimensions $L_{\mathrm{T}} \times N_{\mathrm{s}}$
$\mathbf{F}_{\mathrm{opt}}/\mathbf{F}_{\mathrm{DBF}}$	Optimal fully digital precoder matrix
$\mathbf{F}_{\mathrm{RF}}$	Analog precoder with dimensions $N_{\mathrm{T}} \times L_{\mathrm{T}}$
$\mathbf{H}$	MmWave channel matrix
$\boldsymbol{\Lambda}$	Lagrange multiplier matrix with dimensions $N_{\mathrm{T}} \times L_{\mathrm{T}}$
$\mathbf{R}_{\eta}$	Combined noise covariance matrix
$\boldsymbol{\Sigma}_{\mathrm{H}}$	Rectangular matrix of singular values in channel's SVD with dimensions $N_{\mathrm{R}} \times N_{\mathrm{T}}$
$\mathbf{U}_{\mathrm{H}}$	Left singular matrix of channel's SVD with dimensions $N_{\mathrm{R}} \times N_{\mathrm{R}}$
$\mathbf{V}_{\mathrm{H}}$	Right singular matrix of channel's SVD with dimensions $N_{\mathrm{T}} \times N_{\mathrm{T}}$
$\mathbf{W}_{\mathrm{BB}}$	Baseband combiner with dimensions $L_{\mathrm{R}} \times N_{\mathrm{s}}$
$\mathbf{W}_{\mathrm{opt}}/\mathbf{W}_{\mathrm{DBF}}$	Optimal fully digital combiner matrix
$\mathbf{W}_{\mathrm{RF}}$	Analog combiner with dimensions $N_{\mathrm{R}} \times L_{\mathrm{R}}$
$\mathbf{Z}$	Auxiliary matrix for overall precoder

Acronyms and Abbreviations

5G	Fifth Generation
A/D	Analog/Digital
ADC	Analog-to-Digital Converter
ADMM	Alternating Direction Method of Multipliers
AMP	Approximate Message Passing
AQNM	Additive Quantization Noise Model
AWGN	Additive White Gaussian Noise
BF	Brute Force
CDF	Cumulative Distribution Function
CS	Compressed Sensing
CSI	Channel State Information
DAC	Digital-to-Analog Converter
DFT	Discrete Fourier transform
DM	Dinkelbach Method
EE	Energy Efficiency
EM	Expectation-Maximization
FS	Full Search
GAMP	Generalized Approximate Message Passing
GP	Gradient Pursuit
HBF	Hybrid Beamforming
I.I.D.	Independent and Identically Distributed

MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Square Error
MmWave	Millimeter Wave
MSE	Mean Square Error
OMP	Orthogonal Matching Pursuit
PDF	Probability Density Function
PMF	Probability Mass Function
RF	Radio Frequency
RX	Receiver
SE	Spectral Efficiency
SNR	Signal-to-Noise Ratio
SURE	Stein's Unbiased Risk Estimate
SVD	Singular Value Decomposition
TX	Transmitter
ULA	Uniform Linear Array
VAMP	Vector Approximate Message Passing
W.R.T.	With Respect To

Chapter 1

Introduction

THIS THESIS addresses efficient communication techniques for the Fifth Generation (5G) and beyond Millimeter Wave (MmWave) Hybrid Beamforming (HBF) Multiple-Input Multiple-Output (MIMO) systems. Our main objective is to optimize such systems for Energy Efficiency (EE) maximization with low complexity by exploiting the Analog/Digital (A/D) HBF architecture and provide better solutions than existing baselines in the literature. Low complexity refers to reducing the complexity associated with an algorithm or system design, i.e., providing a fast solution with acceptable accuracy and the least computation. In this introductory chapter, Section 1.1 introduces the motivation of the research work carried out. Section 1.2 summarizes the objectives and key contributions of the research work conducted for this thesis. Section 1.3 provides an outline of the remaining chapters of this thesis.

1.1 Motivation

1.1.1 5G Mobile Communication

For future mobile communication systems, there is a wide scope to identify the technical needs and possible solutions to transform and revolutionize the wireless connectivity ecosystem for a better inter-connected society. The 5G technology is set to address the consumer demands and performance enhancements for mobile communication in 2020 and beyond [1–3]. The emerging advanced consumer devices and developed communication systems have resulted in ever-increasing demands on bandwidth and capacity. For instance, Cisco's annual report suggests that mobile video traffic is expected to generate 82% of the global mobile data traffic, and there will be 28.5 billion networked devices and connections by 2022 [4]. Ericsson mobility report [5] forecasts that there will be 8.9 billion mobile subscriptions by the end of 2024 and more than 40% of the world's population is forecast to be covered by 5G in the same year (see Fig. 1.1). The 5G (and beyond) services are expected to be commercially implemented on a *large scale* in the next few years, for example, in North America and North East Asia significant 5G subscriptions are expected early [5]. The 5G (and beyond) standards would require high data rates/throughput, improved coverage, lower latency, high mobility, high reliability and lower infrastructure costs [1, 2].

One of the building blocks for fulfilling the requirements of 5G mobile communications is the use of MIMO technology and spectrum availability. The microwave frequency spectrum at sub-6 GHz frequencies, which we currently make use of for mobile broadband, is limited to a very crowded frequency range enhancing the demand for unused and available spectrum which can be resolved by the use of mmWave frequency spectrum [6, 7].

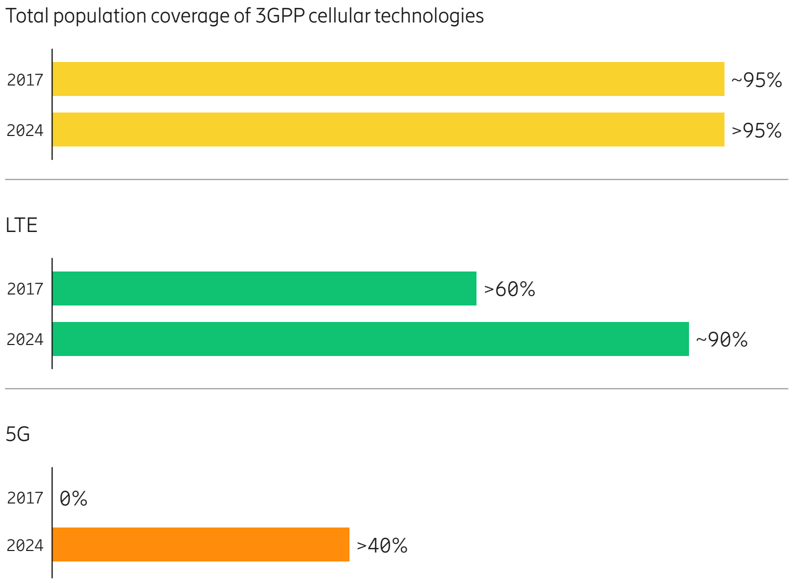
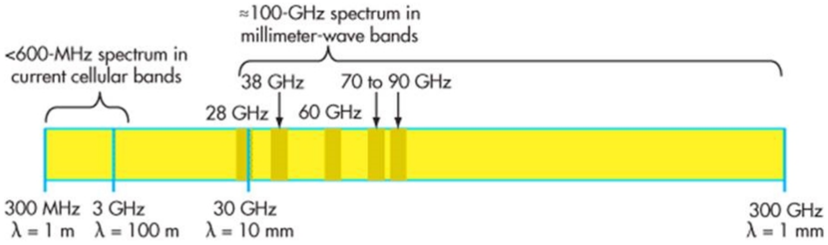


Figure 1.1: World population coverage by technology [5].



Source: Electronic Design

Figure 1.2: Frequency spectrum allocation to mmWave band.

1.1.2 MmWave Channel Characteristics

The use of mmWave frequency bands appears to be a promising technology to meet the needs of the 5G mobile communication systems [8–10]. Mmwave makes use of spectrum from 30 GHz to 300 GHz whereas most consumer wireless systems operate at carrier frequencies below 6 GHz. Fig. 1.2 shows the frequency spectrum allocation to mmWave band. Reference [11] states that the United States Federal Communications Commission (FCC) has freed approximately 30

Channel attributes	**Values**
Bandwidth	100 MHz - 2 GHz
Base Station (BS) antennas	64 - 256
Mobile Station (MS) antennas	4 - 16
Channel Sparsity	High
Spatial Correlation	High
Angular Spread	$<$ 50 degrees
Orientation Sensitivity	High

Table 1.1: Channel attributes at mmWave.

times more bandwidth at mmWave frequencies than is available at cellphone bands for commercial use.

The main benefit of a mmWave band is the larger spectral channels, and larger bandwidth channels means higher data rates. Due to their high data rates, a few existing applications of the mmWave spectrum are in satellite communications, wireless backhaul and radio applications. Also, radar systems occupy some of the mmWave bands, for example, 77 GHz will be used as one band for radar in driverless cars. However, mmWave faces challenges of severe path loss, blocking effects, new hardware constraints and unconventional channel characteristics [11]. Table 1.1 discusses typical mmWave channel characteristics which may be considered as important attributes when considering mmWave frequency channels for future 5G (and beyond) standards. For example, an important characteristic of a mmWave frequency channel is high sparsity, i.e., there are only few non-zero elements in the channel matrix, in both the angle and delay domains [7, 12]. Other important properties are high spatial correlation meaning that some spatial directions during mmWave communication are statistically stronger than others, and high sensitivity towards the orientation angle of the user equipment.

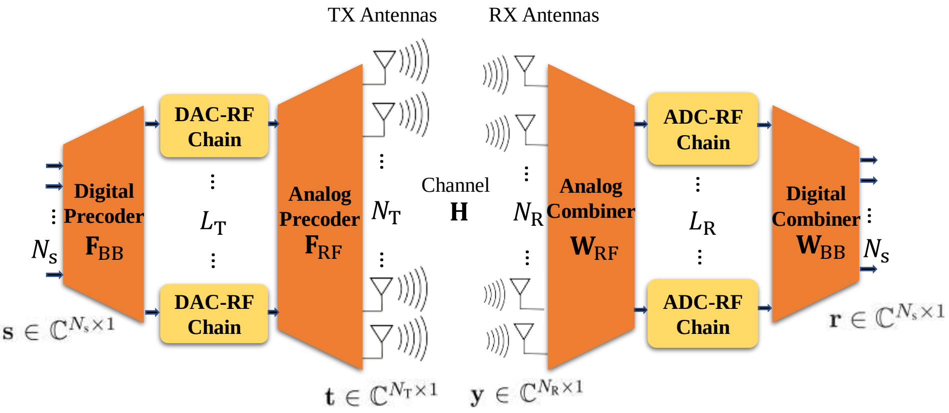

TX Antennas
RX Antennas
Digital Precoder $\mathbf{F}_{\mathrm{BB}}$
N_{s}
$\mathbf{s} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$
DAC-RF Chain
L_{T}
DAC-RF Chain
Analog Precoder $\mathbf{F}_{\mathrm{RF}}$
N_{T}
$\mathbf{t} \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$
Channel $\mathbf{H}$
N_{R}
$\mathbf{y} \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$
Analog Combiner $\mathbf{W}_{\mathrm{RF}}$
ADC-RF Chain
L_{R}
ADC-RF Chain
Digital Combiner $\mathbf{W}_{\mathrm{BB}}$
N_{s}
$\mathbf{r} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$

Figure 1.3: Block diagram of a HBF architecture for mmWave MIMO system.

1.1.3 MmWave MIMO: Potentials and Challenges

The high bandwidths for mmWave communication compared to sub-6 GHz frequency bands must be traded off against increased path loss [13], which can be compensated using large-scale antenna arrays, i.e., MIMO systems [14, 15]. The large number of antenna elements and the high bandwidth makes it hard to use a separate Radio Frequency (RF) chain for each antenna due to the large requirements in power consumption and hardware complexity [15]. Also, using many Digital-to-Analog Converter (DAC)/Analog-to-Digital Converter (ADC) units associated with RF chains, which are power hungry components, would lead to more hardware complexity and high power consumption. Moreover, DACs/ADCs have a relatively higher sampling rate in high frequency systems than at microwave frequencies, and employing high speed converters increases the power consumption and the cost significantly.

These hardware constraints have led to several mmWave-specific MIMO architectures where a mixture of analog and digital signal processing operations are made With Respect To (W.R.T.) the number of antennas or resolution of data converters. The HBF architectures as shown in Fig. 1.3 are one approach for providing enhanced benefits of MIMO communication at mmWave frequencies. This architecture is discussed further in the following chapters. Note that, in such an architecture, the number of RF chains and associated ADCs/DACs are much less than the number of antennas, and enables spatial multiplexing and multi-user communication that enhances the benefits of MIMO. The benefits of using a HBF architecture over conventional beamforming architectures is discussed in the next chapter.

1.2 Objectives and Key Contributions

1.2.1 Objectives

In HBF architectures, the hardware complexity and power consumption is reduced through using fewer RF chains and it can support multi-stream communication with high Spectral Efficiency (SE) [14–23]. Such systems can also be optimized to achieve high EE gains with low complexity but this has not been widely studied in the literature. Thus, the aim of this thesis is to design energy efficient and low complexity communication techniques for the mmWave HBF MIMO systems which may be implemented in 5G standards. In particular, the thesis has the following main objectives:

- Designing energy efficient mmWave hybrid MIMO systems with low complexity by exploiting the structure of complex and power hungry components such as RF chains and DAC/ADC units.

- Exploiting the sparsity of the mmWave channel, and provide an efficient and low complexity solution for sparse channel estimation while considering low resolution sampling.

1.2.2 Key Contributions

The key contributions of this thesis are summarized as follows:

- **EE maximization** by optimizing the number of RF chains unlike existing baseline approaches that use a fixed number of RF chains and aim only for high SE. Fractional programming is used to solve an EE maximization problem and the Dinkelbach Method (DM) based framework is exploited to optimize the number of active RF chains and the data streams. The HBF matrices are designed using a codebook-based fast approximation solution.

- **Sparse channel estimation** algorithm is developed with low resolution sampling at the Receiver (RX) using Compressed Sensing (CS) through Stein's Unbiased Risk Estimate (SURE) based parametric denoiser and Expectation-Maximization (EM) density estimation. An **EE maximization** solution is also developed with low resolution sampling at the Transmitter (TX) where the best subset of the active RF chains and the DAC resolution were selected based on the DM and subset selection optimization approach.

- **EE maximization** by decomposing the HBF matrices into three matrices, which are the analog beamforming matrix, the bit resolution matrix and the baseband beamforming matrix at both the TX and the RX. These matrices are obtained by the solution of an EE maximization problem where the joint TX/RX problem is decoupled into two sub-problems and the corresponding problems are solved by Alternating Direction Method of Multipliers (ADMM). We jointly optimize the HBF and bit resolution matrices unlike existing approaches that optimize either the bit resolution or the HBF matrices.

1.3 Thesis Outline

The remainder of this thesis is organized as follows:

Chapter 2

This chapter provides the background for this thesis. An overview of mmWave MIMO is provided. MIMO beamforming and the advantages of the HBF architecture over conventional architectures are discussed. An overview of convex optimization and CS techniques for mmWave HBF MIMO systems is also described.

Chapter 3

This chapter is mainly based on [24] which proposes a novel architecture with a framework that dynamically activates the optimal number of RF chains. Fractional programming is used to solve an EE maximization problem and the HBF matrices are designed using a codebook-based fast approximation solution. The greedy strategy implemented to compute HBF matrices in this chapter was introduced for mmWave HBF MIMO systems in our work in [17].

Chapter 4

This chapter is in part based on [25] which proposes an efficient sparse mmWave channel estimation algorithm with low resolution ADCs at the RX. The sparsity of the mmWave channel is exploited and the estimation problem is tackled using CS. Also, this chapter reports on results in [26] where an energy efficient mmWave hybrid MIMO system is developed with DACs at the TX where the best subset of the active RF chains and the DAC resolution are selected.

Chapter 5

In [27], bit allocation and hybrid combining at the RX are discussed, and the number of ADC bits and hybrid combiner matrices are jointly optimized for EE maximization. In addition, this chapter is based on [28] which proposes the joint optimization of the bit allocation and the HBF matrices at both the TX and the RX for EE maximization unlike the existing approaches that optimize either the bit resolution or the HBF matrices. The HBF matrix is decomposed into the analog beamforming matrix, the bit resolution matrix and the baseband beamforming matrix at both the TX and the RX. These matrices are obtained through the solution of a joint TX-RX EE maximization problem.

Chapter 6

This chapter concludes this thesis and provides possible future research directions.

Chapter 2

Background

THIS CHAPTER provides a basic technical background for this thesis. This chapter starts by providing an overview of mmWave MIMO systems which includes applications of the mmWave communications, the basics of Additive White Gaussian Noise (AWGN) channel capacity, MIMO channel capacity, mmWave channel models and mmWave channel estimation techniques. Then several MIMO beamforming architectures and the advantages of implementing the HBF architectures over the conventional beamforming architectures are described. The HBF architectures for mmWave MIMO systems reduce the hardware complexity and power consumption using fewer Radio Frequency (RF) chains while supporting multi-stream communication with high Spectral Efficiency (SE). Then an overview of convex optimization and Compressed Sensing (CS) techniques for mmWave HBF MIMO systems is also provided. The study of these signal processing techniques is very important to develop energy efficient and low complexity solutions for mmWave HBF MIMO systems. Finally, a summary of this chapter is provided.

2.1 Overview of MmWave MIMO

This section provides the basics of the AWGN channel capacity and MIMO channel capacity. We then proceed with how the benefits of MIMO systems can be exploited at mmWave frequencies. MmWave makes use of spectrum from 30 GHz to 300 GHz whereas most consumer wireless systems operate at carrier frequencies below 6 GHz. The main benefit of mmWave communication is larger spectral channels and larger bandwidth channels means higher data rates.

However, mmWave faces challenges of severe path loss, blocking effects, new hardware constraints and unconventional channel characteristics. The high bandwidths for mmWave communication compared to microwave bands must be traded off against increased path loss, which can be compensated using large-scale antenna arrays, i.e., MIMO systems. Next, we discuss the applications for mmWave communications and how mmWave propagation with large-scale antenna arrays impacts the hardware complexity and power efficiency.

2.1.1 Applications of the MmWave Communications

As we know, the main benefit of a mmWave band is the larger spectral channels, and larger bandwidth channels means higher data rates. Due to their high data rates, a few existing applications of the mmWave spectrum are in satellite communications, wireless backhaul and radio applications. Also, radar systems occupy some of the mmWave bands, for example, 77 GHz will be used as one band for radar in driverless cars. However, mmWave propagation has the limitation of being affected by blockage effects, for example, from the human body (attenuation from 20 to 35 dB [29]) and building materials such as brick (attenuation of 40 to 80 dB [30, 31]).

In addition to path loss and blockage effects, mmWave wave communication shows hardware constraints and unconventional channel characteristics. For instance, the large number of antenna elements and the high bandwidth makes it hard to use a separate RF chain for each antenna due to the large requirements in power consumption and hardware complexity. Implementing a very large number of antennas, i.e., massive MIMO, would achieve high data rate performance but would increase the hardware complexity and reduce the power efficiency considerably. Also, using many DAC/ADC units associated with RF chains, which are power hungry components, would lead to more hardware complexity and high power consumption. Thus, there is a need to exploit enhanced benefits of MIMO communication at mmWave frequencies through unconventional beamforming architectures such as the Hybrid Beamforming (HBF) architecture. Next we proceed with the basic AWGN channel capacity and MIMO channel capacity, and the benefits of implementing MIMO at mmWave.

2.1.2 AWGN Channel Capacity

The Shannon capacity provides the maximal rate to achieve reliable communication over a noisy channel. Communicating at the rates above this channel capacity fails to provide zero error probability for very large data packet sizes. The following equation provides the basic AWGN channel model [32]:

$$y[m] = x[m] + n[m], \tag{2.1}$$

where $x(m)$ is a complex-valued input, $y(m)$ is the complex-valued output, both at time m, and $n(m)$ denotes the complex Gaussian-distributed noise corrupting the Receiver (RX) which is independent over time with 0 mean and variance σ^2. Similar to (2.1), considering a continuous-time AWGN channel with B Hz bandwidth, $\bar{P}$ W transmit power and $N_0/2$ power spectral density at the RX for the AWGN. For B complex samples per second, the capacity of such a channel can be expressed as

$$C_{\text{AWGN}}(\bar{P}, B) = B \log\left(1 + \frac{\bar{P}}{N_0 B}\right) \text{ (bits/s)} \tag{2.2}$$

$$\implies SE_{\text{AWGN}} = \log(1 + \text{SNR}) \text{ (bits/s/Hz)}, \tag{2.3}$$

where SNR $= \bar{P}/(N_0 B)$ denotes the Signal-to-Noise Ratio (SNR) per degree of freedom. Equation (2.3) represents the maximum achievable SE for the AWGN channel in terms of SNR.

The dependence of the capacity C_{AWGN} can be observed in two ways: (a) linear dependency on B for a fixed SNR $= \bar{P}/(N_0 B)$, and (b) SNR decreases with the bandwidth for a given received power $\bar{P}$. However, when the bandwidth is large such that SNR at each frequency is small, we have

$$\begin{aligned} B \log\left(1 + \frac{\bar{P}}{N_0 B}\right) &\approx B \left(\frac{\bar{P}}{N_0 B}\right) \log_2 e \\ &= \frac{\bar{P}}{N_0} \log_2 e, \end{aligned} \tag{2.4}$$

which shows that the capacity is proportional to the total received power and increasing B does not have a significant impact on capacity. When B tends to infinity, we reach the limit of $C_{\text{inf}} = \frac{\bar{P}}{N_0} \log_2 e$, where there is no bandwidth dependence and the capacity has a finite value.

Moreover, [32] suggests that a frequency selective AWGN channel can be converted into a number of independent sub-carriers. The transformed channel can be treated as a collection of sub-channels, where each sub-channel is an AWGN channel and the total power constraint is across the sub-channels. Some power is allocated to each sub-channel which add up to the total power constraint and power allocation can be chosen appropriately to maximize rate [32]. The optimal power allocation can be computed using the waterfilling power allocation approach [32]. Transmitter (TX) allocates more power to the sub-carriers which are stronger where there are better channel conditions and the weaker sub-carriers are either allocated lesser power or no power at all. The waterfilling power allocations in MIMO channel capacity are also described in the following subsection.

2.1.3 MIMO Channel Capacity

A MIMO system is a multi-antenna system as shown in Fig. 2.1 with a channel matrix $\mathbf{H} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$ with N_T TX antennas and N_R RX antennas, and assume that the Channel State Information (CSI) is known to both the TX and the RX perfectly. Using the same time-invariant and narrowband channel, RX antennas receive both the direct components such as H_{11}, H_{22} etc., and indirect components such as H_{21}, H_{12} etc., which are the entries of the channel matrix. The TX data is divided into N_s streams where the number of streams N_s is always less than or equal to the number of antennas. The received signal $\mathbf{y} \in \mathbb{C}^{N_\mathrm{R} \times 1}$ can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.5}$$

where $\mathbf{x} \in \mathbb{C}^{N_\mathrm{T} \times 1}$ is the transmitted signal, and $\mathbf{n}$ is the the Gaussian noise with Independent and Identically Distributed (I.I.D.) entries and complex Gaussian distribution, i.e., $\mathbf{n} \sim \mathcal{CN}(0, N_0 \mathbf{I}_{N_\mathrm{R}})$.

The Singular Value Decomposition (SVD) of the channel matrix $\mathbf{H}$ can be

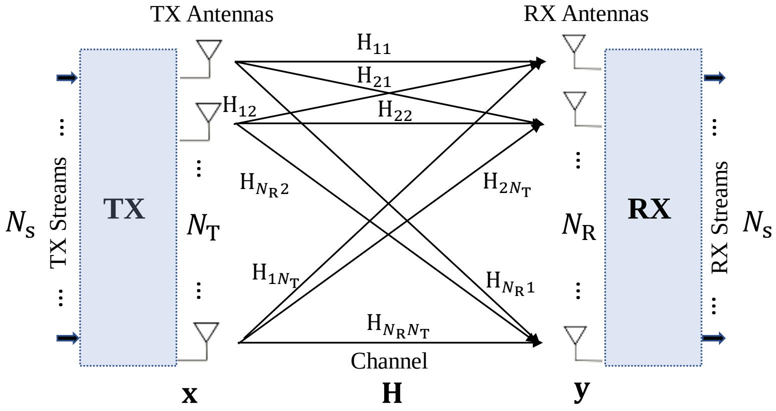


Figure 2.1: Block diagram of a $N_T \times N_R$ MIMO system.

expressed as follows:

$$\mathbf{H} = \mathbf{U}_\mathrm{H}\mathbf{\Sigma}_\mathrm{H}\mathbf{V}_\mathrm{H}^H, \tag{2.6}$$

where $\mathbf{U}_\mathrm{H} \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{R}}$ and $\mathbf{V}_\mathrm{H} \in \mathbb{C}^{N_\mathrm{T}\times N_\mathrm{T}}$ are unitary matrices, and $\mathbf{\Sigma}_\mathrm{H} \in \mathbb{R}^{N_\mathrm{R}\times N_\mathrm{T}}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are $\lambda_1 \geq \lambda_1 \geq ... \geq \lambda_{l_\mathrm{min}}$ (where $l_\mathrm{min} = \min(N_\mathrm{T}, N_\mathrm{R})$) which are non-negative real numbers and whose non-diagonal elements are zero. The λ_i^2 values represent the eigenvalues of the matrix $\mathbf{H}\mathbf{H}^H$ and also for the matrix $\mathbf{H}^H\mathbf{H}$, and we have

$$\mathbf{H}\mathbf{H}^H = \mathbf{U}_\mathrm{H}\mathbf{\Lambda}_\mathrm{H}\mathbf{\Lambda}_\mathrm{H}^T\mathbf{U}_\mathrm{H}^H, \tag{2.7}$$

The SVD can be re-written as the sum of rank-one matrices as follows:

$$\mathbf{H} = \sum_{i=1}^{l_\mathrm{min}} \lambda_i \mathbf{u}_i \mathbf{v}_i^H. \tag{2.8}$$

The rank of the channel matrix **H** is equal to the number of non zero singular values. Following the SVD, the MIMO channel capacity can be expressed as

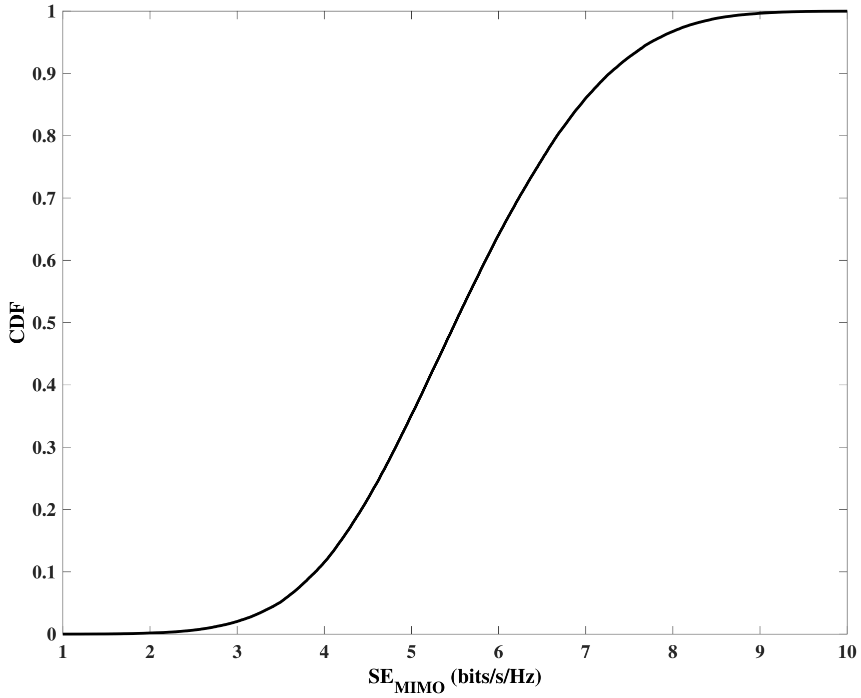


Figure 2.2: CDF plot for 2×2 MIMO system with Rayleigh fading channel.

follows [32]:

$$C_{\text{MIMO}} = B \sum_{i=1}^{l_{\min}} \log\left(1 + \frac{P_i^* \lambda_i^2}{N_0}\right) \text{ (bits/s)}, \tag{2.9}$$

$$\implies SE_{\text{MIMO}} = \sum_{i=1}^{l_{\min}} \log\left(1 + \frac{P_i^* \lambda_i^2}{N_0}\right) \text{ (bits/s/Hz)}, \tag{2.10}$$

where $P_i^*, ..., P_{l_{\min}}^*$ represent the waterfilling power allocations such as $P_i^* = (\mu - N_0/\lambda_i^2)^+$, where μ satisfies the total power constraint $\sum_i P_i^* = P$. Note that each of the non-zero λ_i entries can support a data stream which allows the MIMO channel to support spatial multiplexing with multiple streams.

For Rayleigh fading channel which has complex Gaussian distribution $\mathcal{CN}(0,1)$, Fig. 2.2 shows the variations of Cumulative Distribution Function (CDF) for a 2×2 MIMO system. Note that y-axis represents the cumulative probability which is an increasing function and varies between 0 to 1 With Respect To (W.R.T.) SE in bits/s/Hz at the x-axis. It can be observed that the

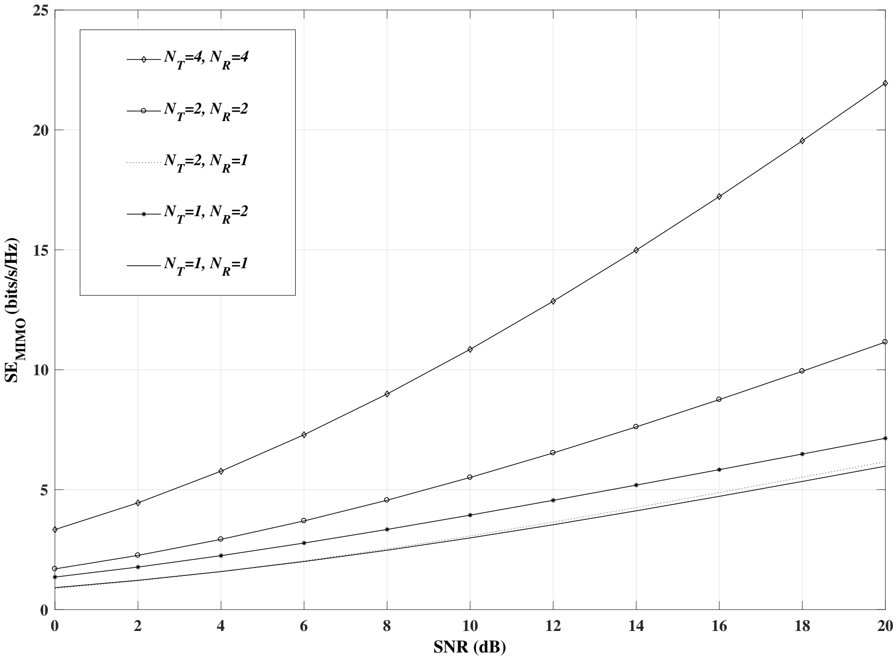


Figure 2.3: SE w.r.t. SNR for a MIMO channel with different number of N_T TX and N_R RX antennas.

average SE is about 5.5 bits/s/Hz and the 10%-ile SE is about 4 bits/s/Hz. Fig. 2.3 shows the variations of MIMO SE w.r.t. SNR for different numbers of N_T TX antennas and N_R RX antennas. It can be observed that the SE increases with increases in SNR and higher number of antennas show higher capacity values for a given SNR. For example, at 10 dB SNR, the case of $N_\mathrm{T} = 4$ and $N_\mathrm{R} = 4$ has 5 bits/s/Hz higher SE than the case of $N_\mathrm{T} = 2$ and $N_\mathrm{R} = 2$. This plot provides a basic example of the benefits of implementing large-scale antenna arrays, i.e., MIMO systems, to achieve higher SE.

Note that, for a MIMO system, the antennas may all be located at one TX/RX which is called as the single user MIMO system or each antenna may belong to a different TX/RX which is called as the multi-user MIMO system. Fig. 2.4 shows the basic block diagram of a downlink multi-user MIMO system where we show

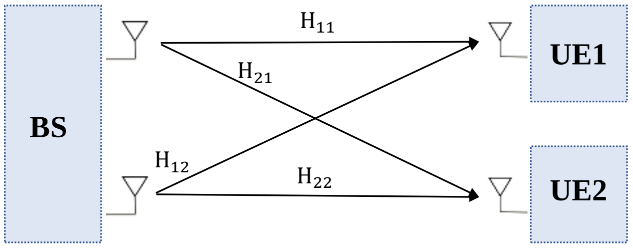


Figure 2.4: Block diagram of a downlink multi-user MIMO system.

one base station (BS) at the TX unit and two users (UE) at the RX side. In the following, we proceed to discuss how MIMO approaches can be implemented efficiently at mmWave frequencies.

2.1.4 MmWave MIMO Channel Models

Due to high frequency, i.e., small wavelength, mmWave channel characteristics are different than that of microwave. By Friis' Law [33], the received power P_{R} is related to the transmit power P_{T} as follows:

$$P_{\mathrm{R}} = G_{\mathrm{R}} G_{\mathrm{T}} \left(\frac{\lambda}{4\pi d} \right)^2 P_{\mathrm{T}}, \tag{2.11}$$

where G_{R} and G_{T} are RX and TX antenna gains, respectively, λ is the wavelength and d is the distance between the TX and the RX. Note that for unit gains, i.e., $G_{\mathrm{R}} = G_{\mathrm{T}} = 1$, the ratio $P_{\mathrm{T}}/P_{\mathrm{R}}$ is inversely proportional to the square of the wavelength. It indicates that when there are no directional antenna gains, for high frequency propagation such as mmWave, the path loss is expected to be higher than for lower sub-6 GHz frequencies such as the microwave frequency bands. This higher path loss associated with mmWave spectrum can be compensated by directional transmission using large scale antenna arrays such as MIMO systems.

The mmWave MIMO systems can be modelled using similar channel models as used for microwave frequency spectrum [32] taking the mmWave-specific channel characteristics into account. Consider N_{T} TX antennas and N_{R} RX antennas, $\mathbf{a}_{\mathrm{T}}(\phi_{il}^{t})$ and $\mathbf{a}_{\mathrm{R}}(\phi_{il}^{r})$ being the normalized transmit and receive array response/steering vectors [16], where ϕ_{il}^{t} and ϕ_{il}^{r} denote the azimuth angles of departure and arrival, respectively. For carrier wavelength λ, d inter-element spacing, and a Uniform Linear Array (ULA) geometry with N_{Z} antenna elements (N_{T} at the TX and N_{R} at the RX) to compute the array response vector $\mathbf{a}_{\mathrm{Z}}$ ($\mathbf{a}_{\mathrm{T}}$ at the TX and $\mathbf{a}_{\mathrm{R}}$ at the RX) as follows [34]:

$$\mathbf{a}_{\mathrm{Z}}(\phi) = \frac{1}{\sqrt{N_{\mathrm{Z}}}}[1, e^{j\frac{2\pi}{\lambda}d\sin(\phi)}, ..., e^{j(N_{\mathrm{Z}}-1)\frac{2\pi}{\lambda}d\sin(\phi)}]^{T}. \tag{2.12}$$

It is useful to represent the channel in the frequency domain, however, as the channel response is time-varying in general so the channel matrix $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ can be expressed at time t and frequency f [15] as

$$\mathbf{H}(t, f) = \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{l=1}^{N_{\mathrm{ray}}} \alpha_{il} e^{j2\pi(\nu_{il}t - \tau_{il}f)} \mathbf{a}_{\mathrm{R}}(\phi_{il}^{r}) \mathbf{a}_{\mathrm{T}}(\phi_{il}^{t})^{H}, \tag{2.13}$$

where N_{cl} is the number of clusters, N_{ray} is the number of rays in each cluster, the number of paths can be classified as clustered multipaths, i.e., the product of $N_{\mathrm{cl}}N_{\mathrm{ray}}$. The parameter α_{il} is the complex gain, τ_{il} is delay, and ν_{il} is Doppler shift which is determined by the angle of arrival or departure. The above equation (2.13) can be approximated as follows, when Doppler shifts associated with all the paths are small over a signal duration T, i.e., $\nu_{il}T << 1 \,\forall\, i = 1, .., N_{\mathrm{cl}}; l = 1, .., N_{\mathrm{ray}}$:

$$\mathbf{H}(f) = \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{l=1}^{N_{\mathrm{ray}}} \alpha_{il} e^{j2\pi\tau_{il}f} \mathbf{a}_{\mathrm{R}}(\phi_{il}^{r}) \mathbf{a}_{\mathrm{T}}(\phi_{il}^{t})^{H}. \tag{2.14}$$

Additionally if the bandwidth of the channel B is sufficiently small so that $\tau_{il}B << 1 \, \forall \, i = 1, .., N_{\text{cl}}; l = 1, .., N_{\text{ray}}$ then we obtain the narrowband spatial model for the channel matrix as follows:

$$\mathbf{H} = \sum_{i=1}^{N_{\text{cl}}} \sum_{l=1}^{N_{\text{ray}}} \alpha_{il} \mathbf{a}_{\text{R}}(\phi_{il}^{r}) \mathbf{a}_{\text{T}}(\phi_{il}^{t})^{H}. \tag{2.15}$$

The antenna elements at the TX and the RX can be modeled as ideal sectored elements [35] and then antenna element gains can be evaluated over ideal sectors. In (3.1), the transmit and receive antenna element gains are considered unity over ideal sectors defined by $\phi_{il}^{t} \in [\phi_{\text{min}}^{t}, \phi_{\text{max}}^{t}]$ and $\phi_{il}^{r} \in [\phi_{\text{min}}^{r}, \phi_{\text{max}}^{r}]$, respectively.

Note that the fading channel models used in traditional MIMO becomes inaccurate for mmWave channel modeling due to the high free-space path loss changes in material reflection coefficients and blockage effects plus the use of large tightly-packed antenna arrays. The existing literature mostly addresses the narrowband clustered channel model [36, 37] for mmWave propagation due to different channel settings such as number of multipaths, amplitudes, etc. such as in [15], [16].

Furthermore, the large scale antenna arrays and highly directional characteristic of propagation at mmWave leads to beamspace representation of the mmWave MIMO channels. For L_{T} number of RF chains at the TX and L_{R} number of RF chains at the RX, the beamspace representation [38,39] of the narrowband channel can be written as follows:

$$\mathbf{H} = \mathbf{D}_{\text{R}} \mathbf{H}_{\text{v}} \mathbf{D}_{\text{T}}^{H}, \tag{2.16}$$

where $\mathbf{H}_{\text{v}} \in \mathbb{C}^{L_{\text{R}} \times L_{\text{T}}}$ represents a sparse matrix with a few non-zero entries, while $\mathbf{D}_{\text{R}} \in \mathbb{C}^{N_{\text{R}} \times L_{\text{R}}}$ and $\mathbf{D}_{\text{T}} \in \mathbb{C}^{N_{\text{T}} \times L_{\text{T}}}$ are the Discrete Fourier transform (DFT)

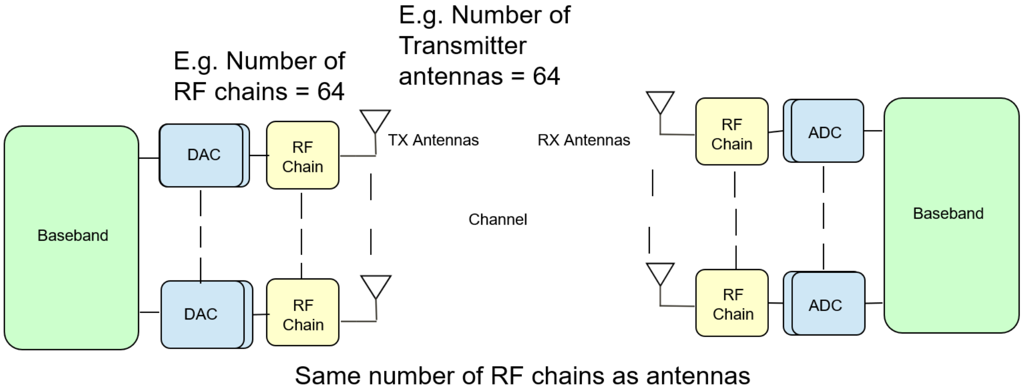

E.g. Number of RF chains = 64
E.g. Number of Transmitter antennas = 64
Baseband
DAC
DAC
RF Chain
RF Chain
TX Antennas
Channel
RX Antennas
RF Chain
RF Chain
ADC
ADC
Baseband
Same number of RF chains as antennas

Figure 2.5: MmWave MIMO System with Fully Digital Beamforming.

matrices. In the next section, we discuss the beamforming techniques that can be applied to design the mmWave MIMO systems.

2.2 MIMO Beamforming Architectures for the MmWave Band

2.2.1 Conventional Beamforming

At 6-sub GHz microwave frequencies, digital or baseband processing plays a vital role in MIMO communication. However, for MIMO communication at mmWave frequencies, the large number of antenna elements and the high bandwidth makes it hard to use a separate RF chain for each antenna due to the large requirements in power consumption and hardware complexity [15].

A conventional fully digital beamforming architecture used for sub-6 GHz frequencies is shown in Fig. 2.5, which has a digital/baseband unit and DAC/ADCs with one RF chain associated per antenna, i.e., there are same number of RF chains as the number of antennas. As digital beamforming architecture requires a dedicated RF chain per antenna with the electronic

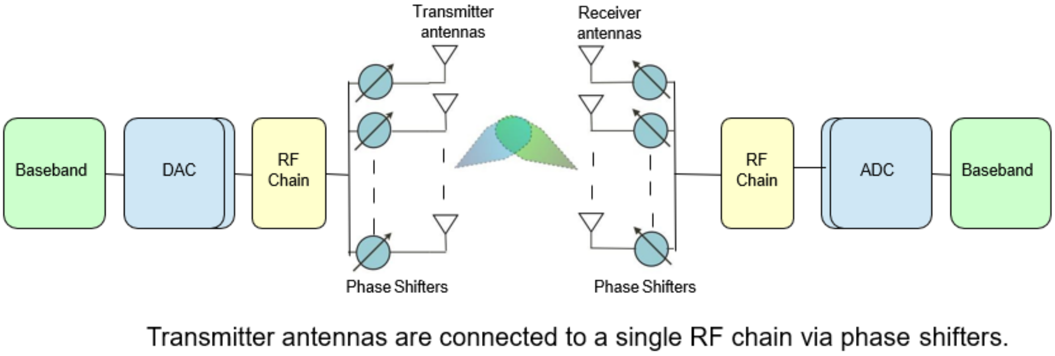


Figure 2.6: MmWave MIMO System with Analog Only Beamforming.

components such as DACs and ADCs that enhances the hardware complexity and power consumption with the increase in antenna size [14, 15]. Thus, a digital beamforming architecture currently seems impractical to be implemented for large scale antenna arrays in the mmWave band due to high power consumption and hardware complexity.

As an alternative, an analog beamforming approach could be considered to solve this problem. The analog beamforming architecture, shown in Fig. 2.6, has a digital/baseband unit and involves a network of analog phase shifters with a single RF chain in the system [40, 41], i.e., all the TX/RX antennas are connected with a single RF chain only. This approach is highly advantageous to reduce hardware complexity and power consumption. However, the analog only beamforming approach only supports single-user and single-stream transmission, i.e., it cannot support multi-stream and multi-user communication which are typical benefits associated with MIMO. Moreover, the capacity performance is usually significantly worse than the fully digital beamforming. Thus a more adaptable beamforming approach is needed for mmWave MIMO systems that could compensate the limitations associated with the conventional beamforming architectures.

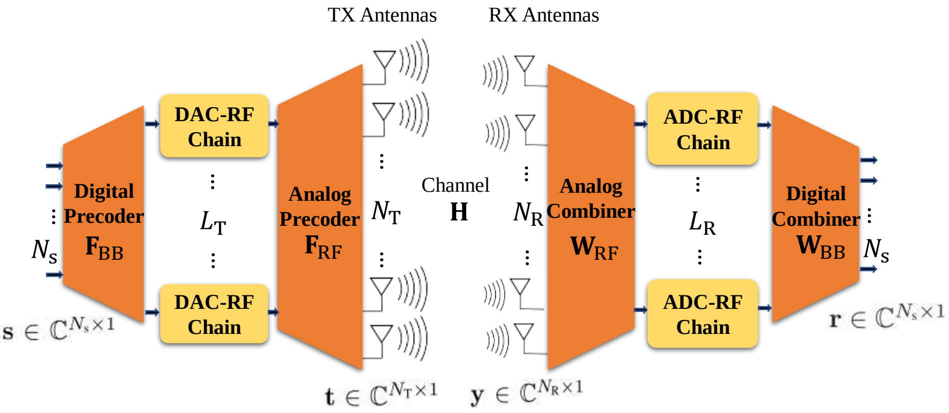

TX Antennas
RX Antennas
DAC-RF Chain
DAC-RF Chain
Digital Precoder $\mathbf{F}_{\text{BB}}$
N_s
$\mathbf{s} \in \mathbb{C}^{N_s \times 1}$
L_T
Analog Precoder $\mathbf{F}_{\text{RF}}$
N_T
Channel $\mathbf{H}$
N_R
Analog Combiner $\mathbf{W}_{\text{RF}}$
ADC-RF Chain
ADC-RF Chain
L_R
Digital Combiner $\mathbf{W}_{\text{BB}}$
N_s
$\mathbf{r} \in \mathbb{C}^{N_s \times 1}$
$\mathbf{t} \in \mathbb{C}^{N_T \times 1}$
$\mathbf{y} \in \mathbb{C}^{N_R \times 1}$

Figure 2.7: Block diagram of a mmWave MIMO system with HBF architecture.

2.2.2 Hybrid Beamforming

The performance of the mmWave MIMO systems can be significantly improved through the use of Analog/Digital (A/D) hybrid beamforming architectures, as shown in Fig. 2.7. This architecture is discussed in detail with the system and channel model parameters in the following chapter. From Fig 2.4, we can notice that in a A/D HBF architecture, the number of RF chains and associated ADCs/DACs are much less than the number of antennas, i.e., L_{T} (number of TX RF chains) $\leq N_{\text{T}}$ (number of TX antennas) and L_{R} (number of RX RF chains) $\leq N_{\text{R}}$ (number of RX antennas) [42, 43]. Unlike the conventional beamforming architectures, the A/D HBF enables spatial multiplexing and multi-user communication that enhances the benefits of MIMO. There are several A/D hybrid transceiver solutions which have been recently proposed to enable mmWave MIMO systems [16, 17, 44]. Given the CSI, several algorithms can be designed for HBF approach to provide a capacity efficient system. Generally, beamforming at the TX can be referred to as precoding and at the RX as combining such as in [16, 17]. These precoders and combiners decompose into product of analog and digital matrices with different constraints. We can notice from the existing literature that the mmWave HBF MIMO systems can be implemented to provide satisfying rate performance by avoiding the discussed limitations of a fully digital solution [16, 17, 44].

Furthermore, we can reduce the power consumption by implementing low resolution quantization for both conventional and A/D HBF architectures. To that end some approaches have been applied for EE maximization such as in [26]. We will show later in Chapter 3 that optimizing the number of RF chains further leverages the Energy Efficiency (EE) metric and reduces the gap between the SE of A/D hybrid and fully digital beamforming architectures with high resolution sampling. Further in Chapters 4 and 5, we will study what happens when low resolution quantization can be implemented at both the TX and the RX. Optimizing bit resolution with the precoding and combining design can provide a highly energy efficient solution.

Fig. 2.8 shows the SE plot w.r.t. SNR for different beamforming approaches for TX antennas $N_{\mathrm{T}} = 64$, RX antennas $N_{\mathrm{R}} = 16$ and Number of TX/RX chains, $L_{\mathrm{T}} = L_{\mathrm{R}} = 4$. For the channel parameters, there are 10 rays for each cluster and there are 8 clusters in total, i.e., $N_{\mathrm{ray}} = 10$ and $N_{\mathrm{cl}} = 8$ in (2.15). The average power of each cluster is unity, i.e., $\sigma_{\alpha,i} = 1$. The azimuth and elevation angles of departure and arrival are computed on the basis of the Laplacian distribution with uniformly distributed mean angles and angle spread as 7.5°. The mean angles are sectored within the range of 60° to 120° in the azimuth domain, and 80° to 100° in the elevation domain. The antenna elements are spaced by distance $d = \lambda/2$ where $\lambda/2$ can be based on a standard frequency value such as 28 GHz. The system bandwidth is normalized to 1 Hz and the signal to noise ratio (SNR) is $1/\sigma_{\mathrm{n}}^2$. It can be observed that the HBF approach performs similar to the fully digital beamforming and better than the analog only beamforming. For example,

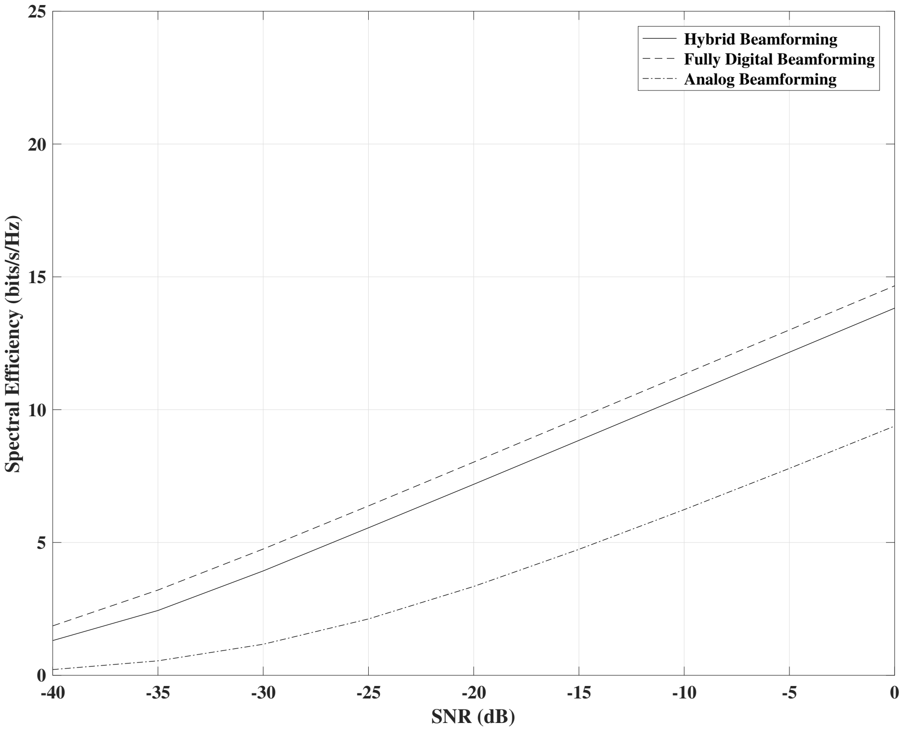


Figure 2.8: SE w.r.t. SNR for Conventional and Hybrid Beamforming Approaches for TX antennas $N_{\mathrm{T}} = 64$, RX antennas $N_{\mathrm{R}} = 16$ and Number of TX/RX chains, $L_{\mathrm{T}} = L_{\mathrm{R}} = 4$.

at 0 dB SNR, HBF has SE close to the fully digital beamforming and 5 bits/s/Hz better than the analog only beamforming. These plots are for high resolution sampling, however, in the following subsection we discuss about the advantages of using low resolution sampling in mmWave HBF MIMO systems.

2.2.3 Low Resolution Quantization

The DACs and ADCs associated with RF chains are power hungry components as well and the large number of antennas in mmWave MIMO systems make it hard to use many converting units [15]. The converting units with high bit resolution may achieve highly capacity efficient system but implementing low resolution quantization such as 1-bit to 3-bits can improve the EE of mmWave hybrid MIMO systems. Designing techniques for EE maximization but keeping high SE have been the main objective of this thesis which we will discuss later in the technical chapters. In the following we discuss the state of the art in ADCs and factors affecting the ADC performance.

State of the art in ADCs

Reference [45] discusses the developments in low power ADCs and factors impacting the ADC power efficiency. The system architecture and its performance is affected by the efficiency and speed of converting analog to digital digital signals. A very high conversion rate can be expected from the modern sampling devices but power dissipation is a key concern in mixed-signal or RF applications. For instance, the high-speed 6-8-bit ADCs achieve sampling rates in excess 20 GS/s, at power dissipations of 1.2 W and 10 W, respectively. To avoid draining battery of a device within a short span of time, designing ADCs and RF chains based on an available power budget, i.e., optimizing the power consumption associated with such power consuming devices, would lead to a power efficient consumer device. There are several surveys on ADC performance in the literature [46–48].

Recent developments in ADCs target mainly low to moderate resolution as the high resolution designs with signal-to-noise-and-distortion ratio (SNDR) $>$ 85 dB do not follow the implied 2x increase in power per bit [45]. Besides the power or energy efficiency of an ADC, the available signal bandwidth also proves to be an important parameter. Bandwidth versus SNDR for an ADC can be plotted and it can be observed that for all resolutions, the parts with highest bandwidth achieve a considerable performance [45]. Taking into account additional nonidealities such as quantization noise, thermal noise and differential non-linearity also impact the ADC performance. There is certainly a performance trade-off between the power efficiency and bandwidth, e.g. [49] achieves high bandwidth but average power efficiency and on the other hand, [50] shows high power efficiency with low bandwidth. Thus, designing ADCs for high speed limits will sacrifice on the power efficiency and vice-versa. Besides potentially increasing sampling speed by utilizing sub-circuits, the goal should be to improve the power efficiency with the use of low to moderate bit resolutions and optimizing the bit resolution depending upon the current need would maximize the power efficiency of such a system. In addition to the ADC performance and trends discussed in [45], up-to-date architectural trends and specifications affecting the ADC performance are discussed in [51]. Furthermore, [52, 53] provide discussion about RF technology for millimeter wave in 5G applications which may be useful in order to understand the power efficiency terms associated with the RF components of a HBF design.

For the case of 1-bit ADCs, there is negligible power consumption in comparison to the other circuit components. The communication fundamentals at 1-bit ADCs are different than the conventional full bit resolution sampling [54, 55]. From [55], we can notice that the low SNR capacity difference between 1-bit resolution sampling and infinite/full-bit resolution sampling is only 1.96 dB. While at high SNR values, maximum achievable rate is $2^{2N_{\mathrm{R}}}$ bits/s/Hz providing the rank of the channel is at least N_{R}, i.e., the number of RX antennas. There are several implications of using 1-bit or low resolution sampling and there is a need of developing different HBF optimization solutions which take into account the low resolution sampling such as performed in Chapters 4 and 5. In addition estimating mmWave CSI with 1-bit ADCs at the RX is a challenging problem which has been addressed in part of Chapter 4 of this thesis.

Next, we discuss the power model for both full resolution and low resolution sampling cases used in the following chapters in the thesis.

2.2.4 Power Model for the HBF Architecture

Measuring the energy consumed for each hardware entity in the HBF architecture plays an important role when designing an energy efficient mmWave A/D hybrid MIMO system. Following [14, 56] total power P for a A/D HBF system with a fully-connected structure and full resolution sampling as discussed in Chapter 3 later can be described as follows, where we include the power consumed by the RX components as well:

$$\begin{aligned} P = \beta \mathrm{tr}(\mathbf{P}_{\mathrm{TX}}) + 2P_{\mathrm{CP}} + N_{\mathrm{T}}P_{\mathrm{T}} + N_{\mathrm{R}}P_{\mathrm{R}} + L_{\mathrm{T}} \times \\ (P_{\mathrm{RF}} + N_{\mathrm{T}}P_{\mathrm{PS}}) + L_{\mathrm{R}}(P_{\mathrm{RF}} + N_{\mathrm{R}}P_{\mathrm{PS}}) \ \text{(W)}, \end{aligned} \tag{2.17}$$

where β represents the reciprocal of amplifier efficiency; the common parameters at the TX and the RX are P_{CP}, P_{RF}, and P_{PS} which represent the circuit power, i.e., is the power required by all circuit components at the TX, the power per RF chain, and the power per phase shifter, respectively. P_{T} and P_{R} represent the power per antenna element at the TX and the RX, respectively. Other entities can be noted from the description of Fig. 2.7, such as L_{T} and L_{R} being the number of RF chains at the TX and the RX, respectively, and N_{T} and N_{R} being the number of antennas at the TX and the RX, respectively.

For instance, from (2.17), we can observe the variation of P w.r.t. the term $\mathrm{tr}(\mathbf{P}_{\mathrm{TX}})$ which represents the transmit power constraint and $\mathbf{P}_{\mathrm{TX}}$ is a diagonal matrix of power allocation values with $\mathrm{tr}(\mathbf{P}_{\mathrm{TX}}) = P_{\max}$, where $P_{\max}$ is the maximum allocated power. Fig. 2.9 shows that variation of P w.r.t. $\mathrm{tr}(\mathbf{P}_{\mathrm{TX}})$

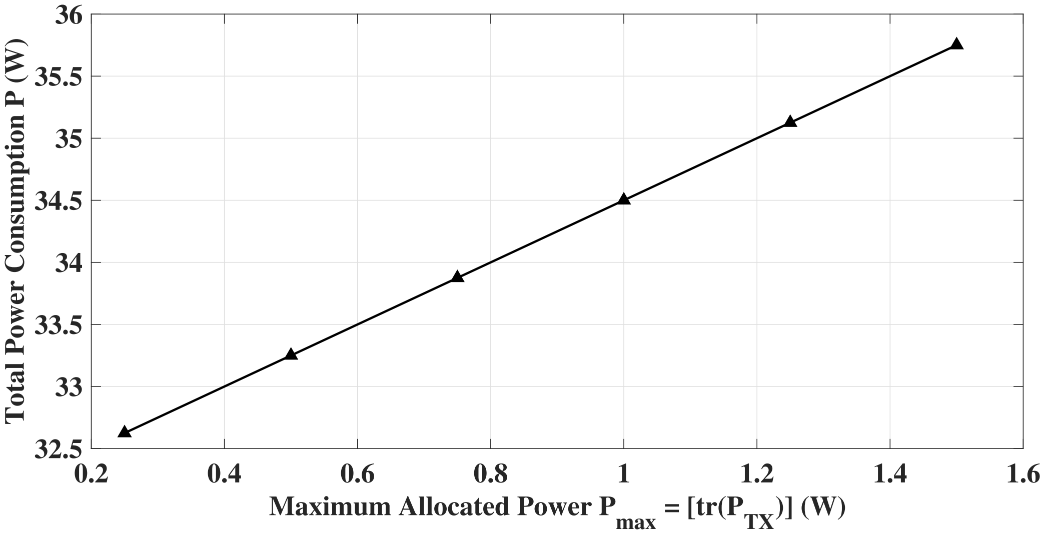


Figure 2.9: Total power consumption P versus transmit power constraint $\mathrm{tr}(\mathbf{P}_{\mathrm{TX}})$.

which is a linear relationship between these terms, e.g., at $P_{\mathrm{max}} = 1\mathrm{W}$, the value of the total power consumption P is 34.5 W. The typical simulation values for fixed power terms and system parameters are provided in Table 2.1. Note that we provide further discussion about the terms used in the power model such as in (2.17) in the following chapters.

In the case of low resolution quantization at both the TX and the RX as discussed later in Chapter 5, the total power consumption can be expressed as

$$P \triangleq P_{\mathrm{TX}}(\mathbf{F}_{\mathrm{RF}}, \mathbf{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}}) + P_{\mathrm{RX}}(\mathbf{\Delta}_{\mathrm{RX}}) \text{ (W)}, \tag{2.18}$$

where the matrices $\mathbf{\Delta}_{\mathrm{TX}}$ and $\mathbf{\Delta}_{\mathrm{RX}}$ represent diagonal matrices with values depending on the bit resolution of each DAC and ADC, respectively. The matrix $\mathbf{F}_{\mathrm{BB}}$ denotes the baseband precoder matrix which has dimensions of $L_{\mathrm{T}} \times N_{\mathrm{s}}$ (N_{s} being the number of streams) using its L_{T} transmit chains and $\mathbf{F}_{\mathrm{RF}}$ denotes the RF precoder matrix which has dimensions of $N_{\mathrm{T}} \times L_{\mathrm{T}}$ using the phase shifting

Power Terms	**Values**
Circuit power of the TX	$P_{\text{CP}} = 10$ W
Power per RF chain	$P_{\text{RF}} = 100$ mW
Power per phase shifter	$P_{\text{PS}} = 10$ mW
Power per antenna at the TX/RX	$P_{\text{T}} = P_{\text{R}} = 100$ mW

(a) Typical values of the power terms.

System Parameters	**Values**
Number of TX antennas	$N_{\text{T}} = 64$
Number of RX antennas	$N_{\text{R}} = 16$
Number of TX/RX RF chains	$L_{\text{T}} = L_{\text{R}} = 4$
Reciprocal of amplifier efficiency	$\beta = 1/0.4$

(b) System parameter values.

Table 2.1: Simulation parameter values to compute power consumption P in (2.17).

network. Specifically, each diagonal entry of $\mathbf{\Delta}_{\text{TX}}$ is given by:

$$[\mathbf{\Delta}_{\text{TX}}]_{ii} = \sqrt{1 - \frac{\pi\sqrt{3}}{2} 2^{-2b_i^t}} \in [m, M] \, \forall\, i = 1, \ldots, L_{\text{T}}, \tag{2.19}$$

and each diagonal entry of $\mathbf{\Delta}_{\text{RX}}$ is given by:

$$[\mathbf{\Delta}_{\text{RX}}]_{ii} = \sqrt{1 - \frac{\pi\sqrt{3}}{2} 2^{-2b_i^r}} \in [m, M] \, \forall\, i = 1, \ldots, L_{\text{R}}, \tag{2.20}$$

where in the following thesis, for simplicity, we assume that the range $[m, M]$ is the same for each of the DACs/ADCs. The resolution parameter b is denoted as $b_i^t \,\forall\, i = 1, \ldots, L_{\text{T}}$ and $b_i^r \,\forall\, i = 1, \ldots, L_{\text{R}}$ at the TX and the RX, respectively. The power consumption at the TX is as follows:

$$P_{\text{TX}}(\mathbf{F}_{\text{RF}}, \mathbf{\Delta}_{\text{TX}}, \mathbf{F}_{\text{BB}}) = \text{tr}(\mathbf{F}\mathbf{F}^H) + P_{\text{DT}}(\mathbf{\Delta}_{\text{TX}}) + N_{\text{T}}P_{\text{T}} + N_{\text{T}}L_{\text{T}}P_{\text{PT}} + P_{\text{CT}} \text{ (W)}, \tag{2.21}$$

where P_{PT} is the power per phase shifter, P_{T} is the power per antenna element,

$P_{\mathrm{DT}}(\boldsymbol{\Delta}_{\mathrm{TX}})$ is the power associated with the total quantization operation at the TX, and following (2.19) and [57], we have

$$P_{\mathrm{DT}}(\boldsymbol{\Delta}_{\mathrm{TX}}) = P_{\mathrm{DAC}} \sum_{i=1}^{L_{\mathrm{T}}} 2^{b_i^t} = P_{\mathrm{DAC}} \sum_{i=1}^{L_{\mathrm{T}}} \left(\frac{\pi\sqrt{3}}{2(1-[\boldsymbol{\Delta}_{\mathrm{TX}}]_{ii}^2)} \right)^{\frac{1}{2}} \text{ (W)}, \tag{2.22}$$

where P_{DAC} is the power consumed per bit in the DAC and P_{CT} is the power required by all circuit components at the TX. Similarly, the total power consumption at the RX is,

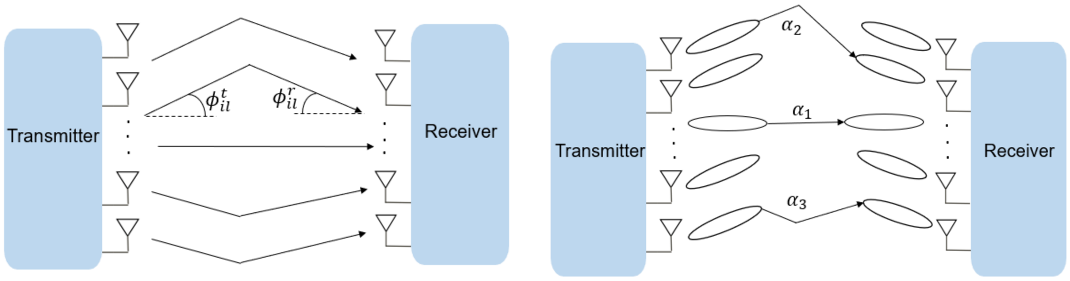
$$P_{\mathrm{RX}}(\boldsymbol{\Delta}_{\mathrm{RX}}) = P_{\mathrm{DR}}(\boldsymbol{\Delta}_{\mathrm{RX}}) + N_{\mathrm{R}} P_{\mathrm{R}} + N_{\mathrm{R}} L_{\mathrm{R}} P_{\mathrm{PR}} + P_{\mathrm{CR}} \text{ (W)}, \tag{2.23}$$

where, at the RX, P_{PR} is the power per phase shifter, P_{R} is the power per antenna element, P_{DR} is the power associated with the total quantization operation, and following (2.20) and [57], we have

$$P_{\mathrm{DR}}(\boldsymbol{\Delta}_{\mathrm{RX}}) = P_{\mathrm{ADC}} \sum_{i=1}^{L_{\mathrm{R}}} 2^{b_i^r} = P_{\mathrm{ADC}} \sum_{i=1}^{L_{\mathrm{R}}} \left(\frac{\pi\sqrt{3}}{2(1-[\boldsymbol{\Delta}_{\mathrm{RX}}]_{ii}^2)} \right)^{\frac{1}{2}} \text{ (W)}, \tag{2.24}$$

where P_{ADC} is the power consumed per bit in the ADC and P_{CR} is the power required by all RX circuit components. Similar to Fig. 2.9, we can observe the variation of total power consumption P in (2.18) for the low resolution quantization case for different parameter settings.

In the next two sections, we discuss the basics of convex optimization and compressed sensing approaches which are useful in developing efficient algorithms for mmWave HBF MIMO systems with both full and low resolution sampling.



(a) Narrowband Channel Model in (2.15).

(b) Sparse Channel Model in (2.16).

Figure 2.10: Sparsity Characteristics of a MmWave Channel.

2.3 Overview of Convex Optimization

A general form of optimization problems is given in the following equation:

$$\begin{aligned} &\min \ f_0(\mathbf{x}) \\ &\text{subject to } f_i(\mathbf{x}) \leq a_i, \ \ i = 1, ..., m, \end{aligned} \tag{2.25}$$

where vector $\mathbf{x}$ is the optimization variable, $f_0 : \mathbf{R}^n \rightarrow \mathbf{R}$ is an objective function and the functions f_i represent constraints on the optimization problem with a_i as the limits/bounds $\forall i = 1, .., m$. A solution of the problem in (2.25) is obtained when an optimal vector $\hat{\mathbf{x}}$ has the smallest objective value among all vectors that satisfy the constraints, i.e., for any $\mathbf{b}$ with $f_1(\mathbf{b}) < a_1$,...,$f_m(\mathbf{b}) < a_m$ we have $f_0(\mathbf{b}) \geq f_0(\hat{\mathbf{x}})$. A solution to the optimization problem in (2.25) corresponds to an optimal choice that has minimum cost or in some cases, maximum utility among all the choices that meet the constraint requirements.

Signal processing algorithms play a vital role in solving the optimization problems such as in (2.25). The effectiveness of these algorithms depends on the objective, constraint functions, number of variables and constraints and sparsity. A sparse problem is one where each constraint function depends on only a small number of the variables [58]. In terms of mmWave channel in (2.15), the number of clusters and rays is small, thus the beamspace representation of narrowband channel in (2.16) includes a sparse matrix $\mathbf{H}_{\mathrm{v}}$ which has few non-zero entries. The sparse nature of the MIMO channel at mmWave is represented by the sparse nature of the beamspace channel matrix $\mathbf{H}_{\mathrm{v}}$. The DFT matrices in (2.16) correspond to the array response vectors with virtual angle of arrivals and angle of departures corresponding to the uniformly spaced normalized angles. Fig. 2.10 shows the sparsity characteristics of mmWave channel where communication with a narrowband channel model such as in (2.15) is shown in Fig. 2.10 (a), and Fig. 2.10 (b) shows the sparsity through a few non-zero entries of the sparse matrix $\mathbf{H}_{\mathrm{v}} \in \mathbb{C}^{L_{\mathrm{R}} \times L_{\mathrm{T}}}$ in (2.16). In addition to the sparse mmWave channel, we mainly focus on mathematical convex optimization problems in the following chapters. Reference [58] suggests that we can easily solve optimization problems with many variables and constraints and by exploiting the problem's structure, such as sparsity in the case of a mmWave channel, we can solve far larger problems with many more variables and constraints.

A convex optimization problem can be written as follows:

$$\begin{aligned} &\min \;\; f_0(\mathbf{x}) \\ &\text{subject to } f_i(\mathbf{x}) \leq a_i, \;\; i = 1, ..., m, \end{aligned} \tag{2.26}$$

where constraint functions $f_i \,\forall\, i = 1, ..., m$, are convex which satisfy

$$f_i(\alpha\mathbf{x} + \beta\mathbf{y}) \leq \alpha f_i(\mathbf{x}) + \beta f_i(\mathbf{y}) \,\forall\, \alpha, \beta \in \mathbf{R}, \tag{2.27}$$

with $\alpha + \beta = 1$, $\alpha \geq 0$ and $\beta \geq 0$. This expression in (2.26) is a general convex optimization problem, and least squares and linear programming problems are special cases of this problem [58]. There are several reliable approaches to solve convex optimization problems such as interior point methods [58], however, solving non-linear convex optimization problems need to be solved more carefully due to the non-linear nature of such problems.

In order to address non-linear optimization, we consider *local* optimization and *global* optimization methods. In *local* optimization, we aim to seek a point which is *locally optimal* which means that it minimizes the objective function among all feasible points nearby, but is not guaranteed to have a lower objective value than all other feasible points. The main drawback of finding local optima is requiring an accurate initial guess for the optimization variable and the choice of algorithm and its parameters also effect the solution of such a problem. While in *global* optimization, we seek to find the true global solution of the optimization problem such as for (2.25), but the cost is computation time which can be prohibitively large even for small number of parameters.

Convex optimization plays a vital role even when the problem is non-convex. Firstly, we can combine a local optimization method with convex optimization. To begin with a non convex problem can be converted into an approximate convex problem which can be solved exactly without an initial guess. This point can be used as the starting point for a local optimization method that is applied to the original non convex problem. Furthermore, we can consider convex optimization for a sparse problem such as when $\mathbf{x}$ is a sparse vector with few non-zero entries in (2.26) that satisfies some constraints. Global optimization methods require a less computationally complex lower bound on the optimal value of the non convex problem. We can use *relaxation* where each non convex constraint is replaced with a less strict convex constraint, or *Lagrangian relaxation* where the problem, i.e., Lagrangian dual problem [58], is convex and provides a lower bound on the optimal value of the non convex problem.

It is also worth noting that we can express the maximization optimization problem as follows:

$$\begin{aligned} &\max \;\; f_0(\mathbf{x}) \\ &\text{subject to } f_i(\mathbf{x}) \leq 0, \;\; i = 1, ..., m, \\ &\qquad\qquad h_i(\mathbf{x}) = 0, \;\; i = 1, ..., n, \end{aligned} \tag{2.28}$$

which can be solved by minimizing the function $-f_0$ subject to the given constraints. For example, the optimal value of (2.28) can be expressed as

$$\hat{\mathbf{x}} = \sup\{f_0(\mathbf{x}) \,|\, f_i(\mathbf{x}) \leq 0, \, i = 1, ..., m; \, h_i(\mathbf{x}) = 0, \, i = 1, ..., n\}, \tag{2.29}$$

where sup (or supremum) refers to the largest value and a feasible point $\mathbf{x}$ is $\boldsymbol{\epsilon}$-suboptimal (where a $\boldsymbol{\epsilon}$-suboptimal set refers to the set of feasible points with objective value within $\boldsymbol{\epsilon}$ of optimal) if $f_0(\mathbf{x}) \geq \hat{\mathbf{x}} - \boldsymbol{\epsilon}$. We make use of the maximization optimization problems in the following chapters where the EE ratio, $EE = R\,(\text{bits/s/Hz})/P\,(\text{W})$, is required to be maximized based on given hardware constraints on rate R and total power consumption P. The expressions of rate and power contain the matrices related to the system hardware which are constrained and the EE optimization problems containing these expressions can be maximized or minimized using concepts of convex optimization in order to achieve maximum EE. In the next section, we proceed with the basics of CS methods which we use in the following chapters to solve optimization problems.

2.4 Overview of Compressed Sensing

The fundamental idea behind CS is that instead of compressing the sampling data that is sampled at a high rate, there is a need to directly sense the data in a compressed form at a lower sampling rate. Thus CS has large implications in signal processing fields such as medical imaging, sensor networks and sub-Nyquist sampling systems [59]. Moreover, CS techniques also have applications in mobile communication systems.

A basic mathematical equation for CS can be expressed as follows:

$$\mathbf{A}\mathbf{x} = \mathbf{y}, \tag{2.30}$$

where the observed data $y \in \mathbb{C}^m$ is connected to the signal $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{A} \in \mathbb{C}^{m \times N}$ models the linear measurement process. By solving the linear system in (2.30), we try to recover the vector $\mathbf{x}$. Note that the number of measurements $m \geq$ signal length N, otherwise the linear system in (2.30) is under-determined and there exist infinitely many solutions, i.e., without additional information it is impossible to recover $\mathbf{x}$ from $\mathbf{y}$ in this case. However, under certain assumptions, it is possible to reconstruct signals using efficient algorithms when $m < N$, such as in the case of sparsity [60]. If a signal is sparse in nature, it means that there are less unknowns and CS algorithms can be implemented to reconstruct such a signal. For example, in terms of mmWave channel estimation (discussed in detail in Chapter 4), we would need fewer number of training symbols to obtain the channel as at mmWave, MIMO channel is sparse in nature. The main problem exists in the determining the locations of the non-zero entries of the vector $\mathbf{x}$ which are not known a priori. The essential points to be discussed for applying

CS for (2.30) is to define/design matrix $\mathbf{A}$ and reconstructing $\mathbf{x}$ efficiently. Note that matrix $\mathbf{A}$ should ideally be designed for all signal samples $\mathbf{x}$ simultaneously.

The algorithmic approach l_0-minimization is the most basic CS approach, where we can reconstruct $\mathbf{x}$ as a solution of the following optimization problem:

$$\min \|\mathbf{z}\|_0 \;\text{ subject to } \mathbf{Az} = \mathbf{y}, \tag{2.31}$$

where we search for the sparsest vector consistent with the measured data $\mathbf{y} = \mathbf{Ax}$. A more popular approach is called as l_1-minimization or basis pursuit, where we aim to find the minimizer of the following problem:

$$\min \|\mathbf{z}\|_1 \;\text{ subject to } \mathbf{Az} = \mathbf{y}, \tag{2.32}$$

where l_1 norm, i.e. $\|.\|$, is a convex function which can be solved by efficient methods from convex optimization, discussed in the previous section. This basis pursuit technique can be interpreted as convex *relaxation* of l_0-minimization method. Furthermore, there are iterative hard thresholding method and greedy startegies such as Orthogonal Matching Pursuit (OMP) and Gradient Pursuit (GP) [61–63] to recover sparse vectors, which we discuss in more detail in the following chapters.

Specifically in terms of sparse approximation, let us form the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with columns $\mathbf{a}_1, ..., \mathbf{a}_N$, then solving (2.31) provides the sparsest representation of $\mathbf{y}$. Further tolerating a representation error, say η, the optimization problem in (2.31) can be written as

$$\min \|\mathbf{z}\|_0 \;\text{ subject to }\; \|\mathbf{Az} - \mathbf{y}\| \leq \eta. \tag{2.33}$$

The equation (2.33) is NP-hard in general, i.e., non-deterministic polynomial-time hardness, and all the standard CS algorithms, including l_1-minimization can be applied in this context as well. Moreover, the conditions on $\mathbf{A}$ remain valid which ensures that the sparsest vector $\mathbf{x}$ is recovered exactly or approximately. However, a note worthy difference between CS and sparse approximation is that we are interested in computing the error $\|\mathbf{x} - \hat{\mathbf{x}}\|$ (where $\hat{\mathbf{x}}$ is reconstructed vector) in CS, whereas in sparse approximation, we are interested in computing $\|\mathbf{y} - \hat{\mathbf{y}}\|$ and aim to approximate given $\mathbf{y}$ with a sparse expansion $\hat{\mathbf{y}} = \sum_j \hat{x}_j \mathbf{a}_j$.

While computing sparse representations, convex optimization techniques play a key role, however, greedy strategies can also be used to solve such problems. As mentioned above, OMP and GP are the standard examples of greedy methods. Other approaches such as matching pursuit, conjugate gradient pursuit and order recursive matching pursuit [59, 63] can also be implemented while dealing with sparse approximation problems in mmWave hybrid MIMO systems. These algorithms rely on iterative approximation of signal coefficients and support, either by iteratively identifying the support of that signal until its convergence or by iteratively obtaining an improved estimate of the sparse signal.

Some greedy methods can have similar performance as that of the convex optimization algorithms. The state of the art OMP algorithm, used to compute precoders and combiners in mmWave hybrid MIMO systems, begins by finding the column of matrix most correlated with the measurements. It then repeats this step by correlating the columns with the signal residual obtained by subtracting the contribution of a partial estimate of the signal from the original measurement vector. The stopping criterion for this algorithm can be a limit on the number of iterations, for example, number of RF chains in HBF MIMO systems. For exactly k-sparse $\mathbf{x}$ with noise-free measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$, OMP will recover x

exactly in k iterations [59]. A good alternative to OMP is GP which is described in detail in [59,63] and we implement it in Chapter 3 in the context of mmWave HBF MIMO systems as a faster approximation solution and with lower complexity than the state of the art OMP algorithm. In Chapters 4 and 5, we also take into account low resolution sampling in the optimization problems and solve them efficiently with suitable convex optimization and CS approaches. We provide step-by-step details on these approaches in these chapters. As a starting point, [64] can be studied to understand 1-bit CS and related reconstruction algorithms.

2.5 Summary

In this chapter, firstly we discussed AWGN channel capacity and its relation to SNR per degree of freedom. We then proceeded with the derivation of MIMO channel capacity equation and waterfilling power allocation. The benefits of MIMO communication are discussed with plot of capacity versus SNR where more TX/RX antennas resulted into higher capacity. Then the advantages of MIMO with high frequency mmWave technology, mmWave channel and beamspace representation of mmWave channel were discussed.

We then proceeded with the advantages of using a HBF approach over conventional beamforming approaches mainly in terms of hardware complexity and power consumption. A SE versus SNR plot showed that HBF approach shows higher SE performance than the conventional approaches for given simulation parameters. We also discussed the use of low-resolution quantization in mmWave MIMO systems with HBF architecture. We then discussed the overview of convex optimization while focussing on the basics of such problems, non-linear optimization methods and sparsity of mmWave channels. A brief overview of CS approaches was also discussed with the basic mathematical equations, sparse approximation and related algorithmic approaches such as greedy strategies, for example, OMP algorithm.

In the following three contribution chapters, we use these basic concepts and equations discussed in this chapter. Firstly, we begin with a RF chain selection problem for EE maximization with full-bit resolution sampling and in the next two chapters, we introduce low resolution sampling in these systems.

Chapter 3

EE Maximization by Dynamic RF Chain Selection with Low Complexity Hybrid Beamforming

3.1 Introduction

The performance of mmWave MIMO systems can be significantly improved through the use of Analog/Digital (A/D) HBF architectures where the number of RF chains and associated Analog-to-Digital Converter (ADC) and Digital-to-Analog Converter (DAC) are much less than the number of antennas [42, 43]. The hardware complexity and power consumption is reduced through using fewer RF chains and it can still support multi-stream communication with high performance in terms of the achieved SE [14, 16–23]. Such systems can also be optimized to achieve high EE gains [24, 26, 27, 44, 65, 66].

To implement the A/D HBF system which uses RF precoders based on the phase shifting networks, we can use the most popular structures such as the fully-connected and the partially-connected configurations. The fully-connected structure connects all the antennas to each RF chain whereas the partially-connected structure connects only a subset of the antennas requiring fewer phase shifters [56]. The use of a partially-connected structure at the transceiver can further reduce the power consumption [44], for example, [66] uses a partially-connected structure to evaluate the energy and rate performance where the partially-connected structure is optimized to achieve high EE. This chapter mainly uses the fully-connected structure to demonstrate the contributions of the proposed techniques for a mmWave hybrid MIMO system. However, the EE performance using the partially-connected structure is also observed via simulations.

An alternative solution to reduce the power consumption and hardware complexity is by reducing the bit resolution [15] of the DACs and the ADCs. Some approaches have been applied in A/D hybrid mmWave MIMO systems for EE maximization with low resolution sampling [26,27,66]. For EE maximization, [26] selects the best subset of the active RF chains and the DAC resolution using Dinkelbach Method (DM) and subset selection optimization approach, [27] proposes to jointly optimize the ADC bit resolution and A/D hybrid combiner matrices and [66] implements low resolution DACs with the number of RF chains optimization.

Reference [67] makes use of switches and phase shifters to execute analog beamforming for the A/D hybrid model, and then the EE and SE performance is investigated. Given the distinct system and channel model characteristics at mmWave compared to microwave, EE and SE performance needs to be analyzed for the A/D HBF architecture with both high resolution and low resolution sampling cases. Firstly in this chapter, we proceed with the mmWave channel and hybrid MIMO system model and then we discuss EE maximization by optimizing the number of RF chains for a full resolution sampling case. In the following chapters, we discuss channel estimation and EE maximization for the low resolution sampling cases based on this channel and system model. Next, we describe the model for the mmWave A/D HBF MIMO system, and based on this a literature review and the main contributions of this chapter are presented.

Let us consider a single user MIMO system with N_{T} antennas at the TX, sending N_{s} data streams to a system with N_{R} RX antennas. The fading channel models used in traditional MIMO become inaccurate for mmWave channel modeling due to the high free-space path loss and large tightly-packed antenna arrays. The existing literature mostly addresses the narrowband clustered channel model [36, 37] for mmWave propagation due to different channel settings such as number of multipaths, amplitudes, etc. such as in [15, 16].

For N_{cl} clusters and N_{ray} propagation paths in each cluster and for a ULA antenna elements, the mmWave channel matrix is defined as follows:

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{N_{\mathrm{cl}}N_{\mathrm{ray}}}} \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{l=1}^{N_{\mathrm{ray}}} \alpha_{il}\mathbf{a}_{\mathrm{R}}(\phi_{il}^{r})\mathbf{a}_{\mathrm{T}}(\phi_{il}^{t})^{H}, \tag{3.1}$$

where α_{il} denotes the gain of l-th ray in i-th cluster and it is assumed that α_{il} are i.i.d. $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$, where $\sigma_{\alpha,i}^2$ is average power of the i-th cluster such that $\sum_{i=1}^{N_{\mathrm{cl}}} \sigma_{\alpha,i}^2 = \gamma$, where $\gamma = \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{N_{\mathrm{cl}}N_{\mathrm{ray}}}}$, is the normalization factor satisfying $\mathbb{E}\{\|\mathbf{H}\|_F^2\} = 1/\sqrt{N_{\mathrm{cl}}N_{\mathrm{ray}}}$. Further, $\mathbf{a}_{\mathrm{R}}(\phi_{il}^{r})$ and $\mathbf{a}_{\mathrm{T}}(\phi_{il}^{t})$ represent the normalized receive and transmit array response vectors, where ϕ_{il}^{t} and ϕ_{il}^{r} are the azimuth angles of departure and arrival, respectively. The antenna elements at the TX and the RX can be modeled as ideal sectored elements [35] and then antenna element gains can be evaluated over ideal sectors. In (3.1), the transmit and receive antenna element gains are considered unity over ideal sectors defined by $\phi_{il}^{t} \in [\phi_{\min}^{t}, \phi_{\max}^{t}]$ and $\phi_{il}^{r} \in [\phi_{\min}^{r}, \phi_{\max}^{r}]$, respectively. For a N_{Z}-element ULA on Z-axis, the array response vector can be expressed as [34]: $\mathbf{a}_{\mathrm{Z}}(\phi) = \frac{1}{\sqrt{N_{\mathrm{Z}}}} e^{jm\frac{2\pi}{\lambda}d\sin(\phi)^{T}}$, where $0 \leq m \leq (N_{\mathrm{Z}} - 1)$ is a real integer, d is the inter-element spacing in

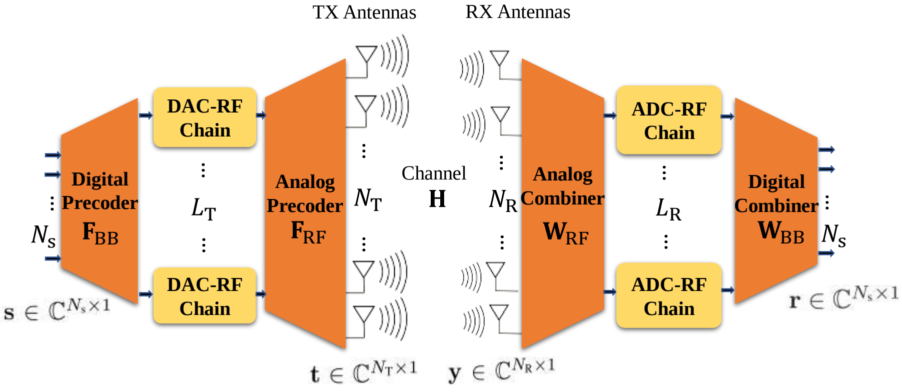

TX Antennas
RX Antennas
Digital Precoder $\mathbf{F}_{\mathrm{BB}}$
N_{s}
$\mathbf{s} \in \mathbb{C}^{N_{\mathrm{s}}\times 1}$
DAC-RF Chain
L_{T}
DAC-RF Chain
Analog Precoder $\mathbf{F}_{\mathrm{RF}}$
N_{T}
$\mathbf{t} \in \mathbb{C}^{N_{\mathrm{T}}\times 1}$
Channel $\mathbf{H}$
N_{R}
$\mathbf{y} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$
Analog Combiner $\mathbf{W}_{\mathrm{RF}}$
ADC-RF Chain
L_{R}
ADC-RF Chain
Digital Combiner $\mathbf{W}_{\mathrm{BB}}$
N_{s}
$\mathbf{r} \in \mathbb{C}^{N_{\mathrm{s}}\times 1}$

Figure 3.1: Block diagram of a mmWave MIMO system with HBF architecture.

wavelengths and λ is the signal wavelength. The array response vectors can also be computed using other array geometries such as rectangular array and circular array. Note that, we assume perfect channel knowledge at the TX and the RX [16,44,65] for the EE maximization work and consider channel estimation errors in Chapter 4 when proposing an efficient channel estimation algorithm.

The beamspace representation [38, 39] of the narrowband channel can be written as follows:

$$\mathbf{H} = \mathbf{D}_{\mathrm{R}}\mathbf{H}_{\mathrm{v}}\mathbf{D}_{\mathrm{T}}^{H}, \tag{3.2}$$

where $\mathbf{H}_{\mathrm{v}} \in \mathbb{C}^{L_{\mathrm{R}}\times L_{\mathrm{T}}}$ is a sparse matrix with a few non-zero entries, $\mathbf{D}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}}\times L_{\mathrm{R}}}$ and $\mathbf{D}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}$ are the Discrete Fourier transform (DFT) matrices.

In large-scale MIMO communication systems, based on the A/D hybrid precoding scheme, the number of RF chains is larger than or equal to the number of baseband data streams and smaller than or equal to the number of TX antennas. L_{T} denotes the number of available RF chains at the TX with the limitation that $N_{\mathrm{s}} \leq L_{\mathrm{T}} \leq N_{\mathrm{T}}$ and similarly L_{R} is for the RX with the condition $N_{\mathrm{s}} \leq L_{\mathrm{R}} \leq N_{\mathrm{R}}$. We consider the number of RF chains at the RX to be same as at the TX, i.e., $L_{\mathrm{R}} = L_{\mathrm{T}}$. Fig. 3.1 shows the block diagram of a mmWave single user fully-connected A/D HBF MIMO system with digital baseband precoding and associated switches, followed by RF chains and associated DACs, and constrained RF precoding implemented using phase shifters network at the TX, and vice-versa at the RX. This basic system setup can be considered with upgrades for both full resolution as shown in this chapter and low resolution cases as shown in the following two chapters.

The matrix $\mathbf{F}_{\mathrm{BB}}$ denotes the baseband precoder matrix which has dimensions of $L_{\mathrm{T}} \times N_{\mathrm{s}}$ using its L_{T} transmit chains and $\mathbf{F}_{\mathrm{RF}}$ denotes the RF precoder matrix which has dimensions of $N_{\mathrm{T}} \times L_{\mathrm{T}}$ using the phase shifting network. Similarly at the RX, the matrices $\mathbf{W}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{RF}}$ denote the $L_{\mathrm{R}} \times N_{\mathrm{s}}$ baseband combiner and the $N_{\mathrm{R}} \times L_{\mathrm{R}}$ RF combiner, respectively. The TX symbol vector $\mathbf{s} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ is such that $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \frac{1}{N_{\mathrm{s}}}\mathbf{I}_{N_{\mathrm{s}}}$. All elements of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ are of constant modulus. The power constraint at the TX is satisfied by $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = P_{\max}$, where $P_{\max}$ is the maximum allocated power. We assume a unit magnitude and continuous phase constraint on the phase shifters [16, 44].

Consider a narrowband propagation channel with $\mathbf{H}$ as the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ channel matrix as shown in (3.1), which is assumed to be known to both the TX and the RX, then the received signal can be expressed as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{n}, \tag{3.3}$$

where $\mathbf{y}$ is the $N_{\mathrm{R}} \times 1$ received vector and $\mathbf{n}$ is a $N_{\mathrm{R}} \times 1$ noise vector with entries which are modeled as Independent and Identically Distributed (I.I.D.) $\mathcal{CN}(0, \sigma_{\mathrm{n}}^2)$. After the application of the combining matrices, the received signal can be written as follows:

$$\tilde{\mathbf{y}} = \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{y} = \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{n}. \tag{3.4}$$

In the following subsection, we discuss the literature review related to the mmWave A/D HBF MIMO system design and our contributions in this chapter.

3.1.1 Literature Review

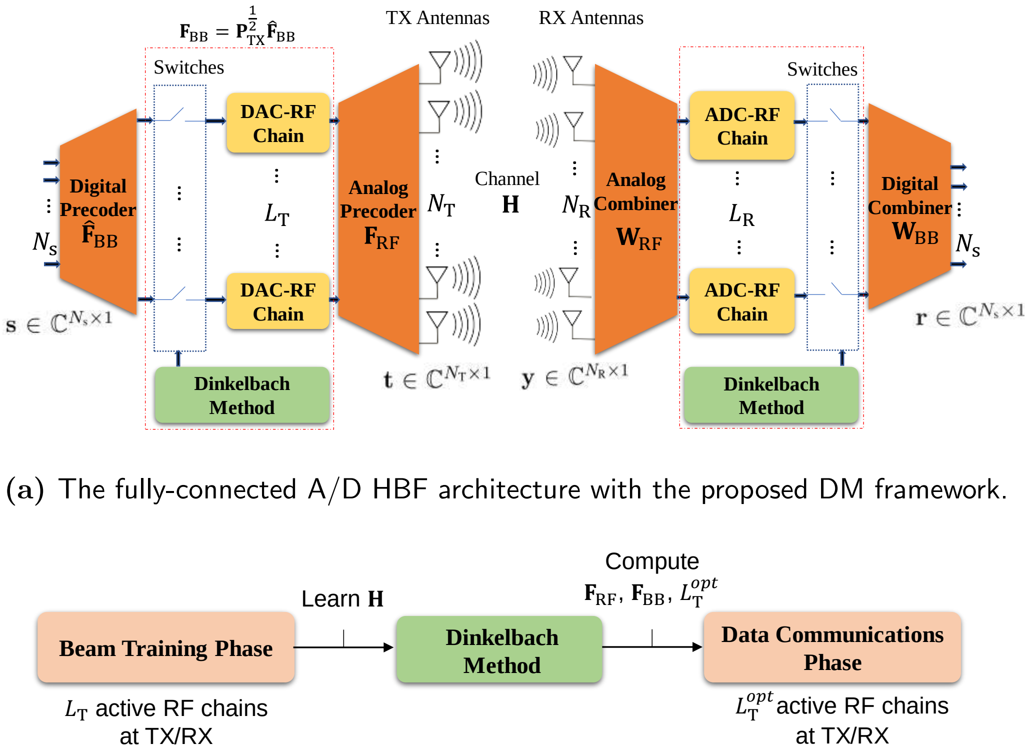
Reference [16] proposes a spectrally efficient A/D hybrid precoder design by maximizing the desired rate for fully-connected limited RF chain systems. However, it does not consider the energy consumption. For an energy efficient system, [68] considers a sub-connected architecture, where each RF chain is connected to only a subset of the TX antennas requiring fewer phase shifters, but it does not discuss how to design an energy efficient precoder with a fully-connected architecture. Reference [56] considers both fully-connected and partially-connected structures to design a A/D hybrid precoder where the partially-connected structure seems to outperform fully-connected structure in terms of EE. However, it only considers the design of the precoder matrices and there is no emphasis on optimizing the number of RF chains which is a key factor for an energy efficient system.

The RF chains consume a large amount of power in wireless communication systems and increase the cost for these systems [69]. Reference [65] performs an energy efficient optimization to design a A/D hybrid precoder where to calculate the optimal number of RF chains, the full precoding solution is computed for all possible numbers of RF chains. This is referred to as the Brute Force (BF) technique throughout in this chapter. References [16] and [65] use OMP to optimize the precoder matrices. Alternative greedy strategies to OMP can be exploited to lower the complexity. A mmWave A/D hybrid MIMO system can be used for 5G mmWave MIMO applications such as cellular backhaul connections when we jointly optimize the number of RF chains and the A/D hybrid precoder and combiner matrices leading to a highly energy efficient system.

3.1.2 Contributions

This chapter proposes an energy efficient A/D HBF framework, where the RF precoder and baseband precoder matrices, and RF combiner and baseband combiner matrices are optimized along with the number of active RF chains but with low complexity. We use power allocation, and the DM is implemented to optimize the number of RF chains. Fig. 3.2 shows the novel architecture with proposed framework for a mmWave fully-connected A/D HBF MIMO system which is an update to the system setup shown in Fig. 3.1. For this architecture, let $\mathbf{F}_{\mathrm{BB}} = \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\mathrm{BB}}$ denotes the baseband precoder matrix which inputs to the DAC-RF chain block where $\mathbf{P}_{\mathrm{TX}} \in \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ is a diagonal matrix of power allocation values with $\mathrm{tr}(\mathbf{P}_{\mathrm{TX}}) = P_{\max}$, $\hat{\mathbf{F}}_{\mathrm{BB}}$ is the digital precoding matrix before the switches, and $\mathbf{F}_{\mathrm{RF}}$ denotes the RF precoder matrix. In this novel architecture, for a certain number of RF chains implemented in the hardware, the DM block drives digital switches to activate only those RF chains that we obtain as an optimal solution from the proposed method. In practice the digital switches would be a part of the digital processor. If the DM block is replaced by another method used to optimize the number of RF chains, the number of active RF chains and associated DACs/ADCs may be different.

To compute the A/D hybrid precoders and combiners, the proposed approach incorporates a codebook-based approach through one of the greedy strategies, i.e., GP [63]. Simulations show that the proposed GP-based approach is a faster and less complex approach to compute the precoder and combiner matrices than the state of the art OMP. Furthermore, the proposed framework can also be incorporated with the existing codebook-free solutions such as Alternating Direction Method of Multipliers (ADMM) [44] and SVD based solution [42]. The objective is to achieve better EE performance for codebook-free approaches


$\mathbf{F}_{\text{BB}} = \mathbf{P}_{\text{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\text{BB}}$
TX Antennas
RX Antennas
Switches
Switches
DAC-RF Chain
DAC-RF Chain
L_{T}
Digital Precoder $\hat{\mathbf{F}}_{\text{BB}}$
N_{s}
$\mathbf{s} \in \mathbb{C}^{N_{\text{s}}\times 1}$
Analog Precoder $\mathbf{F}_{\text{RF}}$
N_{T}
Channel $\mathbf{H}$
N_{R}
Analog Combiner $\mathbf{W}_{\text{RF}}$
ADC-RF Chain
ADC-RF Chain
L_{R}
Digital Combiner $\mathbf{W}_{\text{BB}}$
N_{s}
$\mathbf{r} \in \mathbb{C}^{N_{\text{s}}\times 1}$
Dinkelbach Method
Dinkelbach Method
$\mathbf{t} \in \mathbb{C}^{N_{\text{T}}\times 1}$
$\mathbf{y} \in \mathbb{C}^{N_{\text{R}}\times 1}$
Learn $\mathbf{H}$
Compute
$\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, L_{\text{T}}^{opt}$
Beam Training Phase
Dinkelbach Method
Data Communications Phase
L_{T} active RF chains at TX/RX
L_{T}^{opt} active RF chains at TX/RX

(a) The fully-connected A/D HBF architecture with the proposed DM framework.

(b) Block diagram of the beam tracking phase and the data communications phase.

Figure 3.2: System model for a mmWave A/D hybrid MIMO system with the proposed DM framework.

over the fixed number of RF chains case. The proposed energy efficient and low complexity A/D hybrid precoder framework with a fully-connected architecture can be used in designing 5G mmWave MIMO systems effectively and efficiently, such as in 5G cellular systems and wireless backhaul networks.

The main contributions of this chapter can be summarized as follows:

1. The chapter proposes a novel algorithmic framework, where the number of active RF chains is dynamically adapted on a frame-by-frame basis. This is carried out using a low complexity alternative to the BF optimization [65] based on the current channel conditions measured in the A/D HBF architecture.

2. We develop a reduced complexity DM based solution to find the optimal number of RF chains and streams for the mmWave MIMO system for the current channel conditions.

3. A GP-based approach is proposed as a lower complexity approximation solution to compute the precoder and combiner matrices than the state of the art OMP solution.

In the following Section 3.3, we discuss the low complexity designs of A/D HBF matrices, i.e., $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$. Section 3.4 discusses the proposed EE maximization approach via dynamic power allocation. Section 3.5 provides the simulation results. Section 3.6 concludes this chapter.

3.2 Low Complexity A/D HBF Design

The combined problem of designing the precoders and combiners and the number of RF chains can be partitioned into three sub-problems:

- to optimize the A/D hybrid precoders $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$,
- to optimize the A/D hybrid combiners $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ and
- to optimize the number of RF chains, i.e., obtaining L_{T}^{opt} at the TX and L_{R}^{opt} at the RX.

Firstly in this section, we focus on designing the A/D hybrid precoder matrices $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ as shown in Subsection 3.3.1 and the hybrid combiner matrices $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ as shown in Subsection 3.3.2 by assuming that L_{T}^{opt} and L_{R}^{opt} are computed from the proposed DM based solution in Section 3.4 already. In the next section, we propose the DM based solution for optimizing the number of RF chains at the TX and consider that $L_{\mathrm{R}}^{opt} = L_{\mathrm{T}}^{opt}$.

3.2.1 A/D Hybrid Precoding at the TX

It is known that the precoding matrix for the digital beamformer is given based on the Singular Value Decomposition (SVD) of the channel matrix. We consider channel's SVD as $\mathbf{H} = \mathbf{U}_{\mathrm{H}}\mathbf{\Sigma}_{\mathrm{H}}\mathbf{V}_{\mathrm{H}}^{H}$, where $\mathbf{U}_{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ and $\mathbf{V}_{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ are unitary matrices, and $\mathbf{\Sigma}_{\mathrm{H}} \in \mathbb{R}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. The optimal fully digital precoding matrix $\mathbf{F}_{\mathrm{opt}} = \mathbf{V}_{\mathrm{H1}}\mathbf{P}_{\mathrm{TX}}^{(1/2)}$ where the matrix $\mathbf{V}_{\mathrm{H1}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{s}}}$ consists of the N_{s} columns of the right singular matrix $\mathbf{V}_{\mathrm{H}}$ [16] and $\mathbf{P}_{\mathrm{TX}}$ is a diagonal matrix where each diagonal entry represents the power of each transmission stream for the digital precoding case with $\|\mathbf{F}_{\mathrm{opt}}\|_F^2 = \mathrm{tr}(\mathbf{P}_{\mathrm{TX}}) = P_{\mathrm{max}}$. We discuss about $\mathbf{P}_{\mathrm{TX}}$ in more details in the next section. In this section we assume that $\mathbf{P}_{\mathrm{TX}}$ is known.

In order to design the near-optimal A/D hybrid precoder, it can be assumed that the decomposition $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ can be made sufficiently close to the optimal fully digital precoding matrix $\mathbf{F}_{\mathrm{opt}}$ [16]. The Euclidean distance problem is a good approximation, so we can consider the Euclidean distance between the A/D hybrid precoder $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and the channel's optimal fully digital precoder $\mathbf{F}_{\mathrm{opt}}$ to optimize the A/D hybrid precoder matrices. We can define $\mathcal{F}_{\mathrm{RF}}$ to be a set of basis vectors $\mathbf{a}_{\mathrm{T}}(\tilde{\phi}_{il}^{t})$ in order to find the best low dimensional representation of the optimal matrix $\mathbf{F}_{\mathrm{opt}}$ where $\tilde{\phi}_{il}^{t}$ are the angles from the DFT codebook. The problem to design the A/D hybrid precoders can be stated as follows [16, 17]:

$$\begin{aligned}(\mathbf{F}_{\mathrm{RF}}^{opt}, \mathbf{F}_{\mathrm{BB}}^{opt}) = \underset{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}}{\arg\min} \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2,& \\ \text{subject to } \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = P_{\mathrm{max}}.&\end{aligned} \tag{3.5}$$

We consider two stages in the system model as shown in Fig. 3.2: a) the beam training phase, and b) the data communications phase. In stage a), firstly L_{T} available RF chains are activated and the channel is computed which provides us the optimal beamformer, i.e., $\mathbf{F}_{\mathrm{opt}}$. Then the SVD of the channel is computed and the proposed DM is performed to obtain L_{T}^{opt}. In stage b), the optimal analog and digital precoder matrices $\mathbf{F}_{\mathrm{RF}}^{opt}$ and $\mathbf{F}_{\mathrm{BB}}^{opt}$, respectively, are obtained using L_{T}^{opt}. Note that, if we assume that the TX is active for stage a) a small proportion of time, for example, < 10%, then the overall transmit energy consumption is dominated by stage b). The previous problem can be cast in the following form, given by:

$$\begin{aligned} \tilde{\mathbf{F}}_{\mathrm{BB}}^{opt} &= \arg\min_{\tilde{\mathbf{F}}_{\mathrm{BB}}} \|\mathbf{F}_{\mathrm{opt}} - \tilde{\mathbf{D}}_{\mathrm{T}}\tilde{\mathbf{F}}_{\mathrm{BB}}\|_F^2, \\ &\text{subject to } \|\mathrm{diag}(\tilde{\mathbf{F}}_{\mathrm{BB}}\tilde{\mathbf{F}}_{\mathrm{BB}}^H)\|_0 = L_{\mathrm{T}}^{opt}, \ \|\tilde{\mathbf{D}}_{\mathrm{T}}\tilde{\mathbf{F}}_{\mathrm{BB}}\|_F^2 = P_{\max}, \end{aligned} \tag{3.6}$$

where $\tilde{\mathbf{D}}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times L_{\mathrm{T}}^{opt}}$ is the matrix composed by the L_{T}^{opt} columns of the DFT matrix $\mathbf{D}_{\mathrm{T}}$ and $\tilde{\mathbf{F}}_{\mathrm{BB}}$ is a $L_{\mathrm{T}}^{opt} \times N_{\mathrm{s}}$ matrix. The matrices $\tilde{\mathbf{D}}_{\mathrm{T}}$ and $\tilde{\mathbf{F}}_{\mathrm{BB}}$ act as auxiliary variables from which we obtain $\mathbf{F}_{\mathrm{RF}}^{opt}$ and $\mathbf{F}_{\mathrm{BB}}^{opt}$, respectively. The sparsity constraint $\|\mathrm{diag}(\tilde{\mathbf{F}}_{\mathrm{BB}}\tilde{\mathbf{F}}_{\mathrm{BB}}^H)\|_0 = L_{\mathrm{T}}^{opt}$ suggests that $\tilde{\mathbf{F}}_{\mathrm{BB}}$ can not have more than L_{T}^{opt} non-zero rows. Thus, only L_{T}^{opt} columns of the DFT matrix $\mathbf{D}_{\mathrm{T}}$ are effectively selected which is given by $\tilde{\mathbf{D}}_{\mathrm{T}}$. Therefore, L_{T}^{opt} non-zero rows of $\tilde{\mathbf{F}}_{\mathrm{BB}}$ will give us the baseband precoder matrix $\mathbf{F}_{\mathrm{BB}}^{opt}$ and the columns of $\tilde{\mathbf{D}}_{\mathrm{T}}$ will provide the RF precoder matrix $\mathbf{F}_{\mathrm{RF}}^{opt}$. The optimal number of RF chains, i.e., L_{T}^{opt}, is obtained from the proposed optimization solution derived in Section 3.4.

As shown in [16], (3.6) basically reformulates (3.5) into a sparsity constrained reconstruction problem with one variable. The problem can be now addressed as a sparse approximation problem [61] and OMP [62] can be used as an algorithmic solution. To develop fast approximate OMP algorithms that are less complex, [63] proposes improvements to greedy strategies using directional pursuit methods

Algorithm 1 Proposed A/D Hybrid Precoder Design by GP

```
1: Input: $\mathbf{F}_{\mathrm{opt}}$, $\tilde{\mathbf{D}}_{\mathrm{T}}$, $L_{\mathrm{T}}^{opt}$
2: $\mathbf{F}_{\mathrm{RF}} = \mathbf{0}_{N_{\mathrm{T}} \times L_{\mathrm{T}}^{opt}}$, $\Gamma = \varnothing$
3: $\mathbf{F}_{\mathrm{res}} = \mathbf{F}_{\mathrm{opt}}$, $\mathbf{F}_{\mathrm{BB}} = \mathbf{0}_{L_{\mathrm{T}}^{opt} \times N_{\mathrm{s}}}$
4: for $i \leq L_{\mathrm{T}}^{opt}$
5:     $\mathbf{\Psi} = \tilde{\mathbf{D}}_{\mathrm{T}}^{H} \mathbf{F}_{\mathrm{res}}$
6:     $k = \arg\max_{l=1,\ldots,L_{\mathrm{T}}^{opt}} (\mathbf{\Psi}\mathbf{\Psi}^{H})_{l,l}$
7:     $\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{F}_{\mathrm{RF}} \mid \tilde{\mathbf{D}}_{\mathrm{T}}^{(k)}\right]$
8:     $\mathbf{D} = \mathbf{F}_{\mathrm{RF}}^{H} \mathbf{F}_{\mathrm{res}}$
9:     $\mathbf{C} = \mathbf{F}_{\mathrm{RF}} \mathbf{D}$
10:    $g = \frac{\mathrm{tr}\{\mathbf{F}_{\mathrm{res}}^{H} \mathbf{C}\}}{\|\mathbf{C}\|_F^2}$
11:    $\Gamma = \Gamma \cup k$
12:    $\mathbf{F}_{\mathrm{BB}}|_{\Gamma} = \mathbf{F}_{\mathrm{BB}}|_{\Gamma} - g\mathbf{D}$
13:    $\mathbf{F}_{\mathrm{res}} = \mathbf{F}_{\mathrm{res}} - g\mathbf{C}$
14: end for
15: $\mathbf{F}_{\mathrm{BB}} = \sqrt{P_{\max}} \frac{\mathbf{F}_{\mathrm{BB}}}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2}$
```

and discusses optimization schemes on basis of gradient, conjugate gradient and approximate conjugate gradient approaches. GP approach is implemented as an alternative solution to the optimization objective exhibiting similar performance as OMP, faster processing time and lower complexity. GP avoids matrix inversion by using only one matrix vector multiplication per iteration.

Algorithm 1 starts by finding that column of $\tilde{\mathbf{D}}_{\mathrm{T}}$, which is denoted as k as shown in Step 6, along which the optimal precoder has the maximum projection, which is denoted as $\tilde{\mathbf{D}}_{\mathrm{T}}^{(k)}$. It then concatenates that selected column vector to the RF precoder $\mathbf{F}_{\mathrm{RF}}$ as shown in Step 7. The gradient direction in Step 8 is computed at each iteration and the step-size is determined explicitly making use of the gradient direction, as shown in Step 10. The index set Γ is updated at each iteration as shown in Step 11 which is used to generate the baseband precoder matrix $\mathbf{F}_{\mathrm{BB}}$. The residual precoding matrix is computed at Step 13 and the algorithm continues until all L_{T}^{opt} RF chains have been used. Finally the RF precoder matrix $\mathbf{F}_{\mathrm{RF}}$ and the baseband precoder matrix $\mathbf{F}_{\mathrm{BB}}$ are obtained at the end of the algorithm. The transmit power constraint is satisfied at Step 15.

3.2.2 A/D Hybrid Combining at the RX

The A/D hybrid combiner design has a similar mathematical formulation except that the transmit power constraint no longer applies. One may note here that by assuming the A/D hybrid precoders $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ to be fixed, the A/D hybrid combiners $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ can be designed in order to minimize the Mean Square Error (MSE) between the transmitted and processed received signals by using the linear Minimum Mean Square Error (MMSE) RX [16, 17]. The optimization of the number of RF chains at the RX can be performed similarly as at the TX. The design problem for combining matrices can be written as follows:

$$\begin{aligned}(\mathbf{W}_{\mathrm{RF}}^{opt}, \mathbf{W}_{\mathrm{BB}}^{opt}) = \underset{\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}}{\arg\min}\ \mathbb{E}\Big[\|\mathbf{s}-\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{y}\|_2^2\Big],\\ \text{s.t. } \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}},\end{aligned} \tag{3.7}$$

where $\mathcal{W}_{\mathrm{RF}}$ is defined similarly to $\mathcal{F}_{\mathrm{RF}}$ for TX. Following the steps in [16] and similar to the precoder optimization, the MMSE estimation problem may be further written as follows:

$$\begin{aligned}\tilde{\mathbf{W}}_{\mathrm{BB}}^{opt} = \underset{\tilde{\mathbf{W}}_{\mathrm{BB}}}{\arg\min}\|\mathbb{E}[\mathbf{y}\mathbf{y}^{H}]^{\frac{1}{2}}\mathbf{W}_{\mathrm{mmse}} - \mathbb{E}[\mathbf{y}\mathbf{y}^{H}]^{\frac{1}{2}}\tilde{\mathbf{D}}_{\mathrm{R}}\tilde{\mathbf{W}}_{\mathrm{BB}}\|_F^2\\ \text{subject to } \|\mathrm{diag}(\tilde{\mathbf{W}}_{\mathrm{BB}}\tilde{\mathbf{W}}_{\mathrm{BB}}^{H})\|_0 = L_{\mathrm{R}}^{opt},\end{aligned} \tag{3.8}$$

where $\tilde{\mathbf{D}}_{\mathrm{R}}$ is the DFT matrix and $\tilde{\mathbf{W}}_{\mathrm{BB}}$ is a $L_{\mathrm{R}}^{opt} \times N_{\mathrm{s}}$ matrix. The exact solution to (3.8) yields $\mathbf{W}_{\mathrm{mmse}}^{H}$ as follows [16]:

$$\mathbf{W}_{\mathrm{mmse}}^{H} = \Big(\mathbf{F}_{\mathrm{BB}}^{H}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{H}^{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} + \sigma_{\mathrm{n}}^{2}N_{\mathrm{s}}\mathbf{I}_{N_{\mathrm{s}}}\Big)^{-1}\mathbf{F}_{\mathrm{BB}}^{H}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}. \tag{3.9}$$

Algorithm 2 Proposed A/D Hybrid Combiner Design by GP

```
1: Input: $\mathbf{W}_{\text{mmse}}$, $\tilde{\mathbf{D}}_{\text{R}}$, $L_{\text{R}}^{opt}$
2: $\mathbf{W}_{\text{RF}} = \mathbf{0}_{N_{\text{R}} \times L_{\text{R}}^{opt}}$, $\Gamma = \varnothing$
3: $\mathbf{W}_{\text{res}} = \mathbf{W}_{\text{mmse}}$, $\mathbf{W}_{\text{BB}} = \mathbf{0}_{L_{\text{R}}^{opt} \times N_{\text{s}}}$
4: for $i \leq L_{\text{R}}^{opt}$
5:     $\mathbf{\Psi} = \tilde{\mathbf{D}}_{\text{R}}^{H} \mathbb{E}[\mathbf{y}\mathbf{y}^{H}] \mathbf{W}_{\text{res}}$
6:     $k = \arg\max_{l=1,\ldots,L_{\text{R}}^{opt}} (\mathbf{\Psi}\mathbf{\Psi}^{H})_{l,l}$
7:     $\mathbf{W}_{\text{RF}} = \left[ \mathbf{W}_{\text{RF}} \mid \tilde{\mathbf{D}}_{\text{R}}^{(k)} \right]$
8:     $\mathbf{D} = \mathbf{W}_{\text{RF}}^{H} \mathbf{W}_{\text{res}}$
9:     $\mathbf{C} = \mathbf{W}_{\text{RF}} \mathbf{D}$
10:    $g = \frac{\text{tr}\{\mathbf{W}_{\text{res}}^{H} \mathbf{C}\}}{\|\mathbf{C}\|_{F}^{2}}$
11:    $\Gamma = \Gamma \cup k$
12:    $\mathbf{W}_{\text{BB}}|_{\Gamma} = \mathbf{W}_{\text{BB}}|_{\Gamma} - g\mathbf{D}$
13:    $\mathbf{W}_{\text{res}} = \mathbf{W}_{\text{res}} - g\mathbf{C}$
14: end for
```

Similar to the sparsity reconstruction problem for the TX, L_{R}^{opt} non-zero rows of $\tilde{\mathbf{W}}_{\text{BB}}$ will give us the baseband combiner matrix $\mathbf{W}_{\text{BB}}^{opt}$ and the corresponding L_{R}^{opt} columns of $\mathbf{D}_{\text{R}}$ will provide the RF combiner matrix $\mathbf{W}_{\text{RF}}^{opt}$. This sparse signal recovery problem can again be solved by the GP algorithm.

Algorithm 2 provides the pseudo code of the GP solution to find the combiner matrices. It should be noted that step 15 of Algorithm 1 does not need to be replicated here as there is no power constraint at the RX unlike at the TX. Similarly, it starts by finding that column of $\tilde{\mathbf{D}}_{\text{R}}$, which is denoted as k as shown in Step 6, along which the optimal combiner has the maximum projection where the received signal is required as well for computation, which is denoted as $\tilde{\mathbf{D}}_{\text{R}}^{(k)}$. It then concatenates that selected column vector to the RF combiner $\mathbf{W}_{\text{RF}}$ as shown in Step 7. The gradient direction in Step 8 is computed at each iteration and the step-size is determined explicitly making use of the gradient direction as shown in Step 10. Similar to the TX case, the index set Γ is updated at each iteration in Step 11 which is used to generate baseband combiner matrix $\mathbf{W}_{\text{BB}}$. The residual precoding matrix is computed at Step 13. Finally the RF combiner matrix $\mathbf{W}_{\mathrm{RF}}$ and the baseband combiner matrix $\mathbf{W}_{\mathrm{BB}}$ are obtained at the end of the algorithm. In the next section we discuss on obtaining the optimal number of RF chains.

3.3 EE Maximization via Dynamic Power Allocation

In this section we derive the proposed approach which aims at the maximization of the EE by dynamic power allocation in the baseband domain. In terms of achievable information rate R and consumed power P, the EE for the A/D hybrid design can be computed as follows:

$$\mathrm{EE}(\mathbf{P}_{\mathrm{TX}}) \triangleq \frac{R(\mathbf{P}_{\mathrm{TX}})}{P(\mathbf{P}_{\mathrm{TX}})} \text{ (bits/Hz/J)}, \tag{3.10}$$

where R represents the information rate in bits/s/Hz and P is the required power in Watts (W).

The proposed design, as depicted in Fig. 3.2, describes a A/D hybrid system for the TX and the RX, with a certain number of RF chains L_{T} implemented in the hardware. The selection mechanism between the available RF chains is implemented in the baseband domain, as part of the digital processor. This procedure is driven by the DM block, which describes the optimal power scheme for each channel realization.

The power allocation at the TX can be described mathematically by using a diagonal sparse matrix $\mathbf{P}_{\mathrm{TX}} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ where $\mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}} \subset \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ denotes the set of $L_{\mathrm{T}} \times L_{\mathrm{T}}$ diagonal sparse matrices. To represent the baseband selection mechanism we consider that $[\mathbf{P}_{\mathrm{TX}}]_{kk} \in [0, P_{\max}]$, for $k = 1, \ldots, L_{\mathrm{T}}$, where $P_{\max} = \mathrm{tr}(\mathbf{P}_{\mathrm{TX}})$. The diagonal entries of $\mathbf{P}_{\mathrm{TX}}$ with a zero value represent an open switch in Fig. 3.2. Thus, the non-zero diagonal values of $\mathbf{P}_{\mathrm{TX}}$ determine the number of the active RF chains for the TX, i.e., $L_{\mathrm{T}}^{opt} = \|\mathbf{P}_{\mathrm{TX}}\|_0$. If we increase the number of RF chains we might achieve a higher information rate but there is also higher power consumption. Hence, maximizing the EE ratio in (3.10) while considering different constraints on the precoder design provides us the optimal number of RF chains.

3.3.1 Problem Formulation

For a point-to-point A/D hybrid MIMO system, as shown in Fig. 3.2, the overall achievable rate With Respect To (W.R.T.) the active RF chains can be expressed as follows:

$$\begin{aligned}R(\mathbf{P}_{\mathrm{TX}}, \mathbf{P}_{\mathrm{RX}}) = \log \Big| &\mathbf{I}_{N_s} + \frac{1}{\sigma_{\mathrm{n}}^2} \mathbf{W}_{\mathrm{BB}}^H \mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}} \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \times \\ &\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \hat{\mathbf{F}}_{\mathrm{BB}} \hat{\mathbf{F}}_{\mathrm{BB}}^H \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}^H \mathbf{W}_{\mathrm{RF}} \mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}} \mathbf{W}_{\mathrm{BB}} \Big|,\end{aligned} \tag{3.11}$$

where $\mathbf{P}_{\mathrm{TX}} \in \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ is the diagonal matrix describing the power allocation for the TX. For the RX, we use the diagonal matrix $\mathbf{P}_{\mathrm{RX}} \in \{0,1\}^{L_{\mathrm{R}} \times L_{\mathrm{R}}}$ which takes only values from $\{0,1\}$, since it only represents a switching network, hence, $L_{\mathrm{R}}^{opt} = \|\mathbf{P}_{\mathrm{RX}}\|_0$.

Based on [16], it is reasonable to assume that $\hat{\mathbf{F}}_{\mathrm{BB}} \hat{\mathbf{F}}_{\mathrm{BB}}^H \approx \mathbf{I}_{L_{\mathrm{T}}}$ and $\mathbf{W}_{\mathrm{BB}} \mathbf{W}_{\mathrm{BB}}^H \approx \mathbf{I}_{L_{\mathrm{R}}}$, then

$$\begin{aligned}R(\mathbf{P}_{\mathrm{TX}}, \mathbf{P}_{\mathrm{RX}}) = \log \Big| &\mathbf{I}_{L_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2} \mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}} \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \\ &\mathbf{P}_{\mathrm{TX}} \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}^H \mathbf{W}_{\mathrm{RF}} \mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}} \Big|.\end{aligned} \tag{3.12}$$

To simplify this problem, we decompose it into two successive sub-problems, one for the TX and one for the RX. Specifically, to obtain $\mathbf{P}_{\mathrm{TX}}$ we assume that the RX has activated all the switches, i.e., $\mathbf{P}_{\mathrm{RX}} = \mathbf{I}_{L_{\mathrm{R}}}$. So,

$$R(\mathbf{P}_{\mathrm{TX}}) = \log \left| \mathbf{I}_{L_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2} \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{TX}} \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}^H \mathbf{W}_{\mathrm{RF}} \right|. \tag{3.13}$$

Once we obtain $\mathbf{P}_{\mathrm{TX}}$, we can estimate $\mathbf{P}_{\mathrm{RX}}$ based on the following formulation:

$$\begin{aligned} R(\mathbf{P}_{\mathrm{RX}}) = \log \Big| \mathbf{I}_{L_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2} \mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}} \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \\ \mathbf{P}_{\mathrm{TX}} \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}^H \mathbf{W}_{\mathrm{RF}} \mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}} \Big|. \end{aligned} \tag{3.14}$$

Maximizing EE at the RX using (3.14) results into a non-trivial integer programming problem. Therefore in the following we will focus our analysis on the EE maximization at the TX in order to obtain L_{T}^{opt}. We consider the optimal number of RF chains at the RX to be same as at the TX, i.e., $L_{\mathrm{R}}^{opt} = L_{\mathrm{T}}^{opt}$.

Measuring the energy consumed for each entity in the precoder and the combiner is important to design an energy efficient mmWave A/D hybrid MIMO system. In this chapter, we use the power model described in Section 2.2.4 for the case of full resolution sampling, so that the total power P for an A/D HBF system can be described as follows, where we include the power consumed by the RX components as well:

$$\begin{aligned} P = \beta \mathrm{tr}(\mathbf{P}_{\mathrm{TX}}) + 2P_{\mathrm{CP}} + N_{\mathrm{T}} P_{\mathrm{T}} + N_{\mathrm{R}} P_{\mathrm{R}} + L_{\mathrm{T}}^{opt} \times \\ (P_{\mathrm{RF}} + N_{\mathrm{T}} P_{\mathrm{PS}}) + L_{\mathrm{R}}^{opt} (P_{\mathrm{RF}} + N_{\mathrm{R}} P_{\mathrm{PS}}) \ (\mathrm{W}), \end{aligned} \tag{3.15}$$

where β represents the reciprocal of amplifier efficiency; the common parameters at the TX and the RX are P_{CP}, P_{RF}, and P_{PS} which represent the circuit power, i.e., is the power required by all circuit components at the TX, the power per

RF chain, and the power per phase shifter, respectively. P_{T} and P_{R} represent the power per antenna element at the TX and the RX, respectively.

For simplicity we remove the sub-index term "TX" from $\mathbf{P}_{\mathrm{TX}}$. Hence, we consider the problem (3.10) expressed w.r.t. the power allocation matrix $\mathbf{P} \in \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ as follows:

$$\max_{\mathbf{P} \in \mathfrak{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}} \frac{R(\mathbf{P})}{P(\mathbf{P})} \text{ s. t. } P(\mathbf{P}) \leq P'_{\max} \text{ and } R(\mathbf{P}) \geq R_{\min}. \tag{3.16}$$

The first constraint term in (3.16) sets the upper bound for the total power budget of the communication system, i.e., $P'_{\max} = \beta P_{\max} + 2P_{\mathrm{CP}} + N_{\mathrm{T}} P_{\mathrm{T}} + N_{\mathrm{R}} P_{\mathrm{R}} + L_{\mathrm{T}} \times (P_{\mathrm{RF}} + N_{\mathrm{T}} P_{\mathrm{PS}}) + L_{\mathrm{R}}(P_{\mathrm{RF}} + N_{\mathrm{R}} P_{\mathrm{PS}})$.

3.3.2 DM Based Proposed Solution

Fractional programming theory provides us several options to obtain the solution of (3.16). One computational efficient algorithm is the Dinkelbach's algorithm which has been introduced firstly in [70, 71]. Dinkelbach's algorithm replaces the fractional cost function of (3.16) with a sequence of easier difference-based problems. The simulation results in Section 3.5 suggest that this method can achieve good performance. Specifically, the cost function of (3.16) is replaced by a sequence of problems:

$$\max_{\mathbf{P}^{(m)} \in \mathfrak{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}} \left\{ R(\mathbf{P}^{(m)}) - \nu^{(m)} P(\mathbf{P}^{(m)}) \right\}, \tag{3.17}$$

where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$, for $m = 1, 2, \ldots, I_{\max}$, where $I_{\max}$ is the number of maximum iterations. Dinkelbach's algorithm is an iterative algorithm, where at each step an update of $\nu^{(m)}$ is obtained based on the estimated rate and power from the previous iteration. To simplify the implementation of this algorithm we desire a rate expression that does not require explicit formulas for the precoder and combiner matrices, thus avoiding re-running Algorithms 1 and 2 for each possible choice of active RF chains.

In order to proceed with the Dinkelbach's algorithm in our context, let us first elaborate on the information rate and power expressions. Considering the SVD of the channel as $\mathbf{H} = \mathbf{U}_{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{H}}\mathbf{V}_{\mathrm{H}}^{H}$ as shown in Section 3.2, (3.13) is expressed as:

$$\begin{aligned} R(\mathbf{P}) = \log \Bigg| \mathbf{I}_{N_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^{2}} \mathbf{W}_{\mathrm{RF}}^{H}\mathbf{U}_{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{H}}\mathbf{V}_{\mathrm{H}}^{H}\mathbf{F}_{\mathrm{RF}} \times \\ \mathbf{P}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{V}_{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{H}}^{H}\mathbf{U}_{\mathrm{H}}^{H}\mathbf{W}_{\mathrm{RF}} \Bigg|. \end{aligned} \tag{3.18}$$

Following the analysis of [16], it can be proven that $\mathbf{V}_{\mathrm{H}}^{H}\mathbf{F}_{\mathrm{RF}} \approx [\mathbf{I}_{L_{\mathrm{T}}}\, \mathbf{0}_{(N_{\mathrm{T}}-L_{\mathrm{T}})\times L_{\mathrm{T}}}^{T}]^{T}$ and $\mathbf{U}_{\mathrm{H}}^{H}\mathbf{W}_{\mathrm{RF}} \approx [\mathbf{I}_{L_{\mathrm{R}}}\, \mathbf{0}_{(N_{\mathrm{R}}-L_{\mathrm{R}})\times L_{\mathrm{R}}}^{T}]^{T}$, hence,

$$R(\mathbf{P}) = \log \left| \mathbf{I}_{N_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^{2}} \bar{\boldsymbol{\Sigma}}^{2}\mathbf{P} \right|, \tag{3.19}$$

where $\bar{\boldsymbol{\Sigma}} \in \mathbb{R}^{L_{\mathrm{R}}\times L_{\mathrm{T}}}$ with $[\bar{\boldsymbol{\Sigma}}]_{kk} = [\boldsymbol{\Sigma}_{\mathrm{H}}]_{kk}$ for $k = 1, \ldots, L_{\mathrm{T}}$, assuming $L_{\mathrm{T}} = L_{\mathrm{R}}$, while its remaining entries are zero. Since the involved matrices in (3.19) are diagonal, the information rate is decomposed into L_{T} parallel streams, as follows:

$$R(\mathbf{P}) \approx \sum_{k=1}^{L_{\mathrm{T}}} \log \left(1 + \frac{1}{\sigma_{\mathrm{n}}^{2}} [\bar{\boldsymbol{\Sigma}}^{2}]_{kk}[\mathbf{P}]_{kk} \right) \text{ (bits/s/Hz).} \tag{3.20}$$

Recall that L_{T} and L_{R} have preset values based on the hardware design and describe the available RF chains at the TX and the RX, respectively. Considering only the TX, the consumed power w.r.t. the diagonal power allocation matrix can be written as:

$$P_{\mathrm{TX}}(\mathbf{P}) = P_{\mathrm{static}} + \sum_{k=1}^{L_{\mathrm{T}}} (\beta[\mathbf{P}]_{kk} + P_{\mathrm{RF}} + N_{\mathrm{T}} P_{\mathrm{PS}}) \tag{3.21}$$

$$= P_{\mathrm{static}} + \sum_{k=1}^{L_{\mathrm{T}}} \beta'[\mathbf{P}]_{kk} \ (\mathrm{W}), \tag{3.22}$$

where $P_{\mathrm{static}} \triangleq P_{\mathrm{CP}} + N_{\mathrm{T}} P_{\mathrm{T}}$ is independent of the power allocation matrix $\mathbf{P}$ and $\beta' \triangleq \beta + \frac{P_{\mathrm{RF}} + N_{\mathrm{T}} P_{\mathrm{PS}}}{P_{\max}}$. The equivalence between (3.21) and (3.22) is justified since $\sum_{k=1}^{L_{\mathrm{T}}} [\mathbf{P}]_{kk} = \mathrm{tr}(\mathbf{P}) = P_{\max}$.

Based on (3.20) and (3.22), the m-th DM step can be expressed as follows:

$$\{\mathbf{P}^{(m)}, \nu^{(m)}\} = \arg \max_{\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}} \mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)}), \tag{3.23}$$

where

$$\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)}) \triangleq \sum_{k=1}^{L_{\mathrm{T}}} \log \left(1 + \frac{1}{\sigma_{\mathrm{n}}^2} [\bar{\mathbf{\Sigma}}^2]_{kk} [\mathbf{P}^{(m)}]_{kk} \right) - \nu^{(m)} \sum_{k=1}^{L_{\mathrm{T}}} \beta' [\mathbf{P}^{(m)}]_{kk}. \tag{3.24}$$

Note that problem (3.23) is a non-convex one because of the constraint $\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$. To proceed, first we alleviate this constraint, thus (3.23) can be efficiently solved by any standard interior-point method (for example, CVX [72]). Step 3 of Algorithm 3 shows that after alleviating this constraint, (3.23) is solved via CVX to update $\mathbf{P}^{(m)}$. Then we impose the constraint by hard-thresholding the entries of $\mathbf{P}^{(m)}$, i.e., $\mathbf{P}_{\mathrm{th}}^{(m)}$, as shown in Step 4 of Algorithm 3. The thresholding sets to zero all entries of $\mathbf{P}^{(m)}$ that are lower than a given tolerance value ϵ_{th}.

Algorithm 3 starts by initializing the number of available RF chains L_{T}. We update $\mathbf{P}^{(m)}$ by solving the relaxation of (3.23) via CVX as shown in Step 3.

Algorithm 3 Proposed DM for RF Chain Selection

```
1: **Initialize:** $\mathbf{P}^{(0)}, \nu^{(0)}$ satisfying $\mathcal{G}(\mathbf{P}^{(0)}, \nu^{(0)}) \geq 0$, $L_{\mathrm{T}}$, tolerance $\epsilon$
2: $m = 0$
3: **while** $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})| > \epsilon$ **do**
4:     Update $\mathbf{P}^{(m)}$ by solving the relaxation of (3.23) via
       CVX [72].
5:     Thresholding $\mathbf{P}^{(m)}$ as $\mathbf{P}_{\mathrm{th}}^{(m)}$.
6:     Counting non-zero values of $\mathbf{P}_{\mathrm{th}}^{(m)}$ provides $L_{\mathrm{T}}^{opt}$.
7:     Compute $R(\mathbf{P}^{(m)})$ and $P_{\mathrm{TX}}(\mathbf{P}^{(m)})$.
8:     Compute $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$
       where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$.
9:     Update $\nu^{(m)}$ with $R(\mathbf{P}^{(m)})/P_{\mathrm{TX}}(\mathbf{P}^{(m)})$.
10:    $m = m + 1$
11: **end while**
12: Obtain $L_{\mathrm{T}}^{opt} = \|\mathbf{P}_{\mathrm{th}}^{(m)}\|_0$
```

Steps 4-5 show that $\mathbf{P}^{(m)}$ is thresholded as $\mathbf{P}_{\mathrm{th}}^{(m)}$ and counting its non-zero values provides us the optimal number of RF chains which keeps updating within the loop but obtained as $\|\mathbf{P}_{\mathrm{th}}^{(m)}\|_0$ after the loop ends as shown in Step 11. $R(\mathbf{P}^{(m)})$ and $P_{\mathrm{TX}}(\mathbf{P}^{(m)})$ are computed in Step 6 and $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$ is computed based on (3.24) in Step 7 where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$. Steps 8 shows the update in $\nu^{(m)}$ with $R(\mathbf{P}^{(m)})/P_{\mathrm{TX}}(\mathbf{P}^{(m)})$. The loop continues until $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})|$ is less than a given tolerance ϵ. We consider that the optimal number of RF chains provides the number of data streams as well, i.e., $N_{\mathrm{s}} = L_{\mathrm{T}}^{opt}$.

3.3.3 Full Search (FS) Approach

To show that the loss performance is not much in Dinkelbach optimization we also consider a Full Search (FS) approach which resolves the non-convexity issue of (3.23) with convex approximation providing a modified version of the proposed Dinklbach optimization solution which iterates over all the possible number of RF chains. The steps are stated in Algorithm 4 where the maximum EE is obtained and the corresponding number of RF chains are considered to be optimal at the

Algorithm 4 FS Approach for RF Chain Selection

```
Initialize: $L_{\mathrm{T}}$, tolerance $\epsilon$, $\mathrm{EE}^{(0)} = 0$
for $i = 1 : L_{\mathrm{T}}$
  while $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})| > \epsilon$ do
    Compute $\mathbf{P}^{(m)}$ subject to $i$ RF chains
    $\rightarrow$ obtain $L_{\mathrm{T}}^{opt}$ from $\mathbf{P}_{\mathrm{th}}^{(m)}$.
    Compute $R(\mathbf{P}^{(m)})$, $P_{\mathrm{TX}}(\mathbf{P}^{(m)})$ and $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$.
    Update $\nu^{(m)}$ and compute $\mathrm{EE}^{(m)}$
    $= R(\mathbf{P}^{(m)})/P_{\mathrm{TX}}(\mathbf{P}^{(m)})$.
    $m = m + 1$
  end while
  Obtain $L_{\mathrm{T}}^{(i)} = L_{\mathrm{T}}^{opt}$ and $\mathrm{EE}^{(i)}$ based on $\mathrm{EE}^{(m)}$ value.
  if $\mathrm{EE}^{(i)} \geq$ previous $\mathrm{EE}^{(i-1)}$
    Update $\mathrm{EE}^{(i)}$ and $L_{\mathrm{T}}^{(i)}$
  end if
end for
```

end of the algorithm. In Table 3.4 of Section 3.5, we show that the proposed DM has similar performance to the FS approach, while the complexity for computing FS increases significantly.

3.3.4 Brute Force (BF) Approach

The solution to achieve optimal number of RF chains at each realization is also provided in [65] which we call as the BF approach. To make the A/D HBF system energy efficient, BF approach, at each realization (current channel condition), makes a search on all the possible number of RF chains, i.e., $L_{\mathrm{T}} = \{1, 2, 3, ..., N_{\mathrm{T}}\}$, and computes best EE while designing the precoder and combiner matrices, and chooses the corresponding number of RF chains as the optimal number of RF chains. We, in our work, mitigate that need of searching for all possible number of RF chains and then finding an optimal solution, and thus providing equally a high energy efficient and low complexity solution. The observations made in the next section support this statement.

3.4 Simulation Results

This section shows the performance of the proposed DM compared to the existing state of the art solutions such as the BF approach, digital beamforming, analog beamforming and modified version of the proposed solution, i.e., FS approach. For simulations, the proposed DM and the FS approach consider $L_{\mathrm{T}} = L_{\mathrm{R}} = \mathrm{length}\big(\mathrm{eig}(\mathbf{H}\mathbf{H}^H)\big)$ and the BF approach uses the same precoding and combining matrices as the DM solution. The tolerance values considered in both the DM solution and the FS approach algorithms are $\epsilon = 10^{-4}$ and $\epsilon_{\mathrm{th}} = 10^{-6}$. The fully digital beamforming solution uses the same number of RF chains as antennas, i.e., $L_{\mathrm{T}} = N_{\mathrm{T}}$ and $L_{\mathrm{R}} = N_{\mathrm{R}}$, and precoding and combining matrices are $\mathbf{F}_{\mathrm{opt}}$ and $\mathbf{W}_{\mathrm{mmse}}$, respectively, as shown in Subsections 3.3.1 and 3.3.2. The analog beamforming solution implements a single RF chain, i.e., $L_{\mathrm{T}} = L_{\mathrm{R}} = 1$, and the precoding and combining matrices are computed as the phases of the first singular vectors, i.e., $\mathbf{F} = \mathbf{V}_{\mathrm{H}}(1 : N_{\mathrm{T}}, 1)/\mathrm{abs}(\mathbf{V}_{\mathrm{H}})$ and $\mathbf{W} = \mathbf{U}_{\mathrm{H}}(1 : N_{\mathrm{R}}, 1)/\mathrm{abs}(\mathbf{U}_{\mathrm{H}})$, respectively.

The performance of the codebook-free designs such as ADMM [44] and SVD based [42] solutions when incorporated with the proposed framework, using L_{T}^{opt} RF chains, are also observed over the case when fixed number of RF chains are used to compute the precoder and combiner matrices. The comparison between GP and OMP algorithms is also observed through observing the variations in run time w.r.t. the number of RF chains and computational complexities.

3.4.1 System Setup

For the channel parameters, there are 10 rays for each cluster and there are 8 clusters in total, i.e., $N_{\mathrm{ray}} = 10$ and $N_{\mathrm{cl}} = 8$ in (3.1). The average power of each

Circuit power of the TX	$P_{\text{CP}} = 10$ W
Power per RF chain	$P_{\text{RF}} = 100$ mW
Power per phase shifter	$P_{\text{PS}} = 10$ mW

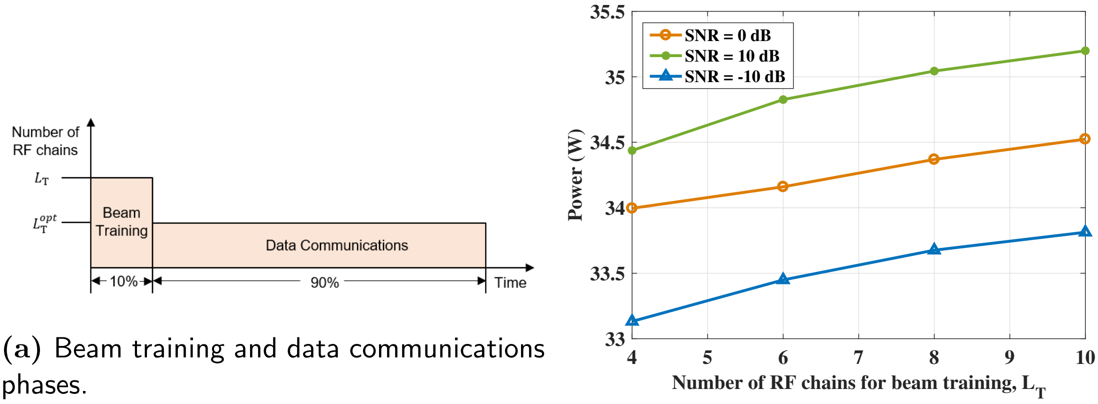
(a) Typical values of the power terms [73].

Number of RF chains, L_{T}	Maximum consumed power (W)
4	34.50
8	38.50
64	94.50

(b) Maximum consumed power in (3.15) for different values of L_{T} for a 64×16 system with $\text{tr}(\mathbf{FF}^H) = 1$.

Table 3.1: Simulation parameters for the power expressions of different precoding solutions.

cluster is unity, i.e., $\sigma_{\alpha,i} = 1$. The azimuth and elevation angles of departure and arrival are computed on the basis of the Laplacian distribution [74] with uniformly distributed mean angles and angle spread as 7.5°. The mean angles are sectored within the range of 60° to 120° in the azimuth domain, and 80° to 100° in the elevation domain. The 64 antenna elements at the TX, i.e., $N_{\text{T}} = 64$, and 16 at the RX, i.e., $N_{\text{R}} = 16$, in the ULA, antenna elements are spaced by distance $d = \lambda/2$ where $\lambda/2$ can be based on a standard frequency value such as 28 GHz [65]. The system bandwidth is normalized to 1 Hz in the simulations. The Signal-to-Noise Ratio (SNR) is $1/\sigma_{\text{n}}^2$. All the simulation results are averaged over 1000 random channel realizations. To illustrate the achievable EE of different precoding solutions, the parameters in the power expressions for each precoder design are set as shown in Table 3.1.(a). For a typical case, the power per power amplifier, $P_{\text{PA}} = 300$ mW, and maximum achievable power, $P_{\max} = 1$ W. Table 3.1.(b) shows the maximum power which can be consumed as determined in (3.15) for different number of RF chains in a 64×16 fully-connected system. The amplifier efficiency $1/\beta$ is considered as 0.4 and the minimum desired rate in (3.16), $R_{\min} = 1$ bits/s/Hz.



(a) Beam training and data communications phases.

(b) Overall power consumption performance for 10% beam training and 90% data communications phases.

Figure 3.3: Beam training and data communications phases and associated power consumption performance for a fully-connected 64×16 system.

3.4.2 Beam Training and Data Communications Phases Analysis

Based on the described communication phases in Fig. 3.2.(b), there are L_{T} active RF chains during the beam training phase. Once the Dinkelbach or FS optimization is performed then we obtain the optimal number L_{T}^{opt} RF chains for the data communications phase. Considering that α represents the ratio between the two phases, the power consumption performance for both the stages is given by:

$$\text{Power} = \alpha \times P(L_{\mathrm{T}}) + (1-\alpha) \times P(L_{\mathrm{T}}^{opt}) \text{ (W)}, \tag{3.25}$$

where $P(L_{\mathrm{T}})$ is the power consumption with (3.15) using L_{T} RF chains and $P(L_{\mathrm{T}}^{opt})$ is using the optimal number of RF chains, L_{T}^{opt}. For example, as shown in Fig. 3.3.(a), when we consider that the beam training phase is active for 10% of the time with L_{T} RF chains, i.e., $\alpha = 0.1$, and the data communications phase is active for the remaining 90% time with L_{T}^{opt} RF chains, i.e., $1-\alpha = 0.9$. The

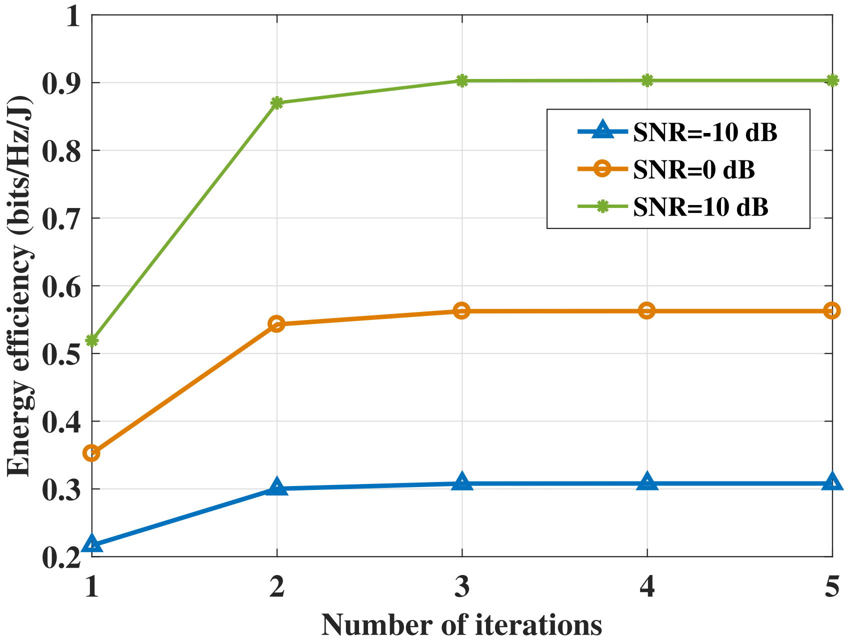
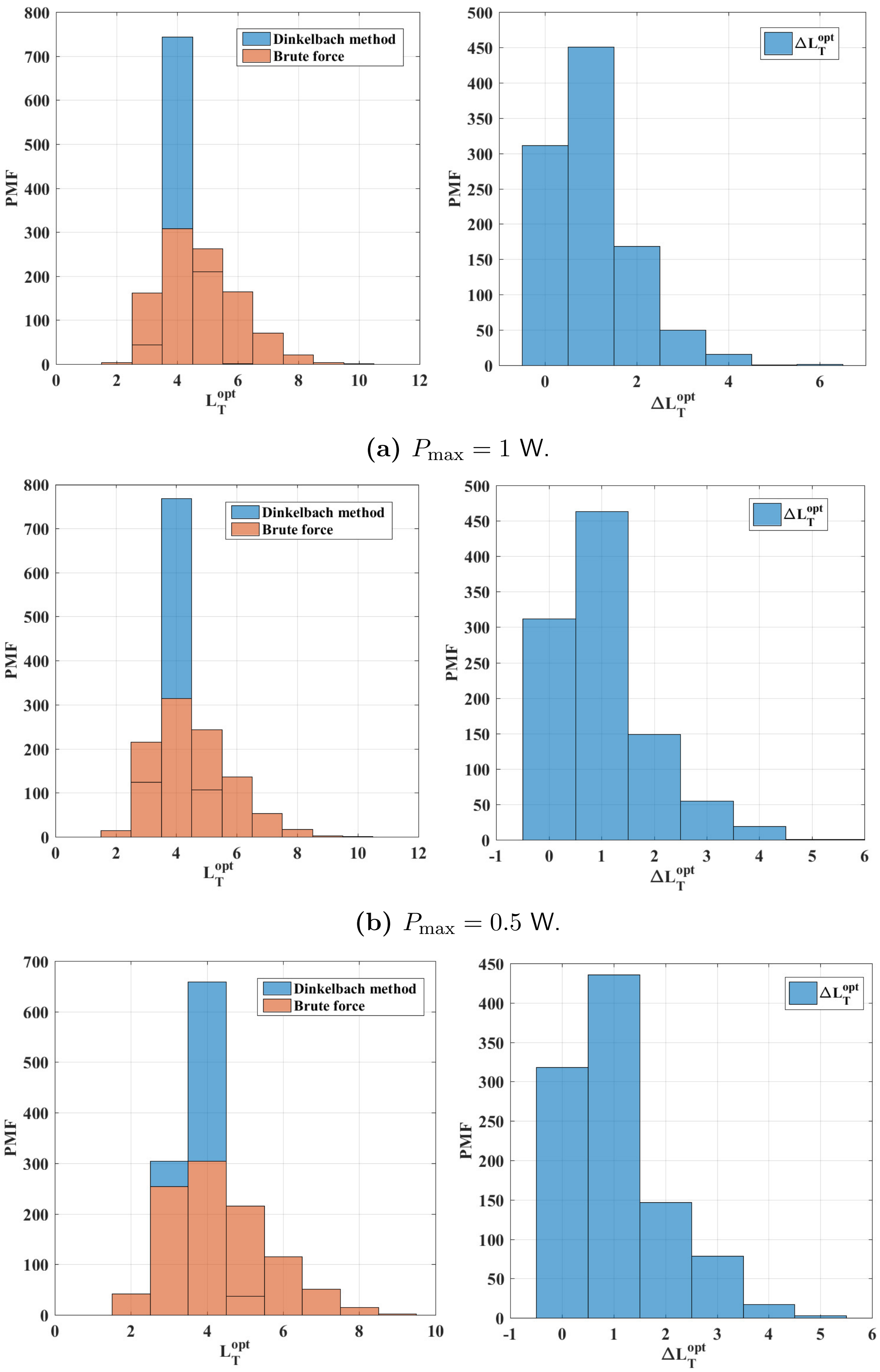


Figure 3.4: Convergence of the proposed DM for different SNR levels for a fully-connected 64×16 system.

performance is observed with three SNR cases in Fig. 3.3.(b). It can be observed that the overall power consumption increases with the increase in the number of RF chains in the beam training phase and high SNR values have higher power consumption levels. For example, at $L_{\mathrm{T}} = 6$, the power consumption at SNR $=$ 0 dB is about 0.65 W higher than at SNR $= -10$ dB.

3.4.3 Convergence of the Proposed DM Solution

Fig. 3.4 shows the convergence of the Dinkelbach optimization solution as proposed in Algorithm 3 to obtain the optimal number of RF chains. It can be observed that the EE for different SNR levels increases with the iterations used to find the optimal number of RF chains. The proposed solution converges rapidly and needs only 2 iterations to converge and achieve an optimal solution at each realization.



(a) $P_{\max} = 1$ W.

(b) $P_{\max} = 0.5$ W.

(c) $P_{\max} = 0.25$ W.

Figure 3.5: PMF plots of the DM and BF solutions at different $P_{\max}$ values for the optimal number of RF chains L_{T}^{opt} and their difference $\Delta L_{\mathrm{T}}^{opt}$ for 64×16 system and SNR = 10 dB.

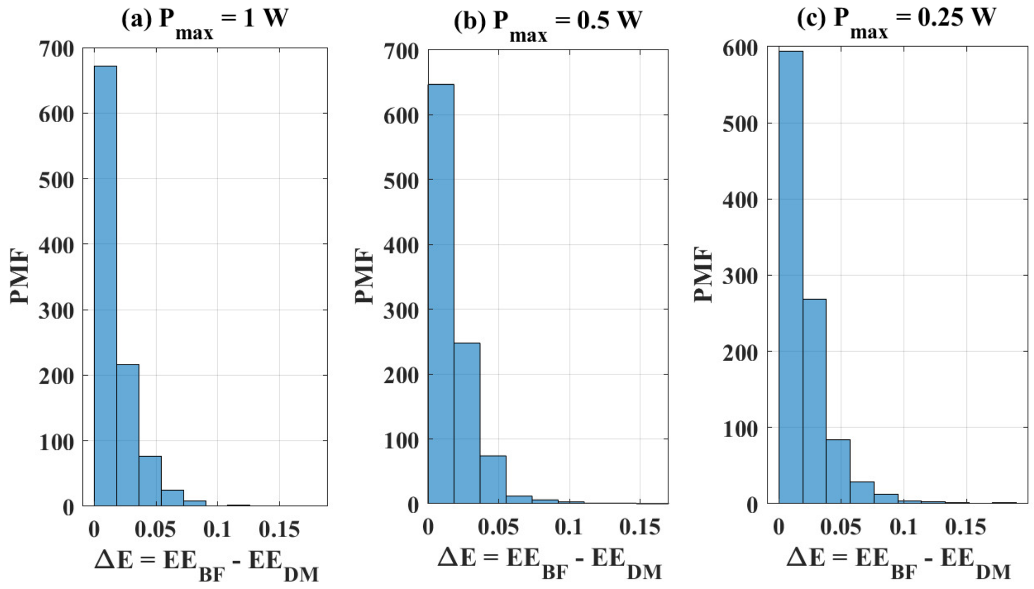


Figure 3.6: PMF plots of EE difference between DM and BF solutions at different P_{max} values for a 64×16 system and SNR $= 10$ dB.

3.4.4 Proposed DM versus BF Approach

The comparison is made to the BF method [65] in detail in terms of the Probability Mass Function (PMF) for RF chain selection, EE performance and the computational complexity. The PMF plots indicate the histogram that for how many realizations (on y-axis) a particular value of the variable defined on x-axis is achieved. Figs. 3.5 and 3.6 show the PMF of the distribution of the proposed DM and the BF approach over the optimal number of RF chains, i.e., L_{T}^{opt}, their difference, i.e., $\Delta L_{\text{T}}^{opt} = |L_{\text{T}}^{opt}{}_{\text{BF}} - L_{\text{T}}^{opt}{}_{\text{DM}}|$, and the EE difference, i.e., $\Delta\text{E} = |\text{EE}_{\text{BF}} - \text{EE}_{\text{DM}}|$, at each channel realization.

Fig. 3.5 shows that for how many channel realizations, the beamforming solutions such as the DM and the BF approach find a particular optimal number of RF chains for different values of P_{max}. It gives us an idea on how close the proposed DM solution is to the BF technique, in terms of finding the optimal

Algorithm	Complexity Order
DM	$\mathcal{O}\big(L_{\mathrm{T}}^{opt}\big)$
BF	$\mathcal{O}\big(L_{\mathrm{T}}^{opt}N_{\mathrm{T}}\big)$
FS	$\mathcal{O}\big(L_{\mathrm{T}}^{opt}L_{\mathrm{T}}\big)$

(a) Complexity orders of the DM, the BF and FS approaches.

No. of TX antennas, N_{T}	Time (s): DM	Time (s): BF
80	0.429	0.613
96	0.438	1.06
112	0.454	1.76
128	0.455	2.69

(b) Run times of the DM and the BF approach w.r.t. N_{T} at SNR $= 10$ dB and $P_{\max} = 1$.

Table 3.2: Computational complexity comparison between the DM, the BF and FS approaches.

number of RF chains. For example, at $P_{\max} = 1$ W, the DM solution chooses $L_{\mathrm{T}}^{opt} = 4$ for ≈ 750 different channel realizations whereas BF chooses 4 RF chains for ≈ 300 realizations and the difference (at each realization) between chosen optimal number of RF chains by both the methods, i.e., $\Delta L_{\mathrm{T}}^{opt}$ is 0 for ≈ 450 different realizations. Also, for example, the EE difference between the two methods, ΔE, at $P_{\max} = 1$ W is close to 0 bits/Hz/J for ≈ 650 channel realizations as observed from Fig. 3.6.

Table 3.2.(a) shows the computational complexities used by the solutions of the DM, the BF and FS approaches w.r.t. the number of the RF chains. We can observe that complexity for the solution of the DM requires complexity order of only $\mathcal{O}(L_{\mathrm{T}}^{opt})$ per iteration. Since the number of the required iterations is usually very small, the overall complexity of the DM is much less than the BF approach which depends on the product of the number of RF chains and the number of antennas. Also, it clearly suggests that the complexity for FS approach increases significantly as the search is made for all possible number of RF chains L_{T}.

For further comparison of the proposed method to the BF approach, we verify

Algorithm	Complexity Order
OMP	$\mathcal{O}\big((L_{\mathrm{T}}^{opt})^4\big) + \mathcal{O}\big((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}}\big)$
GP	$\mathcal{O}\big((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}}\big)$

(a) Complexity orders of GP and OMP.

No. of RF chains at the TX	Time (μs): OMP	Time (μs): GP
8	1.6	1.1
16	5.8	2.8
24	10	5.0
32	16.4	8.0

(b) Run time comparison w.r.t. the number of RF chains for 64×16 mmWave system with $N_{\mathrm{cl}} = 8$, $N_{\mathrm{ray}} = 10$, and SNR = 10 dB.

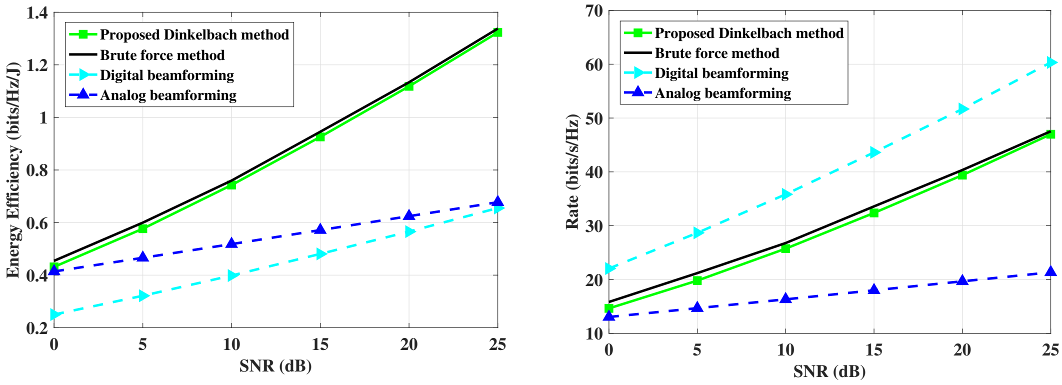
Table 3.3: Computational complexity comparison between GP and OMP solutions.

the run time results as shown in Table 3.2.(b). At SNR = 10 dB and $P_{\max} = 1$, the run time is much less for the proposed solution w.r.t. the number of TX antennas. These results are reported from MATLAB simulation runtime for an independent channel realization. For example, for a large number of antennas, i.e., $N_{\mathrm{T}} = 128$, the proposed solution consumes ≈ 6 times less run time than the BF solution. The observations support the statement that the proposed solution has low complexity while still optimizing the number of RF chains.

3.4.5 Proposed GP versus OMP

Concerning the complexity for deriving the beamforming matrices, recall that OMP requires inversion of a matrix with size $k \times k$, at each one of the L_{T}^{opt} iterations in total, with $k = 1, \ldots, L_{\mathrm{T}}^{opt}$. This operation has cubic complexity order w.r.t. the size of the matrix, i.e., $\mathcal{O}(k^3)$, in general. So, for L_{T}^{opt} iterations, the total cost would be:

$$\sum_{k=1}^{L_{\mathrm{T}}^{opt}} \mathcal{O}(k^3) = \mathcal{O}\big((L_{\mathrm{T}}^{opt})^4\big). \tag{3.26}$$



(a) EE w.r.t. SNR.

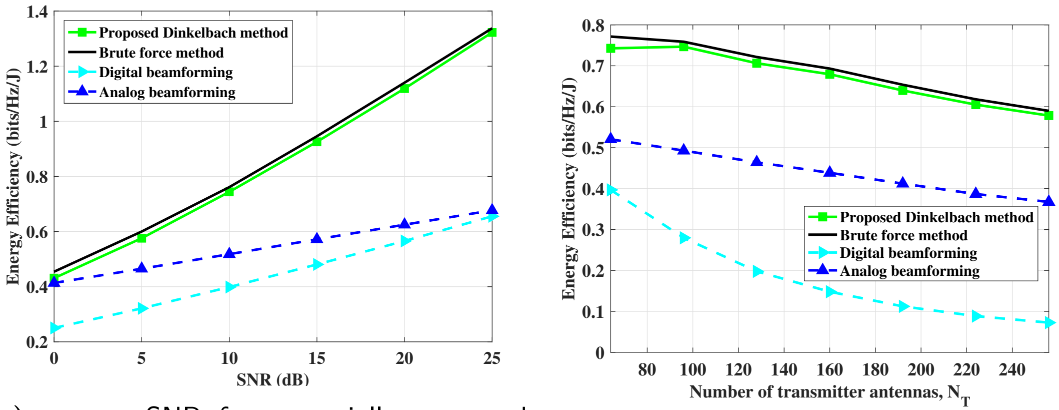
(b) Rate w.r.t. SNR.

Figure 3.7: EE and rate performance of different solutions w.r.t. SNR for a fully-connected 64×16 system at $P_{\max} = 1$ W.

Additionally, a matrix-matrix product is required at each iteration with total cost $\mathcal{O}\big((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}}\big)$. On the other side, the proposed GP algorithm requires only matrix-matrix multiplications at each iteration, hence the complexity order is $\mathcal{O}\big((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}}\big)$. This complexity reduction is justified by the substitution of the matrix inversion with a gradient step. The derived complexity orders are summarized in Table 3.3.(a). In Table 3.3.(b) we show the MATLAB$^{\mathrm{TM}}$ run time comparison (in μs) between OMP and GP w.r.t. the number of RF chains at the TX for a 64×16 mmWave MIMO system with SNR $= 10$ dB. As the time difference between both the algorithmic solutions is considerable with the increase in the number of RF chains, the obtained values indicate that GP consumes much less time than OMP leading to a lower complexity system.

3.4.6 EE and SE Performance of Proposed DM

Fig. 3.7 shows the EE and SE performance of the proposed solution, the BF solution, the full digital solution and the analog beamforming solution w.r.t. SNR for a 64×16 mmWave MIMO system. It can be clearly observed from Fig. 3.7.(a) that the proposed solution is as energy efficient as the BF solution, and better



(a) w.r.t. SNR for a partially-connected structure.

(b) w.r.t. N_{T} for a fully-connected structure.

Figure 3.8: EE performance of different solutions for a 64×16 hybrid mmWave MIMO system at $P_{\max} = 1$ W.

than the fully digital and analog beamforming solutions. For example, at 10 dB, the proposed solution has merely a EE difference of ≈ 0.01 bits/Hz/J with the BF, but shows ≈ 0.35 bits/Hz/J and ≈ 0.25 bits/Hz/J better EE than the fully digital and analog beamforming solutions, respectively. Also, for example, in Fig. 3.7.(b) the proposed design at 10 dB shows a ≈ 10 bits/s/Hz less SE than the fully digital solution, ≈ 10 bits/s/Hz better than analog beamforming and approximately the same performance as the BF method.

Fig. 3.8.(a) shows the EE comparison among the solutions with partially-connected structures where each RF chain is connected to $N_{\mathrm{T}}/L_{\mathrm{T}}^{opt}$ antennas through phase shifters. We can observe similar EE performance characteristics as in Fig. 3.7.(a); for example, the proposed solution has approximately the same EE performance as the BF method, ≈ 0.4 bits/Hz/J and ≈ 0.32 bits/Hz/J better than the fully digital and analog beamforming solutions, respectively, at SNR $= 15$ dB. Fig. 3.8.(b) shows the EE performance comparison w.r.t. the number of TX antennas, N_{T}, for a fully-connected structure. We can observe that the performance starts decreasing with the increase in the number of antenna

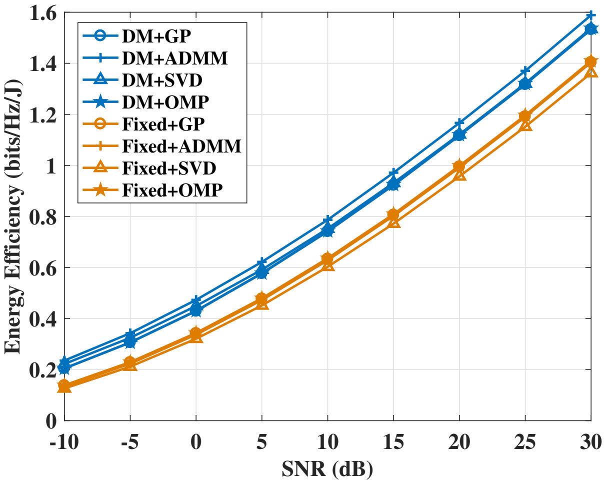


Figure 3.9: EE performance gains w.r.t. SNR at $N_{\mathrm{T}} = 64$ over the fixed number of RF chains case.

SNR (dB)	$\lvert\mathrm{EE}_{\mathrm{DM}} - \mathrm{EE}_{\mathrm{FS}}\rvert$ (bits/Hz/J)
-10	0.013
-5	0.018
0	0.043
5	0.108
10	0.189

Table 3.4: EE performance difference between the DM and the FS approach.

elements. For example, at $N_{\mathrm{T}} = 64$, the EE for the proposed DM is close to that of the BF solution which is ≈ 0.35 bits/Hz/J and ≈ 0.25 bits/Hz/J better than the fully digital beamforming and analog beamforming solutions, respectively. At $N_{\mathrm{T}} = 256$, the EE performance for the proposed DM solution is decreased to ≈ 0.56 bits/Hz/J and close to the BF solution, and ≈ 0.5 bits/Hz/J and ≈ 0.2 bits/Hz/J better than the fully digital beamforming and analog beamforming solutions, respectively.

Fig. 3.9 shows the EE gain of the DM based framework when used with codebook-based GP and OMP techniques, and when incorporated with codebook-free ADMM [44] and SVD [42] techniques, over the case of a fixed number of RF chains, in this case, 8. The codebook-free technique such as ADMM performs better than the codebook-based techniques such as GP and OMP, while SVD shows a similar performance. The EE performance of GP and OMP techniques are same. Table 3.4 shows EE performance comparison between the proposed DM approach (Algorithm 3), i.e., $\mathrm{EE}_{\mathrm{DM}}$, and the FS approach (Algorithm 4), i.e., $\mathrm{EE}_{\mathrm{FS}}$, where we can observe that the difference between their EE is considerably low. It states that the FS approach shows very similar performance to the proposed method. From implementation perspective, we already showed in Table 3.2 (a) that the FS approach has higher computational complexity than the proposed DM solution.

3.5 Summary

This chapter proposes an energy efficient A/D HBF framework with a novel architecture for a mmWave MIMO system, where we optimize the active number of RF chains through fractional programming. The proposed DM based framework reduces the complexity significantly and achieves almost the same EE performance as the state of the art BF approach. Both approaches achieve higher EE performance when compared with the fully digital beamforming and the analog beamforming solutions. In particular, the proposed solution only needs to compute the precoder and combiner matrices once, after the number of active RF chains are decided through the Dinkelbach optimization solution.

The modified version of the proposed solution, i.e., FS approach, shows very similar performance to the proposed DM but the complexity increases significantly. The codebook-free designs such as ADMM and SVD based solutions, when incorporated with the proposed framework also achieve better EE performance over the fixed number of RF chains case. It is also shown that GP incorporated with the proposed DM is a faster and less complex approximation solution to compute the precoder and combiner matrices than OMP.

For this chapter, we focus on maximizing the EE but extending these techniques to consider both estimated channels and frequency selective channels can be considered for future work. Also, this chapter optimizes the number of RF chains and streams to provide an energy efficient solution, however it considers full resolution sampling. In the following chapter, we discuss channel estimation and EE maximization solutions for the mmWave hybrid MIMO system with low resolution sampling.

Chapter 4

Sparse MmWave Channel Estimation and EE Maximization with Low Resolution DACs/ADCs

4.1 Introduction

IN PREVIOUS Chapter 3 we discussed that A/D HBF architectures reduce the hardware complexity through fewer RF chains and support multi-stream communication with good capacity performance [14, 16, 17]. Furthermore, optimizing the number of RF chains and streams provides an energy efficient mmWave hybrid MIMO system. However the large number of antenna elements associated with mmWave MIMO systems makes it hard to use many ADCs, which is a power hungry component [15]. Moreover, ADCs have much higher sampling rates for wide bandwidth mmWave systems than at microwave frequencies, and employing high speed ADCs increases the power consumption and the cost significantly [46, 75]. Implementing low resolution quantization such as 1-bit to 3-bit resolution in hybrid MIMO systems further improves the EE of such systems [15]. For example, the use of 1-bit ADCs in MIMO systems has been discussed in [76] and [77], and channel estimation is investigated as well. In that work, the channel is known perfectly to the TX and the RX while in practical scenarios, the CSI is not known and should be estimated by both the TX and the RX. In this chapter we discuss the role of low resolution quantization in mmWave HBF MIMO systems for sparse channel estimation and EE maximization. In this section, we first proceed with the literature review of the sparse channel estimation and EE maximization associated with the low resolution quantization and then we discuss our contributions in this chapter.

4.1.1 Literature Review

In terms of the sparse channel estimation, references [78–80] estimate the sparse mmWave channel using signal processing tools for high resolution ADCs, but the use of low resolution ADCs at the RX can significantly reduce the power consumption without significantly affecting the capacity of the system [81]. Recently, [82] and [83] considered 1-bit ADC quantization systems and the sparsity in the angle domain is exploited to be able to use CS techniques to recover the channel parameters. The proposed adaptive technique in [82] fails to provide good estimation of the channel at low SNR values. Reference [83] proposes only an Expectation-Maximization (EM) algorithm which has high complexity since each iteration requires a matrix inverse computation and convergence of the algorithm requires many iterations. To observe the effect of low resolution ADCs, an Additive Quantization Noise Model (AQNM) is considered in [57] and [84]. The effect of AQNM is investigated in [57] for the case of a point-to-point mmWave MIMO system, while in [84] the desired rate of the uplink was derived for the case of mmWave fading channels. References [85] and [86] also implement the EM algorithm for a MIMO channel. Further improvements to the EM algorithm are proposed using EM-Generalized Approximate Message Passing (GAMP) [87] and Vector Approximate Message Passing (VAMP) [88]. The use of EM-GAMP has been exploited for a broadband mmWave MIMO channel model with low resolution ADCs at the RX in [89].

In terms of the EE maximization, the existing literature mostly discusses low resolution DACs/ADCs with a large or full number of RF chains (one RF chain per antenna) or full or high resolution sampling with a small number of RF chains. As the power consumption of DACs/ADCs increases exponentially with the number of bits, to further reduce the power consumption one can consider a combined analog and digital hybrid structure with small number of RF chains and low resolution DACs/ADCs as discussed briefly in Chapter 3. To observe the effect of low resolution ADCs, an AQNM is considered in [57] for the case of a point-to-point mmWave MIMO system and in [84] for the case of mmWave fading channels. Reference [67] assumes fully digital precoding at the TX, and baseband and RF combining with low resolution sampling at the RX. Reference [90] develops the idea of a mixed-ADC architecture where a better energy-rate trade off is achieved with the use of a combination of low and high resolution ADCs than using only full resolution or low resolution systems.

Most of the literature studies the use of low resolution sampling only at the RX side, assuming fully digital or hybrid TX with high resolution DACs. Given the use of wide bandwidths in typical mmWave systems at the TX, employing low resolution DACs at the TX can also help to reduce the power consumption. So EE approaches that are mainly focused on ADCs at the RX can also be applied to the DACs at the TX considering the TX specific system model parameters. Reference [91] uses low resolution DACs which can be implemented to reduce the power consumption for a hybrid MIMO architecture. Reference [92] employs low resolution DACs at the base station for a narrowband multi-user MIMO system.

References [44, 65] consider the EE optimization problem for hybrid transceivers but with full resolution sampling at the DACs/ADCs.

This chapter exploits the low resolution sampling at the conversion units and provides more efficient solutions in terms of EE and channel estimation than existing baselines in the literature. The details of the contributions are discussed in the following subsection.

4.1.2 Contributions

In section 4.3, we exploit the Stein's Unbiased Risk Estimate (SURE) based GAMP solution combined with EM steps called the EM-SURE-GAMP in a mmWave MIMO system with low resolution sampling at the RX. Reference [93] describes the advantages of the SURE based parametric denoiser when incorporated with the Approximate Message Passing (AMP) framework. This novel solution avoids strong assumptions on the channel statistics where SURE, depending on the noisy observation, is minimized to adaptively optimize the denoiser within the parametric class at each iteration. The proposed solution is compared with the EM-GAMP solution for a narrowband channel model and improved MSE performance is observed for both low and high SNR regimes. The unknown channel parameters are modeled by a Bernoulli Gaussian distribution for both the techniques.

In Section 4.4, we proceed with a A/D hybrid transmit beamformer with low resolution DACs. The analog and digital parts are connected with a predefined number of RF chains which can be in active or inactive state. Assuming that the power consumption of the TX is determined mainly by the DACs of the RF chains, deactivating specific RF chains in an intelligent manner would increase the EE of the beamformer. Therefore, in this work, we derive an optimal approach in terms of EE maximization, which selects the best subset between the available RF chains. We implement an iterative method to overcome the non-convexity of the fractional programming optimization problem. The proposed approach capitalizes from sparse-based subset selection techniques to provide an efficient solution to the problem. We also implement an exhaustive search approach (for example, in [65]) which expresses the upper bound for EE maximization and clearly shows the performance trade-offs. In the next section, we discuss the mmWave A/D HBF MIMO system model with low resolution DACs/ADCs.

This chapter proceeds by discussing the system model for a mmWave HBF MIMO system with low resolution sampling in Section 4.2. Then it proposes an efficient sparse channel estimation algorithm for a mmWave HBF MIMO system with low resolution ADCs in Section 4.3, and EE maximization approach for mmWave HBF MIMO system with low resolution DACs in Section 4.4. Section 4.5 provides the simulation results and Section 4.6 concludes this chapter.

4.2 MmWave HBF MIMO System with Low Resolution DACs/ADCs

The system setup in Fig. 4.1 shows the updated system model (of Fig. 3.1) with low resolution DACs and ADCs at the TX and the RX, respectively. We already know that the number of TX RF chains L_{T} is usually smaller than the number of the TX antennas N_{T} and similarly for the RX $L_{\mathrm{R}} \leq N_{\mathrm{R}}$ for a HBF system. After the RF/analog precoding, each phase shifter is connected to all the antenna elements, and similarly at the RX, each phase shifter is connected to all the antenna elements before the analog combining unit.

At the TX, the low resolution DACs are associated with the RF chains after the baseband precoding unit and before the analog precoding unit. At the RX, the low resolution ADCs are associated with the RF chains after the analog combining unit and before the baseband combining unit. The analog precoder and combiner matrices, $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$, are based on phase shifters, i.e., the elements of these matrices have unit modulus and continuous phase over $0 - 2\pi$ radians. Thus, $\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}^{N_{\mathrm{T}} \times L_{\mathrm{T}}}$ and $\mathbf{W}_{\mathrm{RF}} \in \mathcal{W}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}$ where the set $\mathcal{F}$ and $\mathcal{W}$ represent the set of possible phase shifts in $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$, respectively. The sets $\mathcal{F}$ and $\mathcal{W}$ for variables a and b, respectively, are defined as $\mathcal{F} = \{a \in \mathbb{C} \mid |a| = 1\}$ and $\mathcal{W} = \{b \in \mathbb{C} \mid |b| = 1\}$.

4.2.1 System with Low Resolution ADCs for Channel Estimation

We consider the channel model in (3.1) and the beamspace representation of the channel with ULA setup [38,39] as shown in (3.2) and express it as $\mathbf{H} = \mathbf{D}_{\mathrm{R}}\mathbf{Z}\mathbf{D}_{\mathrm{T}}^{H}$, where $\mathbf{Z} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ represents a sparse channel matrix with a few non-zero entries assumed to follow Bernoulli-Gaussian distribution, while $\mathbf{D}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ and $\mathbf{D}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ are the DFT matrices at the RX and the TX, respectively. For N_{T} TX antennas and N_{R} RX antennas, the channel matrix $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ can be further expressed as follows:

$$\mathbf{H} = \mathbf{D}_{\mathrm{R}}\mathbf{Z}\mathbf{D}_{\mathrm{T}}^{H} = \sum_{i=1}^{N_{\mathrm{R}}} \sum_{k=1}^{N_{\mathrm{T}}} [\mathbf{Z}]_{ik} \mathbf{a}_{\mathrm{R}}(\phi_i) \mathbf{a}_{\mathrm{T}}^{H}(\theta_k), \tag{4.1}$$

where $\mathbf{a}_{\mathrm{T}}(\theta_k) = \frac{1}{\sqrt{N_{\mathrm{T}}}}[1, e^{-j\theta_k}, \ldots, e^{-j(k-1)\theta_k}]^T$ is the steering vector of the TX with $\theta_k = k/N_{\mathrm{T}}$ the normalized uniformly spaced spatial angles. Specifically, each element of the sparse matrix $[\mathbf{Z}]_{ik}$ is assumed to follow the Bernoulli-Gaussian

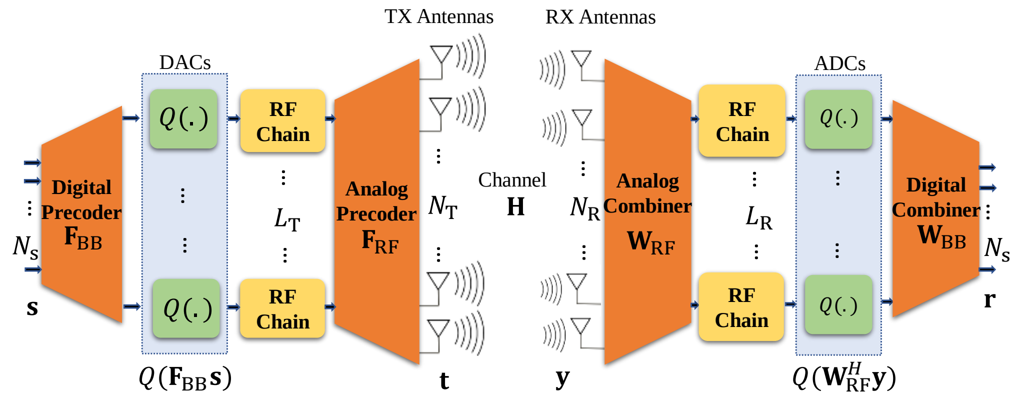

TX Antennas
RX Antennas
DACs
ADCs
Digital Precoder $\mathbf{F}_{\text{BB}}$
N_s
$\mathbf{s}$
$Q(.)$
RF Chain
L_T
Analog Precoder $\mathbf{F}_{\text{RF}}$
N_T
Channel $\mathbf{H}$
N_R
Analog Combiner $\mathbf{W}_{\text{RF}}$
L_R
Digital Combiner $\mathbf{W}_{\text{BB}}$
$\mathbf{r}$
$Q(\mathbf{F}_{\text{BB}}\mathbf{s})$
$\mathbf{t}$
$\mathbf{y}$
$Q(\mathbf{W}_{\text{RF}}^H\mathbf{y})$

Figure 4.1: A mmWave A/D hybrid MIMO system with low resolution DACs/ADCs at the TX/RX.

distribution [89], i.e.,

$$p([\mathbf{Z}]_{ik}) = (1-\eta)\delta([\mathbf{Z}]_{ik}) + \frac{\eta}{\sqrt{2\pi}\sigma_{\text{h}}}e^{-\frac{|[\mathbf{Z}]_{ik}|^2}{2\sigma_{\text{h}}^2}}$$

where $\delta(\cdot)$ is the Dirac delta function and $\eta = \frac{L}{N_{\text{T}}N_{\text{R}}}$ denotes the sparsity of the virtual channel where L is the number of channel paths.

We consider the system setup in Fig. 4.1 with low resolution ADCs at the RX and assume that the channel is quasi-static, i.e., it remains static during a period of time, which includes both channel training and data transmission phases. During the training phase, at each training instance t, the TX generates the vector $\mathbf{s}(t) \in \mathbb{C}^{N_{\text{s}}\times 1}$ following $\mathbb{E}[\mathbf{s}(t)\mathbf{s}(t)^H] = \frac{1}{N_{\text{s}}}\mathbf{I}_{N_{\text{s}}}$, which is the input to the RF precoder, $\mathbf{F}_{\text{RF}}(t) \in \mathbb{C}^{L_{\text{T}}\times N_{\text{T}}}$. This signal is transmitted through the sparsely modeled channel $\hat{\mathbf{H}}$ and the received vector is processed by the RF combiner $\mathbf{W}_{\text{RF}}(t) \in \mathbb{C}^{N_{\text{R}}\times L_{\text{R}}}$. The elements of the RF precoders and combiners have equal norm as they represent the TX and the RX phase shifters. For the case of number of streams equal to the number of RF chains, the baseband matrices, $\mathbf{F}_{\text{BB}}(t) \in \mathbb{C}^{L_{\text{T}}\times N_{\text{s}}}$ at the TX and $\mathbf{W}_{\text{BB}}(t) \in \mathbb{C}^{L_{\text{R}}\times N_{\text{s}}}$ at the RX, are identity matrices so we consider only RF/analog processing to formulate the channel estimation problem. Similar to (3.4) and considering low resolution ADCs at the RX and full resolution sampling at the TX, the received signal after RF/analog processing, $\mathbf{y}_{\mathrm{c}}(t) \in \mathbb{C}^{L_{\mathrm{R}}\times 1}$ for $t = 1, \ldots, T$, is expressed as follows:

$$\mathbf{y}_{\mathrm{c}}(t) = \mathbf{W}_{\mathrm{RF}}^{H}(t)[\mathbf{H}\mathbf{t}(t) + \mathbf{n}(t)] = \mathbf{W}_{\mathrm{RF}}^{H}(t)\mathbf{H}\mathbf{F}_{\mathrm{RF}}(t)\mathbf{s}(t) + \mathbf{W}_{\mathrm{RF}}^{H}(t)\mathbf{n}(t), \tag{4.2}$$

where $\mathbf{t}(t) = \mathbf{F}_{\mathrm{RF}}(t)\mathbf{s}(t)$ is the transmitted signal at time instance t, $\mathbf{n}(t)$ is the noise vector following the complex Gaussian distribution with i.i.d. entries, i.e., $\mathbf{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_{\mathrm{R}}})$. By concatenating all the T training sequences the into the real-valued equivalent form we have

$$\bar{\mathbf{y}} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{y}}_{\mathrm{c}}) \\ \mathrm{Im}(\bar{\mathbf{y}}_{\mathrm{c}}) \end{bmatrix} = \bar{\mathbf{\Psi}} \begin{bmatrix} \mathrm{Re}(\mathbf{z}_{\mathrm{c}}) \\ \mathrm{Im}(\mathbf{z}_{\mathrm{c}}) \end{bmatrix} + \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{n}}_{\mathrm{c}}) \\ \mathrm{Im}(\bar{\mathbf{n}}_{\mathrm{c}}) \end{bmatrix} \tag{4.3}$$

where the concatenated received signal $\bar{\mathbf{y}}_{\mathrm{c}} = [\mathbf{y}_{\mathrm{c}}(1), \cdots, \mathbf{y}_{\mathrm{c}}(t)]^T \in \mathbb{C}^{TL_{\mathrm{R}}\times 1}$, the system matrix $\bar{\mathbf{\Psi}} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{\Psi}}_{\mathrm{c}}) & -\mathrm{Im}(\bar{\mathbf{\Psi}}_{\mathrm{c}}) \\ \mathrm{Im}(\bar{\mathbf{\Psi}}_{\mathrm{c}}) & \mathrm{Re}(\bar{\mathbf{\Psi}}_{\mathrm{c}}) \end{bmatrix}^T \in \mathbb{R}^{2TL_{\mathrm{R}}\times 2N_{\mathrm{R}}N_{\mathrm{T}}}$, where $\bar{\mathbf{\Psi}}_{\mathrm{c}} = [\mathbf{\Psi}_{\mathrm{c}}(1), \cdots, \mathbf{\Psi}_{\mathrm{c}}(t)]^T$ with $\mathbf{\Psi}_{\mathrm{c}}(t) = \{\mathbf{s}^T(t)\mathbf{F}_{\mathrm{RF}}^{T}(t)\mathbf{D}_{\mathrm{T}} \otimes \mathbf{W}_{\mathrm{RF}}^{H}(t)\mathbf{D}_{\mathrm{R}}\} \in \mathbb{C}^{TL_{\mathrm{R}}\times N_{\mathrm{R}}N_{\mathrm{T}}}$, $\mathbf{z}_{\mathrm{c}}$ contains the entries of the sparse channel matrix $\mathbf{Z}$, i.e.,

$$\mathbf{z}_{\mathrm{c}} = \mathrm{vec}(\mathbf{Z}) = \mathrm{vec}([\mathbf{Z}_{11}, \mathbf{Z}_{12}, \ldots, \mathbf{Z}_{21}, \mathbf{Z}_{22}, \ldots, \mathbf{Z}_{N_{\mathrm{R}}N_{\mathrm{T}}}]^T), \tag{4.4}$$

and $\bar{\mathbf{n}}_{\mathbf{c}} = [\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{n}(1), \cdots, \mathbf{W}_{\mathrm{RF}}^{H}\mathbf{n}(t)]^T \in \mathbb{C}^{TL_{\mathrm{R}}\times 1}$.

Now, let us denote the K-level quantization of $\bar{\mathbf{y}} \in \mathbb{R}^{2TL_{\mathrm{R}}\times 1}$ as

$$\bar{\mathbf{q}} = \mathcal{Q}(\bar{\mathbf{y}}), \tag{4.5}$$

where $\bar{\mathbf{q}} = [q_1, \dots, q_{2TL_{\mathrm{R}}}]^T \in \mathbb{R}^{2TL_{\mathrm{R}} \times 1}$. Each output element takes one of the K distinct values i.e., $q_i^1, \dots, q_i^K$. with $q_i^k = -(M+1) + k\Delta$ depending on the quantizer lower and upper thresholds $[l_i^k, u_i^k]$. The lower and upper quantizer boundary values are set to $q_{\min} = -\kappa\sqrt{\mathbb{E}\{y_i^2\}}$ and $q_{\max} = \kappa\sqrt{\mathbb{E}\{y_i^2\}}$, $\forall i$ and for $\kappa \in [1, 5]$, respectively. The quantizer's step-size is given by $\Delta = \frac{q_{\max} - q_{\min}}{M}$, while the average power $\mathbb{E}\{y_i^2\}$ can be obtained via an automatic gain control circuit.

4.2.2 System with Low Resolution DACs for EE Maximization

We now consider how to extend Section 4.2.1 to study the AQNM to represent the introduced distortion of the quantization noise at the TX. Given that $Q(\cdot)$ denotes a uniform scalar quantizer then for the scalar input s we have that,

$$Q(s) \approx \delta s + \epsilon, \tag{4.6}$$

where

$$\delta = \sqrt{1 - \frac{\pi\sqrt{3}}{2} 2^{-2b}} \tag{4.7}$$

is the multiplicative distortion parameter for bit sampling resolution equal to b and ϵ is the additive quantization noise with $\epsilon \sim \mathcal{CN}(0, \sigma_\epsilon^2)$, where

$$\sigma_\epsilon = \sqrt{1 - \frac{\pi\sqrt{3}}{2} 2^{-2b}} \sqrt{\frac{\pi\sqrt{3}}{2} 2^{-2b}} = \delta(1 - \delta^2). \tag{4.8}$$

Let $\mathbf{s} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ is the normalized data vector, then based on the AQNM the vector containing the complex output of all the DACs can be expressed as:

$$Q(\mathbf{F}_{\mathrm{BB}}\mathbf{s}) \approx \delta \mathbf{F}_{\mathrm{BB}}\mathbf{s} + \boldsymbol{\epsilon}, \tag{4.9}$$

where $Q(\mathbf{F}_{\mathrm{BB}}\mathbf{s}) \in \mathbb{C}^{L_{\mathrm{T}}\times 1}$ and $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{L_{\mathrm{T}}\times N_s}$ is the baseband part of transmit beamformer. The second term of (4.9) expresses the additive quantization noise for all RF chains with $\boldsymbol{\epsilon} \in \mathcal{CN}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{L_{\mathrm{T}}})$. This leads us to the following expression for the transmitted signal, as seen at the output of the A/D hybrid TX:

$$\mathbf{t} = \mathbf{F}_{\mathrm{RF}}\left(\delta \mathbf{F}_{\mathrm{BB}}\mathbf{s} + \boldsymbol{\epsilon}\right) = \delta \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{F}_{\mathrm{RF}}\boldsymbol{\epsilon}, \tag{4.10}$$

where $\mathbf{F}_{\mathrm{RF}}$ is the analog precoding matrix at the TX.

Note that, in this context, we consider the low resolution quantization at the TX and full resolution sampling at the RX. So we can express the RX combining matrix as $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{s}}}$ which includes both the RF and digital processing at the RX. For such a system, the output RX signal is expressed as follows:

$$\mathbf{r} = \mathbf{W}^H\mathbf{H}\mathbf{t} + \mathbf{W}^H\mathbf{n} \tag{4.11}$$

$$\implies \mathbf{r} = \underbrace{\delta \mathbf{W}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}}_{\mathbf{H}_{\mathrm{eff}}(L_{\mathrm{T}},\delta)}\mathbf{s} + \underbrace{\mathbf{W}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\boldsymbol{\epsilon} + \mathbf{W}^H\mathbf{n}}_{\boldsymbol{\eta}}, \tag{4.12}$$

where $\mathbf{H}_{\mathrm{eff}}(L_{\mathrm{T}}, \delta)$ is the effective channel which is a function of the number of RF chains L_{T} and the distortion δ, $\boldsymbol{\eta}$ is the combined effect of the Gaussian and quantization noise with $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\eta)$, while $\mathbf{R}_\eta$ is the combined noise covariance matrix with,

$$\mathbf{R}_\eta(L_{\mathrm{T}}, \delta) = \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^H] = \sigma_\epsilon^2 \mathbf{W}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}^H\mathbf{W} + \sigma_{\mathrm{n}}^2\mathbf{W}^H\mathbf{W}, \tag{4.13}$$

which is also a function of the number of RF chains L_{T} and the distortion δ. Note that unlike what is common in the existing literature, in this work we also take into account the cross-terms of the noise covariance matrix $\mathbf{R}_\eta$. We believe this is a more realistic scenario since it can also incorporate system impairments such as phase noise into the problem formulation for the EE maximization case. In the next section, we first proceed with the proposed sparse channel estimation for the HBF system with low resolution ADCs at the RX and then discuss the EE maximization for the HBF system with low resolution DACs at the TX.

4.3 Proposed Sparse Channel Estimation

4.3.1 Problem Formulation

Following the beamspace representation of the sparse mmWave channel in (3.2), the system model of (4.2) can be rewritten into an equivalent form for the channel estimation problem, i.e.,

$$\mathbf{y}_{\mathrm{c}}(t) = \big(\underbrace{\mathbf{s}^T(t)\mathbf{F}_{\mathrm{RF}}^T(t)\mathbf{D}_{\mathrm{T}} \otimes \mathbf{W}_{\mathrm{RF}}^H(t)\mathbf{D}_{\mathrm{R}}}_{\mathbf{\Psi}_{\mathrm{c}}(t)\in\mathbb{C}^{L_{\mathrm{R}}\times N_{\mathrm{R}}N_{\mathrm{T}}}}\big)\mathbf{z}_{\mathrm{c}} + \mathbf{W}_{\mathrm{RF}}^H(t)\mathbf{n}(t). \tag{4.14}$$

Now the sparse estimation techniques can be utilized to recover the sparse vector $\mathbf{z}_{\mathrm{c}}$. Concerning the analog/RF beamforming matrices, these are designed as random matrices [94] as we require sensing matrix to be random to be able to apply CS. The TX and the RX share a pseudo-random key so the RX can predict the precoding matrix. In particular, the angles of precoding/combining matrices are generated as random variables following a uniform distribution, i.e., $\tilde{\phi}_i(t) \sim \mathcal{U}(0, 2\pi)$. Then, for each training instance t and $\forall\, k = 1, \ldots, N_{\mathrm{T}}, i = 1, \ldots, L_{\mathrm{T}}$, we use the matrix:

$$[\mathbf{F}_{\mathrm{RF}}(t)]_{ki} = \frac{1}{\sqrt{N_{\mathrm{T}}}} e^{j(k-1)\sin(\tilde{\phi}_i(t))}, \tag{4.15}$$

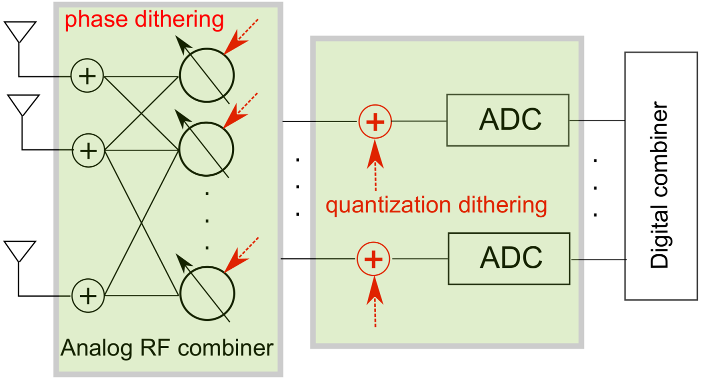


Figure 4.2: Dithered beamforming architecture where the random control signals are represented by red dashed arrows [96].

for the precoder, and accordingly for the combiner at the RX:

$$[\mathbf{W}_{\mathrm{RF}}(t)]_{ki} = \frac{1}{\sqrt{N_{\mathrm{R}}}} e^{j(k-1)\sin(\tilde{\phi}_i(t))}. \tag{4.16}$$

Before proceeding, let us describe in more detail two main issues that render the channel estimation problem more challenging in the case of channel estimation at the RX of a hybrid MIMO system and low resolution quantization. The first issue comes from the channel subspace sampling limitation [95] which prevents the direct estimation of the CSI due to the beamforming matrices. In the conventional case, where the beamforming matrices are composed by DFT columns, the resulting measurement matrix $\mathbf{\Psi}_{\mathrm{c}}$ has a block structure with areas of similar values [96]. This implies that $\mathrm{rank}(\mathbf{\Psi}_{\mathrm{c}}) = \mathrm{rank}(\mathbf{W}_{\mathrm{RF}}^{H}(t)\mathbf{D}_{\mathrm{R}}) \leq N_{\mathrm{T}}N_{\mathrm{R}}$. Moreover, taking into account the quantization of the received signal, the overall system, given by (4.5), is a non-linear one due to the staircase ADCs, especially for the low resolution cases, i.e., 1-3 bit.

To overcome the quantization non-linearity effects at the RX, we employ quantization dithering [97]. Dithering is a commonly used technique where an external signal is injected to the input to combat the non-linear quantization effects, improve the robustness and the asymptotic stability of the system [98,99]. In a design like ours, two external signals are injected at the MIMO RX, one in the spatial angles and another in the amplitude, as shown in Fig. 4.2 and discussed in [96]. Therefore, the use of dithering is two-fold: first we improve the properties of the measurement matrix by introducing randomness into the signal capturing process. Afterwards, the outputs of the RF combiner are perturbed by adding random analog memory-less signals to overcome the stair-case effects of low resolution ADCs. In this work we consider a simple type of dithering termed as non-subtractive random dithering. The additional dithering term in noise can be considered as an artificial noise and the concept is similar to the method of stochastic resonance. The concept of stochastic resonance comes into existence when there are system nonlinearities and the increases in levels of unpredictable fluctuations, e.g., random noise, cause an increase in a metric of the quality of signal transmission or detection performance rather than a decrease. For instance, [100] suggested that detection performance can be improved by adding an independent noise to the data under certain conditions. Specifically, we assume that a Gaussian random signal with zero mean, i.e., $\bar{\mathbf{d}} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{d}}^2\mathbf{I})$ is added to the input, thus, the overall system is described as:

$$\bar{\mathbf{r}} = \mathcal{Q}\big(\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}} + \bar{\mathbf{n}} + \bar{\mathbf{d}}\big) \in \mathbb{R}^{2TN_{\mathrm{R}}\times 1}, \tag{4.17}$$

where $\bar{\mathbf{d}} \in \mathbb{R}^{2TN_{\mathrm{R}}\times 1}$ is the control signal. The overall noise can be modelled as $\bar{\mathbf{n}} + \bar{\mathbf{d}} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2\mathbf{I}\big)$, where $\sigma^2 = \sigma_{\mathrm{n}}^2 + \sigma_{\mathrm{d}}^2$.

4.3.2 EM-SURE-GAMP Based Proposed Solution

To solve the non-linear sparse channel estimation problem of (4.14), we obtain an approximation of the maximum a-posteriori channel estimator via the EM algorithm [83], for l-th iteration, i.e.,

$$\mathbb{E}_{\bar{\mathbf{y}}|\bar{\mathbf{r}},\mathbf{z}_{\mathrm{c}}}\left\{\frac{\partial}{\partial \mathbf{z}_{\mathrm{c}}}\ln p(\bar{\mathbf{r}},\bar{\mathbf{y}}|\mathbf{z}_{\mathrm{c}(l)})\right\}=0, \tag{4.18}$$

where the conditional Probability Density Function (PDF) $p(\bar{\mathbf{r}},\bar{\mathbf{y}}|\mathbf{z}_{\mathrm{c}})$ involving $\bar{\mathbf{r}}$ and $\bar{\mathbf{y}}$ random variables is computed based on [101]. The EM algorithm which solves (4.17), is described by the following steps for the $(l+1)$-th iteration:

- **E-step:** In the first step compute a vector $\mathbf{b}_l=[b_{1(l)},\dots,b_{2TN_{\mathrm{R}}(l)}]$ with entries

$$b_{i(l)}=-\frac{\sigma}{\sqrt{2\pi}}\frac{e^{-\frac{(l_i-[\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}(l)}]_i)^2}{2\sigma^2}}-e^{-\frac{(u_i-[\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}(l)}]_i)^2}{2\sigma^2}}}{\mathrm{erf}(\frac{-l_i+[\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}(l)}]_i}{\sqrt{2}\sigma})-\mathrm{erf}(\frac{-u_i+[\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}(l)}]_i}{\sqrt{2}\sigma})} \tag{4.19}$$

  where l_i,u_i are the lower and upper bounds of the quantizer for $[\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}(l)}]_i$ respectively; $\mathrm{erf}(\cdot)$ is the error function.

- **M-step:** In the second step estimate the sparse channel $\mathbf{z}_{\mathrm{c}(l+1)}\in\mathbb{R}^{2N_{\mathrm{R}}N_{\mathrm{T}}\times 1}$ by solving the linear system of equations:

$$\mathbf{A}\mathbf{z}_{\mathrm{c}(l+1)}=\boldsymbol{\beta}_l, \tag{4.20}$$

  with $\boldsymbol{\beta}_l\triangleq\bar{\mathbf{\Psi}}^T\bar{\mathbf{\Psi}}\mathbf{z}_{\mathrm{c}(l)}+\mathbf{b}_l$ and $\mathbf{A}\triangleq\bar{\mathbf{\Psi}}^T\bar{\mathbf{\Psi}}+\mathbf{C}_{\mathrm{h}}^{-1}$ where $\mathbf{C}_{\mathrm{h}}^{-1}$ is the correlation matrix based on the channel known statistics.

The performance of the EM algorithm is determined by which solution we use for the linear system of equations in (4.20). Given that prior PDF of the CSI, i.e., $p([\mathbf{Z}]_{ik})$, is known, several sparse solvers can be employed for the estimation of $\mathbf{z}_{\mathrm{c}}$,

Algorithm 5 Proposed EM-SURE-GAMP Algorithm for Channel Estimation

```
**Initialization**: $\hat{\mathbf{z}}_1 = \mathbf{0}, \boldsymbol{\xi}_0 = \mathbf{0}, c_1 = \frac{1}{2N_\mathrm{R}N_\mathrm{T}}, \tau_{\mathrm{z}(1)} = 1$.
// start GAMP iteration [108]
**for** t = $1, \ldots, T_\mathrm{max}$ **do**
    $\boldsymbol{\gamma}_t = \mathbf{A}\hat{\mathbf{z}}_t$
    $\tau_{\mathrm{p}(t)} = \frac{1}{2N_\mathrm{R}N_\mathrm{T}}\|\mathbf{A}\|_F^2\tau_{\mathrm{z}(t)}$
    $\mathbf{p}_t = \boldsymbol{\gamma}_t - \tau_{\mathrm{p}(t)}\boldsymbol{\xi}_{t-1}$
// Compute EM-steps from (4.20)
    Update $\boldsymbol{\delta}_l$ using EM-steps.
// start parametric SURE-AMP steps [93]
    Compute estimate of conditional probability distribution $p(\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l)$ as $\boldsymbol{\xi}_t = \mathbb{E}_{p(\boldsymbol{\gamma}_t|\mathbf{p}_t,\tau_{\mathrm{p}(t)},\boldsymbol{\delta}_l)}[\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l]$
    Compute $\tau_{\xi(t)} = \frac{1}{2N_\mathrm{R}N_\mathrm{T}\tau_{\mathrm{p}(t)}}\left[1 - \frac{\mathrm{Var}_{p(\boldsymbol{\gamma}_t|\mathbf{p}_t,\tau_{\mathrm{p}(t)},\boldsymbol{\delta}_l)}[\boldsymbol{\gamma}_t|\mathbf{p}_t,\tau_{\mathrm{p}(t)},\boldsymbol{\delta}_l]}{\tau_{\mathrm{p}(t)}}\right]$ using variance of $p(\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l)$
    $\frac{1}{\tau_{\beta(t)}} = \frac{1}{2N_\mathrm{R}N_\mathrm{T}}\|\mathbf{A}\|_F^2\tau_{\xi(t)}$
    Compute noisy version of signal $\hat{\mathbf{z}}_t$ as $\boldsymbol{\beta}_t = \hat{\mathbf{z}}_t + \tau_{\beta(t)}\mathbf{A}^*\boldsymbol{\xi}_t$
    Select $\boldsymbol{\theta}_t = H_t(\boldsymbol{\beta}_t, c_t)$ where parameter selection function $H_t$ is designed as a function of noisy data $\boldsymbol{\beta}_t$ and effective noise covariance $c_t$
    Compute new signal estimate $\hat{\mathbf{z}}_{t+1} = f_t(\boldsymbol{\beta}_t, c_t|\boldsymbol{\theta}_t)$ by denoising $\boldsymbol{\beta}_l$ using parametric denoising function $f_t(\cdot|\boldsymbol{\theta}_t)$
    $\tau_{\mathrm{z}(t+1)} = \tau_{\beta(t)}f'_t(\boldsymbol{\beta}_t, c_t|\boldsymbol{\theta}_t)$
    Estimate effective noise variance $c_{t+1} = \frac{1}{2N_\mathrm{R}N_\mathrm{T}}\|\tau_{\beta(t)}\boldsymbol{\xi}_t\|_2^2$
**end for**
```

e.g., AMP [102], CoSaMP [103], SGP [104], offering trade-offs between complexity, performance and prior knowledge. Since the matrix dimensions are expected to be very large in the massive MIMO case, matrix inversion is prohibitively complex.

The linear channel estimation problem in (4.20) can be considered similar to the noisy quantized CS problem [105]; among the numerous existing algorithms for sparse inverse linear problems, the AMP-based solver has been shown to converge faster, i.e., in few iterations, with predictable dynamics together with low computational complexity. In its original formulation for l_1-minimization [102], AMP is designed as a variant of a soft-thresholding iterative algorithm;

in [106, 107] extensions of AMP have been used to handle a wide class of random sensing matrices and for sparse learning applications.

Note that the tendency of a GAMP algorithm [108] is to approach a computationally difficult problem by a sequence of simple scalar estimation problems with matrix multipliers. When the matrix $\mathbf{A}$ is very large with i.i.d. sub-Gaussian entries, GAMP is characterized by scalar state evolution [108] and when this state evolution approaches a unique fixed point, GAMP converges to the Minimum Mean Square Error (MMSE) solution. However, in practice, $\mathbf{A}$ may not be very large with i.i.d. sub-Gaussian entries, even in that case the estimate provided by the GAMP algorithm after a few iterations are often very close to the MMSE [87].

In our context, since the channel noise model in (4.17) is quantized Gaussian as it is modeled as the quantization function, we need to adopt the generalized version of AMP, i.e., GAMP [108], whose computation is detailed in the Algorithm 5 where the expectation is over the posterior probability $p(\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l)$ which is dependent on the quantizer function $\mathcal{Q}$ through (4.19). In particular, this algorithm performs a sequence of MMSE estimations on the product $\mathbf{A}\hat{\mathbf{z}}_t$ (which is denoted as $\boldsymbol{\gamma}_t$) where $\hat{\mathbf{z}}_t$ refers to the estimate of the vector $\mathbf{z}_{\mathrm{c}(l+1)}$ for the M-step in (4.20) and l is the EM iteration index. The vector $\boldsymbol{\delta}_l$ is updated using the EM-steps as indicated in (4.20). In Algorithm 5, lines 7 and 8 represent the estimate and variance of conditional probability distribution $p(\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l)$ denoted as $\mathbb{E}_{p(\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l)}[\cdot]$ (the value results in $\boldsymbol{\xi}_t$ which is used in following steps) and $\mathrm{Var}_{p(\boldsymbol{\gamma}_t|\mathbf{p}_t, \tau_{\mathrm{p}(t)}, \boldsymbol{\delta}_l)}[\cdot]$, respectively. Regarding the MMSE estimator for $\hat{\mathbf{z}}_t$, standard AMP [102] is based on the assumption that the prior $p(\hat{\mathbf{z}}_t)$ is precisely defined and, therefore, it is possible to derive the associated MMSE estimator.

In this case, we utilize a variant, named SURE-GAMP, which derives specific MMSE estimators tailored for the dithered system model in (4.17) as follows. The

SURE approach [93] aims to find the denoiser within a class with the least MSE by optimizing the free parameters $\boldsymbol{\theta}_t$ of some piecewise kernel functions $f_t(\cdot|\boldsymbol{\theta}_t)$ in order to obtain an optimal adaptive non linearity; moreover, the optimization of the denoiser does not require knowledge of the prior distribution. In the simulations, SURE-GAMP uses a family of parameterized denoising functions for the class of Bernoulli Gaussian signals, which can be analyzed through the Gaussian-mixture distribution as well [89]. At each iteration, the parametric SURE-GAMP algorithm adaptively chooses the best denoiser, i.e. the one with the least MSE, by selecting the parameters $\boldsymbol{\theta}^t$ which correspond to the minimum of the selection function H_t, such as in line 11 of Algorithm 5, dependent on the noisy data $\boldsymbol{\beta}_t$ and the estimate of the effective noise variance c_t which leads to solving the following optimization problem:

$$\begin{aligned}\theta_t &= H_t(\boldsymbol{\beta}_t, c_t) \\ &= \arg\min_{\theta} \mathbb{E}[f(\boldsymbol{\beta}_t, c_t|\boldsymbol{\theta}) - \boldsymbol{\beta}_t)^2 + 2c_t f'(\boldsymbol{\beta}_t, c_t|\boldsymbol{\theta})]\end{aligned} \tag{4.21}$$

In [93], the authors have shown that this optimization is equivalent to solving a linear system of equations whose dimension equals the number of kernel functions which are the number n_{ker} of basis functions representing $f(\cdot|\boldsymbol{\theta})$ ($n_{\text{ker}} = 3$, in the simulations). Therefore, the overall complexity of SURE-GAMP is dominated by the matrix-vector multiplications in lines 3 and 10 of Algorithm 5, whose order is $\mathcal{O}((N_{\text{R}}N_{\text{T}})^2)$. The EM steps are combined with the SURE-GAMP algorithm to avoid the need of specifying a prior probability on $\mathbf{z}_{\text{c}(l+1)}$. The algorithm converges after a few iterations when a solution close to minimum MSE is achieved. In the next section we proceed with the low resolution DACs case for EE maximization and the simulation results for both the sparse channel estimation and EE maximization problems are presented in Section 4.5.

4.4 Proposed EE Maximization Approach

4.4.1 Problem Formulation

Similar to the EE equation in (3.10) where it is a function of the diagonal sparse matrix $\mathbf{P}_{\mathrm{TX}}$, we can again define the EE of a point-to-point MIMO system as the ratio of the information rate and the total consumed power [109]. Note that, in this context we consider low resolution DACs at the TX, so the rate and power quantities depend on the distortion of the DACs, i.e., δ and the number of the RF chains, i.e., L_{T}, thus the EE can be expressed as

$$\mathrm{EE}(L_{\mathrm{T}}, \delta) \triangleq \frac{R(L_{\mathrm{T}}, \delta)}{P(L_{\mathrm{T}}, \delta)} \text{ (bits/Joule).} \tag{4.22}$$

Exploiting the linearity property of the quantization model in (4.9), the information rate $R(L_{\mathrm{T}}, \delta)$ is expressed as:

$$R(L_{\mathrm{T}}, \delta) = \log_2 |\mathbf{I}_{N_{\mathrm{s}}} + \frac{1}{N_{\mathrm{s}}} \mathbf{R}_{\eta}^{-1} \mathbf{H}_{\mathrm{eff}} \mathbf{H}_{\mathrm{eff}}^{H}| \text{ (bits/s/Hz),} \tag{4.23}$$

where the values of L_{T} and δ will affect the noise covariance matrix $\mathbf{R}_{\eta}(L_{\mathrm{T}}, \delta)$ and the effective channel $\mathbf{H}_{\mathrm{eff}}(L_{\mathrm{T}}, \delta)$.

Concerning the power consumption model as described in Section 2.2.4 for the case of low resolution sampling at the TX and δ being the distortion of the DACs, we consider that the total power consumption $P(L_{\mathrm{T}}, \delta)$ is proportional to:

$$P(L_{\mathrm{T}}, \delta) \propto L_{\mathrm{T}} \left[P_{\mathrm{DAC}} \left(\frac{\pi\sqrt{3}}{2(1-\delta^2)} \right)^{1/2} + N_{\mathrm{T}} P_{\mathrm{PS}} \right] \text{ (W),} \tag{4.24}$$

where P_{DAC} and P_{S} depend upon the DAC and phase-shifter power consumption values, respectively.

Given the expressions (4.23) and (4.24), we can now define the EE maximization problem as a fractional programming problem:

$$\arg\max_{L_{\text{T}},\delta} \ \text{EE}(L_{\text{T}},\delta) \text{ subject to } P(L_{\text{T}},\delta) \leq P_{\text{max}}, \tag{4.25}$$

where P_{max} is the maximum available power budget. Our goal, by solving (4.25), is to obtain the number of RF chains and bit resolution in an optimal manner. To obtain a solution to (4.25) we have developed an iterative procedure that approximates the initial fractional problem with a convex-concave optimization, using the Dinkelbach approximation [70] and subset selection. The Dinkelbach approach makes an iterative approximation of the fractional problem with a sequence of non-fractional but constrained optimization ones. Although simpler, each one of these problems is still non-convex. However, by decomposing the contribution of each RF chain to the EE performance of the system, we can employ subset selection methods which minimize the number of RF chains by solving an ℓ_1 approximation to the non-convex problem.

Before proceeding with the description of the proposed technique, we derive a technique based on exhaustive search for EE maximization, which will serve as an upper bound for comparison with the proposed method.

4.4.2 Upper Bound on EE via Exhaustive Search

To obtain an upper bound, we consider the case where $L_{\text{T}} = N_{\text{T}}$. This simplifies the computation of the beamformers at the RX and the RX, by using the SVD of the channel. However, since we change the number of the RF chains/antennas, the channel and its SVD, has to be updated at each time. Specifically, an exhaustive search approach is needed to obtain the optimum EE over all possible values of $(L_{\mathrm{T}}, \delta) \in \{1, \ldots, b_{\max}\} \times \{1, \ldots, L_{\mathrm{T}}\}$. For each set value (L_{T}, δ), the SVD of the effective channel has to be obtained, i.e.,

$$\mathbf{H}_{\mathrm{eff}}(L_{\mathrm{T}}, \delta) = \delta \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{H}, \tag{4.26}$$

where $\mathbf{U} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ and $\mathbf{V} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ are unitary matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. We assume that the rank of the channel is r.

Hence, the rate expression in (4.23) becomes:

$$\begin{aligned} R(L_{\mathrm{T}}, \delta) &= \log_2 |\mathbf{I}_{N_{\mathrm{s}}} + \frac{\delta^2}{N_{\mathrm{s}}} \mathbf{R}_{\eta}^{-1} \mathbf{W}^{H} \mathbf{H} \mathbf{F} \mathbf{F}^{H} \mathbf{H}^{H} \mathbf{W}| \\ &= \log_2 |\mathbf{I}_{N_{\mathrm{s}}} + \frac{\delta^2}{N_{\mathrm{s}}} \mathbf{R}_{\eta}^{-1} \mathbf{\Sigma}\mathbf{\Sigma}^{H}| \\ &= \sum_{i=1}^{r} \log_2 \left(1 + \frac{\delta^2}{N_{\mathrm{s}}} [\mathbf{R}_{\eta}^{-1}]_{ii} [\mathbf{\Sigma}\mathbf{\Sigma}^{H}]_{ii}\right), \end{aligned} \tag{4.27}$$

where $\mathbf{R}_{\eta}$ becomes a diagonal matrix with entries $[\mathbf{R}_{\eta}]_{ii} = \sigma_{\epsilon}^2 [\mathbf{\Sigma}\mathbf{\Sigma}^{H}]_{ii} + \sigma_n^2$. Based on (4.27), the rate expression is decomposed into the singular value domain, thus, the number of the rank r represents the *virtual* number of RF chains. So, the goal here is to reduce the number of virtual RF chains r, alongside with the distortion δ which depends on the bit resolution b.

Algorithm 6 shows the exhaustive search approach (similar to [65] and the concept of exhaustive search discussed in Section 3.4.4 of Chapter 3), called the Brute Force (BF) technique, thus, it provides the solution to achieve the optimal number of RF chains and the optimal number of associated DAC bits at each channel realization. It makes a search of all the possible number of RF

Algorithm 6 Brute Force (BF) Approach for RF Chain and DAC Bit Resolution Selection

```
Input: $b_{\max}$, $\mathbf{H}$
Begin:
1. for $b = 1, ..., b_{\max}$
2.     Compute $\delta(b)$ based on (4.7)
3.     for $l_{\mathrm{t}} = 1, ..., N_{\mathrm{T}}$
4.         Compute the SVD of $\mathbf{H}_{\mathrm{eff}}(l_t, \delta(b_i))$ based on (4.26)
5.         Compute $\mathrm{EE}(l_{\mathrm{t}}, \delta(b))$ based on (4.23) and (4.24)
6.     end
7. end
8. Find the $L_{\mathrm{T}}^{opt}$ and $b^{opt}$ such as $\mathrm{EE}(L_{\mathrm{T}}^{\mathrm{opt}}, \delta(b^{\mathrm{opt}})) > \mathrm{EE}(l_t, \delta(b))\, \forall (b, l_t)$
Output: $L_{\mathrm{T}}^{\mathrm{opt}}$ and $b^{\mathrm{opt}}$
```

chains/antennas, i.e., $l_{\mathrm{t}} = \{1, ..., N_{\mathrm{T}}\}$ and over the available bit resolution, i.e., $b = 1, ..., b_{\max}$, where $b_{\max}$ is the highest achievable resolution. It then finds the best EE out of all possible efficiency values and chooses the corresponding optimal number of active RF chains L_{T}^{opt} and the optimal resolution sampling b^{opt} for the TX. This method provides the best possible EE performance assuming that the SVD of $\mathbf{H}$ is perfectly known at the TX.

4.4.3 Proposed Dinkelbach Method (DM) with Subset Selection Optimization

Let us now consider an optimal design where we seek the sampling resolution for each DAC and the optimal number of active RF chains L_{T} that will maximize the EE of the TX. We consider a variable number of RF chains, i.e., by using switches to activate/deactivate each one independently [66], then the problem becomes:

$$\arg\max_{\mathbf{S},\delta} \frac{R(\mathbf{S}, \delta)}{P(\mathbf{S}, \delta)} \text{ subject to } P(\mathbf{S}, \delta) \leq P_{\max}, \tag{4.28}$$

where $\mathbf{S} \in \{0, 1\}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ is a diagonal binary matrix representing switches which activate or deactivate the RF chains. Hence, the resulting optimization problem of (4.28) has two unknown quantities to be recovered, the matrices $\mathbf{S}$ and δ. We transform the problem into a subset selection based problem considering sparse optimization and compressive sampling.

We consider the problem to be equivalent to finding only a sparse selection vector, $\mathrm{diag}(\mathbf{S}) \in \{0,1\}^{L_{\mathrm{T}} \times 1}$, where each unity value represents one active RF chain with a predefined resolution, while the zero value represents an inactive RF chain. It is important to note that based on the proposed architecture, the optimization problem does not consider a predefined number of active/inactive RF chains, but this quantity is an optimization variable. Incorporating this selection procedure into our formulation, the received signal $\hat{\mathbf{r}} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ after the baseband RX, which is the modified expression of the output RX signal $\mathbf{r}$ in (4.12), is expressed as

$$\hat{\mathbf{r}} = \delta \mathbf{W}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{S} \mathbf{F}_{\mathrm{BB}} \mathbf{s} + \boldsymbol{\eta}, \tag{4.29}$$

where $\mathbf{S} \in \{0,1\}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ is a diagonal selection matrix composed by zeros and ones, with $[\mathbf{S}]_{kk} \in \{0,1\}$ and $[\mathbf{S}]_{kl} = 0$ for $k \neq l$; the term $\delta \mathbf{W}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{S} \mathbf{F}_{\mathrm{BB}}$ is the effective channel $\hat{\mathbf{H}}_{\text{eff}} \in \mathbb{C}^{N_{\mathrm{s}} \times N_{\mathrm{s}}}$ in this case, including hybrid TX precoding and RX combining and quantization distortion. The parameter that we aim to optimize in (4.29) is now the entries of the diagonal selection matrix $\mathbf{S} \in \{0,1\}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$. The effective channel can be decomposed as:

$$\begin{aligned}
\hat{\mathbf{H}}_{\text{eff}} &= \delta \mathbf{W}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{S} \mathbf{F}_{\mathrm{BB}} && (4.30)\\
&= \sum_{i=1}^{L_{\mathrm{T}}} [\mathbf{S}]_{ii} [\delta \mathbf{W}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}}]_i [\mathbf{F}_{\mathrm{BB}}^T]_i^T \\
&= \sum_{i=1}^{L_{\mathrm{T}}} [\mathbf{S}]_{ii} \mathbf{a}_i \mathbf{b}_i^T, && (4.31)
\end{aligned}$$

where $\mathbf{b}_i \triangleq [\mathbf{F}_{\mathrm{BB}}^T]_i \in \mathbb{C}^{N_s \times 1}$, $\mathbf{a}_i \triangleq [\delta \mathbf{R}_\eta^{-\frac{1}{2}} \mathbf{W}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}}]_i \in \mathbb{C}^{N_s \times 1}$ and where

$[\mathbf{S}]_{ii} \in \{0, 1\}$ determines the state of the i-th RF chain. Based on (4.31), the received signal can be equivalently expressed as the following measurement vector:

$$\hat{\mathbf{r}} = \sum_{i=1}^{L_\mathrm{T}} [\mathbf{S}]_{ii} \mathbf{a}_i (\mathbf{b}_i^T \mathbf{s}) + \hat{\boldsymbol{\eta}}, \tag{4.32}$$

where $\hat{\boldsymbol{\eta}} \triangleq \mathbf{S}\boldsymbol{\eta}$ whose noise covariance matrix, which is the modified expression of the noise covariance matrix $\mathbf{R}_\eta$ in (4.13), can be expressed in terms of the selection matrix as

$$\hat{\mathbf{R}}_\eta = \mathbb{E}[\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}^H] = \sigma_\epsilon^2 \mathbf{W}^H \mathbf{H} \mathbf{F}_\mathrm{RF} \mathbf{S} \mathbf{F}_\mathrm{BB} \mathbf{F}_\mathrm{BB}^H \mathbf{S} \mathbf{F}_\mathrm{RF}^H \mathbf{H}^H \mathbf{W} + \sigma_\mathrm{n}^2 \mathbf{W}^H \mathbf{W}. \tag{4.33}$$

The problem becomes equivalent with the estimation of $\mathbf{S}$ that maximizes the EE of the hybrid precoder. It can be shown that the rate and power equations for such scenario can be expressed as:

$$R(\mathbf{S}, \delta) = \log_2 \left| \mathbf{I}_{N_\mathrm{s}} + \frac{1}{N_\mathrm{s}} \sum_{i=1}^{L_\mathrm{T}} [\mathbf{S}]_{ii} \mathbf{a}_i^H \mathbf{a}_i \mathbf{b}_i \mathbf{b}_i^H \right|, \tag{4.34}$$

and

$$P(\mathbf{S}, \delta) \propto \sum_{i=1}^{L_\mathrm{T}} [\mathbf{S}]_{ii} \left[P_\mathrm{DAC} \left(\frac{\pi\sqrt{3}}{2(1-\delta^2)} \right)^{1/2} + N_\mathrm{T} P_\mathrm{PS} \right] \tag{4.35}$$

$$= L_\mathrm{T} \left[P_\mathrm{DAC} \left(\frac{\pi\sqrt{3}}{2(1-\delta^2)} \right)^{1/2} + N_\mathrm{T} P_\mathrm{PS} \right] \ (\mathrm{W}). \tag{4.36}$$

The problem of maximizing EE (4.28) is a concave-convex fractional problem and one solution method is the Dinkelbach approximation [70]. The DM, as discussed in Chapter 3 already, is an iterative and parametric algorithm, where a sequence of easier problems converge to the global solution. Let

Algorithm 7 Proposed Dinkelbach Method (DM) for RF Chain and DAC Bit Selection

```
Input: $\kappa^{(0)}$, $\mathbf{H}$
Begin:
1. for $b = 1, ..., b_{\max}$
2.   Compute $\mathbf{H}_{\text{eff}}(N_{\text{T}}, \delta(b))$ in (4.26).
3.     for $m = 1, 2, \ldots, I_{\max}$
4.       Obtain $\mathbf{S}^{(m)}$ by solving (4.37) given $\kappa^{(m-1)}$.
5.       Calculate $R(\mathbf{S}^{(m)}, \delta^{(m)})$ in (4.34) and $P(\mathbf{S}^{(m)}, \delta^{(m)})$ in (4.35).
6.       Compute $\kappa^{(m)} = R(\mathbf{S}^{(m)}, \delta^{(m)})/P(\mathbf{S}^{(m)}, \delta^{(m)})$.
7.     end
8. end
Output: Optimal ${L_{\text{T}}}^{opt}$ and $b^{opt}$
```

$\kappa^{(m)} = R(\mathbf{S}^{(m)}, \delta^{(m)})/P(\mathbf{S}^{(m)}, \delta^{(m)}) \in \mathbb{R}$, for $m = 1, 2, \ldots, I_{\max}$, where $I_{\max}$ is the maximum number of iterations, then each iteration step of Dinkelbach can be expressed as:

$$\mathbf{S}^{(m)}(\kappa^{(m)}) \triangleq \arg\max_{\mathbf{S}\in\mathcal{S}} \left\{R(\mathbf{S}, \delta) - \kappa^{(m)} P(\mathbf{S}, \delta)\right\}, \tag{4.37}$$

where $\mathcal{S}$ is the set of diagonal matrices with the feasible bit allocations which satisfy $P(\mathbf{S}, \delta) \leq P_{\max}$. Algorithm 7 summarizes the Dinkelbach algorithm via the subset selection approach where the optimal number of RF chains and associated sampling resolution is obtained.

Computational Complexity of the Proposed DM

It can be observed that the DM via subset selection approach requires complexity order of only $b_{\max}\mathcal{O}(L_{\text{T}}^3)$ per iteration and the BF approach requires complexity order of $b_{\max}\mathcal{O}(L_{\text{T}}^2 N_{\text{T}})$. Since the number of the required iterations is usually very small (as shown in Fig. 4.7) as the **F** and **W** matrices are required to be computed in Algorithm 6 and not Algorithm 7, the overall complexity of the DM via the subset selection approach is much less than the BF approach.

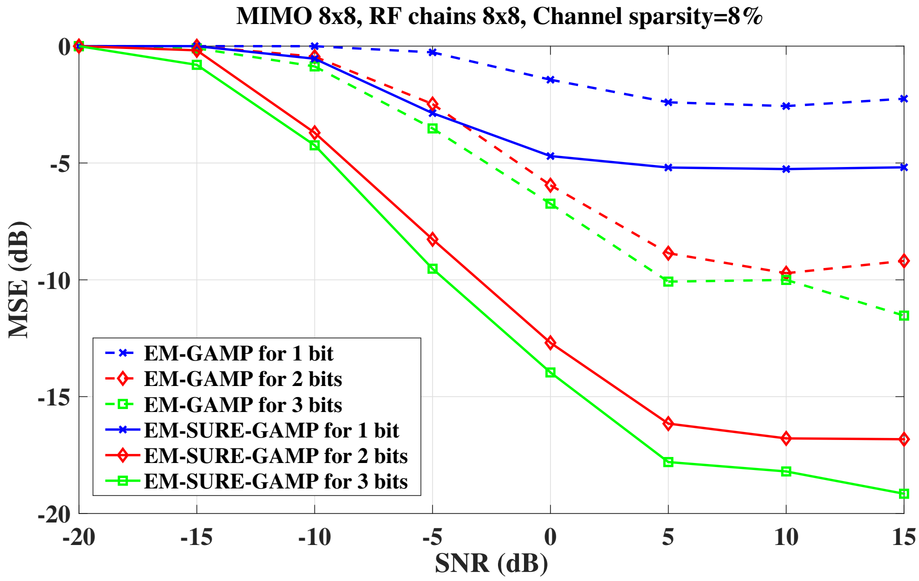


Figure 4.3: MSE versus SNR performance for the proposed EM-SURE-GAMP channel estimation.

4.5 Simulation Results

In this section, we evaluate the MSE performance of the proposed EM-SURE-GAMP technique for sparse channel estimation, and the EE and SE performance of the proposed DM for EE maximization using computer simulation results. The computer simulation results have been averaged over 100 and 1000 Monte-Carlo realizations, for the sparse channel estimation case and the EE maximization case, respectively. Note that MSE performance results of the proposed EM-SURE-GAMP algorithm are compared with the EM-GAMP solution. Reference [108] suggests the computation of the minimum MSE of the estimate; combined with EM steps we can plot the MSE results of EM-GAMP algorithm to compare with the proposed EM-SURE-GAMP solution.

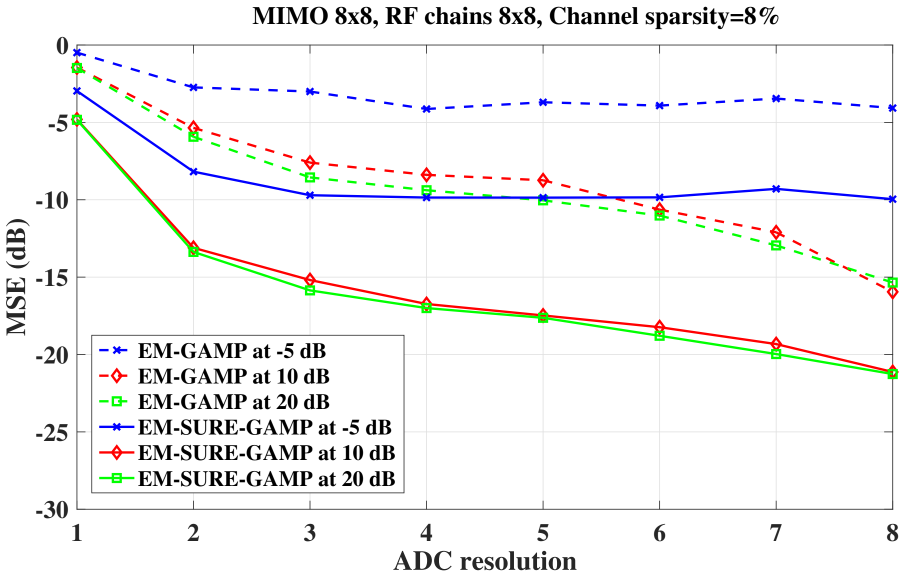


Figure 4.4: MSE versus the number of ADC bits for the proposed EM-SURE-GAMP channel estimation.

4.5.1 MSE Performance of Proposed EM-SURE-GAMP

System Setup: Following the condition $L_\mathrm{T} \leq N_\mathrm{T}$ and $L_\mathrm{R} \leq N_\mathrm{R}$ for a hybrid MIMO architecture, we consider a simple case of $N_\mathrm{T} = 8$, $N_\mathrm{R} = 8$, and the number of RF chains and streams equal to the number of antennas, i.e., $L_\mathrm{T} = L_\mathrm{R} = N_\mathrm{s} = 8$. It provides us easier computation for the analog precoder and combiner matrices. We can also consider fewer RF chains and streams than the number of antennas [17] to observe the channel estimation performance plots. The number of multipaths is 5 and due to low overload probability, the value of κ used in the quantization is 4. We run the proposed algorithm for $T_\mathrm{max} = 1$ and 100 EM iterations.

Fig. 4.3 shows the MSE variations w.r.t. SNR when comparing the EM-SURE-GAMP algorithm with EM-GAMP for 1-bit, 2-bit and 3-bit resolution ADCs. We can observe that the EM-SURE-GAMP algorithm achieves better

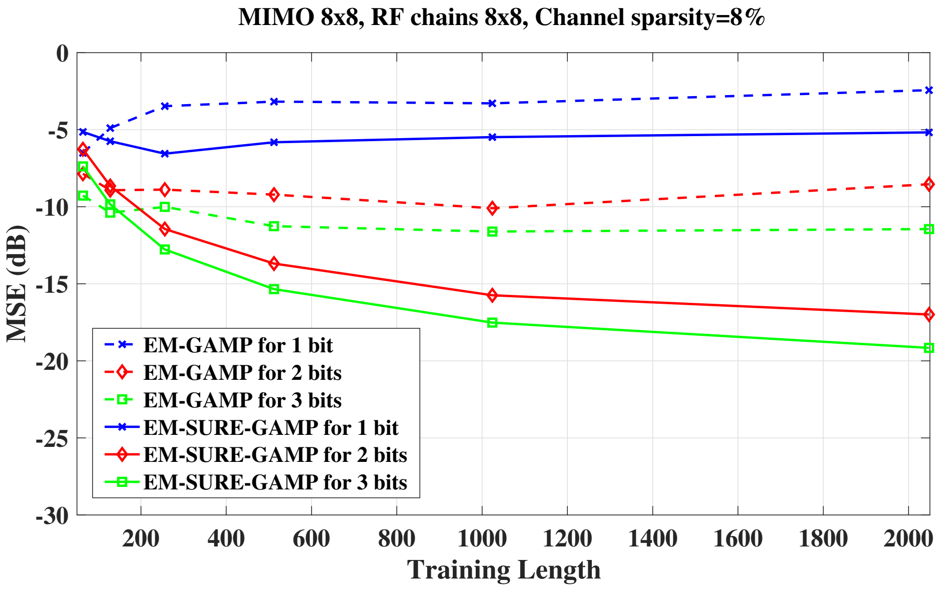


Figure 4.5: MSE versus the training length T for the proposed EM-SURE-GAMP channel estimation.

MSE performance for both low and high SNR regimes. For example, at SNR = 10 dB, the SURE algorithm variant outperforms EM-GAMP by about 3 dB in MSE terms for 1-bit quantization. For 2- and 3-bit, the MSE gain is around 2 dB.

Fig. 4.4 again shows that EM-SURE-GAMP performs better than EM-GAMP when MSE is plotted against the number of quantization bits for different values of SNR such as -5 dB, 10 dB and 20 dB. The training length for Fig. 4.3 and Fig. 4.4 is $T = 2^{11}$ and EM-SURE-GAMP shows good performance for a channel sparsity level of 8%, i.e., the ratio of non-zero entries of the beamspace channel to the product $N_{\mathrm{R}} \times N_{\mathrm{T}}$. It can be observed that, for example, with 3-bit resolution, a significant gain in MSE for the SURE variant of around $6 - 7$ dB compared to EM-GAMP is observed for all SNR values.

Fig. 4.5 shows that the EM-SURE-GAMP solution outperforms EM-GAMP solution w.r.t. the training length for a range of training sequence lengths of 64

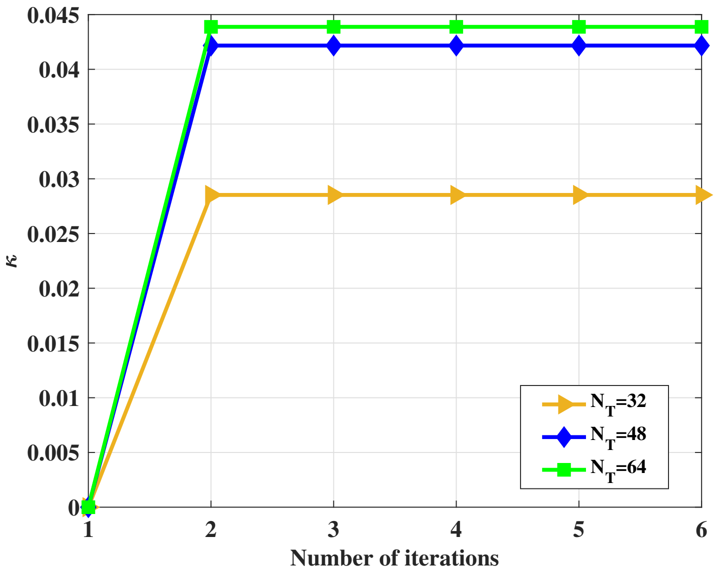


Figure 4.6: Convergence of the proposed DM for different number of TX antennas at SNR = 30 dB, $N_{\mathrm{R}} = 32$, $L_{\mathrm{T}} = 32$ and $N_{\mathrm{s}} = 8$.

to 2048 and converges more quickly than EM-GAMP for a channel sparsity level of 8%, 15 dB SNR, when 1-bit, 2-bit and 3-bit ADC resolutions are considered.

4.5.2 EE and SE Performance of Proposed DM

System Setup: $N_{\mathrm{T}} = 64$, $N_{\mathrm{R}} = 32$, $L_{\mathrm{T}} = 32$ (the number of available RF chains), $N_{\mathrm{s}} = 8$, $N_{\mathrm{cl}} = 2$, $N_{\mathrm{ray}} = 10$, and $\sigma^2_{\alpha,i} = 1$. The azimuth angles of departure and arrival are computed with uniformly distributed mean angles; each cluster follows a Laplacian distribution with mean angles equal to zero. The antenna elements in the ULA are spaced by distance $d = \lambda/2$. Concerning the quantization model, since DACs have the same sampling resolution for each RF chain the quantization distortion parameter is the same for all DACs and the highest bit resolution $b_{\max} = 8$. The typical values of power terms for the power model in (4.24) of Subsection 3.2.2 are $P_{\mathrm{PS}} = 10$ mW, $P_{\mathrm{DAC}} = 0.1$ W and $P_{\max} = 1$ W. We solve the sparse approximation problem for the RF and baseband precoding matrices $\mathbf{F}_{\mathrm{RF}}$

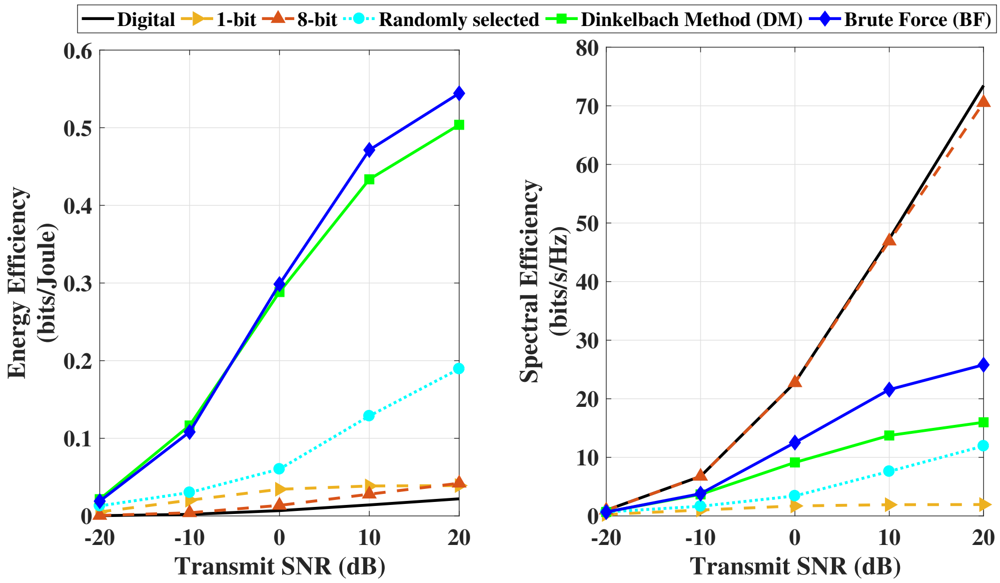


Figure 4.7: EE and SE performance comparison w.r.t. transmit SNR (dB) at $N_\mathrm{T} = 64$, $N_\mathrm{R} = 32$, $L_\mathrm{T} = 32$ and $N_\mathrm{s} = 8$.

and $\mathbf{F}_\mathrm{BB}$ using Orthogonal Matching Pursuit (OMP) [16, 17], and the combiner matrix $\mathbf{W}$ is the product of $1/\sqrt{N_\mathrm{s}}$ and the first N_s columns of the matrix $\mathbf{U}$.

For comparison with the proposed DM via subset selection solution, we have considered the digital beamforming architecture $(L_\mathrm{T} = N_\mathrm{T})$ with 8-bit DACs, which represents the optimum from the achievable SE perspective, combined analog and digital hybrid precoding with L_T RF chains for 1-bit and 8-bit DACs, which represent the lowest and the highest SE cases. We also compare with the hybrid beamforming for L_T RF chains with a random resolution selected for each DAC from the range $[1, 8]$-bit, and hybrid beamforming with the optimal number of active RF chains L_T^{opt} and corresponding optimal sampling resolution b^{opt} obtained from the BF approach.

Fig. 4.6 shows the convergence of the DM based solution as proposed in Algorithm 7 to obtain the optimal number of active RF chains and corresponding optimal sampling resolution. It can be observed that the performance curves

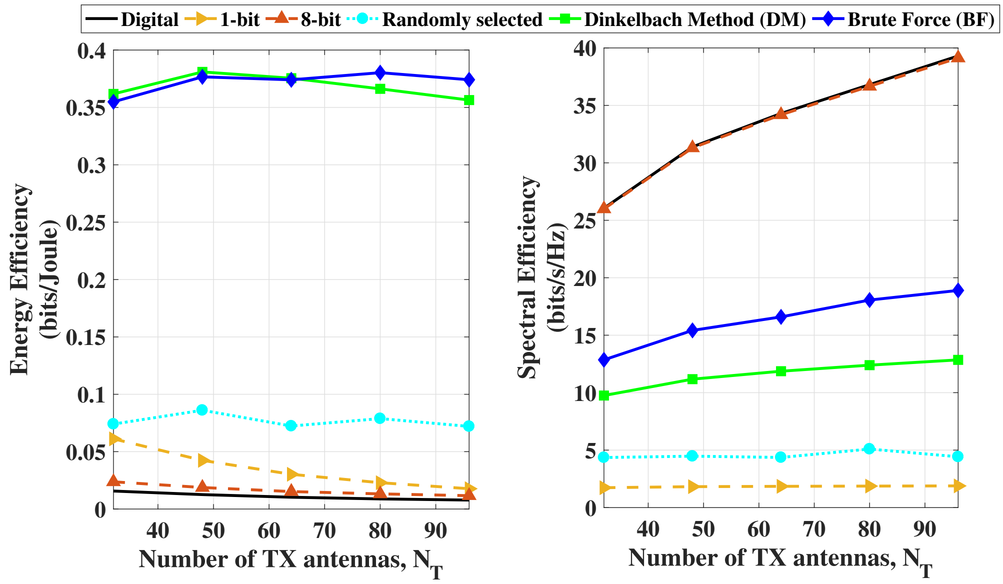


Figure 4.8: EE and SE performance comparison w.r.t. the number of TX antennas N_T at SNR $= 5$ dB, $N_\mathrm{R} = 32$, $L_\mathrm{T} = 32$ and $N_\mathrm{s} = 8$.

based on the current EE κ (step 6 of Algorithm 7) for different numbers of TX antennas increase w.r.t. the number of iterations. The proposed solution converges rapidly and needs only 2-3 iterations to converge.

It can be clearly observed from Fig. 4.7 that the proposed solution achieves a similar EE performance w.r.t. SNR as the BF approach and outperforms the hybrid 1-bit and hybrid 8-bit quantized DACs, plus the hybrid randomly selected resolution and digital beamforming with full-bit (8-bit) quantization. For example, at 10 dB SNR, the EE for the proposed DM solution is approximating the BF solution performance, about 0.3 bits/Joule better than the randomly selected resolution with hybrid beamforming, about 0.35 bits/Joule better than the hybrid 1-bit and about 0.38 bits/Joule better than the hybrid 8-bit and digital beamforming baselines. The proposed solution also achieves SE performance higher than the randomly selected and 1-bit quantization baselines. Only the digital beamforming and 8-bit hybrid baselines have better SE performance, but

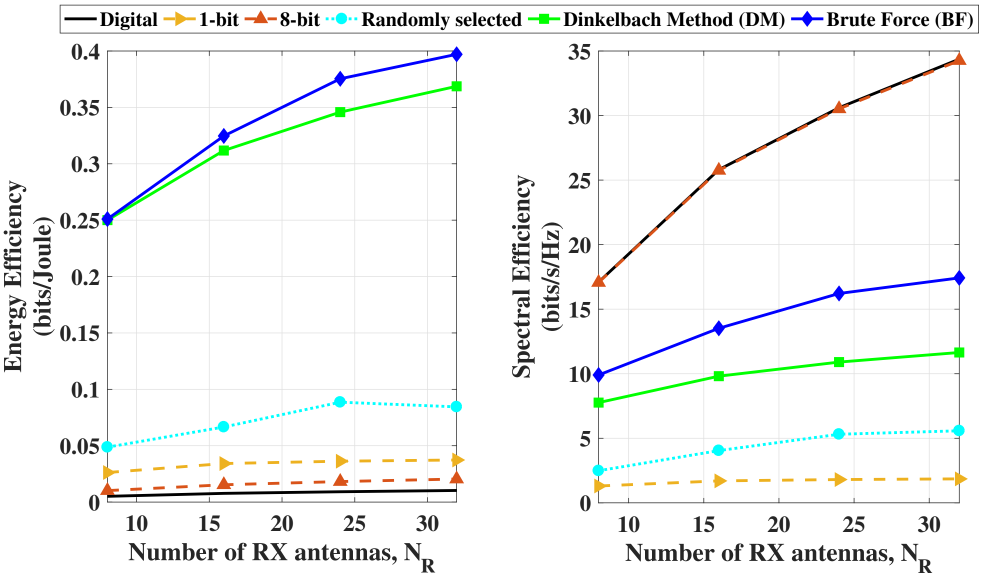


Figure 4.9: EE and SE performance comparison w.r.t. the number of RX antennas at SNR = 5 dB, $N_\mathrm{T} = 64$, $L_\mathrm{T} = 32$ and $N_\mathrm{s} = 8$.

this is achieved by using higher rate 8-bit quantization DACs. For example, at 0 dB SNR, the proposed solution outperforms randomly selected quantization by about 7 bits/s/Hz, 1-bit hybrid by about 9 bits/s/Hz. Concerning the lower SE performance of the proposed technique and the BF approach, this is due to the fact that BF has no constraint in the overall power consumption.

Fig. 4.8 shows similar performance behavior when plotting EE and SE w.r.t. the number of TX antennas at 5 dB SNR. For example, for $N_\mathrm{T} = 80$, the proposed solution demonstrates EE performance close to the BF approach, The DM performs about 0.3 bits/Joule and about 7.5 bits/s/Hz better than the hybrid randomly selected resolution baseline and about 0.35 bits/Joule and 10 bits/s/Hz better than the 1-bit hybrid baseline. Fig. 4.9 plots the performance comparison of the proposed solution with the baselines w.r.t. number of RX antennas at 5 dB SNR. Similar to above plots, it achieves high SE and has almost the same EE performance as the BF approach. For example, for $N_\mathrm{R} = 16$, the proposed solution demonstrates EE performance close to the BF approach. The DM solution performs about 0.25 bits/Joule and 5 bits/s/Hz better than the randomly selected resolution baseline, and about 0.275 bits/Joule and about 7.5 bits/s/Hz better than the 1-bit hybrid baseline.

4.6 Summary

In this chapter we discussed sparse channel estimation and EE maximization solutions with low resolution sampling at the ADCs and the DACs, respectively.

Firstly in Section 4.3, we propose an efficient algorithm based on the AMP framework to estimate the sparse mmWave channel in a hybrid MIMO system with low resolution ADCs at the RX. The EM-SURE-GAMP algorithm is proposed and exploited to estimate the channel which provides the flexibility to avoid strong assumptions on the channel priors where SURE, depending on the noisy observation, is minimized to adaptively optimize the denoiser within the parametric class at each iteration. When compared with the state of the art EM-GAMP solution, the MSE of the proposed solution performs better w.r.t. low and high SNR regimes, w.r.t. the number of ADC bits, and w.r.t. the training length.

Secondly, in Section 4.4, we consider a mmWave hybrid MIMO system with analog and digital parts connected with fewer number of RF chains than the transmitting antennas, while TX DACs operate with low resolution sampling. We consider the case where all DACs have the same sampling resolution for each RF chain and aim to optimize the number of active RF chains and associated resolution of DACs. The proposed method achieves similar EE performance with the upper bound of the derived exhaustive search approach, while it exhibits lower computational complexity and fast convergence.

In the next chapter, we use the low resolution sampling at both the TX and the RX simultaneously, and include the joint DAC/ADC bit allocation and HBF optimization for EE maximization with varying bit resolutions unlike the EE maximization case in this chapter.

Chapter 5

EE Maximization by Joint Bit Allocation and Hybrid Beamforming Optimization

5.1 Introduction

The A/D HBF MIMO systems reduce the hardware complexity and power consumption through using fewer RF chains and optimizing the number of RF chains with full resolution sampling provides an energy efficient system. An alternative solution to reduce the power consumption and hardware complexity is by reducing the bit resolution [15] of the DACs and the ADCs. In the previous chapter, an efficient sparse mmWave channel estimation algorithm is designed for a HBF MIMO system with low resolution sampling at the ADCs. Furthermore, the low resolution sampling is implemented at the DACs and bit resolution with active RF chains selection is optimized to achieve high EE gains. The EE maximization work in Chapter 4 discusses a low resolution sampling setup but all DACs choose the same sampling resolution for each RF chain. In this chapter, we provide the flexibility in choosing the bit resolution for each DAC and ADC, and a joint optimization problem is formulated involving both the TX and the RX. We jointly optimize the HBF and DAC/ADC bit resolution matrices, unlike the existing approaches that optimize either DAC/ADC bit resolution or HBF matrices. The proposed design provides high flexibility, given that the analog precoder/combiner is codebook-free, thus there is no restriction on the angular vectors and different bit resolutions can be assigned to each DAC/ADC. We proceed with the literature review in the next subsection and then discuss the contributions of this chapter in detail in the following subsection.

5.1.1 Literature Review

As we know, to observe the effect of ADC resolution and bandwidth on rate, an AQNM is considered in [57] for a mmWave MIMO system under a RX power constraint. Reference [84] uses AQNM and shows the significance of low resolution ADCs on decreasing the rate. Recent work on A/D hybrid MIMO systems with low resolution sampling dynamically adjusts the ADC resolution [110]. Most of the literature such as in [25, 57, 84, 90, 110–112] imposes low resolution only at the RX side, and mostly assumed a fully digital or hybrid TX with high resolution DACs. However, there is a need to conduct research on optimizing the bit resolution problem for the TX side as well.

Furthermore, the existing literature mostly develops systems based on high resolution ADCs with a small number of RF chains or low resolution ADCs with a large number of RF chains. Either way, only fixed resolution DACs/ADCs are taken into account. References [44,65] consider EE optimization problems for A/D hybrid transceivers but with fixed and high resolution at the DACs/ADCs. The power model in [65]takes into account the power consumed at every RF chain and a constant power term for site-cooling, baseband processing and synchronization at the TX and [44] considers the RF hardware losses and some computational power expenditure.

Some approaches have been applied in A/D hybrid mmWave MIMO systems for EE maximization and low complexity with both full and low resolution sampling cases [24,26]. Reference [24] proposes an energy efficient A/D hybrid beamforming framework with a novel architecture for a mmWave MIMO system. The number of active RF chains are optimized dynamically by fractional programming to maximize EE performance but the DAC/ADC bit resolutions are fixed. Reference [26] proposes a novel EE maximization technique that selects the best subset of the active RF chains and DAC resolution which can also be extended to low resolution ADCs at the RX. Reference [111] suggests implementing fixed and low resolution ADCs with a small number of RF chains. Reference [90] works on the idea of a mixed-ADC architecture where a better energy-rate trade off is achieved by combining low and high resolution ADCs, but still with a fixed resolution for each ADC and without considering A/D hybrid beamforming. An A/D hybrid beamforming system with fixed and low resolution ADCs has been analyzed for channel estimation in [25].

One can implement varying resolution ADCs at the RX [112] which may provide a better solution than the RX with fixed and low resolution ADCs. Similarly, exploring low resolution DACs at the TX can also help reduce the power consumption. Thus, research that is focused on ADCs at the RX can also be applied to the TX DACs considering the TX specific system model parameters. Similar to using different ADC resolutions at the RX [112], which could provide a better solution than fixed low resolution ADCs, one can design a variable DAC resolution TX. Extra care is needed when deciding the number of bits used as the total DAC/ADC power consumption can be dominated by only a few high resolution DACs/ADCs. From [113], we notice that a good trade off between the power consumption and the performance may be to consider the range of 1-8 bits for I- and Q-channels, where 8-bit represents the full-bit resolution DACs/ADCs.

Reference [91] uses low resolution DACs for a single user MIMO system while [92] employs low resolution DACs at the base station for a narrowband multi-user MIMO system. Reference [114] also discusses fixed and low resolution DACs architecture for multi-user MIMO systems. Reference [115] considers a single user MIMO system with quantized hybrid precoding including the RF quantized noise term beside the AWGN while evaluating EE and SE performance. The existing literature still does not consider adjusting the resolution associated with DACs/ADCs dynamically. It is possible to consider both the TX and the RX simultaneously where we can design an optimization problem to find the optimal number of quantized bits to achieve high EE performance. When designing for high EE, the complexity of the solution also needs to be taken into account while providing improvements over the existing literature.

5.1.2 Contributions

This chapter designs an optimal EE solution for a mmWave A/D hybrid MIMO system by introducing a novel TX decomposition of the A/D hybrid precoder to three parts representing the analog precoder matrix, the DAC bit resolution matrix and the digital precoder matrix, respectively. A similar decomposition at the RX represents the analog combiner matrix, the ADC bit resolution matrix and the digital combiner matrix. Our aim is to minimize the distance between the decomposition, which is expressed as the product of three matrices, and the corresponding fully digital precoder or combiner matrix. The joint problem is decomposed into a series of sub-problems which are solved using the Alternating

Direction Method of Multipliers (ADMM). We implement an exhaustive search approach [65] to evaluate the upper bound for EE maximization.

In [27], we addressed bit allocation and hybrid combining at the RX only, where we jointly optimized the number of ADC bits and hybrid combiner matrices for EE maximization. A novel decomposition of the hybrid combiner to three parts was introduced: the analog combiner matrix, the bit resolution matrix and the baseband combiner matrix, and these matrices were computed using the ADMM approach in order to solve the matrix factorization problem. In addition to [27], the main contributions of this chapter can be listed as follows:

- This chapter designs an optimal EE solution for a mmWave A/D hybrid beamforming MIMO system by introducing the novel matrix decomposition that is applied to the hybrid beamforming matrices at both the TX and the RX. This decomposition defines three matrices, which are the analog beamforming matrix, the bit resolution matrix and the baseband beamforming matrix at both the TX and the RX. These matrices are obtained by the solution of an EE maximization problem and the DAC/ADC bit resolution is adjusted dynamically unlike fixed bit resolution in the existing literature.

- The joint TX-RX problem is a difficult problem to solve due to non-convex constraints and non-convex cost functions. Firstly we address the joint TX-RX problem unlike in the existing literature. Then we decouple it into two sub-problems dealing with the TX and the RX separately, where the corresponding problems at the TX and the RX are solved by the alternating minimization technique such as ADMM [116] to obtain the unknown precoder/combiner and DAC/ADC bit resolution matrices.

- This work jointly optimizes the hybrid beamforming and DAC/ADC bit resolution matrices, unlike the existing approaches that optimize either DAC/ADC bit resolution or hybrid beamforming matrices. Moreover, the proposed design has high flexibility, given that the analog precoder/combiner is codebook-free, thus there is no restriction on the angular vectors and different bit resolutions can be assigned to each DAC/ADC.

The performance of the proposed technique is investigated through extensive simulation results, achieving increased EE compared to the baseline techniques with fixed DAC/ADC bit resolutions and number of RF chains, and an exhaustive search based approach which is an upper bound for EE maximization. In the next section, we present the channel and system models where the channel model is based on a mmWave channel setup and the system model defines the low resolution quantization at both the TX and the RX.

5.2 MmWave HBF MIMO System with Low Resolution DACs and ADCs

We consider the same channel model as in (3.1) and similar mmWave MIMO HBF system model as shown in Fig. 4.1. In addition, Fig. 5.1 shows the block diagram of beam tracking phase and data communications phase in this context. Note that, unlike the previous chapter where we discuss the low resolution sampling at the TX for EE maximization and the RX for channel estimation, we consider the low resolution sampling both at the DACs and the ADCs simultaneously in this chapter. We follow the same definition of the channel model and system model parameters as in the previous chapters. We again use ULA antennas for simplicity and model the antenna elements at the RX as ideal sectored elements [35]. We assume that the CSI is known at both the TX and the RX. The matrices

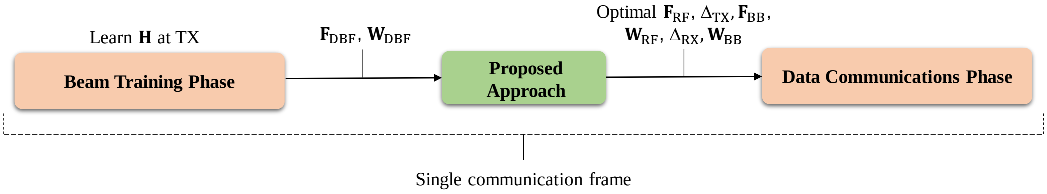

Learn **H** at TX
Beam Training Phase
$\mathbf{F}_{\mathrm{DBF}}$, $\mathbf{W}_{\mathrm{DBF}}$
Proposed Approach
Optimal $\mathbf{F}_{\mathrm{RF}}$, Δ_{TX}, $\mathbf{F}_{\mathrm{BB}}$, $\mathbf{W}_{\mathrm{RF}}$, Δ_{RX}, $\mathbf{W}_{\mathrm{BB}}$
Data Communications Phase
Single communication frame

Figure 5.1: Block diagram of the beam tracking phase and the data communications phase.

$\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}^{N_{\mathrm{T}} \times L_{\mathrm{T}}}$ and $\mathbf{W}_{\mathrm{RF}} \in \mathcal{W}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}$ where the set $\mathcal{F}$ and $\mathcal{W}$ represent the set of possible phase shifts in $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$, respectively. The sets $\mathcal{F}$ and $\mathcal{W}$ for variables f and w, respectively, are defined as $\mathcal{F} = \{f \in \mathbb{C} \mid |f| = 1\}$ and $\mathcal{W} = \{w \in \mathbb{C} \mid |w| = 1\}$.

Note that, we optimize the DAC and ADC resolution and the precoder and combiner matrices at the TX and the RX on a frame-by-frame basis. As shown in Fig. 5.1, we consider two stages in the system model: i) the beam training phase, and ii) the data communications phase. In stage i), firstly, the channel $\mathbf{H}$ is computed which provides us the optimal beamforming matrices, i.e., $\mathbf{F}_{\mathrm{DBF}}$ at the TX and $\mathbf{W}_{\mathrm{DBF}}$ at the RX. In stage ii), the optimal precoding and DAC bit resolution matrices $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{\Delta}_{\mathrm{TX}}$ at the TX, respectively, and the optimal combining and ADC bit resolution matrices $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}$ and $\mathbf{\Delta}_{\mathrm{RX}}$ at the RX are obtained. These two phases consist of one communication frame where the frame duration is smaller than the channel coherence time. Furthermore, if we assume that the TX/RX is active for stage i) a small proportion of time, for example, $< 10\%$, then the overall transmit energy consumption is dominated by stage ii).

Similar to the previous chapter, we consider the linear AQNM to represent the distortion of quantization [57]. Given that $Q(\cdot)$ denotes a uniform scalar quantizer then for the scalar complex input $x \in \mathbb{C}$ that is applied to both the real and imaginary parts, we have, $Q(x) \approx \delta x + \epsilon$, where $\delta = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b}} \in [m, M]$ is the multiplicative distortion parameter for a bit resolution equal to b [117], where m and M denote the minimum and maximum value of the range. The resolution parameter b is denoted as $b_i^t \, \forall \, i = 1, \ldots, L_\mathrm{T}$ and $b_i^r \, \forall \, i = 1, \ldots, L_\mathrm{R}$ at the TX and the RX, respectively. Note that the introduced error in the above linear approximation decreases for larger resolutions. However, our proposed solution focuses on EE maximization and this linear approximation does not impact the performance significantly as observed from the simulation results in Section 5.5. The parameter ϵ is the additive quantization noise with $\epsilon \sim \mathcal{CN}(0, \sigma_\epsilon^2)$, where $\sigma_\epsilon = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b}}\sqrt{\frac{\pi\sqrt{3}}{2}2^{-2b}}$. The matrices $\mathbf{\Delta}_\mathrm{TX}$ and $\mathbf{\Delta}_\mathrm{RX}$ represent diagonal matrices with values depending on the bit resolution of each DAC and ADC, respectively. Specifically, each diagonal entry of $\mathbf{\Delta}_\mathrm{TX}$ is given by:

$$[\mathbf{\Delta}_\mathrm{TX}]_{ii} = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b_i^t}} \in [m, M] \, \forall \, i = 1, \ldots, L_\mathrm{T}, \tag{5.1}$$

and each diagonal entry of $\mathbf{\Delta}_\mathrm{RX}$ is given by:

$$[\mathbf{\Delta}_\mathrm{RX}]_{ii} = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b_i^r}} \in [m, M] \, \forall \, i = 1, \ldots, L_\mathrm{R}, \tag{5.2}$$

where, for simplicity, we assume that the range $[m, M]$ is the same for each of the DACs/ADCs. The additive quantization noise for the DACs and ADCs are written as complex Gaussian vectors $\boldsymbol{\epsilon}_\mathrm{TX} \in \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\epsilon\mathrm{T}})$ and $\boldsymbol{\epsilon}_\mathrm{RX} \in \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\epsilon\mathrm{R}})$ [26] where $\mathbf{C}_{\epsilon\mathrm{T}}$ and $\mathbf{C}_{\epsilon\mathrm{R}}$ are the diagonal covariance matrices for DACs and ADCs, respectively. The covariance matrix entries are as follows:

$$[\mathbf{C}_{\epsilon\mathrm{T}}]_{ii} = \left(1 - \frac{\pi\sqrt{3}}{2}2^{-2b_i^t}\right)\left(\frac{\pi\sqrt{3}}{2}2^{-2b_i^t}\right) \forall i = 1, .., L_\mathrm{T}, \tag{5.3}$$

and

$$[\mathbf{C}_{\epsilon\mathrm{R}}]_{ii}=\left(1-\frac{\pi\sqrt{3}}{2}2^{-2b_i^r}\right)\left(\frac{\pi\sqrt{3}}{2}2^{-2b_i^r}\right)\forall i=1,..,L_{\mathrm{R}}. \tag{5.4}$$

Note that while optimizing the EE of the TX side, it is considered that the RX parameters, which includes the analog combiner matrix, the ADC bit resolution matrix and the baseband combiner matrix is known to the TX and vice-versa.

Let us consider $\mathbf{s} \in \mathbb{C}^{N_{\mathrm{s}}\times 1}$ as the normalized data vector, then based on the AQNM, the vector containing the complex output of all the DACs can be expressed as follows:

$$Q(\mathbf{F}_{\mathrm{BB}}\mathbf{s}) \approx \mathbf{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \boldsymbol{\epsilon}_{\mathrm{TX}} \in \mathbb{C}^{L_{\mathrm{T}}\times 1}, \tag{5.5}$$

This leads us to the following linear approximation for the transmitted signal $\mathbf{t} \in \mathbb{C}^{N_{\mathrm{T}}\times 1}$, as seen at the output of the A/D hybrid TX in Fig. 4.1:

$$\mathbf{t} = \mathbf{F}_{\mathrm{RF}}\mathbf{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{F}_{\mathrm{RF}}\boldsymbol{\epsilon}_{\mathrm{TX}}. \tag{5.6}$$

After the effect of the wireless mmWave channel $\mathbf{H}$ and the Gaussian noise $\mathbf{n}$ with independent and identically distributed entries and complex Gaussian distribution, i.e., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{n}}^2\mathbf{I}_{N_{\mathrm{R}}})$, the received signal $\mathbf{y} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ is expressed as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{t} + \mathbf{n} = \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{H}\mathbf{F}_{\mathrm{RF}}\boldsymbol{\epsilon}_{\mathrm{TX}} + \mathbf{n}. \tag{5.7}$$

When the analog combiner matrix $\mathbf{W}_{\mathrm{RF}}$ and ADC quantization based on AQNM are applied to the received signal $\mathbf{y}$, we obtain the following:

$$Q(\mathbf{W}_{\mathrm{RF}}^H\mathbf{y}) \approx \mathbf{\Delta}_{\mathrm{RX}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{y} + \boldsymbol{\epsilon}_{\mathrm{RX}} \in \mathbb{C}^{L_{\mathrm{R}}\times 1}. \tag{5.8}$$

After the application of the baseband combiner matrix $\mathbf{W}_{\mathrm{BB}}$, the output signal $\mathbf{r} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ at the RX, as shown in Fig. 4.1, can be expressed as follows:

$$\mathbf{r} = \mathbf{W}_{\mathrm{BB}}^{H} \boldsymbol{\Delta}_{\mathrm{RX}}^{H} \mathbf{W}_{\mathrm{RF}}^{H} \mathbf{y} + \mathbf{W}_{\mathrm{BB}}^{H} \boldsymbol{\epsilon}_{\mathrm{RX}}. \tag{5.9}$$

Considering the A/D hybrid precoder matrix $\mathbf{F} = \mathbf{F}_{\mathrm{RF}} \boldsymbol{\Delta}_{\mathrm{TX}} \mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{s}}}$ and the A/D hybrid combiner matrix $\mathbf{W} = \mathbf{W}_{\mathrm{RF}} \boldsymbol{\Delta}_{\mathrm{RX}} \mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{s}}}$, we can express the RX output signal $\mathbf{r}$ in (5.9) as follows:

$$\mathbf{r} = \mathbf{W}^{H} \mathbf{H} \mathbf{F} \mathbf{s} + \underbrace{\mathbf{W}^{H} \mathbf{H} \mathbf{F}_{\mathrm{RF}} \boldsymbol{\epsilon}_{\mathrm{TX}} + \mathbf{W}_{\mathrm{BB}}^{H} \boldsymbol{\epsilon}_{\mathrm{RX}} + \mathbf{W}^{H} \mathbf{n}}_{\boldsymbol{\eta}}, \tag{5.10}$$

where $\boldsymbol{\eta}$ is the combined effect of the additive white Gaussian RX noise and quantization noise that has covariance matrix, $\mathbf{R}_{\eta} \in \mathbb{C}^{N_{\mathrm{s}} \times N_{\mathrm{s}}}$, given by,

$$\mathbf{R}_{\eta} = \mathbf{W}^{H} \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{C}_{\epsilon \mathrm{T}} \mathbf{F}_{\mathrm{RF}}^{H} \mathbf{H}^{H} \mathbf{W} + \mathbf{W}_{\mathrm{BB}}^{H} \mathbf{C}_{\epsilon \mathrm{R}} \mathbf{W}_{\mathrm{BB}} + \sigma_{\mathrm{n}}^{2} \mathbf{W}^{H} \mathbf{W}. \tag{5.11}$$

In the following sections, we discuss the joint optimization solution to compute the optimal DAC/ADC bit resolution matrices and the optimal precoder/combiner matrices.

5.3 Joint DAC Bit Allocation and A/D Hybrid Precoding Optimization

Let us consider a point-to-point MIMO system with a linear quantization model. We define the EE as the ratio of the information rate R, i.e. SE, and the total consumed power P [109] as:

$$EE \triangleq \frac{R}{P} \text{ (bits/Hz/J).} \tag{5.12}$$

For the given point-to-point MIMO system, the SE is defined as,

$$R \triangleq \log_2 \left| \mathbf{I}_{N_\text{s}} + \frac{\mathbf{R}_\eta^{-1}}{N_\text{s}} \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H \mathbf{W} \right| \text{ (bits/s/Hz),} \tag{5.13}$$

where $\mathbf{F} = \mathbf{F}_\text{RF} \mathbf{\Delta}_\text{TX} \mathbf{F}_\text{BB}$ and $\mathbf{W} = \mathbf{W}_\text{RF} \mathbf{\Delta}_\text{RX} \mathbf{W}_\text{BB}$.

Similar to the power model at the TX in [26] and following Section 2.2.4 for the case of low resolution quantization and the power consumption at both the TX and the RX, the total consumed power for the system is expressed as

$$P \triangleq P_\text{TX}(\mathbf{F}_\text{RF}, \mathbf{\Delta}_\text{TX}, \mathbf{F}_\text{BB}) + P_\text{RX}(\mathbf{\Delta}_\text{RX}) \text{ (W),} \tag{5.14}$$

where the power consumption at the TX is as follows:

$$P_\text{TX}(\mathbf{F}_\text{RF}, \mathbf{\Delta}_\text{TX}, \mathbf{F}_\text{BB}) = \text{tr}(\mathbf{F}\mathbf{F}^H) + P_\text{DT}(\mathbf{\Delta}_\text{TX}) + N_\text{T} P_\text{T} + N_\text{T} L_\text{T} P_\text{PT} + P_\text{CT} \text{ (W),} \tag{5.15}$$

where P_PT is the power per phase shifter, P_T is the power per antenna element, $P_\text{DT}(\mathbf{\Delta}_\text{TX})$ is the power associated with the total quantization operation at the TX, and following (5.1) and [57], we have

$$P_\text{DT}(\mathbf{\Delta}_\text{TX}) = P_\text{DAC} \sum_{i=1}^{L_\text{T}} 2^{b_i^t} = P_\text{DAC} \sum_{i=1}^{L_\text{T}} \left(\frac{\pi\sqrt{3}}{2(1 - [\mathbf{\Delta}_\text{TX}]_{ii}^2)} \right)^{\frac{1}{2}} \text{ (W),} \tag{5.16}$$

where P_DAC is the power consumed per bit in the DAC and P_CT is the power required by all circuit components at the TX. Similarly, the total power consumption at the RX is,

$$P_{\mathrm{RX}}(\mathbf{\Delta}_{\mathrm{RX}})=P_{\mathrm{DR}}(\mathbf{\Delta}_{\mathrm{RX}})+N_{\mathrm{R}}P_{\mathrm{R}}+N_{\mathrm{R}}L_{\mathrm{R}}P_{\mathrm{PR}}+P_{\mathrm{CR}}\ (\mathrm{W}), \tag{5.17}$$

where, at the RX, P_{PR} is the power per phase shifter, P_{R} is the power per antenna element, P_{DR} is the power associated with the total quantization operation, and following (5.2) and [57], we have

$$P_{\mathrm{DR}}(\mathbf{\Delta}_{\mathrm{RX}})=P_{\mathrm{ADC}}\sum_{i=1}^{L_{\mathrm{R}}}2^{b_i^r}=P_{\mathrm{ADC}}\sum_{i=1}^{L_{\mathrm{R}}}\left(\frac{\pi\sqrt{3}}{2(1-[\mathbf{\Delta}_{\mathrm{RX}}]_{ii}^2)}\right)^{\frac{1}{2}}\ (\mathrm{W}), \tag{5.18}$$

where P_{ADC} is the power consumed per bit in the ADC and P_{CR} is the power required by all RX circuit components.

The maximization of EE is given by

$$\max_{\mathbf{F}_{\mathrm{RF}},\mathbf{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\mathbf{\Delta}_{\mathrm{RX}},\mathbf{W}_{\mathrm{BB}}} \frac{R(\mathbf{F}_{\mathrm{RF}},\mathbf{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\mathbf{\Delta}_{\mathrm{RX}},\mathbf{W}_{\mathrm{BB}})}{P_{\mathrm{TX}}(\mathbf{F}_{\mathrm{RF}},\mathbf{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}})+P_{\mathrm{RX}}(\mathbf{\Delta}_{\mathrm{RX}})}$$
$$\text{subject to } \mathbf{F}_{\mathrm{RF}}\in\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}},\mathbf{\Delta}_{\mathrm{TX}}\in\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}}\times L_{\mathrm{T}}},\mathbf{W}_{\mathrm{RF}}\in\mathcal{W}^{N_{\mathrm{R}}\times L_{\mathrm{R}}},\mathbf{\Delta}_{\mathrm{RX}}\in\mathcal{D}_{\mathrm{RX}}^{L_{\mathrm{R}}\times L_{\mathrm{R}}}, \tag{5.19}$$

when the SE R is given by (5.13) and the power P in (5.14). The problem to be addressed involves a fractional cost function that both the numerator and the denominator parts are non-convex functions of the optimizing variables. Furthermore the optimization problem involves non-convex constraint sets. Thus, it is in general a very difficult problem to be addressed. It is interesting that the corresponding problem for a fully digital transceiver that admits a much simpler form is in general intractable due to the coupling of the TX-RX design [118]. To that end, we start by decoupling the TX-RX design problem.

Let us first express the EE maximization problem in the following relaxed form:

$$\begin{aligned}\min_{\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{RX}},\mathbf{W}_{\mathrm{BB}}} \quad & -R(\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{RX}},\mathbf{W}_{\mathrm{BB}}) \\ & +\gamma_{\mathrm{T}}P_{\mathrm{TX}}(\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}})+\gamma_{\mathrm{R}}P_{\mathrm{RX}}(\boldsymbol{\Delta}_{\mathrm{RX}}) \\ \text{subject to } & \mathbf{F}_{\mathrm{RF}}\in\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}},\boldsymbol{\Delta}_{\mathrm{TX}}\in\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}}\times L_{\mathrm{T}}}, \\ & \mathbf{W}_{\mathrm{RF}}\in\mathcal{W}^{N_{\mathrm{R}}\times L_{\mathrm{R}}},\boldsymbol{\Delta}_{\mathrm{RX}}\in\mathcal{D}_{\mathrm{RX}}^{L_{\mathrm{R}}\times L_{\mathrm{R}}},\end{aligned} \tag{5.20}$$

where the parameters $\gamma_{\mathrm{T}}\in(0,\gamma_{\mathrm{T}}^{max}]\subset\mathbb{R}^+$ and $\gamma_{\mathrm{R}}\in(0,\gamma_{\mathrm{R}}^{max}]\subset\mathbb{R}^+$ are introducing a trade-off between the achieved rate and the power consumption at the TX's and the RX's side, respectively. Such an approach has been used in the past to tackle fractional optimization problems [70]. In the concave/convex case, the equivalence of the relaxed problem with the original fractional one is theoretically established. Unfortunately, a similar result for the case considered in the present work is not easy to be derived due to the complexity of the addressed problem. Thus, in the present work, we rely on line search methods in order to optimally tune these parameters.

Having simplified the original problem, we may now proceed by temporally decoupling the designs at the TX's and the RX's side. Under the assumption that the RX can perform optimal nearest-neighbor decoding based on the received signals, the optimal precoding matrices are designed such that the mutual information achieved by Gaussian signaling over the wireless channel is maximized [16]. The mutual information is given by

$$I\triangleq\log_2\left|\mathbf{I}_{N_{\mathrm{s}}}+\frac{\mathbf{Q}_{\eta'}^{-1}}{N_{\mathrm{s}}}\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H\right| \text{ (bits/s/Hz)}, \tag{5.21}$$

where again $\mathbf{F}=\mathbf{F}_{\mathrm{RF}}\boldsymbol{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}$ and and $\mathbf{Q}_{\eta'}$ is the covariance matrix of the sum of noise and transmit quantization noise variables, i.e. $\eta' = \mathbf{F}_{\mathrm{RF}}\boldsymbol{\epsilon}_{\mathrm{TX}} + \mathbf{n}$, given by

$$\mathbf{Q}_{\eta'} = \mathbf{F}_{\mathrm{RF}}\mathbf{C}_{\epsilon\mathrm{T}}\mathbf{F}_{\mathrm{RF}}^{H} + \sigma_{\mathrm{n}}^{2}\mathbf{I}_{N_{\mathrm{R}}}. \tag{5.22}$$

Based on (5.20)-(5.21), the precoding matrices may be derived as the solution to the following optimization problem:

$$\begin{aligned}(\mathcal{P}_{1\mathrm{T}}) : \min_{\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}}} & -I(\mathbf{F}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}}) + \gamma_T P_{\mathrm{TX}}(\mathbf{F}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}}),\\ & \text{subject to } \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}, \boldsymbol{\Delta}_{\mathrm{TX}} \in \mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}}\times L_{\mathrm{T}}},\end{aligned}$$

Now provided that the optimal precoding matrix $\mathbf{F}^{\star} = \mathbf{F}_{\mathrm{RF}}^{\star}\boldsymbol{\Delta}_{\mathrm{TX}}^{\star}\mathbf{F}_{\mathrm{BB}}^{\star}$ is derived from solving $(\mathcal{P}_{1\mathrm{T}})$, we can plug in these resulted precoding matrices in the cost function of (5.20) resulting in an optimization problem dependent only on the decoder matrices at the RX's side, defined as,

$$\begin{aligned}(\mathcal{P}_{1\mathrm{R}}) : \min_{\mathbf{W}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{RX}},\mathbf{W}_{\mathrm{BB}}} & - \tilde{R}(\mathbf{W}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{RX}}, \mathbf{W}_{\mathrm{BB}}) + \gamma_R P_{RX}(\boldsymbol{\Delta}_{\mathrm{RX}})\\ & \text{subject to } \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}^{N_{\mathrm{R}}\times L_{\mathrm{R}}}, \boldsymbol{\Delta}_{\mathrm{RX}} \in \mathcal{D}_{\mathrm{RX}}^{L_{\mathrm{R}}\times L_{\mathrm{R}}},\end{aligned} \tag{5.23}$$

where $\tilde{R}(\mathbf{W}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{RX}}, \mathbf{W}_{\mathrm{BB}}) = R(\mathbf{F}_{\mathrm{RF}}^{\star}, \boldsymbol{\Delta}_{\mathrm{TX}}^{\star}, \mathbf{F}_{\mathrm{BB}}^{\star}, \mathbf{W}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{RX}}, \mathbf{W}_{\mathrm{BB}})$.

Thus, the precoding and decoding matrices can be derived as the solutions to the two decoupled problems $(\mathcal{P}_{1\mathrm{T}})-(\mathcal{P}_{1\mathrm{R}})$ above. In the following subsections, the solutions to these problems are developed. We start first with the development of the solution to TX's side one $(\mathcal{P}_{1\mathrm{T}})$ and then the solution for the RX's side $(\mathcal{P}_{1\mathrm{R}})$ counterpart follows.

5.3.1 Problem Formulation at the TX

Focusing on the TX side, we seek the bit resolution matrix $\mathbf{\Delta}_{\mathrm{TX}}$ and the hybrid precoding matrices $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$ that solve $(\mathcal{P}_{1\mathrm{T}})$. The set $\mathcal{D}_{\mathrm{TX}}$ represents the finite states of the quantizer and is defined as,

$$\mathcal{D}_{\mathrm{TX}} = \left\{\mathbf{\Delta}_{\mathrm{TX}} \in \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}} \middle| m \leq [\mathbf{\Delta}_{\mathrm{TX}}]_{ii} \leq M \;\forall\, i = 1, ..., L_{\mathrm{T}}\right\}.$$

Note that $P_{\mathrm{TX}}(\mathbf{F}_{\mathrm{RF}}, \mathbf{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}}) > 0$, as defined in (5.15), since the power required by all circuit components is always larger than zero, i.e., $P_{\mathrm{CP}} > 0$.

Since dealing with the part of the cost function of $(\mathcal{P}_{1\mathrm{T}})$ that involves the mutual information expression is a difficult task due to the perplexed form of the latter, we adopt the approach in [16] where the maximization of the mutual information I can be approximated by finding the minimum Euclidean distance of the hybrid precoder to the one of the fully digital transceiver for the full-bit resolution sampling case, denoted by $\mathbf{F}_{\mathrm{DBF}}$, i.e., $\|\mathbf{F}_{\mathrm{DBF}} - \mathbf{F}_{\mathrm{RF}}\mathbf{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\|_F^2$ [16]. Therefore, motivated by the previous, $(\mathcal{P}_{1\mathrm{T}})$ can be approximated to finding the solution of the following problem:

$$(\mathcal{P}_2): \min_{\mathbf{F}_{\mathrm{RF}}, \mathbf{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}}} \frac{1}{2}\|\mathbf{F}_{\mathrm{DBF}} - \mathbf{F}_{\mathrm{RF}}\mathbf{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\|_F^2 + \gamma_{\mathrm{T}} P_{\mathrm{TX}}(\mathbf{F}),$$
$$\text{subject to } \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}^{N_{\mathrm{T}} \times L_{\mathrm{T}}}, \mathbf{\Delta}_{\mathrm{TX}} \in \mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}.$$

For a point-to-point MIMO system the optimal $\mathbf{F}_{\mathrm{DBF}}$ is given by $\mathbf{F}_{\mathrm{DBF}} = \mathbf{V}\sqrt{\mathbf{P}}$ where the orthonormal matrix $\mathbf{V} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ is derived via the channel matrix singular value decomposition (SVD), i.e. $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ and $\mathbf{P}$ is a diagonal power allocation matrix with real positive diagonal entries derived by the so-called "water-filling algorithm" [32].

Problem $(\mathcal{P}_2)$ is still very difficult to address as it is non-convex due to the non-convex cost function that involves the product of three matrix variables and non-convex constraints. In the next section, an efficient algorithmic solution based on the ADMM is proposed.

5.3.2 Proposed ADMM Solution at the TX

In the following we develop an iterative procedure for solving $(\mathcal{P}_2)$ based on the ADMM approach [116]. This method is a variant of the standard augmented Lagrangian method that uses partial updates (similar to the Gauss-Seidel method for the solution of linear equations) to solve constrained optimization problems. While it is mainly known for its good performance for a number of convex optimization problems, recently it has been successfully applied to non-convex matrix factorization as well [116, 119, 120]. Motivated by this, in the following ADMM based solutions are developed that are tailored for the non-convex matrix factorization problem $(\mathcal{P}_2)$.

We first transform $(\mathcal{P}_2)$ into a form that can be addressed via ADMM. By using the auxiliary variable $\mathbf{Z}$, $(\mathcal{P}_2)$ can be written as:

$$(\mathcal{P}_3) : \min_{\mathbf{Z},\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}}} \frac{1}{2}\|\mathbf{F}_{\mathrm{DBF}}-\mathbf{Z}\|_F^2+\mathbb{1}_{\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\mathbf{F}_{\mathrm{RF}}\}+\mathbb{1}_{\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\boldsymbol{\Delta}_{\mathrm{TX}}\}+\gamma_{\mathrm{T}}P_{\mathrm{TX}}(\mathbf{F}),$$
$$\text{subject to } \mathbf{Z} = \mathbf{F}_{\mathrm{RF}}\boldsymbol{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}.$$

Problem $(\mathcal{P}_3)$ formulates the A/D hybrid precoder matrix design as a matrix factorization problem. That is, the overall precoder $\mathbf{Z}$ is sought so that it minimizes the Euclidean distance to the optimal, fully digital precoder $\mathbf{F}_{\mathrm{DBF}}$ while supporting decomposition into three factors: the analog precoder matrix $\mathbf{F}_{\mathrm{RF}}$, the DAC bit resolution matrix $\boldsymbol{\Delta}_{\mathrm{TX}}$ and the digital precoder matrix $\mathbf{F}_{\mathrm{BB}}$.

The augmented Lagrangian function of $(\mathcal{P}_3)$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{Z}, \mathbf{F}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Lambda}) =& \frac{1}{2}\|\mathbf{F}_{\mathrm{DBF}}-\mathbf{Z}\|_F^2+\mathbb{1}_{\mathcal{F}^{N_\mathrm{T}\times L_\mathrm{T}}}\{\mathbf{F}_{\mathrm{RF}}\}+\mathbb{1}_{\mathcal{D}_{\mathrm{TX}}^{L_\mathrm{T}\times L_\mathrm{T}}}\{\boldsymbol{\Delta}_{\mathrm{TX}}\}\\ &+\frac{\alpha}{2}\|\mathbf{Z}+\boldsymbol{\Lambda}/\alpha-\mathbf{F}_{\mathrm{RF}}\boldsymbol{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\|_F^2+\gamma_\mathrm{T}P_{\mathrm{TX}}(\mathbf{F}),\end{aligned} \quad (5.24)$$

where α is a scalar penalty parameter and $\boldsymbol{\Lambda} \in \mathbb{C}^{N_\mathrm{T}\times L_\mathrm{T}}$ is the Lagrange Multiplier matrix. According to the ADMM approach [116], the solution to $(\mathcal{P}_3)$ is derived by the following iterative steps where n denotes the iteration index:

$$\begin{aligned}&(\mathcal{P}_{3\mathrm{A}}): \mathbf{Z}_{(n)} = \arg\min_{\mathbf{Z}} \mathcal{L}(\mathbf{Z}, \mathbf{F}_{\mathrm{RF}(n-1)}, \boldsymbol{\Delta}_{\mathrm{TX}(n-1)}, \mathbf{F}_{\mathrm{BB}(n-1)}, \boldsymbol{\Lambda}_{(n-1)}),\\ &(\mathcal{P}_{3\mathrm{B}}): \mathbf{F}_{\mathrm{RF}(n)} = \arg\min_{\mathbf{F}_{\mathrm{RF}}} \mathcal{L}(\mathbf{Z}_{(n)}, \mathbf{F}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{TX}(n-1)}, \mathbf{F}_{\mathrm{BB}(n-1)}, \boldsymbol{\Lambda}_{(n-1)}),\\ &(\mathcal{P}_{3\mathrm{C}}): \boldsymbol{\Delta}_{\mathrm{TX}(n)} = \arg\min_{\boldsymbol{\Delta}_{\mathrm{TX}}} \mathcal{L}(\mathbf{Z}_{(n)}, \mathbf{F}_{\mathrm{RF}(n)}, \boldsymbol{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}(n-1)}, \boldsymbol{\Lambda}_{(n-1)})+\gamma_\mathrm{T}P_{\mathrm{TX}}(\mathbf{F}),\\ &(\mathcal{P}_{3\mathrm{D}}): \mathbf{F}_{\mathrm{BB}(n)} = \arg\min_{\mathbf{F}_{\mathrm{BB}}} \mathcal{L}(\mathbf{Z}_{n}, \mathbf{F}_{\mathrm{RF}(n)}, \boldsymbol{\Delta}_{\mathrm{TX}(n)}, \mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Lambda}_{(n-1)}),\\ &\boldsymbol{\Lambda}_{(n)} = \boldsymbol{\Lambda}_{(n-1)} + \alpha\left(\mathbf{Z}_{(n)} - \mathbf{F}_{\mathrm{RF}(n)}\boldsymbol{\Delta}_{\mathrm{TX}(n)}\mathbf{F}_{\mathrm{BB}(n)}\right).\end{aligned} \quad (5.25)$$

In order to apply the ADMM iterative procedure, we have to solve the optimization problems $(\mathcal{P}_{3\mathrm{A}})$-$(\mathcal{P}_{3\mathrm{D}})$. We may start from problem $(\mathcal{P}_{3\mathrm{A}})$ which can be written as follows:

$$(\mathcal{P}'_{3\mathrm{A}}): \mathbf{Z}_{(n)}=\arg\min_{\mathbf{Z}} \frac{1}{2}\|(1+\alpha)\mathbf{Z}-\mathbf{F}_{\mathrm{DBF}}+\boldsymbol{\Lambda}_{(n-1)}-\alpha\mathbf{F}_{\mathrm{RF}(n-1)}\boldsymbol{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\|_F^2.$$

Problem $(\mathcal{P}'_{3\mathrm{A}})$ can be directly solved by equating the gradient of the augmented Lagrangian (5.24) w.r.t. $\mathbf{Z}$ being set to zero. Therefore, we have

$$\mathbf{Z}_{(n)}=\frac{1}{\alpha+1}\Big(\mathbf{F}_{\mathrm{DBF}}-\boldsymbol{\Lambda}_{(n-1)}+\alpha\mathbf{F}_{\mathrm{RF}(n-1)}\boldsymbol{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\Big). \quad (5.26)$$

We may now proceed to solve $(\mathcal{P}_{3\mathrm{B}})$ which can be written in the following simplified form by keeping only the terms of the augmented Lagrangian that are dependent on $\mathbf{F}_{\mathrm{RF}}$:

$$(\mathcal{P}'_{3\mathrm{B}}): \ \mathbf{F}_{\mathrm{RF}(n)} = \arg\min_{\mathbf{F}_{\mathrm{RF}}} \mathbb{1}_{\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\mathbf{F}_{\mathrm{RF}}\} + \frac{\alpha}{2}\|\mathbf{Z}_{(n)} + \mathbf{\Lambda}_{(n-1)}/\alpha \\ - \mathbf{F}_{\mathrm{RF}}\mathbf{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\|_F^2.$$

The solution to problem $(\mathcal{P}'_{3B})$ does not admit a closed form and thus, it is approximated by solving the unconstrained problem and then projecting onto the set $\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}$, i.e.,

$$\mathbf{F}_{\mathrm{RF}(n)} = \Pi_{\mathcal{F}}\Big\{ \left(\mathbf{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\right) \mathbf{F}^H_{\mathrm{BB}(n-1)}\mathbf{\Delta}^H_{\mathrm{TX}(n-1)} \\ \left(\alpha\mathbf{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\mathbf{F}^H_{\mathrm{BB}(n-1)}\mathbf{\Delta}^H_{\mathrm{TX}(n-1)}\right)^{-1}\Big\}, \tag{5.27}$$

where $\Pi_{\mathcal{F}}$ projects the solution onto the set $\mathcal{F}$. This is computed by solving the following optimization problem [121]:

$$(\mathcal{P}''_{3\mathrm{B}}): \qquad \min_{\mathbf{A}_{\mathcal{F}}} \|\mathbf{A}_{\mathcal{F}} - \mathbf{A}\|_F^2, \text{subject to } \mathbf{A}_{\mathcal{F}} \in \mathcal{F},$$

where $\mathbf{A}$ is an arbitrary matrix and $\mathbf{A}_{\mathcal{F}}$ is its projection onto the set $\mathcal{F}$. The solution to $(\mathcal{P}''_{3\mathrm{B}})$ is given by the phase of the complex elements of $\mathbf{A}$. Thus, for $\mathbf{A}_{\mathcal{F}} = \Pi_{\mathcal{F}}\{\mathbf{A}\}$ we have

$$\mathbf{A}_{\mathcal{F}}(x,y) = \begin{cases} 0, & \mathbf{A}(x,y) = 0 \\ \frac{\mathbf{A}(x,y)}{|\mathbf{A}(x,y)|}, & \mathbf{A}(x,y) \neq 0 \end{cases}, \tag{5.28}$$

where $\mathbf{A}_{\mathcal{F}}(x,y)$ and $\mathbf{A}(x,y)$ are the elements at the xth row-yth column of matrices

Algorithm 8 Proposed ADMM Solution for the A/D Hybrid Precoder Design

```
Initialize: $\mathbf{Z}$, $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{\Delta}_{\mathrm{TX}}$, $\mathbf{F}_{\mathrm{BB}}$ with random values, $\mathbf{\Lambda}$ with zeros, $\alpha = 1$ and $n = 1$
while The termination criteria of (5.30) are not met or $n \leq N_{\max}$ do
    Update $\mathbf{Z}_{(n)}$ using solution (5.26),
        $\mathbf{F}_{\mathrm{RF}(n)}$ using solution (5.27),
        $\mathbf{\Delta}_{\mathrm{TX}(n)}$ by solving $(\mathcal{P}_{3\mathrm{C}}^{''})$ using CVX [72],
        $\mathbf{F}_{\mathrm{BB}(n)}$ using solution (5.29), and
        update $\mathbf{\Lambda}_{(n)}$ using solution (5.25).
    $n \leftarrow n + 1$
end while
return $\mathbf{F}_{\mathrm{RF}}^{\star}$, $\mathbf{\Delta}_{\mathrm{TX}}^{\star}$, $\mathbf{F}_{\mathrm{BB}}^{\star}$
```

$\mathbf{A}_{\mathcal{F}}$ and $\mathbf{A}$, respectively. While, this is an approximate solution, it turns out that it behaves remarkably well, as verified in the simulation results of Section 5.5. This is due to the interesting property that ADMM is observed to converge even in cases where the alternating minimization steps are not carried out exactly [116]. There are theoretical results that support this statement [122, 123], though an exact analysis for the case considered here is beyond the scope of this chapter.

In a similar manner, $(\mathcal{P}_{3\mathrm{C}})$ may be re-written as,

$$(\mathcal{P}_{3\mathrm{C}}') : \mathbf{\Delta}_{\mathrm{TX}(n)} = \arg\min_{\mathbf{\Delta}_{\mathrm{TX}}} \mathbb{1}_{\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}}\{\mathbf{\Delta}_{\mathrm{TX}}\} + \frac{\alpha}{2}\|\mathbf{Z}_{(n)} + \mathbf{\Lambda}_{(n-1)}/\alpha - \mathbf{F}_{\mathrm{RF}(n)}\mathbf{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}(n-1)}\|_F^2 + \gamma_{\mathrm{T}} P_{\mathrm{TX}}(\mathbf{F}).$$

To solve the above problem, we can write:

$$(\mathcal{P}_{3\mathrm{C}}^{''}) : \mathbf{\Delta}_{\mathrm{TX}(n)} = \arg\min_{\mathbf{\Delta}_{\mathrm{TX}}} \|\mathbf{y}_{\mathrm{c}} - \mathbf{\Psi}_{\mathrm{T}} \mathrm{vec}(\mathbf{\Delta}_{\mathrm{TX}})\|_2^2 + \gamma_{\mathrm{T}} P_{\mathrm{TX}}(\mathbf{F}),$$
$$\text{subject to } \mathbf{\Delta}_{\mathrm{TX}} \in \mathcal{D}_{\mathrm{TX}},$$

The minimization problem in $(\mathcal{P}_{3\mathrm{C}}^{''})$ consists of $\mathbf{y}_c = \mathrm{vec}(\mathbf{Z}_n + \Lambda_{n-1}/\alpha)$, $\mathbf{\Psi}_{\mathrm{T}} = \mathbf{F}_{\mathrm{BB}(n-1)} \otimes \mathbf{F}_{\mathrm{RF}(n)}$ ($\otimes$ being the Khatri-Rao product) and is solved using CVX [72].

The solution of problem $(\mathcal{P}_{3\text{D}})$ may be written in the following form:

$$(\mathcal{P}'_{3\text{D}}): \ \mathbf{F}_{\text{BB}(n)} = \arg\min_{\mathbf{F}_{\text{BB}}} \frac{\alpha}{2} \|\mathbf{Z}_{(n)} + \mathbf{\Lambda}_{(n-1)}/\alpha - \mathbf{F}_{\text{RF}(n)}\mathbf{\Delta}_{\text{TX}(n)}\mathbf{F}_{\text{BB}}\|_F^2.$$

It is straightforward to see that the solution for $(\mathcal{P}'_{3\text{D}})$ can be obtained by equating the gradient to zero and solving the resulting equation w.r.t. the matrix variable $\mathbf{F}_{\text{BB}}$, i.e.,

$$\mathbf{F}_{\text{BB}(n)} = \left(\alpha \mathbf{\Delta}^H_{\text{TX}(n)}\mathbf{F}^H_{\text{RF}(n)}\mathbf{F}_{\text{RF}(n)}\mathbf{\Delta}_{\text{TX}(n)}\right)^{-1} \mathbf{\Delta}^H_{\text{TX}(n)}\mathbf{F}^H_{\text{RF}(n)} \left(\mathbf{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\right). \quad (5.29)$$

Algorithm 8 provides the complete procedure to obtain the optimal analog precoder matrix $\mathbf{F}_{\text{RF}}$, the optimal bit resolution matrix $\mathbf{\Delta}_{\text{TX}}$ and the optimal baseband (or digital) precoder matrix $\mathbf{F}_{\text{BB}}$. It starts the alternating minimization procedure by initializing the entries of the matrices $\mathbf{Z}$, $\mathbf{F}_{\text{RF}}$, $\mathbf{\Delta}_{\text{TX}}$, $\mathbf{F}_{\text{BB}}$ with random values and the entries of the Lagrange multiplier matrix $\mathbf{\Lambda}$ with zeros. For iteration index n, $\mathbf{Z}_{(n)}$, $\mathbf{F}_{\text{RF}(n)}$, $\mathbf{\Delta}_{\text{TX}(n)}$ and $\mathbf{F}_{\text{BB}(n)}$ are updated using Step 3 which shows the steps to be used to obtain the matrices. A termination criterion related to either the maximum permitted number of iterations ($N_{\max}$) is considered or the ADMM solution meeting the following criteria is considered:

$$\|\mathbf{Z}_{(n)} - \mathbf{Z}_{(n-1)}\|_F \leq \epsilon^z \ \& \ \|\mathbf{Z}_{(n)} - \mathbf{F}_{\text{RF}(n)}\mathbf{\Delta}_{\text{TX}(n)}\mathbf{F}_{\text{BB}(n)}\|_F \leq \epsilon^p, \quad (5.30)$$

where ϵ^z and ϵ^p are the corresponding tolerances. Upon convergence, the number of bits for each DAC is obtained by using (5.1) and quantizing to the nearest integer value. The optimal hybrid precoding matrices $\mathbf{F}^\star_{\text{RF}}$, $\mathbf{\Delta}^\star_{\text{TX}}$, $\mathbf{F}^\star_{\text{BB}}$ are obtained at the end of this algorithm.

Computational Complexity Analysis of Algorithm 8

When running Algorithm 8, mainly Step 3, while updating $\mathbf{\Delta}_{\mathrm{TX}(n)}$ by solving $(\mathcal{P}''_{3\mathrm{C}})$ using CVX, involves multiplication by $\mathbf{\Psi}_{\mathrm{T}}$ whose dimensions are $L_{\mathrm{T}}N_{\mathrm{T}} \times N_{\mathrm{s}}L_{\mathrm{T}}$. In general, the solution of $(\mathcal{P}''_{3\mathrm{C}})$ can be upper-bounded by $\mathcal{O}((L_{\mathrm{T}}^2 N_{\mathrm{T}} N_{\mathrm{s}})^3)$ which can be improved significantly by exploiting the structure of $\mathbf{\Psi}_{\mathrm{T}}$.

In the following section, we discuss the joint optimization problem at the RX and the solution to obtain the analog combiner matrix $\mathbf{W}_{\mathrm{RF}}$, the ADC bit resolution matrix $\mathbf{\Delta}_{\mathrm{RX}}$ and the digital combiner matrix $\mathbf{W}_{\mathrm{BB}}$.

5.4 Joint ADC Bit Allocation and A/D Hybrid Combining Optimization

5.4.1 Problem Formulation at the RX

Let us now move to the derivation of the solution to $(\mathcal{P}_{1\mathrm{R}})$. The set $\mathcal{D}_{\mathrm{RX}}$ represents the finite states of the ADC quantizer and is defined as,

$$\mathcal{D}_{\mathrm{RX}} = \left\{ \mathbf{\Delta}_{\mathrm{RX}} \in \mathbb{R}^{L_{\mathrm{R}} \times L_{\mathrm{R}}} \middle| m \leq [\mathbf{\Delta}_{\mathrm{RX}}]_{ii} \leq M \; \forall \, i = 1, ..., L_{\mathrm{R}} \right\}.$$

Due to the perplexed form of the function $\tilde{R}(\mathbf{W}_{\mathrm{RF}}, \mathbf{\Delta}_{\mathrm{RX}}, \mathbf{W}_{\mathrm{BB}})$, we follow the same arguments the under of which we approximated $(\mathcal{P}_2)$ by $(\mathcal{P}_{1\mathrm{T}})$, in order to approximate $(\mathcal{P}_{1\mathrm{R}})$ by

$$(\mathcal{P}_5) : \min_{\mathbf{W}_{\mathrm{RF}}, \mathbf{\Delta}_{\mathrm{RX}}, \mathbf{W}_{\mathrm{BB}}} \frac{1}{2} \| \mathbf{W}_{\mathrm{DBF}} - \mathbf{W}_{\mathrm{RF}} \mathbf{\Delta}_{\mathrm{RX}} \mathbf{W}_{\mathrm{BB}} \|_F^2 + \gamma_{\mathrm{R}} P_{\mathrm{RX}}(\mathbf{\Delta}_{\mathrm{RX}}),$$
$$\text{subject to } \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}, \mathbf{\Delta}_{\mathrm{RX}} \in \mathcal{D}_{\mathrm{RX}}^{L_{\mathrm{R}} \times L_{\mathrm{R}}},$$

where $\mathbf{W}_{\text{DBF}}$ is the optimal solution for the fully digital RX which is given by $\mathbf{W}_{\text{DBF}} = \sqrt{\tilde{\mathbf{P}}}\tilde{\mathbf{U}}$, where $\tilde{\mathbf{U}} \in \mathbb{C}^{N_{\text{R}}\times N_{\text{s}}}$ is the orthonormal singular vector matrix which can be derived by the SVD of the equivalent channel matrix $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{F}^{\star} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^{H}$, and $\tilde{\mathbf{P}}$ is diagonal power allocation matrix. Problem $(\mathcal{P}_5)$ is also non-convex due to the non-convex cost function and non-convex set of constraints, as well, and for its solution an ADMM-based solution similar to the case of $(\mathcal{P}_2)$ is derived in the following subsection.

5.4.2 Proposed ADMM Solution at the RX

In the following we develop an iterative procedure for solving $(\mathcal{P}_5)$ based on ADMM [116]. We first transform $(\mathcal{P}_5)$ into an amenable form. By using the auxiliary variable $\mathbf{Z}$, $(\mathcal{P}_5)$ can be written as:

$$\begin{aligned}(\mathcal{P}_6):\ \min_{\mathbf{Z},\mathbf{W}_{\text{RF}},\mathbf{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}} &\frac{1}{2}\|\mathbf{W}_{\text{DBF}}-\mathbf{Z}\|_F^2 + \mathbb{1}_{\mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}}\{\mathbf{W}_{\text{RF}}\} + \mathbb{1}_{\mathcal{D}_{\text{RX}}^{L_{\text{R}}\times L_{\text{R}}}}\{\mathbf{\Delta}_{\text{RX}}\} \\ &+ \gamma_{\text{R}} P_{\text{RX}}(\mathbf{\Delta}_{\text{RX}}), \\ \text{subject to } &\mathbf{Z} = \mathbf{W}_{\text{RF}}\mathbf{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}}.\end{aligned}$$

Problem $(\mathcal{P}_6)$ formulates the A/D hybrid combiner matrix design as a matrix factorization problem. That is, the overall combiner $\mathbf{Z}$ is sought so that it minimizes the Euclidean distance to the optimal, fully digital combiner $\mathbf{W}_{\text{DBF}}$ while supporting the decomposition into the analog combiner matrix $\mathbf{W}_{\text{RF}}$, the quantization error matrix $\mathbf{\Delta}_{\text{RX}}$ and the digital combiner matrix $\mathbf{W}_{\text{BB}}$. The augmented Lagrangian function of $(\mathcal{P}_6)$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{Z},\mathbf{W}_{\text{RF}},\mathbf{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}},\mathbf{\Lambda}) = &\frac{1}{2}\|\mathbf{W}_{\text{DBF}}-\mathbf{Z}\|_F^2 + \mathbb{1}_{\mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}}\{\mathbf{W}_{\text{RF}}\} + \mathbb{1}_{\mathcal{D}_{\text{RX}}^{L_{\text{R}}\times L_{\text{R}}}}\{\mathbf{\Delta}_{\text{RX}}\} \\ &+ \frac{\alpha}{2}\|\mathbf{Z}+\mathbf{\Lambda}/\alpha - \mathbf{W}_{\text{RF}}\mathbf{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}}\|_F^2 + \gamma_{\text{R}} P_{\text{RX}}(\mathbf{\Delta}_{\text{RX}}),\end{aligned} \quad (5.31)$$

Algorithm 9 Proposed ADMM Solution for the A/D Hybrid Combiner Design

```
Initialize: $\mathbf{Z}$, $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{\Delta}_{\mathrm{RX}}$, $\mathbf{W}_{\mathrm{BB}}$ with random values, $\mathbf{\Lambda}$ with zeros, $\alpha = 1$ and $n = 1$
while $n \leq N_{\max}$ do
    Update $\mathbf{Z}_{(n)}$ using solution (5.33),
        $\mathbf{W}_{\mathrm{RF}(n)}$ using solution (5.34),
        $\mathbf{\Delta}_{\mathrm{RX}(n)}$ by solving $(\mathcal{P}_{6\mathrm{C}})$ using CVX [72],
        $\mathbf{W}_{\mathrm{BB}(n)}$ using solution (5.35), and
        update $\mathbf{\Lambda}_{(n)}$ using solution (5.32).
    $n \leftarrow n + 1$
end while
return $\mathbf{W}^{\star}_{\mathrm{RF}}$, $\mathbf{\Delta}^{\star}_{\mathrm{RX}}$, $\mathbf{W}^{\star}_{\mathrm{BB}}$
```

where α is a scalar penalty parameter and $\mathbf{\Lambda} \in \mathbb{C}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}$ is the Lagrange Multiplier matrix. According to the ADMM approach [116], the solution to $(\mathcal{P}_6)$ is derived by the following iterative steps:

$$\begin{aligned}
(\mathcal{P}_{6\mathrm{A}}) : \mathbf{Z}_{(n)} &= \arg\min_{\mathbf{Z}} \frac{1}{2} \|(1+\alpha)\mathbf{Z} - \mathbf{W}_{\mathrm{DBF}} + \mathbf{\Lambda}_{(n-1)} \\
&\quad - \alpha \mathbf{W}_{\mathrm{RF}(n-1)} \mathbf{\Delta}_{\mathrm{RX}(n-1)} \mathbf{W}_{\mathrm{BB}(n-1)}\|_F^2, \\
(\mathcal{P}_{6\mathrm{B}}) : \mathbf{W}_{\mathrm{RF}(n)} &= \arg\min_{\mathbf{W}_{\mathrm{RF}}} \mathbb{1}_{\mathcal{W}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}}\{\mathbf{W}_{\mathrm{RF}}\} + \frac{\alpha}{2} \times \|\mathbf{Z}_{(n)} + \mathbf{\Lambda}_{(n-1)}/\alpha \\
&\quad - \mathbf{W}_{\mathrm{RF}} \mathbf{\Delta}_{\mathrm{RX}(n-1)} \mathbf{W}_{\mathrm{BB}(n-1)}\|_F^2, \\
(\mathcal{P}_{6\mathrm{C}}) : \mathbf{\Delta}_{\mathrm{RX}(n)} &= \arg\min_{\mathbf{\Delta}_{\mathrm{RX}}} \|\mathbf{y}_{\mathrm{c}} - \mathbf{\Psi}_{\mathrm{R}} \mathrm{vec}(\mathbf{\Delta}_{\mathrm{RX}})\|_2^2 + \gamma_{\mathrm{R}} P_{\mathrm{RX}}(\mathbf{\Delta}_{\mathrm{RX}}) \\
&\quad \text{subject to } \mathbf{\Delta}_{\mathrm{RX}} \in \mathcal{D}_{\mathrm{RX}}, \\
(\mathcal{P}_{6\mathrm{D}}) : \mathbf{W}_{\mathrm{BB}(n)} &= \arg\min_{\mathbf{W}_{\mathrm{BB}}} \frac{\alpha}{2} \|\mathbf{Z}_{(n)} + \mathbf{\Lambda}_{(n-1)}/\alpha - \mathbf{W}_{\mathrm{RF}(n)} \mathbf{\Delta}_{\mathrm{RX}(n)} \mathbf{W}_{\mathrm{BB}}\|_F^2, \\
\mathbf{\Lambda}_{(n)} &= \mathbf{\Lambda}_{(n-1)} + \alpha \left(\mathbf{Z}_{(n)} - \mathbf{W}_{\mathrm{RF}(n)} \mathbf{\Delta}_{\mathrm{RX}(n)} \mathbf{W}_{\mathrm{BB}(n)} \right),
\end{aligned} \tag{5.32}$$

where n denotes the iteration index, $\mathbf{y}_{\mathrm{c}} = \mathrm{vec}(\mathbf{Z}_{(n)} + \mathbf{\Lambda}_{(n-1)}/\alpha)$ and $\mathbf{\Psi}_{\mathrm{R}} = \mathbf{W}_{\mathrm{BB}(n-1)} \otimes \mathbf{W}_{\mathrm{RF}(n)}$ ($\otimes$ is the Khatri-Rao product).

We solve the optimization problems $(\mathcal{P}_{6\mathrm{A}})$-$(\mathcal{P}_{6\mathrm{D}})$ in a similar way to the derivations in Section 5.3 for the TX. The solution for $\mathbf{Z}_{(n)}$ is:

$$\mathbf{Z}_{(n)} = \frac{1}{\alpha+1}\big(\mathbf{W}_{\mathrm{DBF}} - \mathbf{\Lambda}_{(n-1)} + \alpha \mathbf{W}_{\mathrm{RF}\,(n-1)} \mathbf{\Delta}_{\mathrm{RX}(n-1)} \mathbf{W}_{\mathrm{BB}\,(n-1)}\big). \tag{5.33}$$

The equation for $\mathbf{W}_{\mathrm{RF}\,(n)}$ is as follows:

$$\begin{aligned}\mathbf{W}_{\mathrm{RF}\,(n)} = \Pi_{\mathcal{W}}\Big\{&\big(\mathbf{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\big)\mathbf{W}_{\mathrm{BB}}{}_{(n-1)}^{H}\mathbf{\Delta}_{\mathrm{RX}(n-1)}^{H}\\ &\big\{\alpha\mathbf{\Delta}_{\mathrm{RX}(n-1)}\mathbf{W}_{\mathrm{BB}\,(n-1)}\mathbf{W}_{\mathrm{BB}}{}_{(n-1)}^{H}\mathbf{\Delta}_{\mathrm{RX}(n-1)}^{H}\big\}^{-1}\Big\}.\end{aligned} \tag{5.34}$$

The solution to $\mathbf{\Delta}_{\mathrm{RX}(n)}$ is obtained by solving ($\mathcal{P}_{6\mathrm{C}}$) using CVX [72]. The matrix $\mathbf{W}_{\mathrm{BB}\,(n)}$ is obtained as follows:

$$\mathbf{W}_{\mathrm{BB}\,(n)} = \big\{\alpha\mathbf{\Delta}_{\mathrm{RX}(n)}^{H}\mathbf{W}_{\mathrm{RF}}{}_{(n)}^{H}\mathbf{W}_{\mathrm{RF}\,(n)}\mathbf{\Delta}_{\mathrm{RX}(n)}\big\}^{-1}\mathbf{\Delta}_{\mathrm{RX}(n)}^{H}\mathbf{W}_{\mathrm{RF}}{}_{(n)}^{H}\big(\mathbf{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\big). \tag{5.35}$$

Algorithm 9 provides the complete procedure to obtain $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{\Delta}_{\mathrm{RX}}$ and $\mathbf{W}_{\mathrm{BB}}$. It starts by initializing the entries of the matrices $\mathbf{Z}$, $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{\Delta}_{\mathrm{RX}}$, $\mathbf{W}_{\mathrm{BB}}$ with random values and the entries of the Lagrange multiplier matrix $\mathbf{\Lambda}$ with zeros. For iteration index n, $\mathbf{Z}_{(n)}$, $\mathbf{W}_{\mathrm{RF}\,(n)}$, $\mathbf{\Delta}_{\mathrm{RX}(n)}$, $\mathbf{W}_{\mathrm{BB}\,(n)}$ are updated at each iteration step by using the solution in (5.33), (5.34), solving ($\mathcal{P}_{6\mathrm{C}}$) using CVX, (5.35) and (5.32), respectively. The operator $\Pi_{\mathcal{W}}$ projects the solution onto the set $\mathcal{W}$. This procedure is identical to problem ($\mathcal{P}_{3\mathrm{B}}^{''}$) in Section 5.3, except that the set $\mathcal{W}$ replaces $\mathcal{F}$. A termination criterion is defined using a maximum number of iterations ($N_{\max}$) or a fidelity criterion similar to (5.30). Upon convergence, the number of bits for each ADC is obtained by using (5.2) and quantizing to the nearest integer value. The optimal hybrid combining matrices $\mathbf{W}_{\mathrm{RF}}^{\star}$, $\mathbf{\Delta}_{\mathrm{RX}}^{\star}$, $\mathbf{W}_{\mathrm{BB}}^{\star}$ are obtained at the end of this algorithm.

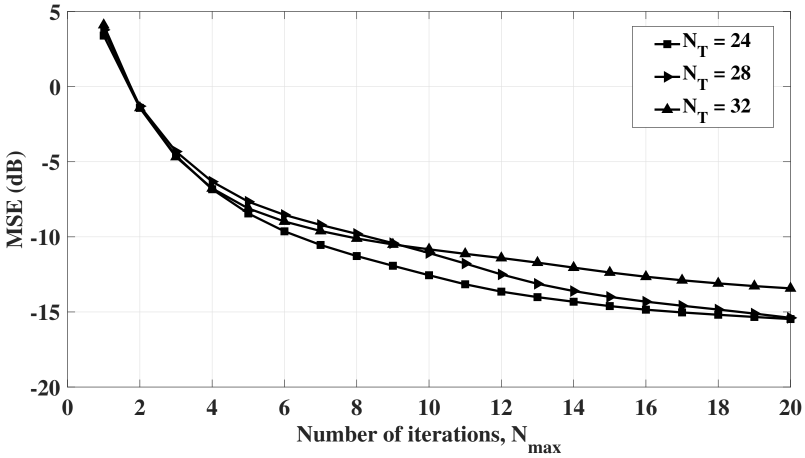


Figure 5.2: Convergence of the proposed ADMM solution at the TX for different N_T at $\gamma_\mathrm{T} = 0.001$.

Computational Complexity Analysis of Algorithm 9

Similar to Algorithm 8 for the TX, the complexity of the solution of ($\mathcal{P}_{6\mathrm{C}}$) can be upper-bounded by $\mathcal{O}((L_\mathrm{R}^2 N_\mathrm{R} N_\mathrm{s})^3)$ which can be improved significantly by exploiting the structure of $\boldsymbol{\Psi}_\mathrm{R}$.

Once the optimal DAC and ADC bit resolution matrices, i.e., $\boldsymbol{\Delta}_\mathrm{TX}$ and $\boldsymbol{\Delta}_\mathrm{RX}$, and optimal hybrid precoding and combining matrices, i.e., $\mathbf{F}_\mathrm{RF}$, $\mathbf{F}_\mathrm{BB}$ and $\mathbf{W}_\mathrm{RF}$, $\mathbf{W}_\mathrm{BB}$, are obtained then they can be plugged into (5.13) and (5.14) to obtain the maximum EE in (5.12). In the next section, we discuss the simulation results based on the proposed solution at the TX and the RX, and comparison with existing benchmark techniques.

5.5 Simulation Results

In this section, we evaluate the performance of the proposed ADMM solution using computer simulation results. All the results have been averaged over 1000

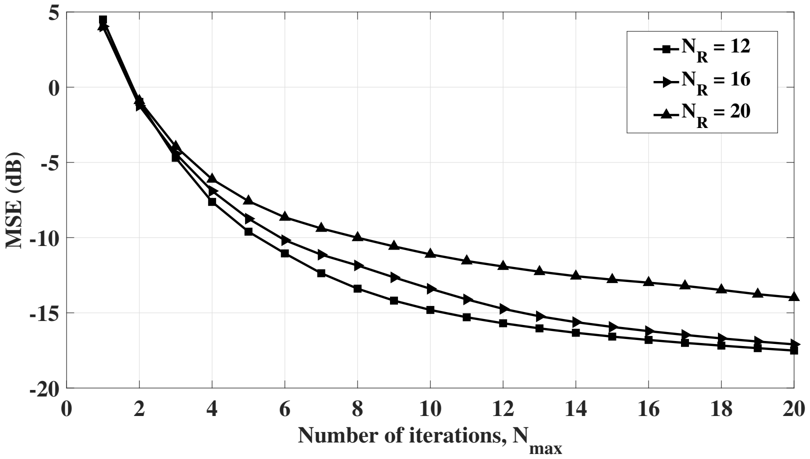


Figure 5.3: Convergence of the proposed ADMM solution at the RX for different N_{R} at $\gamma_{\mathrm{R}} = 0.5$.

Monte-Carlo realizations. For comparison with the proposed ADMM solution, we consider several existing benchmark techniques as follows.

5.5.1 Benchmark Techniques

Digital Beamforming with Full-bit Resolution

We consider the conventional fully digital beamforming architecture, where the number of RF chains at the TX/RX is equal to the number of TX/RX antennas, i.e., $L_{\mathrm{T}} = N_{\mathrm{T}}$ and $L_{\mathrm{R}} = N_{\mathrm{R}}$. In terms of the resolution sampling, we consider full-bit resolution, i.e., $M = 8$-bit, which represents the best case from the achievable SE perspective.

A/D HBF with 1-bit and 8-bit Resolutions

We also consider a A/D HBF architecture with $L_{\mathrm{T}} < N_{\mathrm{T}}$ and $L_{\mathrm{R}} < N_{\mathrm{R}}$, for two cases of DAC/ADC bit resolution: a) 1-bit resolution which usually shows reasonable EE performance, and b) 8-bit resolution which usually shows high SE results.

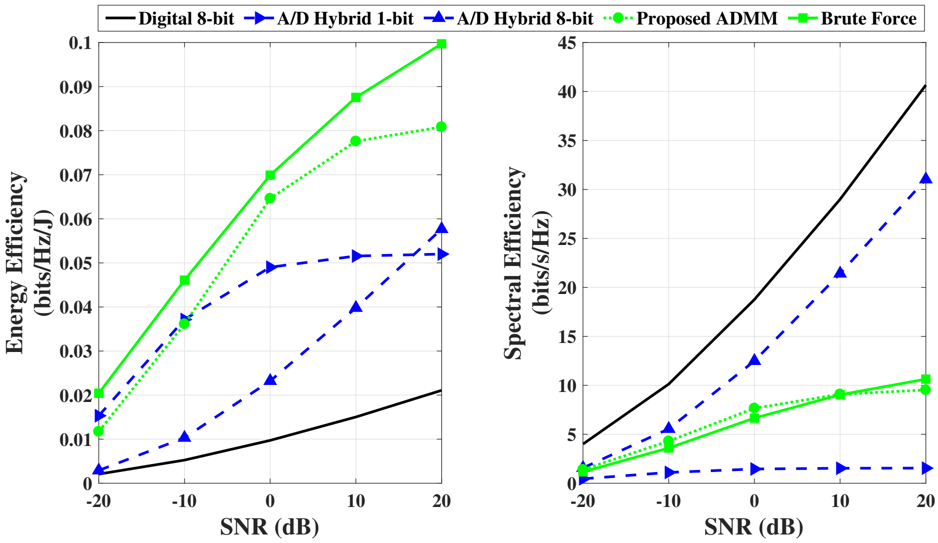


Figure 5.4: EE and SE performance w.r.t. SNR at $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$.

BF with A/D HBF

We also implement an exhaustive search approach as an upper bound for EE maximization called Brute Force (BF), based on [65]. Firstly the EE problem is split into TX and RX optimization problems similar to those for the proposed ADMM approach. Then it makes a search over all the possible DAC and ADC bit resolutions in the range of $[m, M]$ associated with the each RF chain from 1 to L_{T} and 1 to L_{R} at the TX and the RX, respectively. It then finds the best EE out of all the possible cases and chooses the corresponding optimal resolution for each DAC and ADC. This method provides the best possible EE performance and serves as upper bound for EE maximization by the ADMM approach.

Complexity Comparison with the BF Approach

The proposed ADMM solution has lower complexity than the upper bound BF approach because the BF technique involves a search over all the possible DAC/ADC bit resolutions while the proposed ADMM solution directly optimizes the number of bits at each DAC/ADC. We constrain the number of RF chains

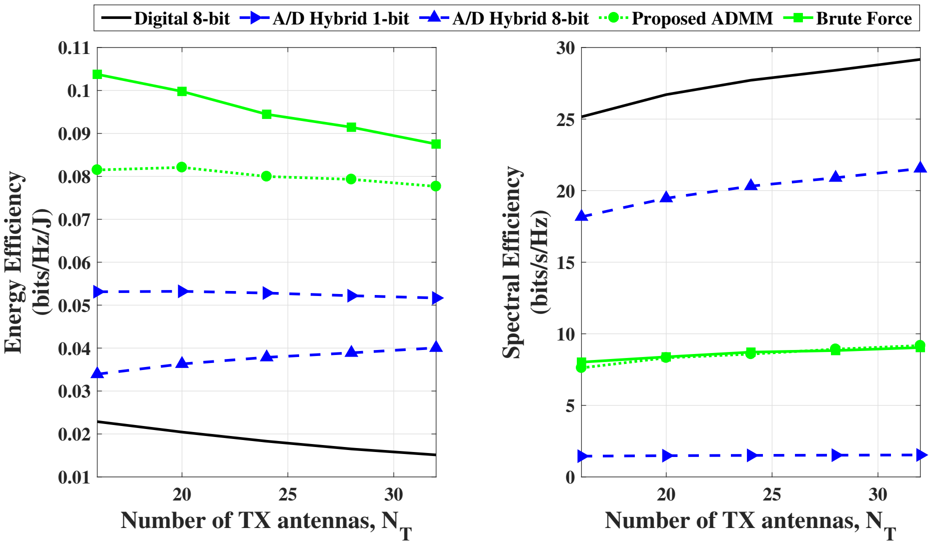


Figure 5.5: EE and SE performance w.r.t. N_T at SNR $= 10$ dB, $\gamma_\mathrm{T} = 0.001$ and $\gamma_\mathrm{R} = 0.5$.

$L_\mathrm{T} = L_\mathrm{R} = 5$ for the BF approach due to the high complexity order which is $\mathcal{O}(M^{L_\mathrm{T}})$ and $\mathcal{O}(M^{L_\mathrm{R}})$ at the TX and the RX, respectively.

5.5.2 System Setup

We set the following parameters, unless specified otherwise, to obtain the desired results: $N_\mathrm{T} = 32$, $L_\mathrm{T} = 5$, $N_\mathrm{s} = L_\mathrm{T}$, $L_\mathrm{R} = L_\mathrm{T}$, $N_\mathrm{R} = 5$, $N_\mathrm{cl} = 2$, $N_\mathrm{ray} = 3$, $N_\mathrm{max} = 20$, $m = 1$, $M = 8$, $\gamma_\mathrm{T}^{max} = 0.1$, $\gamma_\mathrm{R}^{max} = 1$, $\alpha = 1$ and $\sigma_{\alpha,i}^2 = 1$. The azimuth angles of departure and arrival are computed with uniformly distributed mean angles, and each cluster follows a Laplacian distribution about the mean angle. The antenna elements in the ULA are spaced by distance $d = \lambda/2$. The SNR is given by the inverse of the noise variance, i.e., $1/\sigma_\mathrm{n}^2$. The transmit vector $\mathbf{s}$ is composed of the normalized i.i.d. Gaussian symbols. The values used for the power terms [73] in the power model equations in (5.15) and (5.17) are $P_\mathrm{DAC} = P_\mathrm{ADC} = 100$ mW, $P_\mathrm{CT} = P_\mathrm{CR} = 10$ W, $P_\mathrm{T} = P_\mathrm{R} = 100$ mW and $P_\mathrm{PT} = P_\mathrm{PR} = 10$ mW.

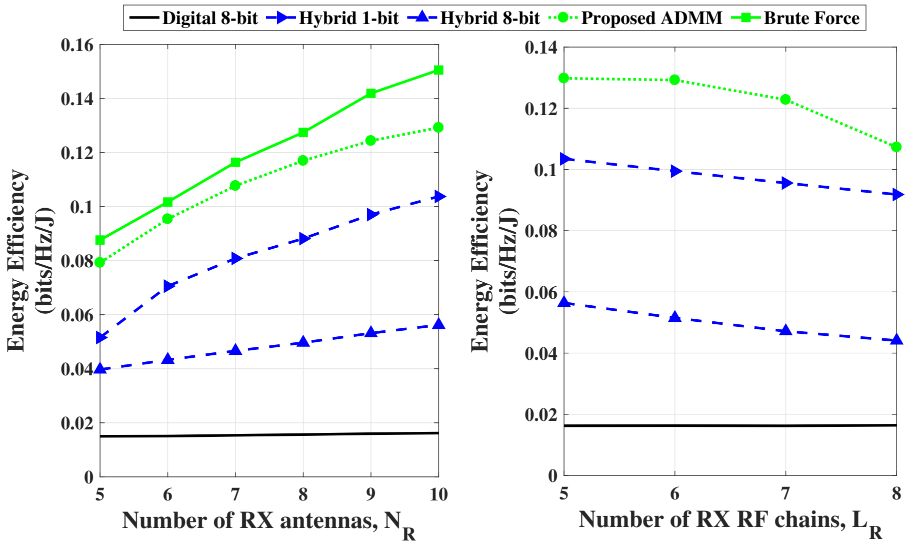


Figure 5.6: EE performance w.r.t. N_{R} and L_{R} at SNR $= 10$ dB, $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$.

5.5.3 Convergence of the Proposed ADMM Solution

Figs. 5.2 and 5.3 show the convergence of the ADMM solution at the TX and the RX as proposed in Algorithm 8 and Algorithm 9, respectively, to obtain the optimal bit resolution at each DAC/ADC and the corresponding optimal precoder/combiner matrices. It can be observed from Fig. 5.2 that the proposed solution converges rapidly within 16 iterations and the normalized mean square error (NMSE) at the TX, $\|\mathbf{F}_{\mathrm{DBF}} - \mathbf{F}_{\mathrm{RF}(N_{\max})}\boldsymbol{\Delta}_{\mathrm{TX}(N_{\max})}\mathbf{F}_{\mathrm{BB}(N_{\max})}\|_F^2/\|\mathbf{F}_{\mathrm{DBF}}\|_F^2$, goes as low as -15 dB. Similarly, in Fig. 5.3, the proposed solution again converges rapidly and the NMSE at the RX, $\|\mathbf{W}_{\mathrm{DBF}}-\mathbf{W}_{\mathrm{RF}(N_{\max})}\boldsymbol{\Delta}_{\mathrm{RX}(N_{\max})}\mathbf{W}_{\mathrm{BB}(N_{\max})}\|_F^2/\|\mathbf{W}_{\mathrm{DBF}}\|_F^2$, goes as low as -17 dB. A lower number of TX/RX antennas shows lower NMSE for a given number of iterations as expected, since fewer parameters are required to be estimated.

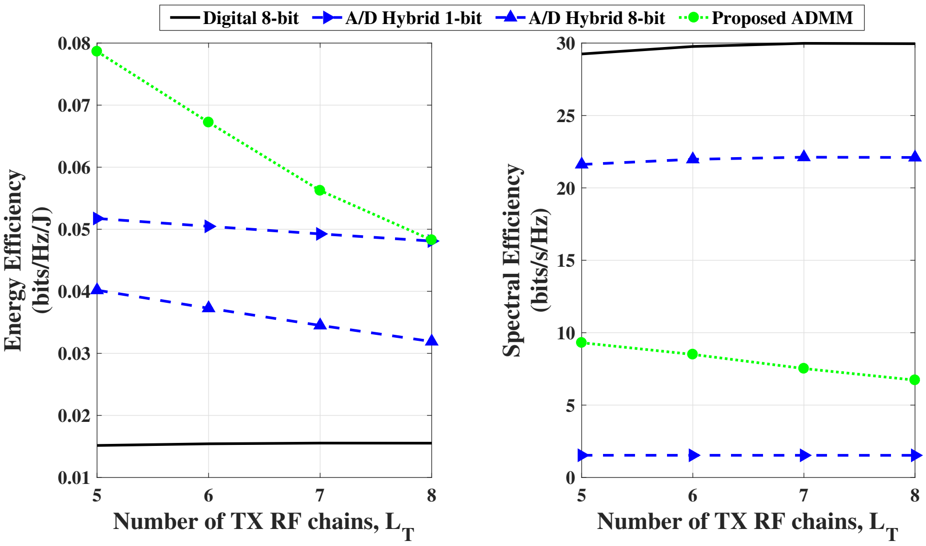


Figure 5.7: EE and SE performance w.r.t. $L_{\rm T}$ at SNR $= 10$ dB, $\gamma_{\rm T} = 0.001$ and $\gamma_{\rm R} = 0.5$.

5.5.4 EE and SE performance of Proposed ADMM

Fig. 5.4 shows the performance of the proposed ADMM solution compared with existing benchmark techniques w.r.t. SNR at $\gamma_{\rm T} = 0.001$ and $\gamma_{\rm R} = 0.5$. The proposed ADMM solution achieves high EE which is computed by (5.12) after obtaining the optimal DAC and ADC bit resolution matrices, i.e., $\mathbf{\Delta}_{\rm TX}$ and $\mathbf{\Delta}_{\rm RX}$, and optimal hybrid precoding and combining matrices, i.e., $\mathbf{F}_{\rm RF}$, $\mathbf{F}_{\rm BB}$ and $\mathbf{W}_{\rm RF}$, $\mathbf{W}_{\rm BB}$. The results are plugged into (5.13) and (5.14) to evaluate rate and power respectively. The EE for the proposed solution has similar performance to the BF approach and is better than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines. For example, at SNR $= 10$ dB, the proposed ADMM solution outperforms the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.03 bits/Hz/J, 0.04 bits/Hz/J and 0.065 bits/Hz/J, respectively.

The proposed solution also exhibits better SE, which is the rate in (5.13) after obtaining the optimal DAC and ADC bit resolution matrices, and optimal hybrid precoding and combining matrices, than the hybrid 1-bit and has similar

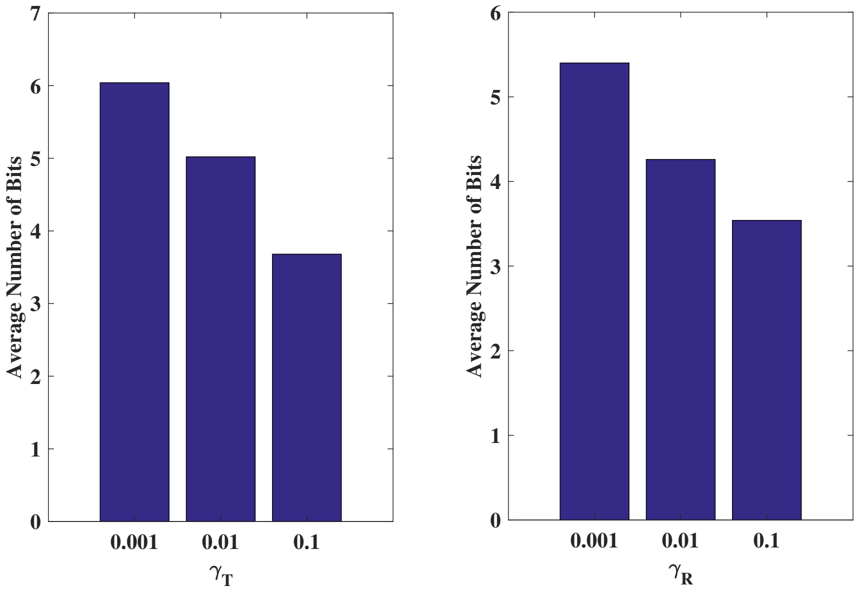


Figure 5.8: Average number of bits for proposed ADMM w.r.t. γ_{T} and γ_{R} at the TX and the RX, respectively, at SNR $= 10$ dB.

performance to the BF approach for high and low SNR regions and hybrid 8-bit baseline for low SNR region. Note that the proposed ADMM solution enables the selection of different resolutions for different DACs/ADCs and thus, it offers a better trade-off for EE versus SE than existing approaches which are based on a fixed DAC/ADC bit resolution.

Fig. 5.5 shows the EE (from (5.12)) and SE (from (5.13)) performance results w.r.t. the number of TX antennas N_{T} at 10 dB SNR, $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$. The proposed ADMM solution again achieves high EE and performs similar to the BF approach and better than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines. For example, at $N_{\mathrm{T}} = 20$, the proposed ADMM solution outperforms hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.03 bits/Hz/J, 0.045 bits/Hz/J and 0.06 bits/Hz/J, respectively. The proposed ADMM solution also exhibits SE performance similar to the BF approach and better than the hybrid 1-bit baseline.

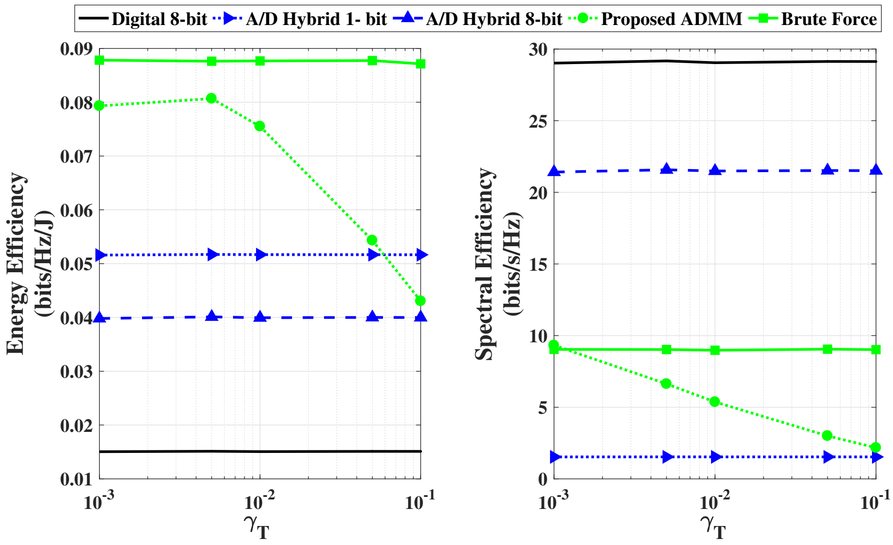


Figure 5.9: EE and SE performance w.r.t. γ_T at SNR = 10 dB.

Fig. 5.6 shows the EE performance results w.r.t. the number of RX antennas N_R and the number of RX RF chains L_R, respectively, at 10 dB SNR, $\gamma_T = 0.001$ and $\gamma_R = 0.5$. The proposed ADMM solution again achieves high EE which decreases with increase in the number of RX RF chains,and performs similar to the BF approach (for versus N_R) and better than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines. For example, at $N_R = 7$, the proposed ADMM solution outperforms hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.03 bits/Hz/J, 0.06 bits/Hz/J and 0.09 bits/Hz/J, respectively. Also, for example, at $L_R = 6$, the proposed ADMM solution outperforms hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.025 bits/Hz/J, 0.08 bits/Hz/J and 0.115 bits/Hz/J, respectively. Note that, due to the high complexity of the BF approach, we do not plot results for this approach w.r.t. L_T and L_R.

Fig. 5.7 shows the EE and SE performance results w.r.t. the number of TX RF chains L_T at 10 dB SNR, $\gamma_T = 0.001$ and $\gamma_R = 0.5$. The proposed ADMM solution achieves high EE, though this decreases with increase in the number of

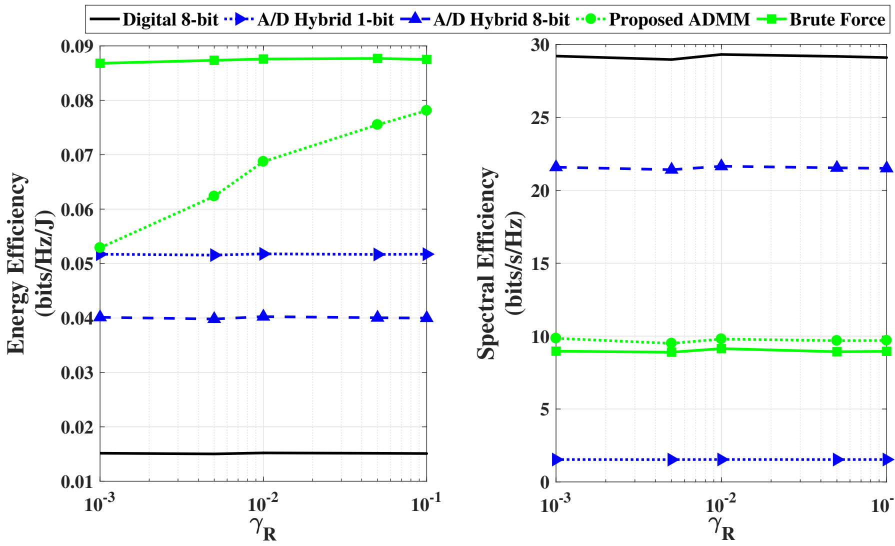


Figure 5.10: EE and SE performance w.r.t. γ_{R} at SNR $= 10$ dB.

TX RF chains ADMM achieves better EE performance than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit resolution baselines. Also, the proposed ADMM solution exhibits SE performance better than the hybrid 1-bit baseline.

Furthermore, we investigate the performance over the trade-off parameters γ_{T} and γ_{R} introduced in $(\mathcal{P}_2)$ and $(\mathcal{P}_5)$, respectively. Fig. 5.8 shows the bar plot of the average of the optimal number of bits selected by the proposed ADMM solution for each DAC versus γ_{T} and for each ADC versus γ_{R}. It can be observed that the average optimal number decreases with the increase in γ_{T} and γ_{R}, for example, the average number of DAC bits is around 6 for $\gamma_{\mathrm{T}} = 0.001$, 5 for $\gamma_{\mathrm{T}} = 0.01$ and 4 for $\gamma_{\mathrm{T}} = 0.1$. Similarly, at the RX, the average number of ADC bits is about 5 for $\gamma_{\mathrm{R}} = 0.001$, 4 for $\gamma_{\mathrm{R}} = 0.01$ and 3 for $\gamma_{\mathrm{R}} = 0.1$. This is because increasing γ_{T} or γ_{R} gives more weight to the power consumption.

Figs. 5.9 and 5.10 show the EE and SE plots for several solutions w.r.t. γ_{T} and γ_{R} at the TX and the RX, respectively. It can be observed that the proposed solution achieves higher EE performance than the fixed bit allocation solutions such as the digital full-bit, the hybrid 1-bit and the hybrid 8-bit baselines and

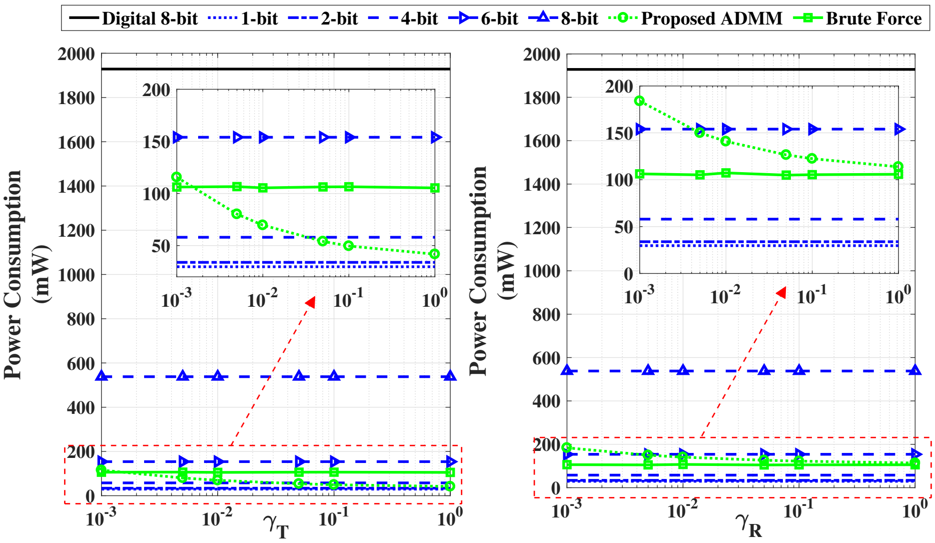


Figure 5.11: Power consumption w.r.t. γ_{T} and γ_{R} at the TX and the RX, respectively, at SNR $= 10$ dB.

achieves comparable EE and SE results to the BF approach. These curves also show that adjusting γ_{T} and γ_{R} values allow the system to vary the energy-rate trade-off. Note that the TX also accounts for the extra power term, i.e., $\mathrm{tr}(\mathbf{F}\mathbf{F}^H)$ as shown in (5.15) which means that the selected γ_{T} parameter at the TX is lower than the selected γ_{R} parameter at the RX. Fig. 5.11 shows that the power consumption in the proposed case is low and decreases with the increase in the trade-off parameter γ_{T} and γ_{R} values unlike digital 8-bit, fixed bit resolution hybrid baselines and the BF approach.

5.6 Summary

This chapter proposes an energy efficient mmWave A/D hybrid MIMO system which can vary dynamically the DAC and ADC bit resolutions at the TX and the RX, respectively. This method uses the decomposition of the A/D hybrid precoder/combiner matrix into three parts representing the analog precoder/combiner matrix, the DAC/ADC bit resolution matrix and the digital precoder/combiner matrix. These three matrices are optimized by a novel ADMM solution which outperforms the EE of the digital full-bit, the hybrid 1-bit beamforming and the hybrid 8-bit beamforming baselines, for example, by 3%, 4% and 6.5%, respectively, for a typical value of 10 dB SNR. There is an energy-rate trade-off with the BF approach which yields the upper bound for EE maximization and the proposed ADMM solution exhibits lower computational complexity. Moreover, the proposed ADMM solution enables the selection of the optimal resolution for each DAC/ADC and thus, it offers better trade-off for data rate versus EE than existing approaches that are based on a fixed DAC/ADC bit resolution.

In the next section, we conclude the PhD research work and provide future work that will be carried out in relation to the research associated with the mmWave A/D HBF MIMO systems.

Chapter 6

Conclusions and Future Work

This thesis contributed to the field of energy efficient and low complexity solutions for mmWave MIMO systems with HBF architectures. Both full resolution and low resolution sampling cases are considered. In this concluding chapter, we first summarize the key findings of this thesis in Section 6.1. Then we proceed with the potential improvements and future work in Section 6.2.

6.1 Conclusions

In this thesis, we optimized mmWave HBF MIMO systems to achieve high EE gains with low complexity, which has not been widely studied in the literature. These communication techniques may be implemented in 5G and Beyond 5G standards. In a nutshell, we successfully designed energy efficient mmWave HBF MIMO systems with low complexity by exploiting the structure of complex and power hungry components such as RF chains in Chapter 3 and DAC/ADC converting units in Chapters 4 and 5. We also exploited the sparsity of the mmWave channel in part of Chapter 4 and provided an efficient and low complexity solution for sparse channel estimation while employing low resolution sampling. In following subsections we summarize the key findings of this thesis.

6.1.1 EE Maximization by Dynamic RF Chain Selection

The research work in Chapter 3 proposes an energy efficient A/D HBF framework with a novel architecture for a mmWave MIMO system, where we optimize the active number of RF chains through fractional programming. The proposed DM based framework reduces the complexity significantly and achieves almost the same EE performance as the state of the art BF approach. Both approaches achieve higher EE performance when compared with the fully digital beamforming and the analog beamforming solutions. In particular, the proposed solution only needs to compute the precoder and combiner matrices once, after the number of active RF chains are decided through the Dinkelbach optimization solution.

The modified version of the proposed solution, i.e., FS approach, shows very similar performance to the proposed DM but the complexity increases significantly. The codebook-free designs such as ADMM and SVD based solutions, when incorporated with the proposed framework also achieve better EE performance over the fixed number of RF chains case. It is also shown that GP incorporated with the proposed DM is a faster and less complex approximation solution to compute the precoder and combiner matrices than OMP.

6.1.2 Channel Estimation and EE Maximization with Low Resolution Sampling

The research work in Chapter 4 discussed sparse channel estimation and EE maximization solutions with low resolution sampling at the ADCs and the DACs, respectively. An algorithm based on EM density estimation, plus the SURE parametric denoiser with the GAMP framework is proposed for a mmWave hybrid MIMO system with low resolution sampling at the RX. We exploit this EM-

SURE-GAMP algorithm to estimate the channel which provides the flexibility to avoid strong assumptions on the channel priors where SURE, depending on the noisy observation, is minimized to adaptively optimize the denoiser within the parametric class at each iteration. When compared with the state of the art EM-GAMP solution, the MSE of the proposed solution performs better with respect to low and high SNR regimes, with respect to the number of ADC bits, and with respect to the training length.

Furthermore in Chapter 4, we consider low resolution sampling at the TX. We consider the case where all DACs have the same sampling resolution for each RF chain and aim to optimize the number of active RF chains and associated resolution of DACs. The proposed method achieves similar EE performance with the upper bound of the derived exhaustive search approach, i.e., BF approach, while it exhibits lower computational complexity and fast convergence.

6.1.3 EE Maximization by Joint Bit Allocation and HBF Optimization

The research work in Chapter 5 proposes an energy efficient mmWave A/D hybrid MIMO system which can vary dynamically the DAC and ADC bit resolutions at the TX and the RX, respectively. This method uses the decomposition of the A/D hybrid precoder/combiner matrix into three parts representing the analog precoder/combiner matrix, the DAC/ADC bit resolution matrix and the digital precoder/combiner matrix. These three matrices are optimized by a novel ADMM solution which outperforms the EE of the digital full-bit, the hybrid 1-bit beamforming and the hybrid 8-bit beamforming baselines, for example, by 3%, 4% and 6.5%, respectively, for a typical value of 10 dB SNR.

Furthermore, there is an energy-rate trade-off with the BF approach which yields the upper bound for EE maximization and the proposed ADMM solution exhibits lower computational complexity. Moreover, the proposed ADMM solution enables the selection of the optimal resolution for each DAC/ADC and thus, it offers better trade-off for data rate versus EE than existing approaches that are based on a fixed DAC/ADC bit resolution.

6.2 Future Work

In this section, we introduce several potential research problems which can be studied for future work and as an improvement over the existing techniques. These research problems may be considered for beyond 5G and Sixth Generation (6G) communication standards.

6.2.1 EE Maximization with Combined TX-RX Optimization for Bit Allocation and RF Chain Selection

For Chapter 3, we focus on maximizing the EE but extending these techniques to consider both estimated channels and frequency selective channels can be considered for future work. Furthermore, the research work about EE maximization in Chapter 4 considers the case where all DACs have the same sampling resolution for each RF chain and select the best subset of the active RF chains and the DAC resolution at the TX. This work can be extended for a combined problem at the TX and the RX. For example, we present bit allocation and hybrid combining optimization solution for the RX in [27] and extend the problem for EE maximization for the case of joint TX and RX problem in [28].

Similarly, we can implement our technique for EE maximization used at the TX in [26] for a combined TX-RX problem. An EE maximization problem, with rate and power consisting of system and channel model parameters at both the TX and the RX, can be provided. The joint TX and RX problem can be decoupled to deal with the TX and the RX separately. The corresponding problems can be solved by technique based on the DM and subset selection optimization such as in [26]. We can also implement similar exhaustive search approach as the BF approach, for example as shown in [28], to serve as an upper bound on the EE performance and show the performance trade-offs. This future work has been listed as "under preparation" in Appendix A.1, i.e., A.1.4, at the end of this thesis.

6.2.2 Channel Estimation with Low Resolution Sampling for Phase Shifters- and Lens-Based Hybrid MIMO

The research work about channel estimation in Chapter 4 uses the GAMP framework with EM density estimation and the SURE parameteric denoiser to estimate the sparse channel with low MSE and with low computational complexity. This work can be further extended with VAMP framework and performance trade-offs in terms of MSE and complexity can be observed for a phase shifters-based hybrid MIMO system. Moreover, the narrowband channel model can be replaced by wideband channel model and EM-based density estimation can be improved with more advanced CS approaches to achieve higher accuracy and lesser complexity.

Furthermore, we know that MIMO systems with beamforming capabilities are required to compensate for the high path-loss at mmWave frequencies. Recently, a practical two-stage Rotman lens beamformer has demonstrated increased antenna gain with reduced implementation complexity, since the conventional beam selection network was omitted. In a related future work, we can adopt this lens-based MIMO system with HBF architecture and investigate its performance in terms of channel estimation with low resolution sampling at the RX. Although this design is characterised by low-complexity and low-cost, the analog beamformer and the ADCs introduce several impairments to the received signal. To mitigate these effects, we can develop a robust maximum a posteriori (MAP) estimator based on the EM iterative algorithm. This sparse channel estimation method for lens-based MIMO system can be compared with the conventional EM and minimum MSE approaches in the medium to high SNR regimes and for different bit resolution values.

Bibliography

[1] NGMN 5G White Paper, pp. 1-125, Feb. 2015.

[2] 2020 Networld White Paper for Research Beyond 5G, pp. 1-43, Oct. 2015.

[3] X. Li, A. Gani, R. Salleh, and O. Zakaria, "The future of mobile wireless communication networks," *IEEE Intern. Conf. Commun. Soft. Netw.*, Macau, pp. 554-557, Feb. 2009.

[4] Cisco visual networking index forecast 2017–22, `https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html`

[5] Ericsson Mobility Report, pp. 1-32, Nov. 2018.

[6] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101-107, June 2011.

[7] S. Rangan et al., "Millimeter-wave cellular wireless networks: potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366385, Mar. 2014.

[8] J. G. Andrews et al., "What will 5G be?", *IEEE Journ. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, June 2014.

[9] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: it will work!", *IEEE Access*, vol. 1, pp. 335-349, 2013.

[10] F. Boccardi et al., “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 7480, Feb. 2014.

[11] IEEE 5G and Beyond Technology Roadmap Whitepaper, pp. 1-39, Oct. 2017.

[12] M. Akdeniz, et al., “Millimeter wave channel modeling and cellular capacity evaluation,” *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1164-1179, June 2014.

[13] T. S. Rappaport et al., “Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design,” *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029-3056, Sept. 2015.

[14] S. Han et al., “Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,” *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.

[15] R. W. Heath et al., “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[16] O. E. Ayach et al., “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[17] A. Kaushik et al.,“Sparse hybrid precoding and combining in millimeter wave MIMO systems,” in *Proc. IET Radio Prop. Tech. 5G*, Durham, UK, pp. 1-7, Oct. 2016.

[18] T. E. Bogale et al., “On the number of rf chains and phase shifters and scheduling design with hybrid analog digital beamforming,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311-3326, May 2016.

[19] S. Payami et al., “Hybrid beamforming for large antenna arrays with phase shifter selection,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258-7271, Nov. 2016.

[20] S. Payami et al., “Hybrid beamforming with a reduced number of phase shifters for massive mimo systems,” *IEEE Trans. Veh. Tech.*, vol. 67, no. 6, pp. 4843-4851, June 2018.

[21] A. Li and C. Masouros, “Hybrid analog-digital millimeter-wave mumimo transmission with virtual path selection,” *IEEE Commun. Letters*, vol. 21, no. 2, pp. 438-441, Feb. 2017.

[22] C. G. Tsinos et al., “Hybrid Analog-Digital Transceiver Designs for mmWave Amplify-and-Forward Relaying Systems,” *Int. Conf. Telecommun. Sig. Process. (TSP), Athens, Greece*, pp. 1-6, 2018.

[23] C. G. Tsinos et al., “Hybrid analog-digital transceiver designs for cognitive radio millimeter wave systems,” *Asilomar Conf. Sig. Syst. Comput., Pacific Grove, CA*, pp. 1785-1789, 2016.

[24] A. Kaushik et al., “Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid Beamforming in Millimeter Wave MIMO Systems,” *IEEE Trans. Green Commun. Netw.*, accepted, July 2019.

[25] A. Kaushik et al., “Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs”, *IEEE EUSIPCO, Rome, Italy*, Sept. 2018.

[26] A. Kaushik et al., “Energy Efficiency maximization of millimeter wave hybrid MIMO systems with low resolution DACs,” *IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, pp. 1-6, May 2019.

[27] A. Kaushik et al., “Energy Efficient Bit Allocation and Hybrid Combining for Millimeter Wave MIMO Systems,” *IEEE Global Commun. Conf. (GLOBECOM)*, Hawaii, USA, accepted, Dec. 2019.

[28] A. Kaushik et al., “Joint Bit Allocation and Hybrid Beamforming Optimization for Energy Efficient Millimeter Wave MIMO Systems,” *J. Sel. Areas Commun.*, arXiv, Sept. 2019.

[29] J. S. Lu et al., “Modeling human blockers in millimeter wave radio links,” *ZTE Commun.*, vol. 10, no. 4, pp. 23-28, Dec. 2012.

[30] A. V. Alejos et al., “Measurement and analysis of propagation mechanisms at 40 GHz: Viability of site shielding forced by obstacles,” *IEEE Trans. Veh. Tech.*, vol. 57, no. 6, pp. 3369-3380, Nov. 2008.

[31] H. Zhao et al., “28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York City,” *IEEE Int. Conf. Commun. (ICC)*, pp. 5163-5167, 2013.

[32] D. Tse and P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, UK, 2004.

[33] T. S. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.

[34] C. Balanis, *Antenna Theory*, Wiley, NY, USA, 1997.

[35] S. Singh et al., “Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control”, *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513-1527, Oct. 2011.

[36] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501-513, Apr. 2016.

[37] L. Liang et al., "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653-656, Dec. 2014.

[38] J. Brady et al., "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Ant. Prop.*, vol. 61, no. 7, pp. 3814-3827, Jul. 2013.

[39] L. Dai et al., "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE/CIC Int. Conf. Commun. China (ICCC)*, pp. 1-6, July 2016.

[40] O. E. Ayach et al., "The capacity optimality of beam steering in large millimeter wave MIMO systems," *IEEE 13th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, pp. 100-104, June 2012.

[41] D. J. Love and R. W. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102-1110, July 2003.

[42] X. Zhang et al., "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005.

[43] J. Ahmadi-Shokouh et al., "Optimal receive soft antenna selection for MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5893-5903, Dec. 2009.

[44] C. G. Tsinos et al., "On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems", *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980-1995, Sept. 2017.

[45] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," *IEEE Custom Integrated Circuits Conf.*, San Jose, CA, pp. 105-112, 2008.

[46] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journ. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539-550, Apr. 1999.

[47] P. B. Kenington and L. Astier, "Power Consumption of A/D Converters for Software Radio Applications," *IEEE Trans. Vehicular Techn.*, vol. 49, pp. 643-650, Mar. 2000.

[48] Y. Chiu, "Scaling of analog-to-digital converters into ultra-deep submicron CMOS," *Proc. CICC*, pp. 375-382, Sept. 2005.

[49] P. Schvan, et al, "A 24GS/s 6b ADC in 90nm CMOS," *ISSCC Dig. Techn. Papers*, pp. 544-545, Feb. 2008.

[50] M. van Elzakker, et al., "A 1.9μW 4.4fJ/Conversion-step 10b 1MS/s Charge-Redistribution ADC," *ISSCC Dig. Techn. Papers*, pp. 244-245, Feb. 2008.

[51] B. Murmann, "The Race for the Extra Decibel: A Brief Review of Current ADC Performance Trajectories," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 58-66, 2015.

[52] T. Cameron, Analog Devices, `https://www.analog.com/media/en/technical-documentation/tech-articles/Bits-to-Beams-RF-Technology-Evolution-for-5G-mmwave-Radios.pdf`

[53] T. Cameron, Analog Devices, `https://www.analog.com/media/en/technical-documentation/white-papers/RF-Technology-for-the-5G-Millimeter-Wave-Radio.pdf`

[54] O. Dabeer, J. Singh, and U. Madhow, "On the limits of communication performance with one-bit analog-to-digital conversion," *Proc. IEEE 7th Workshop Signal Process. Adv. Wireless Commun.*, pp. 1-5, 2006.

[55] A. Mezghani and J. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," *Proc. IEEE Int. Symp. Info. Theory*, pp. 1286-1289, 2007.

[56] X. Yu et al., "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems", *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485-500, Feb. 2016.

[57] O. Orhan et al., "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," *2015 Info. Theory Appl. Workshop, San Diego, CA*, pp. 191-198, 2015.

[58] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, UK, 2004.

[59] Y. C. Eldar and G. Kutyniok, "Compressed Sensing: Theory and Applications," Cambridge University Press, UK, 2012.

[60] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," Springer, NY, USA, 2013.

[61] J. A. Tropp et al., "Algorithms for simultaneous sparse approximation-part I: greedy pursuit", *Signal Process.*, vol. 86, no. 3, pp. 572-588, Mar. 2006.

[62] J. Tropp, and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit”, *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[63] T. Blumensath, and M. E. Davies, “Gradient pursuits”, *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370-2382, June 2008.

[64] Z. Li et al., “A survey on one-bit compressed sensing: theory and applications,” *Front. Comput. Sci.*, vol. 12, no. 2, pp. 217-230, 2018.

[65] R. Zi et al., “Energy efficiency optimization of 5G radio frequency chain systems”, *IEEE Journ. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.

[66] E. Vlachos et al.,“Energy efficient transmitter with low resolution DACs for massive MIMO with partially connected hybrid architecture”, *IEEE Veh. Tech. Conf. (VTC)-Spring*, Porto, Portugal, pp. 1-5, Jun. 2018.

[67] R. Mendez-Rial et al., “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?”, *IEEE Access*, vol. 4, pp. 247-267, Jan. 2016.

[68] X. Gao et al., “Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays”, *IEEE Journ. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998-1009, Apr. 2016.

[69] E. Bjornson et al., “Optimal design of energy-efficient multi-user MIMO systems: is massive MIMO the answer?,” *IEEE Trans. Wireless Commun.*, vol. 14, no.6, pp. 3059-3075, Jun. 2015.

[70] W. Dinkelbach, “On nonlinear fractional programming”, *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.

[71] R. Jagannathan, “On some properties of programming problems in parametric form pertaining to fractional programming”, *Management Science*, vol. 12, no. 7, 1966.

[72] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs”, in *Recent Adv. Learning and Control*, Springer-Verlag Ltd., pp. 95-110, 2008.

[73] T. S. Rappaport et al., “Millimeter wave wireless communications,” *Prentice-Hall*, Sept. 2014.

[74] H. Xu et al., “Spatial and temporal characteristics of 60-GHz indoor channels”, *IEEE Journ. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620-630, Aug. 2002.

[75] B. Le et al., “Analog-to-digital converters,” *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 69-77, Nov. 2005.

[76] J. Choi et al., “Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs,” *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005-2018, May. 2016.

[77] S. Jacobsson et al., “One-bit massive MIMO: channel estimation and high-order modulations,” *IEEE Int. Conf. Commun. Workshop*, pp. 1304-1309, June 2015.

[78] A. Alkhateeb et al., “Channel estimation and hybrid precoding for millimeter wave cellular systems”, *IEEE Journ. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.

[79] J. Lee et al., “Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications,” *IEEE Global Commun. Conf.*, pp. 3326-3331, Dec. 2014.

[80] P. Schniter, and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: the sparse way,” *IEEE Asilomar Conf. Sig. Sys. Comp.*, pp. 273-277, Nov. 2014.

[81] J. Mo et al., “Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information,” *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498-5512, Oct. 2015.

[82] C. Rusu et al., “Low resolution adaptive compressed sensing for mmWave MIMO receivers,” *IEEE Asilomar Conf. Sig. Sys. Comp.*, pp. 1138-1143, Nov. 2015.

[83] J. Mo et al., “Channel estimation in millimeter wave MIMO systems with one-bit quantization,” *IEEE Asilomar Conf. Sig. Sys. Comp.*, pp. 957-961, Nov. 2014.

[84] L. Fan et al., “Uplink achievable rate for massive MIMO systems with low-resolution ADC”, *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2186-2189, Oct. 2015.

[85] M. T. Ivrlac, and J. A. Nossek, “On MIMO channel estimation with single-bit signal-quantization,” *Proc. IEEE Smart Antenn. Workshop*, Feb. 2007.

[86] A. Mezghani et al., “Multiple parameter estimation with quantized channel output,” *Proc. Int. ITG Workshop Smart Antenn.*, pp. 143-150, Oct. 2010.

[87] J. Vila, and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, pp. 4658-4672, May 2014.

[88] S. Rangan et al., "Vector approximate message passing", *arXiv:1610.03082*, Oct. 2016.

[89] J. Mo et al., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141-1154, Mar. 2018.

[90] J. Zhang et al., "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327-1338, June 2017.

[91] A. Mezghani et al., "Transmit processing with low resolution D/A-converters," *Int. Conf. Electronics, Circuits, and Systems*, Tunisia, pp. 683-686, Dec. 2009.

[92] S. Jacobsson et al., "Quantized precoding for massive MU-MIMO," in *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670-4684, Nov. 2017.

[93] C. Guo, and M. E. Davies, "Near optimal compressed sensing without priors: parametric SURE approximate message passing", *IEEE Trans. Signal Process.*, vol. 63, no. 8, Apr. 2015.

[94] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[95] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," *IEEE Int. Workshop Sig. Process. Adv. Wireless Commun. (SPAWC)*, pp. 146-150, June 2015.

[96] E. Vlachos and J. Thompson, “Dithered Beamforming for Channel Estimation in Mmwave-Based Massive Mimo,” *IEEE Int. Conf. Acous. Speech Sig. Process. (ICASSP)*, Calgary, Canada, pp. 3604-3608, Apr. 2018.

[97] R. A. Wannamaker, *The theory of dithered quantization*, National Library Canada, 1997.

[98] F.A. Malekzadeh et al., “Analog Dithering Techniques for Wireless Transmitters,” *Analog Circ. Sig. Process.*, Springer, New York, 2012.

[99] H. C. Papadopoulos et al., “Sequential signal encoding from noisy measurements using quantizers with dynamic bias control,” *IEEE Trans. Info. Theory*, vol. 47, no. 3, pp. 978-1002, Mar. 2001.

[100] S. Kay, “Can detectability be improved by adding noise?,” *IEEE Signal Process. Lett.*, vol. 7, no. 1, pp. 8-10, Jan. 2000.

[101] O. Dabeer and E. Masry, “Multivariate signal parameter estimation under dependent noise from 1-bit dithered quantized data,” *IEEE Trans. Info. Theory*, vol. 54, no. 4, pp. 1637-1654, Apr. 2008.

[102] D. Donoho et al., “Message-passing algorithms for compressed sensing,” *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp.18 914-18 919, 2009.

[103] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Tech. Rep., California Institute of Technology*, Pasadena, CA, 2008.

[104] E. Vlachos et al., “Stochastic gradient pursuit for adaptive equalization of sparse multipath channels,” *IEEE J. Emerg. Sel. Topics Circ. Sys.*, vol. 2, no. 3, pp. 413-423, Sept. 2012.

[105] U. S. Kamilov et al., “Message-passing de-quantization with applications to compressed sensing,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270-6281, 2012.

[106] M. Al-Shoukairi et al., “A GAMP-based low complexity sparse bayesian learning algorithm,” *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 294-308, Jan. 2018.

[107] J. Ma and L. Ping, “Orthogonal AMP,” *IEEE Access*, vol. 5, pp. 2020-2033, 2017.

[108] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” *IEEE Int. Symp. Info. Theory*, pp. 2168-2172, Jul. 2011.

[109] A. Zappone and E. Jorswieck, “Energy Efficiency in Wireless Networks via Fractional Programming Theory,” *Found. Trends Commun. Info. Theory*, vol. 11, no. 3-4, pp. 185-396, 2015.

[110] J. Choi et al., “Resolution-adaptive hybrid MIMO architectures for millimeter wave communications”, *IEEE Trans. Sig. Process.*, vol. 65, no. 23, pp. 6201-6216, Dec. 2017.

[111] J. Mo et al., “Achievable rates of hybrid architectures with few-bit ADC receivers,” *VDE Int. ITG Workshop Smart Antennas*, pp. 1-8, 2016.

[112] T.-C. Zhang et al., “Mixed-ADC massive MIMO detectors: Performance analysis and design optimization,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738-7752, Nov. 2016.

[113] J. Singh, et al., “On the limits of communication with low-precision analog-to-digital conversion at the receiver, *IEEE Trans. Wireless Commun.*, vol. 57, no. 12, pp. 3629-3639, Dec. 2009.

[114] C. G. Tsinos et al., “Symbol-Level Precoding with Low Resolution DACs for Large-Scale Array MU-MIMO Systems,” *Int. Workshop Sig. Process. Adv. Wireless Commun., Kalamata, Greece*, pp. 1-5, 2018.

[115] L. N. Ribeiro et al., ”Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs,” in *IEEE J. Sel. Topics Sig. Process.*, vol. 12, no. 2, pp. 298-312, May 2018.

[116] S. Boyd et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2011.

[117] A. Mezghani and J. A. Nossek, “Capacity lower bound of MIMO channels with output quantization and correlated noise,” *IEEE Int. Symp. Info. Theory (ISIT)*, Cambridge, USA, Jul. 2012.

[118] D. P. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” *IEEE J. Sel. Areas Commun.*, vol. 24,no. 8, pp. 1439-1451, Aug. 2006.

[119] C. G. Tsinos et al., “Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization,” *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 160-174, June 2017.

[120] C. G. Tsinos and B. Ottersten, “An efficient algorithm for unit-modulus quadratic programs with application in beamforming for wireless sensor networks,” *IEEE Sig. Process. Letters*, vol. 25, no. 2, pp. 169-173, Feb. 2018.

[121] D. P. Bertsekas, “Nonlinear programming,” 2nd Edition, Athena Scientific, 1999.

[122] J. Eckstein and D. P. Bertsekas, “On the DouglasRachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 55, no. 1-3, pp. 293-318, 1992.

[123] E. G. Golshtein and N. Tretyakov, “Modified lagrangians in convex programming and their generalizations,” in *Point-to-Set Maps and Mathematical Programming, Springer*, pp. 86-97, 1979.

Appendix A

List of Publications

This chapter comprises a list of the research papers that have been published, submitted or under preparation during the PhD project.

A.1 Journal Papers

First Author

1. **A. Kaushik**, J. Thompson, E. Vlachos, C. Tsinos and S. Chatzinotas, "Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid Beamforming in Millimeter Wave MIMO Systems," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 4, pp. 886-900, Dec. 2019.

2. **A. Kaushik**, E. Vlachos, C. Tsinos, J. Thompson and S. Chatzinotas, "Joint Bit Allocation and Hybrid Beamforming Optimization for Energy Efficient Millimeter Wave MIMO Systems," *IEEE Transactions on Green Communications and Networking*, pp. 1-15, Jan. 2020. (Under Review) [**EiC Invited Paper**]

3. **A. Kaushik**, J. Thompson and E. Vlachos, "Next Generation Wireless Communications with Energy Efficient Millimeter Wave Hybrid MIMO Systems," *Proceedings of the Royal Society A*. (Under Preparation) [**Invited Paper**]

4. **A. Kaushik**, J. Thompson and E. Vlachos, "Energy Efficiency Maximization by Joint Bit Allocation and Active RF Chain Selection in Millimeter Wave Hybrid MIMO Systems," *IEEE Transactions on Green Communications and Networking*. (Under Preparation) [**EiC Invited Paper**]

<u>**Co-Author**</u>

5. Y. Hu, L. Zhao, Z. Yan, **A. Kaushik**, Y. Hou and J. Thompson, "GatedNet: Neural Network Decoding for Decoding over Impulsive Noise Channels," *IEEE Communications Letters*, vol. 23, no. 8, pp. 1381-1384, Aug. 2019.

6. H. Bian, R. Liu, **A. Kaushik**, Y. Hu and J. Thompson, "Design of Segmented CRC-Aided Spinal Codes for IoT Applications," *IET Communications*, pp. 1-8, Sept. 2019. (Under Review)

7. X. Gao, R. Liu and **A. Kaushik**, "Hierarchical Multi-Agent Optimization for Resource Allocation in Cloud Computing Systems," *IEEE Transactions on Parallel and Distributed Systems*, pp. 1-14, Nov. 2019. (Under Review)

8. Y. Hu, R. Liu, **A. Kaushik** and J. Thompson, "Performance Analysis of Downlink NOMA Systems Based on Rateless Codes with Delay Constraints," *IEEE Transactions on Wireless Communications*, pp. 1-15, Dec. 2019. (Under Review)

A.2 Conference Papers

<u>**First Author**</u>

1. **A. Kaushik**, J. Thompson and E. Vlachos, "Energy Efficiency Maximization in Millimeter Wave Hybrid MIMO Systems for 5G and Beyond," *IEEE International Conference on Communications and Networking*, Hammamet, Tunisia, pp. 1-7, Mar. 2020. [**Invited Paper**]

2. **A. Kaushik**, C. Tsinos, E. Vlachos and J. Thompson, "Energy Efficient Bit Allocation and Hybrid Combining for Millimeter Wave MIMO Systems," *IEEE Global Communications Conference*, HI, USA, pp. 1-6, Dec. 2019.

3. **A. Kaushik**, E. Vlachos and J. Thompson, "Energy Efficiency Maximization of Millimeter Wave Hybrid MIMO Systems with Low Resolution DACs," *IEEE International Conference on Communications*, Shanghai, China, pp. 1-6, May 2019.

4. **A. Kaushik**, E. Vlachos, J. Thompson and A. Perelli, "Efficient Channel Estimation in Millimeter Wave Hybrid MIMO Systems with Low Resolution ADCs," *IEEE European Signal Processing Conference*, Rome, Italy, pp. 1825-1829, Sept. 2018.

5. **A. Kaushik**, J. Thompson and M. Yaghoobi, "Sparse Hybrid Precoding and Combining in Millimeter Wave MIMO Systems," *IET Radio Propagation and Technologies*, Durham, UK, pp. 1-7, Oct. 2016.

Co-Author

6. Y. Hu, R. Liu, X. Yue, **A. Kaushik** and M. Kadoch, “Performance Analysis of Rateless-Coded Non-Orthogonal Multiple Access over Nakagami-m Fading Channels with Delay Constrains,” *IEEE International Conference on Communications*, Dublin, Ireland, pp. 1-6, June 2020.

7. Y. Hu, R. Liu, **A. Kaushik**, J. Thompson and X. Yue, “Performance Analysis of Rateless-Coded Non-Orthogonal Multiple Access,” *IEEE International Wireless Communications and Mobile Computing Conference*, Tangier, Morocco, pp. 1-6, June 2019.

8. H. Bian, R. Liu, **A. Kaushik** and R. Duan, “Segmented CRC-Aided Spinal Codes with a Novel Sliding Window Decoding Algorithm,” *IEEE International Wireless Communications and Mobile Computing Conference*, Tangier, Morocco, pp. 1-5, June 2019.

9. Y. Hu, R. Liu, **A. Kaushik**, X. Shi and J. Thompson, “A Low Complexity Decoding Algorithm for Spinal Codes with Efficiently Distributed Symbols,” *NASA/ESA Conference on Adaptive Hardware and Systems*, Edinburgh, UK, pp. 184-191, Aug. 2018.

10. E. Vlachos, **A. Kaushik** and J. Thompson, “Energy Efficient Transmitter with Low Resolution DACs for Massive MIMO with Partially Connected Hybrid Architecture,” *IEEE Vehicular Technology Conference-Spring*, Porto, Portugal, pp. 1-5, June 2018. [**Invited Paper**]

Appendix B

Attached Publications

This chapter contains all work as the **first author**, either published or submitted for publication to peer-reviewed journals and conferences, i.e., A.1.1-2 & A.2.1-5.

Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid Beamforming in Millimeter Wave MIMO Systems

Aryan Kaushik, John Thompson, *Fellow, IEEE*, Evangelos Vlachos, *Member, IEEE*, Christos Tsinos, *Member, IEEE*, and Symeon Chatzinotas, *Senior Member, IEEE*

***Abstract*—This paper proposes a novel architecture with a framework that dynamically activates the optimal number of radio frequency (RF) chains used to implement hybrid beamforming in a millimeter wave (mmWave) multiple-input and multiple-output (MIMO) system. We use fractional programming to solve an energy efficiency maximization problem and exploit the Dinkelbach method (DM)-based framework to optimize the number of active RF chains and data streams. This solution is updated dynamically based on the current channel conditions, where the analog/digital (A/D) hybrid precoder and combiner matrices at the transmitter and the receiver, respectively, are designed using a codebook-based fast approximation solution called gradient pursuit (GP). The GP algorithm shows less run time and complexity while compared to the state-of-the-art orthogonal matching pursuit (OMP) solution. The energy and spectral efficiency performance of the proposed framework is compared with the existing state-of-the-art solutions, such as the brute force (BF), the digital beamformer, and the analog beamformer. The codebook-free approaches to design the precoders and combiners, such as alternating direction method of multipliers (ADMMs) and singular value decomposition (SVD)-based solution are also shown to be incorporated into the proposed framework to achieve better energy efficiency performance.**

***Index Terms*—RF chain selection, energy efficiency optimization, low complexity, hybrid precoding and combining, millimeter wave MIMO, 5G wireless.**

I. Introduction

THE EMERGING advanced consumer devices and developed communication systems have resulted in ever-increasing demands on bandwidth and capacity. For instance, Cisco's annual report suggests that mobile video traffic is expected to generate 74% of the global mobile data traffic by 2020 [1]. The microwave frequency spectrum at sub-6 GHz frequencies, which we currently make use of for mobile broadband, is limited to a very crowded frequency range enhancing the demand for an unused available spectrum which can be resolved by the use of millimeter wave (mmWave) frequency spectrum [2], [3]. The use of mmWave frequency bands appears to be a promising technology to meet the needs of fifth generation (5G) wireless communication systems such as increased capacity, high data rates, improved coverage, lower latency, high mobility, high reliability and lower infrastructure costs [4]–[6]. A few existing applications of the mmWave spectrum are in satellite communications, wireless backhaul, radio applications and radar communication. However, mmWave faces challenges of severe path loss, blocking effects, new hardware constraints and unconventional channel characteristics.

The high bandwidths for mmWave communication compared to sub-6 GHz frequency bands must be traded off against increased path loss [7], which can be compensated using large-scale antenna arrays [8], [9]. The large number of antenna elements and the high bandwidth makes it hard to use a separate radio frequency (RF) chain for each antenna due to the large requirements in power consumption and hardware complexity [8]. A conventional fully digital beamforming architecture used for sub-6 GHz frequencies requires a dedicated RF chain per antenna with the electronic components such as digital-to-analog converters (DACs) and analog-to-digital converters (ADCs) that enhances the hardware complexity and power consumption with the increase in antenna size [8], [9]. Thus, a digital beamforming architecture seems currently impractical to be implemented for large scale antenna arrays in the mmWave band.

As an alternative, an analog beamforming approach could be considered to solve this problem. The analog beamforming architecture involves a network of analog phase shifters with a single RF chain in the system [10], [11], which is highly advantageous to reduce hardware complexity and power consumption. But analog only beamforming approach cannot support multi-stream communication and the capacity performance is usually worse than the fully digital one. Furthermore, the support of multi-user communications is very difficult.

The performance of the mmWave multiple-input multiple-output (MIMO) systems can be significantly improved through

Manuscript received July 16, 2018; revised December 26, 2018, April 5, 2019, and July 23, 2019; accepted July 23, 2019. Date of publication July 29, 2019; date of current version November 20, 2019. This work was supported in part by the Engineering and Physical Sciences Research Council under Grant EP/P000703/1 and Grant EP/L026147/1, and in part by ECLECTIC project under FNR CORE Framework. The associate editor coordinating the review of this paper and approving it for publication was R. Vaze. *(Corresponding author: Aryan Kaushik.)*

A. Kaushik, J. Thompson, and E. Vlachos are with the Institute for Digital Communications, University of Edinburgh, Edinburgh EH9 3JL, U.K. (e-mail: a.kaushik@ed.ac.uk; j.s.thompson@ed.ac.uk; e.vlachos@ed.ac.uk).

C. Tsinos and S. Chatzinotas are with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 1511 Luxembourg, Luxembourg (e-mail: christos.tsinos@uni.lu; symeon.chatzinotas@uni.lu).

Digital Object Identifier 10.1109/TGCN.2019.2931613

the use of analog/digital (A/D) hybrid beamforming architectures where the number of RF chains and associated ADCs/DACs are much less than the number of antennas [12], [13]. The A/D hybrid beamforming also enables spatial multiplexing and multi-user MIMO communication, and A/D hybrid transceiver solutions have recently been proposed to enable mmWave MIMO systems [14]–[16]. The A/D hybrid beamforming system can be implemented to provide satisfying rate performance by avoiding the discussed limitations of a fully digital solution [14]–[16]. One should note that we can reduce the power consumption by implementing low resolution quantization for both conventional and A/D hybrid beamforming architectures. To that end some approaches have been applied for energy efficiency maximization such as in [17]. Optimizing the number of RF chains further leverages the energy efficiency metric and reduces the gap between the spectral efficiency of A/D hybrid and fully digital beamforming architectures. Reference [18] suggests that the A/D hybrid beamforming architecture with low resolution DACs along with optimizing the number of RF chains shows better energy efficiency performance than the conventional digital beamforming architecture for 1-bit and 3-bits sampling resolutions.

To implement the A/D hybrid beamforming system which uses RF precoders based on the phase shifting networks, we can use the most popular structures such as the fully-connected and the partially-connected. The fully-connected structure connects all the antennas to each RF chain whereas the partially-connected structure connects only a subset of the antennas requiring less number of phase shifters [19]. The use of a partially-connected structure at the transceiver can further reduce the power consumption [16], for instance, our previous work [18] uses a partially-connected structure to evaluate the energy and rate performance where the partially-connected structure is opted to achieve high energy efficiency. This paper mainly uses the fully-connected structure to demonstrate the contributions of the proposed framework for a mmWave MIMO system. However, the energy efficiency performance using the partially-connected structure is also observed via simulations. We can observe from recent literature that there are works considering the energy efficient design of a A/D hybrid transceiver, however there is lack of works that optimize the number of RF chains which we discuss in the following subsection.

A. Literature Review

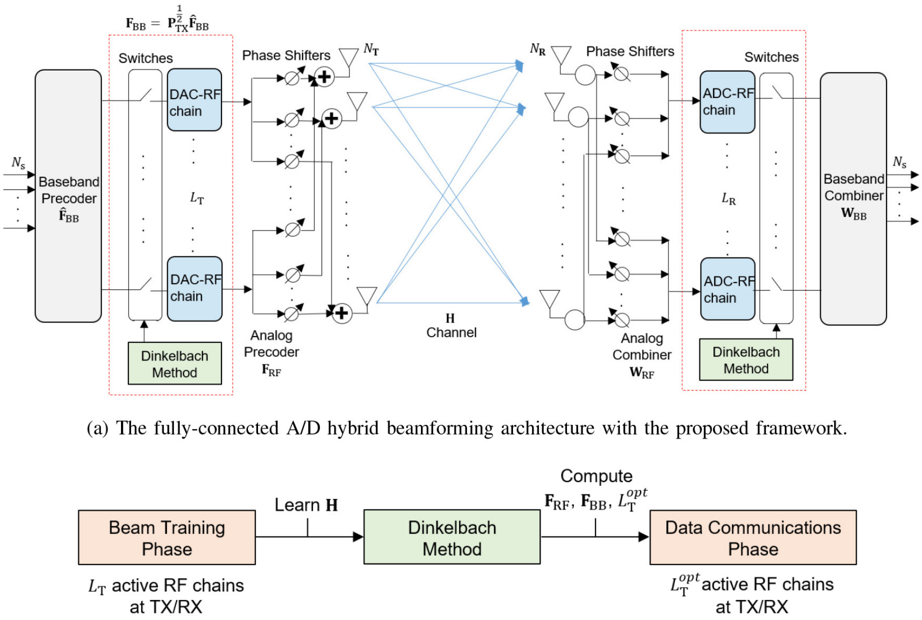
Reference [15] proposes a spectrally efficient A/D hybrid precoder design by maximizing the desired rate for fully-connected limited RF chain systems. However, it does not consider the energy consumption. For an energy efficient system, [20] considers a sub-connected architecture, where each RF chain is connected to only a subset of transmitter (TX) antennas requiring fewer phase shifters, but it does not discuss how to design an energy efficient precoder with a fully-connected architecture. Reference [19] considers both fully-connected and partially-connected structures to design a A/D hybrid precoder where the partially-connected structure seems to outperform fully-connected structure in terms of energy efficiency. However, it only considers the design of the precoder matrices and there is no emphasis on optimizing the number of RF chains which is a key factor for an energy efficient system.

The RF chains consume a large amount of power in wireless communication systems and increase the cost for these systems [21]. Reference [22] performs an energy efficient optimization to design a A/D hybrid precoder where to calculate the optimal number of RF chains, the full precoding solution is computed for all possible numbers of RF chains. This is referred to as the brute force (BF) technique throughout in this paper. References [15] and [22] use orthogonal matching pursuit (OMP) to optimize the precoder matrices. Alternative greedy strategies to OMP can be exploited to lower the complexity. A mmWave A/D hybrid MIMO system can be used for 5G mmWave MIMO applications such as cellular backhaul connections when we jointly optimize the number of RF chains and the A/D hybrid precoder and combiner matrices leading to a highly energy efficient system.

B. Contributions

This paper proposes an energy efficient A/D hybrid beamforming framework, where the RF precoder and baseband precoder matrices, and RF combiner and baseband combiner matrices are optimized along with the number of active RF chains but with low complexity. We use power allocation, and Dinkelbach method (DM) is implemented to optimize the number of RF chains. Fig. 1 shows the novel architecture with proposed framework for a mmWave single user fully-connected A/D hybrid beamforming MIMO system with digital baseband precoding and associated switches, followed by RF chains and associated DACs, and constrained RF precoding implemented using phase shifters network at the TX, and vice-versa at the receiver (RX). In this novel architecture, for a certain number of RF chains implemented in the hardware, the DM block drives digital switches to activate only those RF chains that we obtain as an optimal solution from the proposed method. In practice the digital switches would be a part of the digital processor. If the DM block is replaced by another method used to optimize the number of RF chains, the number of active RF chains and associated DACs/ADCs may be different.

To compute the A/D hybrid precoders and combiners, the proposed approach incorporates a codebook-based approach through one of the greedy strategies, i.e., gradient pursuit (GP) [23]. Simulations show that the proposed GP-based approach is a faster and less complex approach to compute the precoder and combiner matrices than the state of the art OMP. Furthermore, the proposed framework can also be incorporated with the existing codebook-free solutions such as alternating direction method of multipliers (ADMM) [16] and singular value decomposition (SVD) based solution [12]. The objective is to achieve better energy efficiency performance for codebook-free approaches over the fixed number of RF chains case. The proposed energy efficient and low complexity A/D


$\mathbf{F}_{\mathrm{BB}} = \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\mathrm{BB}}$
Switches
Phase Shifters
N_{T}
N_{R}
Phase Shifters
Switches
DAC-RF chain
DAC-RF chain
L_{T}
N_s
Baseband Precoder $\hat{\mathbf{F}}_{\mathrm{BB}}$
Dinkelbach Method
Analog Precoder $\mathbf{F}_{\mathrm{RF}}$
$\mathbf{H}$ Channel
ADC-RF chain
ADC-RF chain
L_{R}
Analog Combiner $\mathbf{W}_{\mathrm{RF}}$
Dinkelbach Method
Baseband Combiner $\mathbf{W}_{\mathrm{BB}}$
N_s
Beam Training Phase
Learn $\mathbf{H}$
Dinkelbach Method
Compute $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$, L_{T}^{opt}
Data Communications Phase
L_{T} active RF chains at TX/RX
L_{T}^{opt} active RF chains at TX/RX

(a) The fully-connected A/D hybrid beamforming architecture with the proposed framework.

(b) Block diagram of the beam tracking phase and the data communications phase.

Fig. 1. System model for a mmWave A/D hybrid MIMO system with the proposed framework.

hybrid precoder framework with a fully-connected architecture can be used in designing 5G mmWave MIMO systems effectively and efficiently, such as in 5G cellular systems and wireless backhaul networks.

The main contributions of this paper can be summarized as follows:

1) The paper proposes a novel algorithmic framework, where the number of active RF chains is dynamically adapted on a frame-by-frame basis. This is carried out using a low complexity alternative to brute force optimization [22] based on the current channel conditions measured in the A/D hybrid beamforming architecture.
2) We develop a reduced complexity DM based solution to find the optimal number of RF chains and streams for the mmWave MIMO system for the current channel conditions.
3) A GP-based approach is proposed as a lower complexity approximation solution to compute the precoder and combiner matrices than the state of the art OMP solution.

Outline: Section II describes the channel and system model implemented for the novel architecture. Section III discusses the low complexity design of the A/D hybrid precoder and combiner matrices using GP algorithm. Section IV provides the energy efficiency maximization problem and we solve the optimization problem via the DM based solution used in the framework where Section IV-A discusses the energy efficiency computation, while Section IV-B describes the energy efficient and low complexity solution to optimize the number of RF chains and activate that many RF chains in the system (as shown in Fig. 1). Section V provides the simulation results. The conclusions are provided in Section VI.

Notations: $\mathbf{A}$, $\mathbf{a}$ and a stand for a matrix, a vector and a scalar, respectively; $\mathbf{A}^{(i)}$ represents the i^{th} column of $\mathbf{A}$; transpose, complex conjugate transpose and pseudo inverse of $\mathbf{A}$ are denoted as $\mathbf{A}^T$, $\mathbf{A}^H$ and $\mathbf{A}^{\dagger}$, respectively; $\|\mathbf{A}\|_F$, tr($\mathbf{A}$) and $|\mathbf{A}|$ represent the Frobenius norm, trace, and determinant of $\mathbf{A}$, respectively; $\|\mathbf{a}\|_p$ is the p-norm of $\mathbf{a}$; $[\mathbf{A}|\mathbf{B}]$ denotes horizontal concatenation; $x \cup y$ denotes the union of x and y union disjoint sets; $\mathbf{A}|_{\Gamma}$ denotes a matrix consisting of rows of matrix $\mathbf{A}$ with indices from Γ set; diag($\mathbf{A}$) generates a vector by the diagonal elements of $\mathbf{A}$; $\mathbf{I}_N$ and $\mathbf{0}_{X\times Y}$ represent $N \times N$ identity matrix and $X \times Y$ all-zeros matrix, respectively; $\mathcal{CN}(\mathbf{a};\mathbf{A})$ denotes a vector of complex Gaussian random variables with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$, and i.i.d. shows that the entries of a vector of random variables are independent and identically distributed. $\mathbf{X} \in \mathbb{C}^{A\times B}$ and $\mathbf{X} \in \mathbb{R}^{A\times B}$ denote $A \times$ B size $\mathbf{X}$ matrix with complex and real entries, respectively; the expectation operator and the real part of a complex variable are denoted as $\mathbb{E}\{\cdot\}$ and $\mathbb{R}\{\cdot\}$, respectively.

II. MmWave A/D Hybrid MIMO Model

A. MmWave Channel Model

Let us consider a single user MIMO system with N_{T} antennas at the TX, sending N_{s} data streams to a system with N_{R} RX antennas. The fading channel models used in traditional MIMO becomes inaccurate for mmWave channel modeling due to the high free-space path loss and large tightly-packed antenna arrays. The existing literature mostly addresses the narrowband clustered channel model [24], [25] for mmWave propagation due to different channel settings such as number of multipaths, amplitudes, etc. such as in [8], [15].

For N_{cl} clusters and N_{ray} propagation paths in each cluster and for a uniform linear array (ULA) antenna elements, the mmWave channel matrix is defined as follows:

$$\mathbf{H} = \sqrt{\frac{N_{\text{T}}N_{\text{R}}}{N_{\text{cl}}N_{\text{ray}}}} \sum_{i=1}^{N_{\text{cl}}} \sum_{l=1}^{N_{\text{ray}}} \alpha_{il}\mathbf{a}_{\text{R}}(\phi_{il}^{r})\mathbf{a}_{\text{T}}\left(\phi_{il}^{t}\right)^{H}, \tag{1}$$

where α_{il} denotes the gain of l-th ray in i-th cluster and it is assumed that α_{il} are i.i.d. $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$, where $\sigma_{\alpha,i}^2$ is average power of the i-th cluster such that $\sum_{i=1}^{N_{\text{cl}}} \sigma_{\alpha,i}^2 = \gamma$, where $\gamma = \sqrt{\frac{N_{\text{T}}N_{\text{R}}}{N_{\text{cl}}N_{\text{ray}}}}$, is the normalization factor satisfying $\mathbb{E}\{\|\mathbf{H}\|_F^2\} = 1/\sqrt{N_{\text{cl}}N_{\text{ray}}}$. Further, $\mathbf{a}_{\text{R}}(\phi_{il}^{r})$ and $\mathbf{a}_{\text{T}}(\phi_{il}^{t})$ represent the normalized receive and transmit array response vectors, where ϕ_{il}^{t} and ϕ_{il}^{r} are the azimuth angles of departure and arrival, respectively. The antenna elements at the TX and the RX can be modeled as ideal sectored elements [26] and then antenna element gains can be evaluated over ideal sectors. In (1), the transmit and receive antenna element gains are considered unity over ideal sectors defined by $\phi_{il}^{t} \in [\phi_{\min}^{t}, \phi_{\max}^{t}]$ and $\phi_{il}^{r} \in [\phi_{\min}^{r}, \phi_{\max}^{r}]$, respectively. For a N_{Z}-element ULA on Z-axis, the array response vector can be expressed as [27]: $\mathbf{a}_{\text{Z}}(\phi) = \frac{1}{\sqrt{N_{\text{Z}}}} e^{jm\frac{2\pi}{\lambda}d\sin(\phi)}{}^{T}$, where $0 \leq m \leq (N_{\text{Z}} - 1)$ is a real integer, d is the inter-element spacing in wavelengths and λ is the signal wavelength. The array response vectors can also be computed using other array geometries such as rectangular array and circular array. As mentioned above, we assume perfect channel knowledge at the TX and the RX [15], [16], [22]. However, this work can also be extended to consider channel estimation errors, for example, [28] proposes an efficient channel estimation algorithm for hybrid architecture mmWave systems.

The beamspace representation [29], [30] of the narrowband channel can be written as follows:

$$\mathbf{H} = \mathbf{D}_{\text{R}}\mathbf{H}_{\text{v}}\mathbf{D}_{\text{T}}^{H}, \tag{2}$$

where $\mathbf{H}_{\text{v}} \in \mathbb{C}^{L_{\text{R}} \times L_{\text{T}}}$ represents a sparse matrix with a few non-zero entries, while $\mathbf{D}_{\text{R}} \in \mathbb{C}^{N_{\text{R}} \times L_{\text{R}}}$ and $\mathbf{D}_{\text{T}} \in \mathbb{C}^{N_{\text{T}} \times L_{\text{T}}}$ are the discrete Fourier transform (DFT) matrices.

B. A/D Hybrid MIMO System Model

In large-scale MIMO communication systems, based on the A/D hybrid precoding scheme, the number of RF chains is larger than or equal to the number of baseband data streams and smaller than or equal to the number of TX antennas. L_{T} denotes the number of available RF chains at the TX with the limitation that $N_{\text{s}} \leq L_{\text{T}} \leq N_{\text{T}}$ and similarly L_{R} is for the RX with the condition $N_{\text{s}} \leq L_{\text{R}} \leq N_{\text{R}}$. We consider the number of RF chains at the RX to be same as at the TX, i.e., $L_{\text{R}} = L_{\text{T}}$.

Let $\mathbf{F}_{\text{BB}} = \mathbf{P}_{\text{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\text{BB}}$ denote the baseband precoder matrix which inputs to the DAC-RF chain block where $\mathbf{P}_{\text{TX}} \in \mathbb{R}^{L_{\text{T}} \times L_{\text{T}}}$ is a diagonal matrix of power allocation values with $\text{tr}(\mathbf{P}_{\text{TX}}) = P_{\max}$, $\hat{\mathbf{F}}_{\text{BB}}$ is the digital precoding matrix before the switches, and $\mathbf{F}_{\text{RF}}$ denotes the RF precoder matrix. $\mathbf{F}_{\text{BB}}$ has dimensions of $L_{\text{T}} \times N_{\text{s}}$ using its L_{T} transmit chains and $\mathbf{F}_{\text{RF}}$ has dimensions of $N_{\text{T}} \times L_{\text{T}}$ using the phase shifting network. Similarly at the RX, the matrices $\mathbf{W}_{\text{BB}}$ and $\mathbf{W}_{\text{RF}}$ denote the $L_{\text{R}} \times N_{\text{s}}$ baseband combiner and the $N_{\text{R}} \times L_{\text{R}}$ RF combiner, respectively. The TX symbol vector $\mathbf{s} \in \mathbb{C}^{N_{\text{s}} \times 1}$ is such that $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \frac{1}{N_{\text{s}}}\mathbf{I}_{N_{\text{s}}}$. All elements of $\mathbf{F}_{\text{RF}}$ and $\mathbf{W}_{\text{RF}}$ are of constant modulus. The power constraint at the TX is satisfied by $\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 = P_{\max}$, where $P_{\max}$ is the maximum allocated power. We assume a unit magnitude and continuous phase constraint on the phase shifters [15].

Consider a narrowband propagation channel with $\mathbf{H}$ as the $N_{\text{R}} \times N_{\text{T}}$ channel matrix, which is assumed to be known to both the TX and the RX, then the received signal can be expressed as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{n}, \tag{3}$$

where $\mathbf{y}$ is the $N_{\text{R}} \times 1$ received vector and $\mathbf{n}$ is a $N_{\text{R}} \times 1$ noise vector with entries which are modeled as i.i.d. $\mathcal{CN}(0, \sigma_{\text{n}}^2)$. After the application of the combining matrices, the received signal can be written as follows:

$$\tilde{\mathbf{y}} = \mathbf{W}_{\text{BB}}^{H}\mathbf{W}_{\text{RF}}^{H}\mathbf{y} = \mathbf{W}_{\text{BB}}^{H}\mathbf{W}_{\text{RF}}^{H}\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{W}_{\text{BB}}^{H}\mathbf{W}_{\text{RF}}^{H}\mathbf{n}. \tag{4}$$

In the following section, we discuss the low complexity designs of A/D hybrid precoders, i.e., $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$, and A/D hybrid combiners, i.e., $\mathbf{W}_{\text{RF}}, \mathbf{W}_{\text{BB}}$.

III. Low Complexity A/D Hybrid Precoders and Combiners Design

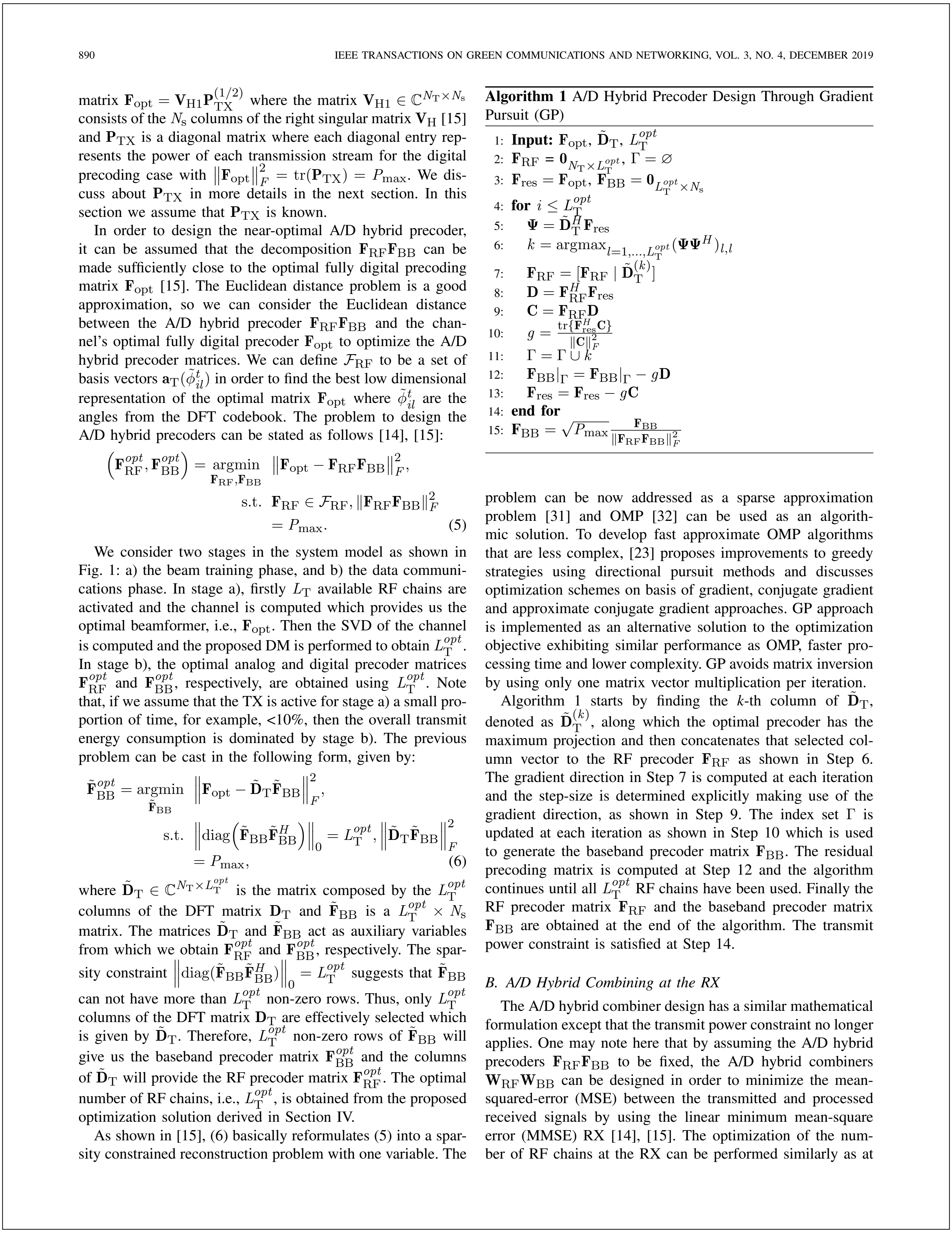
The combined problem of designing the precoders and combiners and the number of RF chains can be partitioned into three sub-problems:

- to optimize the A/D hybrid precoders $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$,
- to optimize the A/D hybrid combiners $\mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}$ and
- to optimize the number of RF chains, i.e., obtaining L_{T}^{opt} at the TX and L_{R}^{opt} at the RX.

Firstly in this section, we focus on designing the A/D hybrid precoder matrices $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$ as shown in Section III-A and the hybrid combiner matrices $\mathbf{W}_{\text{RF}}$ and $\mathbf{W}_{\text{BB}}$ as shown in Section III-B by assuming that L_{T}^{opt} and L_{R}^{opt} are computed from the proposed DM based solution in Section IV already. In the next section, we propose the DM based solution for optimizing the number of RF chains at the TX and consider that $L_{\text{R}}^{opt} = L_{\text{T}}^{opt}$.

A. A/D Hybrid Precoding at the TX

It is known that the precoding matrix for the digital beamformer is given based on the singular value decomposition (SVD) of the channel matrix. We consider channel's SVD as $\mathbf{H} = \mathbf{U}_{\text{H}}\mathbf{\Sigma}_{\text{H}}\mathbf{V}_{\text{H}}^{H}$, where $\mathbf{U}_{\text{H}} \in \mathbb{C}^{N_{\text{R}} \times N_{\text{R}}}$ and $\mathbf{V}_{\text{H}} \in \mathbb{C}^{N_{\text{T}} \times N_{\text{T}}}$ are unitary matrices, and $\mathbf{\Sigma}_{\text{H}} \in \mathbb{R}^{N_{\text{R}} \times N_{\text{T}}}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. The optimal fully digital precoding



matrix $\mathbf{F}_{\rm opt} = \mathbf{V}_{\rm H1}\mathbf{P}_{\rm TX}^{(1/2)}$ where the matrix $\mathbf{V}_{\rm H1} \in \mathbb{C}^{N_{\rm T}\times N_{\rm s}}$ consists of the $N_{\rm s}$ columns of the right singular matrix $\mathbf{V}_{\rm H}$ [15] and $\mathbf{P}_{\rm TX}$ is a diagonal matrix where each diagonal entry represents the power of each transmission stream for the digital precoding case with $\|\mathbf{F}_{\rm opt}\|_F^2 = {\rm tr}(\mathbf{P}_{\rm TX}) = P_{\rm max}$. We discuss about $\mathbf{P}_{\rm TX}$ in more details in the next section. In this section we assume that $\mathbf{P}_{\rm TX}$ is known.

In order to design the near-optimal A/D hybrid precoder, it can be assumed that the decomposition $\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}$ can be made sufficiently close to the optimal fully digital precoding matrix $\mathbf{F}_{\rm opt}$ [15]. The Euclidean distance problem is a good approximation, so we can consider the Euclidean distance between the A/D hybrid precoder $\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}$ and the channel's optimal fully digital precoder $\mathbf{F}_{\rm opt}$ to optimize the A/D hybrid precoder matrices. We can define $\mathcal{F}_{\rm RF}$ to be a set of basis vectors $\mathbf{a}_{\rm T}(\tilde{\phi}_{il}^t)$ in order to find the best low dimensional representation of the optimal matrix $\mathbf{F}_{\rm opt}$ where $\tilde{\phi}_{il}^t$ are the angles from the DFT codebook. The problem to design the A/D hybrid precoders can be stated as follows [14], [15]:

$$\left(\mathbf{F}_{\rm RF}^{opt}, \mathbf{F}_{\rm BB}^{opt}\right) = \underset{\mathbf{F}_{\rm RF},\mathbf{F}_{\rm BB}}{\rm argmin} \ \left\|\mathbf{F}_{\rm opt} - \mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\right\|_F^2, \quad \text{s.t.} \ \mathbf{F}_{\rm RF} \in \mathcal{F}_{\rm RF}, \|\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\|_F^2 = P_{\rm max}. \tag{5}$$

We consider two stages in the system model as shown in Fig. 1: a) the beam training phase, and b) the data communications phase. In stage a), firstly $L_{\rm T}$ available RF chains are activated and the channel is computed which provides us the optimal beamformer, i.e., $\mathbf{F}_{\rm opt}$. Then the SVD of the channel is computed and the proposed DM is performed to obtain $L_{\rm T}^{opt}$. In stage b), the optimal analog and digital precoder matrices $\mathbf{F}_{\rm RF}^{opt}$ and $\mathbf{F}_{\rm BB}^{opt}$, respectively, are obtained using $L_{\rm T}^{opt}$. Note that, if we assume that the TX is active for stage a) a small proportion of time, for example, <10%, then the overall transmit energy consumption is dominated by stage b). The previous problem can be cast in the following form, given by:

$$\tilde{\mathbf{F}}_{\rm BB}^{opt} = \underset{\tilde{\mathbf{F}}_{\rm BB}}{\rm argmin} \ \left\|\mathbf{F}_{\rm opt} - \tilde{\mathbf{D}}_{\rm T}\tilde{\mathbf{F}}_{\rm BB}\right\|_F^2, \quad \text{s.t.} \ \left\|{\rm diag}\left(\tilde{\mathbf{F}}_{\rm BB}\tilde{\mathbf{F}}_{\rm BB}^H\right)\right\|_0 = L_{\rm T}^{opt}, \left\|\tilde{\mathbf{D}}_{\rm T}\tilde{\mathbf{F}}_{\rm BB}\right\|_F^2 = P_{\rm max}, \tag{6}$$

where $\tilde{\mathbf{D}}_{\rm T} \in \mathbb{C}^{N_{\rm T}\times L_{\rm T}^{opt}}$ is the matrix composed by the $L_{\rm T}^{opt}$ columns of the DFT matrix $\mathbf{D}_{\rm T}$ and $\tilde{\mathbf{F}}_{\rm BB}$ is a $L_{\rm T}^{opt} \times N_{\rm s}$ matrix. The matrices $\tilde{\mathbf{D}}_{\rm T}$ and $\tilde{\mathbf{F}}_{\rm BB}$ act as auxiliary variables from which we obtain $\mathbf{F}_{\rm RF}^{opt}$ and $\mathbf{F}_{\rm BB}^{opt}$, respectively. The sparsity constraint $\left\|{\rm diag}(\tilde{\mathbf{F}}_{\rm BB}\tilde{\mathbf{F}}_{\rm BB}^H)\right\|_0 = L_{\rm T}^{opt}$ suggests that $\tilde{\mathbf{F}}_{\rm BB}$ can not have more than $L_{\rm T}^{opt}$ non-zero rows. Thus, only $L_{\rm T}^{opt}$ columns of the DFT matrix $\mathbf{D}_{\rm T}$ are effectively selected which is given by $\tilde{\mathbf{D}}_{\rm T}$. Therefore, $L_{\rm T}^{opt}$ non-zero rows of $\tilde{\mathbf{F}}_{\rm BB}$ will give us the baseband precoder matrix $\mathbf{F}_{\rm BB}^{opt}$ and the columns of $\tilde{\mathbf{D}}_{\rm T}$ will provide the RF precoder matrix $\mathbf{F}_{\rm RF}^{opt}$. The optimal number of RF chains, i.e., $L_{\rm T}^{opt}$, is obtained from the proposed optimization solution derived in Section IV.

As shown in [15], (6) basically reformulates (5) into a sparsity constrained reconstruction problem with one variable. The problem can be now addressed as a sparse approximation problem [31] and OMP [32] can be used as an algorithmic solution. To develop fast approximate OMP algorithms that are less complex, [23] proposes improvements to greedy strategies using directional pursuit methods and discusses optimization schemes on basis of gradient, conjugate gradient and approximate conjugate gradient approaches. GP approach is implemented as an alternative solution to the optimization objective exhibiting similar performance as OMP, faster processing time and lower complexity. GP avoids matrix inversion by using only one matrix vector multiplication per iteration.

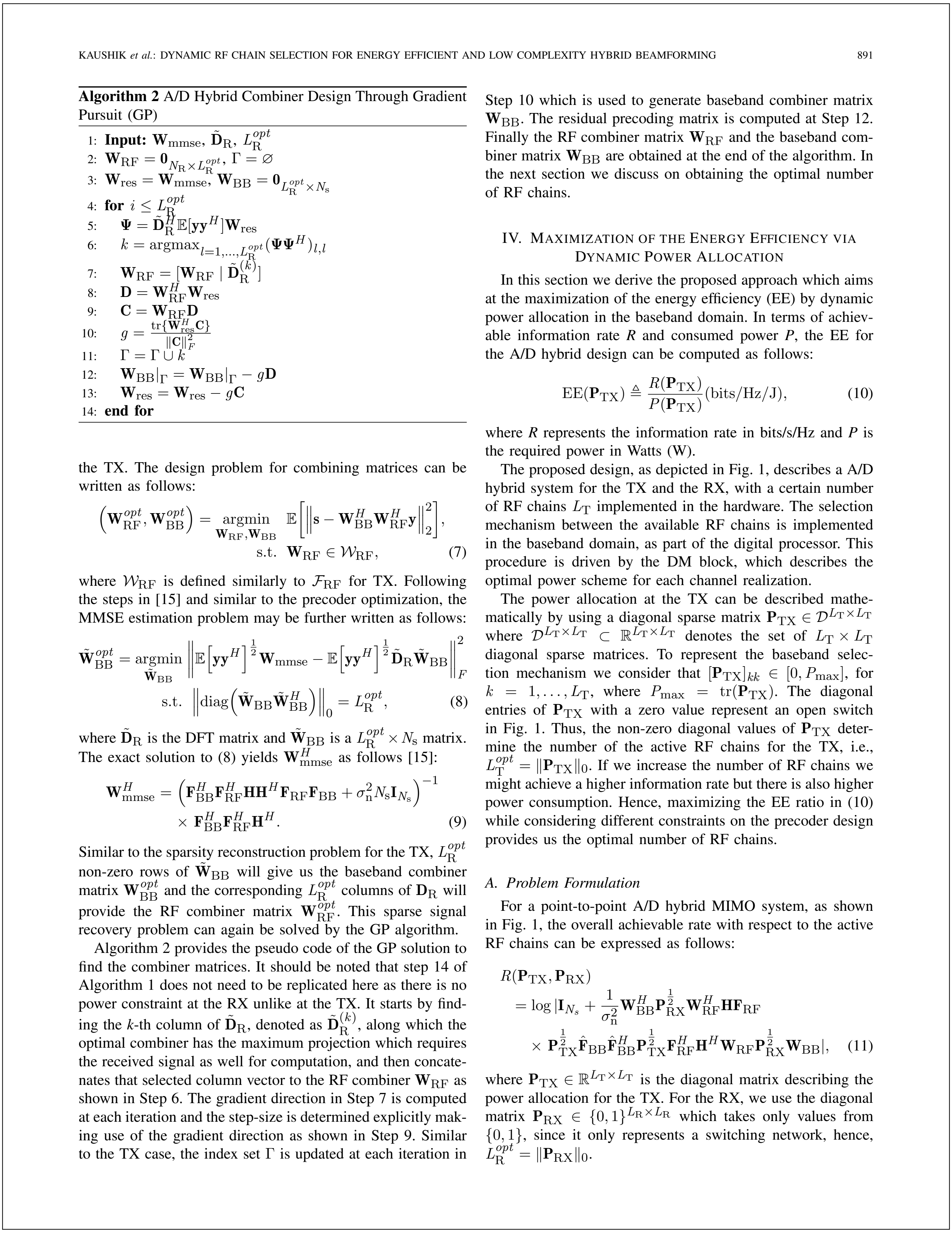
Algorithm 1 A/D Hybrid Precoder Design Through Gradient Pursuit (GP)

```
1:  Input: $\mathbf{F}_{\rm opt}$, $\tilde{\mathbf{D}}_{\rm T}$, $L_{\rm T}^{opt}$
2:  $\mathbf{F}_{\rm RF} = \mathbf{0}_{N_{\rm T}\times L_{\rm T}^{opt}}$, $\Gamma = \varnothing$
3:  $\mathbf{F}_{\rm res} = \mathbf{F}_{\rm opt}$, $\mathbf{F}_{\rm BB} = \mathbf{0}_{L_{\rm T}^{opt}\times N_{\rm s}}$
4:  for $i \le L_{\rm T}^{opt}$
5:      $\mathbf{\Psi} = \tilde{\mathbf{D}}_{\rm T}^H\mathbf{F}_{\rm res}$
6:      $k = {\rm argmax}_{l=1,\dots,L_{\rm T}^{opt}}(\mathbf{\Psi}\mathbf{\Psi}^H)_{l,l}$
7:      $\mathbf{F}_{\rm RF} = [\mathbf{F}_{\rm RF} \mid \tilde{\mathbf{D}}_{\rm T}^{(k)}]$
8:      $\mathbf{D} = \mathbf{F}_{\rm RF}^H\mathbf{F}_{\rm res}$
9:      $\mathbf{C} = \mathbf{F}_{\rm RF}\mathbf{D}$
10:     $g = \frac{{\rm tr}\{\mathbf{F}_{\rm res}^H\mathbf{C}\}}{\|\mathbf{C}\|_F^2}$
11:     $\Gamma = \Gamma \cup k$
12:     $\mathbf{F}_{\rm BB}|_\Gamma = \mathbf{F}_{\rm BB}|_\Gamma - g\mathbf{D}$
13:     $\mathbf{F}_{\rm res} = \mathbf{F}_{\rm res} - g\mathbf{C}$
14: end for
15: $\mathbf{F}_{\rm BB} = \sqrt{P_{\rm max}}\frac{\mathbf{F}_{\rm BB}}{\|\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\|_F^2}$
```

Algorithm 1 starts by finding the k-th column of $\tilde{\mathbf{D}}_{\rm T}$, denoted as $\tilde{\mathbf{D}}_{\rm T}^{(k)}$, along which the optimal precoder has the maximum projection and then concatenates that selected column vector to the RF precoder $\mathbf{F}_{\rm RF}$ as shown in Step 6. The gradient direction in Step 7 is computed at each iteration and the step-size is determined explicitly making use of the gradient direction, as shown in Step 9. The index set Γ is updated at each iteration as shown in Step 10 which is used to generate the baseband precoder matrix $\mathbf{F}_{\rm BB}$. The residual precoding matrix is computed at Step 12 and the algorithm continues until all $L_{\rm T}^{opt}$ RF chains have been used. Finally the RF precoder matrix $\mathbf{F}_{\rm RF}$ and the baseband precoder matrix $\mathbf{F}_{\rm BB}$ are obtained at the end of the algorithm. The transmit power constraint is satisfied at Step 14.

B. A/D Hybrid Combining at the RX

The A/D hybrid combiner design has a similar mathematical formulation except that the transmit power constraint no longer applies. One may note here that by assuming the A/D hybrid precoders $\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}$ to be fixed, the A/D hybrid combiners $\mathbf{W}_{\rm RF}\mathbf{W}_{\rm BB}$ can be designed in order to minimize the mean-squared-error (MSE) between the transmitted and processed received signals by using the linear minimum mean-square error (MMSE) RX [14], [15]. The optimization of the number of RF chains at the RX can be performed similarly as at



Algorithm 2 A/D Hybrid Combiner Design Through Gradient Pursuit (GP)

```
1: Input: $\mathbf{W}_{\mathrm{mmse}}$, $\tilde{\mathbf{D}}_{\mathrm{R}}$, $L_{\mathrm{R}}^{opt}$
2: $\mathbf{W}_{\mathrm{RF}} = \mathbf{0}_{N_{\mathrm{R}} \times L_{\mathrm{R}}^{opt}}$, $\Gamma = \varnothing$
3: $\mathbf{W}_{\mathrm{res}} = \mathbf{W}_{\mathrm{mmse}}$, $\mathbf{W}_{\mathrm{BB}} = \mathbf{0}_{L_{\mathrm{R}}^{opt} \times N_{\mathrm{s}}}$
4: for $i \leq L_{\mathrm{R}}^{opt}$
5:     $\boldsymbol{\Psi} = \tilde{\mathbf{D}}_{\mathrm{R}}^{H} \mathbb{E}[\mathbf{y}\mathbf{y}^{H}] \mathbf{W}_{\mathrm{res}}$
6:     $k = \mathrm{argmax}_{l=1,\ldots,L_{\mathrm{R}}^{opt}} (\boldsymbol{\Psi}\boldsymbol{\Psi}^{H})_{l,l}$
7:     $\mathbf{W}_{\mathrm{RF}} = [\mathbf{W}_{\mathrm{RF}} \mid \tilde{\mathbf{D}}_{\mathrm{R}}^{(k)}]$
8:     $\mathbf{D} = \mathbf{W}_{\mathrm{RF}}^{H} \mathbf{W}_{\mathrm{res}}$
9:     $\mathbf{C} = \mathbf{W}_{\mathrm{RF}} \mathbf{D}$
10:    $g = \frac{\mathrm{tr}\{\mathbf{W}_{\mathrm{res}}^{H}\mathbf{C}\}}{\|\mathbf{C}\|_F^2}$
11:    $\Gamma = \Gamma \cup k$
12:    $\mathbf{W}_{\mathrm{BB}}|_{\Gamma} = \mathbf{W}_{\mathrm{BB}}|_{\Gamma} - g\mathbf{D}$
13:    $\mathbf{W}_{\mathrm{res}} = \mathbf{W}_{\mathrm{res}} - g\mathbf{C}$
14: end for
```

the TX. The design problem for combining matrices can be written as follows:

$$\left(\mathbf{W}_{\mathrm{RF}}^{opt}, \mathbf{W}_{\mathrm{BB}}^{opt}\right) = \underset{\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}}{\operatorname{argmin}} \ \mathbb{E}\left[\left\|\mathbf{s} - \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{y}\right\|_2^2\right], \quad \text{s.t. } \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}, \tag{7}$$

where $\mathcal{W}_{\mathrm{RF}}$ is defined similarly to $\mathcal{F}_{\mathrm{RF}}$ for TX. Following the steps in [15] and similar to the precoder optimization, the MMSE estimation problem may be further written as follows:

$$\tilde{\mathbf{W}}_{\mathrm{BB}}^{opt} = \underset{\tilde{\mathbf{W}}_{\mathrm{BB}}}{\operatorname{argmin}} \left\|\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]^{\frac{1}{2}}\mathbf{W}_{\mathrm{mmse}} - \mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]^{\frac{1}{2}}\tilde{\mathbf{D}}_{\mathrm{R}}\tilde{\mathbf{W}}_{\mathrm{BB}}\right\|_F^2 \quad \text{s.t. } \left\|\mathrm{diag}\left(\tilde{\mathbf{W}}_{\mathrm{BB}}\tilde{\mathbf{W}}_{\mathrm{BB}}^{H}\right)\right\|_0 = L_{\mathrm{R}}^{opt}, \tag{8}$$

where $\tilde{\mathbf{D}}_{\mathrm{R}}$ is the DFT matrix and $\tilde{\mathbf{W}}_{\mathrm{BB}}$ is a $L_{\mathrm{R}}^{opt} \times N_{\mathrm{s}}$ matrix. The exact solution to (8) yields $\mathbf{W}_{\mathrm{mmse}}^{H}$ as follows [15]:

$$\mathbf{W}_{\mathrm{mmse}}^{H} = \left(\mathbf{F}_{\mathrm{BB}}^{H}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{H}^{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} + \sigma_{\mathrm{n}}^{2}N_{\mathrm{s}}\mathbf{I}_{N_{\mathrm{s}}}\right)^{-1} \times \mathbf{F}_{\mathrm{BB}}^{H}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}. \tag{9}$$

Similar to the sparsity reconstruction problem for the TX, L_{R}^{opt} non-zero rows of $\tilde{\mathbf{W}}_{\mathrm{BB}}$ will give us the baseband combiner matrix $\mathbf{W}_{\mathrm{BB}}^{opt}$ and the corresponding L_{R}^{opt} columns of $\mathbf{D}_{\mathrm{R}}$ will provide the RF combiner matrix $\mathbf{W}_{\mathrm{RF}}^{opt}$. This sparse signal recovery problem can again be solved by the GP algorithm.

Algorithm 2 provides the pseudo code of the GP solution to find the combiner matrices. It should be noted that step 14 of Algorithm 1 does not need to be replicated here as there is no power constraint at the RX unlike at the TX. It starts by finding the *k*-th column of $\tilde{\mathbf{D}}_{\mathrm{R}}$, denoted as $\tilde{\mathbf{D}}_{\mathrm{R}}^{(k)}$, along which the optimal combiner has the maximum projection which requires the received signal as well for computation, and then concatenates that selected column vector to the RF combiner $\mathbf{W}_{\mathrm{RF}}$ as shown in Step 6. The gradient direction in Step 7 is computed at each iteration and the step-size is determined explicitly making use of the gradient direction as shown in Step 9. Similar to the TX case, the index set Γ is updated at each iteration in Step 10 which is used to generate baseband combiner matrix $\mathbf{W}_{\mathrm{BB}}$. The residual precoding matrix is computed at Step 12. Finally the RF combiner matrix $\mathbf{W}_{\mathrm{RF}}$ and the baseband combiner matrix $\mathbf{W}_{\mathrm{BB}}$ are obtained at the end of the algorithm. In the next section we discuss on obtaining the optimal number of RF chains.

IV. Maximization of the Energy Efficiency via Dynamic Power Allocation

In this section we derive the proposed approach which aims at the maximization of the energy efficiency (EE) by dynamic power allocation in the baseband domain. In terms of achievable information rate *R* and consumed power *P*, the EE for the A/D hybrid design can be computed as follows:

$$\mathrm{EE}(\mathbf{P}_{\mathrm{TX}}) \triangleq \frac{R(\mathbf{P}_{\mathrm{TX}})}{P(\mathbf{P}_{\mathrm{TX}})} (\text{bits/Hz/J}), \tag{10}$$

where *R* represents the information rate in bits/s/Hz and *P* is the required power in Watts (W).

The proposed design, as depicted in Fig. 1, describes a A/D hybrid system for the TX and the RX, with a certain number of RF chains L_{T} implemented in the hardware. The selection mechanism between the available RF chains is implemented in the baseband domain, as part of the digital processor. This procedure is driven by the DM block, which describes the optimal power scheme for each channel realization.

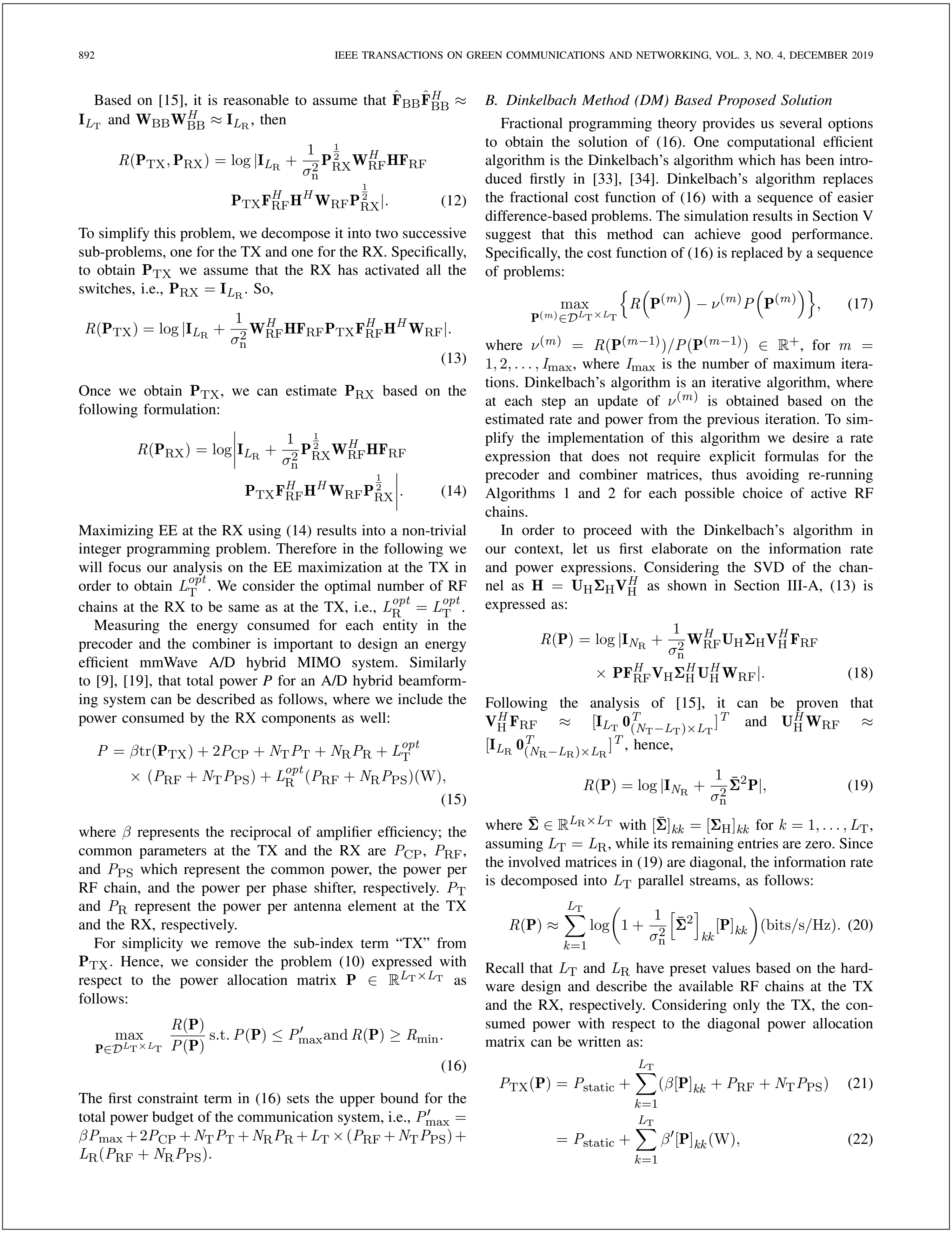
The power allocation at the TX can be described mathematically by using a diagonal sparse matrix $\mathbf{P}_{\mathrm{TX}} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ where $\mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}} \subset \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ denotes the set of $L_{\mathrm{T}} \times L_{\mathrm{T}}$ diagonal sparse matrices. To represent the baseband selection mechanism we consider that $[\mathbf{P}_{\mathrm{TX}}]_{kk} \in [0, P_{\max}]$, for $k = 1, \ldots, L_{\mathrm{T}}$, where $P_{\max} = \mathrm{tr}(\mathbf{P}_{\mathrm{TX}})$. The diagonal entries of $\mathbf{P}_{\mathrm{TX}}$ with a zero value represent an open switch in Fig. 1. Thus, the non-zero diagonal values of $\mathbf{P}_{\mathrm{TX}}$ determine the number of the active RF chains for the TX, i.e., $L_{\mathrm{T}}^{opt} = \|\mathbf{P}_{\mathrm{TX}}\|_0$. If we increase the number of RF chains we might achieve a higher information rate but there is also higher power consumption. Hence, maximizing the EE ratio in (10) while considering different constraints on the precoder design provides us the optimal number of RF chains.

A. Problem Formulation

For a point-to-point A/D hybrid MIMO system, as shown in Fig. 1, the overall achievable rate with respect to the active RF chains can be expressed as follows:

$$\begin{aligned} R(\mathbf{P}_{\mathrm{TX}}, \mathbf{P}_{\mathrm{RX}}) &= \log\left|\mathbf{I}_{N_s} + \frac{1}{\sigma_{\mathrm{n}}^{2}}\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\right. \\ &\quad \left. \times \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\mathrm{BB}}\hat{\mathbf{F}}_{\mathrm{BB}}^{H}\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}\mathbf{W}_{\mathrm{RF}}\mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}}\mathbf{W}_{\mathrm{BB}}\right|, \end{aligned} \tag{11}$$

where $\mathbf{P}_{\mathrm{TX}} \in \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ is the diagonal matrix describing the power allocation for the TX. For the RX, we use the diagonal matrix $\mathbf{P}_{\mathrm{RX}} \in \{0,1\}^{L_{\mathrm{R}} \times L_{\mathrm{R}}}$ which takes only values from $\{0, 1\}$, since it only represents a switching network, hence, $L_{\mathrm{R}}^{opt} = \|\mathbf{P}_{\mathrm{RX}}\|_0$.



Based on [15], it is reasonable to assume that $\hat{\mathbf{F}}_{\mathrm{BB}}\hat{\mathbf{F}}_{\mathrm{BB}}^{H} \approx \mathbf{I}_{L_{\mathrm{T}}}$ and $\mathbf{W}_{\mathrm{BB}}\mathbf{W}_{\mathrm{BB}}^{H} \approx \mathbf{I}_{L_{\mathrm{R}}}$, then

$$R(\mathbf{P}_{\mathrm{TX}}, \mathbf{P}_{\mathrm{RX}}) = \log|\mathbf{I}_{L_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2}\mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{TX}}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}\mathbf{W}_{\mathrm{RF}}\mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}}|. \quad (12)$$

To simplify this problem, we decompose it into two successive sub-problems, one for the TX and one for the RX. Specifically, to obtain $\mathbf{P}_{\mathrm{TX}}$ we assume that the RX has activated all the switches, i.e., $\mathbf{P}_{\mathrm{RX}} = \mathbf{I}_{L_{\mathrm{R}}}$. So,

$$R(\mathbf{P}_{\mathrm{TX}}) = \log|\mathbf{I}_{L_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{TX}}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}\mathbf{W}_{\mathrm{RF}}|. \quad (13)$$

Once we obtain $\mathbf{P}_{\mathrm{TX}}$, we can estimate $\mathbf{P}_{\mathrm{RX}}$ based on the following formulation:

$$R(\mathbf{P}_{\mathrm{RX}}) = \log\left|\mathbf{I}_{L_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2}\mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{TX}}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}\mathbf{W}_{\mathrm{RF}}\mathbf{P}_{\mathrm{RX}}^{\frac{1}{2}}\right|. \quad (14)$$

Maximizing EE at the RX using (14) results into a non-trivial integer programming problem. Therefore in the following we will focus our analysis on the EE maximization at the TX in order to obtain L_{T}^{opt}. We consider the optimal number of RF chains at the RX to be same as at the TX, i.e., $L_{\mathrm{R}}^{opt} = L_{\mathrm{T}}^{opt}$.

Measuring the energy consumed for each entity in the precoder and the combiner is important to design an energy efficient mmWave A/D hybrid MIMO system. Similarly to [9], [19], that total power P for an A/D hybrid beamforming system can be described as follows, where we include the power consumed by the RX components as well:

$$P = \beta\mathrm{tr}(\mathbf{P}_{\mathrm{TX}}) + 2P_{\mathrm{CP}} + N_{\mathrm{T}}P_{\mathrm{T}} + N_{\mathrm{R}}P_{\mathrm{R}} + L_{\mathrm{T}}^{opt} \times (P_{\mathrm{RF}} + N_{\mathrm{T}}P_{\mathrm{PS}}) + L_{\mathrm{R}}^{opt}(P_{\mathrm{RF}} + N_{\mathrm{R}}P_{\mathrm{PS}})(\mathrm{W}), \quad (15)$$

where β represents the reciprocal of amplifier efficiency; the common parameters at the TX and the RX are P_{CP}, P_{RF}, and P_{PS} which represent the common power, the power per RF chain, and the power per phase shifter, respectively. P_{T} and P_{R} represent the power per antenna element at the TX and the RX, respectively.

For simplicity we remove the sub-index term "TX" from $\mathbf{P}_{\mathrm{TX}}$. Hence, we consider the problem (10) expressed with respect to the power allocation matrix $\mathbf{P} \in \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ as follows:

$$\max_{\mathbf{P} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}} \frac{R(\mathbf{P})}{P(\mathbf{P})} \text{ s.t. } P(\mathbf{P}) \leq P'_{\max} \text{and } R(\mathbf{P}) \geq R_{\min}. \quad (16)$$

The first constraint term in (16) sets the upper bound for the total power budget of the communication system, i.e., $P'_{\max} = \beta P_{\max} + 2P_{\mathrm{CP}} + N_{\mathrm{T}}P_{\mathrm{T}} + N_{\mathrm{R}}P_{\mathrm{R}} + L_{\mathrm{T}} \times (P_{\mathrm{RF}} + N_{\mathrm{T}}P_{\mathrm{PS}}) + L_{\mathrm{R}}(P_{\mathrm{RF}} + N_{\mathrm{R}}P_{\mathrm{PS}})$.

B. Dinkelbach Method (DM) Based Proposed Solution

Fractional programming theory provides us several options to obtain the solution of (16). One computational efficient algorithm is the Dinkelbach's algorithm which has been introduced firstly in [33], [34]. Dinkelbach's algorithm replaces the fractional cost function of (16) with a sequence of easier difference-based problems. The simulation results in Section V suggest that this method can achieve good performance. Specifically, the cost function of (16) is replaced by a sequence of problems:

$$\max_{\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}} \left\{R\left(\mathbf{P}^{(m)}\right) - \nu^{(m)}P\left(\mathbf{P}^{(m)}\right)\right\}, \quad (17)$$

where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$, for $m = 1, 2, \ldots, I_{\max}$, where $I_{\max}$ is the number of maximum iterations. Dinkelbach's algorithm is an iterative algorithm, where at each step an update of $\nu^{(m)}$ is obtained based on the estimated rate and power from the previous iteration. To simplify the implementation of this algorithm we desire a rate expression that does not require explicit formulas for the precoder and combiner matrices, thus avoiding re-running Algorithms 1 and 2 for each possible choice of active RF chains.

In order to proceed with the Dinkelbach's algorithm in our context, let us first elaborate on the information rate and power expressions. Considering the SVD of the channel as $\mathbf{H} = \mathbf{U}_{\mathrm{H}}\mathbf{\Sigma}_{\mathrm{H}}\mathbf{V}_{\mathrm{H}}^{H}$ as shown in Section III-A, (13) is expressed as:

$$R(\mathbf{P}) = \log|\mathbf{I}_{N_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{U}_{\mathrm{H}}\mathbf{\Sigma}_{\mathrm{H}}\mathbf{V}_{\mathrm{H}}^{H}\mathbf{F}_{\mathrm{RF}} \times \mathbf{P}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{V}_{\mathrm{H}}\mathbf{\Sigma}_{\mathrm{H}}^{H}\mathbf{U}_{\mathrm{H}}^{H}\mathbf{W}_{\mathrm{RF}}|. \quad (18)$$

Following the analysis of [15], it can be proven that $\mathbf{V}_{\mathrm{H}}^{H}\mathbf{F}_{\mathrm{RF}} \approx [\mathbf{I}_{L_{\mathrm{T}}}\, \mathbf{0}_{(N_{\mathrm{T}}-L_{\mathrm{T}})\times L_{\mathrm{T}}}^{T}]^{T}$ and $\mathbf{U}_{\mathrm{H}}^{H}\mathbf{W}_{\mathrm{RF}} \approx [\mathbf{I}_{L_{\mathrm{R}}}\, \mathbf{0}_{(N_{\mathrm{R}}-L_{\mathrm{R}})\times L_{\mathrm{R}}}^{T}]^{T}$, hence,

$$R(\mathbf{P}) = \log|\mathbf{I}_{N_{\mathrm{R}}} + \frac{1}{\sigma_{\mathrm{n}}^2}\bar{\mathbf{\Sigma}}^2\mathbf{P}|, \quad (19)$$

where $\bar{\mathbf{\Sigma}} \in \mathbb{R}^{L_{\mathrm{R}} \times L_{\mathrm{T}}}$ with $[\bar{\mathbf{\Sigma}}]_{kk} = [\mathbf{\Sigma}_{\mathrm{H}}]_{kk}$ for $k = 1, \ldots, L_{\mathrm{T}}$, assuming $L_{\mathrm{T}} = L_{\mathrm{R}}$, while its remaining entries are zero. Since the involved matrices in (19) are diagonal, the information rate is decomposed into L_{T} parallel streams, as follows:

$$R(\mathbf{P}) \approx \sum_{k=1}^{L_{\mathrm{T}}} \log\left(1 + \frac{1}{\sigma_{\mathrm{n}}^2}\left[\bar{\mathbf{\Sigma}}^2\right]_{kk}[\mathbf{P}]_{kk}\right)(\text{bits/s/Hz}). \quad (20)$$

Recall that L_{T} and L_{R} have preset values based on the hardware design and describe the available RF chains at the TX and the RX, respectively. Considering only the TX, the consumed power with respect to the diagonal power allocation matrix can be written as:

$$P_{\mathrm{TX}}(\mathbf{P}) = P_{\mathrm{static}} + \sum_{k=1}^{L_{\mathrm{T}}}(\beta[\mathbf{P}]_{kk} + P_{\mathrm{RF}} + N_{\mathrm{T}}P_{\mathrm{PS}}) \quad (21)$$

$$= P_{\mathrm{static}} + \sum_{k=1}^{L_{\mathrm{T}}}\beta'[\mathbf{P}]_{kk}(\mathrm{W}), \quad (22)$$

Algorithm 3 Dinkelbach Method (DM) Based Solution

```
1:  Initialize: $\mathbf{P}^{(0)}, \nu^{(0)}$ satisfying $\mathcal{G}(\mathbf{P}^{(0)}, \nu^{(0)}) \geq 0$, $L_{\mathrm{T}}$,
    tolerance $\epsilon$
2:  $m = 0$
3:  while $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})| > \epsilon$ do
4:      Update $\mathbf{P}^{(m)}$ by solving the relaxation of (23) via
        CVX [35].
5:      Thresholding $\mathbf{P}^{(m)}$ as $\mathbf{P}_{\mathrm{th}}^{(m)}$.
6:      Counting non-zero values of $\mathbf{P}_{\mathrm{th}}^{(m)}$ provides $L_{\mathrm{T}}^{opt}$.
7:      Compute $R(\mathbf{P}^{(m)})$ and $P_{\mathrm{TX}}(\mathbf{P}^{(m)})$.
8:      Compute $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$
        where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$.
9:      Update $\nu^{(m)}$ with $R(\mathbf{P}^{(m)})/P_{\mathrm{TX}}(\mathbf{P}^{(m)})$.
10:     $m = m$+1
11: end while
12: Obtain $L_{\mathrm{T}}^{opt} = \|\mathbf{P}_{\mathrm{th}}^{(m)}\|_0$
```

where $P_{\text{static}} \triangleq P_{\text{CP}} + N_{\text{T}} P_{\text{T}}$ is independent of the power allocation matrix $\mathbf{P}$ and $\beta' \triangleq \beta + \frac{P_{\text{RF}} + N_{\text{T}} P_{\text{PS}}}{P_{\max}}$. The equivalence between (21) and (22) is justified since $\sum_{k=1}^{L_{\text{T}}} [\mathbf{P}]_{kk} = \text{tr}(\mathbf{P}) = P_{\max}$.

Based on (20) and (22), the m-th Dinkelbach method (DM) step can be expressed as follows:

$$\left\{\mathbf{P}^{(m)}, \nu^{(m)}\right\} = \arg \max_{\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\text{T}} \times L_{\text{T}}}} \mathcal{G}\left(\mathbf{P}^{(m)}, \nu^{(m)}\right), \quad (23)$$

where

$$\mathcal{G}\left(\mathbf{P}^{(m)}, \nu^{(m)}\right) \triangleq \sum_{k=1}^{L_{\text{T}}} \log\left(1 + \frac{1}{\sigma_{\text{n}}^2}\left[\bar{\mathbf{\Sigma}}^2\right]_{kk}\left[\mathbf{P}^{(m)}\right]_{kk}\right) - \nu^{(m)} \sum_{k=1}^{L_{\text{T}}} \beta' \left[\mathbf{P}^{(m)}\right]_{kk}. \quad (24)$$

Note that problem (23) is a non-convex one because of the constraint $\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\text{T}} \times L_{\text{T}}}$. To proceed, first we alleviate this constraint, thus (23) can be efficiently solved by any standard interior-point method (for example, CVX [35]). Step 3 of Algorithm 3 shows that after alleviating this constraint, (23) is solved via CVX to update $\mathbf{P}^{(m)}$. Then we impose the constraint by hard-thresholding the entries of $\mathbf{P}^{(m)}$, i.e., $\mathbf{P}_{\text{th}}^{(m)}$, as shown in Step 4 of Algorithm 3. The thresholding sets to zero all entries of $\mathbf{P}^{(m)}$ that are lower than a given tolerance value ϵ_{th}.

Algorithm 3 starts by initializing the number of available RF chains L_{T}. We update $\mathbf{P}^{(m)}$ by solving the relaxation of (23) via CVX as shown in Step 3. Steps 4-5 show that $\mathbf{P}^{(m)}$ is thresholded as $\mathbf{P}_{\text{th}}^{(m)}$ and counting its non-zero values provides us the optimal number of RF chains which keeps updating within the loop but obtained as $\|\mathbf{P}_{\text{th}}^{(m)}\|_0$ after the loop ends as shown in Step 11. $R(\mathbf{P}^{(m)})$ and $P_{\text{TX}}(\mathbf{P}^{(m)})$ are computed in Step 6 and $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$ is computed based on (24) in Step 7 where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$. Steps 8 shows the update in $\nu^{(m)}$ with $R(\mathbf{P}^{(m)})/P_{\text{TX}}(\mathbf{P}^{(m)})$. The loop continues until $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})|$ is less than a given tolerance ϵ. We consider that the optimal number of RF chains provides the number of data streams as well, i.e., $N_{\text{s}} = L_{\text{T}}^{opt}$.

Algorithm 4 Full Search (FS) Approach

```
1:  Initialize: $L_{\mathrm{T}}$, tolerance $\epsilon$, $\mathrm{EE}^{(0)} = 0$
2:  for $i = 1 : L_{\mathrm{T}}$
3:    while $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})| > \epsilon$ do
4:        Compute $\mathbf{P}^{(m)}$ subject to $i$ RF chains
          $\rightarrow$ obtain $L_{\mathrm{T}}^{opt}$ from $\mathbf{P}_{\mathrm{th}}^{(m)}$.
5:        Compute $R(\mathbf{P}^{(m)})$, $P_{\mathrm{TX}}(\mathbf{P}^{(m)})$ and $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$.
6:        Update $\nu^{(m)}$ and compute $\mathrm{EE}^{(m)}$
          $= R(\mathbf{P}^{(m)})/P_{\mathrm{TX}}(\mathbf{P}^{(m)})$.
7:        $m = m$+1
8:    end while
9:    Obtain $L_{\mathrm{T}}^{(i)} = L_{\mathrm{T}}^{opt}$ and $\mathrm{EE}^{(i)}$ based on $\mathrm{EE}^{(m)}$ value.
10:   if $\mathrm{EE}^{(i)} \geq$ previous $\mathrm{EE}^{(i-1)}$
11:       Update $\mathrm{EE}^{(i)}$ and $L_{\mathrm{T}}^{(i)}$
12:   end if
13: end for
```

C. Full Search (FS) Approach

To show that the loss performance is not much in Dinkelbach optimization we also consider a full search (FS) approach which resolves the non-convexity issue of (23) with convex approximation providing a modified version of the proposed Dinklbach optimization solution which iterates over all the possible number of RF chains. The steps are stated in Algorithm 4 where the maximum energy efficiency "EE" is obtained and the corresponding number of RF chains are considered to be optimal at the end of the algorithm. In Table IV of Section V, we show that the proposed DM has similar performance to the FS approach, while the complexity for computing FS increases significantly.

D. Brute Force (BF) Approach

The solution to achieve optimal number of RF chains at each realization is also provided in [22] which we call as the brute force (BF) approach. To make the A/D hybrid beamforming system energy efficient, BF approach, at each realization (current channel condition), makes a search on all the possible number of RF chains, i.e., $L_{\text{T}} = \{1, 2, 3, \ldots, N_{\text{T}}\}$, and computes best energy efficiency while designing the precoder and combiner matrices, and chooses the corresponding number of RF chains as the optimal number of RF chains. We, in our work, mitigate that need of searching for all possible number of RF chains and then finding an optimal solution, and thus providing equally a high energy efficient and low complexity solution. The observations made in the next section support this statement.

V. Simulation Results

This section shows the performance of the proposed DM compared to the existing state of the art solutions such as the BF approach, digital beamforming, analog beamforming and modified version of the proposed solution, i.e., FS approach. For simulations, the proposed DM and the FS approach consider $L_\mathrm{T} = L_\mathrm{R} = \mathrm{length}(\mathrm{eig}(\mathbf{HH}^H))$ and the BF approach uses the same precoding and combining matrices as the DM solution. The tolerance values considered in both the DM solution and the FS approach algorithms are $\epsilon = 10^{-4}$ and $\epsilon_\mathrm{th} = 10^{-6}$. The fully digital beamforming solution uses the same number of RF chains as antennas, i.e., $L_\mathrm{T} = N_\mathrm{T}$ and $L_\mathrm{R} = N_\mathrm{R}$, and precoding and combining matrices are $\mathbf{F}_\mathrm{opt}$ and $\mathbf{W}_\mathrm{mmse}$, respectively, as shown in Sections III-A and III-B. The analog beamforming solution implements a single RF chain, i.e., $L_\mathrm{T} = L_\mathrm{R} = 1$, and the precoding and combining matrices are computed as the phases of the first singular vectors, i.e., $\mathbf{F} = \mathbf{V}_\mathrm{H}(1 : N_\mathrm{T}, 1)/\mathrm{abs}(\mathbf{V}_\mathrm{H})$ and $\mathbf{W} = \mathbf{U}_\mathrm{H}(1 : N_\mathrm{R}, 1)/\mathrm{abs}(\mathbf{U}_\mathrm{H})$, respectively.

The performance of the codebook-free designs such as ADMM [16] and SVD based [12] solutions when incorporated with the proposed framework, using L_T^{opt} RF chains, are also observed over the case when fixed number of RF chains are used to compute the precoder and combiner matrices. The comparison between GP and OMP algorithms is also observed through observing the variations in run time with respect to the number of RF chains and computational complexities.

A. System Setup

For the channel parameters, there are 10 rays for each cluster and there are 8 clusters in total, i.e., $N_\mathrm{ray} = 10$ and $N_\mathrm{cl} = 8$ in (1). The average power of each cluster is unity, i.e., $\sigma_{\alpha,i} = 1$. The azimuth and elevation angles of departure and arrival are computed on the basis of the Laplacian distribution [36] with uniformly distributed mean angles and angle spread as 7.5°. The mean angles are sectored within the range of 60° to 120° in the azimuth domain, and 80° to 100° in the elevation domain. The 64 antenna elements at the TX, i.e., $N_\mathrm{T} = 64$, and 16 at the RX, i.e., $N_\mathrm{R} = 16$, in the ULA, antenna elements are spaced by distance $d = \lambda/2$ where $\lambda/2$ can be based on a standard frequency value such as 28 GHz [22]. The system bandwidth is normalized to 1 Hz in the simulations. The signal to noise ratio (SNR) is $1/\sigma_\mathrm{n}^2$. All the simulation results are averaged over 1000 random channel realizations. To illustrate the achievable energy efficiency of different precoding solutions, the parameters in the power expressions for each precoder design are set as shown in Table I(a). For a typical case, the power per power amplifier, $P_\mathrm{PA} = 300$ mW, and maximum achievable power, $P_\mathrm{max} = 1$ W. Table I(b) shows the maximum power which can be consumed as determined in (15) for different number of RF chains in a 64 × 16 fully-connected system. The amplifier efficiency $1/\beta$ is considered as 0.4 and the minimum desired rate in (16), $R_\mathrm{min} = 1$ bits/s/Hz.

B. Beam Training and Data Communications Phases Analysis

Based on the described communication phases in Fig. 1(b), there are L_T active RF chains during the beam training phase. Once the Dinkelbach or FS optimization is performed then we obtain the optimal number L_T^{opt} RF chains for the data communications phase. Considering that α represents the ratio between the two phases, the power consumption performance for both the stages is given by:

$$\mathrm{Power} = \alpha \times P(L_\mathrm{T}) + (1-\alpha) \times P\left(L_\mathrm{T}^{opt}\right)(\mathrm{W}), \quad (25)$$

where $P(L_\mathrm{T})$ is the power consumption with (15) using L_T RF chains and $P(L_\mathrm{T}^{opt})$ is using the optimal number of RF chains, L_T^{opt}. For example, as shown in Fig. 2(a), when

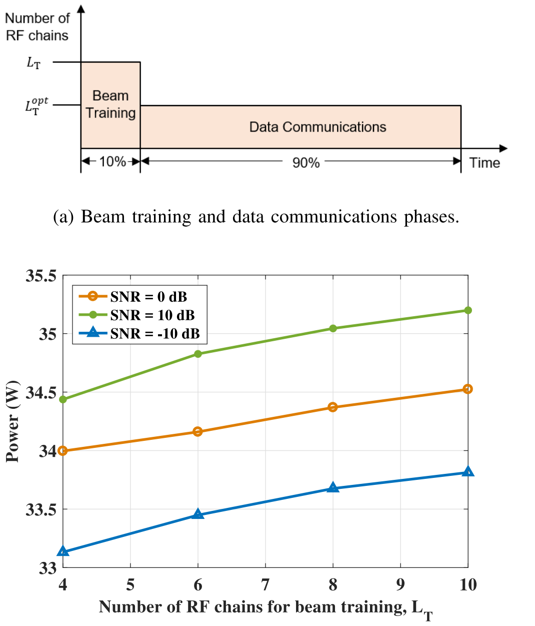
TABLE I
SIMULATION PARAMETERS FOR THE POWER EXPRESSIONS OF DIFFERENT PRECODING SOLUTIONS

Common power of TX	P_CP = 10 W
Power per RF chain	P_RF = 100 mW
Power per phase shifter	P_PS = 10 mW

(a) Typical values of the power terms [37].

Number of RF chains, L_T	Maximum consumed power (W)
4	34.50
8	38.50
64	94.50

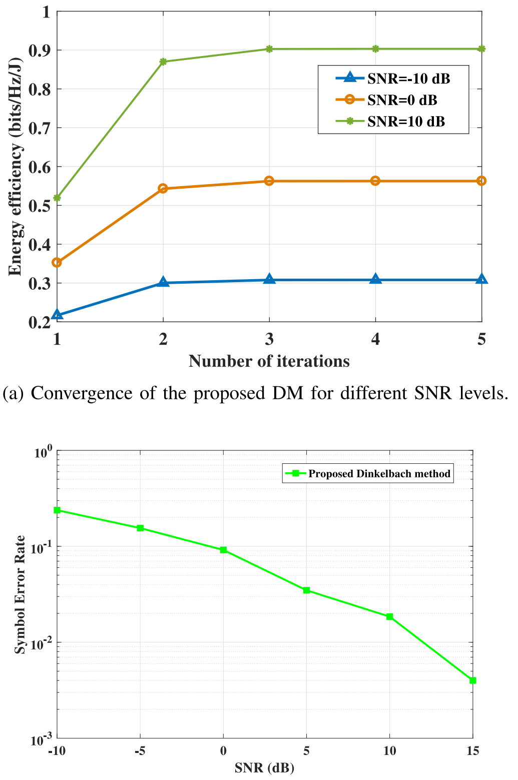
(b) Maximum consumed power in (15) for different values of L_T for a 64 × 16 system with $\mathrm{tr}(\mathbf{FF}^H) = 1$.



(a) Beam training and data communications phases.

(b) Overall power consumption performance for 10% beam training and 90% data communications phases.

Fig. 2. Beam training and data communications phases and associated power consumption performance for a fully-connected 64 × 16 system.



(a) Convergence of the proposed DM for different SNR levels.

(b) Symbol error rate (SER) vs SNR for the proposed DM with Q-PSK modulation.

Fig. 3. Convergence and SER accuracy performance of the proposed DM solution for a fully-connected 64 × 16 system.

we consider that the beam training phase is active for 10% of the time with L_{T} RF chains, i.e., $\alpha = 0.1$, and the data communications phase is active for the remaining 90% time with L_{T}^{opt} RF chains, i.e., $1 - \alpha = 0.9$. The performance is observed with three SNR cases in Fig. 2(b). It can be observed that the overall power consumption increases with the increase in the number of RF chains in the beam training phase and high SNR values have higher power consumption levels. For example, at $L_{\mathrm{T}} = 6$, the power consumption at SNR $= 0$ dB is about 0.65 W higher than at SNR $= -10$ dB.

C. Convergence and Accuracy Performance of the DM

Fig. 3(a) shows the convergence of the Dinkelbach optimization solution as proposed in Algorithm 3 to obtain the optimal number of RF chains. It can be observed that the energy efficiency for different SNR levels increases with the iterations used to find the optimal number of RF chains. The proposed solution converges rapidly and needs only 2 iterations to converge and achieve an optimal solution at each realization. To express the accuracy performance of the proposed DM, Fig. 3(b) shows the symbol error rate (SER) versus SNR plot for quadrature phase shift keying (QPSK) modulation where SER decreases with the increase in SNR.

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN DM AND BF SOLUTIONS

Algorithm	Complexity Order
Dinkelbach	$\mathcal{O}(L_{\mathrm{T}}^{opt})$
Brute force	$\mathcal{O}(L_{\mathrm{T}}^{opt} N_{\mathrm{T}})$

(a) Complexity orders of DM and BF.

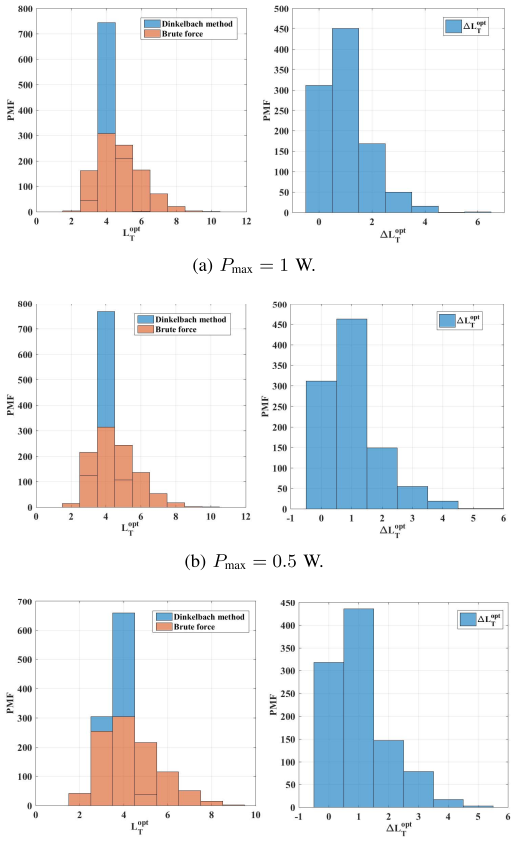
No. of TX antennas, N_{T}	Time (s): DM	Time (s): BF
80	4.29	6.13
96	4.38	10.6
112	4.54	17.6
128	4.55	26.9

(b) Run time comparison w.r.t. N_{T} at SNR = 10 dB and $P_{\max} = 1$.

D. Proposed DM Versus BF Approach

The comparison is made to the BF method [22] in detail in terms of the probability mass function (PMF) for RF chain selection, energy efficiency performance and the computational complexity. The PMF plots indicate the histogram that for how many realizations (on y-axis) a particular value of the variable defined on x-axis is achieved. Figs. 4 and 5 show the PMF of the distribution of the proposed DM and the BF approach over the optimal number of RF chains, i.e., L_{T}^{opt}, their difference, i.e., $\Delta L_{\mathrm{T}}^{opt} = |L_{\mathrm{T}}^{opt}{}_{\mathrm{BF}} - L_{\mathrm{T}}^{opt}{}_{\mathrm{DM}}|$, and the energy efficiency difference, i.e., $\Delta \mathrm{E} = |\mathrm{EE}_{\mathrm{BF}} - \mathrm{EE}_{\mathrm{DM}}|$, at each channel realization. Fig. 4 shows that for how many channel realizations, the beamforming solutions such as the DM and the BF approach find a particular optimal number of RF chains for different values of $P_{\max}$. It gives us an idea on how close the proposed DM solution is to the BF technique, in terms of finding the optimal number of RF chains. For example, at $P_{\max} = 1$ W, the DM solution chooses $L_{\mathrm{T}}^{opt} = 4$ for ≈ 750 different channel realizations whereas BF chooses 4 RF chains for ≈ 300 realizations and the difference (at each realization) between chosen optimal number of RF chains by both the methods, i.e., $\Delta L_{\mathrm{T}}^{opt}$ is 0 for ≈ 450 different realizations. Also, for example, the energy efficiency difference between the two methods, Δ E, at $P_{\max} = 1$ W is close to 0 bits/Hz/J for ≈ 650 channel realizations as observed from Fig. 5.

Table II(a) shows the computational complexities used by the solutions of the BF approach and the DM with respect to the number of the RF chains. We can observe that complexity for the solution of the DM requires complexity order of only $\mathcal{O}(L_{\mathrm{T}}^{opt})$ per iteration. Since the number of the required iterations is usually very small, the overall complexity of the DM is much less than the BF approach which depends on the product of the number of RF chains and the number of antennas. This is also verified by the run time results as shown in Table II(b). At SNR $= 10$ dB and $P_{\max} = 1$, the run time (in seconds) is much less for the proposed solution with respect to (w.r.t.) the number of TX antennas. These results are reported from MATLAB simulation runtime for 10 independent channel realizations. For example, for a large



(a) $P_{\max} = 1$ W.

(b) $P_{\max} = 0.5$ W.

(c) $P_{\max} = 0.25$ W.

Fig. 4. PMF plots of the DM and BF solutions at different $P_{\max}$ values for the optimal number of RF chains L_{T}^{opt} and their difference $\Delta L_{\mathrm{T}}^{opt}$ for 64 × 16 system and SNR = 10 dB.

number of antennas, i.e., $N_{\mathrm{T}} = 128$, the proposed solution consumes ≈ 6 times less run time than the BF solution. The observations support the statement that the proposed solution has low complexity while still optimizing the number of RF chains.

E. Proposed GP Versus OMP

Concerning the complexity for deriving the beamforming matrices, recall that OMP requires inversion of a matrix with size $k \times k$, at each one of the L_{T}^{opt} iterations in total, with $k = 1, \ldots, L_{\mathrm{T}}^{opt}$. This operation has cubic complexity order with respect to the size of the matrix, i.e., $\mathcal{O}(k^3)$, in general. So, for L_{T}^{opt} iterations, the total cost would be:

$$\sum_{k=1}^{L_{\mathrm{T}}^{opt}} \mathcal{O}\left(k^3\right) = \mathcal{O}\left(\left(L_{\mathrm{T}}^{opt}\right)^4\right). \tag{26}$$

Additionally, a matrix-matrix product is required at each iteration with total cost $\mathcal{O}((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}})$. On the other side,

TABLE III
COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN GP AND OMP SOLUTIONS

Algorithm	Complexity Order
OMP	$\mathcal{O}((L_{\mathrm{T}}^{opt})^4) + \mathcal{O}((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}})$
GP	$\mathcal{O}((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}})$

(a) Complexity orders of GP and OMP.

No. of RF chains at the TX	Time (μs): OMP	Time (μs): GP
8	1.6	1.1
16	5.8	2.8
24	10	5.0
32	16.4	8.0

(b) Run time comparison w.r.t. the number of RF chains for 64 × 16 mmWave system with N_{cl} = 8, N_{ray} = 10, and SNR = 10 dB.

the proposed GP algorithm requires only matrix-matrix multiplications at each iteration, hence the complexity order is $\mathcal{O}((L_{\mathrm{T}}^{opt})^3 N_{\mathrm{T}})$. This complexity reduction is justified by the substitution of the matrix inversion with a gradient step. The derived complexity orders are summarized in Table III(a). In Table III(b) we show the MATLAB run time comparison (in μs) between OMP and GP w.r.t. the number of RF chains at the TX for a 64 × 16 mmWave MIMO system with SNR = 10 dB. As the time difference between both the algorithmic solutions is considerable with the increase in the number of RF chains, the obtained values indicate that GP consumes much less time than OMP leading to a lower complexity system.

F. Performance Evaluation

Fig. 6 shows the energy efficiency and spectral efficiency performance of the proposed solution, the BF solution, the full digital solution and the analog beamforming solution w.r.t. SNR for a 64 × 16 mmWave MIMO system. It can be clearly observed from Fig. 6(a) that the proposed solution is as energy efficient as the BF solution, and better than the fully digital and analog beamforming solutions. For example, at 10 dB, the proposed solution has merely a energy efficiency difference of $\approx$ 0.01 bits/Hz/J with the BF, but shows $\approx$ 0.35 bits/Hz/J and $\approx$ 0.25 bits/Hz/J better energy efficiency than the fully digital and analog beamforming solutions, respectively. Also, for example, in Fig. 6(b) the proposed design at 10 dB shows a $\approx$ 10 bits/s/Hz less spectral efficiency than the fully digital solution, $\approx$ 10 bits/s/Hz better than analog beamforming and approximately the same performance as the BF method.

Fig. 7(a) shows the energy efficiency comparison among the solutions with partially-connected structures where each RF chain is connected to $N_{\mathrm{T}}/L_{\mathrm{T}}^{opt}$ antennas through phase shifters. We can observe similar energy efficiency performance characteristics as in Fig. 6(a); for example, the proposed solution has approximately the same energy efficiency performance as the BF method, $\approx$ 0.4 bits/Hz/J and $\approx$ 0.32 bits/Hz/J better than the fully digital and analog beamforming solutions, respectively, at SNR = 15 dB. Fig. 7(b) shows the energy efficiency performance comparison w.r.t. the number of TX antennas, N_{T}, for a fully-connected structure. We

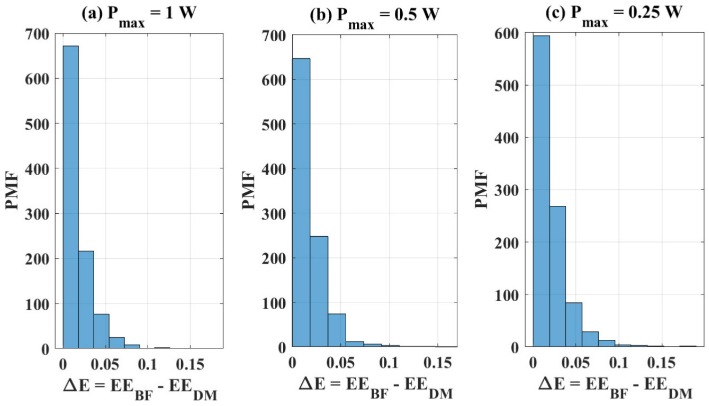
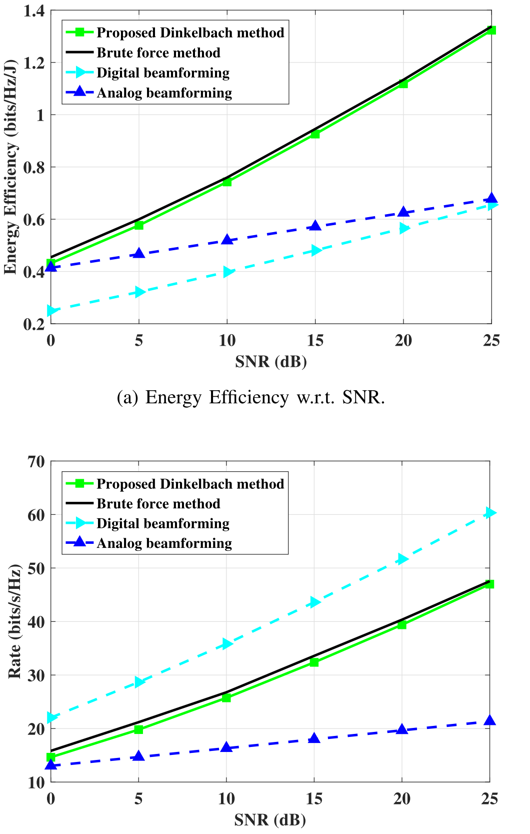


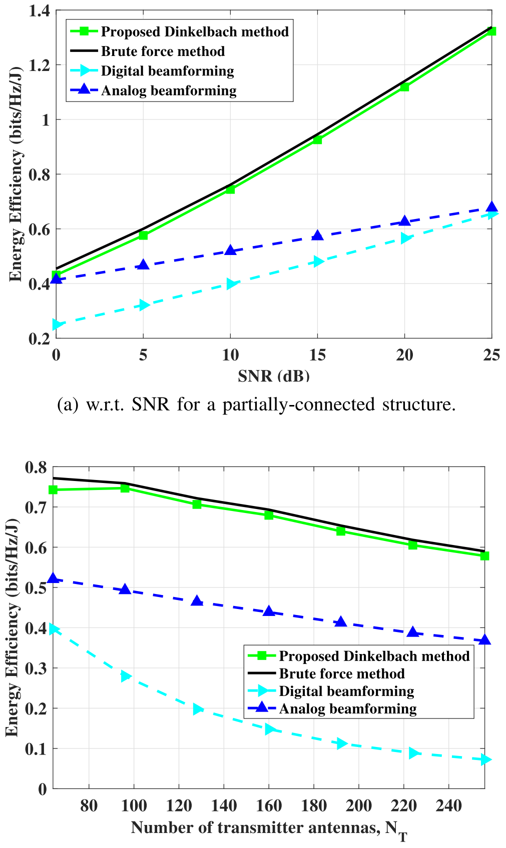
Fig. 5. PMF plots of energy efficiency difference between DM and BF solutions at different $P_{\max}$ values for a 64 × 16 system and SNR = 10 dB.



(a) Energy Efficiency w.r.t. SNR.

(b) Rate w.r.t. SNR.

Fig. 6. Energy efficiency and rate performance of different solutions w.r.t. SNR for a fully-connected 64 × 16 system at $P_{\max} = 1$ W.



(a) w.r.t. SNR for a partially-connected structure.

(b) w.r.t. $N_{\rm T}$ for a fully-connected structure.

Fig. 7. Energy efficiency performance of different solutions for a 64 × 16 hybrid mmWave MIMO system at $P_{\max} = 1$ W.

can observe that the performance starts decreasing with the increase in the number of antenna elements. For example, at $N_{\rm T} = 64$, the energy efficiency for the proposed DM is close to that of the BF solution which is ≈ 0.35 bits/Hz/J and ≈ 0.25 bits/Hz/J better than the fully digital beamforming and analog beamforming solutions, respectively. At $N_{\rm T} = 256$, the

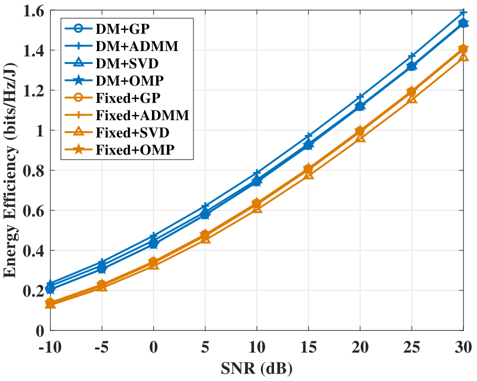


Fig. 8. Energy efficiency performance gains w.r.t. SNR at $N_{\mathrm{T}} = 64$ over the fixed number of RF chains case.

TABLE IV
ENERGY EFFICIENCY AND COMPUTATIONAL COMPLEXITY COMPARISONS BETWEEN THE PROPOSED DM AND THE FS APPROACH

SNR (dB)	$\lvert \mathrm{EE}_{\mathrm{DM}} - \mathrm{EE}_{\mathrm{FS}} \rvert$ (bits/Hz/J)
-10	0.013
-5	0.018
0	0.043
5	0.108
10	0.189

(a) Energy efficiency performance difference between the DM and the FS approach.

Algorithm	Complexity Order
Dinkelbach	$\mathcal{O}(L_{\mathrm{T}}^{opt})$
Full search	$\mathcal{O}(L_{\mathrm{T}}^{opt} L_{\mathrm{T}})$

(b) Complexity orders of the DM and the FS approach.

energy efficiency performance for the proposed DM solution is decreased to ≈ 0.56 bits/Hz/J and close to the BF solution, and ≈ 0.5 bits/Hz/J and ≈ 0.2 bits/Hz/J better than the fully digital beamforming and analog beamforming solutions, respectively.

Fig. 8 shows the energy efficiency gain of the DM based framework when used with codebook-based GP and OMP techniques, and when incorporated with codebook-free ADMM [16] and SVD [12] techniques, over the case of a fixed number of RF chains, in this case, 8. The codebook-free technique such as ADMM performs better than the codebook-based techniques such as GP and OMP, while SVD shows a similar performance. The energy efficiency performance of GP and OMP techniques are same. Table IV(a) shows energy efficiency performance comparison between the proposed DM approach (Algorithm 3), i.e., $\mathrm{EE}_{\mathrm{DM}}$, and the FS approach (Algorithm 4), i.e., $\mathrm{EE}_{\mathrm{FS}}$, where we can observe that the difference between their energy efficiency is considerably low. It states that FS approach shows very similar performance to the proposed method. From implementation perspective, Table IV(b) clearly suggests that the complexity for FS approach increases significantly as the search is made for all possible number of RF chains L_{T}.

VI. Conclusion

This paper proposes an energy efficient A/D hybrid beamforming framework with a novel architecture for a mmWave MIMO system, where we optimize the active number of RF chains through fractional programming. The proposed DM based framework reduces the complexity significantly and achieves almost the same energy efficiency performance as the state of the art BF approach. Both approaches achieve higher energy efficiency performance when compared with the fully digital beamforming and the analog beamforming solutions. In particular, the proposed solution only needs to compute the precoder and combiner matrices once, after the number of active RF chains are decided through the Dinkelbach optimization solution. The modified version of the proposed solution, i.e., FS approach, shows very similar performance to the proposed DM but the complexity increases significantly. The codebook-free designs such as ADMM and SVD based solutions, when incorporated with the proposed framework also achieve better energy efficiency performance over the fixed number of RF chains case. It is also shown that GP incorporated with the proposed DM is a faster and less complex approximation solution to compute the precoder and combiner matrices than OMP. For this paper, we focus on maximizing the energy efficiency but extending these techniques to consider both estimated channels and frequency selective channels can be considered for future work.

Acknowledgment

The authors gratefully acknowledge the Editor for coordinating the review, the anonymous reviewers for helping in revising this paper and Dr. Pan Cao for helpful discussions concerning this work.

References

[1] Cisco Visual Mobile, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016*, vol. 1, Cisco, San Jose, CA, USA, 2016.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

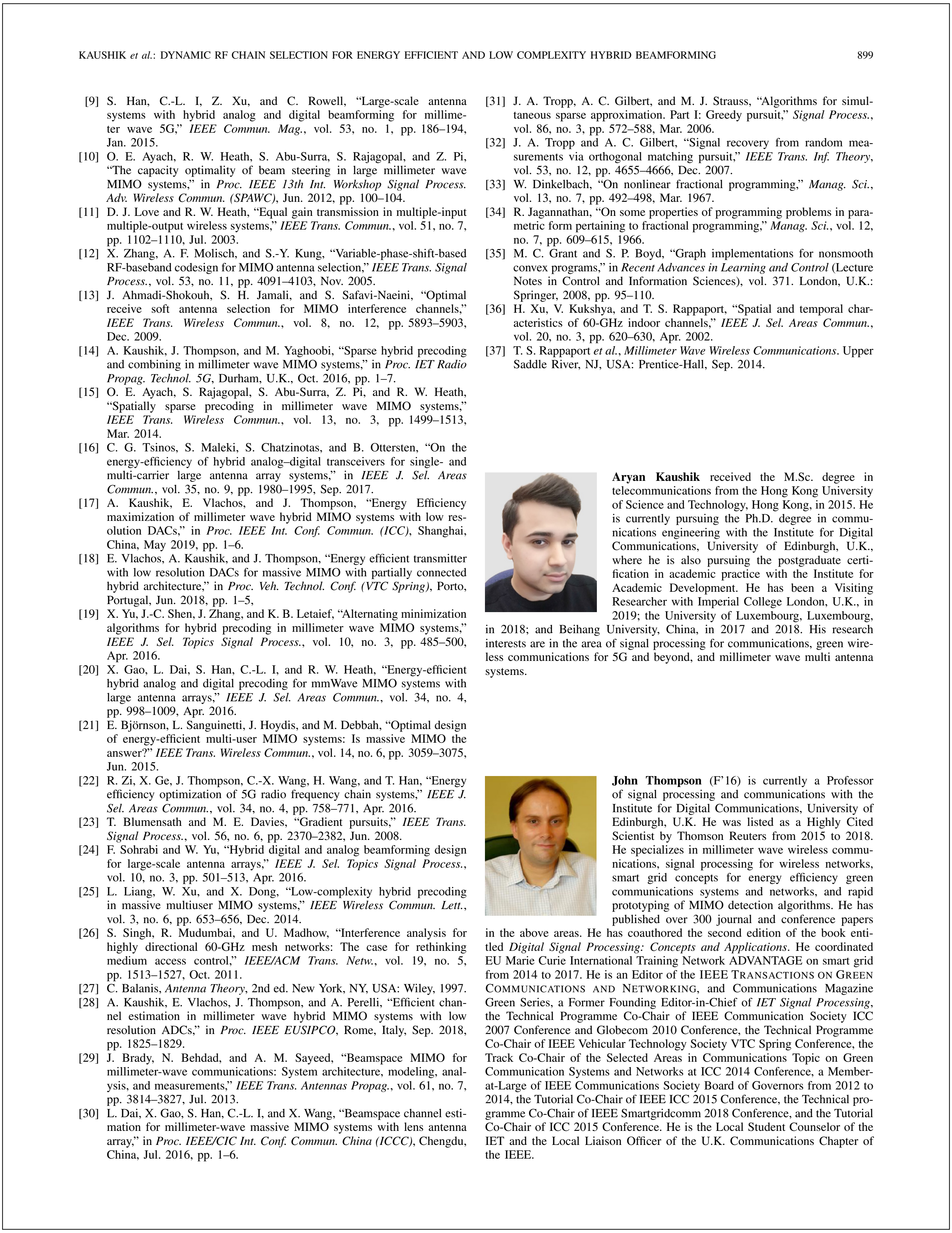
[4] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[6] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[7] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[8] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.



[9] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[10] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE 13th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2012, pp. 100–104.

[11] D. J. Love and R. W. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

[12] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[13] J. Ahmadi-Shokouh, S. H. Jamali, and S. Safavi-Naeini, "Optimal receive soft antenna selection for MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5893–5903, Dec. 2009.

[14] A. Kaushik, J. Thompson, and M. Yaghoobi, "Sparse hybrid precoding and combining in millimeter wave MIMO systems," in *Proc. IET Radio Propag. Technol. 5G*, Durham, U.K., Oct. 2016, pp. 1–7.

[15] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[16] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog–digital transceivers for single- and multi-carrier large antenna array systems," in *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980–1995, Sep. 2017.

[17] A. Kaushik, E. Vlachos, and J. Thompson, "Energy Efficiency maximization of millimeter wave hybrid MIMO systems with low resolution DACs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[18] E. Vlachos, A. Kaushik, and J. Thompson, "Energy efficient transmitter with low resolution DACs for massive MIMO with partially connected hybrid architecture," in *Proc. Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–5,

[19] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[20] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[21] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[22] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758–771, Apr. 2016.

[23] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370–2382, Jun. 2008.

[24] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[25] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[26] S. Singh, R. Mudumbai, and U. Madhow, "Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513–1527, Oct. 2011.

[27] C. Balanis, *Antenna Theory*, 2nd ed. New York, NY, USA: Wiley, 1997.

[28] A. Kaushik, E. Vlachos, J. Thompson, and A. Perelli, "Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs," in *Proc. IEEE EUSIPCO*, Rome, Italy, Sep. 2018, pp. 1825–1829.

[29] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[30] L. Dai, X. Gao, S. Han, C.-L. I, and X. Wang, "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Chengdu, China, Jul. 2016, pp. 1–6.

[31] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[32] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[33] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[34] R. Jagannathan, "On some properties of programming problems in parametric form pertaining to fractional programming," *Manag. Sci.*, vol. 12, no. 7, pp. 609–615, 1966.

[35] M. C. Grant and S. P. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control* (Lecture Notes in Control and Information Sciences), vol. 371. London, U.K.: Springer, 2008, pp. 95–110.

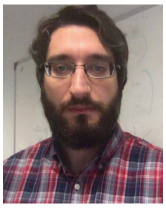
[36] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.

[37] T. S. Rappaport *et al.*, *Millimeter Wave Wireless Communications*. Upper Saddle River, NJ, USA: Prentice-Hall, Sep. 2014.

Aryan Kaushik received the M.Sc. degree in telecommunications from the Hong Kong University of Science and Technology, Hong Kong, in 2015. He is currently pursuing the Ph.D. degree in communications engineering with the Institute for Digital Communications, University of Edinburgh, U.K., where he is also pursuing the postgraduate certification in academic practice with the Institute for Academic Development. He has been a Visiting Researcher with Imperial College London, U.K., in 2019; the University of Luxembourg, Luxembourg, in 2018; and Beihang University, China, in 2017 and 2018. His research interests are in the area of signal processing for communications, green wireless communications for 5G and beyond, and millimeter wave multi antenna systems.

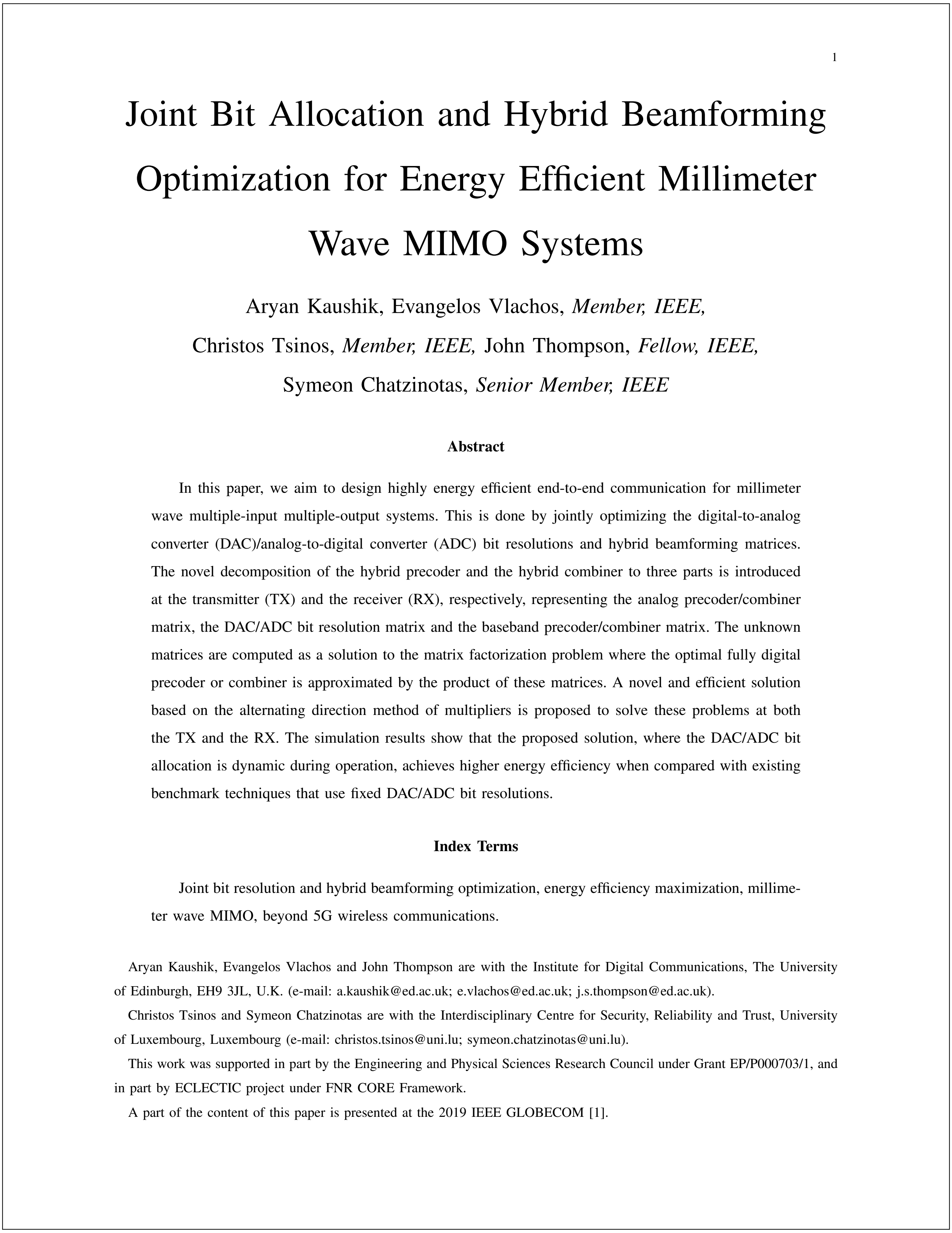
John Thompson (F'16) is currently a Professor of signal processing and communications with the Institute for Digital Communications, University of Edinburgh, U.K. He was listed as a Highly Cited Scientist by Thomson Reuters from 2015 to 2018. He specializes in millimeter wave wireless communications, signal processing for wireless networks, smart grid concepts for energy efficiency green communications systems and networks, and rapid prototyping of MIMO detection algorithms. He has published over 300 journal and conference papers in the above areas. He has coauthored the second edition of the book entitled *Digital Signal Processing: Concepts and Applications*. He coordinated EU Marie Curie International Training Network ADVANTAGE on smart grid from 2014 to 2017. He is an Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and Communications Magazine Green Series, a Former Founding Editor-in-Chief of *IET Signal Processing*, the Technical Programme Co-Chair of IEEE Communication Society ICC 2007 Conference and Globecom 2010 Conference, the Technical Programme Co-Chair of IEEE Vehicular Technology Society VTC Spring Conference, the Track Co-Chair of the Selected Areas in Communications Topic on Green Communication Systems and Networks at ICC 2014 Conference, a Member-at-Large of IEEE Communications Society Board of Governors from 2012 to 2014, the Tutorial Co-Chair of IEEE ICC 2015 Conference, the Technical Programme Co-Chair of IEEE Smartgridcomm 2018 Conference, and the Tutorial Co-Chair of ICC 2015 Conference. He is the Local Student Counselor of the IET and the Local Liaison Officer of the U.K. Communications Chapter of the IEEE.

Evangelos Vlachos (M'19) is currently a Research Associate of signal processing for communications with the Institute for Digital Communications, University of Edinburgh, U.K. His current research focus is on the next-generation 5G wireless networks, developing efficient low-power, and low-complexity techniques, suitable for the future millimeter wave massive MIMO systems. From 2015 to 2016, he was a Post-Doctoral Researcher of computer engineering and informatics with the Laboratory of Signal Processing and Telecommunications, University of Patras, Greece, working on distributed learning for signal processing over networks, where he was a Post-Doctoral Researcher of signal processing with the Visualization and Virtual Reality Group in 2016. He participated in six research projects funded by EU. He was a recipient of the Best Paper Award from IEEE ICME in 2017.



Christos Tsinos (M'14) is currently a Research Associate with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg. His current research interests include signal processing for mmWave, massive MIMO, cognitive radio and satellite communications, and hyperspectral image processing. From 2014 to 2015, he was a Post-Doctoral Researcher with the University of Patras, Greece. He is currently the Principal Investigator of the project "Energy and Complexity Efficient Millimeter-Wave Large-Array Communications" funded under FNR CORE Framework and a member of the Technical Chamber of Greece.

Symeon Chatzinotas (SM'13) is currently a Senior Research Scientist and the Deputy Head of the Research Group SIGCOM with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg. He has worked in numerous research and development projects for the Institute of Informatics and Telecommunications, the National Center for Scientific Research "Demokritos," the Institute of Telematics and Informatics, the Center of Research and Technology Hellas, and Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey, U.K. He is a Visiting Professor with the University of Parma, Italy. He has authored over 300 technical papers in refereed international journals, conferences, and scientific books. His research interests are on multiuser information theory, cooperative/cognitive communications, cross-layer wireless network optimization, and content delivery networks. He was a co-recipient of the 2014 IEEE Distinguished Contributions to Satellite Communications Award, the CROWNCOM 2015 Best Paper Award, and the 2018 EURASIP JWCN Best Paper Award.



Joint Bit Allocation and Hybrid Beamforming Optimization for Energy Efficient Millimeter Wave MIMO Systems

Aryan Kaushik, Evangelos Vlachos, *Member, IEEE,*
Christos Tsinos, *Member, IEEE,* John Thompson, *Fellow, IEEE,*
Symeon Chatzinotas, *Senior Member, IEEE*

Abstract

In this paper, we aim to design highly energy efficient end-to-end communication for millimeter wave multiple-input multiple-output systems. This is done by jointly optimizing the digital-to-analog converter (DAC)/analog-to-digital converter (ADC) bit resolutions and hybrid beamforming matrices. The novel decomposition of the hybrid precoder and the hybrid combiner to three parts is introduced at the transmitter (TX) and the receiver (RX), respectively, representing the analog precoder/combiner matrix, the DAC/ADC bit resolution matrix and the baseband precoder/combiner matrix. The unknown matrices are computed as a solution to the matrix factorization problem where the optimal fully digital precoder or combiner is approximated by the product of these matrices. A novel and efficient solution based on the alternating direction method of multipliers is proposed to solve these problems at both the TX and the RX. The simulation results show that the proposed solution, where the DAC/ADC bit allocation is dynamic during operation, achieves higher energy efficiency when compared with existing benchmark techniques that use fixed DAC/ADC bit resolutions.

Index Terms

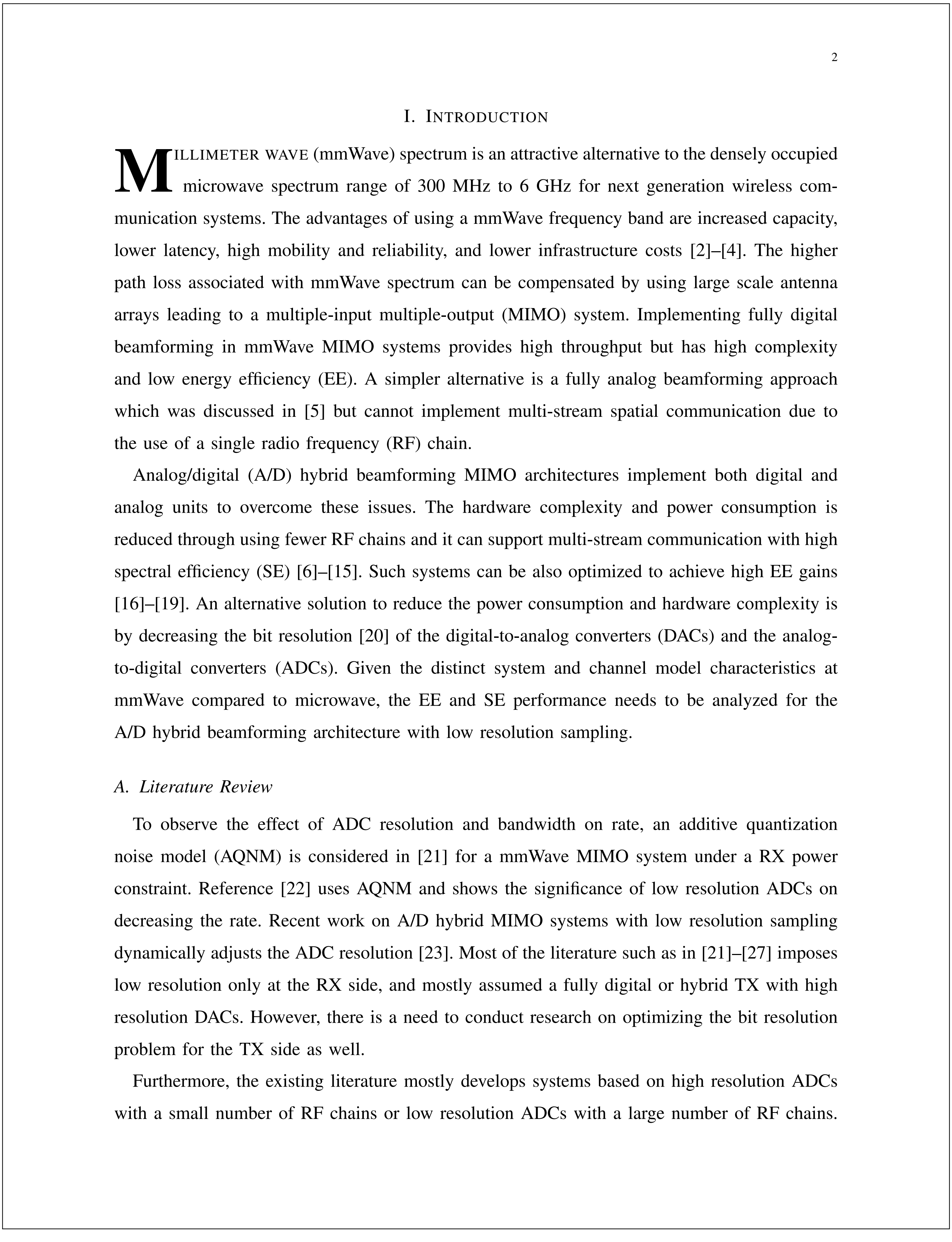
Joint bit resolution and hybrid beamforming optimization, energy efficiency maximization, millimeter wave MIMO, beyond 5G wireless communications.

Aryan Kaushik, Evangelos Vlachos and John Thompson are with the Institute for Digital Communications, The University of Edinburgh, EH9 3JL, U.K. (e-mail: a.kaushik@ed.ac.uk; e.vlachos@ed.ac.uk; j.s.thompson@ed.ac.uk).

Christos Tsinos and Symeon Chatzinotas are with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg (e-mail: christos.tsinos@uni.lu; symeon.chatzinotas@uni.lu).

This work was supported in part by the Engineering and Physical Sciences Research Council under Grant EP/P000703/1, and in part by ECLECTIC project under FNR CORE Framework.

A part of the content of this paper is presented at the 2019 IEEE GLOBECOM [1].



I. Introduction

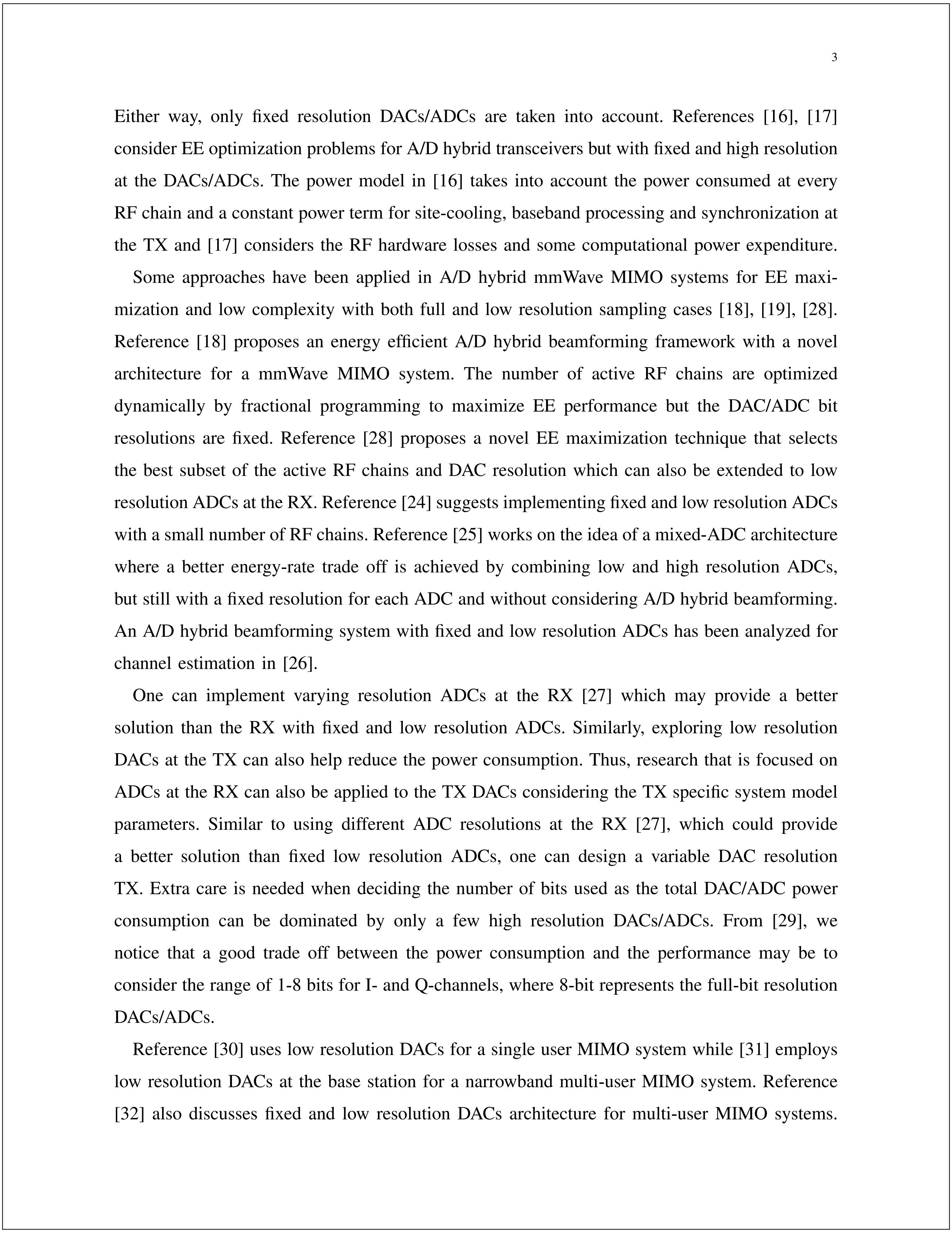
Millimeter wave (mmWave) spectrum is an attractive alternative to the densely occupied microwave spectrum range of 300 MHz to 6 GHz for next generation wireless communication systems. The advantages of using a mmWave frequency band are increased capacity, lower latency, high mobility and reliability, and lower infrastructure costs [2]–[4]. The higher path loss associated with mmWave spectrum can be compensated by using large scale antenna arrays leading to a multiple-input multiple-output (MIMO) system. Implementing fully digital beamforming in mmWave MIMO systems provides high throughput but has high complexity and low energy efficiency (EE). A simpler alternative is a fully analog beamforming approach which was discussed in [5] but cannot implement multi-stream spatial communication due to the use of a single radio frequency (RF) chain.

Analog/digital (A/D) hybrid beamforming MIMO architectures implement both digital and analog units to overcome these issues. The hardware complexity and power consumption is reduced through using fewer RF chains and it can support multi-stream communication with high spectral efficiency (SE) [6]–[15]. Such systems can be also optimized to achieve high EE gains [16]–[19]. An alternative solution to reduce the power consumption and hardware complexity is by decreasing the bit resolution [20] of the digital-to-analog converters (DACs) and the analog-to-digital converters (ADCs). Given the distinct system and channel model characteristics at mmWave compared to microwave, the EE and SE performance needs to be analyzed for the A/D hybrid beamforming architecture with low resolution sampling.

A. Literature Review

To observe the effect of ADC resolution and bandwidth on rate, an additive quantization noise model (AQNM) is considered in [21] for a mmWave MIMO system under a RX power constraint. Reference [22] uses AQNM and shows the significance of low resolution ADCs on decreasing the rate. Recent work on A/D hybrid MIMO systems with low resolution sampling dynamically adjusts the ADC resolution [23]. Most of the literature such as in [21]–[27] imposes low resolution only at the RX side, and mostly assumed a fully digital or hybrid TX with high resolution DACs. However, there is a need to conduct research on optimizing the bit resolution problem for the TX side as well.

Furthermore, the existing literature mostly develops systems based on high resolution ADCs with a small number of RF chains or low resolution ADCs with a large number of RF chains.

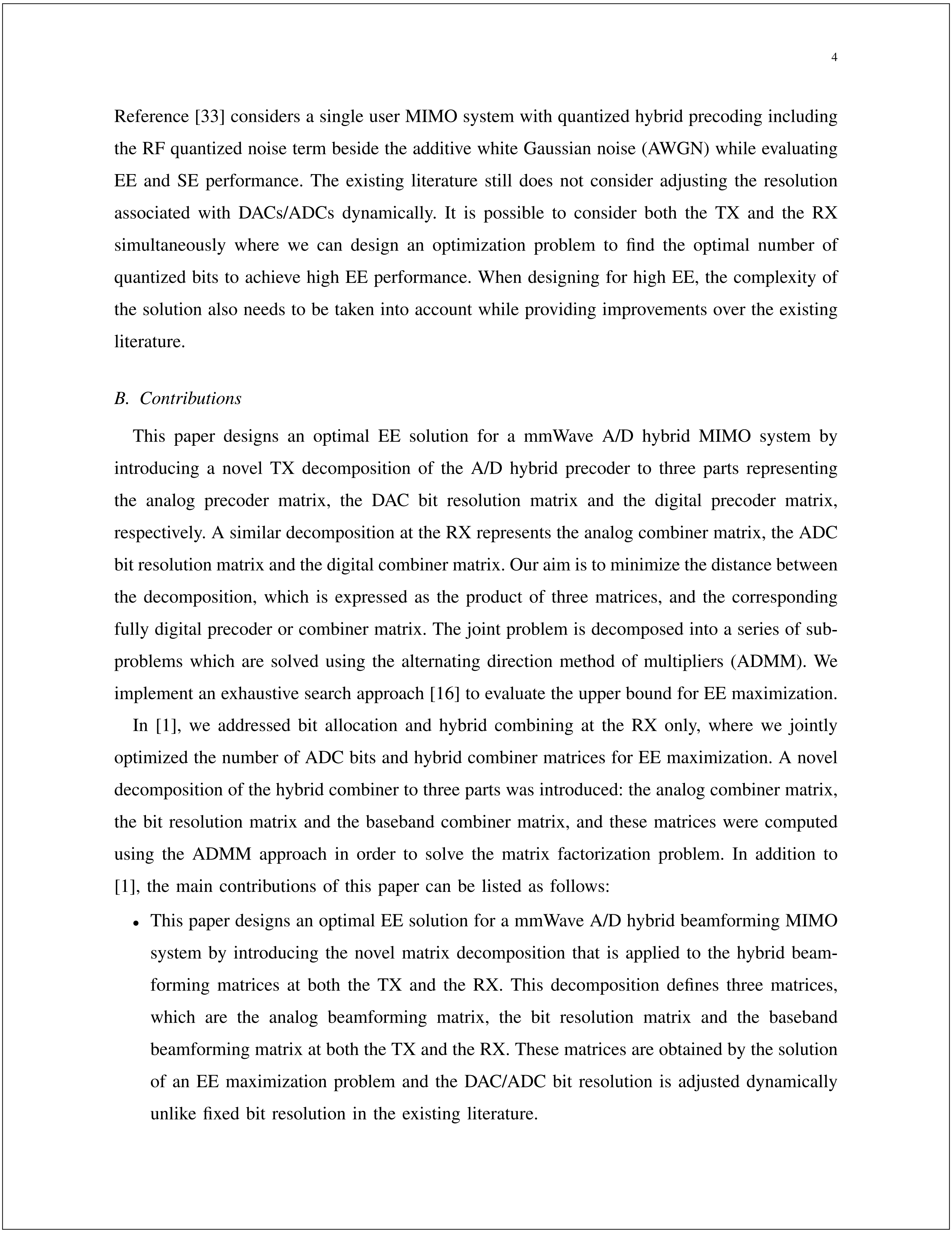


Either way, only fixed resolution DACs/ADCs are taken into account. References [16], [17] consider EE optimization problems for A/D hybrid transceivers but with fixed and high resolution at the DACs/ADCs. The power model in [16] takes into account the power consumed at every RF chain and a constant power term for site-cooling, baseband processing and synchronization at the TX and [17] considers the RF hardware losses and some computational power expenditure.

Some approaches have been applied in A/D hybrid mmWave MIMO systems for EE maximization and low complexity with both full and low resolution sampling cases [18], [19], [28]. Reference [18] proposes an energy efficient A/D hybrid beamforming framework with a novel architecture for a mmWave MIMO system. The number of active RF chains are optimized dynamically by fractional programming to maximize EE performance but the DAC/ADC bit resolutions are fixed. Reference [28] proposes a novel EE maximization technique that selects the best subset of the active RF chains and DAC resolution which can also be extended to low resolution ADCs at the RX. Reference [24] suggests implementing fixed and low resolution ADCs with a small number of RF chains. Reference [25] works on the idea of a mixed-ADC architecture where a better energy-rate trade off is achieved by combining low and high resolution ADCs, but still with a fixed resolution for each ADC and without considering A/D hybrid beamforming. An A/D hybrid beamforming system with fixed and low resolution ADCs has been analyzed for channel estimation in [26].

One can implement varying resolution ADCs at the RX [27] which may provide a better solution than the RX with fixed and low resolution ADCs. Similarly, exploring low resolution DACs at the TX can also help reduce the power consumption. Thus, research that is focused on ADCs at the RX can also be applied to the TX DACs considering the TX specific system model parameters. Similar to using different ADC resolutions at the RX [27], which could provide a better solution than fixed low resolution ADCs, one can design a variable DAC resolution TX. Extra care is needed when deciding the number of bits used as the total DAC/ADC power consumption can be dominated by only a few high resolution DACs/ADCs. From [29], we notice that a good trade off between the power consumption and the performance may be to consider the range of 1-8 bits for I- and Q-channels, where 8-bit represents the full-bit resolution DACs/ADCs.

Reference [30] uses low resolution DACs for a single user MIMO system while [31] employs low resolution DACs at the base station for a narrowband multi-user MIMO system. Reference [32] also discusses fixed and low resolution DACs architecture for multi-user MIMO systems.



Reference [33] considers a single user MIMO system with quantized hybrid precoding including the RF quantized noise term beside the additive white Gaussian noise (AWGN) while evaluating EE and SE performance. The existing literature still does not consider adjusting the resolution associated with DACs/ADCs dynamically. It is possible to consider both the TX and the RX simultaneously where we can design an optimization problem to find the optimal number of quantized bits to achieve high EE performance. When designing for high EE, the complexity of the solution also needs to be taken into account while providing improvements over the existing literature.

B. Contributions

This paper designs an optimal EE solution for a mmWave A/D hybrid MIMO system by introducing a novel TX decomposition of the A/D hybrid precoder to three parts representing the analog precoder matrix, the DAC bit resolution matrix and the digital precoder matrix, respectively. A similar decomposition at the RX represents the analog combiner matrix, the ADC bit resolution matrix and the digital combiner matrix. Our aim is to minimize the distance between the decomposition, which is expressed as the product of three matrices, and the corresponding fully digital precoder or combiner matrix. The joint problem is decomposed into a series of sub-problems which are solved using the alternating direction method of multipliers (ADMM). We implement an exhaustive search approach [16] to evaluate the upper bound for EE maximization.

In [1], we addressed bit allocation and hybrid combining at the RX only, where we jointly optimized the number of ADC bits and hybrid combiner matrices for EE maximization. A novel decomposition of the hybrid combiner to three parts was introduced: the analog combiner matrix, the bit resolution matrix and the baseband combiner matrix, and these matrices were computed using the ADMM approach in order to solve the matrix factorization problem. In addition to [1], the main contributions of this paper can be listed as follows:

- This paper designs an optimal EE solution for a mmWave A/D hybrid beamforming MIMO system by introducing the novel matrix decomposition that is applied to the hybrid beamforming matrices at both the TX and the RX. This decomposition defines three matrices, which are the analog beamforming matrix, the bit resolution matrix and the baseband beamforming matrix at both the TX and the RX. These matrices are obtained by the solution of an EE maximization problem and the DAC/ADC bit resolution is adjusted dynamically unlike fixed bit resolution in the existing literature.

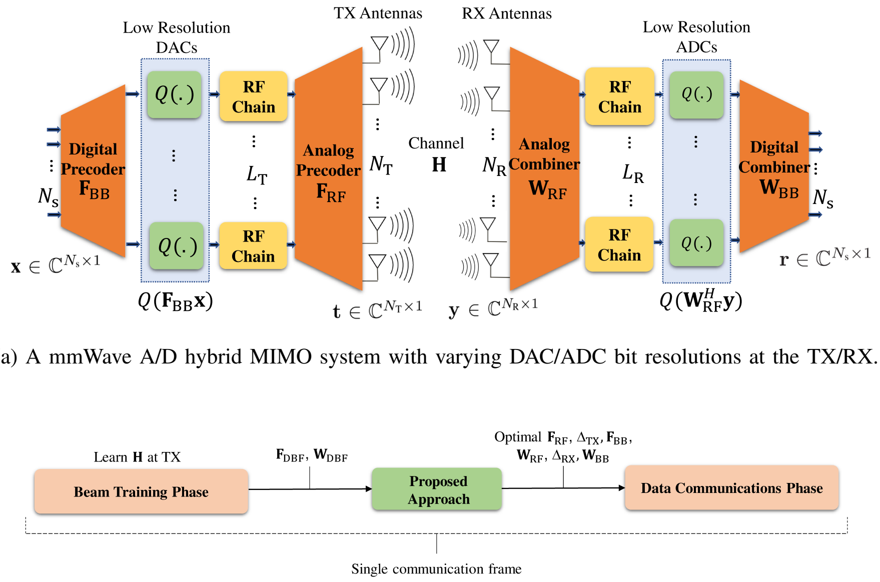
- The joint TX-RX problem is a difficult problem to solve due to non-convex constraints and non-convex cost functions. Firstly we address the joint TX-RX problem unlike in the existing literature. Then we decouple it into two sub-problems dealing with the TX and the RX separately, where the corresponding problems at the TX and the RX are solved by the alternating minimization technique such as ADMM [34] to obtain the unknown precoder/combiner and DAC/ADC bit resolution matrices.
- This work jointly optimizes the hybrid beamforming and DAC/ADC bit resolution matrices, unlike the existing approaches that optimize either DAC/ADC bit resolution or hybrid beamforming matrices. Moreover, the proposed design has high flexibility, given that the analog precoder/combiner is codebook-free, thus there is no restriction on the angular vectors and different bit resolutions can be assigned to each DAC/ADC.

The performance of the proposed technique is investigated through extensive simulation results, achieving increased EE compared to the baseline techniques with fixed DAC/ADC bit resolutions and number of RF chains, and an exhaustive search based approach which is an upper bound for EE maximization.

C. Notation and Organization

$\mathbf{A}$, $\mathbf{a}$ and a stand for a matrix, a vector, and a scalar, respectively. The trace, transpose and complex conjugate transpose of $\mathbf{A}$ are denoted as tr$(\mathbf{A})$, $\mathbf{A}^T$ and $\mathbf{A}^H$, respectively; $\|\mathbf{A}\|_F$ represents the Frobenius norm of $\mathbf{A}$; $|a|$ represents the determinant of a; $\mathbf{I}_N$ represents $N \times N$ identity matrix; $\mathcal{CN}(\mathbf{a}; \mathbf{A})$ denotes a complex Gaussian vector having mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$; $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{R}^+$ denote the sets of complex numbers, real numbers and positive real numbers, respectively; $\mathbf{X} \in \mathbb{C}^{A\times B}$ and $\mathbf{X} \in \mathbb{R}^{A\times B}$ denote $A \times B$ size $\mathbf{X}$ matrix with complex and real entries, respectively; $[\mathbf{A}]_k$ denotes the k-th column of matrix $\mathbf{A}$ while $[\mathbf{A}]_{kl}$ the matrix entry at the k-th row and l-th column; the indicator function $\mathbb{1}_{\mathcal{S}}\left\{\mathbf{A}\right\}$ of a set $\mathcal{S}$ that acts over a matrix $\mathbf{A}$ is defined as $0 \,\forall\, \mathbf{A} \in \mathcal{S}$ and $\infty \,\forall\, \mathbf{A} \notin \mathcal{S}$.

Section II presents the channel and system models where the channel model is based on a mmWave channel setup and the system model defines the low resolution quantization at both the TX and the RX. Sections III and IV present the problem formulation for the proposed technique at the TX and the RX, respectively, and the solution to obtain an energy efficient system. Section V verifies the proposed technique through simulation results and Section VI concludes the paper.


Low Resolution DACs
Digital Precoder $\mathbf{F}_{\text{BB}}$
N_{s}
$\mathbf{x} \in \mathbb{C}^{N_{\text{s}}\times 1}$
$Q(.)$
$Q(\mathbf{F}_{\text{BB}}\mathbf{x})$
RF Chain
L_{T}
Analog Precoder $\mathbf{F}_{\text{RF}}$
TX Antennas
N_{T}
$\mathbf{t} \in \mathbb{C}^{N_{\text{T}}\times 1}$
Channel $\mathbf{H}$
RX Antennas
N_{R}
$\mathbf{y} \in \mathbb{C}^{N_{\text{R}}\times 1}$
Analog Combiner $\mathbf{W}_{\text{RF}}$
RF Chain
L_{R}
Low Resolution ADCs
$Q(.)$
$Q(\mathbf{W}_{\text{RF}}^H\mathbf{y})$
Digital Combiner $\mathbf{W}_{\text{BB}}$
N_{s}
$\mathbf{r} \in \mathbb{C}^{N_{\text{s}}\times 1}$
Learn $\mathbf{H}$ at TX
Beam Training Phase
$\mathbf{F}_{\text{DBF}}$, $\mathbf{W}_{\text{DBF}}$
Proposed Approach
Optimal $\mathbf{F}_{\text{RF}}$, Δ_{TX}, $\mathbf{F}_{\text{BB}}$, $\mathbf{W}_{\text{RF},i}$, Δ_{RX}, $\mathbf{W}_{\text{BB}}$
Data Communications Phase
Single communication frame

(a) A mmWave A/D hybrid MIMO system with varying DAC/ADC bit resolutions at the TX/RX.

(b) Block diagram of the beam tracking phase and the data communications phase.

Fig. 1: System model for mmWave hybrid MIMO with varying DAC/ADC bit resolution.

II. MmWave A/D Hybrid MIMO System

A. MmWave Channel Model

MmWave channels can be modeled by a narrowband clustered channel model due to different channel settings such as the number of multipaths, amplitudes, etc., with N_{cl} clusters and N_{ray} propagation paths in each cluster [6]. Considering a single user mmWave system with N_{T} antennas at the TX, transmitting N_{s} data streams to N_{R} antennas at the RX, the mmWave channel matrix can be written as follows:

$$\mathbf{H} = \sqrt{\frac{N_{\text{T}}N_{\text{R}}}{N_{\text{cl}}N_{\text{ray}}}} \sum_{i=1}^{N_{\text{cl}}} \sum_{l=1}^{N_{\text{ray}}} \alpha_{il}\mathbf{a}_{\text{R}}(\phi_{il}^r)\mathbf{a}_{\text{T}}(\phi_{il}^t)^H, \tag{1}$$

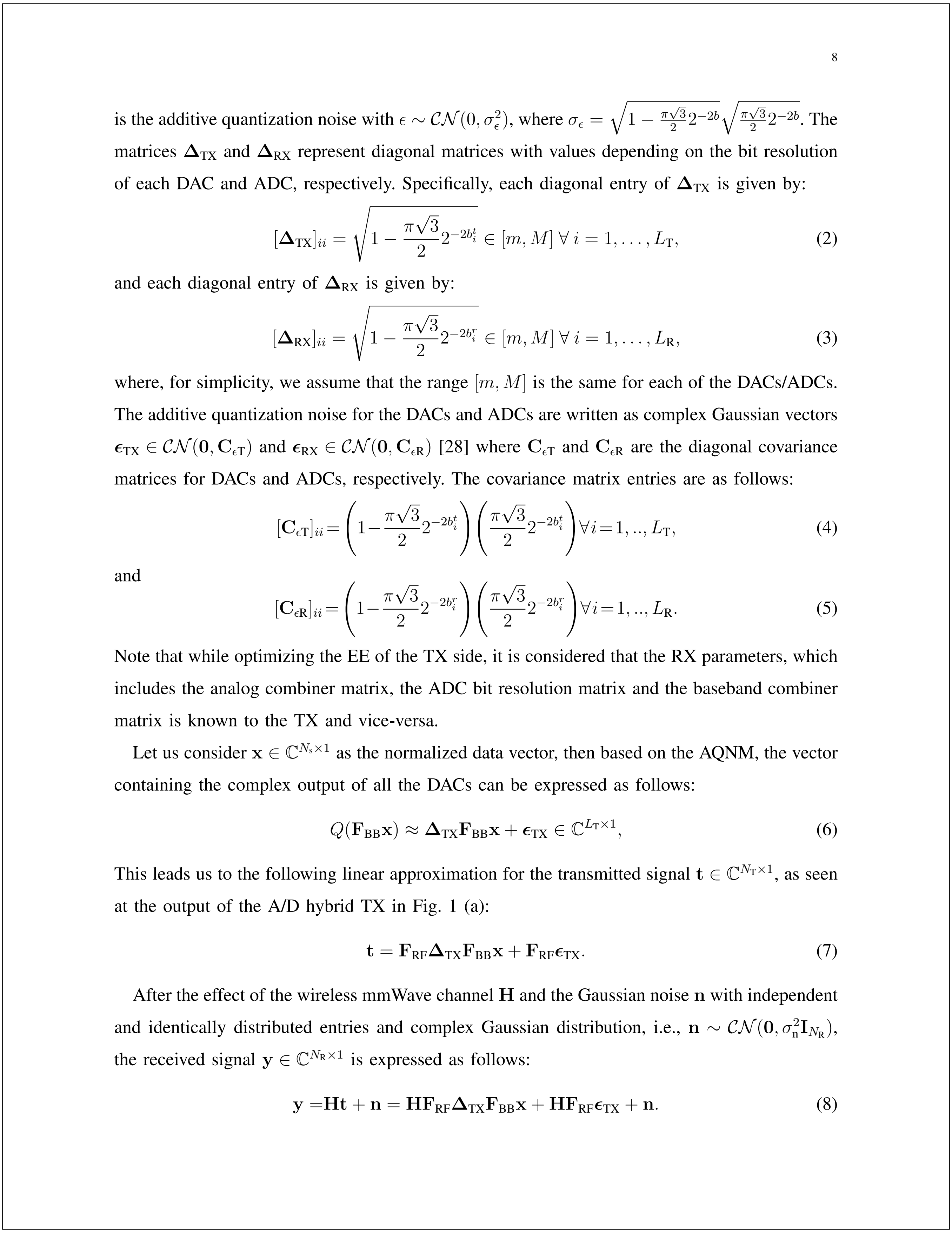
where $\alpha_{il} \in \mathcal{CN}(0, \sigma_{\alpha,i}^2)$ is the gain term with $\sigma_{\alpha,i}^2$ being the average power of the i^{th} cluster. Furthermore, $\mathbf{a}_{\text{T}}(\phi_{il}^t)$ and $\mathbf{a}_{\text{R}}(\phi_{il}^r)$ represent the normalized transmit and receive array response vectors [6], where ϕ_{il}^t and ϕ_{il}^r denote the azimuth angles of departure and arrival, respectively. We use uniform linear array (ULA) antennas for simplicity and model the antenna elements at the RX as ideal sectored elements [35]. We assume that the channel state information (CSI) is known at both the TX and the RX.

B. A/D Hybrid MIMO System Model

Based on the A/D hybrid beamforming scheme in the large scale mmWave MIMO communication systems, the number of TX RF chains L_T follows the limitation $N_\text{s} \leq L_\text{T} \leq N_\text{T}$ and similarly for L_R RF chains at the RX, $N_\text{s} \leq L_\text{R} \leq N_\text{R}$ [6], [7]. As shown in Fig. 1 (a), the matrices $\mathbf{F}_\text{RF} \in \mathbb{C}^{N_\text{T} \times L_\text{T}}$ and $\mathbf{F}_\text{BB} \in \mathbb{C}^{L_\text{T} \times N_\text{s}}$ denote the analog precoder and baseband precoder matrices, respectively. Similarly, the matrices $\mathbf{W}_\text{RF} \in \mathbb{C}^{N_\text{R} \times L_\text{R}}$ and $\mathbf{W}_\text{BB} \in \mathbb{C}^{L_\text{R} \times N_\text{s}}$ denote the analog combiner and baseband combiner matrices, respectively. The analog precoder and combiner matrices, $\mathbf{F}_\text{RF}$ and $\mathbf{W}_\text{RF}$, are based on phase shifters, i.e., the elements that have unit modulus and continuous phase. Thus, $\mathbf{F}_\text{RF} \in \mathcal{F}^{N_\text{T} \times L_\text{T}}$ and $\mathbf{W}_\text{RF} \in \mathcal{W}^{N_\text{R} \times L_\text{R}}$ where the set $\mathcal{F}$ and $\mathcal{W}$ represent the set of possible phase shifts in $\mathbf{F}_\text{RF}$ and $\mathbf{W}_\text{RF}$, respectively. The sets $\mathcal{F}$ and $\mathcal{W}$ for variables f and w, respectively, are defined as $\mathcal{F} = \{f \in \mathbb{C} \mid |f| = 1\}$ and $\mathcal{W} = \{w \in \mathbb{C} \mid |w| = 1\}$.

Note that, we optimize the DAC and ADC resolution and the precoder and combiner matrices at the TX and the RX on a frame-by-frame basis. As shown in Fig. 1 (b), we consider two stages in the system model: i) the beam training phase, and ii) the data communications phase. In stage i), firstly, the channel $\mathbf{H}$ is computed which provides us the optimal beamforming matrices, i.e., $\mathbf{F}_\text{DBF}$ at the TX and $\mathbf{W}_\text{DBF}$ at the RX. In stage ii), the optimal precoding and DAC bit resolution matrices $\mathbf{F}_\text{RF}$, $\mathbf{F}_\text{BB}$ and $\mathbf{\Delta}_\text{TX}$ at the TX, respectively, and the optimal combining and ADC bit resolution matrices $\mathbf{W}_\text{RF}$, $\mathbf{W}_\text{BB}$ and $\mathbf{\Delta}_\text{RX}$ at the RX are obtained. These two phases consist of one communication frame where the frame duration is smaller than the channel coherence time. Furthermore, if we assume that the TX/RX is active for stage i) a small proportion of time, for example, $< 10\%$, then the overall transmit energy consumption is dominated by stage ii).

We consider the linear AQNM to represent the distortion of quantization [21]. Given that $Q(\cdot)$ denotes a uniform scalar quantizer then for the scalar complex input $x \in \mathbb{C}$ that is applied to both the real and imaginary parts, we have, $Q(x) \approx \delta x + \epsilon$, where $\delta = \sqrt{1 - \frac{\pi\sqrt{3}}{2} 2^{-2b}} \in [m, M]$ is the multiplicative distortion parameter for a bit resolution equal to b [38], where m and M denote the minimum and maximum value of the range. The resolution parameter b is denoted as $b_i^t \,\forall\, i = 1, \ldots, L_\text{T}$ and $b_i^r \,\forall\, i = 1, \ldots, L_\text{R}$ at the TX and the RX, respectively. Note that the introduced error in the above linear approximation decreases for larger resolutions. However, our proposed solution focuses on EE maximization and this linear approximation does not impact the performance significantly as observed from the simulation results in Section V. The parameter ϵ



is the additive quantization noise with $\epsilon \sim \mathcal{CN}(0,\sigma_\epsilon^2)$, where $\sigma_\epsilon = \sqrt{1-\frac{\pi\sqrt{3}}{2}2^{-2b}}\sqrt{\frac{\pi\sqrt{3}}{2}2^{-2b}}$. The matrices $\mathbf{\Delta}_{\text{TX}}$ and $\mathbf{\Delta}_{\text{RX}}$ represent diagonal matrices with values depending on the bit resolution of each DAC and ADC, respectively. Specifically, each diagonal entry of $\mathbf{\Delta}_{\text{TX}}$ is given by:

$$[\mathbf{\Delta}_{\text{TX}}]_{ii} = \sqrt{1-\frac{\pi\sqrt{3}}{2}2^{-2b_i^t}} \in [m, M] \;\forall\, i = 1,\ldots,L_{\text{T}}, \tag{2}$$

and each diagonal entry of $\mathbf{\Delta}_{\text{RX}}$ is given by:

$$[\mathbf{\Delta}_{\text{RX}}]_{ii} = \sqrt{1-\frac{\pi\sqrt{3}}{2}2^{-2b_i^r}} \in [m, M] \;\forall\, i = 1,\ldots,L_{\text{R}}, \tag{3}$$

where, for simplicity, we assume that the range $[m, M]$ is the same for each of the DACs/ADCs. The additive quantization noise for the DACs and ADCs are written as complex Gaussian vectors $\boldsymbol{\epsilon}_{\text{TX}} \in \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\epsilon\text{T}})$ and $\boldsymbol{\epsilon}_{\text{RX}} \in \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\epsilon\text{R}})$ [28] where $\mathbf{C}_{\epsilon\text{T}}$ and $\mathbf{C}_{\epsilon\text{R}}$ are the diagonal covariance matrices for DACs and ADCs, respectively. The covariance matrix entries are as follows:

$$[\mathbf{C}_{\epsilon\text{T}}]_{ii} = \left(1-\frac{\pi\sqrt{3}}{2}2^{-2b_i^t}\right)\left(\frac{\pi\sqrt{3}}{2}2^{-2b_i^t}\right)\forall i = 1,..,L_{\text{T}}, \tag{4}$$

and

$$[\mathbf{C}_{\epsilon\text{R}}]_{ii} = \left(1-\frac{\pi\sqrt{3}}{2}2^{-2b_i^r}\right)\left(\frac{\pi\sqrt{3}}{2}2^{-2b_i^r}\right)\forall i = 1,..,L_{\text{R}}. \tag{5}$$

Note that while optimizing the EE of the TX side, it is considered that the RX parameters, which includes the analog combiner matrix, the ADC bit resolution matrix and the baseband combiner matrix is known to the TX and vice-versa.

Let us consider $\mathbf{x} \in \mathbb{C}^{N_s \times 1}$ as the normalized data vector, then based on the AQNM, the vector containing the complex output of all the DACs can be expressed as follows:

$$Q(\mathbf{F}_{\text{BB}}\mathbf{x}) \approx \mathbf{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}\mathbf{x} + \boldsymbol{\epsilon}_{\text{TX}} \in \mathbb{C}^{L_{\text{T}} \times 1}, \tag{6}$$

This leads us to the following linear approximation for the transmitted signal $\mathbf{t} \in \mathbb{C}^{N_{\text{T}} \times 1}$, as seen at the output of the A/D hybrid TX in Fig. 1 (a):

$$\mathbf{t} = \mathbf{F}_{\text{RF}}\mathbf{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}\mathbf{x} + \mathbf{F}_{\text{RF}}\boldsymbol{\epsilon}_{\text{TX}}. \tag{7}$$

After the effect of the wireless mmWave channel $\mathbf{H}$ and the Gaussian noise $\mathbf{n}$ with independent and identically distributed entries and complex Gaussian distribution, i.e., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\text{n}}^2\mathbf{I}_{N_{\text{R}}})$, the received signal $\mathbf{y} \in \mathbb{C}^{N_{\text{R}} \times 1}$ is expressed as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{t} + \mathbf{n} = \mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}\mathbf{x} + \mathbf{H}\mathbf{F}_{\text{RF}}\boldsymbol{\epsilon}_{\text{TX}} + \mathbf{n}. \tag{8}$$

When the analog combiner matrix $\mathbf{W}_{\text{RF}}$ and ADC quantization based on AQNM are applied to the received signal $\mathbf{y}$, we obtain the following:

$$Q(\mathbf{W}_{\text{RF}}^{H}\mathbf{y}) \approx \boldsymbol{\Delta}_{\text{RX}}^{H}\mathbf{W}_{\text{RF}}^{H}\mathbf{y} + \boldsymbol{\epsilon}_{\text{RX}} \in \mathbb{C}^{L_{\text{R}}\times 1}. \tag{9}$$

After the application of the baseband combiner matrix $\mathbf{W}_{\text{BB}}$, the output signal $\mathbf{r} \in \mathbb{C}^{N_{\text{s}}\times 1}$ at the RX, as shown in Fig. 1 (a), can be expressed as follows:

$$\mathbf{r} = \mathbf{W}_{\text{BB}}^{H}\boldsymbol{\Delta}_{\text{RX}}^{H}\mathbf{W}_{\text{RF}}^{H}\mathbf{y} + \mathbf{W}_{\text{BB}}^{H}\boldsymbol{\epsilon}_{\text{RX}}. \tag{10}$$

Considering the A/D hybrid precoder matrix $\mathbf{F} = \mathbf{F}_{\text{RF}}\boldsymbol{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_{\text{T}}\times N_{\text{s}}}$ and the A/D hybrid combiner matrix $\mathbf{W} = \mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_{\text{R}}\times N_{\text{s}}}$, we can express the RX output signal $\mathbf{r}$ in (10) as follows:

$$\mathbf{r} = \mathbf{W}^{H}\mathbf{H}\mathbf{F}\mathbf{x} + \underbrace{\mathbf{W}^{H}\mathbf{H}\mathbf{F}_{\text{RF}}\boldsymbol{\epsilon}_{\text{TX}} + \mathbf{W}_{\text{BB}}^{H}\boldsymbol{\epsilon}_{\text{RX}} + \mathbf{W}^{H}\mathbf{n}}_{\boldsymbol{\eta}}, \tag{11}$$

where $\boldsymbol{\eta}$ is the combined effect of the additive white Gaussian RX noise and quantization noise that has covariance matrix, $\mathbf{R}_{\eta} \in \mathbb{C}^{N_{\text{s}}\times N_{\text{s}}}$, given by,

$$\mathbf{R}_{\eta} = \mathbf{W}^{H}\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{C}_{\epsilon\text{T}}\mathbf{F}_{\text{RF}}^{H}\mathbf{H}^{H}\mathbf{W} + \mathbf{W}_{\text{BB}}^{H}\mathbf{C}_{\epsilon\text{R}}\mathbf{W}_{\text{BB}} + \sigma_{\text{n}}^{2}\mathbf{W}^{H}\mathbf{W}. \tag{12}$$

In the following sections, we discuss the joint optimization solution to compute the optimal DAC/ADC bit resolution matrices and the optimal precoder/combiner matrices.

III. Joint DAC Bit Allocation and A/D Hybrid Precoding Design

Let us consider a point-to-point MIMO system with a linear quantization model. We define the EE as the ratio of the information rate R, i.e. SE, and the total consumed power P [39] as:

$$EE \triangleq \frac{R}{P} \text{ (bits/Hz/J)}. \tag{13}$$

For the given point-to-point MIMO system, the SE is defined as,

$$R \triangleq \log_2 \left| \mathbf{I}_{N_{\text{s}}} + \frac{\mathbf{R}_{\eta}^{-1}}{N_{\text{s}}}\mathbf{W}^{H}\mathbf{H}\mathbf{F}\mathbf{F}^{H}\mathbf{H}^{H}\mathbf{W} \right| \text{ (bits/s/Hz)}, \tag{14}$$

where $\mathbf{F} = \mathbf{F}_{\text{RF}}\boldsymbol{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}$ and $\mathbf{W} = \mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}}$.

Similar to the power model at the TX in [28], the total consumed power for the system is expressed as:

$$P \triangleq P_{\text{TX}}(\mathbf{F}_{\text{RF}}, \boldsymbol{\Delta}_{\text{TX}}, \mathbf{F}_{\text{BB}}) + P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}) \text{ (W)}, \tag{15}$$

where the power consumption at the TX is as follows:

$$P_{\text{TX}}(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}) = \text{tr}(\mathbf{F}\mathbf{F}^{H}) + P_{\text{DT}}(\boldsymbol{\Delta}_{\text{TX}}) + N_{\text{T}}P_{\text{T}} + N_{\text{T}}L_{\text{T}}P_{\text{PT}} + P_{\text{CT}} \text{ (W)}, \tag{16}$$

where P_{PT} is the power per phase shifter, P_{T} is the power per antenna element, $P_{\text{DT}}(\boldsymbol{\Delta}_{\text{TX}})$ is the power associated with the total quantization operation at the TX, and following (2) and [21], we have

$$P_{\text{DT}}(\boldsymbol{\Delta}_{\text{TX}}) = P_{\text{DAC}}\sum_{i=1}^{L_{\text{T}}} 2^{b_i} = P_{\text{DAC}}\sum_{i=1}^{L_{\text{T}}} \left(\frac{\pi\sqrt{3}}{2(1-[\boldsymbol{\Delta}_{\text{TX}}]_{ii}^{2})}\right)^{\frac{1}{2}} \text{ (W)}, \tag{17}$$

where P_{DAC} is the power consumed per bit in the DAC and P_{CT} is the power required by all circuit components at the TX. Similarly, the total power consumption at the RX is,

$$P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}) = P_{\text{DR}}(\boldsymbol{\Delta}_{\text{RX}}) + N_{\text{R}}P_{\text{R}} + N_{\text{R}}L_{\text{R}}P_{\text{PR}} + P_{\text{CR}} \text{ (W)}, \tag{18}$$

where, at the RX, P_{PR} is the power per phase shifter, P_{R} is the power per antenna element, P_{DR} is the power associated with the total quantization operation, and following (3) and [21], we have

$$P_{\text{DR}}(\boldsymbol{\Delta}_{\text{RX}}) = P_{\text{ADC}}\sum_{i=1}^{L_{\text{R}}} 2^{b_i} = P_{\text{ADC}}\sum_{i=1}^{L_{\text{R}}} \left(\frac{\pi\sqrt{3}}{2(1-[\boldsymbol{\Delta}_{\text{RX}}]_{ii}^{2})}\right)^{\frac{1}{2}} \text{ (W)}, \tag{19}$$

where P_{ADC} is the power consumed per bit in the ADC and P_{CR} is the power required by all RX circuit components.

The maximization of EE is given by

$$\begin{aligned} &\max_{\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}},\mathbf{W}_{\text{RF}},\boldsymbol{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}} \frac{R(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}},\mathbf{W}_{\text{RF}},\boldsymbol{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}})}{P_{\text{TX}}(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}) + P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}})} \\ &\text{subject to } \mathbf{F}_{\text{RF}} \in \mathcal{F}^{N_{\text{T}}\times L_{\text{T}}}, \boldsymbol{\Delta}_{\text{TX}} \in \mathcal{D}_{\text{TX}}^{L_{\text{T}}\times L_{\text{T}}}, \mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}, \boldsymbol{\Delta}_{\text{RX}} \in \mathcal{D}_{\text{RX}}^{L_{\text{R}}\times L_{\text{R}}}, \end{aligned} \tag{20}$$

when the SE R is given by (14) and the power P in (15). The problem to be addressed involves a fractional cost function that both the numerator and the denominator parts are non-convex functions of the optimizing variables. Furthermore the optimization problem involves non-convex constraint sets. Thus, it is in general a very difficult problem to be addressed. It is interesting that the corresponding problem for a fully digital transceiver that admits a much simpler form is in general intractable due to the coupling of the TX-RX design [40]. To that end, we start by decoupling the TX-RX design problem.

Let us first express the EE maximization problem in the following relaxed form:

$$\begin{aligned}\min_{\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}},\mathbf{W}_{\text{RF}},\boldsymbol{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}} \quad & -R(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}},\mathbf{W}_{\text{RF}},\boldsymbol{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}) \\ & + \gamma_{\text{T}} P_{\text{TX}}(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}) + \gamma_{\text{R}} P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}) \\ \text{subject to } & \mathbf{F}_{\text{RF}} \in \mathcal{F}^{N_{\text{T}}\times L_{\text{T}}}, \boldsymbol{\Delta}_{\text{TX}} \in \mathcal{D}_{\text{TX}}^{L_{\text{T}}\times L_{\text{T}}}, \mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}, \boldsymbol{\Delta}_{\text{RX}} \in \mathcal{D}_{\text{RX}}^{L_{\text{R}}\times L_{\text{R}}}, \quad (21)\end{aligned}$$

where the parameters $\gamma_{\text{T}} \in (0, \gamma_{\text{T}}^{max}] \subset \mathbb{R}^+$ and $\gamma_{\text{R}} \in (0, \gamma_{\text{R}}^{max}] \subset \mathbb{R}^+$ are introducing a trade-off between the achieved rate and the power consumption at the TX's and the RX's side, respectively. Such an approach has been used in the past to tackle fractional optimization problems [41]. In the concave/convex case, the equivalence of the relaxed problem with the original fractional one is theoretically established. Unfortunately, a similar result for the case considered in the present paper is not easy to be derived due to the complexity of the addressed problem. Thus, in the present paper, we rely on line search methods in order to optimally tune these parameters.

Having simplified the original problem, we may now proceed by temporally decoupling the designs at the TX's and the RX's side. Under the assumption that the RX can perform optimal nearest-neighbor decoding based on the received signals, the optimal precoding matrices are designed such that the mutual information achieved by Gaussian signaling over the wireless channel is maximized [6]. The mutual information is given by

$$I \triangleq \log_2 \left| \mathbf{I}_{N_{\text{s}}} + \frac{\mathbf{Q}_{\eta'}^{-1}}{N_{\text{s}}} \mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H \right| \text{ (bits/s/Hz)}, \quad (22)$$

where again $\mathbf{F} = \mathbf{F}_{\text{RF}}\boldsymbol{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}$ and and $\mathbf{Q}_{\eta'}$ is the covariance matrix of the sum of noise and transmit quantization noise variables, i.e. $\eta' = \mathbf{F}_{\text{RF}}\boldsymbol{\epsilon}_{\text{TX}} + \mathbf{n}$, given by

$$\mathbf{Q}_{\eta'} = \mathbf{F}_{\text{RF}}\mathbf{C}_{\epsilon\text{T}}\mathbf{F}_{\text{RF}}^H + \sigma_{\text{n}}^2\mathbf{I}_{N_{\text{R}}}. \quad (23)$$

Based on (21)-(22), the precoding matrices may be derived as the solution to the following optimization problem:

$$\begin{aligned}(\mathcal{P}_{1\text{T}}): \min_{\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}} \quad & -I(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}) + \gamma_T P_{\text{TX}}(\mathbf{F}_{\text{RF}},\boldsymbol{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}), \\ & \text{subject to } \mathbf{F}_{\text{RF}} \in \mathcal{F}^{N_{\text{T}}\times L_{\text{T}}}, \boldsymbol{\Delta}_{\text{TX}} \in \mathcal{D}_{\text{TX}}^{L_{\text{T}}\times L_{\text{T}}},\end{aligned}$$

Now provided that the optimal precoding matrix $\mathbf{F}^{\star} = \mathbf{F}_{\text{RF}}^{\star}\mathbf{\Delta}_{\text{TX}}^{\star}\mathbf{F}_{\text{BB}}^{\star}$ is derived from solving $(\mathcal{P}_{1\text{T}})$, we can plug in these resulted precoding matrices in the cost function of (21) resulting in an optimization problem dependent only on the decoder matrices at the RX's side, defined as,

$$(\mathcal{P}_{1\text{R}}): \min_{\mathbf{W}_{\text{RF}},\mathbf{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}} -\tilde{R}(\mathbf{W}_{\text{RF}},\mathbf{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}) + \gamma_R P_{RX}(\mathbf{\Delta}_{\text{RX}})$$
$$\text{subject to } \mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}, \mathbf{\Delta}_{\text{RX}} \in \mathcal{D}_{\text{RX}}^{L_{\text{R}}\times L_{\text{R}}}, \tag{24}$$

where $\tilde{R}(\mathbf{W}_{\text{RF}},\mathbf{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}}) = R(\mathbf{F}_{\text{RF}}^{\star},\mathbf{\Delta}_{\text{TX}}^{\star},\mathbf{F}_{\text{BB}}^{\star},\mathbf{W}_{\text{RF}},\mathbf{\Delta}_{\text{RX}},\mathbf{W}_{\text{BB}})$.

Thus, the precoding and decoding matrices can be derived as the solutions to the two decoupled problems $(\mathcal{P}_{1\text{T}}) - (\mathcal{P}_{1\text{R}})$ above. In the following subsections, the solutions to these problems are developed. We start first with the development of the solution to TX's side one $(\mathcal{P}_{1\text{T}})$ and then the solution for the RX's side $(\mathcal{P}_{1\text{R}})$ counterpart follows.

A. Problem Formulation at the TX

Focusing on the TX side, we seek the bit resolution matrix $\mathbf{\Delta}_{\text{TX}}$ and the hybrid precoding matrices $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$ that solve $(\mathcal{P}_{1\text{T}})$. The set $\mathcal{D}_{\text{TX}}$ represents the finite states of the quantizer and is defined as,

$$\mathcal{D}_{\text{TX}} = \left\{\mathbf{\Delta}_{\text{TX}} \in \mathbb{R}^{L_{\text{T}}\times L_{\text{T}}} \middle| m \le [\mathbf{\Delta}_{\text{TX}}]_{ii} \le M \;\forall\, i = 1, ..., L_{\text{T}}\right\}.$$

Note that $P_{\text{TX}}(\mathbf{F}_{\text{RF}},\mathbf{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}) > 0$, as defined in (16), since the power required by all circuit components is always larger than zero, i.e., $P_{\text{CP}} > 0$.

Since dealing with the part of the cost function of $(\mathcal{P}_{1\text{T}})$ that involves the mutual information expression is a difficult task due to the perplexed form of the latter, we adopt the approach in [6] where the maximization of the mutual information I can be approximated by finding the minimum Euclidean distance of the hybrid precoder to the one of the fully digital transceiver for the full-bit resolution sampling case, denoted by $\mathbf{F}_{\text{DBF}}$, i.e., $||\mathbf{F}_{\text{DBF}} - \mathbf{F}_{\text{RF}}\mathbf{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}||_F^2$ [6]. Therefore, motivated by the previous, $(\mathcal{P}_{1\text{T}})$ can be approximated to finding the solution of the following problem:

$$(\mathcal{P}_2): \min_{\mathbf{F}_{\text{RF}},\mathbf{\Delta}_{\text{TX}},\mathbf{F}_{\text{BB}}} \frac{1}{2}||\mathbf{F}_{\text{DBF}} - \mathbf{F}_{\text{RF}}\mathbf{\Delta}_{\text{TX}}\mathbf{F}_{\text{BB}}||_F^2 + \gamma_{\text{T}} P_{\text{TX}}(\mathbf{F}),$$
$$\text{subject to } \mathbf{F}_{\text{RF}} \in \mathcal{F}^{N_{\text{T}}\times L_{\text{T}}}, \mathbf{\Delta}_{\text{TX}} \in \mathcal{D}_{\text{TX}}^{L_{\text{T}}\times L_{\text{T}}}.$$

For a point-to-point MIMO system the optimal $\mathbf{F}_{\text{DBF}}$ is given by $\mathbf{F}_{\text{DBF}} = \mathbf{V}\sqrt{\mathbf{P}}$ where the orthonormal matrix $\mathbf{V} \in \mathbb{C}^{N_{\text{R}}\times N_{\text{T}}}$ is derived via the channel matrix singular value decomposition

(SVD), i.e. $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ and $\mathbf{P}$ is a diagonal power allocation matrix with real positive diagonal entries derived by the so-called "water-filling algorithm" [42].

Problem $(\mathcal{P}_2)$ is still very difficult to address as it is non-convex due to the non-convex cost function that involves the product of three matrix variables and non-convex constraints. In the next section, an efficient algorithmic solution based on the ADMM is proposed.

B. Proposed ADMM Solution at the TX

In the following we develop an iterative procedure for solving $(\mathcal{P}_2)$ based on the ADMM approach [34]. This method is a variant of the standard augmented Lagrangian method that uses partial updates (similar to the Gauss-Seidel method for the solution of linear equations) to solve constrained optimization problems. While it is mainly known for its good performance for a number of convex optimization problems, recently it has been successfully applied to non-convex matrix factorization as well [34], [43], [44]. Motivated by this, in the following ADMM based solutions are developed that are tailored for the non-convex matrix factorization problem $(\mathcal{P}_2)$.

We first transform $(\mathcal{P}_2)$ into a form that can be addressed via ADMM. By using the auxiliary variable $\mathbf{Z}$, $(\mathcal{P}_2)$ can be written as:

$$(\mathcal{P}_3):\min_{\mathbf{Z},\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}}}\frac{1}{2}\|\mathbf{F}_{\mathrm{DBF}}-\mathbf{Z}\|_F^2+\mathbb{1}_{\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\mathbf{F}_{\mathrm{RF}}\}+\mathbb{1}_{\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\boldsymbol{\Delta}_{\mathrm{TX}}\}+\gamma_{\mathrm{T}}P_{\mathrm{TX}}(\mathbf{F}),$$
$$\text{subject to } \mathbf{Z}=\mathbf{F}_{\mathrm{RF}}\boldsymbol{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}.$$

Problem $(\mathcal{P}_3)$ formulates the A/D hybrid precoder matrix design as a matrix factorization problem. That is, the overall precoder $\mathbf{Z}$ is sought so that it minimizes the Euclidean distance to the optimal, fully digital precoder $\mathbf{F}_{\mathrm{DBF}}$ while supporting decomposition into three factors: the analog precoder matrix $\mathbf{F}_{\mathrm{RF}}$, the DAC bit resolution matrix $\boldsymbol{\Delta}_{\mathrm{TX}}$ and the digital precoder matrix $\mathbf{F}_{\mathrm{BB}}$. The augmented Lagrangian function of $(\mathcal{P}_3)$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{Z},\mathbf{F}_{\mathrm{RF}},\boldsymbol{\Delta}_{\mathrm{TX}},\mathbf{F}_{\mathrm{BB}},\boldsymbol{\Lambda})=&\frac{1}{2}\|\mathbf{F}_{\mathrm{DBF}}-\mathbf{Z}\|_F^2+\mathbb{1}_{\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\mathbf{F}_{\mathrm{RF}}\}+\mathbb{1}_{\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\boldsymbol{\Delta}_{\mathrm{TX}}\}\\&+\frac{\alpha}{2}\|\mathbf{Z}+\boldsymbol{\Lambda}/\alpha-\mathbf{F}_{\mathrm{RF}}\boldsymbol{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}}\|_F^2+\gamma_{\mathrm{T}}P_{\mathrm{TX}}(\mathbf{F}),\end{aligned}\tag{25}$$

where α is a scalar penalty parameter and $\boldsymbol{\Lambda}\in\mathbb{C}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}$ is the Lagrange Multiplier matrix. According to the ADMM approach [34], the solution to $(\mathcal{P}_3)$ is derived by the following iterative steps where n denotes the iteration index:

$$(\mathcal{P}_{3\mathrm{A}}):\mathbf{Z}_{(n)}=\arg\min_{\mathbf{Z}}\mathcal{L}(\mathbf{Z},\mathbf{F}_{\mathrm{RF}(n-1)},\boldsymbol{\Delta}_{\mathrm{TX}(n-1)},\mathbf{F}_{\mathrm{BB}(n-1)},\boldsymbol{\Lambda}_{(n-1)}),$$

$$
\begin{aligned}
&(\mathcal{P}_{3\mathrm{B}}): \mathbf{F}_{\mathrm{RF}(n)} = \arg\min_{\mathbf{F}_{\mathrm{RF}}} \mathcal{L}(\mathbf{Z}_{(n)}, \mathbf{F}_{\mathrm{RF}}, \boldsymbol{\Delta}_{\mathrm{TX}(n-1)}, \mathbf{F}_{\mathrm{BB}(n-1)}, \boldsymbol{\Lambda}_{(n-1)}),\\
&(\mathcal{P}_{3\mathrm{C}}): \boldsymbol{\Delta}_{\mathrm{TX}(n)} = \arg\min_{\boldsymbol{\Delta}_{\mathrm{TX}}} \mathcal{L}(\mathbf{Z}_{(n)}, \mathbf{F}_{\mathrm{RF}(n)}, \boldsymbol{\Delta}_{\mathrm{TX}}, \mathbf{F}_{\mathrm{BB}(n-1)}, \boldsymbol{\Lambda}_{(n-1)}) + \gamma_{\mathrm{T}} P_{\mathrm{TX}}(\mathbf{F}),\\
&(\mathcal{P}_{3\mathrm{D}}): \mathbf{F}_{\mathrm{BB}(n)} = \arg\min_{\mathbf{F}_{\mathrm{BB}}} \mathcal{L}(\mathbf{Z}_{n}, \mathbf{F}_{\mathrm{RF}(n)}, \boldsymbol{\Delta}_{\mathrm{TX}(n)}, \mathbf{F}_{\mathrm{BB}}, \boldsymbol{\Lambda}_{(n-1)}),\\
&\boldsymbol{\Lambda}_{(n)} = \boldsymbol{\Lambda}_{(n-1)} + \alpha\left(\mathbf{Z}_{(n)} - \mathbf{F}_{\mathrm{RF}(n)}\boldsymbol{\Delta}_{\mathrm{TX}(n)}\mathbf{F}_{\mathrm{BB}(n)}\right). \qquad (26)
\end{aligned}
$$

In order to apply the ADMM iterative procedure, we have to solve the optimization problems $(\mathcal{P}_{3\mathrm{A}})$-$(\mathcal{P}_{3\mathrm{D}})$. We may start from problem $(\mathcal{P}_{3\mathrm{A}})$ which can be written as follows:

$$(\mathcal{P}'_{3\mathrm{A}}): \ \mathbf{Z}_{(n)} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|(1+\alpha)\mathbf{Z} - \mathbf{F}_{\mathrm{DBF}} + \boldsymbol{\Lambda}_{(n-1)} - \alpha\mathbf{F}_{\mathrm{RF}(n-1)}\boldsymbol{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\|_F^2.$$

Problem $(\mathcal{P}'_{3\mathrm{A}})$ can be directly solved by equating the gradient of the augmented Lagrangian (25) w.r.t. $\mathbf{Z}$ being set to zero. Therefore, we have

$$\mathbf{Z}_{(n)} = \frac{1}{\alpha+1}\Big(\mathbf{F}_{\mathrm{DBF}} - \boldsymbol{\Lambda}_{(n-1)} + \alpha\mathbf{F}_{\mathrm{RF}(n-1)}\boldsymbol{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\Big). \qquad (27)$$

We may now proceed to solve $(\mathcal{P}_{3\mathrm{B}})$ which can be written in the following simplified form by keeping only the terms of the augmented Lagrangian that are dependent on $\mathbf{F}_{\mathrm{RF}}$:

$$(\mathcal{P}'_{3\mathrm{B}}): \ \mathbf{F}_{\mathrm{RF}(n)} = \arg\min_{\mathbf{F}_{\mathrm{RF}}} \mathbb{1}_{\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}}\{\mathbf{F}_{\mathrm{RF}}\} + \frac{\alpha}{2}\|\mathbf{Z}_{(n)} + \boldsymbol{\Lambda}_{(n-1)}/\alpha - \mathbf{F}_{\mathrm{RF}}\boldsymbol{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\|_F^2.$$

The solution to problem $(\mathcal{P}'_{3B})$ does not admit a closed form and thus, it is approximated by solving the unconstrained problem and then projecting onto the set $\mathcal{F}^{N_{\mathrm{T}}\times L_{\mathrm{T}}}$, i.e.,

$$\mathbf{F}_{\mathrm{RF}(n)} = \Pi_{\mathcal{F}}\Big\{\left(\boldsymbol{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\right)\mathbf{F}^{H}_{\mathrm{BB}(n-1)}\boldsymbol{\Delta}^{H}_{\mathrm{TX}(n-1)}\left(\alpha\boldsymbol{\Delta}_{\mathrm{TX}(n-1)}\mathbf{F}_{\mathrm{BB}(n-1)}\mathbf{F}^{H}_{\mathrm{BB}(n-1)}\boldsymbol{\Delta}^{H}_{\mathrm{TX}(n-1)}\right)^{-1}\Big\}, \qquad (28)$$

where $\Pi_{\mathcal{F}}$ projects the solution onto the set $\mathcal{F}$. This is computed by solving the following optimization problem [45]:

$$(\mathcal{P}''_{3\mathrm{B}}): \qquad \min_{\mathbf{A}_{\mathcal{F}}} \|\mathbf{A}_{\mathcal{F}} - \mathbf{A}\|_F^2, \text{subject to } \mathbf{A}_{\mathcal{F}} \in \mathcal{F},$$

where $\mathbf{A}$ is an arbitrary matrix and $\mathbf{A}_{\mathcal{F}}$ is its projection onto the set $\mathcal{F}$. The solution to $(\mathcal{P}''_{3\mathrm{B}})$ is given by the phase of the complex elements of $\mathbf{A}$. Thus, for $\mathbf{A}_{\mathcal{F}} = \Pi_{\mathcal{F}}\{\mathbf{A}\}$ we have

$$\mathbf{A}_{\mathcal{F}}(x,y) = \begin{cases} 0, & \mathbf{A}(x,y) = 0 \\ \frac{\mathbf{A}(x,y)}{|\mathbf{A}(x,y)|}, & \mathbf{A}(x,y) \neq 0 \end{cases}, \qquad (29)$$

Algorithm 1 Proposed ADMM Solution for the A/D Hybrid Precoder Design

```
Initialize: $\mathbf{Z}$, $\mathbf{F}_{\mathrm{RF}}$, $\boldsymbol{\Delta}_{\mathrm{TX}}$, $\mathbf{F}_{\mathrm{BB}}$ with random values, $\boldsymbol{\Lambda}$ with zeros, $\alpha = 1$ and $n = 1$
while The termination criteria of (31) are not met or $n \leq N_{\mathrm{max}}$ do
    Update $\mathbf{Z}_{(n)}$ using solution (27),
          $\mathbf{F}_{\mathrm{RF}(n)}$ using solution (28),
          $\boldsymbol{\Delta}_{\mathrm{TX}(n)}$ by solving $(\mathcal{P}''_{\mathrm{3C}})$ using CVX [48],
          $\mathbf{F}_{\mathrm{BB}(n)}$ using solution (30), and
          update $\boldsymbol{\Lambda}_{(n)}$ using solution (26).
    $n \leftarrow n + 1$
end while
return $\mathbf{F}^{\star}_{\mathrm{RF}}$, $\boldsymbol{\Delta}^{\star}_{\mathrm{TX}}$, $\mathbf{F}^{\star}_{\mathrm{BB}}$
```

where $\mathbf{A}_{\mathcal{F}}(x, y)$ and $\mathbf{A}(x, y)$ are the elements at the xth row-yth column of matrices $\mathbf{A}_{\mathcal{F}}$ and $\mathbf{A}$, respectively. While, this is an approximate solution, it turns out that it behaves remarkably well, as verified in the simulation results of Section V. This is due to the interesting property that ADMM is observed to converge even in cases where the alternating minimization steps are not carried out exactly [34]. There are theoretical results that support this statement [46], [47], though an exact analysis for the case considered here is beyond the scope of this paper.

In a similar manner, $(\mathcal{P}_{\mathrm{3C}})$ may be re-written as,

$$
\begin{aligned}
(\mathcal{P}'_{\mathrm{3C}}) : \boldsymbol{\Delta}_{\mathrm{TX}(n)} = \arg\min_{\boldsymbol{\Delta}_{\mathrm{TX}}} & \mathbb{1}_{\mathcal{D}_{\mathrm{TX}}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}}\{\boldsymbol{\Delta}_{\mathrm{TX}}\} + \frac{\alpha}{2}\|\mathbf{Z}_{(n)} + \boldsymbol{\Lambda}_{(n-1)}/\alpha - \mathbf{F}_{\mathrm{RF}(n)}\boldsymbol{\Delta}_{\mathrm{TX}}\mathbf{F}_{\mathrm{BB}(n-1)}\|_F^2 \\
& + \gamma_{\mathrm{T}} P_{\mathrm{TX}}(\mathbf{F}).
\end{aligned}
$$

To solve the above problem, we can write:

$$
\begin{aligned}
(\mathcal{P}''_{\mathrm{3C}}) : \boldsymbol{\Delta}_{\mathrm{TX}(n)} = & \arg\min_{\boldsymbol{\Delta}_{\mathrm{TX}}} \|\mathbf{y}_{\mathrm{c}} - \boldsymbol{\Psi}_{\mathrm{T}} \mathrm{vec}(\boldsymbol{\Delta}_{\mathrm{TX}})\|_2^2 + \gamma_{\mathrm{T}} P_{\mathrm{TX}}(\mathbf{F}), \\
& \text{subject to } \boldsymbol{\Delta}_{\mathrm{TX}} \in \mathcal{D}_{\mathrm{TX}},
\end{aligned}
$$

The minimization problem in $(\mathcal{P}''_{\mathrm{3C}})$ consists of $\mathbf{y}_c = \mathrm{vec}(\mathbf{Z}_n + \Lambda_{n-1}/\alpha)$, $\boldsymbol{\Psi}_{\mathrm{T}} = \mathbf{F}_{\mathrm{BB}(n-1)} \otimes \mathbf{F}_{\mathrm{RF}(n)}$ ($\otimes$ being the Khatri-Rao product) and is solved using CVX [48].

The solution of problem $(\mathcal{P}_{\mathrm{3D}})$ may be written in the following form:

$$
(\mathcal{P}'_{\mathrm{3D}}) : \ \mathbf{F}_{\mathrm{BB}(n)} = \arg\min_{\mathbf{F}_{\mathrm{BB}}} \frac{\alpha}{2}\|\mathbf{Z}_{(n)} + \boldsymbol{\Lambda}_{(n-1)}/\alpha - \mathbf{F}_{\mathrm{RF}(n)}\boldsymbol{\Delta}_{\mathrm{TX}(n)}\mathbf{F}_{\mathrm{BB}}\|_F^2.
$$

It is straightforward to see that the solution for $(\mathcal{P}'_{3\text{D}})$ can be obtained by equating the gradient to zero and solving the resulting equation w.r.t. the matrix variable $\mathbf{F}_{\text{BB}}$, i.e.,

$$\mathbf{F}_{\text{BB}(n)} = \left(\alpha \boldsymbol{\Delta}^{H}_{\text{TX}(n)} \mathbf{F}^{H}_{\text{RF}(n)} \mathbf{F}_{\text{RF}(n)} \boldsymbol{\Delta}_{\text{TX}(n)}\right)^{-1} \boldsymbol{\Delta}^{H}_{\text{TX}(n)} \mathbf{F}^{H}_{\text{RF}(n)} \left(\boldsymbol{\Lambda}_{(n-1)} + \alpha \mathbf{Z}_{(n)}\right). \tag{30}$$

Algorithm 1 provides the complete procedure to obtain the optimal analog precoder matrix $\mathbf{F}_{\text{RF}}$, the optimal bit resolution matrix $\boldsymbol{\Delta}_{\text{TX}}$ and the optimal baseband (or digital) precoder matrix $\mathbf{F}_{\text{BB}}$. It starts the alternating minimization procedure by initializing the entries of the matrices $\mathbf{Z}$, $\mathbf{F}_{\text{RF}}$, $\boldsymbol{\Delta}_{\text{TX}}$, $\mathbf{F}_{\text{BB}}$ with random values and the entries of the Lagrange multiplier matrix $\boldsymbol{\Lambda}$ with zeros. For iteration index n, $\mathbf{Z}_{(n)}$, $\mathbf{F}_{\text{RF}(n)}$, $\boldsymbol{\Delta}_{\text{TX}(n)}$ and $\mathbf{F}_{\text{BB}(n)}$ are updated using Step 3 which shows the steps to be used to obtain the matrices. A termination criterion related to either the maximum permitted number of iterations (N_{max}) is considered or the ADMM solution meeting the following criteria is considered:

$$\left\|\mathbf{Z}_{(n)} - \mathbf{Z}_{(n-1)}\right\|_F \leq \epsilon^z \ \& \ \|\mathbf{Z}_{(n)} - \mathbf{F}_{\text{RF}(n)} \boldsymbol{\Delta}_{\text{TX}(n)} \mathbf{F}_{\text{BB}(n)}\|_F \leq \epsilon^p, \tag{31}$$

where ϵ^z and ϵ^p are the corresponding tolerances. Upon convergence, the number of bits for each DAC is obtained by using (2) and quantizing to the nearest integer value. The optimal hybrid precoding matrices $\mathbf{F}^{\star}_{\text{RF}}$, $\boldsymbol{\Delta}^{\star}_{\text{TX}}$, $\mathbf{F}^{\star}_{\text{BB}}$ are obtained at the end of this algorithm.

Computational complexity analysis of Algorithm 1: When running Algorithm 1, mainly Step 3, while updating $\boldsymbol{\Delta}_{\text{TX}(n)}$ by solving $(\mathcal{P}''_{3\text{C}})$ using CVX, involves multiplication by $\boldsymbol{\Psi}_{\text{T}}$ whose dimensions are $L_{\text{T}}N_{\text{T}} \times N_{\text{s}}L_{\text{T}}$. In general, the solution of $(\mathcal{P}''_{3\text{C}})$ can be upper-bounded by $\mathcal{O}((L^2_{\text{T}}N_{\text{T}}N_{\text{s}})^3)$ which can be improved significantly by exploiting the structure of $\boldsymbol{\Psi}_{\text{T}}$.

In the following section, we discuss the joint optimization problem at the RX and the solution to obtain the analog combiner matrix $\mathbf{W}_{\text{RF}}$, the ADC bit resolution matrix $\boldsymbol{\Delta}_{\text{RX}}$ and the digital combiner matrix $\mathbf{W}_{\text{BB}}$.

IV. Joint ADC Bit Allocation and A/D Hybrid Combining Optimization

A. Problem Formulation at the RX

Let us now move to the derivation of the solution to $(\mathcal{P}_{1\text{R}})$. The set $\mathcal{D}_{\text{RX}}$ represents the finite states of the ADC quantizer and is defined as,

$$\mathcal{D}_{\text{RX}} = \left\{\boldsymbol{\Delta}_{\text{RX}} \in \mathbb{R}^{L_{\text{R}} \times L_{\text{R}}} \middle| m \leq [\boldsymbol{\Delta}_{\text{RX}}]_{ii} \leq M \ \forall \, i = 1, ..., L_{\text{R}}\right\}.$$

Due to the perplexed form of the function $\tilde{R}(\mathbf{W}_{\text{RF}}, \boldsymbol{\Delta}_{\text{RX}}, \mathbf{W}_{\text{BB}})$, we follow the same arguments the under of which we approximated $(\mathcal{P}_2)$ by $(\mathcal{P}_{1\text{T}})$, in order to approximate $(\mathcal{P}_{1\text{R}})$ by

$$(\mathcal{P}_5): \min_{\mathbf{W}_{\text{RF}}, \boldsymbol{\Delta}_{\text{RX}}, \mathbf{W}_{\text{BB}}} \frac{1}{2}\|\mathbf{W}_{\text{DBF}} - \mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}}\|_F^2 + \gamma_{\text{R}} P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}),$$
$$\text{subject to } \mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}} \times L_{\text{R}}}, \boldsymbol{\Delta}_{\text{RX}} \in \mathcal{D}_{\text{RX}}^{L_{\text{R}} \times L_{\text{R}}},$$

where $\mathbf{W}_{\text{DBF}}$ is the optimal solution for the fully digital RX which is given by $\mathbf{W}_{\text{DBF}} = \sqrt{\tilde{\mathbf{P}}}\tilde{\mathbf{U}}$, where $\tilde{\mathbf{U}} \in \mathbb{C}^{N_{\text{R}} \times N_{\text{s}}}$ is the orthonormal singular vector matrix which can be derived by the SVD of the equivalent channel matrix $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{F}^{\star} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^H$, and $\tilde{\mathbf{P}}$ is diagonal power allocation matrix. Problem $(\mathcal{P}_5)$ is also non-convex due to the non-convex cost function and non-convex set of constraints, as well, and for its solution an ADMM-based solution similar to the case of $(\mathcal{P}_2)$ is derived in the following subsection.

B. Proposed ADMM Solution at the RX

In the following we develop an iterative procedure for solving $(\mathcal{P}_5)$ based on ADMM [34]. We first transform $(\mathcal{P}_5)$ into an amenable form. By using the auxiliary variable $\mathbf{Z}$, $(\mathcal{P}_5)$ can be written as:

$$(\mathcal{P}_6): \min_{\mathbf{Z}, \mathbf{W}_{\text{RF}}, \boldsymbol{\Delta}_{\text{RX}}, \mathbf{W}_{\text{BB}}} \frac{1}{2}\|\mathbf{W}_{\text{DBF}} - \mathbf{Z}\|_F^2 + \mathbb{1}_{\mathcal{W}^{N_{\text{R}} \times L_{\text{R}}}}\{\mathbf{W}_{\text{RF}}\} + \mathbb{1}_{\mathcal{D}_{\text{RX}}^{L_{\text{R}} \times L_{\text{R}}}}\{\boldsymbol{\Delta}_{\text{RX}}\} + \gamma_{\text{R}} P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}),$$
$$\text{subject to } \mathbf{Z} = \mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}}.$$

Problem $(\mathcal{P}_6)$ formulates the A/D hybrid combiner matrix design as a matrix factorization problem. That is, the overall combiner $\mathbf{Z}$ is sought so that it minimizes the Euclidean distance to the optimal, fully digital combiner $\mathbf{W}_{\text{DBF}}$ while supporting the decomposition into the analog combiner matrix $\mathbf{W}_{\text{RF}}$, the quantization error matrix $\boldsymbol{\Delta}_{\text{RX}}$ and the digital combiner matrix $\mathbf{W}_{\text{BB}}$. The augmented Lagrangian function of $(\mathcal{P}_6)$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{Z}, \mathbf{W}_{\text{RF}}, \boldsymbol{\Delta}_{\text{RX}}, \mathbf{W}_{\text{BB}}, \boldsymbol{\Lambda}) =& \frac{1}{2}\|\mathbf{W}_{\text{DBF}} - \mathbf{Z}\|_F^2 + \mathbb{1}_{\mathcal{W}^{N_{\text{R}} \times L_{\text{R}}}}\{\mathbf{W}_{\text{RF}}\} + \mathbb{1}_{\mathcal{D}_{\text{RX}}^{L_{\text{R}} \times L_{\text{R}}}}\{\boldsymbol{\Delta}_{\text{RX}}\} \\ &+ \frac{\alpha}{2}\|\mathbf{Z} + \boldsymbol{\Lambda}/\alpha - \mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{\text{RX}}\mathbf{W}_{\text{BB}}\|_F^2 + \gamma_{\text{R}} P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}), \end{aligned} \tag{32}$$

where α is a scalar penalty parameter and $\boldsymbol{\Lambda} \in \mathbb{C}^{N_{\text{R}} \times L_{\text{R}}}$ is the Lagrange Multiplier matrix. According to the ADMM approach [34], the solution to $(\mathcal{P}_6)$ is derived by the following iterative steps:

$$(\mathcal{P}_{6\text{A}}): \mathbf{Z}_{(n)} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|(1+\alpha)\mathbf{Z} - \mathbf{W}_{\text{DBF}} + \boldsymbol{\Lambda}_{(n-1)} - \alpha\mathbf{W}_{\text{RF}(n-1)}\boldsymbol{\Delta}_{\text{RX}(n-1)}\mathbf{W}_{\text{BB}(n-1)}\|_F^2,$$

Algorithm 2 Proposed ADMM Solution for the A/D Hybrid Combiner Design

```
Initialize: $\mathbf{Z}$, $\mathbf{W}_{\text{RF}}$, $\boldsymbol{\Delta}_{\text{RX}}$, $\mathbf{W}_{\text{BB}}$ with random values, $\boldsymbol{\Lambda}$ with zeros, $\alpha = 1$ and $n = 1$
while $n \leq N_{\text{max}}$ do
    Update $\mathbf{Z}_{(n)}$ using solution (34),
        $\mathbf{W}_{\text{RF}(n)}$ using solution (35),
        $\boldsymbol{\Delta}_{\text{RX}(n)}$ by solving ($\mathcal{P}_{\text{6C}}$) using CVX [48],
        $\mathbf{W}_{\text{BB}(n)}$ using solution (36), and
        update $\boldsymbol{\Lambda}_{(n)}$ using solution (33).
    $n \leftarrow n + 1$
end while
return $\mathbf{W}^{\star}_{\text{RF}}$, $\boldsymbol{\Delta}^{\star}_{\text{RX}}$, $\mathbf{W}^{\star}_{\text{BB}}$
```

$$(\mathcal{P}_{\text{6B}}) : \mathbf{W}_{\text{RF}(n)} = \arg\min_{\mathbf{W}_{\text{RF}}} \mathbb{1}_{\mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}}\{\mathbf{W}_{\text{RF}}\} + \frac{\alpha}{2}\left\|\mathbf{Z}_{(n)} + \boldsymbol{\Lambda}_{(n-1)}/\alpha - \mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{\text{RX}(n-1)}\mathbf{W}_{\text{BB}(n-1)}\right\|_F^2,$$

$$(\mathcal{P}_{\text{6C}}) : \boldsymbol{\Delta}_{\text{RX}(n)} = \arg\min_{\boldsymbol{\Delta}_{\text{RX}}} \|\mathbf{y}_{\text{c}} - \boldsymbol{\Psi}_{\text{R}}\text{vec}(\boldsymbol{\Delta}_{\text{RX}})\|_2^2 + \gamma_{\text{R}}P_{\text{RX}}(\boldsymbol{\Delta}_{\text{RX}}) \text{ subject to } \boldsymbol{\Delta}_{\text{RX}} \in \mathcal{D}_{\text{RX}},$$

$$(\mathcal{P}_{\text{6D}}) : \mathbf{W}_{\text{BB}(n)} = \arg\min_{\mathbf{W}_{\text{BB}}} \frac{\alpha}{2}\|\mathbf{Z}_{(n)} + \boldsymbol{\Lambda}_{(n-1)}/\alpha - \mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{\text{RX}(n)}\mathbf{W}_{\text{BB}}\|_F^2,$$

$$\boldsymbol{\Lambda}_{(n)} = \boldsymbol{\Lambda}_{(n-1)} + \alpha\left(\mathbf{Z}_{(n)} - \mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{\text{RX}(n)}\mathbf{W}_{\text{BB}(n)}\right), \tag{33}$$

where n denotes the iteration index, $\mathbf{y}_{\text{c}} = \text{vec}(\mathbf{Z}_{(n)} + \boldsymbol{\Lambda}_{(n-1)}/\alpha)$ and $\boldsymbol{\Psi}_{\text{R}} = \mathbf{W}_{\text{BB}(n-1)} \otimes \mathbf{W}_{\text{RF}(n)}$ ($\otimes$ is the Khatri-Rao product).

We solve the optimization problems ($\mathcal{P}_{\text{6A}}$)-($\mathcal{P}_{\text{6D}}$) in a similar way to the derivations in Section III for the TX. The solution for $\mathbf{Z}_{(n)}$ is:

$$\mathbf{Z}_{(n)} = \frac{1}{\alpha + 1}\left(\mathbf{W}_{\text{DBF}} - \boldsymbol{\Lambda}_{(n-1)} + \alpha\mathbf{W}_{\text{RF}(n-1)}\boldsymbol{\Delta}_{\text{RX}(n-1)}\mathbf{W}_{\text{BB}(n-1)}\right). \tag{34}$$

The equation for $\mathbf{W}_{\text{RF}(n)}$ is as follows:

$$\begin{aligned}\mathbf{W}_{\text{RF}(n)} = \Pi_{\mathcal{W}}\Big\{&\left(\boldsymbol{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\right)\mathbf{W}_{\text{BB}(n-1)}^H\boldsymbol{\Delta}_{\text{RX}(n-1)}^H \\ &\left\{\alpha\boldsymbol{\Delta}_{\text{RX}(n-1)}\mathbf{W}_{\text{BB}(n-1)}\mathbf{W}_{\text{BB}(n-1)}^H\boldsymbol{\Delta}_{\text{RX}(n-1)}^H\right\}^{-1}\Big\}.\end{aligned} \tag{35}$$

The solution to $\boldsymbol{\Delta}_{\text{RX}(n)}$ is obtained by solving ($\mathcal{P}_{\text{6C}}$) using CVX [48]. The matrix $\mathbf{W}_{\text{BB}(n)}$ is obtained as follows:

$$\mathbf{W}_{\text{BB}(n)} = \left\{\alpha\boldsymbol{\Delta}_{\text{RX}(n)}^H\mathbf{W}_{\text{RF}(n)}^H\mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{\text{RX}(n)}\right\}^{-1}\boldsymbol{\Delta}_{\text{RX}(n)}^H\mathbf{W}_{\text{RF}(n)}^H\left(\boldsymbol{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}\right). \tag{36}$$

Algorithm 2 provides the complete procedure to obtain $\mathbf{W}_{\text{RF}}$, $\mathbf{\Delta}_{\text{RX}}$ and $\mathbf{W}_{\text{BB}}$. It starts by initializing the entries of the matrices $\mathbf{Z}$, $\mathbf{W}_{\text{RF}}$, $\mathbf{\Delta}_{\text{RX}}$, $\mathbf{W}_{\text{BB}}$ with random values and the entries of the Lagrange multiplier matrix $\mathbf{\Lambda}$ with zeros. For iteration index n, $\mathbf{Z}_{(n)}$, $\mathbf{W}_{\text{RF}(n)}$, $\mathbf{\Delta}_{\text{RX}(n)}$, $\mathbf{W}_{\text{BB}(n)}$ are updated at each iteration step by using the solution in (34), (35), solving $(\mathcal{P}_{6\text{C}})$ using CVX, (36) and (33), respectively. The operator $\Pi_{\mathcal{W}}$ projects the solution onto the set $\mathcal{W}$. This procedure is identical to problem $(\mathcal{P}''_{3\text{B}})$ in Section III, except that the set $\mathcal{W}$ replaces $\mathcal{F}$. A termination criterion is defined using a maximum number of iterations (N_{max}) or a fidelity criterion similar to (31). Upon convergence, the number of bits for each ADC is obtained by using (3) and quantizing to the nearest integer value. The optimal hybrid combining matrices $\mathbf{W}^{\star}_{\text{RF}}$, $\mathbf{\Delta}^{\star}_{\text{RX}}$, $\mathbf{W}^{\star}_{\text{BB}}$ are obtained at the end of this algorithm.

Computational complexity analysis of Algorithm 2: Similar to Algorithm 1 for the TX, the complexity of the solution of $(\mathcal{P}_{6\text{C}})$ can be upper-bounded by $\mathcal{O}((L_{\text{R}}^2 N_{\text{R}} N_{\text{s}})^3)$ which can be improved significantly by exploiting the structure of $\mathbf{\Psi}_{\text{R}}$.

Once the optimal DAC and ADC bit resolution matrices, i.e., $\mathbf{\Delta}_{\text{TX}}$ and $\mathbf{\Delta}_{\text{RX}}$, and optimal hybrid precoding and combining matrices, i.e., $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{RF}}$, $\mathbf{W}_{\text{BB}}$, are obtained then they can be plugged into (14) and (15) to obtain the maximum EE in (13). In the next section, we discuss the simulation results based on the proposed solution at the TX and the RX, and comparison with existing benchmark techniques.

V. Simulation Results

In this section, we evaluate the performance of the proposed ADMM solution using computer simulation results. All the results have been averaged over 1000 Monte-Carlo realizations. For comparison with the proposed ADMM solution, we consider following benchmark techniques:

1) Digital beamforming with 8-bit resolution: We consider the conventional fully digital beamforming architecture, where the number of RF chains at the TX/RX is equal to the number of TX/RX antennas, i.e., $L_{\text{T}} = N_{\text{T}}$ and $L_{\text{R}} = N_{\text{R}}$. In terms of the resolution sampling, we consider full-bit resolution, i.e., $M = 8$-bit, which represents the best case from the achievable SE perspective.

2) A/D Hybrid beamforming with 1-bit and 8-bit resolutions: We also consider a A/D hybrid beamforming architecture with $L_{\text{T}} < N_{\text{T}}$ and $L_{\text{R}} < N_{\text{R}}$, for two cases of DAC/ADC bit resolution: a) 1-bit resolution which usually shows reasonable EE performance, and b) 8-bit resolution which usually shows high SE results.

Power Terms	**Values**
Power per bit in the DAC/ADC	$P_{\text{DAC}} = P_{\text{ADC}} = 100$ mW
Circuit power at the TX/RX	$P_{\text{CT}} = P_{\text{CR}} = 10$ W
Power per phase shifter at the TX/RX	$P_{\text{PT}} = P_{\text{PR}} = 10$ mW
Power per antenna at the TX/RX	$P_{\text{T}} = P_{\text{R}}$ = 100 mW

(a) Typical values of the power terms [49] used in (16) and (18).

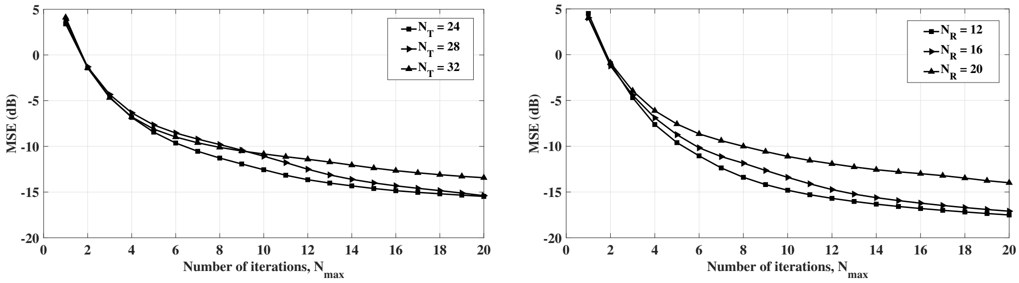
System Parameters	**Values**
Number of clusters	$N_{\text{cl}} = 2$
Number of rays	$N_{\text{ray}} = 3$
Number of TX antennas	$N_{\text{T}} = 32$
Number of RX antennas	$N_{\text{R}} = 5$
Number of TX/RX RF chains	$L_{\text{T}} = L_{\text{R}} = 5$
Number of data streams	$N_{\text{s}} = 5$
Bit resolution range	$[m, M] = [1, 8]$
Maximum number of ADMM iterations	$N_{\max} = 20$
Maximum TX/RX trade-off parameter	$\gamma_{\text{T}}^{max} = 0.1$; γ_{R}^{max} = 1

(b) System parameter values.

TABLE I: Summary of the simulation parameter values.

3) Brute force with A/D hybrid beamforming: We also implement an exhaustive search approach as an upper bound for EE maximization called brute force (BF), based on [16]. Firstly the EE problem is split into TX and RX optimization problems similar to those for the proposed ADMM approach. Then it makes a search over all the possible DAC and ADC bit resolutions in the range of $[m, M]$ associated with the each RF chain from 1 to L_{T} and 1 to L_{R} at the TX and the RX, respectively. It then finds the best EE out of all the possible cases and chooses the corresponding optimal resolution for each DAC and ADC. This method provides the best possible EE performance and serves as upper bound for EE maximization by the ADMM approach.

Complexity comparison with the BF approach: The proposed ADMM solution has lower complexity than the upper bound BF approach because the BF technique involves a search over all the possible DAC/ADC bit resolutions while the proposed ADMM solution directly optimizes the number of bits at each DAC/ADC. We constrain the number of RF chains $L_{\text{T}} = L_{\text{R}} = 5$ for the BF approach due to the high complexity order which is $\mathcal{O}(M^{L_{\text{T}}})$ and $\mathcal{O}(M^{L_{\text{R}}})$ at the TX and the RX, respectively.



(a) At the TX for different N_T at $\gamma_\text{T} = 0.001$.

(b) At the RX for different N_R at $\gamma_\text{R} = 0.5$.

Fig. 2: Convergence of the proposed ADMM solution at the TX and the RX.

System setup: Table 1 summarizes the simulation values used for the system and power terms, and in addition, we consider $\alpha = 1$ and $\sigma^2_{\alpha,i}$ = 1. The azimuth angles of departure and arrival are computed with uniformly distributed mean angles, and each cluster follows a Laplacian distribution about the mean angle. The antenna elements in the ULA are spaced by distance $d = \lambda/2$. The signal-to-noise ratio (SNR) is given by the inverse of the noise variance, i.e., $1/\sigma^2_\text{n}$. The transmit vector $\mathbf{x}$ is composed of the normalized i.i.d. Gaussian symbols. Under this assumption the covariance matrix of $\mathbf{x}$ is an identity matrix.

Convergence of the proposed ADMM solution: Figs. 2 (a) and 2 (b) show the convergence of the ADMM solution at the TX and the RX as proposed in Algorithm 1 and Algorithm 2, respectively, to obtain the optimal bit resolution at each DAC/ADC and the corresponding optimal precoder/combiner matrices. It can be observed from Fig. 2 (a) that the proposed solution converges rapidly within 16 iterations and the normalized mean square error (NMSE) at the TX, $\left\|\mathbf{F}_\text{DBF} - \mathbf{F}_{\text{RF}(N_\text{max})}\mathbf{\Delta}_{\text{TX}(N_\text{max})}\mathbf{F}_{\text{BB}(N_\text{max})}\right\|^2_F \,/\, \|\mathbf{F}_\text{DBF}\|^2_F$, goes as low as -15 dB. Similarly, in Fig. 2 (b), the proposed solution again converges rapidly and the NMSE at the RX, $\left\|\mathbf{W}_\text{DBF} - \mathbf{W}_{\text{RF}(N_\text{max})}\mathbf{\Delta}_{\text{RX}(N_\text{max})}\mathbf{W}_{\text{BB}(N_\text{max})}\right\|^2_F \,/\, \|\mathbf{W}_\text{DBF}\|^2_F$, goes as low as -17 dB. A lower number of TX/RX antennas shows lower NMSE for a given number of iterations as expected, since fewer parameters are required to be estimated.

Fig. 3 shows the performance of the proposed ADMM solution compared with existing benchmark techniques w.r.t. SNR at $\gamma_\text{T} = 0.001$ and $\gamma_\text{R} = 0.5$. The proposed ADMM solution achieves high EE which is computed by (13) after obtaining the optimal DAC and ADC bit resolution matrices, i.e., $\mathbf{\Delta}_\text{TX}$ and $\mathbf{\Delta}_\text{RX}$, and optimal hybrid precoding and combining matrices, i.e., $\mathbf{F}_\text{RF}$, $\mathbf{F}_\text{BB}$ and $\mathbf{W}_\text{RF}$, $\mathbf{W}_\text{BB}$. The results are plugged into (14) and (15) to evaluate rate and

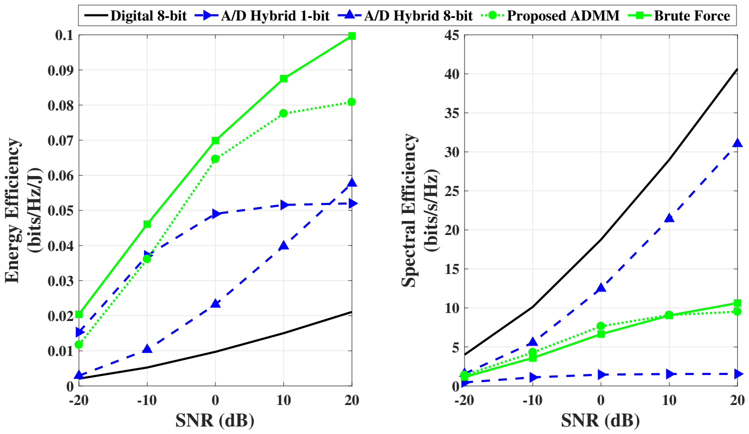


Fig. 3: EE and SE performance w.r.t. SNR at $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$.

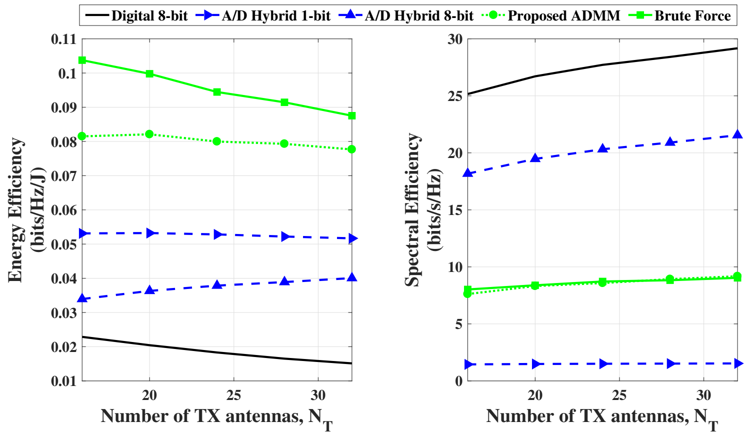


Fig. 4: EE and SE performance w.r.t. N_{T} at SNR = 10 dB, $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$.

power respectively. The EE for the proposed solution has similar performance to the BF approach and is better than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines, e.g., at SNR = 10 dB, the proposed ADMM solution outperforms the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.03 bits/Hz/J, 0.04 bits/Hz/J and 0.065 bits/Hz/J, respectively.

The proposed solution also exhibits better SE, which is the rate in (14) after obtaining the optimal DAC and ADC bit resolution matrices, and optimal hybrid precoding and combining matrices, than the hybrid 1-bit and has similar performance to the BF approach for high and

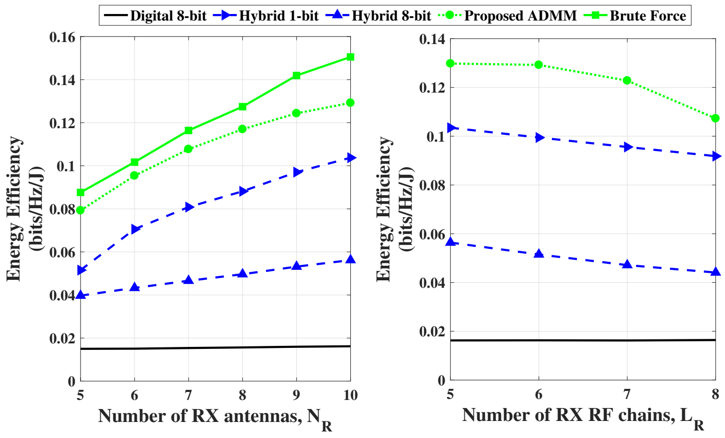


Fig. 5: EE performance w.r.t. N_R and L_R at SNR = 10 dB, $\gamma_\mathrm{T} = 0.001$ and $\gamma_\mathrm{R} = 0.5$.

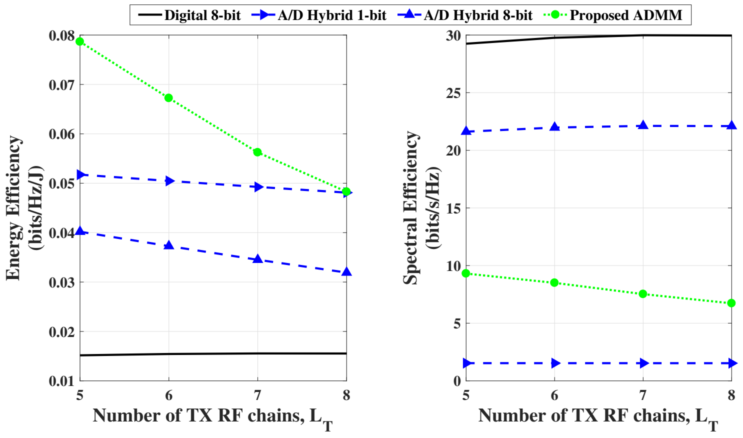


Fig. 6: EE and SE performance w.r.t. L_T at SNR = 10 dB, $\gamma_\mathrm{T} = 0.001$ and $\gamma_\mathrm{R} = 0.5$.

low SNR regions and hybrid 8-bit baseline for low SNR region. Note that the proposed ADMM solution enables the selection of different resolutions for different DACs/ADCs and thus, it offers a better trade-off for EE versus SE than existing approaches which are based on a fixed DAC/ADC bit resolution.

Fig. 4 shows the EE (from (13)) and SE (from (14)) performance results w.r.t. the number of TX antennas N_T at 10 dB SNR, $\gamma_\mathrm{T} = 0.001$ and $\gamma_\mathrm{R} = 0.5$. The proposed ADMM solution again achieves high EE and performs similar to the BF approach and better than the hybrid 1-bit, the

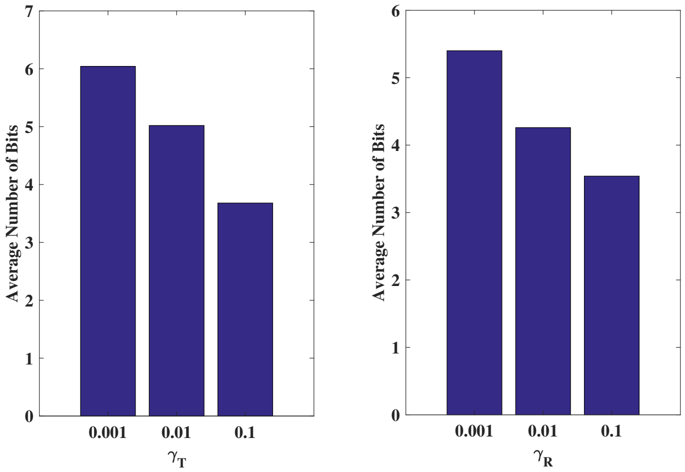


Fig. 7: Average number of bits for proposed ADMM w.r.t. $\gamma_{\rm T}$ and $\gamma_{\rm R}$ at the TX and the RX, respectively, at SNR = 10 dB.

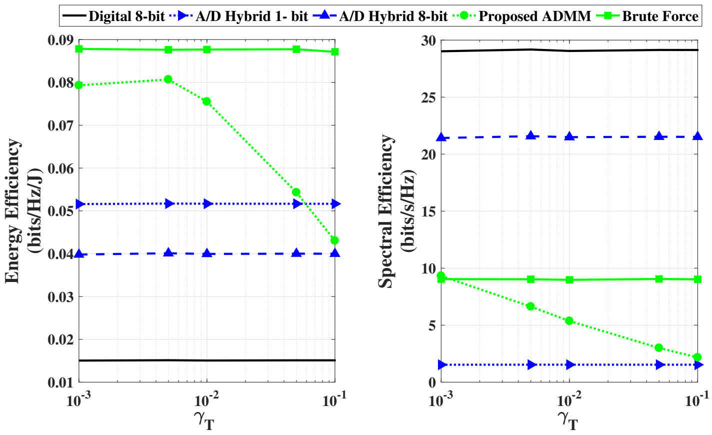


Fig. 8: EE and SE performance w.r.t. $\gamma_{\rm T}$ at SNR = 10 dB.

hybrid 8-bit and the digital full-bit baselines. For example, at $N_{\rm T} = 20$, the proposed ADMM solution outperforms hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.03 bits/Hz/J, 0.045 bits/Hz/J and 0.06 bits/Hz/J, respectively. The proposed ADMM solution also exhibits SE performance similar to the BF approach and better than the hybrid 1-bit baseline.

Fig. 5 shows the EE performance results w.r.t. the number of RX antennas $N_{\rm R}$ and the number

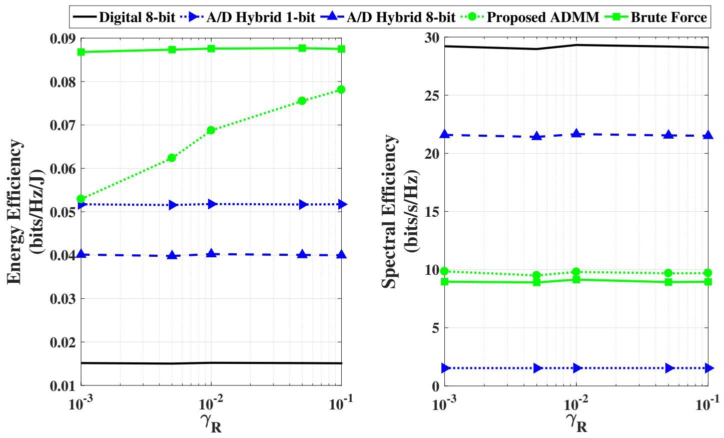


Fig. 9: EE and SE performance w.r.t. γ_{R} at SNR = 10 dB.

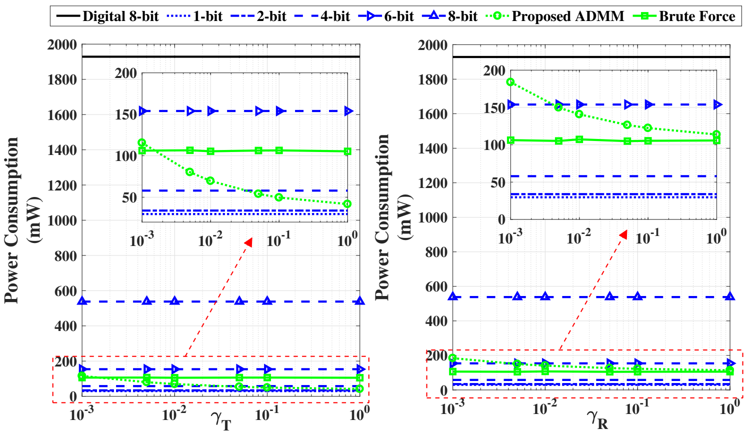


Fig. 10: Power consumption w.r.t. γ_{T} and γ_{R} at the TX and RX, respectively, at SNR = 10 dB.

of RX RF chains L_{R}, respectively, at 10 dB SNR, $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$. The proposed ADMM solution again achieves high EE which decreases with increase in the number of RX RF chains, and performs similar to the BF approach (for versus N_{R}) and better than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines. For example, at $N_{\mathrm{R}} = 7$, the proposed ADMM solution outperforms hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.03 bits/Hz/J, 0.06 bits/Hz/J and 0.09 bits/Hz/J, respectively. Also, e.g., at $L_{\mathrm{R}} = 6$, the proposed ADMM solution outperforms hybrid 1-bit, the hybrid 8-bit and the digital full-bit baselines by about 0.025 bits/Hz/J, 0.08 bits/Hz/J and 0.115 bits/Hz/J, respectively. Due to the high complexity of the BF approach, we do not plot results for this approach w.r.t. L_{T} and L_{R}.

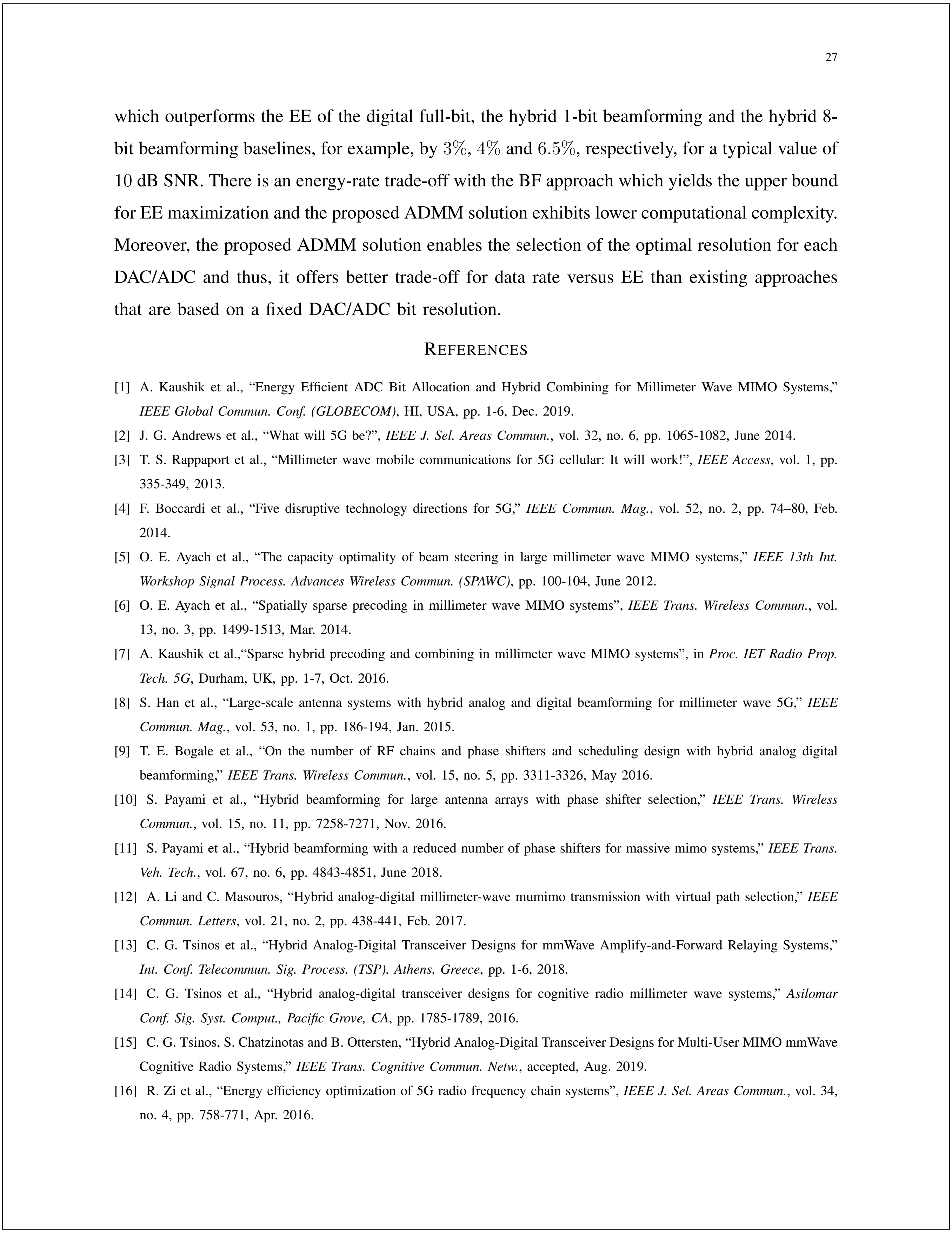
Fig. 6 shows the EE and SE performance results w.r.t. the number of TX RF chains L_{T} at 10 dB SNR, $\gamma_{\mathrm{T}} = 0.001$ and $\gamma_{\mathrm{R}} = 0.5$. The proposed ADMM solution achieves high EE, though this decreases with increase in the number of TX RF chains ADMM achieves better EE performance than the hybrid 1-bit, the hybrid 8-bit and the digital full-bit resolution baselines. Also, the proposed ADMM solution exhibits SE performance better than the hybrid 1-bit baseline.

Furthermore, we investigate the performance over the trade-off parameters γ_{T} and γ_{R} introduced in ($\mathcal{P}_2$) and ($\mathcal{P}_5$), respectively. Fig. 7 shows the bar plot of the average of the optimal number of bits selected by the proposed ADMM solution for each DAC versus γ_{T} and for each ADC versus γ_{R}. It can be observed that the average optimal number decreases with the increase in γ_{T} and γ_{R}, for example, the average number of DAC bits is around 6 for $\gamma_{\mathrm{T}} = 0.001$, 5 for $\gamma_{\mathrm{T}} = 0.01$ and 4 for $\gamma_{\mathrm{T}} = 0.1$. Similarly, at the RX, the average number of ADC bits is about 5 for $\gamma_{\mathrm{R}} = 0.001$, 4 for $\gamma_{\mathrm{R}} = 0.01$ and 3 for $\gamma_{\mathrm{R}} = 0.1$. This is because increasing γ_{T} or γ_{R} gives more weight to the power consumption.

Figs. 8 and 9 show the EE and SE plots for several solutions w.r.t. γ_{T} and γ_{R} at the TX and the RX, respectively. It can be observed that the proposed solution achieves higher EE performance than the fixed bit allocation solutions such as the digital full-bit, the hybrid 1-bit and the hybrid 8-bit baselines and achieves comparable EE and SE results to the BF approach. These curves also show that adjusting γ_{T} and γ_{R} values allow the system to vary the energy-rate trade-off. Note that the TX also accounts for the extra power term, i.e., tr$(\mathbf{F}\mathbf{F}^H)$ as shown in (16) which means that the selected γ_{T} parameter at the TX is lower than the selected γ_{R} parameter at the RX. Fig. 10 shows that the power consumption in the proposed case is low and decreases with the increase in the trade-off parameter γ_{T} and γ_{R} values unlike digital 8-bit, fixed bit resolution hybrid baselines and the BF approach.

VI. Conclusion

This paper proposes an energy efficient mmWave A/D hybrid MIMO system which can vary dynamically the DAC and ADC bit resolutions at the TX and the RX, respectively. This method uses the decomposition of the A/D hybrid precoder/combiner matrix into three parts representing the analog precoder/combiner matrix, the DAC/ADC bit resolution matrix and the digital precoder/combiner matrix. These three matrices are optimized by a novel ADMM solution



which outperforms the EE of the digital full-bit, the hybrid 1-bit beamforming and the hybrid 8-bit beamforming baselines, for example, by 3%, 4% and 6.5%, respectively, for a typical value of 10 dB SNR. There is an energy-rate trade-off with the BF approach which yields the upper bound for EE maximization and the proposed ADMM solution exhibits lower computational complexity. Moreover, the proposed ADMM solution enables the selection of the optimal resolution for each DAC/ADC and thus, it offers better trade-off for data rate versus EE than existing approaches that are based on a fixed DAC/ADC bit resolution.

REFERENCES

[1] A. Kaushik et al., “Energy Efficient ADC Bit Allocation and Hybrid Combining for Millimeter Wave MIMO Systems,” *IEEE Global Commun. Conf. (GLOBECOM)*, HI, USA, pp. 1-6, Dec. 2019.

[2] J. G. Andrews et al., “What will 5G be?”, *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, June 2014.

[3] T. S. Rappaport et al., “Millimeter wave mobile communications for 5G cellular: It will work!”, *IEEE Access*, vol. 1, pp. 335-349, 2013.

[4] F. Boccardi et al., “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[5] O. E. Ayach et al., “The capacity optimality of beam steering in large millimeter wave MIMO systems,” *IEEE 13th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, pp. 100-104, June 2012.

[6] O. E. Ayach et al., “Spatially sparse precoding in millimeter wave MIMO systems”, *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[7] A. Kaushik et al.,“Sparse hybrid precoding and combining in millimeter wave MIMO systems”, in *Proc. IET Radio Prop. Tech. 5G*, Durham, UK, pp. 1-7, Oct. 2016.

[8] S. Han et al., “Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,” *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.

[9] T. E. Bogale et al., “On the number of RF chains and phase shifters and scheduling design with hybrid analog digital beamforming,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311-3326, May 2016.

[10] S. Payami et al., “Hybrid beamforming for large antenna arrays with phase shifter selection,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258-7271, Nov. 2016.

[11] S. Payami et al., “Hybrid beamforming with a reduced number of phase shifters for massive mimo systems,” *IEEE Trans. Veh. Tech.*, vol. 67, no. 6, pp. 4843-4851, June 2018.

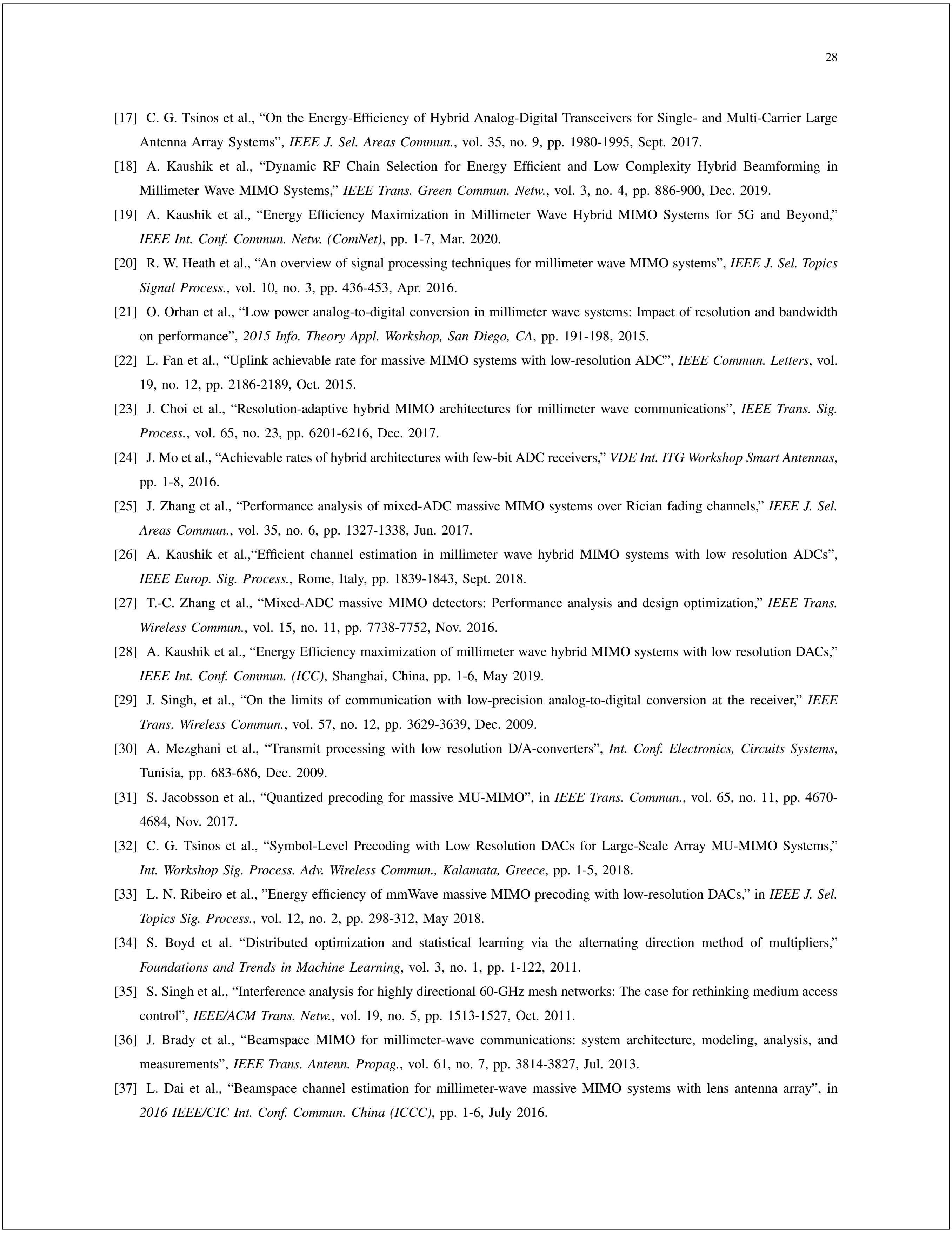
[12] A. Li and C. Masouros, “Hybrid analog-digital millimeter-wave mumimo transmission with virtual path selection,” *IEEE Commun. Letters*, vol. 21, no. 2, pp. 438-441, Feb. 2017.

[13] C. G. Tsinos et al., “Hybrid Analog-Digital Transceiver Designs for mmWave Amplify-and-Forward Relaying Systems,” *Int. Conf. Telecommun. Sig. Process. (TSP), Athens, Greece*, pp. 1-6, 2018.

[14] C. G. Tsinos et al., “Hybrid analog-digital transceiver designs for cognitive radio millimeter wave systems,” *Asilomar Conf. Sig. Syst. Comput., Pacific Grove, CA*, pp. 1785-1789, 2016.

[15] C. G. Tsinos, S. Chatzinotas and B. Ottersten, “Hybrid Analog-Digital Transceiver Designs for Multi-User MIMO mmWave Cognitive Radio Systems,” *IEEE Trans. Cognitive Commun. Netw.*, accepted, Aug. 2019.

[16] R. Zi et al., “Energy efficiency optimization of 5G radio frequency chain systems”, *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.



[17] C. G. Tsinos et al., “On the Energy-Efficiency of Hybrid Analog-Digital Transceivers for Single- and Multi-Carrier Large Antenna Array Systems”, *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980-1995, Sept. 2017.

[18] A. Kaushik et al., “Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid Beamforming in Millimeter Wave MIMO Systems,” *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 886-900, Dec. 2019.

[19] A. Kaushik et al., “Energy Efficiency Maximization in Millimeter Wave Hybrid MIMO Systems for 5G and Beyond,” *IEEE Int. Conf. Commun. Netw. (ComNet)*, pp. 1-7, Mar. 2020.

[20] R. W. Heath et al., “An overview of signal processing techniques for millimeter wave MIMO systems”, *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[21] O. Orhan et al., “Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance”, *2015 Info. Theory Appl. Workshop, San Diego, CA*, pp. 191-198, 2015.

[22] L. Fan et al., “Uplink achievable rate for massive MIMO systems with low-resolution ADC”, *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2186-2189, Oct. 2015.

[23] J. Choi et al., “Resolution-adaptive hybrid MIMO architectures for millimeter wave communications”, *IEEE Trans. Sig. Process.*, vol. 65, no. 23, pp. 6201-6216, Dec. 2017.

[24] J. Mo et al., “Achievable rates of hybrid architectures with few-bit ADC receivers,” *VDE Int. ITG Workshop Smart Antennas*, pp. 1-8, 2016.

[25] J. Zhang et al., “Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327-1338, Jun. 2017.

[26] A. Kaushik et al.,“Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs”, *IEEE Europ. Sig. Process.*, Rome, Italy, pp. 1839-1843, Sept. 2018.

[27] T.-C. Zhang et al., “Mixed-ADC massive MIMO detectors: Performance analysis and design optimization,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738-7752, Nov. 2016.

[28] A. Kaushik et al., “Energy Efficiency maximization of millimeter wave hybrid MIMO systems with low resolution DACs,” *IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, pp. 1-6, May 2019.

[29] J. Singh, et al., “On the limits of communication with low-precision analog-to-digital conversion at the receiver,” *IEEE Trans. Wireless Commun.*, vol. 57, no. 12, pp. 3629-3639, Dec. 2009.

[30] A. Mezghani et al., “Transmit processing with low resolution D/A-converters”, *Int. Conf. Electronics, Circuits Systems*, Tunisia, pp. 683-686, Dec. 2009.

[31] S. Jacobsson et al., “Quantized precoding for massive MU-MIMO”, in *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670-4684, Nov. 2017.

[32] C. G. Tsinos et al., “Symbol-Level Precoding with Low Resolution DACs for Large-Scale Array MU-MIMO Systems,” *Int. Workshop Sig. Process. Adv. Wireless Commun., Kalamata, Greece*, pp. 1-5, 2018.

[33] L. N. Ribeiro et al., ”Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs,” in *IEEE J. Sel. Topics Sig. Process.*, vol. 12, no. 2, pp. 298-312, May 2018.

[34] S. Boyd et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2011.

[35] S. Singh et al., “Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control”, *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513-1527, Oct. 2011.

[36] J. Brady et al., “Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements”, *IEEE Trans. Antenn. Propag.*, vol. 61, no. 7, pp. 3814-3827, Jul. 2013.

[37] L. Dai et al., “Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array”, in *2016 IEEE/CIC Int. Conf. Commun. China (ICCC)*, pp. 1-6, July 2016.

[38] A. Mezghani and J. A. Nossek, “Capacity lower bound of MIMO channels with output quantization and correlated noise,” *IEEE Int. Symp. Info. Theory (ISIT)*, Cambridge, USA, Jul. 2012.

[39] A. Zappone and E. Jorswieck, “Energy Efficiency in Wireless Networks via Fractional Programming Theory,” *Found. Trends Commun. Info. Theory*, vol. 11, no. 3-4, pp 185-396, 2015.

[40] D. P. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” *IEEE J. Sel. Areas Commun.*, vol. 24,no. 8, pp. 1439-1451, Aug. 2006.

[41] W. Dinkelbach, “On nonlinear fractional programming”, *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.

[42] D. Tse and P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, UK, 2004.

[43] C. G. Tsinos et al., “Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization,” *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 160-174, June 2017.

[44] C. G. Tsinos and B. Ottersten, “An efficient algorithm for unit-modulus quadratic programs with application in beamforming for wireless sensor networks,” *IEEE Sig. Process. Letters*, vol. 25, no. 2, pp. 169-173, Feb. 2018.

[45] D. P. Bertsekas, “Nonlinear programming,” Athena Scientific, USA, Sept. 1999.

[46] J. Eckstein and D. P. Bertsekas, “On the DouglasRachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 55, no. 1-3, pp. 293-318, 1992.

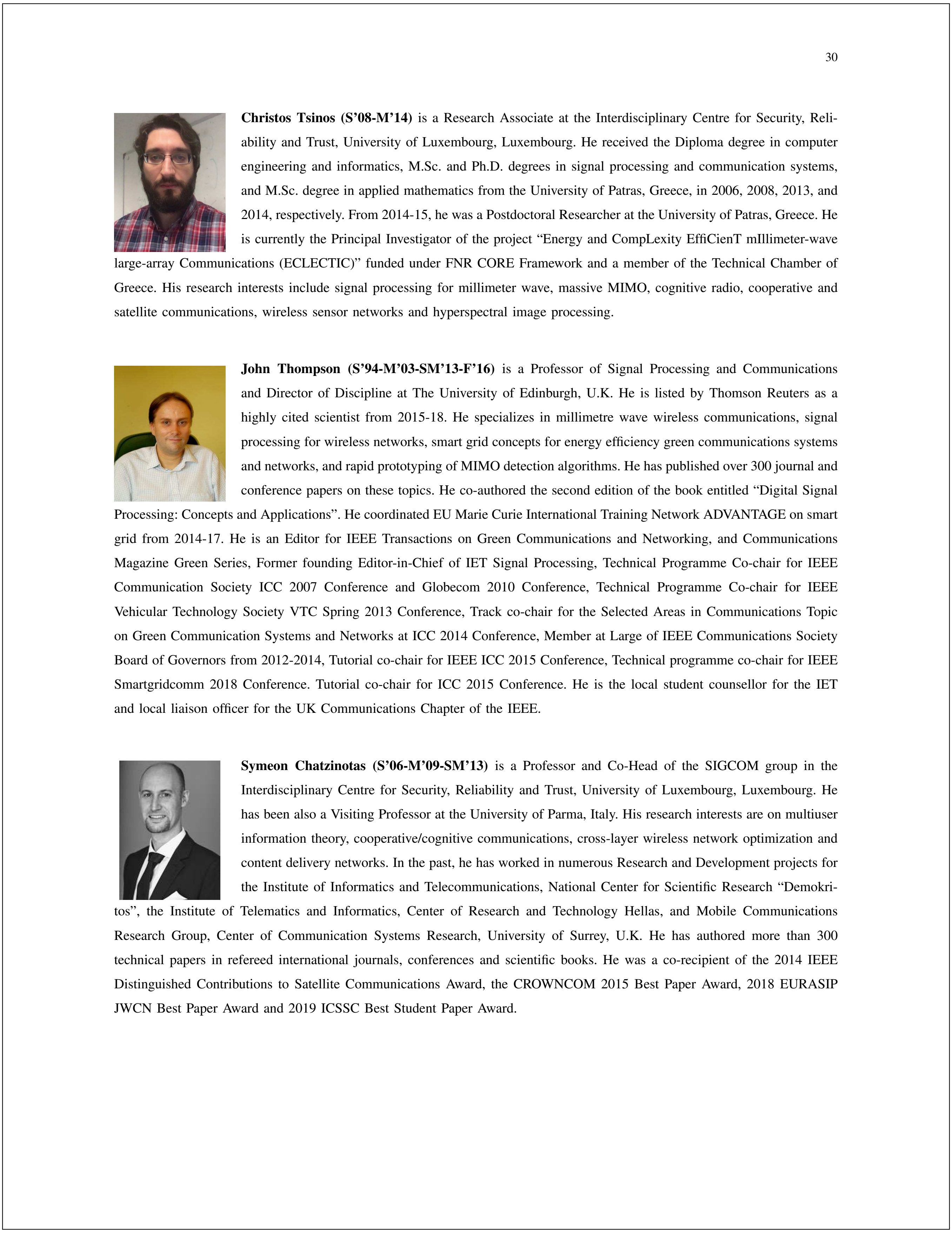
[47] E. G. Golshtein and N. Tretyakov, “Modified lagrangians in convex programming and their generalizations,” in *Point-to-Set Maps and Mathematical Programming, Springer*, pp. 86-97, 1979.

[48] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs”, in *Recent Adv. Learning and Control*, Springer-Verlag Ltd., pp. 95-110, 2008.

[49] T. S. Rappaport et al., “Millimeter wave wireless communications,” Prentice-Hall, NJ, USA, Sept. 2014.

Aryan Kaushik is a Research Fellow in Communications and Radar Transmission at the Department of Electronic and Electrical Engineering, University College London, U.K. He completed his Ph.D. degree in communications engineering at the Institute for Digital Communications, The University of Edinburgh, U.K., in 2019, where he also pursued the postgraduate certification in academic practice with the Institute for Academic Development. He received M.Sc. degree in telecommunications from The Hong Kong University of Science and Technology, Hong Kong, in 2015. He has held visiting research appointments at the Imperial College London, U.K., from 2019-20, University of Luxembourg, Luxembourg, in 2018, and Beihang University, China, in the period of 2017-19. His research interests include signal processing for communications, dual communications and radar transmission, energy efficient wireless communications and millimeter wave massive MIMO systems.

Evangelos Vlachos (M’19) is a Research Associate at the Industrial Systems Institute, Athena Research Centre, Patras, Greece. He has been a Research Associate in Signal Processing for Communications at the Institute for Digital Communications, The University of Edinburgh, U.K., from 2017-19. He received the Diploma, M.Sc. and Ph.D. degrees from the Computer Engineering and Informatics Department, University of Patras, Greece, in 2005, 2009, and 2015, respectively. From 2015-16, he was a Postdoctoral Researcher at the Laboratory of Signal Processing and Telecommunications in the University of Patras, Greece, where in 2016, he worked at the Visualization and Virtual Reality Group. He received best paper award from IEEE ICME in 2017 and participated in six research projects funded by the EU. His research interests include wireless communications, machine learning and optimization, adaptive control and filtering algorithms.



Christos Tsinos (S'08-M'14) is a Research Associate at the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg. He received the Diploma degree in computer engineering and informatics, M.Sc. and Ph.D. degrees in signal processing and communication systems, and M.Sc. degree in applied mathematics from the University of Patras, Greece, in 2006, 2008, 2013, and 2014, respectively. From 2014-15, he was a Postdoctoral Researcher at the University of Patras, Greece. He is currently the Principal Investigator of the project "Energy and CompLexity EffiCienT mIllimeter-wave large-array Communications (ECLECTIC)" funded under FNR CORE Framework and a member of the Technical Chamber of Greece. His research interests include signal processing for millimeter wave, massive MIMO, cognitive radio, cooperative and satellite communications, wireless sensor networks and hyperspectral image processing.

John Thompson (S'94-M'03-SM'13-F'16) is a Professor of Signal Processing and Communications and Director of Discipline at The University of Edinburgh, U.K. He is listed by Thomson Reuters as a highly cited scientist from 2015-18. He specializes in millimetre wave wireless communications, signal processing for wireless networks, smart grid concepts for energy efficiency green communications systems and networks, and rapid prototyping of MIMO detection algorithms. He has published over 300 journal and conference papers on these topics. He co-authored the second edition of the book entitled "Digital Signal Processing: Concepts and Applications". He coordinated EU Marie Curie International Training Network ADVANTAGE on smart grid from 2014-17. He is an Editor for IEEE Transactions on Green Communications and Networking, and Communications Magazine Green Series, Former founding Editor-in-Chief of IET Signal Processing, Technical Programme Co-chair for IEEE Communication Society ICC 2007 Conference and Globecom 2010 Conference, Technical Programme Co-chair for IEEE Vehicular Technology Society VTC Spring 2013 Conference, Track co-chair for the Selected Areas in Communications Topic on Green Communication Systems and Networks at ICC 2014 Conference, Member at Large of IEEE Communications Society Board of Governors from 2012-2014, Tutorial co-chair for IEEE ICC 2015 Conference, Technical programme co-chair for IEEE Smartgridcomm 2018 Conference. Tutorial co-chair for ICC 2015 Conference. He is the local student counsellor for the IET and local liaison officer for the UK Communications Chapter of the IEEE.

Symeon Chatzinotas (S'06-M'09-SM'13) is a Professor and Co-Head of the SIGCOM group in the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg. He has been also a Visiting Professor at the University of Parma, Italy. His research interests are on multiuser information theory, cooperative/cognitive communications, cross-layer wireless network optimization and content delivery networks. In the past, he has worked in numerous Research and Development projects for the Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", the Institute of Telematics and Informatics, Center of Research and Technology Hellas, and Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey, U.K. He has authored more than 300 technical papers in refereed international journals, conferences and scientific books. He was a co-recipient of the 2014 IEEE Distinguished Contributions to Satellite Communications Award, the CROWNCOM 2015 Best Paper Award, 2018 EURASIP JWCN Best Paper Award and 2019 ICSSC Best Student Paper Award.

Energy Efficiency Maximization in Millimeter Wave Hybrid MIMO Systems for 5G and Beyond

Aryan Kaushik, John Thompson and Evangelos Vlachos
Institute for Digital Communications, The University of Edinburgh, United Kingdom
Email: {a.kaushik, j.s.thompson, e.vlachos}@ed.ac.uk

(Invited Paper)

***Abstract*—At millimeter wave (mmWave) frequencies, the higher cost and power consumption of hardware components in multiple-input multiple output (MIMO) systems do not allow beamforming entirely at the baseband with a separate radio frequency (RF) chain for each antenna. In such scenarios, to enable spatial multiplexing, hybrid beamforming, which uses phase shifters to connect a fewer number of RF chains to a large number of antennas is a cost effective and energy-saving alternative. This paper describes our research on fully adaptive transceivers that adapt their behaviour on a frame-by-frame basis, so that a mmWave hybrid MIMO system always operates in the most energy efficient manner. Exhaustive search based brute force approach is computationally intensive, so we study fractional programming as a low-cost alternative to solve the problem which maximizes energy efficiency. The performance results indicate that the resulting mmWave hybrid MIMO transceiver achieves significantly improved energy efficiency results compared to the baseline cases involving analogue-only or digital-only signal processing solutions, and shows performance trade-offs with the brute force approach.**

***Index Terms*—energy efficiency, hybrid beamforming, MIMO, millimeter wave, 5G and beyond.**

I. Introduction

Fifth generation (5G) technology is set to address the consumer demands and performance enhancements for mobile communication in 2020 and beyond [1]. There will be 28.5 billion networked devices and connections by 2022 [2] and 8.9 billion mobile subscriptions by the end of 2024 [3]. For such large scale use of mobile devices through 5G and beyond 5G services, the communication systems would require increased capacity, high data rates, improved coverage and also reduced energy consumption. We currently use the microwave frequency spectrum for communication which is congested with a large number of consumer devices raising the demand for an unused and available spectrum. This increased demand on bandwidth and capacity can be resolved by the use of millimeter wave (mmWave) frequency spectrum which ranges from 30-300 GHz [4]. This is beneficial as the larger spectral channels at mmWave would lead to higher data rates. Moreover, the large scale antenna arrays such as the multiple-input multiple-output (MIMO) systems can reduce the high path loss at mmWave frequencies [5], [6]. However, it would be difficult to use one radio frequency (RF) chain per antenna leading to a least energy efficient and highly complex system. Thus, using digital beamforming which needs a dedicated RF chain per antenna is not very practical from energy efficiency (EE) and hardware complexity perspectives. To save power and reduce complexity, analogue beamforming can be used where a network of analogue phase shifters connects the antennas to a single RF chain [7], but multi-stream and multi-user communication can not be supported.

A mmWave MIMO system with hybrid beamforming (HBF) architecture can save power and reduce hardware complexity using fewer number of RF chains than the large number of antennas, and support multi-stream communication with high spectral efficiency (SE) [8]–[12]. Such systems can also be optimized to achieve high EE gains [13] but this has not been widely studied for EE maximization with low complexity. Low resolution sampling can be implemented to save power such as in [14] we discuss EE maximization with low resolution digital-to-analogue converters (DACs) at the transmitter (TX), in [15] with low resolution analogue-to-digital converters (ADCs) at the receiver (RX) and in [16] with low resolution sampling at both the DACs and the ADCs. However, the existing literature mostly considers fixed number of RF chains for high SE performance [8]–[12] and RF chains consume a lot of power which increases the cost of MIMO systems [17]. Reference [13] provides an exhaustive search based brute force (BF) approach where a full precoder design is evaluated for all possible combinations of RF chains, in order to select the number of RF chains that maximizes EE but this is a computationally inefficient solution. Moreover, lower complexity solutions can be implemented to design the HBF matrices than in [8], [13].

Contribution: This paper describes different approaches to performing *dynamic adaptation* of a mmWave hybrid MIMO system on a frame-by-frame basis. Our idea exploits the beam training phase in the communication system to learn the propagation conditions. Based on this, we can choose to adapt the behaviour of the transceiver in order to optimize a performance metric of interest, such as EE. Maximizing EE is challenging mathematically because it is a ratio of two important parameters, namely data rate (or SE) and power. In our recent research, we use the Dinkelbach method (DM) [18] to replace this ratio function by an iterative sequence of problems based on the difference of the numerator and denominator. In this work, we discuss different ways to optimize the transceivers, particularly in relation to the number of activated

Notations	Description
a	Scalar
$\mathbf{a}$	Vector
$\|\mathbf{a}\|_0$	l_0-norm of $\mathbf{a}$
$\mathbf{A}$	Matrix
$\|\mathbf{A}\|$	Determinant of $\mathbf{A}$
$\mathbf{A}^T$	Transpose of $\mathbf{A}$
$\mathbf{A}^H$	Complex conjugate transpose of $\mathbf{A}$
$\mathbf{A}^{(i)}$	i-th column of $\mathbf{A}$
$\|\mathbf{A}\|_F$	Frobenius norm of $\mathbf{A}$
$\mathcal{CN}(\mathbf{a};\mathbf{A})$	Complex Gaussian vector; mean $\mathbf{a}$, covariance $\mathbf{A}$
$\mathbb{C}^{A\times B}$	To represent matrix of size $A \times B$ with complex entries
$\mathbb{E}\{\cdot\}$	Expectation operator
$\mathbf{I}_N$	Identity matrix with size $N \times N$
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{R}\{\cdot\}$	Real part
tr($\mathbf{A}$)	Trace of $\mathbf{A}$
$\mathbf{X} \in \mathbb{C}^{A\times B}$	Complex-valued matrix $\mathbf{X}$ of size $A \times B$
$\mathbf{X} \in \mathbb{R}^{A\times B}$	Real-valued matrix $\mathbf{X}$ of size $A \times B$

TABLE I: List of notations and their description.

RF chains and the sample rate of the system. As a practical example, we present a more detailed discussion of how the Dinkelbach's approach can be used to optimize the EE and simultaneously achieve a low complexity alternative to the exhaustive search based BF approach in [13]. An attractive feature of our approach is that we only need to compute the HBF matrices once, after the number of RF chains is determined by the DM based solution.

Notations and Organization: Table I provides a list of notations used in this paper along with their description. The remainder of the paper is structured as follows: Section II describes the channel model and HBF architecture that is used in the paper. Section III describes the EE maximization problem and we describe different approaches that we have studied to address this problem. In Section IV, we discuss in more detail how the DM can be applied to select the optimal number of RF chains. Section V presents simulation results to show the performance improvements of the DM and finally Section VI presents conclusions to the paper.

II. MmWave MIMO System with HBF

A. MmWave Channel

We use a narrowband clustered channel model due to different channel settings at mmWave such as the number of multipaths, amplitudes, etc. [6]. We consider N_{cl} clusters with N_{ray} paths related to each cluster and for a single user system we have N_{T} TX antennas transmitting N_{s} data streams to N_{R} RX antennas. This mmWave channel can be expressed as

$$\mathbf{H} = \sum_{i=1}^{N_{\text{cl}}} \sum_{l=1}^{N_{\text{ray}}} \alpha_{il} \mathbf{a}_{\text{R}}(\phi_{il}^{r}) \mathbf{a}_{\text{T}}(\phi_{il}^{t})^{H}, \tag{1}$$

where $\alpha_{il} \in \mathcal{CN}(0, \sigma_{\alpha,i}^2)$ is the gain term with $\sigma_{\alpha,i}^2$ being the average power of the i^{th} cluster. The vectors $\mathbf{a}_{\text{T}}(\phi_{il}^{t})$ and $\mathbf{a}_{\text{R}}(\phi_{il}^{r})$ denote the normalized array response vectors at the TX and the RX, respectively [6], with ϕ_{il}^{t} being the azimuth angles of departure and ϕ_{il}^{r} being the azimuth angles of arrival. We assume the transmit and receive arrays are uniform linear arrays (ULAs) of antennas, which are modelled as ideal sectored elements [19].

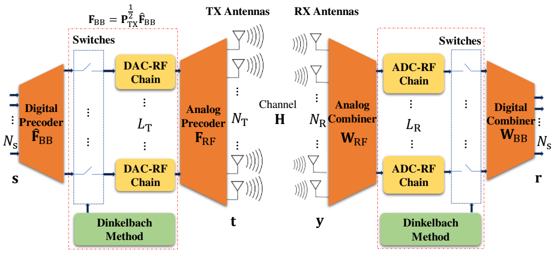

$\mathbf{F}_{\text{BB}} = \mathbf{P}_{\text{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\text{BB}}$
TX Antennas
RX Antennas
Switches
Switches
Digital Precoder $\hat{\mathbf{F}}_{\text{BB}}$
N_{s}
s
DAC-RF Chain
L_{T}
DAC-RF Chain
Analog Precoder $\mathbf{F}_{\text{RF}}$
N_{T}
Channel H
N_{R}
Analog Combiner $\mathbf{W}_{\text{RF}}$
ADC-RF Chain
L_{R}
ADC-RF Chain
Digital Combiner $\mathbf{W}_{\text{BB}}$
N_{s}
r
t
y
Dinkelbach Method
Dinkelbach Method

Fig. 1: A mmWave MIMO system with HBF architecture and the proposed DM framework.

B. MIMO System with HBF Architecture

Fig. 1 shows the system model considered in this paper where L_{T} is the number of available RF chains at the TX and L_{R} at the RX. Based on MIMO communication with HBF, we follow the conditions $N_{\text{s}} \leq L_{\text{T}} \leq N_{\text{T}}$ and $N_{\text{s}} \leq L_{\text{R}} \leq N_{\text{R}}$. The symbol vector $\mathbf{s} \in \mathbb{C}^{N_{\text{s}}\times 1}$ at the TX is such that $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \frac{1}{N_{\text{s}}}\mathbf{I}_{N_{\text{s}}}$. The digital precoder matrix right before the DAC-RF chain blocks is $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{L_{\text{T}}\times N_{\text{s}}} = \mathbf{P}_{\text{TX}}^{\frac{1}{2}}\hat{\mathbf{F}}_{\text{BB}}$ where $\hat{\mathbf{F}}_{\text{BB}}$ is the digital precoder matrix before the switches and $\mathbf{P}_{\text{TX}} \in \mathbb{R}^{L_{\text{T}}\times L_{\text{T}}}$ is a diagonal matrix with entries of power allocation values. We have tr$(\mathbf{P}_{\text{TX}}) = P_{\text{max}}$, where P_{max} is the maximum allocated power. The entries of the analogue precoder matrix $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_{\text{T}}\times L_{\text{T}}}$ are of constant modulus and this matrix models the phase shifting network which is only able to adjust the phase of the incoming signals, not the amplitude [8]. Note that the power constraint at the TX is satisfied by $\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 = P_{\text{max}}$. The matrices $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{L_{\text{R}}\times N_{\text{s}}}$ and $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_{\text{R}}\times L_{\text{R}}}$ denote the digital combiner and the analogue combiner at the RX, respectively. The analogue combiner matrix is also constant modulus.

We assume the channel state information (CSI) to be known at both the TX and the RX. Then the signal received at the RX antennas $\mathbf{y} \in \mathbb{C}^{N_{\text{R}}\times 1}$ can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{n} \in \mathbb{C}^{N_{\text{R}}\times 1} = \mathcal{CN}(0, \sigma_{\text{n}}^2)$ represents independent and identically distributed complex additive noise. After the analogue combiner and digital combiner units, the RX output signal can be expressed as

$$\mathbf{r} = \mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{y} = \mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{n}. \tag{3}$$

The mechanism to select only required number of RF chains L_{T}^{opt} out of the available L_{T} RF chains is implemented during the baseband processing. The proposed DM based solution drives this selection mechanism, which uses dynamic power allocation to decide on how many RF chains should be active during each channel realization. In the next section, we derive a fractional programming problem from the problem which maximizes EE and implement the Dinkelbach's approach to obtain the number of RF chains optimally at the TX/RX.

III. Overview of EE Maximization

In terms of the SE R (bits/s/Hz) and the power consumption P (W), the EE can be written as

$$\mathrm{EE}(\mathbf{P}_{\mathrm{TX}}) \triangleq \frac{R(\mathbf{P}_{\mathrm{TX}})}{P(\mathbf{P}_{\mathrm{TX}})} \text{ (bits/Hz/J).} \tag{4}$$

In (4), $\mathbf{P}_{\mathrm{TX}} \in \mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ represents a square matrix whose diagonal entries contain the transmission power of each data stream at the output of the digitally-computer precoder matrix, while all non-diagonal entries are zero. The notation $\mathcal{D}^{L_{\mathrm{T}} \times L_{\mathrm{T}}} \subset \mathbb{R}^{L_{\mathrm{T}} \times L_{\mathrm{T}}}$ represents the set of possible choices for $L_{\mathrm{T}} \times L_{\mathrm{T}}$ matrices, given the existence of a maximum transmit power constraint.

In order to represent the selection mechanism for RF chains at the digital precoder, we consider $[\mathbf{P}_{\mathrm{TX}}]_{kk} \in [0, P_{\mathrm{max}}] \, \forall \, k = 1, \ldots, L_{\mathrm{T}}$. The diagonal entries of $\mathbf{P}_{\mathrm{TX}}$ with a zero value means an open switch in the selection mechanism shown in Fig. 1. This means that the non-zero diagonal entries of the matrix $\mathbf{P}_{\mathrm{TX}}$ determine the number of the active RF chains currently selected at the TX side, i.e., $L_{\mathrm{T}}^{opt} = \|\mathbf{P}_{\mathrm{TX}}\|_0$. We may achieve high SE by increasing the number of RF chains, however, it increases power consumption as well. Thus, maximizing EE in (4) given suitable constraints on the solution provides us with a practical method for selecting the TX/RX configuration with the best performance trade-off.

The optimization problem in (4) has inspired us to study several different approaches to optimize the performance of a mmWave hybrid MIMO transceiver. As shown in Fig. 2, we deal with two phases in a single communication frame where we assume that at the start of each data frame, a beam training phase provides information to both the TX and RX about the current channel matrix $\mathbf{H}$ and there are L_{T} active RF chains. Based on this knowledge it is possible to adapt the behaviour of the TX and RX before the main data communication phase, where in this paper, the DM based solution is applied to activate only required number of RF chains, i.e., L_{T}^{opt}, which is obtained from the solution of EE maximization problem. In the process, the HBF matrices can be designed through an Euclidean distance minimization problem [8] as discussed in the next section and we also propose a low complexity alternative to design the HBF matrices. Next, we discuss the approaches which we implemented to adapt the behaviour of the TX and RX in order to achieve maximum EE.

*1) **RF Chain Selection:*** In Fig. 1, the analogue precoder and the analogue combiner may connect every RF chain to every TX/RX antenna, which is termed as a *fully-connected* structure. Alternatively, in a structure which is termed as *partially-connected*, each RF chain may only be connected to a subset of all the antennas. In the latter case, we have explored an optimization technique to select the best set of RF chains for data transmission in [20]. A key feature of this approach is that we use a low signal-to-noise ratio (SNR) approximation of the data rate to simplify the optimization approach. A sparse solution for the RF chains is desired and this is obtained by minimizing the number of non-zero entries in the matrix $\mathbf{P}_{\mathrm{TX}}$. This is achieved practically by using a technique called convex relaxation which allows the optimization to be performed efficiently. However, there is lack of research in literature dealing with the selection of RF chains. In a hardware setup, whether its fully-connected or partially-connected, when HBF is implemented on a field-programmable gate array (FPGA) chip, switching on only the needed RF chains would save a lot of power leading to an energy efficient communication system. Following that approach, in [18] we consider a fully-connected structure (as shown in Fig. 1) and the Dinkelbach's approach selects only that number of RF chains which maximizes EE and the complexity is kept minimum. More details of this approach are presented in Section IV below.

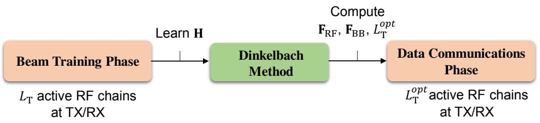

Learn $\mathbf{H}$
Compute
$\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$, L_{T}^{opt}
Beam Training Phase
Dinkelbach Method
Data Communications Phase
L_{T} active RF chains at TX/RX
L_{T}^{opt} active RF chains at TX/RX

Fig. 2: Single communication frame with two phases process: beam training and data communications.

*2) **Sampling Rate Selection:*** A number of papers recently have shown that using limited resolution digital-to-analogue or analogue-to-digital converters in the TX or RX can improve communications efficiency [21]. The reason for this is that the power consumed by a sampling device scales in an exponential manner with the number of quantization bits that are used. The limitation of using limited resolution sampling is that it can limit the overall data rate at high SNR values. However, limited resolution sampling can be particularly attractive for low or medium SNR values where the SE is lower. Reference [14] extends the RF chain selection approach of [18] to the case where the TX uses the fully-connected structure and each RF chain uses fixed resolution DACs at the TX. In that paper, a linear model is used to describe the impact of quantization, through a scaling factor and the addition of a noise term which represents the quantization noise. Similarly, the partially-connected case is with limited resolution sampling studied in [20]. We have recently extended this work to consider the joint optimization of both the HBF matrices design and the bit level resolution of each RF chain [15], [16]. This involves a complex model where the effect of the quantization noise on the data throughput is explicitly modelled and the bit level resolution can be adjusted to optimize the resulting EE. We introduce a novel matrix decomposition that is applied to the HBF matrices at both the TX and RX, i.e., the joint decomposition of a matrix representing analogue beamforming matrix, a second matrix modelling the impact of bit resolution on receiver noise and a third matrix that models digital baseband beamforming. Moreover, we address the joint TX-RX problem unlike in the existing literature and the optimization approach we follow requires the use of the alternating direction method of multipliers to find the best solution for both the HBF matrices and the required bit resolutions at the TX and RX in order to maximize EE.

Next, we describe the Dinkelbach's approach for selecting the number of RF chains optimally and show how this leads to a low-cost solution to EE maximization.

IV. RF Chain Selection for Maximum EE

A. RF Chain Selection Formulation

For MIMO with HBF and point-to-point communication, the SE R given the active number of RF chains is

$$R(\mathbf{P}_{\text{TX}}, \mathbf{P}_{\text{RX}}) = \log \left| \mathbf{I}_{N_s} + \frac{1}{\sigma_{\text{n}}^2} \mathbf{W}_{\text{BB}}^H \mathbf{P}_{\text{RX}}^{\frac{1}{2}} \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \times \mathbf{P}_{\text{TX}}^{\frac{1}{2}} \hat{\mathbf{F}}_{\text{BB}} \hat{\mathbf{F}}_{\text{BB}}^H \mathbf{P}_{\text{TX}}^{\frac{1}{2}} \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \mathbf{P}_{\text{RX}}^{\frac{1}{2}} \mathbf{W}_{\text{BB}} \right|, \quad (5)$$

where the real valued $L_{\text{T}} \times L_{\text{T}}$ matrix $\mathbf{P}_{\text{TX}}$ is the diagonal matrix allocating power at the TX side. At the RX, instead we use the $L_{\text{R}} \times L_{\text{R}}$ real-valued diagonal matrix $\mathbf{P}_{\text{RX}}$ with entries from $\{0, 1\}$, since this matrix represents the activated RF chains, thus, $L_{\text{R}}^{opt} = \|\mathbf{P}_{\text{RX}}\|_0$.

Following [8], we assume that $\hat{\mathbf{F}}_{\text{BB}} \hat{\mathbf{F}}_{\text{BB}}^H \approx \mathbf{I}_{L_{\text{T}}}$ and $\mathbf{W}_{\text{BB}} \mathbf{W}_{\text{BB}}^H \approx \mathbf{I}_{L_{\text{R}}}$, then the SE can be written as

$$R(\mathbf{P}_{\text{TX}}, \mathbf{P}_{\text{RX}}) = \log \left| \mathbf{I}_{L_{\text{R}}} + \frac{1}{\sigma_{\text{n}}^2} \mathbf{P}_{\text{RX}}^{\frac{1}{2}} \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{P}_{\text{TX}} \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \mathbf{P}_{\text{RX}}^{\frac{1}{2}} \right|. \quad (6)$$

The problem in (6) can be simplified by considering the TX side and the RX side separately. To compute the matrix $\mathbf{P}_{\text{TX}}$ it is assumed that the RX has activated all its RF chains, so that $\mathbf{P}_{\text{RX}} = \mathbf{I}_{L_{\text{R}}}$. In that case, the SE can be expressed as

$$R(\mathbf{P}_{\text{TX}}) = \log \left| \mathbf{I}_{L_{\text{R}}} + \frac{1}{\sigma_{\text{n}}^2} \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{P}_{\text{TX}} \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \right|. \quad (7)$$

Once the matrix $\mathbf{P}_{\text{TX}}$ is obtained, the matrix $\mathbf{P}_{\text{RX}}$ can be computed via the following SE expression:

$$R(\mathbf{P}_{\text{RX}}) = \log \left| \mathbf{I}_{L_{\text{R}}} + \frac{1}{\sigma_{\text{n}}^2} \mathbf{P}_{\text{RX}}^{\frac{1}{2}} \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{P}_{\text{TX}} \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \mathbf{P}_{\text{RX}}^{\frac{1}{2}} \right|. \quad (8)$$

Next, we focus on how to maximize the EE for the TX in order to select the optimal number of RF chains L_{T}^{opt}. The alternative of trying to solve (8) to maximize EE at the RX results leads to a complex integer programming optimization problem. In this paper, we will assume that the number of TX and RX spatial streams are the same, so that $L_{\text{R}}^{opt} = L_{\text{T}}^{opt}$.

Following [5], the total consumed power P for a HBF MIMO communication system can be expressed as

$$P = \beta \text{tr}(\mathbf{P}_{\text{TX}}) + 2P_{\text{CP}} + N_{\text{T}} P_{\text{T}} + N_{\text{R}} P_{\text{R}} + L_{\text{T}}^{opt} \times (P_{\text{RF}} + N_{\text{T}} P_{\text{PS}}) + L_{\text{R}}^{opt} (P_{\text{RF}} + N_{\text{R}} P_{\text{PS}}) \text{ (W)}, \quad (9)$$

where the power terms P_{CP}, P_{RF}, P_{PS}, P_{T} and P_{R} represent the power required by the circuit components, the power required by each RF chain, the power required by each phase shifter, the consumed power for each antenna at the TX and that required for each RX antenna, respectively. The parameter β is the reciprocal of amplifier efficiency.

Let us delete the subscript "TX" from $\mathbf{P}_{\text{TX}}$ in order to write simplified expressions. Hence, the EE maximization problem in (4) can be expressed with respect to $\mathbf{P} \in \mathbb{R}^{L_{\text{T}} \times L_{\text{T}}}$ as

$$\max_{\mathbf{P} \in \mathcal{D}^{L_{\text{T}} \times L_{\text{T}}}} \frac{R(\mathbf{P})}{P(\mathbf{P})} \text{ s. t. } P(\mathbf{P}) \leq P'_{\text{max}} \;\&\; R(\mathbf{P}) \geq R_{\text{min}}. \quad (10)$$

Note that the power constraint in (10) provides an upper limit on the power required for the HBF MIMO communication system, i.e., $P'_{\text{max}} = \beta P_{\text{max}} + 2P_{\text{CP}} + N_{\text{T}} P_{\text{T}} + N_{\text{R}} P_{\text{R}} + L_{\text{T}} \times (P_{\text{RF}} + N_{\text{T}} P_{\text{PS}}) + L_{\text{R}} (P_{\text{RF}} + N_{\text{R}} P_{\text{PS}})$. Next, we proceed with the proposed Dinkelbach's approach to obtain both the number of RF chains and the data streams optimally.

B. Dinkelbach's Approach to EE Maximization

In order to obtain a solution to (10) which is a fractional programming problem, we can implement the DM based solution. Dinkelbach's algorithm was first introduced in [22] and it appears to be an efficient algorithm to solve fractional problems. This is verified by the simulation results presented in Section V where we can observe that the Dinkelbach's approach achieves good performance. We can replace the EE ratio in (10) with an iterative sequence of difference-based optimizations as follows:

$$\max_{\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\text{T}} \times L_{\text{T}}}} \left\{ R(\mathbf{P}^{(m)}) - \nu^{(m)} P(\mathbf{P}^{(m)}) \right\} \text{ s. t. } P(\mathbf{P}) \leq P'_{\text{max}} \text{ and } R(\mathbf{P}) \geq R_{\text{min}}. \quad (11)$$

The DM involves a sequence of iterations where the constant $\nu^{(m)}$ is updated at each iteration based on the SE and power values estimated during the previous iteration which is equal to the ratio $R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^+$, for $m = 1, 2, \ldots, I_{\text{max}}$, where I_{max} denotes the maximum number of iterations. In order to reduce complexity compared to the BF method, we wish to use a SE expression that does not depend explicitly on the RF and baseband processing matrices. This avoids the need to compute the HBF matrices each time the number of selected RF chains is updated.

In order to proceed with the DM based solution, let us first update the SE and power expressions. For that, we consider channel's singular value decomposition (SVD) as $\mathbf{H} = \mathbf{U}_{\text{H}} \mathbf{\Sigma}_{\text{H}} \mathbf{V}_{\text{H}}^H$, where $\mathbf{U}_{\text{H}} \in \mathbb{C}^{N_{\text{R}} \times N_{\text{R}}}$ and $\mathbf{V}_{\text{H}} \in \mathbb{C}^{N_{\text{T}} \times N_{\text{T}}}$ are unitary matrices, and $\mathbf{\Sigma}_{\text{H}} \in \mathbb{R}^{N_{\text{R}} \times N_{\text{T}}}$ represents a matrix which is rectangular in nature where the diagonal entries contain the singular values of the channel matrix and all the other entries are zero. Considering the SVD of the channel, (7) is written as

$$R(\mathbf{P}) = \log \left| \mathbf{I}_{N_{\text{R}}} + \frac{1}{\sigma_{\text{n}}^2} \mathbf{W}_{\text{RF}}^H \mathbf{U}_{\text{H}} \mathbf{\Sigma}_{\text{H}} \mathbf{V}_{\text{H}}^H \mathbf{F}_{\text{RF}} \times \mathbf{P} \mathbf{F}_{\text{RF}}^H \mathbf{V}_{\text{H}} \mathbf{\Sigma}_{\text{H}}^H \mathbf{U}_{\text{H}}^H \mathbf{W}_{\text{RF}} \right|. \quad (12)$$

Using the approach given in [8], it can be shown that $\mathbf{V}_{\text{H}}^{H}\mathbf{F}_{\text{RF}} \approx [\mathbf{I}_{L_{\text{T}}}\, \mathbf{0}^{T}_{(N_{\text{T}}-L_{\text{T}})\times L_{\text{T}}}]^{T}$ and $\mathbf{U}_{\text{H}}^{H}\mathbf{W}_{\text{RF}} \approx [\mathbf{I}_{L_{\text{R}}}\, \mathbf{0}^{T}_{(N_{\text{R}}-L_{\text{R}})\times L_{\text{R}}}]^{T}$, hence,

$$R(\mathbf{P}) = \log\left|\mathbf{I}_{N_{\text{R}}} + \frac{1}{\sigma_{\text{n}}^{2}}\bar{\mathbf{\Sigma}}^{2}\mathbf{P}\right|, \tag{13}$$

where the $L_{\text{R}} \times L_{\text{T}}$ matrix $\bar{\mathbf{\Sigma}}$ has diagonal entries $[\bar{\mathbf{\Sigma}}]_{kk} = [\mathbf{\Sigma}_{\text{H}}]_{kk}$ for $k = 1, \ldots, L_{\text{T}}$, assuming $L_{\text{T}} = L_{\text{R}}$. Again, the remaining entries of this matrix are zero. In (13) all of the matrices are diagonal, so it is possible to decompose the SE calculation into L_{T} parallel and orthogonal channels as

$$R(\mathbf{P}) \approx \sum_{k=1}^{L_{\text{T}}} \log\left(1 + \frac{1}{\sigma_{\text{n}}^{2}}[\bar{\mathbf{\Sigma}}^{2}]_{kk}[\mathbf{P}]_{kk}\right) \text{ (bits/s/Hz).} \tag{14}$$

The number of available RF chains at the TX L_{T} and at the RX L_{R} are determined by the hardware setup of the transceiver. For the TX side, the power values in the matrix $\mathbf{P}$ can be written as

$$P_{\text{TX}}(\mathbf{P}) = P_{\text{static}} + \sum_{k=1}^{L_{\text{T}}}(\beta[\mathbf{P}]_{kk} + P_{\text{RF}} + N_{\text{T}}P_{\text{PS}}) \tag{15}$$

$$\Longrightarrow P_{\text{TX}}(\mathbf{P}) = P_{\text{static}} + \sum_{k=1}^{L_{\text{T}}} \beta'[\mathbf{P}]_{kk} \text{ (W)}, \tag{16}$$

where the value of $P_{\text{static}} \triangleq P_{\text{CP}} + N_{\text{T}}P_{\text{T}}$ does not depend on the entries of the matrix $\mathbf{P}$ and $\beta' \triangleq \beta + \frac{P_{\text{RF}}+N_{\text{T}}P_{\text{PS}}}{P_{\text{max}}}$. Simplifying (15) into the form given in (16) is possible as $\sum_{k=1}^{L_{\text{T}}}[\mathbf{P}]_{kk} = \text{tr}(\mathbf{P}) = P_{\text{max}}$.

Following (14)-(16), the m-th DM step can be written as

$$\{\mathbf{P}^{(m)}, \nu^{(m)}\} = \arg \max_{\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\text{T}}\times L_{\text{T}}}} \mathcal{G}(\mathbf{P}^{(m)}\nu^{(m)}),$$

$$\text{s. t. } P(\mathbf{P}) \leq P'_{\text{max}} \text{ and } R(\mathbf{P}) \geq R_{\text{min}}, \tag{17}$$

where $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)}) \triangleq \sum_{k=1}^{L_{\text{T}}} \log\left(1 + \frac{1}{\sigma_{\text{n}}^{2}}[\bar{\mathbf{\Sigma}}^{2}]_{kk}[\mathbf{P}^{(m)}]_{kk}\right) - \nu^{(m)}\sum_{k=1}^{L_{\text{T}}} \beta'[\mathbf{P}^{(m)}]_{kk}$. Note that (17) is generally not convex given the constraint associated with $\mathbf{P}^{(m)}$, i.e., $\mathbf{P}^{(m)} \in \mathcal{D}^{L_{\text{T}}\times L_{\text{T}}}$. Indeed, in the case where the set $\mathcal{D}$ also contains the zero value, the problem (17) is a mixed-integer programming one. To proceed, we alleviate this constraint on $\mathbf{P}^{(m)}$ first, so that (17) can be solved using a standard interior-point method, e.g., using CVX [23]. A theoretical analysis of DM convergence is presented in [26].

In order to explain the steps of Algorithm 1, it begins with the maximum number of RF chains L_{T}. Step 4 shows that we solve (17) to update $\mathbf{P}^{(m)}$ using CVX after alleviating the constraint as mentioned above. Then we apply the constraint again as highlighted in Step 5 of Algorithm 1. This is achieved by setting the values $\mathbf{P}^{(m)}$ to zero when they fall below the tolerance value ϵ_{th} (see Table II for ϵ_{th} value). Step 6 shows that counting the non-zero values of $\mathbf{P}_{\text{th}}^{(m)}$ determines the number of activated RF chains. The DM method keeps updating these values within the loop and finally computes $\|\mathbf{P}_{\text{th}}^{(m)}\|_0$ when the loop ends. Step 7 determines the SE $R(\mathbf{P}^{(m)})$ and the power $P_{\text{TX}}(\mathbf{P}^{(m)})$, and in Step 8 $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$ is computed based on its given expression above, where $\nu^{(m)} = R(\mathbf{P}^{(m-1)})/P(\mathbf{P}^{(m-1)}) \in \mathbb{R}^{+}$. Step 9 is used to update $\nu^{(m)}$ according to the current value $R(\mathbf{P}^{(m)})/P_{\text{TX}}(\mathbf{P}^{(m)})$. The loop terminates when $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})|$ is lower than the specified value ϵ, which is determined empirically (see Table II for ϵ value). The number of spatial streams is then set to be equal to the optimal number of RF chains, i.e., $N_{\text{s}} = L_{\text{T}}^{opt}$.

Algorithm 1 Dinkelbach Method (DM)

```
1: Initialize: $\mathbf{P}^{(0)}$, choose tolerance $\epsilon$, $L_{\text{T}}$ and set $\nu^{(0)}$ with $\mathcal{G}(\mathbf{P}^{(0)}, \nu^{(0)}) \geq 0$.
2: Start Iteration Step $m = 0$.
3: while $|\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})| > \epsilon$ do
4:     Alleviate the constraint on $\mathbf{P}^{(m)}$ and solve (17).
5:     Threshold the entries of $\mathbf{P}^{(m)} \rightarrow$ obtain $\mathbf{P}_{\text{th}}^{(m)}$.
6:     Count non-zero values of $\mathbf{P}_{\text{th}}^{(m)} \rightarrow$ update $L_{\text{T}}^{opt}$.
7:     Calculate $R(\mathbf{P}^{(m)})$ and $P_{\text{TX}}(\mathbf{P}^{(m)})$ using (14)-(16).
8:     Compute $\mathcal{G}(\mathbf{P}^{(m)}, \nu^{(m)})$.
9:     Update the value $\nu^{(m)}$ as $R(\mathbf{P}^{(m)})/P_{\text{TX}}(\mathbf{P}^{(m)})$.
10:    Update $m = m + 1$ for next iteration.
11: end while
12: Compute $L_{\text{T}}^{opt}$ as the value $\|\mathbf{P}_{\text{th}}^{(m)}\|_0$.
```

Once we obtain L_{T}^{opt}, L_{R}^{opt} ($= L_{\text{T}}^{opt}$) and N_{s}, we can design the HBF matrices $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$, $\mathbf{W}_{\text{RF}}$ and $\mathbf{W}_{\text{BB}}$. We assume that as in [8], the matrices $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ can be designed to yield a good approximation of the fully digital precoder $\mathbf{F}_{\text{DBF}}$. Note that the precoder matrix $\mathbf{F}_{\text{DBF}} = \mathbf{V}_{\text{H1}}\mathbf{P}_{\text{TX}}^{(1/2)}$ where the matrix $\mathbf{V}_{\text{H1}} \in \mathbb{C}^{N_{\text{T}}\times N_{\text{s}}}$ consists of the N_{s} columns of the matrix $\mathbf{V}_{\text{H}}$ which contains the right singular eigenvectors [8] with $\|\mathbf{F}_{\text{DBF}}\|_F^2 = \text{tr}(\mathbf{P}_{\text{TX}}) = P_{\text{max}}$. Following [8], the problem to compute the hybrid precoder decomposition $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ through Euclidean distance minimization can be transformed to a sparse approximation problem. To solve that, we use gradient pursuit (GP) algorithm [24] which is implemented as an alternative to the most commonly used orthogonal matching pursuit (OMP) algorithm for HBF design. The GP algorithm has same performance as the OMP algorithm, but it uses only one matrix vector multiplication per iteration to avoid matrix inversion, leading to faster approximation and low complexity [9]. At the RX, the hybrid combiner can be designed with a similar mathematical formulation as at the TX except there is no power constraint. Following the steps in [8], we compute the fully digital combiner matrix $\mathbf{W}_{\text{DBF}}$ and the Euclidean distance minimization problem for the combiner design is transformed to the sparse approximation problem likewise at the TX. The sparse approximation problem at the RX can then be solved by the GP algorithm [9] in order to obtain the hybrid combiner decomposition $\mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}$.

Computational Complexity: The computation for the DM based solution requires only $\mathcal{O}(L_{\text{T}}^{opt})$ operations per iteration. The complexity comparison with the BF approach is provided in Section V. The complexity order in computing beamforming weights for the GP algorithm is $\mathcal{O}\big((L_{\text{T}}^{opt})^{3}N_{\text{T}}\big)$ and for the OMP algorithm equals $\mathcal{O}\big((L_{\text{T}}^{opt})^{4}\big) + \mathcal{O}\big((L_{\text{T}}^{opt})^{3}$ - the GP

System Parameter	Value
Number of clusters	$N_{\text{cl}} = 2$
Number of rays	$N_{\text{ray}} = 10$
Angular spread	$7.5°$
Average power for each cluster	$\sigma_{\alpha,i} = 1$
Mean angles (azimuth domain)	$60° - 120°$
Mean angles (elevation domain)	$80° - 100°$
Normalized system bandwidth	1 Hz
SNR	$1/\sigma_{\text{n}}^2$
Amplifier efficiency	$1/\beta = 0.4$
Minimum desired SE in (10)	$R_{\text{min}} = 1$ bits/s/Hz
Tolerance values	$\epsilon = 10^{-4}$ and $\epsilon_{\text{th}} = 10^{-6}$
Number of available RF chains	$L_{\text{T}} = L_{\text{R}} = \text{length}\big(\text{eig}(\mathbf{HH}^H)\big)$
Spacing between antenna elements	$d = \lambda/2$ (e.g., $\lambda = 1/28$ GHz [13])

(a) Values of the system parameters.

Power Term	Value
Power required by all circuit components	$P_{\text{CP}} = 10$ W
Power required by each RF chain	$P_{\text{RF}} = 100$ mW
Power required by each phase shifter	$P_{\text{PS}} = 10$ mW
Power per TX/RX antenna element	$P_{\text{T}} = P_{\text{R}} = 100$ mW
Maximum allocated power	$P_{\text{max}} = 1$ W

(b) Values of the power terms in (9) [25].

TABLE II: Values of the system parameters and power terms used in the simulations.

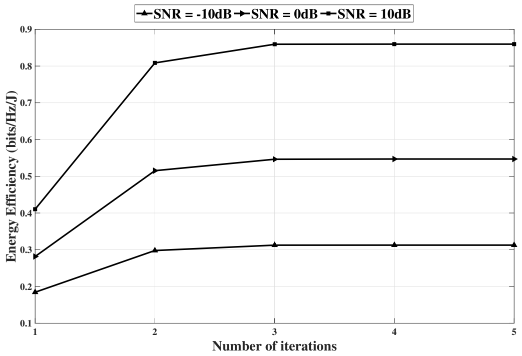


Fig. 3: EE versus number of iterations at $N_{\text{T}} = 32$, $N_{\text{R}} = 8$, $N_{\text{cl}} = 2$, $N_{\text{ray}} = 10$ and $P_{\text{max}} = 16$ W.

method only makes use of matrix multiplies at each step. This reduction in complexity comes from using a gradient computation in place of a full matrix inverse calculation. Reference [9] provides a more detailed complexity comparison. Next, we present simulation results that verify the good performance of the proposed Dinkelbach approach.

V. Simulation Results

This section evaluates the performance of the proposed DM based solution and compares it with existing baseline cases. All results have been averaged over 1,000 Monte-Carlo realizations. In terms of the system setup, Table II (a) provides the values of all the system parameters and Table II (b) provides the values used in the simulations for the power terms in (9).

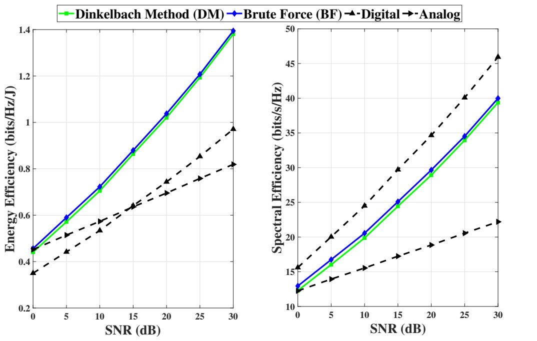


Fig. 4: EE and SE versus SNR at $N_{\text{T}} = 32$, $N_{\text{R}} = 8$, $N_{\text{cl}} = 2$, $N_{\text{ray}} = 10$ and $P_{\text{max}} = 1$ W.

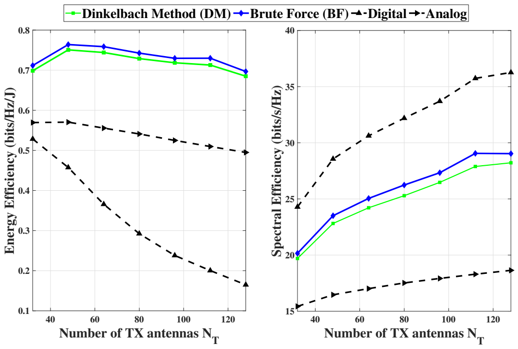


Fig. 5: EE and SE versus N_{T} at SNR = 10 dB, $N_{\text{R}} = 8$, $N_{\text{cl}} = 2$, $N_{\text{ray}} = 10$ and $P_{\text{max}} = 1$ W.

For comparison with the proposed DM based solution, following baseline cases have been considered in this paper.

1) BF Approach: The exhaustive search based approach in [13], i.e., the BF approach, at each realization (current channel realization), computes the EE performance by designing the beamforming matrices for each possible choice of activated RF chains, namely $L_{\text{T}} = \{1, 2, ..., N_{\text{T}}\}$, and then chooses the corresponding number of RF chains corresponding to the highest EE value. In contrast, the proposed DM based solution does not need to iterate for all possible number of RF chains and then find a number of RF chains which is optimal, which reduces the complexity significantly while providing high energy efficient solution. The complexity order of the BF approach is related the number of RF chains multiplied by the total number of antennas, i.e., $\mathcal{O}\big(L_{\text{T}}^{opt} N_{\text{T}}\big)$ which is larger than that of the DM based solution that only requires $\mathcal{O}(L_{\text{T}}^{opt})$ operations per iteration. In simulation, the BF and DM approaches uses the same HBF matrix computation.

2) Digital Beamforming: As mentioned above, the full digital beamforming baseline allocates one active RF chain for each antenna in all simulations, i.e., $L_{\text{T}} = N_{\text{T}}$ and $L_{\text{R}} = N_{\text{R}}$.

3) Analogue Beamforming: In this case, analogue beamforming only implements one active RF chain , i.e., $L_{\mathrm{T}} = L_{\mathrm{R}} = 1$, and the HBF decomposition matrices are designed equal to phases of the first singular vectors.

Fig. 3 graphs the EE performance versus the number of iterations for SNR values of -10, 0 and 10 dB to observe convergence of the proposed DM based solution at $N_{\mathrm{T}} = 32$, $N_{\mathrm{R}} = 8$, $N_{\mathrm{cl}} = 2$, $N_{\mathrm{ray}} = 10$ and $P_{\mathrm{max}} = 16$ W. The DM based solution converges rapidly, requiring typically about two iterations to achieve an optimal solution at each channel realization. Also, the achieved EE results increase with the SNR value, for example, after 2 iterations, the EE value at 10 dB SNR is ≈ 0.55 bits/Hz/J higher than that for -10 dB SNR and ≈ 0.3 bits/Hz/J higher than the result for 0 dB SNR.

Fig. 4 shows the EE and SE performance of the DM method along with the BF approach, and both the analogue and digital baseline cases versus SNR with $N_{\mathrm{T}} = 32$, $N_{\mathrm{R}} = 8$, $N_{\mathrm{cl}} = 2$, $N_{\mathrm{ray}} = 10$ and $P_{\mathrm{max}} = 1$ W. We can observe that the DM based solution has similar EE and SE performance to the BF approach, achieving a much higher EE than the digital baseline case, and higher EE and SE results compared to the analogue baseline. At an SNR value of 20 dB, the DM based solution yield ≈ 0.2 bits/Hz/J higher EE than the digital baseline case, and ≈ 10 bits/s/Hz higher SE and about 0.3 bits/Hz/J higher EE than the analogue baseline case.

Fig. 5 shows the EE and SE performance versus the number of TX antennas, N_{T}, plotted for an SNR of 10 dB, $N_{\mathrm{R}} = 8$, $N_{\mathrm{cl}} = 2$, $N_{\mathrm{ray}} = 10$ and $P_{\mathrm{max}} = 1$ W. It is clear that as the number of antennas increases, the EE results start to decrease for both the proposed DM based solution and the existing baseline cases. For example, at $N_{\mathrm{T}} = 80$, the EE and SE performance of the DM based solution is similar to that of the BF method. Also, the DM based solution has ≈ 0.42 bits/Hz/J higher EE than the digital baseline case, and ≈ 7.5 bits/s/Hz higher SE and about 0.2 bits/Hz/J higher EE than the analogue baseline case.

VI. Conclusion

This paper has discussed the concept of adaptive HBF MIMO systems that adapt their behaviour on a frame-by-frame basis to optimize EE. In particular, a DM based solution has been studied to enable fractional programming to maximize the EE of the candidate transmitter and receiver architectures in a low-cost manner. The DM method described in this paper can achieve EE and SE performance similar to the exhaustive search based BF approach, while reducing the complexity significantly. Once the number of RF chains is selected, the proposed technique needs to compute the HBF matrices only once. Further, the DM solution can also provide significantly improved EE performance when compared with the existing baseline cases, e.g., at 10 dB SNR, it performs $\approx 20\%$ better than the digital beamforming baseline and $\approx 15\%$ better than the analogue beamforming case. Finally it is shown that the GP algorithm, which is used to compute the HBF matrices, is a faster and less complex algorithm in comparison to the state-of-the-art OMP algorithm.

Acknowledgement

The Engineering and Physical Sciences Research Council Grant EP/P000703/1 supported this research work.

References

[1] NetWorld2020 White Paper on Research Beyond 5G, pp. 1-43, Oct. 2015.

[2] Cisco visual networking index forecast 2017-22.

[3] Ericsson Mobility Report, pp. 1-32, Nov. 2018.

[4] S. Rangan et al., "Millimeter-wave cellular wireless networks: potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[5] S. Han et al., "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.

[6] R. W. Heath et al., "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[7] O. E. Ayach et al., "The capacity optimality of beam steering in large millimeter wave MIMO systems," *IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, pp. 100-104, June 2012.

[8] O. E. Ayach et al., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[9] A. Kaushik et al.,"Sparse hybrid precoding and combining in millimeter wave MIMO systems," in *Proc. IET Radio Prop. Tech. 5G*, Durham, UK, pp. 1-7, Oct. 2016.

[10] C. G. Tsinos et al., "Hybrid analog-digital transceiver designs for multi-user MIMO mmWave cognitive radio systems," *IEEE Trans. Cognitive Commun. Netw.*, accepted, Aug. 2019.

[11] A. Li and C. Masouros, "Hybrid analog-digital millimeter-wave mu-mimo transmission with virtual path selection," *IEEE Commun. Letters*, vol. 21, no. 2, pp. 438-441, Feb. 2017.

[12] S. Payami et al., "Hybrid beamforming with a reduced number of phase shifters for massive mimo systems," *IEEE Trans. Veh. Tech.*, vol. 67, no. 6, pp. 4843-4851, June 2018.

[13] R. Zi et al., "Energy efficiency optimization of 5G radio frequency chain systems", *IEEE Journ. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.

[14] A. Kaushik et al., "Energy efficiency maximization of millimeter wave hybrid MIMO systems with low resolution DACs," *IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, pp. 1-6, May 2019.

[15] A. Kaushik et al., "Energy efficient bit allocation and hybrid combining for millimeter wave MIMO systems," *IEEE Global Commun. Conf. (GLOBECOM)*, Hawaii, USA, pp. 1-6, Dec. 2019.

[16] A. Kaushik et al., "Joint bit allocation and hybrid beamforming optimization for energy efficient millimeter wave MIMO systems," *arXiv:1910.01479*, Oct. 2019.

[17] E. Bjornson et al., "Optimal design of energy-efficient multi-user MIMO systems: is massive MIMO the answer?", *IEEE Trans. Wireless Commun.*, vol. 14, no.6, pp. 3059-3075, Jun. 2015.

[18] A. Kaushik et al., "Dynamic RF chain selection for energy efficient and low complexity hybrid beamforming in millimeter wave MIMO systems," *IEEE Trans. Green Commun. Netw.*, accepted, July 2019.

[19] S. Singh et al., "Interference analysis for highly directional 60-GHz mesh networks: the case for rethinking medium access control," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513-1527, Oct. 2011.

[20] E. Vlachos et al., "Energy efficient transmitter with low resolution DACs for massive MIMO with partially connected hybrid architecture," *IEEE Veh. Tech. Conf. (VTC)-Spring*, Porto, Portugal, pp. 1-5, June 2018.

[21] O. Orhan et al., "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," *Info. Theory Appl. Workshop (ITA)*, San Diego, USA, pp. 191-198, 2015.

[22] W. Dinkelbach, "On nonlinear fractional programming", *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.

[23] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs", in *Recent Adv. Learning and Control*, Springer-Verlag Ltd., pp. 95-110, 2008.

[24] T. Blumensath, and M. E. Davies, "Gradient pursuits", *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370-2382, June 2008.

[25] T. S. Rappaport et al., "Millimeter wave wireless communications," *Prentice-Hall*, Sept. 2014.

[26] A. Zappone and E. Jorswieck, "Energy Efficiency in Wireless Networks via Fractional Programming Theory," in *Energy Efficiency in Wireless Networks via Fractional Programming Theory*, Now Foun. Trends, 2015.

Energy Efficient ADC Bit Allocation and Hybrid Combining for Millimeter Wave MIMO Systems

Aryan Kaushik[1], Christos Tsinos[2], Evangelos Vlachos[1], John Thompson[1]
[1]Institute for Digital Communications, The University of Edinburgh, United Kingdom.
[2]Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg.
Emails: {a.kaushik, e.vlachos, j.s.thompson}@ed.ac.uk, christos.tsinos@uni.lu

***Abstract*—Low resolution analog-to-digital converters (ADCs) can be employed to improve the energy efficiency (EE) of a wireless receiver since the power consumption of each ADC is exponentially related to its sampling resolution and the hardware complexity. In this paper, we aim to jointly optimize the sampling resolution, i.e., the number of ADC bits, and analog/digital hybrid combiner matrices which provides highly energy efficient solutions for millimeter wave multiple-input multiple-output systems. A novel decomposition of the hybrid combiner to three parts is introduced: the analog combiner matrix, the bit resolution matrix and the baseband combiner matrix. The unknown matrices are computed as the solution to a matrix factorization problem where the optimal, fully digital combiner is approximated by the product of these matrices. An efficient solution based on the alternating direction method of multipliers is proposed to solve this problem. The simulation results show that the proposed solution achieves high EE performance when compared with existing benchmark techniques that use fixed ADC resolutions.**

***Index Terms*—energy efficient design, optimal bit resolution and hybrid combining, mmWave MIMO.**

I. Introduction

The analog/digital (A/D) hybrid beamforming architectures for millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems reduce the hardware complexity and the power consumption through fewer radio frequency (RF) chains and support multi-stream communication with good capacity performance [1]–[3]. Designing such systems for high energy efficiency (EE) gains would leverage their significance [4], [5]. An alternative solution to reduce the power consumption and hardware complexity is by reducing the resolution sampling [6]. Some approaches have been applied in hybrid mmWave MIMO systems for EE maximization and low complexity with full resolution [7] and low resolution [8].

The existing literature mostly discusses full or high resolution analog-to-digital converters (ADCs) with a small number of RF chains or low resolution ADCs with a large number of RF chains: either way only the fixed resolution ADCs are taken into account. References [4], [5] consider EE optimization problems for A/D hybrid transceivers but with fixed and high resolution digital-to-analog converters (DACs)/ADCs. Reference [8] proposes a novel EE maximization transmission technique with subset selection optimization to find the best subset of the active RF chains and DAC resolution, which can be extended to low resolution ADCs at the receiver (RX). Reference [9] suggests implementing fixed and low resolution ADCs with few RF chains. Reference [10] studies the idea of a mixed-ADC architecture where a better energy-rate trade off is achieved by using mixed resolution ADCs but still with a fixed resolution for each ADC and it does not consider A/D hybrid beamforming. A hybrid beamforming system with fixed and low resolution ADCs has been analyzed for channel estimation in [11]. Varying resolution ADCs can be implemented at the RX [12] which may provide a better solution than fixed and low resolution ADCs. Extra care is needed when deciding the range of number of ADC bits as the total ADC power consumption can be dominated by only a few high resolution ADCs. Thus, a good trade-off between power consumption and performance is to consider the range of 1-8 bits for the varying number of ADC bits.

Contributions: This paper designs an optimal EE solution for a mmWave A/D hybrid receiver MIMO system by introducing the novel decomposition of the A/D hybrid combiner to three parts representing the analog combiner matrix, the bit resolution matrix and baseband combiner matrix. Our aim is to minimize the distance between this decomposition, which is expressed as the product of three matrices, and the fully digital combiner matrix. The joint problem is decomposed into a series of sub-problems which are solved using an alternating optimization framework, i.e., alternating direction method of multipliers (ADMM) is developed to obtain the unknown matrices. The proposed design has high flexibility, given that the analog combiner is codebook-free, thus there is no restriction on the angular vectors and different bit resolutions can be assigned to each ADC. Our proposed solution optimizes the resolution on a packet-by-packet basis for each one of the ADCs unlike existing approaches that are based on fixed resolution sampling. We also implement an exhaustive search approach [4] for comparison which provides the upper bound for EE maximization.

Notation: $\mathbf{A}$, $\mathbf{a}$ and a denote a matrix, a vector and a scalar, respectively. The complex conjugate transpose and transpose of $\mathbf{A}$ are denoted as $\mathbf{A}^H$ and $\mathbf{A}^T$; $|a|$ represents the determinant of a; $\mathbf{I}_N$ represents $N \times N$ identity matrix; $\mathbf{X} \in \mathbb{C}^{A \times B}$ and $\mathbf{X} \in \mathbb{R}^{A \times B}$ denote $A \times B$ size $\mathbf{X}$ matrix with complex and real entries, respectively; $\mathcal{CN}(\mathbf{a}, \mathbf{A})$ denotes a complex Gaussian vector having mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$; $[\mathbf{A}]_{kl}$ is the matrix entry at the k-th row and l-th column. The indicator function $\mathbb{1}_{\mathcal{S}}\{\mathbf{A}\}$ of a set $\mathcal{S}$ that acts over a matrix $\mathbf{A}$ is defined as $0 \; \forall \; \mathbf{A} \in \mathcal{S}$ and $\infty \; \forall \; \mathbf{A} \notin \mathcal{S}$.

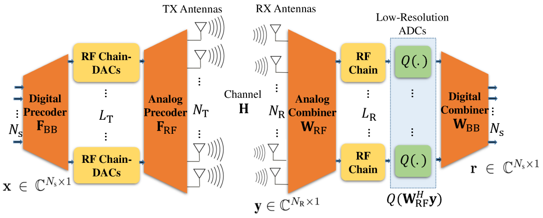

TX Antennas
RX Antennas
Low-Resolution ADCs
Digital Precoder $\mathbf{F}_{\text{BB}}$
N_{s}
RF Chain-DACs
L_{T}
RF Chain-DACs
Analog Precoder $\mathbf{F}_{\text{RF}}$
N_{T}
Channel $\mathbf{H}$
N_{R}
Analog Combiner $\mathbf{W}_{\text{RF}}$
RF Chain
L_{R}
RF Chain
$Q(.)$
$Q(.)$
Digital Combiner $\mathbf{W}_{\text{BB}}$
N_{s}
$\mathbf{x} \in \mathbb{C}^{N_{\text{s}}\times 1}$
$\mathbf{y} \in \mathbb{C}^{N_{\text{R}}\times 1}$
$Q(\mathbf{W}_{\text{RF}}^H\mathbf{y})$
$\mathbf{r} \in \mathbb{C}^{N_{\text{s}}\times 1}$

Fig. 1. A mmWave A/D hybrid MIMO system with low resolution ADCs.

II. A/D Hybrid MmWave MIMO System

A. MmWave Channel Model

MmWave channels can be modeled by a narrowband clustered channel model due to different channel settings such as number of multipaths, amplitudes, etc., with N_{cl} clusters and N_{ray} propagation paths in each cluster [1]. Considering a single user mmWave system with N_{T} antennas at the transmitter (TX), transmitting N_{s} data streams to N_{R} antennas at the RX, the mmWave channel matrix can be written as follows:

$$\mathbf{H} = \sqrt{\frac{N_{\text{T}}N_{\text{R}}}{N_{\text{cl}}N_{\text{ray}}}} \sum_{i=1}^{N_{\text{cl}}} \sum_{l=1}^{N_{\text{ray}}} \alpha_{il} \mathbf{a}_{\text{R}}(\phi_{il}^r)\mathbf{a}_{\text{T}}(\phi_{il}^t)^H, \tag{1}$$

where $\alpha_{il} \in \mathcal{CN}(0, \sigma_{\alpha,i}^2)$ is the gain term with $\sigma_{\alpha,i}^2$ being the average power of the i^{th} cluster. Furthermore, $\mathbf{a}_{\text{T}}(\phi_{il}^t)$ and $\mathbf{a}_{\text{R}}(\phi_{il}^r)$ represent the normalized transmit and receive array response vectors [1], where ϕ_{il}^t and ϕ_{il}^r denote the azimuth angles of departure and arrival, respectively. We use uniform linear array (ULA) antennas for simplicity and model the antenna elements at the RX as ideal sectored elements [13]. However, the proposed technique is not limited to this setup and can be easily extended to the case of wideband channels and uniform planar/circular arrays.

B. A/D Hybrid MIMO System Model

Based on the A/D hybrid beamforming scheme in the large-scale mmWave MIMO communication systems, the number of RX RF chains L_{R} follows the limitation $N_{\text{s}} \leq L_{\text{R}} \leq N_{\text{R}}$ [1], [2]. The matrices $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_{\text{R}}\times L_{\text{R}}}$ and $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{L_{\text{R}}\times N_{\text{s}}}$ denote the analog combiner and baseband (or digital) combiner matrices, respectively. The analog combiner matrix $\mathbf{W}_{\text{RF}}$ is based on phase shifters, i.e., the elements that have unit modulus and continuous phase. Thus, $\mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}}\times L_{\text{R}}}$ where the set $\mathcal{W}$ represents the set of possible phase shifts in $\mathbf{W}_{\text{RF}}$ and for a variable a, is defined as, $\mathcal{W} = \{a \in \mathbb{C} \mid |a| = 1\}$. At the TX, with L_{T} RF chains, the analog precoder matrix is denoted as $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_{\text{T}}\times L_{\text{T}}}$ and the baseband precoder matrix is denoted as $\mathbf{F}_{\text{BB}} \in \mathbb{C}^{L_{\text{T}}\times N_{\text{s}}}$. The received signal $\mathbf{y} \in \mathbb{C}^{N_{\text{R}}\times 1}$ can be expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{x} \in \mathbb{C}^{N_{\text{s}}\times 1}$ is the transmit symbol vector and $\mathbf{n} \in \mathbb{C}^{N_{\text{R}}\times 1}$ is a noise vector with independent and identically distributed entries and follow the complex Gaussian distribution with zero mean and σ_{n}^2 variance, i.e., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\text{n}}^2\mathbf{I}_{N_{\text{R}}})$.

As widely used in the existing literature, we consider the linear additive quantization noise model (AQNM) to represent the distortion of quantization [14]. Given that $Q(\cdot)$ denotes a uniform scalar quantizer then for the scalar complex input $x \in \mathbb{C}$ that is applied to both the real and imaginary parts, we have that,

$$Q(x) \approx \delta x + \epsilon, \tag{3}$$

where $\delta = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b}} \in [m, M]$ is the multiplicative distortion parameter for a bit resolution equal to b [15] where m and M denote the minimum and maximum value of the range. Note that the introduced error in the linear approximation in (3) decreases for larger resolutions. However, our proposed solution focuses on EE maximization and this linear approximation does not impact the performance significantly as observed from the simulation results in Section IV. The parameter ϵ is the additive quantization noise with $\epsilon \sim \mathcal{CN}(0, \sigma_\epsilon^2)$, where $\sigma_\epsilon = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b}}\sqrt{\frac{\pi\sqrt{3}}{2}2^{-2b}}$. Based on AQNM, the vector containing the complex output of all the ADCs can be expressed as follows:

$$Q(\mathbf{W}_{\text{RF}}^H\mathbf{y}) \approx \boldsymbol{\Delta}^H\mathbf{W}_{\text{RF}}^H\mathbf{y} + \boldsymbol{\epsilon}, \tag{4}$$

where $Q(\mathbf{W}_{\text{RF}}^H\mathbf{y}) \in \mathbb{C}^{L_{\text{R}}\times 1}$ and $\boldsymbol{\Delta} = \boldsymbol{\Delta}^H \in \mathbb{C}^{L_{\text{R}}\times L_{\text{R}}}$ is a diagonal matrix with values depending on the ADC resolution b_i of each ADC. Specifically, each diagonal entry of $\boldsymbol{\Delta}$ is given by:

$$[\boldsymbol{\Delta}]_{ii} = \sqrt{1 - \frac{\pi\sqrt{3}}{2}2^{-2b_i}} \in [m, M] \; \forall\, i = 1, \ldots, L_{\text{R}}, \tag{5}$$

where, for simplicity, we assume that the range $[m, M]$ is the same for each one of the ADCs. The second term of (4) expresses the additive quantization noise for all RF chains, with $\boldsymbol{\epsilon} \in \mathcal{CN}(\mathbf{0}, \mathbf{C}_\epsilon)$ [8] where $\mathbf{C}_\epsilon$ is a diagonal covariance matrix with entries as follows:

$$[\mathbf{C}_\epsilon]_{ii} = \left(1 - \frac{\pi\sqrt{3}}{2}2^{-2b_i}\right)\left(\frac{\pi\sqrt{3}}{2}2^{-2b_i}\right) \forall\, i = 1, \ldots, L_{\text{R}}. \tag{6}$$

After the effect of the quantization and application of the baseband combining matrix, the output $\mathbf{r} \in \mathbb{C}^{N_{\text{s}}\times 1}$ at the RX can be expressed as:

$$\mathbf{r} = \mathbf{W}_{\text{BB}}^H\boldsymbol{\Delta}^H\mathbf{W}_{\text{RF}}^H\mathbf{y} + \mathbf{W}_{\text{BB}}^H\boldsymbol{\epsilon}. \tag{7}$$

Based on the received signal expression in (2), we can express (7) as follows:

$$\mathbf{r} = \mathbf{W}_{\text{BB}}^H\boldsymbol{\Delta}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{x} + \underbrace{\mathbf{W}_{\text{BB}}^H\boldsymbol{\Delta}^H\mathbf{W}_{\text{RF}}^H\mathbf{n} + \mathbf{W}_{\text{BB}}^H\boldsymbol{\epsilon}}_{\boldsymbol{\eta}}, \tag{8}$$

where $\boldsymbol{\eta}$ is the combined effect of the Gaussian and the quantization noise with $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\eta)$. Here $\mathbf{R}_\eta \in \mathbb{C}^{L_{\text{R}}\times L_{\text{R}}}$ is the combined noise covariance matrix with,

$$\mathbf{R}_\eta = \sigma_{\text{n}}^2\mathbf{W}_{\text{BB}}^H\boldsymbol{\Delta}^H\mathbf{W}_{\text{RF}}^H\mathbf{W}_{\text{RF}}\boldsymbol{\Delta}\mathbf{W}_{\text{BB}} + \mathbf{W}_{\text{BB}}^H\mathbf{C}_\epsilon\mathbf{W}_{\text{BB}}. \tag{9}$$

III. Bit Allocation and Hybrid Combiner Design

A. Problem Formulation

Let us consider a point-to-point MIMO system with the linear quantization model. We define the EE as the ratio of the information rate and the total consumed power as,

$$EE(\mathbf{W}_{\text{RF}}, \mathbf{\Delta}, \mathbf{W}_{\text{BB}}) \triangleq \frac{R(\mathbf{W}_{\text{RF}}, \mathbf{\Delta}, \mathbf{W}_{\text{BB}})}{P(\mathbf{\Delta})} \text{ (bits/Joule)}, \tag{10}$$

where the information rate is defined as,

$$R(\mathbf{W}_{\text{RF}}, \mathbf{\Delta}, \mathbf{W}_{\text{BB}}) \triangleq \log_2 |\mathbf{I}_{L_\text{R}} + \frac{\mathbf{R}_\eta^{-1}}{N_\text{s}} \mathbf{W}_{\text{BB}}^H \mathbf{\Delta}^H \mathbf{W}_{\text{RF}}^H \mathbf{H}\mathbf{F} \times \mathbf{F}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \mathbf{\Delta} \mathbf{W}_{\text{BB}}| \text{ (bits/s)}, \tag{11}$$

where the A/D hybrid precoder $\mathbf{F} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_\text{T} \times N_\text{s}}$.

Similar to the power model at the TX in [8], the total consumed power at the RX is expressed as:

$$P(\mathbf{\Delta}) = P_\text{D} + N_\text{R} P_\text{R} + N_\text{R} L_\text{R} P_\text{PS} + P_\text{CP} \text{ (W)}, \tag{12}$$

where P_PS is the power per phase shifter, P_R is the power per antenna, P_D is the power associated with the total quantization operation, and following (5) and [14], we have

$$P_\text{D} = P_\text{ADC} \sum_{i=1}^{L_\text{R}} 2^{b_i} = P_\text{ADC} \sum_{i=1}^{L_\text{R}} \left(\frac{\pi\sqrt{3}}{2(1 - [\mathbf{\Delta}]_{ii}^2)} \right)^{\frac{1}{2}} \text{ (W)}, \tag{13}$$

where P_ADC is the power consumed per bit in the ADC and P_CP is the power required by all circuit components.

Considering the rate and power model in (11) and (12), respectively, we can express the following fractional problem:

$$(\mathcal{P}_1): \max_{\mathbf{W}_{\text{RF}}, \mathbf{\Delta}, \mathbf{W}_{\text{BB}}} \frac{R(\mathbf{W}_{\text{RF}}, \mathbf{\Delta}, \mathbf{W}_{\text{BB}})}{P(\mathbf{\Delta})}$$
$$\text{subject to } \mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_\text{R} \times L_\text{R}}, \mathbf{\Delta} \in \mathcal{D}^{L_\text{R} \times L_\text{R}},$$

where the set $\mathcal{D}$ represents the finite states of the quantizer and is defined as,

$$\mathcal{D} = \left\{ \mathbf{\Delta} \in \mathbb{R}^{L_\text{R} \times L_\text{R}} \middle| m \leq [\mathbf{\Delta}]_{ii} \leq M \; \forall \; i = 1, ..., L_\text{R} \right\}.$$

The channel's singular value decomposition (SVD) is written as $\mathbf{H} = \mathbf{U}_\text{H} \mathbf{\Sigma}_\text{H} \mathbf{V}_\text{H}^H$, where $\mathbf{U}_\text{H} \in \mathbb{C}^{N_\text{R} \times N_\text{R}}$ and $\mathbf{V}_\text{H} \in \mathbb{C}^{N_\text{T} \times N_\text{T}}$ are unitary matrices, and $\mathbf{\Sigma}_\text{H} \in \mathbb{R}^{N_\text{R} \times N_\text{T}}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. The optimal, fully digital combiner matrix $\mathbf{W}_\text{opt}$ consists of the N_s columns of the left singular matrix $\mathbf{U}_\text{H}$. Our goal, by solving $(\mathcal{P}_1)$, is to obtain the combiner matrices and the bit resolution matrix in an optimal manner. We introduce the novel decomposition of the A/D hybrid combiner to three parts representing the analog combiner matrix, the bit resolution matrix and digital combiner matrix, i.e., $\mathbf{W}_\text{RF} \mathbf{\Delta} \mathbf{W}_\text{BB}$. So the Euclidean distance $\|\mathbf{W}_\text{opt} - \mathbf{W}_\text{RF} \mathbf{\Delta} \mathbf{W}_\text{BB}\|_F^2$ should be as small as possible for a maximum throughput combiner design. Note that we optimize over the bit resolution matrix with varying resolutions and the choice of combiner matrices at the RX.

Proposition 1. *The maximization of the fractional problem $(\mathcal{P}_1)$ is equivalent with the solution of the following problem:*

$$(\mathcal{P}_2): \min_{\mathbf{W}_{RF}, \mathbf{\Delta}, \mathbf{W}_{BB}} \frac{1}{2} \|\mathbf{W}_{opt} - \mathbf{W}_{RF} \mathbf{\Delta} \mathbf{W}_{BB}\|_F^2 + \gamma P(\mathbf{\Delta}),$$
$$\textit{subject to } \mathbf{W}_{RF} \in \mathcal{W}^{N_R \times L_R}, \mathbf{\Delta} \in \mathcal{D}^{L_R \times L_R},$$

where the parameter $\gamma \in \mathbb{R}^+$ denotes the trade-off between the rate and the power consumption.

Proof. The main idea to prove the equivalence is first to apply the Dinkelbach approach to transform the fractional problem into an affine one [16]. Afterwards, based on [1], [2], the maximization of the rate R can be expressed as minimization of the Euclidean distance between the computed A/D hybrid combiner and the optimal, fully digital combiner $\mathbf{W}_\text{opt}$. The details of this proof are omitted due to space limitations. □

Parameter γ also determines how close is the solution of $(\mathcal{P}_2)$ to $(\mathcal{P}_1)$. In this work, γ is selected after an exhaustive search over all the possible values in the range of [0.001, 0.1] and the value which gives the best result for $(\mathcal{P}_2)$ is selected. Problem $(\mathcal{P}_2)$ is non-convex due to the constraints on the structure of matrix $\mathbf{W}_\text{RF}$. Similar non-convex problems have been recently addressed in the literature via alternating direction method of multipliers (ADMM) based solutions [17]–[19].

B. Proposed ADMM Solution

In the following we develop an iterative procedure for solving $(\mathcal{P}_2)$ based on the ADMM approach [17]. This method, is a variant of the standard augmented Lagrangian method that uses partial updates (similar to the Gauss-Seidel method for the solution of linear equations) to solve constrained optimization problems. This method replaces a constrained minimization problem by a series of unconstrained problems and add a penalty term to the objective function. This penalty improves robustness compared to other optimization methods for constrained problems (for example, the dual ascent method) and in particular achieves convergence without the need of specific assumptions for the objective function, i.e., strict convexity and finiteness. The interested reader may refer to [17] for further information.

We first transform $(\mathcal{P}_2)$ into a form that can be addressed via ADMM. By using the auxiliary variable $\mathbf{Z}$, $(\mathcal{P}_2)$ can be written in the following form:

$$(\mathcal{P}_3): \min_{\mathbf{Z}, \mathbf{W}_\text{RF}, \mathbf{\Delta}, \mathbf{W}_\text{BB}} \frac{1}{2} \|\mathbf{W}_\text{opt} - \mathbf{Z}\|_F^2 + \mathbb{1}_{\mathcal{W}^{N_\text{R} \times L_\text{R}}}\{\mathbf{W}_\text{RF}\}$$
$$+ \mathbb{1}_{\mathcal{D}^{L_\text{R} \times L_\text{R}}}\{\mathbf{\Delta}\} + \gamma P(\mathbf{\Delta}),$$
$$\text{subject to } \mathbf{Z} = \mathbf{W}_\text{RF} \mathbf{\Delta} \mathbf{W}_\text{BB}.$$

Problem $(\mathcal{P}_3)$ formulates the A/D hybrid combiner matrix design as a matrix factorization problem. That is, the overall combiner $\mathbf{Z}$ is sought so that it minimizes the Euclidean distance to the optimal, fully digital combiner $\mathbf{W}_\text{opt}$ while supporting decomposition into three factors: the analog combiner matrix $\mathbf{W}_\text{RF}$, the matrix $\mathbf{\Delta}$ which is related to the resolution of each ADC and the digital combiner matrix $\mathbf{W}_\text{BB}$.

The augmented Lagrangian function of $(\mathcal{P}_3)$ is given by,

$$\mathcal{L}(\mathbf{Z},\mathbf{W}_{\text{RF}},\boldsymbol{\Delta},\mathbf{W}_{\text{BB}},\boldsymbol{\Lambda})=\frac{1}{2}\|\mathbf{W}_{\text{opt}}-\mathbf{Z}\|_F^2+\mathbb{1}_{\mathcal{W}^{N_\text{R}\times L_\text{R}}}\{\mathbf{W}_{\text{RF}}\}$$
$$+\mathbb{1}_{\mathcal{D}^{L_\text{R}\times L_\text{R}}}\{\boldsymbol{\Delta}\}+\frac{\alpha}{2}\|\mathbf{Z}+\boldsymbol{\Lambda}/\alpha-\mathbf{W}_{\text{RF}}\boldsymbol{\Delta}\mathbf{W}_{\text{BB}}\|_F^2+\gamma P(\boldsymbol{\Delta}), \tag{14}$$

where α is a scalar penalty parameter and $\boldsymbol{\Lambda}\in\mathbb{C}^{N_\text{R}\times L_\text{R}}$ is the Lagrange Multiplier matrix. According to ADMM [17], the solution to $(\mathcal{P}_3)$ is derived by the following iterative steps:

$$\begin{aligned}
(\mathcal{P}_{3\text{A}}):\mathbf{Z}_{(n)}&=\arg\min_{\mathbf{Z}}\frac{1}{2}\|(1+\alpha)\mathbf{Z}-\mathbf{W}_{\text{opt}}+\boldsymbol{\Lambda}_{(n-1)}\\
&\quad-\alpha\mathbf{W}_{\text{RF}(n-1)}\boldsymbol{\Delta}_{(n-1)}\mathbf{W}_{\text{BB}(n-1)}\|_F^2,\\
(\mathcal{P}_{3\text{B}}):\mathbf{W}_{\text{RF}(n)}&=\arg\min_{\mathbf{W}_{\text{RF}}}\mathbb{1}_{\mathcal{W}^{N_\text{R}\times L_\text{R}}}\{\mathbf{W}_{\text{RF}}\}+\frac{\alpha}{2}\times\\
&\quad\left\|\mathbf{Z}_{(n)}+\boldsymbol{\Lambda}_{(n-1)}/\alpha-\mathbf{W}_{\text{RF}}\boldsymbol{\Delta}_{(n-1)}\mathbf{W}_{\text{BB}(n-1)}\right\|_F^2,\\
(\mathcal{P}_{3\text{C}}):\boldsymbol{\Delta}_{(n)}&=\arg\min_{\boldsymbol{\Delta}}\|\mathbf{y}_\text{c}-\boldsymbol{\Psi}\text{vec}(\boldsymbol{\Delta})\|_2^2+\gamma P(\boldsymbol{\Delta}),\\
&\quad\text{subject to }\boldsymbol{\Delta}\in\mathcal{D},\\
(\mathcal{P}_{3\text{D}}):\mathbf{W}_{\text{BB}(n)}&=\arg\min_{\mathbf{W}_{\text{BB}}}\frac{\alpha}{2}\|\mathbf{Z}_{(n)}+\boldsymbol{\Lambda}_{(n-1)}/\alpha\\
&\quad-\mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{(n)}\mathbf{W}_{\text{BB}}\|_F^2,\\
\boldsymbol{\Lambda}_{(n)}&=\boldsymbol{\Lambda}_{(n-1)}+\alpha\left(\mathbf{Z}_{(n)}-\mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{(n)}\mathbf{W}_{\text{BB}(n)}\right),
\end{aligned}\tag{15}$$

where n denotes the iteration index, $\mathbf{y}_\text{c}=\text{vec}(\mathbf{Z}_{(n)}+\boldsymbol{\Lambda}_{(n-1)}/\alpha)$ and $\boldsymbol{\Psi}=\mathbf{W}_{\text{BB}(n-1)}\otimes\mathbf{W}_{\text{RF}(n)}$ ($\otimes$ is the Khatri-Rao product).

We solve the optimization problems $(\mathcal{P}_{3\text{A}})$-$(\mathcal{P}_{3\text{D}})$ and the solutions are provided in Algorithm 1. The algorithm provides the complete procedure to obtain the optimal analog combiner matrix $\mathbf{W}_{\text{RF}}$, the optimal bit resolution matrix $\boldsymbol{\Delta}$ and the optimal baseband (or digital) combiner matrix $\mathbf{W}_{\text{BB}}$. It starts by initializing the entries of the matrices $\mathbf{Z}$, $\mathbf{W}_{\text{RF}}$, $\boldsymbol{\Delta}$, $\mathbf{W}_{\text{BB}}$ with random values and the entries of the Lagrange multiplier matrix $\boldsymbol{\Lambda}$ with zeros. For iteration index n, $\mathbf{Z}_{(n)}$, $\mathbf{W}_{\text{RF}(n)}$, $\boldsymbol{\Delta}_{(n)}$ and $\mathbf{W}_{\text{BB}(n)}$ are updated at each iteration step using the solutions provided in Steps 4, 7, 8, 10 and 11 of Algorithm 1. In Step 7, $\Pi_{\mathcal{W}}$ is the operator that projects the solution onto the set $\mathcal{W}$. This is computed by solving the following optimization problem [20]:

$$(\mathcal{P}_4):\qquad\min_{\mathbf{A}_{\mathcal{W}}}\|\mathbf{A}_{\mathcal{W}}-\mathbf{A}\|_F^2,\text{subject to }\mathbf{A}_{\mathcal{W}}\in\mathcal{W},$$

where $\mathbf{A}$ is an arbitrary matrix and $\mathbf{A}_{\mathcal{W}}$ is its projection onto the set $\mathcal{W}$. The solution to $(\mathcal{P}_4)$ is given by the phase of the complex elements of $\mathbf{A}$. Thus, for $\mathbf{A}_{\mathcal{W}}=\Pi_{\mathcal{W}}\{\mathbf{A}\}$ we have

$$\mathbf{A}_{\mathcal{W}}(x,y)=\begin{cases}0, & \mathbf{A}(x,y)=0\\ \frac{\mathbf{A}(x,y)}{|\mathbf{A}(x,y)|}, & \mathbf{A}(x,y)\neq 0\end{cases}, \tag{16}$$

where $\mathbf{A}_{\mathcal{W}}(x,y)$ and $\mathbf{A}(x,y)$ are the elements at the xth row-yth column of matrices $\mathbf{A}_{\mathcal{W}}$ and $\mathbf{A}$, respectively. Furthermore, as shown in Step 8, the minimization problem in $(\mathcal{P}_{3\text{C}})$ is solved by implementing CVX [21]. A termination criterion related to the maximum permitted number of iterations of the ADMM sequence (N_{max}) is considered. Upon convergence, the number of bits for each ADC is obtained by using (5) and quantized to the nearest integer value.

Algorithm 1 Proposed ADMM Solution for the A/D Hybrid Combiner Design

```
1:  Initialize: $\mathbf{Z}$, $\mathbf{W}_{\text{RF}}$, $\boldsymbol{\Delta}$, $\mathbf{W}_{\text{BB}}$ with random values, $\boldsymbol{\Lambda}$ with zeros, $\alpha = 1$ and $n = 1$
2:  while $n \le N_{\text{max}}$ do
3:      $\mathbf{A} = \alpha\mathbf{W}_{\text{RF}(n-1)}\boldsymbol{\Delta}_{(n-1)}\mathbf{W}_{\text{BB}(n-1)}$.
4:      $\mathbf{Z}_{(n)} = \frac{1}{\alpha+1}\big(\mathbf{W}_{\text{opt}} - \boldsymbol{\Lambda}_{(n-1)} + \mathbf{A}\big)$.
5:      $\mathbf{B} = \boldsymbol{\Lambda}_{(n-1)} + \alpha\mathbf{Z}_{(n)}$.
6:      $\mathbf{C} = \alpha\boldsymbol{\Delta}_{(n-1)}\mathbf{W}_{\text{BB}(n-1)}\mathbf{W}_{\text{BB}}{}^H_{(n-1)}\boldsymbol{\Delta}^H_{(n-1)}$.
7:      $\mathbf{W}_{\text{RF}(n)} = \Pi_{\mathcal{W}}\{\mathbf{B}\mathbf{W}_{\text{BB}}{}^H_{(n-1)}\boldsymbol{\Delta}^H_{(n-1)}\mathbf{C}^{-1}\}$.
8:      Update $\boldsymbol{\Delta}_{(n)}$ by solving $(\mathcal{P}_{3\text{C}})$ using CVX [21].
9:      $\mathbf{D} = \alpha\boldsymbol{\Delta}^H_{(n)}\mathbf{W}_{\text{RF}}{}^H_{(n)}\mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{(n)}$.
10:     $\mathbf{W}_{\text{BB}(n)} = \mathbf{D}^{-1}\boldsymbol{\Delta}^H_{(n)}\mathbf{W}_{\text{RF}}{}^H_{(n)}\mathbf{B}$.
11:     $\boldsymbol{\Lambda}_{(n)} = \boldsymbol{\Lambda}_{(n-1)} + \alpha\left(\mathbf{Z}_{(n)} - \mathbf{W}_{\text{RF}(n)}\boldsymbol{\Delta}_{(n)}\mathbf{W}_{\text{BB}(n)}\right)$.
12:     $n \leftarrow n + 1$
13: end while
14: return $\mathbf{W}_{\text{RF}(N_{\text{max}})}$, $\boldsymbol{\Delta}_{(N_{\text{max}})}$, $\mathbf{W}_{\text{BB}(N_{\text{max}})}$
```

Computational complexity analysis of Algorithm 1: In Algorithm 1, mainly Step 8 involves multiplication by $\boldsymbol{\Psi}$ whose dimensions are $L_\text{R}N_\text{R}\times N_\text{s}L_\text{R}$. In general, the solution of $(\mathcal{P}_{3\text{C}})$ can be upper-bounded by $\mathcal{O}((L_\text{R}^2N_\text{R}N_\text{s})^3)$ which can be improved significantly by exploiting the structure of $\boldsymbol{\Psi}$.

IV. Simulation Results

In this section, we evaluate the performance of the proposed ADMM technique using computer simulation results. The results have been averaged over 1,000 Monte-Carlo realizations.

System setup: We set the following parameters, unless specified otherwise, to obtain the desired results: N_T = 32, N_R = 16, L_R = 4, N_s = 4, N_cl = 2, N_ray = 4, N_max = 40, m = 1, M = 8, α = 1 and $\sigma^2_{\alpha,i}$ = 1. The azimuth angles of departure and arrival are computed with uniformly distributed mean angles; each cluster follows a Laplacian distribution about the mean angle. The antenna elements in the ULA are spaced by distance $d=\lambda/2$. The signal-to-noise ratio (SNR) is given by the inverse of the noise variance, i.e., $1/\sigma_\text{n}^2$. The transmit vector $\mathbf{x}$ is composed of the normalized i.i.d. Gaussian symbols. The values used for the terms in the power model in (12) of Section III are P_{ADC} = 100 mW, P_{CP} = 10 W, P_R = 100 mW and P_{PS} = 10 mW. Note that to measure the spectral efficiency (SE) performance, we compute the ratio R/B bits/s/Hz where B represents the bandwidth, and for the simulations we set B = 1 Hz. For simulations, the precoder matrix $\mathbf{F}$ is considered equal to the optimal fully digital precoder matrix [1], [2], i.e., the product of $1/\sqrt{N_\text{s}}$ and first N_s columns of the right singular matrix $\mathbf{V}_\text{H}$.

Convergence of the proposed ADMM solution: Fig. 2 shows the convergence of the ADMM solution as proposed in Algorithm 1 to obtain the optimal bit resolution at each ADC and corresponding optimal combiner matrices. The proposed solution converges rapidly at around 20 iterations and mean square error (MSE), $\left\|\mathbf{W}_{\text{opt}}-\mathbf{W}_{\text{RF}(N_{\text{max}})}\boldsymbol{\Delta}_{(N_{\text{max}})}\mathbf{W}_{\text{BB}(N_{\text{max}})}\right\|_F^2$, goes as low as -20 dB. A lower number of RX antennas shows

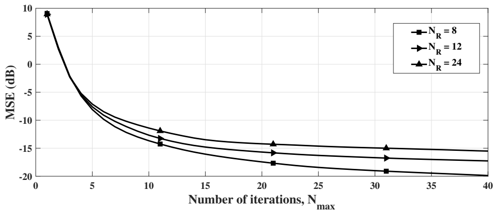


Fig. 2. Convergence of the ADMM solution for different N_{R} at $\gamma = 0.01$.

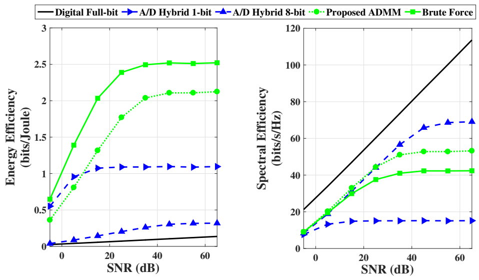


Fig. 3. EE and SE performance w.r.t. SNR at $N_{\mathrm{R}} = 16$ and $\gamma = 0.01$.

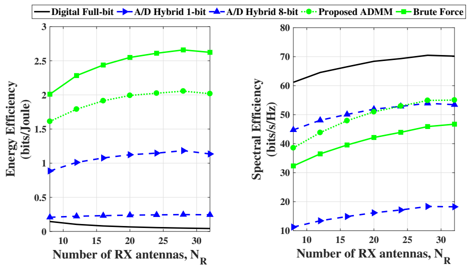


Fig. 4. EE and SE performance w.r.t. N_{R} at SNR = 30 dB and $\gamma = 0.01$.

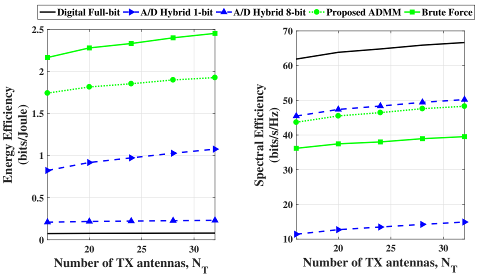


Fig. 5. EE and SE performance w.r.t. N_{T} at SNR = 30 dB and $\gamma = 0.01$.

lower MSE as expected, since fewer parameters are required to be estimated.

Benchmark techniques:

1) Digital combining with full-bit resolution: We consider the conventional fully digital beamforming architecture, where the number of RF chains at the RX is equal to the number of RX antennas, i.e., $L_{\mathrm{R}} = N_{\mathrm{R}}$. The fully digital combining solution may be provided by SVD and waterfilling [22]. In terms of the resolution sampling, we consider full-bit resolution, i.e., $M = 8$-bit, which represents the optimum from the achievable SE perspective.

2) A/D Hybrid combining with 1-bit and 8-bit resolutions: We also consider a A/D hybrid combining architecture with $L_{\mathrm{R}} < N_{\mathrm{R}}$, for two cases of bit resolution: a) 1-bit resolution which usually shows reasonable EE performance, and b) 8-bit resolution which usually shows high SE results.

3) Brute force with A/D hybrid combining: We also implement an exhaustive search approach as an upper bound for EE maximization called brute force (BF), based on [4], which clearly shows the energy-rate performance trade-offs in the simulations. It makes a search over the number of RF chains L_{R} and all the available bit resolutions, i.e., $b = 1, ..., M$. It then finds the best EE out of all the possible cases and chooses the corresponding optimal resolution for each ADC. This method provides the best possible EE performance, but it is computationally intractable for $L_{\mathrm{R}} > 4$.

Fig. 3 shows the performance of the proposed ADMM solution compared with existing benchmark techniques with respect to (w.r.t.) SNR at $N_{\mathrm{R}} = 16$. The proposed ADMM solution achieves high EE which has performance close to the BF approach and better than the 8-bit hybrid, 1-bit hybrid and full-bit digital baselines. For example, at SNR = 20 dB, the proposed ADMM solution outperforms 1-bit hybrid, 8-bit hybrid and full-bit digital baselines by about 0.45 bits/Joule, 1.375 bits/Joule and 1.44 bits/Joule, respectively. It also exhibits better SE than 1-bit hybrid and has similar performance to the 8-bit hybrid baseline.

There is an energy-rate trade-off between the proposed solution and the BF approach as we can achieve better rate with lower EE and vice-versa. Moreover, the proposed solution has lower complexity than the BF approach because the BF involves a search over all the possible bit resolutions while the proposed solution directly optimizes the number of bits to obtain an optimal number of bits at each ADC. We constrain the number of RF chains $L_{\mathrm{R}} = 4$ for the BF approach due to the high complexity order which is $\mathcal{O}(M^{L_{\mathrm{R}}})$. Also note that the proposed approach enables the selection of different resolutions for different ADCs and thus, it offers a better trade-off for EE versus SE than existing approaches which are based on a fixed ADC resolution.

Figs. 4 and 5 show the performance results w.r.t. the number of RX and TX antennas at 30 dB SNR. The proposed ADMM solution again achieves high EE and performs close to the BF approach and better than the 8-bit hybrid, 1-bit hybrid and full-bit digital baselines. For example, at $N_{\mathrm{R}} = 20$, the proposed ADMM solution outperforms 1-bit hybrid, 8-bit hybrid and full-bit digital baselines by about 0.85 bits/Joule, 1.75 bits/Joule and 1.875 bits/Joule, respectively. Also, for $N_{\mathrm{T}} = 20$, the proposed solution outperforms 1-bit hybrid, 8-bit hybrid and full-bit digital baselines by about 1.0 bits/Joule, 1.5 bits/Joule and 1.625 bits/Joule, respectively. The proposed solution also exhibits better SE than 1-bit hybrid and has similar performance to the 8-bit hybrid baseline. Both the

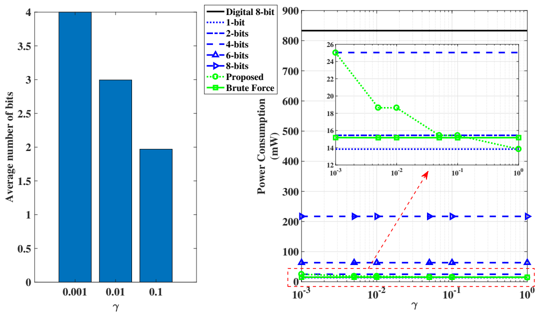


Fig. 6. Average number of bits for proposed ADMM and power consumption w.r.t. γ at SNR = 30 dB.

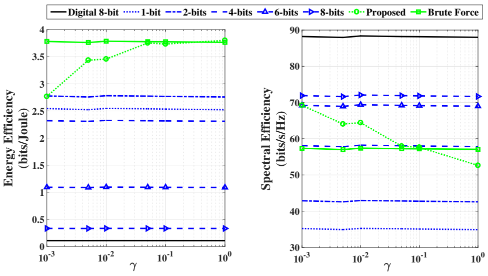


Fig. 7. EE and SE performance w.r.t. γ at SNR = 30 dB.

figures follow the energy-rate trade-off with the BF approach.

Furthermore, we investigate the performance over the trade-off parameter γ introduced in ($\mathcal{P}_2$). Fig. 6 shows the bar plot of average of the optimal number of bits selected by the proposed solution for each ADC versus γ. The average optimal number decreases with the increase in γ, for example, it is 4 for $\gamma = 0.001$, 3 for $\gamma = 0.01$ and 2 for $\gamma = 0.1$. Fig. 6 also shows that the power consumption in the proposed case is considerably low and decreases with the increase in the trade-off parameter γ unlike digital 8-bit, several fixed bit hybrid baselines and the BF approach. Fig. 7 shows the EE and SE plots for several solutions w.r.t. γ. It can be observed that the proposed solution achieves higher EE than the fixed bit allocation solutions and achieves comparable EE and SE results to the BF approach. These curves also show that adjusting γ allows the system to vary the energy-rate trade-off.

V. Conclusion

This paper proposes an energy efficient mmWave A/D hybrid MIMO system which can vary the ADC bit resolution at the RX. This method uses the decomposition of the A/D hybrid combiner matrix into three parts representing the analog combiner matrix, the bit resolution matrix and the digital combiner matrix. These three matrices are optimized by the novel ADMM solution which outperforms the EE of the full-bit digital, 1-bit hybrid combining and 8-bit hybrid combining baselines. There is an energy-rate trade-off with the BF approach which yields the upper bound for EE maximization. The proposed approach enables the selection of the optimal resolution for each ADC and thus, it offers better trade-off for data rate versus EE than existing approaches based on fixed ADC resolution. In future work, we will jointly optimize the DAC and ADC bit resolution and hybrid precoder and combiner matrices at the TX and the RX.

References

[1] O. E. Ayach et al., “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[2] A. Kaushik et al.,“Sparse hybrid precoding and combining in millimeter wave MIMO systems,” *IET Radio Prop. Tech. 5G*, Durham, UK, pp. 1-7, Oct. 2016.

[3] S. Han et al., “Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,” *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.

[4] R. Zi et al., “Energy efficiency optimization of 5G radio frequency chain systems”, *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.

[5] C. G. Tsinos et al., “On the Energy-Efficiency of Hybrid Analog-Digital Transceivers for Single- and Multi-Carrier Large Antenna Array Systems”, in *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980-1995, Sept. 2017.

[6] R. W. Heath et al., “An overview of signal processing techniques for millimeter wave MIMO systems”, *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[7] A. Kaushik et al., “Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid Beamforming in Millimeter Wave MIMO Systems,” *IEEE Trans. Green Commun. Netw.*, accepted, July 2019.

[8] A. Kaushik et al., “Energy Efficiency maximization of millimeter wave hybrid MIMO systems with low resolution DACs,” *IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, pp. 1-6, May 2019.

[9] J. Mo et al., “Achievable rates of hybrid architectures with few-bit ADC receivers,” *VDE Int. ITG Workshop Smart Antennas*, pp. 1-8, 2016.

[10] J. Zhang et al., “Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327-1338, Jun. 2017.

[11] A. Kaushik et al.,“Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs”, *IEEE Europ. Sig. Process.*, Rome, Italy, pp. 1839-1843, Sept. 2018.

[12] T.-C. Zhang et al., “Mixed-ADC massive MIMO detectors: Performance analysis and design optimization,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738-7752, Nov. 2016.

[13] S. Singh et al., “Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control”, *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513-1527, Oct. 2011.

[14] O. Orhan et al., “Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance,” *Info. Theory Appl. Workshop (ITA)*, San Diego, USA, pp. 191-198, Feb. 2015.

[15] A. Mezghani and J. A. Nossek, “Capacity lower bound of MIMO channels with output quantization and correlated noise,” *IEEE Int. Symp. Info. Theory (ISIT)*, Cambridge, USA, Jul. 2012.

[16] W. Dinkelbach, “On nonlinear fractional programming”, *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2011.

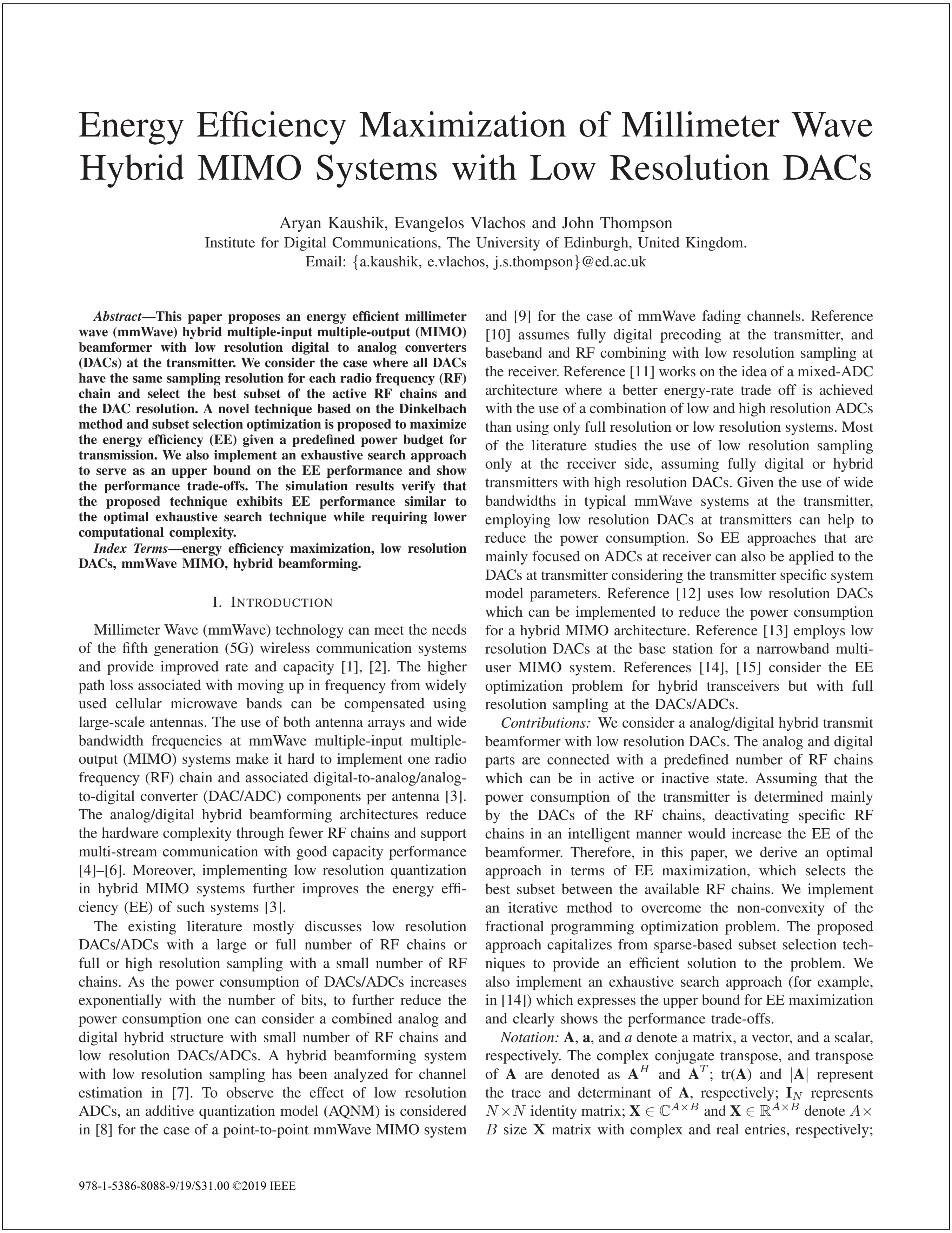
[18] C. G. Tsinos et al., “Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization,” *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 160174, June 2017.

[19] C. G. Tsinos and B. Ottersten, “An efficient algorithm for unit-modulus quadratic programs with application in beamforming for wireless sensor networks,” *IEEE Signal Process. Letters*, vol. 25, no. 2, pp. 169-173, Feb. 2018.

[20] D. P. Bertsekas, “Nonlinear programming,” 1999.

[21] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs”, in *Recent Adv. Learning and Control*, Springer-Verlag Ltd., pp. 95-110, 2008.

[22] T. S. Rappaport et al., “Millimeter wave wireless communications,” *Prentice-Hall*, Sept. 2014.



Energy Efficiency Maximization of Millimeter Wave Hybrid MIMO Systems with Low Resolution DACs

Aryan Kaushik, Evangelos Vlachos and John Thompson
Institute for Digital Communications, The University of Edinburgh, United Kingdom.
Email: {a.kaushik, e.vlachos, j.s.thompson}@ed.ac.uk

***Abstract*—This paper proposes an energy efficient millimeter wave (mmWave) hybrid multiple-input multiple-output (MIMO) beamformer with low resolution digital to analog converters (DACs) at the transmitter. We consider the case where all DACs have the same sampling resolution for each radio frequency (RF) chain and select the best subset of the active RF chains and the DAC resolution. A novel technique based on the Dinkelbach method and subset selection optimization is proposed to maximize the energy efficiency (EE) given a predefined power budget for transmission. We also implement an exhaustive search approach to serve as an upper bound on the EE performance and show the performance trade-offs. The simulation results verify that the proposed technique exhibits EE performance similar to the optimal exhaustive search technique while requiring lower computational complexity.**

***Index Terms*—energy efficiency maximization, low resolution DACs, mmWave MIMO, hybrid beamforming.**

I. Introduction

Millimeter Wave (mmWave) technology can meet the needs of the fifth generation (5G) wireless communication systems and provide improved rate and capacity [1], [2]. The higher path loss associated with moving up in frequency from widely used cellular microwave bands can be compensated using large-scale antennas. The use of both antenna arrays and wide bandwidth frequencies at mmWave multiple-input multiple-output (MIMO) systems make it hard to implement one radio frequency (RF) chain and associated digital-to-analog/analog-to-digital converter (DAC/ADC) components per antenna [3]. The analog/digital hybrid beamforming architectures reduce the hardware complexity through fewer RF chains and support multi-stream communication with good capacity performance [4]–[6]. Moreover, implementing low resolution quantization in hybrid MIMO systems further improves the energy efficiency (EE) of such systems [3].

The existing literature mostly discusses low resolution DACs/ADCs with a large or full number of RF chains or full or high resolution sampling with a small number of RF chains. As the power consumption of DACs/ADCs increases exponentially with the number of bits, to further reduce the power consumption one can consider a combined analog and digital hybrid structure with small number of RF chains and low resolution DACs/ADCs. A hybrid beamforming system with low resolution sampling has been analyzed for channel estimation in [7]. To observe the effect of low resolution ADCs, an additive quantization model (AQNM) is considered in [8] for the case of a point-to-point mmWave MIMO system and [9] for the case of mmWave fading channels. Reference [10] assumes fully digital precoding at the transmitter, and baseband and RF combining with low resolution sampling at the receiver. Reference [11] works on the idea of a mixed-ADC architecture where a better energy-rate trade off is achieved with the use of a combination of low and high resolution ADCs than using only full resolution or low resolution systems. Most of the literature studies the use of low resolution sampling only at the receiver side, assuming fully digital or hybrid transmitters with high resolution DACs. Given the use of wide bandwidths in typical mmWave systems at the transmitter, employing low resolution DACs at transmitters can help to reduce the power consumption. So EE approaches that are mainly focused on ADCs at receiver can also be applied to the DACs at transmitter considering the transmitter specific system model parameters. Reference [12] uses low resolution DACs which can be implemented to reduce the power consumption for a hybrid MIMO architecture. Reference [13] employs low resolution DACs at the base station for a narrowband multi-user MIMO system. References [14], [15] consider the EE optimization problem for hybrid transceivers but with full resolution sampling at the DACs/ADCs.

Contributions: We consider a analog/digital hybrid transmit beamformer with low resolution DACs. The analog and digital parts are connected with a predefined number of RF chains which can be in active or inactive state. Assuming that the power consumption of the transmitter is determined mainly by the DACs of the RF chains, deactivating specific RF chains in an intelligent manner would increase the EE of the beamformer. Therefore, in this paper, we derive an optimal approach in terms of EE maximization, which selects the best subset between the available RF chains. We implement an iterative method to overcome the non-convexity of the fractional programming optimization problem. The proposed approach capitalizes from sparse-based subset selection techniques to provide an efficient solution to the problem. We also implement an exhaustive search approach (for example, in [14]) which expresses the upper bound for EE maximization and clearly shows the performance trade-offs.

Notation: $\mathbf{A}$, $\mathbf{a}$, and a denote a matrix, a vector, and a scalar, respectively. The complex conjugate transpose, and transpose of $\mathbf{A}$ are denoted as $\mathbf{A}^H$ and $\mathbf{A}^T$; tr($\mathbf{A}$) and $|\mathbf{A}|$ represent the trace and determinant of $\mathbf{A}$, respectively; $\mathbf{I}_N$ represents $N \times N$ identity matrix; $\mathbf{X} \in \mathbb{C}^{A\times B}$ and $\mathbf{X} \in \mathbb{R}^{A\times B}$ denote $A \times B$ size $\mathbf{X}$ matrix with complex and real entries, respectively;

978-1-5386-8088-9/19/$31.00 ©2019 IEEE

$\mathcal{CN}(\mathbf{a},\mathbf{A})$ denotes a complex Gaussian vector having mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$; $[\mathbf{A}]_k$ denotes the k-th column of matrix $\mathbf{A}$ and $[\mathbf{A}]_{kl}$ is the matrix entry at the k-th row and l-th column.

II. Hybrid mmWave MIMO

A. MmWave channel and system model

MmWave channels can be modeled by a narrowband clustered channel model due to different channel settings such as number of multipaths, amplitudes, etc., with N_{cl} clusters and N_{ray} propagation paths in each cluster [3], [4]. Considering a single-user mmWave system with N_{T} antennas at the transmitter, transmitting N_{s} data streams to N_{R} antennas at receiver, the mmWave channel matrix can be written as follows:

$$\mathbf{H}=\sum_{i=1}^{N_{\text{cl}}}\sum_{l=1}^{N_{\text{ray}}}\alpha_{il}\mathbf{a}_{\text{R}}(\phi_{il}^{r})\mathbf{a}_{\text{T}}(\phi_{il}^{t})^{H}, \tag{1}$$

where $\alpha_{il}\in\mathcal{CN}(0,\sigma_{\alpha,i}^2)$ is the gain term with $\sigma_{\alpha,i}^2$ being the average power of the i^{th} cluster. Furthermore, $\mathbf{a}_{\text{T}}(\phi_{il}^{t})$ and $\mathbf{a}_{\text{R}}(\phi_{il}^{r})$ represent the normalized transmit and receive array response vectors [3], where ϕ_{il}^{t} and ϕ_{il}^{r} denote the azimuth angles of departure and arrival, respectively. We use uniform linear array (ULA) antennas for simplicity and model the antenna elements at the transmitter as ideal sectored elements [16]. However, the proposed technique is not limited to this setup and can be easily extended to the case of wideband channels and uniform planar arrays.

B. Quantization Model

We consider the linear model approximation (AQNM) to represent the introduced distortion of the quantization noise [18]. Given that $Q(\cdot)$ denotes a uniform scalar quantizer then for the scalar input s we have that,

$$Q(s)\approx\delta x+\epsilon, \tag{2}$$

where

$$\delta=\sqrt{1-\frac{\pi\sqrt{3}}{2}2^{-2b}} \tag{3}$$

is the multiplicative distortion parameter for bit sampling resolution equal to b and ϵ is the additive quantization noise with $\epsilon\sim\mathcal{CN}(0,\sigma_{\epsilon}^2)$, where

$$\sigma_{\epsilon}=\sqrt{1-\frac{\pi\sqrt{3}}{2}2^{-2b}}\sqrt{\frac{\pi\sqrt{3}}{2}2^{-2b}}=\delta(1-\delta^2). \tag{4}$$

C. System Model

In the analog and digital hybrid beamforming architecture, the number of transmitter RF chains L_{T} is usually smaller than the number of the transmitting antennas N_{T}, $L_{\text{T}}\leq N_{\text{T}}$, and similarly for the receiver, the number of RF chains $L_{\text{R}}\leq N_{\text{R}}$ (the number of receiving antennas). After the RF or analog precoding, each phase shifter is connected to all the antenna elements. Fig. 1 shows the system setup.

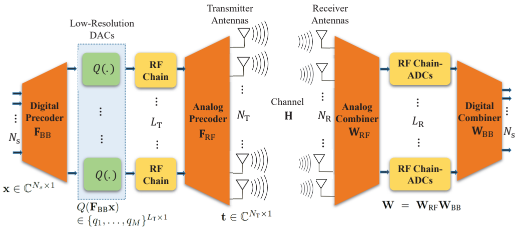

Low-Resolution DACs
Transmitter Antennas
Receiver Antennas
Digital Precoder $\mathbf{F}_{\text{BB}}$
N_{s}
$\mathbf{x}\in\mathbb{C}^{N_s\times1}$
$Q(.)$
$Q(.)$
RF Chain
L_{T}
RF Chain
Analog Precoder $\mathbf{F}_{\text{RF}}$
N_{T}
Channel $\mathbf{H}$
N_{R}
Analog Combiner $\mathbf{W}_{\text{RF}}$
RF Chain-ADCs
L_{R}
RF Chain-ADCs
Digital Combiner $\mathbf{W}_{\text{BB}}$
N_{s}
$Q(\mathbf{F}_{\text{BB}}\mathbf{x})$
$\in\{q_1,\ldots,q_M\}^{L_{\text{T}}\times1}$
$\mathbf{t}\in\mathbb{C}^{N_{\text{T}}\times1}$
$\mathbf{W}=\mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}$

Fig. 1. A mmWave hybrid MIMO system with low resolution DACs.

Let $\mathbf{x}\in\mathbb{C}^{N_{\text{s}}\times1}$ is the normalized data vector, then based on the AQNM the vector containing the complex output of all the DACs can be expressed as:

$$Q(\mathbf{F}_{\text{BB}}\mathbf{x})\approx\delta\mathbf{F}_{\text{BB}}\mathbf{x}+\boldsymbol{\epsilon}, \tag{5}$$

where $Q(\mathbf{F}_{\text{BB}}\mathbf{x})\in\mathbb{C}^{L_{\text{T}}\times1}$ and $\mathbf{F}_{\text{BB}}\in\mathbb{C}^{L_{\text{T}}\times N_s}$ is the baseband part of transmit beamformer. The second term of (5) expresses the additive quantization noise for all RF chains with $\boldsymbol{\epsilon}\in\mathcal{CN}(\mathbf{0},\sigma_{\epsilon}^2\mathbf{I}_{L_{\text{T}}})$. This leads us to the following expression for the transmitted signal, as seen at the output of the analog and digital hybrid transmitter:

$$\mathbf{t}=\mathbf{F}_{\text{RF}}\left(\delta\mathbf{F}_{\text{BB}}\mathbf{x}+\boldsymbol{\epsilon}\right)=\delta\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{x}+\mathbf{F}_{\text{RF}}\boldsymbol{\epsilon}, \tag{6}$$

where $\mathbf{F}_{\text{RF}}$ is the analog precoding matrix at the transmitter.

After the effect of the mmWave channel and the RF processing at the receiver, the received signal is expressed as:

$$\begin{aligned}\mathbf{y}&=\mathbf{W}^H\mathbf{H}\mathbf{t}+\mathbf{W}^H\mathbf{n} &(7)\\ &=\underbrace{\delta\mathbf{W}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}}_{\mathbf{H}_{\text{eff}}(L_{\text{T}},\delta)}\mathbf{x}+\underbrace{\mathbf{W}^H\mathbf{H}\mathbf{F}_{\text{RF}}\boldsymbol{\epsilon}+\mathbf{W}^H\mathbf{n}}_{\boldsymbol{\eta}}, &(8)\end{aligned}$$

where $\mathbf{H}_{\text{eff}}(L_{\text{T}},\delta)$ is the effective channel which is a function of the number of the RF chains L_{T} and the distortion δ, $\mathbf{W}\in\mathbb{C}^{N_{\text{R}}\times N_{\text{s}}}$ is the receiver combining matrix, $\boldsymbol{\eta}$ is the combined effect of the Gaussian and quantization noise with $\boldsymbol{\eta}\sim\mathcal{CN}(\mathbf{0},\mathbf{R}_{\eta})$, while $\mathbf{R}_{\eta}$ is the combined noise covariance matrix with,

$$\mathbf{R}_{\eta}(L_{\text{T}},\delta)=\sigma_{\epsilon}^2\mathbf{W}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{RF}}^H\mathbf{H}^H\mathbf{W}+\sigma_{\text{n}}^2\mathbf{W}^H\mathbf{W}, \tag{9}$$

which is also a function of the number of the RF chains L_{T} and the distortion δ. Note that unlike what is common in the existing literature, in this work we also take into account the cross-terms of the noise covariance matrix $\mathbf{R}_{\eta}$. We believe this is a more realistic scenario since it can also incorporate system impairments such as phase noise into the problem formulation.

III. Energy Efficiency Maximization

The EE of a point-to-point MIMO system is defined as the ratio of the information rate and the total consumed power [22]. Since these quantities depend on the distortion of the DACs δ and the number of the RF chains L_{T}, EE is expressed as

$$\text{EE}(L_{\text{T}},\delta)\triangleq\frac{R(L_{\text{T}},\delta)}{P(L_{\text{T}},\delta)}\text{ (bits/Joule).} \tag{10}$$

Exploiting the linearity property of the quantization model in (5), the information rate $R(L_{\mathrm{T}}, \delta)$ is expressed as:

$$R(L_{\mathrm{T}}, \delta) = \log_2 |\mathbf{I}_{N_{\mathrm{s}}} + \frac{1}{N_{\mathrm{s}}} \mathbf{R}_{\eta}^{-1} \mathbf{H}_{\mathrm{eff}} \mathbf{H}_{\mathrm{eff}}| \text{ (bits/s/Hz)}, \quad (11)$$

where the values of L_{T} and δ will affect the noise covariance matrix $\mathbf{R}_{\eta}(L_{\mathrm{T}}, \delta)$ and the effective channel $\mathbf{H}_{\mathrm{eff}}(L_{\mathrm{T}}, \delta)$.

Concerning the power consumption model, we consider that the total power consumption $P(L_{\mathrm{T}}, \delta)$ is proportional to:

$$P(L_{\mathrm{T}}, \delta) \propto L_{\mathrm{T}} \left[P_{\mathrm{DAC}} \left(\frac{\pi\sqrt{3}}{2(1-\delta^2)} \right)^{1/2} + N_{\mathrm{T}} P_{\mathrm{PS}} \right] \quad (W) \quad (12)$$

where P_{DAC} and P_{S} depend upon the DAC and phase-shifter power consumption values, respectively.

Given the expressions (11) and (12), we can now define the EE maximization problem as a fractional programming problem:

$$\arg \max_{L_{\mathrm{T}}, \delta} \ \mathrm{EE}(L_{\mathrm{T}}, \delta) \text{ subject to } P(L_{\mathrm{T}}, \delta) \le P_{\max}, \quad (13)$$

where $P_{\max}$ is the maximum available power budget. Our goal, by solving (13), is to obtain the number of RF chains and bit resolution in an optimal manner. To obtain a solution to (13) we have developed an iterative procedure that approximates the initial fractional problem with a convex-concave optimization, using Dinkelbach approximation [20] and subset selection. Dinkelbach approach makes an iterative approximation of the fractional problem with a sequence of non-fractional but constrained optimization ones. Although simpler, each one of these problems is still non-convex. However, by decomposing the contribution of each RF chain to the EE performance of the system, we can employ subset selection methods which minimize the number of the RF chains by solving an ℓ_1 approximation to the non-convex problem.

Before proceeding with the description of the proposed technique, we derive a technique based on exhaustive search for EE maximization, which will serve as an upper bound for comparison with the proposed method.

A. Upper Bound on EE via Exhaustive Search

To obtain an upper bound, we consider the case where $L_{\mathrm{T}} = N_{\mathrm{T}}$. This simplifies the computation of the beamformers at the receiver and the receiver, by using the singular value decomposition of the channel (SVD). However, since we change the number of the RF chains/antennas, the channel and its SVD, has to be updated at each time. Specifically, an exhaustive search approach is needed to obtain the optimum EE over all possible values of $(L_{\mathrm{T}}, \delta) \in \{1, \ldots, b_{\max}\} \times \{1, \ldots, L_{\mathrm{T}}\}$. For each set value (L_{T}, δ), the singular value decomposition (SVD) of the effective channel has to be obtained, i.e.,

$$\mathbf{H}_{\mathrm{eff}}(L_{\mathrm{T}}, \delta) = \delta \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H, \quad (14)$$

where $\mathbf{U} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ and $\mathbf{V} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ are unitary matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. We assume that the rank of the channel is r.

```
Algorithm 1: Brute-force approach
Input: $b_{\max}$, $\mathbf{H}$
Begin:
for $b = 1, ..., b_{\max}$
    Compute $\delta(b)$ based on (3)
    for $l_{\mathrm{t}} = 1, ..., N_{\mathrm{T}}$
        Compute the SVD of $\mathbf{H}_{\mathrm{eff}}(l_t, \delta(b_i))$ based on (14)
        Compute $\mathrm{EE}(l_{\mathrm{t}}, \delta(b))$ based on (11) and (12)
    end
end
    Find the $L_{\mathrm{T}}^{opt}$ and $b^{opt}$ such as
        $\mathrm{EE}(L_{\mathrm{T}}^{\mathrm{opt}}, \delta(b^{\mathrm{opt}})) > \mathrm{EE}(l_t, \delta(b)) \quad \forall (b, l_t)$
Output: $L_{\mathrm{T}}^{\mathrm{opt}}$ and $b^{\mathrm{opt}}$
```

Hence, the rate expression in (11) becomes:

$$\begin{aligned} R(L_{\mathrm{T}}, \delta) &= \log_2 |\mathbf{I}_{N_{\mathrm{s}}} + \frac{\delta^2}{N_{\mathrm{s}}} \mathbf{R}_{\eta}^{-1} \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H \mathbf{W}| \\ &= \log_2 |\mathbf{I}_{N_{\mathrm{s}}} + \frac{\delta^2}{N_{\mathrm{s}}} \mathbf{R}_{\eta}^{-1} \mathbf{\Sigma} \mathbf{\Sigma}^H| \\ &= \sum_{i=1}^{r} \log_2 (1 + \frac{\delta^2}{N_{\mathrm{s}}} [\mathbf{R}_{\eta}^{-1}]_{ii} [\mathbf{\Sigma} \mathbf{\Sigma}^H]_{ii}), \quad (15) \end{aligned}$$

where $\mathbf{R}_{\eta}$ becomes a diagonal matrix with entries $[\mathbf{R}_{\eta}]_{ii} = \sigma_{\epsilon}^2 [\mathbf{\Sigma}\mathbf{\Sigma}^H]_{ii} + \sigma_n^2$. Based on (15), the rate expression is decomposed into the singular values domain, thus, the number of the rank r represents the *virtual* number of RF chains. So, the goal here is to reduce the number of virtual RF chains r, alongside with the distortion δ which depends on the bit resolution b.

Algorithm 1 shows the exhaustive search approach (similar to [14]), called the Brute-force technique, thus, it provides the solution to achieve the optimal number of RF chains and the optimal number of associated DAC bits at each channel realization. It makes a search of all the possible number of RF chains/antennas, i.e., $l_{\mathrm{t}} = \{1, ..., N_{\mathrm{T}}\}$ and over the available bit resolution, i.e., $b = 1, ..., b_{\max}$, where $b_{\max}$ is the highest achievable resolution. It then finds the best EE out of all the efficiencies and chooses the corresponding optimal number of active RF chains L_{T}^{opt} and optimal resolution sampling b^{opt} for the transmitter. This method provides the best possible energy efficiency performance assuming that the SVD of $\mathbf{H}$ is perfectly known at the transmitter.

B. Proposed Method

Let us now consider an optimal design where we seek the sampling resolution for each DACs and the optimal number of active RF chains L_{T} that will maximize the EE of the transmitter. We consider a variable number of RF chains, i.e., by using switches to activate/deactivate each one independently [19], then the problem becomes:

$$\arg \max_{\mathbf{S}, \delta} \frac{R(\mathbf{S}, \delta)}{P(\mathbf{S}, \delta)} \text{ subject to } P(\mathbf{S}, \delta) \le P_{\max}, \quad (16)$$

where $\mathbf{S} \in \{0,1\}^{L_{\rm T}\times L_{\rm T}}$ is a diagonal binary matrix representing switches which activate or deactivate the RF chains. Hence, the resulting optimization problem of (16) has two unknown quantities to be recovered, the matrices $\mathbf{S}$ and δ. We transform the problem into a subset selection based problem considering sparse optimization and compressive sampling.

We consider the problem to be equivalent to finding only a sparse selection vector, diag($\mathbf{S}$) $\in \{0,1\}^{L_{\rm T}\times 1}$, where each unity value represents one active RF chain with a predefined resolution, while the zero value represents an inactive RF chain. It is important to note that based on the proposed architecture, the optimization problem does not consider a predefined number of active/inactive RF chains, but this quantity is an optimization variable. Incorporating this selection procedure into our formulation, the received signal $\hat{\mathbf{y}} \in \mathbb{C}^{N_{\rm s}\times 1}$ at the baseband receiver is expressed as:

$$\hat{\mathbf{y}} = \delta \mathbf{W}^H \mathbf{H}\mathbf{F}_{\rm RF}\mathbf{S}\mathbf{F}_{\rm BB}\mathbf{x} + \boldsymbol{\eta}, \qquad (17)$$

where $\mathbf{S} \in \{0,1\}^{L_{\rm T}\times L_{\rm T}}$ is a diagonal selection matrix composed by zeros and ones, with $[\mathbf{S}]_{kk} \in \{0,1\}$ and $[\mathbf{S}]_{kl} = 0$ for $k \neq l$; $\delta \mathbf{W}^H \mathbf{H}\mathbf{F}_{\rm RF}\mathbf{S}\mathbf{F}_{\rm BB}$ is the effective channel $\hat{\mathbf{H}}_{\rm eff} \in \mathbb{C}^{N_{\rm s}\times N_{\rm s}}$ in this case, including hybrid transmitter precoding and receiver combining and quantization distortion. The parameter that we aim to optimize in (17) is now the entries of the diagonal selection matrix $\mathbf{S} \in \{0,1\}^{L_{\rm T}\times L_{\rm T}}$. The effective channel can be decomposed as:

$$\begin{aligned}\hat{\mathbf{H}}_{\rm eff} &= \delta \mathbf{W}^H \mathbf{H}\mathbf{F}_{\rm RF}\mathbf{S}\mathbf{F}_{\rm BB} && (18)\\ &= \sum_{i=1}^{L_{\rm T}} [\mathbf{S}]_{ii} [\delta \mathbf{W}^H \mathbf{H}\mathbf{F}_{\rm RF}]_i [\mathbf{F}_{\rm BB}^T]_i^T \\ &= \sum_{i=1}^{L_{\rm T}} [\mathbf{S}]_{ii} \mathbf{a}_i \mathbf{b}_i^T, && (19)\end{aligned}$$

where $\mathbf{b}_i \triangleq [\mathbf{F}_{\rm BB}^T]_i \in \mathbb{C}^{N_s\times 1}$, $\mathbf{a}_i \triangleq [\delta \mathbf{R}_\eta^{-\frac{1}{2}} \mathbf{W}^H \mathbf{H}\mathbf{F}_{\rm RF}]_i \in \mathbb{C}^{N_s\times 1}$ and where $[\mathbf{S}]_{ii} \in \{0,1\}$ determines the state of the i-th RF chain. Based on (19), the received signal can be equivalently expressed as the following measurement vector:

$$\hat{\mathbf{y}} = \sum_{i=1}^{L_{\rm T}} [\mathbf{S}]_{ii} \mathbf{a}_i (\mathbf{b}_i^T \mathbf{x}) + \hat{\boldsymbol{\eta}}, \qquad (20)$$

where $\hat{\boldsymbol{\eta}} \triangleq \mathbf{S}\boldsymbol{\eta}$ whose noise covariance matrix can be expressed with respect to the selection matrix, i.e.,

$$\hat{\mathbf{R}}_\eta = \sigma_\epsilon^2 \mathbf{W}^H \mathbf{H}\mathbf{F}_{\rm RF}\mathbf{S}\mathbf{F}_{\rm BB}\mathbf{F}_{\rm BB}^H\mathbf{S}\mathbf{F}_{\rm RF}^H\mathbf{H}^H\mathbf{W} + \sigma_{\rm n}^2 \mathbf{W}^H\mathbf{W}. \qquad (21)$$

The problem becomes equivalent with the estimation of $\mathbf{S}$ that maximizes the EE of the hybrid precoder. It can be shown that the rate and power equations for such scenario can be expressed as:

$$R(\mathbf{S},\delta) = \log_2 \left| \mathbf{I}_{N_{\rm s}} + \frac{1}{N_{\rm s}} \sum_{i=1}^{L_{\rm T}} [\mathbf{S}]_{ii} \mathbf{a}_i^H \mathbf{a}_i \mathbf{b}_i \mathbf{b}_i^H \right|, \qquad (22)$$

Algorithm 2: Proposed technique

```
Input: $\kappa^{(0)}$, $\mathbf{H}$
Begin:
for $b = 1, ..., b_{\rm max}$
    Compute $\mathbf{H}_{\rm eff}(N_{\rm T}, \delta(b))$
        for $m = 1, 2, \ldots, I_{\rm max}$
            Obtain $\mathbf{S}^{(m)}$ by solving (25) given $\kappa^{(m-1)}$.
            Calculate $R(\mathbf{S}^{(m)}, \delta^{(m)})$ and $P(\mathbf{S}^{(m)}, \delta^{(m)})$.
            Compute $\kappa^{(m)} = R(\mathbf{S}^{(m)}, \delta^{(m)})/P(\mathbf{S}^{(m)}, \delta^{(m)})$.
        end
end
Output: Optimal ${L_{\rm T}}^{opt}$ and $b^{opt}$
```

and

$$\begin{aligned}P(\mathbf{S},\delta) &\propto \sum_{i=1}^{L_{\rm T}} [\mathbf{S}]_{ii} \left[P_{\rm DAC} \left(\frac{\pi\sqrt{3}}{2(1-\delta^2)} \right)^{1/2} + N_{\rm T} P_{\rm PS} \right] && (23)\\ &= L_{\rm T} \left[P_{\rm DAC} \left(\frac{\pi\sqrt{3}}{2(1-\delta^2)} \right)^{1/2} + N_{\rm T} P_{\rm PS} \right]. \ (W) && (24)\end{aligned}$$

The problem of maximizing EE (16) is a concave-convex fractional problem and one solution method is the Dinkelbach approximation [20]. The Dinkelbach method is an iterative and parametric algorithm, where a sequence of easier problems converge to the global solution. Let $\kappa^{(m)} = R(\mathbf{S}^{(m)}, \delta^{(m)})/P(\mathbf{S}^{(m)}, \delta^{(m)}) \in \mathbb{R}$, for $m = 1, 2, \ldots, I_{\rm max}$, where $I_{\rm max}$ is the number of maximum iterations, then each iteration step of Dinkelbach can be expressed as:

$$\mathbf{S}^{(m)}(\kappa^{(m)}) \triangleq \arg\max_{\mathbf{S}\in\mathcal{S}} \left\{ R(\mathbf{S},\delta) - \kappa^{(m)} P(\mathbf{S},\delta) \right\}, \qquad (25)$$

where $\mathcal{S}$ is the set of diagonal matrices with the feasible bit allocations which satisfy $P(\mathbf{S},\delta) \leq P_{\rm max}$. Algorithm 2 summarizes the Dinkelbach algorithm via the subset selection approach where the optimal number of RF chains and associated sampling resolution is obtained.

Computational Complexity: It can be observed that the Dinkelbach method via subset selection approach requires complexity order of only $b_{\rm max}\mathcal{O}(L_{\rm T}^3)$ per iteration and the Brute-force approach requires complexity order of $b_{\rm max}\mathcal{O}(L_{\rm T}^2 N_{\rm T})$. Since the number of the required iterations is usually very small (as shown in Fig. 2) as $\mathbf{F}$ and $\mathbf{W}$ matrices are required to be computed in Algorithm 1 and not Algorithm 2, the overall complexity of the Dinkelbach method via subset selection approach is much less than the Brute-force approach.

IV. Simulation Results

In this section, we evaluate the performance of the proposed technique using computer simulation results. The simulations are performed with MATLAB$^{\rm TM}$ and all the results have been averaged over 1,000 Monte-Carlo realizations.

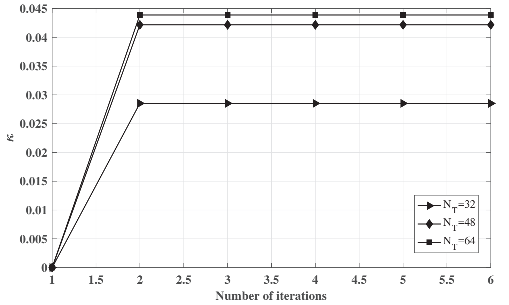


Fig. 2. Convergence of the proposed Dinkelbach method for different number of transmitter antennas at SNR = 30 dB, $N_{\mathrm{R}} = 32$, $L_{\mathrm{T}} = 32$ and $N_{\mathrm{s}} = 8$.

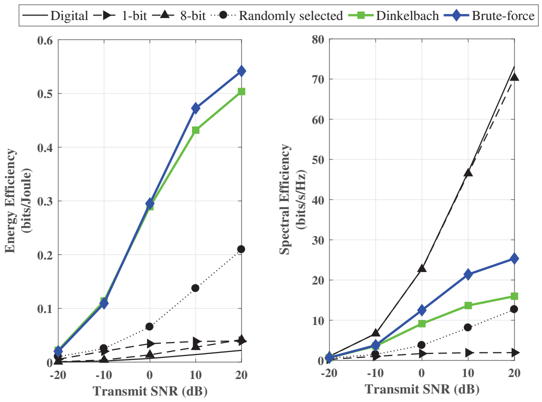


Fig. 3. Energy efficiency and spectral efficiency performance comparison w.r.t. transmit SNR (dB) at $N_{\mathrm{T}} = 64$, $N_{\mathrm{R}} = 32$, $L_{\mathrm{T}} = 32$ and $N_{\mathrm{s}} = 8$.

We set the following baseline parameters for simulation: $N_{\mathrm{T}} = 64$, $N_{\mathrm{R}} = 32$, $L_{\mathrm{T}} = 32$ (the number of available RF chains), $N_{\mathrm{s}} = 8$, $N_{\mathrm{cl}} = 2$, $N_{\mathrm{ray}} = 10$, and $\sigma^2_{\alpha,i} = 1$. The azimuth angles of departure and arrival are computed with uniformly distributed mean angles; each cluster follows a Laplacian distribution with mean angles equal to zero. The antenna elements in the ULA are spaced by distance $d = \lambda/2$.

Concerning the quantization model, since DACs have the same sampling resolution for each RF chain the quantization distortion parameter is the same for all DACs and the highest bit resolution $b_{\max} = 8$. The typical values of power terms for the power model in (12) of Section III are $P_{\mathrm{PS}} = 10$ mW, $P_{\mathrm{DAC}} = 0.1$ W and $P_{\max} = 1$ W. We solve the sparse approximation problem for the RF and baseband precoding matrices $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ using orthogonal matching pursuit (OMP) [4], [6], and the combiner matrix $\mathbf{W}$ is the product of $1/\sqrt{N_{\mathrm{s}}}$ and first N_{s} columns of $\mathbf{U}$ matrix.

For comparison with the proposed Dinkelbach method via subset selection solution, we have considered the digital beamforming architecture ($L_{\mathrm{T}} = N_{\mathrm{T}}$) with 8-bit DACs, which represents the optimum from the achievable spectral efficiency (SE) perspective, combined analog and digital hybrid precoding with L_{T} RF chains for 1-bit and 8-bit DACs, which represent the lowest and the highest SE cases. We also compare with the hybrid beamforming for L_{T} RF chains with a random resolution selected for each DAC from the range $[1, 8]$-bit, and hybrid beamforming with the optimal number of active RF chains L_{T}^{opt} and corresponding optimal sampling resolution b^{opt} obtained from the Brute-force approach.

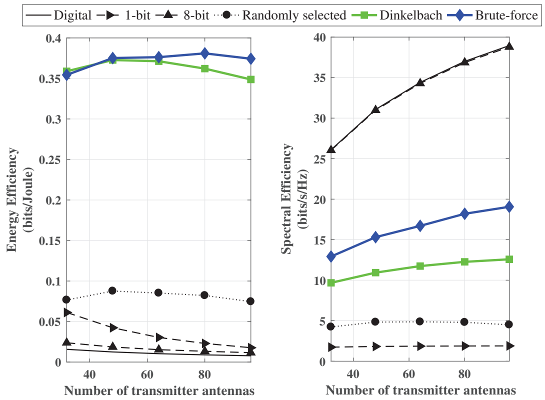


Fig. 4. Energy efficiency and spectral efficiency performance comparison w.r.t. the number of transmitter antennas at SNR = 5 dB, $N_{\mathrm{R}} = 32$, $L_{\mathrm{T}} = 32$ and $N_{\mathrm{s}} = 8$.

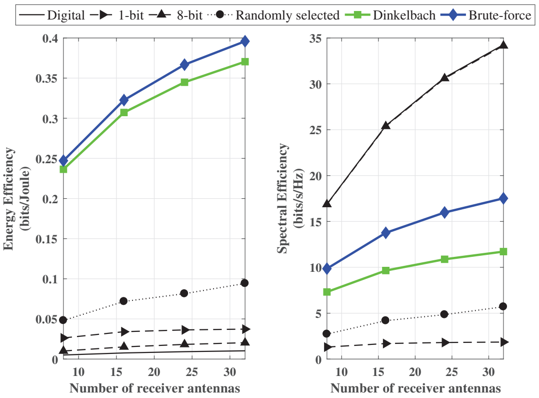


Fig. 5. Energy efficiency and spectral efficiency performance comparison w.r.t. the number of receiver antennas at SNR = 5 dB, $N_{\mathrm{T}} = 64$, $L_{\mathrm{T}} = 32$ and $N_{\mathrm{s}} = 8$.

Fig. 2 shows the convergence of the Dinkelbach method based solution as proposed in Algorithm 2 to obtain the optimal number of active RF chains and corresponding optimal sampling resolution. It can be observed that the performance curves based on the current EE κ (step 6 of Algorithm 2) for different numbers of transmitter antennas increase with respect to (w.r.t.) the number of iterations. The proposed solution converges rapidly and needs only 2-3 iterations to converge, and achieves an optimal solution at each realization.

It can be clearly observed from Fig. 3 that the proposed solution achieves a similar EE performance w.r.t. signal-to-noise ratio (SNR) as the Brute-force approach and outperforms hybrid 1-bit and hybrid 8-bit quantized DACs, plus the hybrid randomly selected resolution and digital beamforming with full-bit (8-bit) quantization. For example, at 10 dB SNR, EE for the proposed solution is approximating the Brute-force solution performance, about 0.3 bits/Joule better than the randomly selected resolution with hybrid beamforming, about 0.35 bits/Joule better than the hybrid 1-bit and about 0.38 bits/Joule better than the hybrid 8-bit and digital beamforming baselines. The proposed solution also achieves SE performance higher than the randomly selected and 1-bit quantization baselines. Digital beamforming and 8-bit hybrid baselines have the highest rate performance by using higher rate 8-bit quantization. For example, at 0 dB SNR, the proposed solution outperforms randomly selected quantization by about 7 bits/s/Hz, 1-bit hybrid by about 9 bits/s/Hz. Concerning the lower SE performance of the proposed technique and the Brute-Force approach, this is due to the fact that Brute-force has no constraint in the overall power consumption.

Fig. 4 shows similar performance behavior when plotting EE and SE w.r.t. the number of transmitter antennas at 5 dB SNR. For example, for $N_{\mathrm{T}} = 80$, the proposed solution has performance close to the Brute-force approach, performs about 0.3 bits/Joule and about 7.5 bits/s/Hz better than the hybrid randomly selected resolution baseline, about 0.35 bits/Joule and 10 bits/s/Hz better than the 1-bit hybrid baseline. Fig. 5 plots the performance comparison of the proposed solution with the baselines w.r.t. number of receiver antennas at 5 dB SNR. Similar to above plots, it achieves high SE and has almost the same EE performance as the Brute-force approach.

V. Conclusion

We consider a mmWave hybrid MIMO system with analog and digital parts connected with fewer number of RF chains than the transmitting antennas, while transmitter DACs operate with low-resolution sampling. We consider the case where all DACs have the same sampling resolution for each RF chain and aim to optimize the number of active RF chains and associated resolution of DACs. The proposed method achieves similar EE performance with the upper bound of the derived exhaustive search approach, while it exhibits lower computational complexity and fast convergence. Future work will include the optimization of energy efficiency with different bit resolutions for every RF chain.

Acknowledgment

The authors gratefully acknowledge partial funding of this work by EPSRC grant EP/P000703/1.

References

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems", *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101-107, Jun. 2011.

[2] J. G. Andrews et al., "What will 5G be", *IEEE Journ. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[3] R. W. Heath et al.,"An overview of signal processing techniques for millimeter wave MIMO systems", *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[4] O. E. Ayach et al., "Spatially sparse precoding in millimeter wave MIMO systems", *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[5] S. Han et al., "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.

[6] A. Kaushik et al.,"Sparse hybrid precoding and combining in millimeter wave MIMO systems", in *Proc. IET Radio Prop. Tech. 5G*, Durham, UK, pp. 1-7, Oct. 2016.

[7] A. Kaushik et al.,"Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs", in *IEEE Europ. Sig. Process.*, Rome, Italy, pp. 1839-1843, Sept. 2018.

[8] O. Orhan et al., "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance", *2015 Info. Theory Appl. Workshop, San Diego, CA*, pp. 191-198, 2015.

[9] L. Fan et al., "Uplink achievable rate for massive MIMO systems with low-resolution ADC", *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2186-2189, Oct. 2015.

[10] R. Mendez-Rial et al., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?", *IEEE Access*, vol. 4, pp. 247-267, Jan. 2016.

[11] J. Zhang et al., "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327-1338, Jun. 2017.

[12] A. Mezghani et al., "Transmit processing with low resolution D/A-converters", in *Proc. Int. Conf. Electronics, Circuits, and Systems*, Tunisia, pp. 683-686, Dec. 2009.

[13] S. Jacobsson et al., "Quantized precoding for massive MU-MIMO", in *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670-4684, Nov. 2017.

[14] R. Zi et al., "Energy efficiency optimization of 5G radio frequency chain systems", *IEEE Journ. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.

[15] C. G. Tsinos et al., "On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems", in *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980-1995, Sept. 2017.

[16] S. Singh et al., "Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control", *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513-1527, Oct. 2011.

[17] C. Balanis, *Antenna Theory*, Wiley, 1997.

[18] J. Mo et al., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," in *IEEE Trans. Sig. Process.*, vol. 66, no. 5, pp. 1141-1154, Mar. 2018.

[19] E. Vlachos et al.,"Energy efficient transmitter with low resolution DACs for massive MIMO with partially connected hybrid architecture", *IEEE Veh. Tech. Conf. (VTC)-Spring*, Porto, Portugal, Jun. 2018.

[20] W. Dinkelbach, "On nonlinear fractional programming", *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.

[21] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs", in *Recent Adv. Learning and Control*, Springer-Verlag Ltd., pp. 95-110, 2008.

[22] A. Zappone and E. Jorswieck, "Energy Efficiency in Wireless Networks via Fractional Programming Theory", Foundations and Trends in Communications and Information Theory: Vol. 11: No. 3-4, pp 185-396, 2015.

2018 26th European Signal Processing Conference (EUSIPCO)

Efficient Channel Estimation in Millimeter Wave Hybrid MIMO Systems with Low Resolution ADCs

Aryan Kaushik, Evangelos Vlachos, John Thompson, and Alessandro Perelli
Institute for Digital Communications, The University of Edinburgh, United Kingdom.
Email: {A.Kaushik, E.Vlachos, J.S.Thompson, A. Perelli}@ed.ac.uk

***Abstract*—This paper proposes an efficient channel estimation algorithm for millimeter wave (mmWave) systems with a hybrid analog-digital multiple-input multiple-output (MIMO) architecture and few-bits quantization at the receiver. The sparsity of the mmWave MIMO channel is exploited for the problem formulation while limited resolution analog-to-digital converters (ADCs) are used in the receiver architecture. The estimation problem can be tackled using compressed sensing through the Stein's unbiased risk estimate (SURE) based parametric denoiser with the generalized approximate message passing (GAMP) framework. Expectation-maximization (EM) density estimation is used to avoid the need of specifying channel statistics resulting the EM-SURE-GAMP algorithm to estimate the channel. SURE, depending on the noisy observation, is minimized to adaptively optimize the denoiser within the parametric class at each iteration. The proposed solution is compared with the expectation-maximization generalized AMP (EM-GAMP) solution and the mean square error (MSE) performs better with respect to low and high signal-to-noise ratio (SNR) regimes, the number of ADC bits, and the training length. The use of the low resolution ADCs reduces power consumption and leads to an efficient mmWave MIMO system.**

***Keywords*—channel estimation, low resolution analog-to-digital converter (ADC), compressed sensing, mmWave MIMO.**

I. Introduction

The large number of antenna elements associated with millimeter wave (mmWave) multiple input multiple output (MIMO) systems makes it hard to use many analog-to-digital converters (ADCs), which is a power hungry component [1]. Moreover, ADCs have much higher sampling rates for wide bandwidth mmWave systems than at microwave frequencies, and employing high speed ADCs increases the power consumption and the cost significantly [2], [3]. Implementing low resolution ADCs such as 1-bit to 3-bits in mmWave MIMO systems efficiently improves the power metric of the system [1]. Fig. 1 shows the hardware block diagram of a mmWave system with a hybrid analog-digital architecture and low resolution ADCs at the receiver. The use of 1-bit ADCs in MIMO systems has been discussed in [4] and [5], and channel estimation is investigated as well. In that work, the channel is known perfectly to the transmitter and the receiver while in practical scenarios, the channel state information (CSI) is not known and should be estimated by both the transmitter and the receiver.

References [6]-[8] estimate the sparse mmWave channel using signal processing tools for high resolution analog to digital converting structures, but the use of low resolution ADCs at the receiver can significantly reduce the power consumption without significantly affecting the capacity of the system [9]. Recently, [10] and [11] considered 1-bit ADC quantization systems and the sparsity in the angle domain is exploited to be able to use compressed sensing (CS) techniques to recover the channel parameters. The proposed adaptive technique in [10] fails to provide good estimation of the channel at low SNR values. Reference [11] proposes only an expectation-maximization (EM) algorithm which has high complexity since each iteration requires a matrix inverse computation and convergence of the algorithm requires many iterations. To observe the effect of low resolution ADCs, an additive quantization model (AQNM) is considered in [12] and [13]. The effect of AQNM is investigated in [12] for the case of a point-to-point mmWave MIMO system, while in [13] the desired rate of the uplink was derived for the case of mmWave fading channels. References [14] and [15] also implement the EM algorithm for a MIMO channel. Further improvements to the EM algorithm are proposed using expectation-maximization generalized approximate message passing (EM-GAMP) [16] and vector approximate message passing (VAMP) [17]. The use of EM-GAMP has been exploited for a broadband mmWave MIMO channel model with low resolution ADCs at the receiver in [18].

Reference [19] describes the advantages of the Stein's unbiased risk estimate (SURE) based parametric denoiser when incorporated with the approximate message passing (AMP) framework. This paper exploits the SURE-generalized AMP solution combined with expectation-maximization (EM) steps called the EM-SURE-GAMP in a mmWave MIMO system. This novel solution avoids strong assumptions on the channel statistics where SURE, depending on the noisy observation, is minimized to adaptively optimize the denoiser within the parametric class at each iteration. The proposed solution is compared with the EM-GAMP solution for a narrowband channel model and improved mean square error (MSE) performance is observed for both low and high signal-to-noise ratio (SNR) regimes. The unknown channel parameters are modeled by a Bernoulli Gaussian distribution for both the techniques.

Notations: x, $\mathbf{x}$, and $\mathbf{X}$, represent a scalar, a vector, and a matrix, respectively; the i^{th} column of $\mathbf{X}$ is $\mathbf{X}^{(i)}$; the transpose of $\mathbf{X}$ is $\mathbf{X}^T$ while the conjugate transpose is $\mathbf{X}^*$; tr($\mathbf{X}$) and $|\mathbf{X}|$, are the trace and determinant of $\mathbf{X}$, while $||\mathbf{X}||_F$ is the Frobenius norm; the p-norm of $\mathbf{x}$ is $||\mathbf{x}||_p$; $\mathbf{X} \otimes \mathbf{Y}$ represents the Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$, diag($\mathbf{X}$) generates a vector of the diagonal elements of $\mathbf{X}$; vec($\mathbf{X}$) is a vector showing all the columns of $\mathbf{X}$, $\mathbf{I}_N$ represents an identity matrix of dimension $N \times N$ and $\mathbf{0}_{A \times B}$ is an all-zeros matrix of dimension $A \times B$. $\mathbb{E}[.]$ represents the expectation of a complex variable. $\mathbb{R}^{A \times B}$ and $\mathbb{C}^{A \times B}$ denote the set of

ISBN 978-90-827970-1-5 © EURASIP 2018

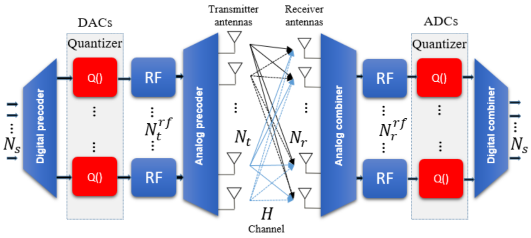


Fig. 1: MmWave system with a hybrid analog-digital MIMO architecture and low resolution ADCs at the receiver.

$A \times B$ matrices with real and complex entries, respectively. A complex Gaussian vector with mean $\mathbf{x}$ and covariance matrix as $\mathbf{X}$ is represented as $\mathcal{CN}(\mathbf{x};\mathbf{X})$, and i.i.d. indicates the entries to be independent and identically distributed.

II. MmWave Hybrid MIMO Model

The high path loss and small number of multi-path components in mmWave MIMO systems restrict use of the fading channels used in the analysis of MIMO systems [1]. Consider a single-user mmWave MIMO system with N_t antennas at the transmitter, with N_s transmitted data streams to N_r receiver antennas. For the number of multipaths computed by the product of N_{cl} clusters and N_{ray} rays in every cluster, the narrowband channel is written as follows:

$$\mathbf{H} = \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{il} \mathbf{a}_r(\phi_{il}^r) \mathbf{a}_t(\phi_{il}^t)^*, \tag{1}$$

α_{il} in (1) is the complex gain of l^{th} ray in i^{th} cluster; $\mathbf{a}_t(\phi_{il}^t)$ and $\mathbf{a}_r(\phi_{il}^r)$ are the normalized transmit and receive array response vectors, where ϕ_{il}^t and ϕ_{il}^r are the elevation angles of departure and arrival, respectively. We modeled the antenna elements as ideal sectored elements at both the transmitter and the receiver [20]. In (1), the transmit and receive antenna element gains are considered unity over the sectors defined by $\phi_{il}^t \in [\phi_{min}^t, \phi_{max}^t]$ and $\phi_{il}^r \in [\phi_{min}^r, \phi_{max}^r]$, respectively. We implement uniform linear array (ULA) geometry. For λ signal wavelength, d inter-element spacing, and a ULA geometry with N_z antenna elements, the array response vector is written as follows [21]:

$$\mathbf{a}_z(\phi) = \frac{1}{\sqrt{N_z}} [1, e^{j\frac{2\pi}{\lambda} d \sin(\phi)}, ..., e^{j(N_z-1)\frac{2\pi}{\lambda} d \sin(\phi)}]^T, \tag{2}$$

Equation (2) can be used to compute the array response vectors at both the transmitter and receiver with the corresponding terms. The beamspace representation [22], [23] of the narrowband channel in (1) can be written as follows:

$$\mathbf{H} = \hat{\mathbf{A}}_r \mathbf{Z} \hat{\mathbf{A}}_t^*, \tag{3}$$

where $\mathbf{Z} \in \mathbb{C}^{N_r \times N_t}$ represents a sparse matrix with a few non-zero entries assumed to follow Bernoulli-Gaussian distribution, while $\hat{\mathbf{A}}_r \in \mathbb{C}^{N_r \times N_r}$ and $\hat{\mathbf{A}}_t \in \mathbb{C}^{N_t \times N_t}$ are DFT matrices.

Let us consider a MIMO $N_t \times N_r$ system with a hybrid analog-digital architecture with N_t^{rf} and N_r^{rf} chains at the transmitter and the receiver, respectively. The number of RF chains is smaller or equal to the number of antennas for both the transmitter $N_t^{rf} \leq N_t$ and the receiver $N_r^{rf} \leq N_r$. We assume that the channel is quasi-static, i.e., it remains static during a period of time, which includes both channel training and data transmission phases. During the training phase, at each time instance t, the transmitter generates a training signal vector $\mathbf{s}(t) \in \mathbb{C}^{N_t^{rf} \times 1}$ following $\mathbb{E}[\mathbf{s}(t)\mathbf{s}(t)^*] = \frac{1}{N_s}\mathbf{I}_{N_s}$, which is the input to the analog RF precoder at transmitter, $\mathbf{F}_{rf}(t) \in \mathbb{C}^{N_t \times N_t^{rf}}$. This signal is transmitted through the channel $\mathbf{H}$ and the received vector is processed by the analog RF combiner at receiver, $\mathbf{W}_{rf}(t) \in \mathbb{C}^{N_r \times N_r^{rf}}$. The elements of the RF precoders and combiners have equal norm as they represent transmitter and receiver phase shifters. For the case of number of streams equal to the number of RF chains, the baseband matrices, $\mathbf{F}_{bb}(t) \in \mathbb{C}^{N_t^{rf} \times N_s}$ at transmitter and $\mathbf{W}_{bb}(t) \in \mathbb{C}^{N_r^{rf} \times N_s}$ at receiver, are identity matrices so we consider only RF/analog processing to formulate the channel estimation problem. The received signal after RF/analog processing, $\mathbf{y}_c(t) \in \mathbb{C}^{N_r \times 1}$ for $t = 1, \ldots, T$, is expressed as:

$$\mathbf{y}_c(t) = \mathbf{W}_{rf}^*(t) \mathbf{H} \mathbf{F}_{rf}(t) \mathbf{s}(t) + \mathbf{n}_c(t), \tag{4}$$

where $\mathbf{n}_c \in \mathbb{C}^{N_r \times 1}$ noise vector following the complex Gaussian distribution with i.i.d. entries, i.e., $\mathbf{n}_c \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$. By concatenating all the T training sequences into the real-valued equivalent form we have:

$$\bar{\mathbf{y}} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{y}}_c) \\ \mathrm{Im}(\bar{\mathbf{y}}_c) \end{bmatrix} = \bar{\mathbf{\Psi}} \begin{bmatrix} \mathrm{Re}(\mathbf{z}_c) \\ \mathrm{Im}(\mathbf{z}_c) \end{bmatrix} + \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{n}}_c) \\ \mathrm{Im}(\bar{\mathbf{n}}_c) \end{bmatrix}, \tag{5}$$

where $\bar{\mathbf{\Psi}} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{\Psi}}_c) & -\mathrm{Im}(\bar{\mathbf{\Psi}}_c) \\ \mathrm{Im}(\bar{\mathbf{\Psi}}_c) & \mathrm{Re}(\bar{\mathbf{\Psi}}_c) \end{bmatrix}^T \in \mathbb{R}^{2TN_r \times 2N_r N_t}$ and $\bar{\mathbf{y}}_\mathbf{c}, \bar{\mathbf{n}}_\mathbf{c}, \bar{\mathbf{\Psi}}_\mathbf{c}$ are the concatenated quantities for the received signal, the AWGN and the system matrix, respectively.

Let us denote the K-level quantization of $\bar{\mathbf{y}} \in \mathbb{R}^{2TN_r \times 1}$ as the function $\mathcal{Q}(.)$,

$$\bar{\mathbf{q}} = \mathcal{Q}(\bar{\mathbf{y}}), \tag{6}$$

where $\bar{\mathbf{q}} = [q_1 \ldots q_{2TN_r}]^T \in \mathbb{R}^{2TN_r \times 1}$. Each output element takes one of K distinct values with,

$$q_i^k = -l_i^k + \frac{\Delta}{2} + (k-1)\Delta, \forall k = 1, ..., K, \tag{7}$$

depending on the quantizer lower and upper thresholds $[l_i^k, u_i^k]$ where $l_i^k = -\kappa\sqrt{\mathbb{E}\{y_i^2\}}$ and $u_i^k = \kappa\sqrt{\mathbb{E}\{y_i\}}$, $\forall i$ and $\kappa \in [1, 5]$. The quantizer's step-size is given by $\Delta = \frac{u_i^k - l_i^k}{K}$, while the average power $\mathbb{E}\{y_i\}$ can be obtained via an automatic gain control (AGC) circuit.

III. Proposed Channel Estimation Solution

A. Problem Formulation

Following the beamspace representation of the sparse mmWave channel in (3), the system model of (4) can be rewritten into an equivalent form for the channel estimation problem, i.e.,

$$\mathbf{y}_c(t) = \underbrace{\left(\mathbf{s}^T(t)\mathbf{F}_{rf}^T(t)\hat{\mathbf{A}}_t \otimes \mathbf{W}_{rf}^*(t)\hat{\mathbf{A}}_r\right)}_{\mathbf{\Psi}_c(t)} \underbrace{\mathrm{vec}(\mathbf{Z})}_{\mathbf{z}} + \mathbf{n}_c(t), \tag{8}$$

thus, sparse estimation techniques can be utilized to recover the sparse vector $\mathbf{z}$.

Concerning the analog RF beamforming matrices, these are designed as random matrices [24] as we require sensing matrix to be random to be able to apply compressed sensing. The transmitter and the receiver share a pseudo-random key so receiver can predict the precoding matrix. In particular, the angles of precoding/combiner matrices are generated as random variables following a uniform distribution, i.e., $\tilde{\phi}_i(t) \sim \mathcal{U}(0, 2\pi)$. Then, for each training instance t and $\forall k = 1, \ldots, N_t, i = 1, \ldots, N_t^{rf}$ we use the matrix:

$$[\mathbf{F}_{rf}(t)]_{ki} = \frac{1}{\sqrt{N_t}} e^{j(k-1)\sin(\tilde{\phi}_i(t))}, \tag{9}$$

for precoding, and accordingly for the combiner at the receiver:

$$[\mathbf{W}_{rf}(t)]_{ki} = \frac{1}{\sqrt{N_t}} e^{j(k-1)\sin(\tilde{\phi}_i(t))}. \tag{10}$$

To overcome the quantization non-linearity effects at the receiver, we employ quantization dithering [25]. In this work we consider a simple type of dithering termed as non-subtractive random dithering. Specifically, we assume that a Gaussian random signal with zero mean, i.e., $\bar{\mathbf{d}} \sim \mathcal{N}(\mathbf{0}, \sigma_d^2 \mathbf{I})$ is added to the input, thus, the overall system is described as:

$$\bar{\mathbf{r}} = \mathcal{Q}\big(\bar{\boldsymbol{\Psi}}\mathbf{z} + \bar{\mathbf{n}} + \bar{\mathbf{d}}\big) \in \mathbb{R}^{2TN_r \times 1}, \tag{11}$$

where $\bar{\mathbf{d}} \in \mathbb{R}^{2TN_r \times 1}$ is the control signal. The overall noise can be modelled as $\bar{\mathbf{n}} + \bar{\mathbf{d}} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2 \mathbf{I}\big)$, where $\sigma^2 = \sigma_n^2 + \sigma_d^2$.

B. EM-SURE-GAMP Solution for Channel Estimation

To solve the non-linear sparse channel estimation problem of (8) we obtain an approximation of the maximum a-posteriori channel estimator via the EM algorithm [11], for l-th iteration, i.e.,

$$\mathbb{E}_{\bar{\mathbf{y}}|\bar{\mathbf{r}},\mathbf{z}} \left\{ \frac{\partial}{\partial \mathbf{z}} \ln p(\bar{\mathbf{r}}, \bar{\mathbf{y}} | \mathbf{z}^l) \right\} = 0, \tag{12}$$

where the conditional probability density function (PDF) involving $\bar{\mathbf{r}}$ and $\bar{\mathbf{y}}$ random variables is given by [26] as follows:

$$p(\bar{\mathbf{r}}, \bar{\mathbf{y}} | \mathbf{z}) = \mathbb{I}_{D(\bar{\mathbf{r}})}(\bar{\mathbf{y}}) \frac{1}{(2\pi\sigma^2)^{2TN_r \times 1/2}} e^{\frac{-\|\bar{\mathbf{y}} - \bar{\boldsymbol{\Psi}}\mathbf{z}\|_2^2}{2\sigma^2}}. \tag{13}$$

The EM algorithm is defined by the following two steps for the $(l+1)$-th iteration:

- **E-step:** Compute $\mathbf{b}^l = [b_1^l, \ldots, b_{2TN_r}^l]$ with

  $$b_i^l = -\frac{\sigma}{\sqrt{2\pi}} \frac{e^{-\frac{(l_i - [\bar{\boldsymbol{\Psi}}\mathbf{z}^l]_i)^2}{2\sigma^2}} - e^{-\frac{(u_i - [\bar{\boldsymbol{\Psi}}\mathbf{z}^l]_i)^2}{2\sigma^2}}}{\operatorname{erf}(\frac{-l_i + [\bar{\boldsymbol{\Psi}}\mathbf{z}^l]_i}{\sqrt{2}\sigma}) - \operatorname{erf}(\frac{-u_i + [\bar{\boldsymbol{\Psi}}\mathbf{z}^l]_i}{\sqrt{2}\sigma})}, \tag{14}$$

  where l_i, u_i are the lower and upper bounds for the i^{th} quantized sample of the quantizer for $[\bar{\boldsymbol{\Psi}}\mathbf{z}^l]_i$ respectively; $\operatorname{erf}(\cdot)$ is the error function.
- **M-step:** Estimate the sparse channel $\mathbf{z}^{l+1} \in \mathbb{R}^{2N_r N_t \times 1}$ via solution of the linear system of equations:

  $$\mathbf{A}\mathbf{z}^{l+1} = \boldsymbol{\delta}_l, \tag{15}$$

  with $\boldsymbol{\delta}_l \triangleq \bar{\boldsymbol{\Psi}}^T \bar{\boldsymbol{\Psi}} \mathbf{z}^l + \mathbf{b}^l$ and $\mathbf{A} \triangleq \bar{\boldsymbol{\Psi}}^T \bar{\boldsymbol{\Psi}} + \mathbf{C}_h^{-1}$, where $\mathbf{C}_h^{-1}$ is the correlation matrix based on the channel known statistics.

Algorithm 1: EM-SURE-GAMP algorithm

```
1  Initialization: $\hat{\mathbf{z}}^1 = \mathbf{0}, \boldsymbol{\xi}^0 = \mathbf{0}, c^1 = \frac{1}{2N_r N_t}, \tau_z^1 = 1.$
2  for t = $1, \ldots, T_{max}$ do
3      $\boldsymbol{\gamma}^t = \mathbf{A}\hat{\mathbf{z}}^t$
4      $\tau_p^t = \frac{1}{2N_r N_t} \|\mathbf{A}\|_F^2 \tau_z^t$
5      $\mathbf{p}^t = \boldsymbol{\gamma}^t - \tau_p^t \boldsymbol{\xi}^{t-1}$
6      Update $\boldsymbol{\delta}_l$ using EM-steps as indicated in (15)
7      $\boldsymbol{\xi}^t = \mathbb{E}_{p(\boldsymbol{\gamma}^t|\mathbf{p}^t, \tau_p^t, \boldsymbol{\delta}_l)}[\boldsymbol{\gamma}^t|\mathbf{p}^t, \tau_p^t, \boldsymbol{\delta}_l]$
8      $\tau_\xi^t = \frac{1}{2N_r N_t \tau_p^t} \left[1 - \frac{\mathrm{Var}_{p(\boldsymbol{\gamma}^t|\mathbf{p}^t,\tau_p^t,\boldsymbol{\delta}_l)}[\boldsymbol{\gamma}^t|\mathbf{p}^t,\tau_p^t,\boldsymbol{\delta}_l]}{\tau_p^t}\right]$
9      $\frac{1}{\tau_\beta^t} = \frac{1}{2N_r N_t} \|\mathbf{A}\|_F^2 \tau_\xi^t$
10     $\boldsymbol{\beta}^t = \hat{\mathbf{z}}^t + \tau_\beta^t \mathbf{A}^* \boldsymbol{\xi}^t$
11     $\boldsymbol{\theta}^t = H_t(\boldsymbol{\beta}^t, c^t)$
12     $\hat{\mathbf{z}}^{t+1} = f_t(\boldsymbol{\beta}^t, c^t | \boldsymbol{\theta}^t)$
13     $\tau_z^{t+1} = \tau_\beta^t f_t'(\boldsymbol{\beta}^t, c^t | \boldsymbol{\theta}^t)$
14     $c^{t+1} = \frac{1}{2N_r N_t} ||\tau_\beta^t \boldsymbol{\xi}^t||_2^2$
15 end for
```

The linear channel estimation problem in (15) can be considered similar to the noisy quantized CS problem [27]; among the numerous existing algorithms for sparse inverse linear problems, AMP-based solver has been shown to converge faster, i.e. in few iterations, with predictable dynamics together with low computational complexity. In its original formulation for l_1-minimization [28], AMP is a designed as a variant of a soft-thresholding iterative algorithm; in [29], [30] extensions of AMP have been used to handle wide class of random sensing matrices and for sparse learning applications. Generally AMP family of algorithms has been proven to converge for the class of right orthogonal random matrices; to reduce the convergence problems with general structured random matrices, damping is often used. However, for our system model we do not need to perform damping on the update of the messages.

In particular, AMP-based algorithms perform a sequence of MMSE estimations of the estimated measurement vector $\boldsymbol{\gamma}^t = \bar{\boldsymbol{\Psi}}\hat{\mathbf{z}}^t$, such as in line 3 of Algorithm 1, where $\hat{\mathbf{z}}^t$ refers to the estimate of the vector $\mathbf{z}^{l+1}$ for the M-step in (15) and l is the EM iteration index. Regarding the MMSE estimator for $\boldsymbol{\gamma}^t$, since the channel noise model in (11) is quantized Gaussian as it is modeled as the quantization function, we need to adopt the generalized version of AMP (GAMP) [31] whose computation is detailed in the Algorithm 1 where the expectation is over the posterior probability $p(\boldsymbol{\gamma}^t|\mathbf{p}^t, \tau_p^t, \boldsymbol{\delta}_l)$ which is dependent on the quantizer function $\mathcal{Q}$ through (14). $\boldsymbol{\delta}_l$ represents the vector of measurements updated using the EM-steps as indicated in (15). In line 8 of Algorithm 1, $\mathrm{Var}_{p(\boldsymbol{\gamma}^t|\mathbf{p}^t,\tau_p^t,\boldsymbol{\delta}_l)}[\cdot]$ represents the Variance of the conditional probability distribution $p(\boldsymbol{\gamma}^t|\mathbf{p}^t, \tau_p^t, \boldsymbol{\delta}_l)$. Regarding the MMSE estimator for $\hat{\mathbf{z}}^t$, standard AMP [28] is based on the assumption that the prior $p(\hat{\mathbf{z}}^t)$ is precisely defined and, therefore, it is possible to derive the associated MMSE estimator.

In this work, we utilize a variant, named SURE-GAMP, which derives specific MMSE estimators tailored for the dithered system model in (11) as follows. The SURE approach [19] aims to find the denoiser within a class with the least MSE by optimizing the free parameters $\boldsymbol{\theta}^t$ of some piecewise kernel functions $f_t(\cdot|\boldsymbol{\theta}^t)$ in order to obtain an optimal adaptive non linearity; moreover, the optimization of the denoiser does not require knowledge of the prior distribution. In the simulations, SURE-GAMP uses a family of parameterized denoising functions for the class of Bernoulli Gaussian signals, which can be analyzed through Gaussian-mixture distribution as well [18]. At each iteration, the parametric SURE-GAMP algorithm adaptively chooses the best denoiser, i.e. the one with the least MSE, by selecting the parameters $\boldsymbol{\theta}^t$ which correspond to the minimum of the selection function H_t, such as in line 11 of Algorithm 1, dependent on the noisy data $\boldsymbol{\beta}^t$ and the estimate of the effective noise variance c^t which leads to solving the following optimization problem:

$$\begin{aligned} \boldsymbol{\theta}^t &= H_t(\boldsymbol{\beta}^t, c^t) \qquad (16)\\ &= \arg\min_{\boldsymbol{\theta}} \mathbb{E}[f(\boldsymbol{\beta}^t, c^t|\boldsymbol{\theta}) - \boldsymbol{\beta}^t)^2 + 2c^t f'(\boldsymbol{\beta}^t, c^t|\boldsymbol{\theta})] \end{aligned}$$

In [19], authors have shown that this optimization is equivalent to solving a linear system of equations whose dimension equals the number of kernel functions which are the number n_{ker} of basis functions representing $f(\cdot|\boldsymbol{\theta})$ ($n_{ker} = 3$, in the simulations). Therefore, the overall complexity of SURE-GAMP is dominated by the matrix-vector multiplications in lines 3 and 10 of Algorithm 1, whose order is $\mathcal{O}((N_r N_t)^2)$. The EM steps as shown in (14) and (15) are combined with the SURE-GAMP algorithm to avoid the need of specifying a prior probability on $\mathbf{z}^{l+1}$. The algorithm converges after a few iterations when the solution close to minimum MSE is achieved.

IV. Simulation Results

This section shows the performance results obtained for the proposed EM-SURE-GAMP algorithm and the comparison is made with the EM-GAMP solution. Reference [31] suggests the computation of the minimum MSE of the estimate; combined with EM steps we can plot the MSE results of EM-GAMP algorithm to compare with the proposed solution. Following the condition $N_t^{rf} \leq N_t$ and $N_r^{rf} \leq N_r$ for a hybrid analog-digital MIMO architecture, we consider a simple case of $N_t = 8$, $N_r = 8$, and the number of RF chains and streams equal to the number of antennas, i.e., $N_t^{rf} = N_r^{rf} = N_s = 8$. It provides us easier computation for the analog precoder and combiner matrices. We can also consider fewer RF chains and streams than the number of antennas [32] to observe the channel estimation performance plots. The number of multipaths is 5 and due to low overload probability, the value of κ used in the quantization (see Section II) is 4. We run the proposed algorithm for $T_{max} = 1$ and 100 EM iterations. The performance results are obtained for 100 Monte-Carlo realizations each.

Fig. 2 shows the mean square error (MSE) variations with respect to (w.r.t.) the SNR when comparing the proposed EM-SURE-GAMP algorithm with EM-GAMP for 1-bit, 2-bits, and 3-bits resolution ADCs. We can observe that the proposed algorithm achieves better MSE performance for both low and high SNR regimes. For example at an SNR of 10 dB, the SURE algorithm variant outperforms EM-GAMP by about 3 dB in MSE terms for 1-bit quantization. For 2- and 3-bits, the MSE gain is around 2 dB.

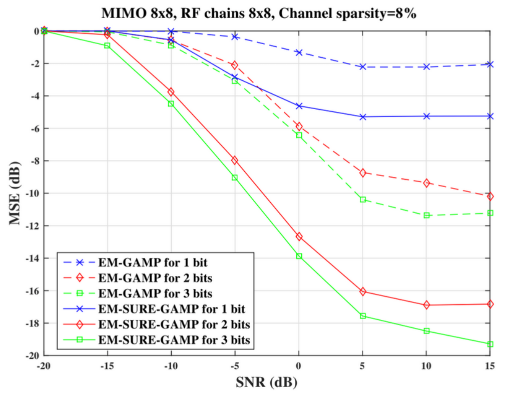


Fig. 2: MSE versus SNR.

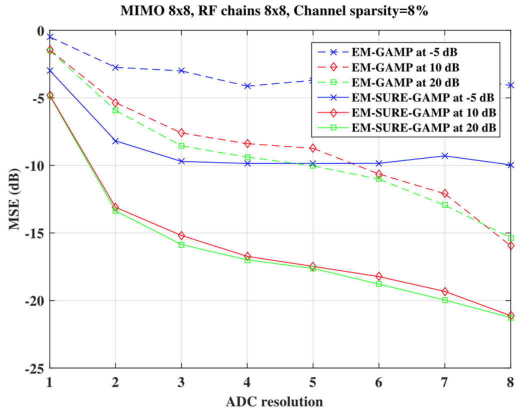


Fig. 3: MSE versus the number of ADC bits.

Fig. 3 again shows that EM-SURE-GAMP performs better than EM-GAMP when MSE is plotted against the number of quantization bits for different values of SNR such as -5 dB, 10 dB, and 20 dB. The training length for Fig. 2 and Fig. 3 is $T = 2^{11}$, and EM-SURE-GAMP exhibits good performance for a channel sparsity level, i.e., ratio of non-zero entries of the beamspace channel and $N_r \times N_t$, of 8%. It can be seen for example that with 3 bits resolution, a significant gain in MSE for the SURE variant of around 6-7 dB compared to EM-GAMP is observed for all SNR values.

Fig. 4 exhibits that the EM-SURE-GAMP solution outperforms EM-GAMP solution w.r.t. the training length for a range of training sequence lengths of 64 to 2048 and converges more quickly than EM-GAMP for a channel sparsity level of 8%, 15 dB SNR, when 1-bit, 2-bits, and 3-bits ADC resolutions are considered.

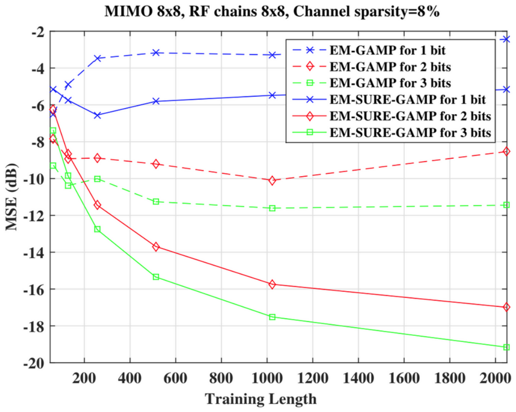


Fig. 4: MSE versus the training length T.

V. Conclusion

This paper proposes an efficient algorithm based on the approximate message passing (AMP) framework to estimate the channel in a mmWave MIMO system with a hybrid analog-digital architecture and low-resolution ADCs at the receiver. EM-SURE-GAMP is exploited to estimate the channel which provides the flexibility to avoid strong assumptions on the channel priors where SURE, depending on the noisy observation, is minimized to adaptively optimize the denoiser within the parametric class at each iteration. When compared with the expectation-maximization generalized AMP (EM-GAMP) solution, the mean square error (MSE) performs better with respect to low and high SNR regimes, the number of ADC bits, and the training length.

References

[1] R. W. Heath et al., “An overview of signal processing techniques for millimeter wave MIMO systems”, *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[2] B. Le et al., “Analog-to-digital converters”, *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 69-77, Nov. 2005.

[3] R. Walden, “Analog-to-digital converter survey and analysis”, *IEEE Journ. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539-550, Apr. 1999.

[4] J. Choi et al., “Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs”, *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005-2018, May. 2016.

[5] S. Jacobsson et al., “One-bit massive MIMO: channel estimation and high-order modulations”, *IEEE Int. Conf. Commun. Workshop*, pp. 1304-1309, June 2015.

[6] A. Alkhateeb et al., “Channel estimation and hybrid precoding for millimeter wave cellular systems”, *IEEE Journ. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.

[7] J. Lee et al., “Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications”, *IEEE Global Commun. Conf.*, pp. 3326-3331, Dec. 2014.

[8] P. Schniter, and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: the sparse way”, *IEEE Asilomar Conf. Sig. Sys. Comp.*, pp. 273-277, Nov. 2014.

[9] J. Mo et al., “Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information”, *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498-5512, Oct. 2015.

[10] C. Rusu et al., “Low resolution adaptive compressed sensing for mmWave MIMO receivers”, *IEEE Asilomar Conf. Sig. Sys. Comp.*, pp. 1138-1143, Nov. 2015.

[11] J. Mo et al., “Channel estimation in millimeter wave MIMO systems with one-bit quantization”, *IEEE Asilomar Conf. Sig. Sys. Comp.*, pp. 957-961, Nov. 2014.

[12] O. Orhan et al., “Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance”, *2015 Info. Theory Appl. Workshop, San Diego, CA*, pp. 191-198, 2015.

[13] L. Fan et al., “Uplink achievable rate for massive MIMO systems with low-resolution ADC”, *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2186-2189, Oct. 2015.

[14] M. T. Ivrlac, and J. A. Nossek, “On MIMO channel estimation with single-bit signal-quantization”, in *Proc. IEEE Smart Antenn. Workshop*, Feb. 2007.

[15] A. Mezghani et al., “Multiple parameter estimation with quantized channel output”, in *Proc. Int. ITG Workshop Smart Antenn.*, pp. 143-150, Oct. 2010.

[16] J. Vila, and P. Schniter, “Expectation-maximization Gaussian-mixture approximate message passing”, *IEEE Trans. Signal Process.*, pp. 4658-4672, May 2014.

[17] S. Rangan et al., “Vector approximate message passing”, *arXiv:1610.03082*, Oct. 2016.

[18] J. Mo et al., “Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs”, in *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141-1154, March 2018.

[19] C. Guo, and M. E. Davies, “Near optimal compressed sensing without priors: parametric SURE approximate message passing”, *IEEE Trans. Signal Process.*, vol. 63, no. 8, Apr. 2015.

[20] S. Singh et al., “Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control”, *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513-1527, Oct. 2011.

[21] C. Balanis, *Antenna Theory*, Wiley, 1997.

[22] J. Brady et al., “Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements”, *IEEE Trans. Antenn. Propag.*, vol. 61, no. 7, pp. 38143827, Jul. 2013.

[23] L. Dai et al., “Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array”, in *2016 IEEE/CIC Int. Conf. Commun. China (ICCC)*, pp. 16, July 2016.

[24] D. L. Donoho, “Compressed sensing”, *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 12891306, Apr. 2006.

[25] R. A. Wannamaker, *The theory of dithered quantization*, National Library Canada, 1997.

[26] O. Dabeer and E. Masry, “Multivariate signal parameter estimation under dependent noise From 1-bit dithered quantized data”, *IEEE Trans. Info. Theory*, vol. 54, no. 4, pp. 1637-1654, April 2008.

[27] U. S. Kamilov et al., “Message-passing de-quantization with applications to compressed sensing”, *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270-6281, 2012.

[28] D. Donoho et al., “Message-passing algorithms for compressed sensing”, *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp.18 914-18 919, 2009.

[29] M. Al-Shoukairi et al., “A GAMP-based low complexity sparse bayesian learning algorithm”, *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 294-308, Jan. 2018.

[30] J. Ma and L. Ping, “Orthogonal AMP”, *IEEE Access*, vol. 5, pp. 2020-2033, 2017.

[31] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing”, *IEEE Int. Symp. Info. Theory Proc.*, pp. 2168-2172, 2011.

[32] A. Kaushik et al., “Sparse hybrid precoding and combining in millimeter wave MIMO systems”, *Radio Propag. Tech. 5G, Durham, UK*, pp. 1-7, Oct. 2016.

Sparse Hybrid Precoding and Combining in Millimeter Wave MIMO Systems

Aryan Kaushik*, **John Thompson***, **Mehrdad Yaghoobi***
*Institute for Digital Communications,
The University of Edinburgh, United Kingdom.
Email: *{A.Kaushik, J.S.Thompson, M.Yaghoobi-Vaighan}@ed.ac.uk

Keywords: Hybrid precoder, energy efficiency, spectral efficiency, millimeter wave (mmWave), multiple-input multiple-output (MIMO).

Abstract

Millimeter wave (mmWave) communication allows us to exploit a new spectrum band between 30 GHz to 300 GHz to meet the growing demands of capacity for fifth generation (5G) wireless communication systems. Multiple-input multiple-output (MIMO) antennas can be used to tackle higher path loss and attenuation at mmWave frequencies compared to microwave bands. Beamforming, called precoding at the transmitter, is performed digitally in conventional microwave frequency MIMO systems, but at mmWave frequencies the higher cost and power consumption of system components means that the system cannot implement one radio frequency (RF) chain per antenna. To enable spatial multiplexing, hybrid precoders using fewer RF chains than antennas emerge as cost-effective and power saving alternative for the transceiver architecture of mmWave MIMO systems. This paper demonstrates the hybrid precoder design with its spectral efficiency and energy efficiency characteristics, and we compare the performance with that of optimal digital precoding (with one RF chain per antenna) and simplified beam steering systems. It also includes two different algorithmic solutions to meet the optimization objective. The orthogonal matching pursuit (OMP) algorithm appears to provide high performance solution to the problem, whereas the gradient pursuit (GP) algorithm is proposed as a cost-effective and fast approximation solution that can still provide equally high performance.

I. Introduction

To advance the state of present wireless communication systems, researchers are primarily concerned about the evolution of fifth generation (5G) networks and even beyond. It is suggested that initial 5G standards may be introduced by 2020 [1]. Such advanced systems systems demand lower latency, lower infrastructure costs, ultra-high reliability, higher mobility, improved range, much higher throughput, and increased capacity of networks [2,3]. The main differences of 5G systems compared to fourth generation (4G) systems will be the use of much greater spectrum allocations, higher aggregate capacity, much higher bit rates, longer battery life, and higher reliability to support many simultaneous users in both licensed and unlicensed RF bands [4]. The emerging advanced consumer devices and developed communication systems have resulted in ever-increasing demands on bandwidth and capacity [5]. The current carrier frequency spectrum has been limited to the very crowded range between 700 MHz and 2.6 GHz leading to the worldwide need for more spectrum and higher capacity. In such scenario, millimeter Wave (mmWave) appears to be a promising technology for future wireless communication systems [4,5]. Utilizing the unused wireless spectrum at much higher frequencies makes mmWave technology different from existing wireless solutions. MmWave offers larger bandwidth channels resulting in much higher data rates, thus supporting much better internet-based access and higher connectivity [4]. MmWave spectrum is currently used for various applications such as satellite communication, radio applications, and backhaul networks. MmWave technology is already a very significant technology for wireless backhaul [6] along with the possibility of self-backhaul in cellular systems. However, mmWave cellular systems do hold certain challenges such as supporting directional communication, susceptibility to shadowing, intermittent connectivity, and processing power consumption by data converters [7].

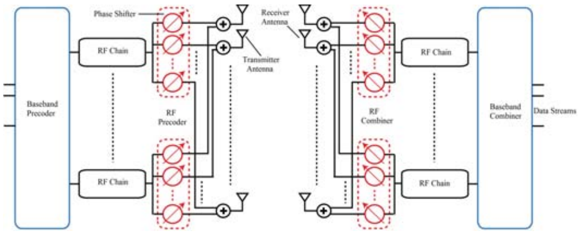


Fig. 1. Hardware block diagram of mmWave single-user fully-connected hybrid beamforming system.

MmWave technology fits very well with multiple-input multiple-output (MIMO) systems as the size of antenna arrays and associated electronics will reduce due to the shorter wavelengths [8]. MIMO technology has already been applied to commercial wireless local area networks and cellular systems at sub-6GHz frequencies. MIMO techniques at mmWave frequencies will be applied differently than at microwave frequencies due to changes in RF propagation and additional hardware constraints. Signal processing for mmWave MIMO systems is of critical importance. At lower frequencies, the signal processing actions are carried out at baseband leading to entirely digital signal processing solutions. While at higher frequencies, there are various hardware constraints making it difficult to have a separate radio frequency (RF) chain dedicated to each antenna. Moreover, the practical implementations of system entities such as RF chains, power amplifiers, low noise amplifiers, and baseband connections are more difficult to construct at mmWave [9], and power consumption is a major issue as these entities become power hungry devices [10]. MmWave frequency systems will exploit polarization and spatial processing techniques such as very directional adaptive beamforming to improve the performance of the system. Deploying a large number of antennas results in high beamforming gain, forming directional beam patterns between transmitter and receiver, which further can assist in overcoming the higher path loss experienced at mmWave frequencies. One of the objectives of this paper is to focus on the sparse nature of the mmWave channel which allows us to use signal processing to enhance performance of mmWave systems towards ultimate performance limits.

One of the simplest approaches to apply MIMO in mmWave systems is analog beamforming which can be implemented at both transmitter and receiver. This approach often connects antenna elements via phase shifters to a single RF chain which supports single stream communication only and does not provide spatial multiplexing gains. Hybrid beamforming can be implemented instead to enable spatial multiplexing and multi-user MIMO communication. Fig. 1 shows the basic structure of a mmWave single-user fully-connected hybrid beamforming system [11] with digital baseband precoding followed by constrained RF precoding implemented using RF phase shifters. The same number of phase shifters as antennas are connected to each RF chain which leads to a fully-connected architecture. Precoding generally refers to beamforming at the transmitter, which may be generalized to support multi-stream (or multi-layer) transmission. At the receiver end, signal combining techniques can be used. One may find the unique advantage associated with hybrid precoding is that, to approach the performance of unconstrained solutions, the digital precoder can correct analog limitations such as cancelling residual multi-stream interference. Although hybrid precoding currently makes compromise on power consumption and hardware complexity yet there is much scope to exploit energy and capacity efficient designs.

Reference [11] proposes a fully-connected hybrid precoder design which leads to a capacity efficient mmWave MIMO system. For an energy efficient design, [12] considers sub-connected architecture, where each RF chain is connected to only a subset of transmitter antennas requiring fewer phase shifters in comparison to the fully-connected architecture. This energy efficient hybrid precoding design is based on successive interference cancellation (SIC) providing near-optimal performance and proposing a low complexity algorithmic solution. Reference [13] considers both fully-connected and partially-connected structures to design a hybrid precoder. The fully-connected structure seems to outperform partially-connected structure in terms of capacity whereas the latter shows higher energy efficiency. In [14] an energy efficient optimization to design the hybrid precoder through the use of optimal number of RF chains is proposed. More generally [15] provides a overview on the relationship between energy efficiency and spectral efficiency for different configurations of a hybrid beamforming system.

This paper mainly exhibits spectral efficiency and energy efficiency characteristics of a hybrid precoder which are helpful in analyzing the throughput and energy variations with respect to the system parameters and the channel parameters. The simulation results are plotted with respect to signal-to-noise ratio (SNR) and the number of RF chains. The solution to the optimization problem implements orthogonal matching pursuit (OMP) at the transmitter and the receiver which appears to be a low complexity solution. Gradient Pursuit (GP) method is introduced as a novel solution to the optimization objective which has the same performance as OMP yet it is a cost-effective and fast approximation solution. The performance and run time comparisons between both the algorithmic solutions are performed and GP is implemented to plot the spectral efficiency and energy efficiency characteristics.

The following notations have been used throughout the paper: $\mathbf{A}$, $\mathbf{a}$, and a stand for a matrix, a vector, and a scalar, respectively; $\mathbf{A}^{(i)}$ represents the i^{th} column of $\mathbf{A}$; transpose and conjugate transpose of $\mathbf{A}$ are denoted as $\mathbf{A}^T$ and $\mathbf{A}^*$, respectively; $||\mathbf{A}||_F$, tr($\mathbf{A}$), and det ($\mathbf{A}$) represent the Frobenius norm, trace, and determinant of $\mathbf{A}$, respectively; $||\mathbf{a}||_p$ is the p-norm of $\mathbf{a}$; $[\mathbf{A}|\mathbf{B}]$ denotes horizontal concatenation; diag($\mathbf{A}$) generates a vector by the diagonal elements of $\mathbf{A}$; $\mathbf{I}_N$ and $\mathbf{0}_{X\times Y}$ represent $N \times N$ identity matrix and $X \times Y$ all-zeros matrix, respectively; $\mathcal{CN}(\mathbf{a};\mathbf{A})$ denotes a complex Gaussian vector having mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$, and i.i.d. shows that the entries of that vector are independent and identically distributed. The expectation and real part of a complex variable are denoted as $\mathcal{E}[.]$ and $\Re\{.\}$, respectively.

II. System and Channel Models

This section presents the mmWave system model and channel model used in this paper.

A. System Model

Considering a single-user mmWave system with N_t antennas at the transmitter end, sending N_s data streams to N_r receiver antennas. N_t^{rf} and N_r^{rf} denote the number of RF chains at the transmitter with the limitation $N_s \leq N_t^{rf} \leq N_t$ and at the receiver with the limitation $N_s \leq N_r^{rf} \leq N_r$, respectively. In other words, in massive MIMO communication systems, based on the function of the RF chains and the hybrid precoding scheme, the number of RF chains is larger than or equal to the number of baseband data streams and smaller than or equal to number of the transmitter antennas. The matrices $\mathbf{F}_{bb}$ and $\mathbf{F}_{rf}$ denote the $N_t^{rf} \times N_s$ baseband precoder and the $N_t \times N_t^{rf}$ RF precoder, respectively. Similarly at the receiver end, the matrices $\mathbf{W}_{bb}$ and $\mathbf{W}_{rf}$ denote the $N_r^{rf} \times N_s$ baseband combiner and the $N_r \times N_r^{rf}$ RF combiner, respectively. Fig. 1 shows the system setup. The signal, $\mathbf{x} = \mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{s}$, is transmitted where $\mathbf{s}$ is the $N_s \times 1$ symbol vector such that $\mathcal{E}[\mathbf{s}\mathbf{s}^*] = \frac{1}{N_s}\mathbf{I}_{N_s}$. All elements of $\mathbf{F}_{rf}$ and $\mathbf{W}_{rf}$ are constrained to have equal norm. The power constraint at the transmitter end is satisfied by $||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F^2 = N_s$. Considering a narrowband block-fading propagation channel with $\mathbf{H}$ as $N_r \times N_t$ channel matrix, which is assumed to be known to both the transmitter and the receiver, a discrete-time model for the received signal is

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the $N_r \times 1$ received vector, ρ is the average received power, and $\mathbf{n}$ is a noise vector with entries which are i.i.d. $\mathcal{CN}(0, \sigma_n^2)$. After combining processing, the processed received signal can be written as follows:

$$\tilde{\mathbf{y}} = \sqrt{\rho}\mathbf{W}_{bb}^*\mathbf{W}_{rf}^*\mathbf{H}\mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{s} + \mathbf{W}_{bb}^*\mathbf{W}_{rf}^*\mathbf{n}, \tag{2}$$

For transmitted symbols following a Gaussian distribution, the achievable spectral efficiency can be expressed as follows:

$$R = \log_2 \det\{\mathbf{I}_{N_s} + \frac{\rho}{N_s}\mathbf{R}_n^{-1}\mathbf{W}_{bb}^*\mathbf{W}_{rf}^*\mathbf{H}\mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{F}_{bb}^*\mathbf{F}_{rf}^*\mathbf{H}^*\mathbf{W}_{rf}\mathbf{W}_{bb}\}, \tag{3}$$

where $\mathbf{R}_n = \sigma_n^2\mathbf{W}_{bb}^*\mathbf{W}_{rf}^*\mathbf{W}_{rf}\mathbf{W}_{bb}$ represents the noise covariance matrix after the combining processing.

B. **Channel Model**

The fading channel models used in traditional MIMO becomes inaccurate for mmWave channel modeling due to the high free-space path loss and large tightly-packed antenna arrays. So the mmWave propagation environment can be characterized by a narrowband clustered channel model, such as the Saleh-Valenzuela model [10]. For N_{cl} clusters and N_{ray} propagation paths each cluster, mmWave channel matrix can be depicted as follows:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl}N_{ray}}}\sum_{i=1}^{N_{cl}}\sum_{l=1}^{N_{ray}}\alpha_{il}\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)^*, \tag{4}$$

where α_{il} denotes the gain of l^{th} ray in i^{th} cluster and it is assumed that α_{il} are i.i.d. $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$, where $\sigma_{\alpha,i}^2$ is average power of the i^{th} cluster such that $\sum_{i=1}^{N_{cl}}\sigma_{\alpha,i}^2 = \gamma$, γ being the normalization factor satisfying $\mathcal{E}[||\mathbf{H}||_F^2] = N_tN_r$, and $\gamma = \sqrt{\frac{N_tN_r}{N_{cl}N_{ray}}}$. Further, $\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ represent the normalized receive and transmit array response vectors, where ϕ_{il}^t and θ_{il}^t are azimuth and elevation angles of departure, respectively, and ϕ_{il}^r and θ_{il}^r are azimuth and elevation angles of arrival, respectively. The antenna elements at the transmitter and the receiver can be modeled as ideal sectored elements [16] and then antenna element gains can be evaluated over the ideal sectors. In (4), the transmit and receive antenna element gains are considered unity over ideal sectors defined by $\phi_{il}^t \in [\phi_{min}^t, \phi_{max}^t]$ and $\theta_{il}^t \in [\theta_{min}^t, \theta_{max}^t]$; $\phi_{il}^r \in [\phi_{min}^r, \phi_{max}^r]$ and $\theta_{il}^r \in [\theta_{min}^r, \theta_{max}^r]$, respectively, and the gains are zero otherwise. This paper considers uniform linear array (ULA) antenna elements for simulations, where for a N_z-element ULA on z-axis, the array response vector can be expressed as follows [17]:

$$\mathbf{a}_z(\phi) = \frac{1}{\sqrt{N_z}}[e^{jm\frac{2\pi}{\lambda}d(\sin(\phi))}]^T, \tag{5}$$

where $0 \leq m \leq (N_z - 1)$ is a real integer counting through antennas, d is inter-element spacing, and λ is the signal wavelength. The array response vectors could also be computed considering a uniform planar array (UPA) of antenna elements in a two-dimensional plane [17].

III. **Hybrid Precoder Design**

It is usually difficult to find a global optimization solution for the joint optimization problem over transmitter and receiver precoders [18]. So, the design can be split into two sub-optimization problems, i.e, one focusing on designing $\mathbf{F}_{rf}\mathbf{F}_{bb}$ for the precoder and the other on designing $\mathbf{W}_{rf}\mathbf{W}_{bb}$ for the combiner. The mutual information obtained through Gaussian signaling over the channel is computed for the hybrid precoder $\mathbf{F}_{rf}\mathbf{F}_{bb}$, measuring the mutual dependence between the two matrices, as follows [11]:

$$\mathcal{I}(\mathbf{F}_{rf}, \mathbf{F}_{bb}) = \log_2 \det(\mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\mathbf{H}\mathbf{F}_{rf}\mathbf{F}_{bb}\mathbf{F}_{bb}^*\mathbf{F}_{rf}^*\mathbf{H}^*) \ , \tag{6}$$

While designing hybrid precoders and combiners for mmWave MIMO systems, we are very much concerned about hardware complexity, spectral efficiency, and energy consumption for baseband processing and analog processing entities such as analog-to-digital converters (ADCs), digital-to-analog converters (DACs), RF chains, phase shifters, and power amplifiers. Sparing use of these entities can lead the system to operate in a very energy efficient manner. For instance, as the number of RF chains increase, more energy would get consumed leading to a decrease in energy efficiency. Measuring the energy efficiency characteristics with respect to the number of RF chains, as shown in Section IV, is quite helpful to design a energy efficient hybrid beamforming system. Meanwhile, the hybrid precoder optimization problem can be formulated as follows:

$$\begin{aligned}(\mathbf{F}_{rf}^{opt}, \mathbf{F}_{bb}^{opt}) = \max_{\mathbf{F}_{rf},\mathbf{F}_{bb}} &\mathcal{I}(\mathbf{F}_{rf}, \mathbf{F}_{bb}),\\ \text{s.t. } &\mathbf{F}_{rf} \in \mathcal{F}_{rf},\\ &||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F^2 = N_s,\end{aligned} \tag{7}$$

where $\mathcal{F}_{rf}$ denotes the set of $N_t \times N_t^{rf}$ matrices having elements of constant magnitude. For such a non-convex constraint, it is difficult to yield general solutions to the problem. So in order to design the near-optimal hybrid precoder, certain assumptions and approximations can be exploited as in [11] to simplify the above problem. Equation (7) can be transformed in terms of the Euclidean distance between $\mathbf{F}_{rf}\mathbf{F}_{bb}$ and the channel's optimal fully digital precoder $\mathbf{F}_{opt}$. The hybrid precoder $\mathbf{F}_{rf}\mathbf{F}_{bb}$ can be located in a constrained space to be as close as possible to the optimal matrix $\mathbf{F}_{opt}$ in the unconstrained space. So the Euclidean distance $||\mathbf{F}_{opt} - \mathbf{F}_{rf}\mathbf{F}_{bb}||_F$ should be as small as possible for maximum throughput. We compute the channel's singular value decomposition (SVD) as $\mathbf{H} = \mathbf{U_H}\Lambda_\mathbf{H}\mathbf{V_H^*}$, where $\mathbf{U_H} \in \mathcal{C}^{N_r \times N_r}$ and $\mathbf{V_H} \in \mathcal{C}^{N_t \times N_t}$ are unitary matrices, and $\Lambda_\mathbf{H} \in \Re^{N_r \times N_t}$ is a rectangular matrix of singular values in decreasing order whose diagonal elements are non-negative real numbers and whose non-diagonal elements are zero. The optimal matrix $\mathbf{F}_{opt}$ is comprised of the first N_s columns of $\mathbf{V_H}$. As the array response vectors $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ are constant-magnitude phase-only vectors and $\mathcal{F}_{rf}$ denotes the set of $N_t \times N_t^{rf}$ matrices having elements of constant magnitude, we can restrict $\mathcal{F}_{rf}$ to be a set of basis vectors $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ in order to find the best low dimensional representation of the optimal matrix $\mathbf{F}_{opt}$. So the hybrid precoder optimization problem can further be stated as follows:

$$\begin{aligned}(\mathbf{F}_{rf}^{opt},\mathbf{F}_{bb}^{opt}) = \min_{\mathbf{F}_{rf},\mathbf{F}_{bb}} & ||\mathbf{F}_{opt}-\mathbf{F}_{rf}\mathbf{F}_{bb}||_F, \\ \text{s.t. } & \mathbf{F}_{rf}^{(i)} \in \{\mathbf{a}_t(\phi_{il}^t,\theta_{il}^t), \forall i, l\}, \\ & ||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F^2 = N_s,\end{aligned} \tag{8}$$

One may note here that the constraint on $\mathbf{F}_{rf}^{(i)}$ may be added into the optimization given (8) to obtain the problem as follows:

$$\begin{aligned}\tilde{\mathbf{F}}_{bb}^{opt} = & \min_{\tilde{\mathbf{F}}_{bb}} ||\mathbf{F}_{opt}-\mathbf{A}_t\tilde{\mathbf{F}}_{bb}||_F, \\ \text{s.t. } & ||\text{diag}(\tilde{\mathbf{F}}_{bb}\tilde{\mathbf{F}}_{bb}^*)||_0 = N_t^{rf}, \\ & ||\mathbf{A}_t\tilde{\mathbf{F}}_{bb}||_F^2 = N_s,\end{aligned} \tag{9}$$

where $\mathbf{A}_t$ is an $N_t \times N_{cl}N_{ray}$ matrix consisting of array response vectors and $\tilde{\mathbf{F}}_{bb}$ is an $N_{cl}N_{ray} \times N_s$ matrix. The matrices $\mathbf{A}_t$ and $\tilde{\mathbf{F}}_{bb}$ help to obtain $\mathbf{F}_{rf}^{opt}$ and $\mathbf{F}_{bb}^{opt}$ as the N_t^{rf} non-zero rows of $\tilde{\mathbf{F}}_{bb}$ will give us the baseband precoder matrix $\mathbf{F}_{bb}^{opt}$ and the corresponding N_t^{rf} columns of $\mathbf{A}_t$ will provide the RF precoder matrix $\mathbf{F}_{rf}^{opt}$. Equation (9) basically reformulates (8) into a sparsity constrained reconstruction problem with one variable. The problem can now be addressed as a sparse approximation problem [19], and orthogonal matching pursuit (OMP) [20] can be used as an algorithmic solution to this problem. The receiver side follows a problem definition, optimization objective, and the same algorithmic solution can be used with minimal changes. As the hybrid combiner design has a similar mathematical formulation except for the extra transmitter power constraint at the transmitter, this paper mainly focuses on hybrid precoder design and the hybrid combiner design has been omitted. One may note here that by assuming the hybrid precoders $\mathbf{F}_{rf}\mathbf{F}_{bb}$ to be fixed, the hybrid combiners $\mathbf{W}_{rf}\mathbf{W}_{bb}$ can be designed in order to minimize the mean-squared-error (MSE) between the transmitted and processed received signals by using the linear minimum mean-square error (MMSE) receiver.

Algorithm 1: Hybrid Precoder Design through Orthogonal Matching Pursuit (OMP) [20]

```
Require: $\mathbf{F}_{opt}$
1: $\mathbf{F}_{rf} = \emptyset$
2: $\mathbf{F}_{res} = \mathbf{F}_{opt}$
3: for $i \leq N_t^{rf}$
4: $\mathbf{\Psi} = \mathbf{A}_t^*\mathbf{F}_{res}$
5: $k = \text{arg max}_{l=1,\ldots,N_{cl}N_{ray}}(\mathbf{\Psi}\mathbf{\Psi}^*)_{l,l}$
6: $\mathbf{F}_{rf} = \left[\mathbf{F}_{rf} \mid \mathbf{A}_t^{(k)}\right]$
7: $\mathbf{F}_{bb} = (\mathbf{F}_{rf}^*\mathbf{F}_{rf})^{-1}\mathbf{F}_{rf}^*\mathbf{F}_{opt}$
8: $\mathbf{F}_{res} = \frac{\mathbf{F}_{opt}-\mathbf{F}_{rf}\mathbf{F}_{bb}}{||\mathbf{F}_{opt}-\mathbf{F}_{rf}\mathbf{F}_{bb}||_F}$
9: end for
10: $\mathbf{F}_{bb} = \sqrt{N_s}\frac{\mathbf{F}_{bb}}{||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F}$
11: return $\mathbf{F}_{rf}, \mathbf{F}_{bb}$
```

Algorithm 1 starts by finding the array response vector $\mathbf{a}_t(\phi_{il}^t,\theta_{il}^t)$ along which the optimal precoder has the maximum projection, and then concatenates that selected column vector into the RF precoder $\mathbf{F}_{rf}$ as shown in step 6. It then continues to find least squares solution to the baseband precoder $\mathbf{F}_{bb}$, and then the residual precoding matrix $\mathbf{F}_{res}$ is computed in order to remove the contribution of the selected vector. Then the algorithm continues to find the column along which $\mathbf{F}_{res}$ has the largest projection until all RF chains have been used. The transmit power constraint is satisfied at step 10, which is applicable for a general case of $N_s \geq 1$.

To develop fast approximate OMP algorithms that require less storage, [21] proposes improvements to greedy strategies using directional pursuit methods, and discusses optimization schemes on the basis of gradient, conjugate gradient, and approximate conjugate gradient approaches. The gradient pursuit (GP) method is introduced as a novel solution to the optimization objective exhibiting the same performance as OMP, cheaper cost consumption, and faster processing time. Unlike OMP where optimum signal approximation is achieved on all the selected atoms, GP makes use of a single gradient direction for the approximation avoiding the need to consider all the atoms and hence leading to reduced computation time. The computation time is considerably less for large MIMO configurations when implementing GP, as shown in section IV. Algorithm 2 starts in the same way as Algorithm 1. There is a index set which is updated at each iteration as shown in step 6 which is used to generate baseband precoder matrix $\mathbf{F}_{bb}$. The gradient direction, as mentioned in step 8, is computed at each iteration and the step-size is determined explicitly making use of the gradient direction, as shown in step 10. Finally the RF precoder matrix $\mathbf{F}_{rf}$ and the baseband precoder matrix $\mathbf{F}_{bb}$ are obtained at the end of the algorithm. The transmit power constraint is satisfied at step 14.

Algorithm 2: Hybrid Precoder Design through Gradient Pursuit (GP) [21]

```
Require: $\mathbf{F}_{opt}$
1: $\mathbf{F}_{rf} = \emptyset$, $\Gamma = \emptyset$
2: $\mathbf{F}_{res} = \mathbf{F}_{opt}$, $\mathbf{F}_{bb} = 0$
3: for $i \leq N_t^{rf}$
4: $\mathbf{\Psi} = \mathbf{A}_t^*\mathbf{F}_{res}$
5: $k = \text{arg max}_{l=1,\ldots,N_{cl}N_{ray}}(\mathbf{\Psi}\mathbf{\Psi}^*)_{l,l}$
6: $\Gamma = \Gamma \cup k$
7: $\mathbf{F}_{rf} = \left[\mathbf{F}_{rf} \mid \mathbf{A}_t^{(k)}\right]$
8: $\mathbf{D} = \mathbf{F}_{rf}^*\mathbf{F}_{res}$
9: $\mathbf{C} = \mathbf{F}_{rf}\mathbf{D}$
10: $g = \frac{\text{tr}\{\mathbf{F}_{res}^*\mathbf{C}\}}{||\mathbf{C}||_F^2}$
11: $\mathbf{F}_{bb}|_\Gamma = \mathbf{F}_{bb}|_\Gamma - g\mathbf{D}$
12: $\mathbf{F}_{res} = \mathbf{F}_{res} - g\mathbf{C}$
13: end for
14: $\mathbf{F}_{bb} = \sqrt{N_s}\frac{\mathbf{F}_{bb}}{||\mathbf{F}_{rf}\mathbf{F}_{bb}||_F}$
15: return $\mathbf{F}_{rf}, \mathbf{F}_{bb}$
```

For the fully connected hybrid precoder design, it is quite interesting to observe the energy performance. Reference [15] suggests that energy efficiency ε can be defined as the ratio between spectral efficiency R and total power consumption P_{tot} as shown in (10). The total power consumption is the sum of power consumed for transmission, and baseband processing and analog processing entities.

$$\varepsilon = \frac{R}{P_{tot}} = \frac{R}{P_{cp} + N_t^{rf} P_{rf} + N_{ps}(P_{ps} + P_{pa})} bits/Hz/J, \quad (10)$$

where $N_{ps}, P_{cp}, P_{rf}, P_{ps}$, and P_{pa} represent the number of phase shifters, the common power of transmitter, the power per RF chain, the power per phase shifter, and the power per power amplifier. The energy consumed by the RF chains is a major concern leading to high value of P_{rf} with substantial increase in each RF chain. In a fully-connected hybrid precoder structure, one can consider that N_{ps} is equal to $N_t^{rf} N_t$ [12,13].

IV. Simulation Results

This section demonstrates the spectral efficiency and energy efficiency characteristics of the hybrid precoder design. For observation, there are 10 rays for each cluster and there are 8 clusters in total, i.e., N_{ray} = 10 and N_{cl} = 8. The average power of each cluster is unity, i.e., $\sigma_{\alpha,i}$ = 1. The azimuth and elevation angles of departure and arrival are computed on the basis of a Laplacian distribution with uniformly distributed mean angles within the range of 60° to 120° in the azimuth domain, and 80° to 100° in the elevation domain. The angle spread which is the standard deviation of the Laplacian distribution of the angles is set to be 7.5°. The antenna elements in the ULA are spaced by half wavelength distance. The symbol vector **s** is generated using quadrature amplitude modulation (QAM) scheme. The signal-to-noise ratio (SNR) is determined as $\frac{\rho}{\sigma_n^2}$ for the plots. All the simulation results are averaged over 5000 random channel realizations.

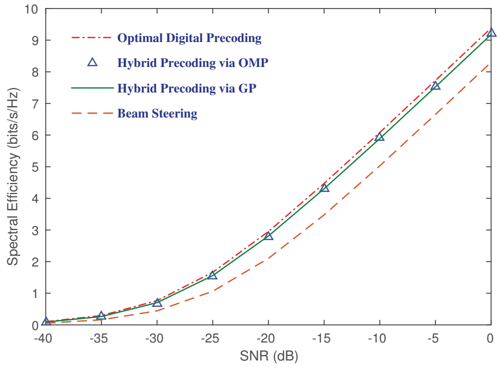


Fig. 2. Spectral efficiency for several precoding solutions for 64 × 16 fully-connected mmWave system with N_s = 1, N_{cl} = 8, and N_{ray} = 10.

Fig. 2 shows the spectral efficiency versus SNR plot for several precoding solutions. For a single-user 64×16 mmWave system with a single stream being transmitted and received, the parameters are set in such a way that the hybrid precoder $\mathbf{F}_{rf}\mathbf{F}_{bb}$ can be made sufficiently close to the optimal precoder $\mathbf{F}_{opt}$. The optimal digital precoder uses N_t RF chains at the transmitter and N_r RF chains at the receiver, while beam steering [22] uses only a single RF chain both at the transmitter and at the receiver ends. Hybrid precoding implements 4 RF chains both at the transmitter and the receiver, i.e., N_t^{rf} = N_r^{rf} = 4. Both OMP and GP algorithmic solutions have been implemented for the hybrid precoder design. It can be observed that hybrid precoding performs slightly worse than optimal digital precoding but it is clearly better than beam steering. Moreover, the hybrid precoder using GP shows the same performance characteristics as that for OMP. GP provides a fast approximation solution as it requires less run time than OMP, which provides us a novel cost-effective solution to design the hybrid precoders.

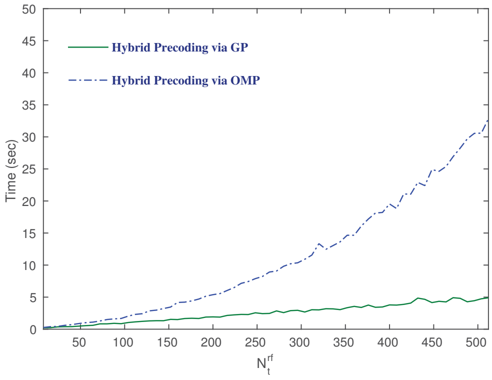


Fig. 3. Time evaluation with respect to number of RF chains for OMP and GP for 512×512 mmWave system with N_{cl}=12, N_{ray} = 20, N_s = 8 and SNR = −25 dB

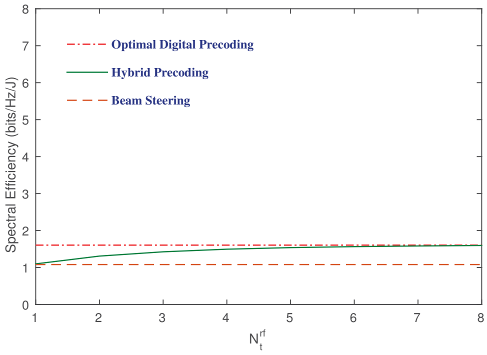


Fig. 4. Spectral efficiency for several fully-connected precoder designs while SNR = −25 dB.

The run time for GP is less than that of OMP for both small and large MIMO configurations. Fig. 3 shows the run time characteristics with respect to the number of RF chains for both GP and OMP for a large 512×512 mmWave system with N_{cl} = 12, N_{ray} = 20, N_s = 8, and SNR = −25dB. The time difference between both the algorithmic solutions is considerable which shows that GP is a better practical solution and more efficient than OMP to design a hybrid precoder. As GP has the same performance but less run time, the rest of

Common power of transmitter	$P_{cp} = 10$ W
Power per RF chain	$P_{rf} = 100$ mW
Power per phase shifter	$P_{ps} = 10$ mW
Power per power amplifier	$P_{pa} = 300$ mW

TABLE I. SIMULATION PARAMETERS FOR THE POWER MODEL [10].

the plots in this paper make use of GP as the algorithmic solution to find the optimum precoder. Fig. 4 plots the spectral efficiency characteristics of the hybrid precoder, the optimal digital precoder, and beam steering system with respect to the number of RF chains at a SNR of -25 dB. It can be observed from Fig. 4 that the spectral efficiency of the hybrid precoder increases gradually and starts approximating the performance of the optimal digital precoder. It also clearly outperforms the beam steering approach in terms of spectral efficiency with increase in number of RF chains for a certain SNR (such as -25 dB).

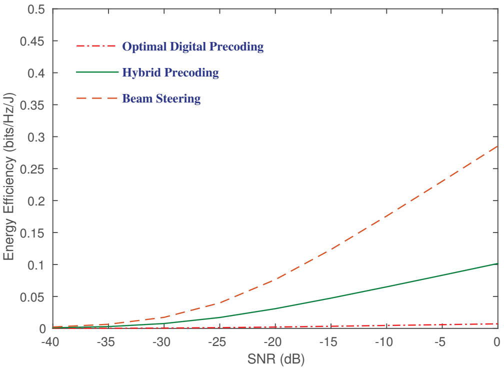


Fig. 5. Energy efficiency for several precoding solutions for 64 × 16 fully-connected mmWave system with N_s= 1, N_{cl} = 8, and N_{ray} = 10.

Fig. 5 shows the energy efficiency versus SNR plot for several precoding solutions. To illustrate the achievable energy efficiency of different precoding solutions, the parameters in (10) are set as as shown in Table I and the other required parameters are same as used to obtain Fig. 2. The energy efficiency performance of the hybrid precoder clearly appears to outperform the optimal digital precoder as the SNR increases. However, the beam steering approach performs better in terms of energy efficiency as only one RF chain is being used in that system which reduces the energy consumption considerably. As N_{ps} is scaled linearly with N_t^{rf} and N_t, the energy consumption will significantly increase with respect to N_t^{rf}. For the same reason, beam steering outperforms hybrid precoding and optimal digital precoding as number of RF chains increases for a certain SNR (such as -25 dB) as shown in Fig. 6. The hybrid precoding performs exactly the same as beam steering in terms of energy efficiency with use of a single RF chain. One should note that, in order to achieve a significant spectral efficiency gain while accepting an increase in the energy consumption, the hybrid precoder solution might be a better approach to follow. For instance, to obtain a gain of 1 bits/s/Hz over the beam steering approach, the hybrid precoder will exhibit 0.11 bits/Hz/J less energy efficiency than beam steering at SNR = -10 dB as observed from Fig. 2 and Fig. 5.

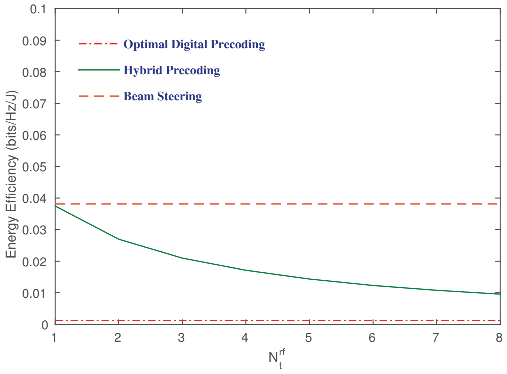


Fig. 6. Energy efficiency for several fully-connected precoder designs while SNR = -25 dB.

V. Conclusion

This paper is focused on evaluating the spectral efficiency and energy efficiency characteristics of a hybrid precoder which help in designing capacity and energy efficient hybrid mmWave communication systems. The spectral efficiency and energy efficiency characteristics of a hybrid precoder are compared with that of optimal digital precoding (with one RF chain per antenna) and simplified beam steering systems. It can be observed that the hybrid precoder design provides near-optimal spectral efficiency, and outperforms the optimal digital precoder significantly in terms of energy efficiency. While compared to the conventional beam steering approach, the hybrid precoder shows notable performance gain in terms of spectral efficiency. However, beam steering outperforms hybrid precoding in terms of energy efficiency with respect to SNR and number of RF chains. The gradient pursuit (GP) method is introduced as a novel algorithmic solution to the optimization objective. The orthogonal matching pursuit (OMP) algorithm appears to provide high performance solution to the problem, whereas the GP algorithm is proposed as a cost-effective and fast approximation solution. GP shows the same performance as OMP but it requires less run time for both small and large MIMO configurations. This research work will be extended to design an energy efficient hybrid precoder with a fully-connected architecture through optimizing the baseband precoder and RF precoder matrices along with optimizing the number of RF chains, and compare the energy performance of the fully optimized hybrid precoder to the hybrid precoder before optimization, the optimal digital precoder, and the simplified beam steering system.

References

[1] Li, X., Gani, A., Salleh, R., and Zakaria, O.: 'The future of mobile wireless communication networks', IEEE Intern. Conf. Commun. Software and Networks, Macau, Feb. 2009, pp. 554-557.

[2] NGMN 5G White Paper, v.1.0, Feb. 2015, pp. 1-125.

[3] 2020 Networld White Paper for Research Beyond 5G, v.1.0, Oct. 2015, pp. 1-43.

[4] Rappaport, T. S., Shu, S., Mayzus, R., Hang, Z., Azar, Y., Wang, K., Wong, G. N., Schulz, J. K., Samimi, M., and Gutierrez: 'Millimeter wave mobile communications for 5G cellular: It will work!', IEEE Access, 2013, 1, pp. 335-349.

[5] Pi, Z., and Khan, F.: 'An introduction to millimeter-wave mobile broadband systems', IEEE Commun. Mag., 2011, 49, (6), pp. 101-107.

[6] Hur, S., Kim, T., Love, D. J., Krogmeier, J. V., Thomas, T. A., and Ghosh, A.: 'Millimeter wave beamforming for wireless backhaul and access in small cell networks', IEEE Trans. Commun., 2013, 61, (10), pp, 4391-4403, .

[7] Rangan, S., Rappaport, T. S., and Erkip, E.: 'Millimeter-wave cellular wireless networks: potentials and challenges', Proc. IEEE, 2014, 102, (3), pp. 366-385.

[8] Biglarbegian, B., Fakharzadeh, M., Busuioc, D., Nezhad-Ahmadi, M.-R., and Safavi-Naeini, S.: 'Optimized microstrip antenna arrays for emerging millimeter-wave wireless applications', IEEE Trans. Antenn. and Propag., 2011, 59, (5), pp. 1742-1747.

[9] Zhang, J. A., Huang, X., Dyadyuk, V., and Guo, Y. J.: 'Massive hybrid antenna array for millimeter-wave cellular communications', IEEE Wireless Commun., 2015, 22, (1), pp. 79-87,

[10] Rappaport, T. S., Heath, R. W., Daniels, R. C., and Murdock, J. N.: 'Millimeter wave wireless communications' (Prentice-Hall, Sept. 2014).

[11] Ayach, O. E., Rajagopal, S., Abu-Surra, S., Pi, Z., and Heath, R. W.: 'Spatially sparse precoding in millimeter wave MIMO systems', IEEE Trans. Wireless Commun., 2014, 13, (3), pp. 1499-1513.

[12] Gao, X., Dai, L., Han, S., I, C.-L., and Heath, R. W.: 'Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays', IEEE J. Sel. Areas Commun., 2016, 34, (4), pp. 1-12.

[13] Yu, X., Shen, J. C., Zhang, J., and Letaief, K. B.: 'Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems', IEEE J. Sel. Topics Signal Process., 2016, 10, (3), pp. 485-500.

[14] Zi, R., Ge, X., Thompson, J., Wang, C.-X., Wang, H., and Han, T.: 'Energy efficiency optimization of 5G radio frequency chain systems', IEEE J. Sel. Areas Commun., 2016, 34, (4), pp. 1-16.

[15] Han, S., I, C.-L, Xu, Z., and Rowell, C.: 'Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G', IEEE Commun. Mag., 2015, 53, (1), pp. 186-194.

[16] Singh, S., Mudumbai, R., and Madhow, U.: 'Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control', IEEE/ACM Trans. Netw., 2011, 19, (5), pp. 1513-1527.

[17] Balanis, C.: 'Antenna Theory' (Wiley, 1997).

[18] Palomar, D. P., Cioffi, J. M., and Lagunas, M. A.: 'Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization', IEEE Trans. Signal Process., 2003, 51, (9), pp. 2381-2401.

[19] Tropp, J. A., Gilbert, A. C., and Strauss, M. J.: 'Algorithms for simultaneous sparse approximation-part I: greedy pursuit", Signal Process., 2006, 86, (3), pp. 572-588.

[20] Tropp, J., and Gilbert, A.: 'Signal recovery from random measurements via orthogonal matching pursuit', IEEE Trans. Info. Theory, 2007, 53, (12), pp. 4655-4666.

[21] Blumensath, T., and Davies, M. E.: 'Gradient pursuits', IEEE Trans. on Signal Process., 2008, 56, (6), pp. 2370-2382.

[22] Ayach, O. E., Heath, R. W., Abu-Surra, S., Rajagopal, S., and Pi, Z.: 'The capacity optimality of beam steering in large millimeter wave MIMO systems', in Proc. 2012 13th IEEE Int. Work. Signal Process. Advances Wireless Commun., Cesme, June 2012, pp. 100-104.