

Copyright  
by  
Junmo Sung  
2020

The Dissertation Committee for Junmo Sung  
certifies that this is the approved version of the following dissertation:

**Compressed-sensing based Channel State Information  
Acquisition in mmWave Hybrid Beamforming  
Communication Systems**

Committee:

Brian L. Evans, Supervisor

Naofal Al-Dhahir

Georgios-Alex Dimakis

Haris Vikalo

Hao Zhu

**Compressed-sensing based Channel State Information  
Acquisition in mmWave Hybrid Beamforming  
Communication Systems**

by

**Junmo Sung**

**DISSERTATION**

Presented to the Faculty of the Graduate School of  
The University of Texas at Austin  
in Partial Fulfillment  
of the Requirements  
for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2020

Dedicated to my mother, my wife and my son

## Acknowledgments

I would like to express my sincerest appreciation to my supervisor Prof. Brian L. Evans for his kind and infinite support not just as a supervisor but also as a great mentor. I would like to extend my gratitude to my dissertation committee: Prof. Naofal Al-Dhahir, Prof. Alexandros G. Dimakis, Prof. Haris Vikalo and Prof. Hao Zhu for their support and comments on this dissertation.

I thank my family: my mother, my wife Wookyung and my son Siwoo for their endless love and faith. I could not have finished this long and tough journey without my wife's dedication and sacrifice. My gratitude and love for her are beyond description.

# Compressed-sensing based Channel State Information Acquisition in mmWave Hybrid Beamforming Communication Systems

Junmo Sung, Ph.D.

The University of Texas at Austin, 2020

Supervisor: Brian L. Evans

Fifth-generation (5G) cellular communications promises 10x higher data rate, 10x reduced latency and high reliability compared against the fourth-generation (4G). The higher data rate is primarily enabled by the use of higher frequency millimeter wave (mmWave) bands. MmWave bands experience high signal power attenuation over distance, which can be overcome by employing large antenna arrays and advanced signal processing techniques to focus radiated power in a beam. Massive number of antennas and accompanying radio frequency (RF) circuitry, however, excessively increase system operating power.

Hybrid analog and digital beamforming (HB) architectures, which can significantly scale down the number of RF transceivers, and low resolution analog-to-digital converters (ADCs) are attractive in reducing power consumption for wireless communication systems with large antenna arrays. However,

reducing power consumption comes at the expense of reducing communication performance. The HB architectures, due to fewer dimensions of the digital beamforming stage and hardware constraints in the analog beamforming stage, suffer fewer degrees of freedom compared with the all digital architecture. Low resolution quantizers inevitably produce higher quantization noise than high resolution quantizers do. Conventional channel state information (CSI) acquisition algorithms employed in all-digital beamforming architectures generally yield degraded performance in such power saving architectures. Therefore I consider compressed sensing techniques to acquire millimeter wave (mmWave) CSI in HB architectures. Compressed sensing is able to exploit the sparsity in angular mmWave channel responses.

In the point-to-point mmWave communications, I develop a deterministic HB codebook design framework for compressed sensing (CS) based channel estimation. The framework is versatile to be applicable to various HB architectures including phase shifters, switches and RF lens. The design approach is to configure analog and digital beamformers in the most favorable forms to CS techniques under the hybrid beamforming constraints. When one tries to reduce measurement overhead of CS-based channel estimation, extra randomness is usually considered: random RF precoder permutation. I propose a computationally efficient algorithm to find a deterministic order of RF precoders that can reduce the overhead down to a half without significant performance loss.

Low-resolution ADC is another means for further power consumption

reduction along with HB architectures. However, the combination of a HB architecture and low-resolution ADCs makes channel estimation in such systems more challenging. Adopting the extremely low resolution, i.e., one-bit ADC, in the HB communication systems, I develop a CS-based channel estimation algorithm that is suitable for one-bit quantization.

In developing 5G NR, a new challenge has been arisen: beam management. Since beamforming became an essential component in 5G, beam search and detection are performed even in the initial access. I investigate CS-based downlink beam detection for mmWave HB systems taking the 3GPP standard into account. With the exhaustive search being a benchmark, the CS approach is evaluated using the random and the discrete Fourier transform (DFT) RF codebooks in terms of the beam detection probability.

Through the research contributions I present in this dissertation, it is shown that compressed sensing is the key to exploit sparsity in angular mmWave channel responses. Compressed sensing is beneficial in not only improving accuracy and reducing latency of CSI acquisition, but also the overall communication performance of the hybrid analog/digital beamforming system.

# Table of Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>List of Tables</b>	<b>xii</b>
<b>List of Figures</b>	<b>xiii</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.1.1 Wireless Communication Systems . . . . .	2
1.1.2 Wireless Channel . . . . .	4
1.1.3 Multi-input Multi-output Communication Systems . . . . .	6
1.1.4 Millimeter Wave Frequency Bands . . . . .	8
1.1.5 Hybrid Analog and Digital Beamforming . . . . .	9
1.1.6 Low-resolution Analog-to-Digital Converter . . . . .	10
1.1.7 Channel Estimation . . . . .	11
1.1.8 Compressed Sensing . . . . .	13
1.2 Motivation . . . . .	14
1.2.1 Compressed Sensing based Channel Estimation . . . . .	14
1.3 Dissertation Summary . . . . .	17
1.3.1 Thesis Statement . . . . .	17
1.3.2 Overview of Contributions . . . . .	18
1.4 Conclusion . . . . .	19
1.5 Notation and Abbreviations . . . . .	20
1.5.1 Notation . . . . .	20
1.5.2 Abbreviations . . . . .	20

<b>Chapter 2.</b>	<b>Versatile Compressive mmWave Hybrid Beamformer Codebook Design Framework</b>	<b>24</b>
2.1	Introduction . . . . .	25
2.1.1	Prior Work . . . . .	26
2.1.2	Contributions . . . . .	30
2.2	System and Channel Models . . . . .	33
2.3	Sparse Formulation . . . . .	36
2.4	Deterministic Sensing Matrix Design . . . . .	37
2.4.1	Total Coherence . . . . .	38
2.4.2	Transmitter Design . . . . .	40
2.4.2.1	Minimization of Off-diagonal Terms . . . . .	40
2.4.2.2	Minimization of Diagonal Terms . . . . .	44
2.4.3	Receiver Design . . . . .	45
2.4.3.1	Minimization of Off-diagonal Terms . . . . .	45
2.4.3.2	Minimization of Diagonal Terms . . . . .	47
2.4.4	Alternative Approach . . . . .	49
2.4.5	Comparison of total coherence . . . . .	51
2.5	Hybrid Architectures . . . . .	55
2.5.1	Phase Shifting Network . . . . .	56
2.5.2	Phase Shifting Network in Subsets . . . . .	56
2.5.3	Switching Network with Analog Precoding . . . . .	58
2.5.4	Switching Network in Subsets with Analog Precoding . . . . .	59
2.5.5	Switching Network . . . . .	60
2.5.6	Switching Network in Subsets . . . . .	60
2.5.7	Continuous Aperture Phased MIMO . . . . .	61
2.5.8	Remarks . . . . .	61
2.6	Numerical Results . . . . .	62
2.6.1	Receive RF Chains . . . . .	63
2.6.2	Granularity of Angle Grids in Dictionary . . . . .	69
2.6.3	Channel Sparsity . . . . .	72
2.7	Conclusion . . . . .	74

<b>Chapter 3. Hybrid Beamformer Codebook Design and Ordering for Compressive mmWave Channel Estimation</b>	<b>76</b>
3.1 Introduction . . . . .	77
3.2 System Model . . . . .	80
3.3 Coherence Minimizing Codebook . . . . .	83
3.4 Precoder Column Permutation and Pilot Design . . . . .	87
3.5 Numerical Results . . . . .	91
3.6 Conclusion . . . . .	96
<b>Chapter 4. Narrowband Channel Estimation for Hybrid Beamforming Millimeter Wave Communication Systems with One-bit ADCs</b>	<b>98</b>
4.1 Introduction . . . . .	99
4.2 System Model . . . . .	102
4.3 Compressed Sensing Channel Estimation . . . . .	105
4.4 Numerical Results . . . . .	107
4.5 Conclusion . . . . .	113
<b>Chapter 5. Compressed-Sensing based Beam Detection in 5G New Radio Initial Access</b>	<b>115</b>
5.1 Introduction . . . . .	116
5.2 System and Channel Models . . . . .	118
5.3 Sparse Formulation . . . . .	122
5.4 Initial Access . . . . .	126
5.5 Numerical Results . . . . .	127
5.6 Conclusion . . . . .	133
<b>Chapter 6. Concluding Remarks</b>	<b>135</b>
6.1 Summary . . . . .	135
6.2 Future Work . . . . .	136
<b>Bibliography</b>	<b>139</b>
<b>Vita</b>	<b>158</b>

## List of Tables

2.1	Comparison among CS-based channel estimation algorithms . . . . .	29
2.2	Brief comparison among non-CS based CE techniques . . . . .	30
2.3	Simulation parameters . . . . .	62
3.1	Mean coherence of random permutation and coherence of the proposed algorithm . . . . .	92
3.2	Simulation parameters . . . . .	93
4.1	Comparison of prior work . . . . .	102
4.2	Comparison among channel estimation algorithms for low-resolution ADCs . . . . .	103
4.3	Simulation parameters . . . . .	110
5.1	Comparison among beam detection algorithms using compressive sensing (CS), exhaustive beam search (ES) and iterative search (IS) . . . . .	117
5.2	Simulation parameters . . . . .	128

## List of Figures

1.1	Features of 5G New Radio . . . . .	1
2.1	General hybrid beamforming architecture block diagram . . .	33
2.2	Brief algorithm description of Section 2.4.2.1 . . . . .	43
2.3	Brief algorithm description of Section 2.4.2.2 . . . . .	44
2.4	Brief algorithm description of Section 2.4.3.1 . . . . .	47
2.5	Brief algorithm description of Section 2.4.3.2 . . . . .	48
2.6	Histograms of the magnitudes of off-diagonal entries . . . . .	53
2.7	Seven types of analog precoders . . . . .	57
2.8	List of legends . . . . .	63
2.9	Evaluated NMSE with combinations of estimation algorithms and hybrid beamforming designs . . . . .	64
2.10	Spectral efficiencies as a function of SNR . . . . .	67
2.11	NMSE and spectral efficiency as a function of the grid size . .	70
2.12	NMSE and spectral efficiency as a function of the number of channel paths . . . . .	73
3.1	Phase shifter based hybrid beamforming architecture . . . . .	79
3.2	Sensing matrix mutual coherence distribution . . . . .	90
3.3	NMSE of random and MTC codebooks as a function of SNR . . .	94
3.4	NMSE as a function of number of frames . . . . .	95
3.5	NMSE as a function of the number of channel paths . . . . .	96
4.1	MIMO hybrid beamforming communication system . . . . .	99
4.2	Magnitude of virtual channel matrix in the angular domain . .	105
4.3	NMSE of different channel estimation algorithms . . . . .	108
4.4	Spectral efficiency of four channel estimation algorithms . . . .	109
4.5	Comparison of true and estimated path gain vectors . . . . .	111
4.6	NMSE as a function of the number of RF chains . . . . .	112

4.7	NMSE as a function of the number of frames . . . . .	113
5.1	General OFDM hybrid beamforming architecture block diagram	118
5.2	A diagram of the phase shifting RF stage of the precoder. . .	119
5.3	Beam detection probability as a function of the transmit SNR.	127
5.4	Empirical CDF of estimated transmit beam detection errors . .	129
5.5	Empirical CDF of estimated receive beam detection errors . .	131
5.6	Beam detection probability as a function of the transmit SNR	132

# Chapter 1

## Introduction

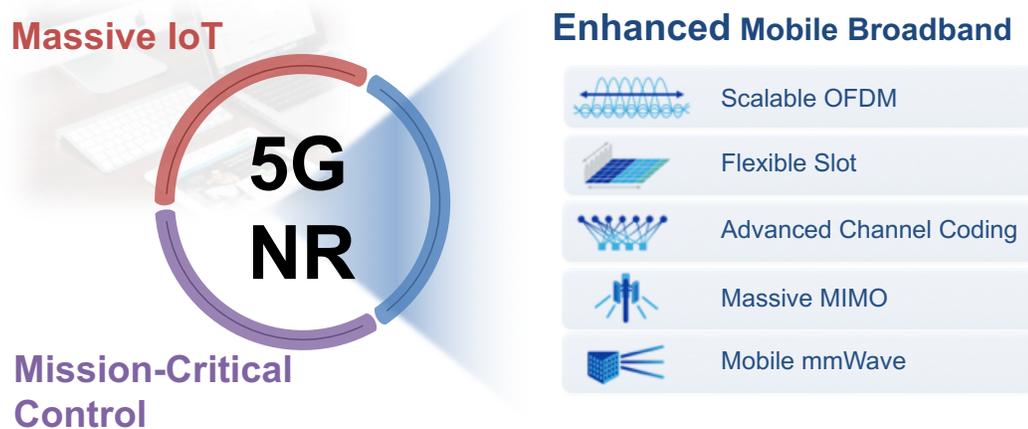


Figure 1.1: Features of 5G New Radio [1]

When compared against fourth-generation (4G) cellular communications, fifth-generation (5G) systems offer higher reliability, lower latency, and higher data rates. These improvements are intended for massive Internet of things (IoTs), mission-critical control, and enhanced mobile broadband, respectively, as shown in Fig. 1.1. To support higher reliability in massive narrowband IoT, for example, higher latency is required because the same data is repeated up to 128 times. For other communications, the latency reduction from 10ms to 1ms has the potential to make remote control of vehicles possible

over 5G connections. The increase in data rates by 10x is enabled by the use of higher millimeter wave (mmWave) frequency bands, e.g. 28 GHz, instead of using bands below 6 GHz.

MmWave communications is promising due to its potential enormous spectrum resource. In addition, large-scale antenna arrays, thanks to the tiny wavelengths of mmWave frequencies, can be packed into a small form factor. Large-scale antenna arrays provide large beamforming gains and help to overcome inherent high propagation losses in these high frequency bands. This particular dissertation is concerned with developing communication architectures and practical algorithms to achieve the 10x increase in data rate using mmWave frequency bands.

## **1.1 Background**

### **1.1.1 Wireless Communication Systems**

A cellular network consists of three large entities: a core network (CN), a radio access network (RAN) and a user equipment (UE). Names of these entities may vary in different standards and generations; however, the roles remain relatively the same. The CN resides on one end of a cellular network communicating with the world of the Internet. The RAN is a bridge between the CN and UEs, in charge of air interface and is composed of multiple base stations (BSs). The UEs include small phones and other devices on the other end of a network that end-users use to access cellular services.

To cover a wide geographic area for wireless communication purposes,

the area can be decomposed into multiple smaller regions, and users in each region are served by one or more stationary wireless transceivers. In cellular networks, a region and a transceiver are referred to as a cell and a base station (BS), respectively. Thus, a BS simultaneously serves a large number of users, and users are served by a single BS in general. A communication link between a UE and a BS is bidirectional, and the link is called *uplink* when a signal direction is from a user to a BS, or *downlink* is in the opposite direction. Division between uplink and downlink is generally done in either the time or frequency domain – they are called time division duplex (TDD) or frequency division duplex (FDD), respectively. There have been research efforts to make the two share the same time-frequency resources in order to maximize the resource usage efficiency.

The most significant evolutions in 5G that have impact on the air interface are the new mmWave frequency bands and the accompanying massive number of antennas to support them. These changes have the potential to provide more 10x the data rate vs. sub-6 GHz bands provided that certain technical challenges can be overcome. The 5G mmWave bands have very different channel characteristics such as high signal power loss during propagation in the air. A massive number of transmit and receive antennas employed on a BS, commonly referred to as massive multi-input multi-output (massive MIMO), require more complicated analog/digital circuit/algorithm designs. However, 5G is not the first challenger to deal with the difficulties. Millimeter wave has been exploited in wireless local and personal area networks (LAN and PAN)

as well as radar applications. A kind of massive MIMO – 64 transmit and 64 receive antennas – also known as full-dimension MIMO (FD-MIMO) was commercially deployed in 4G and has been inherited to the next generation [2, 3].

The first standardization effort for 5G of the cellular networks was released in late 2017 [4], and some updates have been made afterwards. In April 2019, some mobile carriers in South Korea and the United States launched the world’s first 5G service even though it is still debatable who is the first.

### 1.1.2 Wireless Channel

Conveying information over the air is a difficult task, especially at a distance. This is primarily because signals that one wants to deliver get distorted in various manners while propagating through a medium. In addition, the signals are corrupted by thermal noise. In communication systems, the medium is referred to as the channel, and channel estimation is a process to figure out characteristics of the channel. A wireless communication channel is inherently susceptible to distortion as the transmitted signal is most likely to arrive at a receiver after traversing various paths. Along each path, the signal would be reflected, diffracted, scattered and delayed differently. The distortion observed at the receiver is commonly called fading.

Fading is classified into *large-scale fading* and *small-scale fading*. Large-scale fading is mainly due to signal power attenuation through long distance signal propagation (path loss) and signal power fluctuation due to large objects

blocking propagation paths (shadowing). Large scale fading, thus, is location dependent and is modeled for link budget estimation, outage likelihood prediction and cell planning. Small-scale fading is caused by the constructive and destructive superposition of multiple copies of a signal at a receiver as well as mobility of transceivers.

Fading wireless channels can be further classified by two important characteristics: *coherence time* and *coherence bandwidth*. Coherence time is a measure of time duration over which channel remains relatively unchanged and is inversely proportional to Doppler spread. Doppler spread is proportional to the relative velocity of the transmitter and receiver. Channel is called *slow fading* when a symbol period is much shorter than the coherence time, or *fast fading* otherwise. In slow fading channels, multiple symbols can be transmitted and received while a channel impulse response does not change. Coherence bandwidth is a measure of a frequency range over which a channel frequency response is deemed to be flat. Coherence bandwidth is inversely proportional to a delay spread that is caused by multipath propagation. Flatness of channel describes a shape of channel for a given signal bandwidth. In the time domain, it can be interpreted into time period comparison between a symbol and a delay spread. A longer symbol period, or equivalently a narrower signal bandwidth, leads to frequency flat channel. Otherwise, the channel is frequency selective.

Channel reciprocity refers to the fact the wireless channel in the uplink direction is the same as the downlink direction. The most significant implication of channel reciprocity is that a BS can exploits its uplink channel state

information (CSI) for downlink without necessity of feedback from UEs for acquisition of highly quantized downlink CSI. Due to this, TDD and FDD have a great difference from the CSI acquisition standpoint. Channel reciprocity holds for TDD as both links operate in the identical frequency whereas it does not for FDD.

### 1.1.3 Multi-input Multi-output Communication Systems

Multipath, in the past, was deemed to cause detrimental effects on communications, e.g., intersymbol interference (ISI) and frequency selective channel. Many techniques have been developed to address the multipath effects, and employment of multiple antennas is one of them. Multiple antennas at a transmitter and/or a receiver allow spatial diversity that can enhance communication reliability. It can be achieved by maximizing SNR, and equivalently, minimizing the bit error rate (BER).

Another aspect of spatial diversity is beamforming. Multiple antennas with individually variable magnitude and phase can control and shape one or more beams to illuminate desired areas. With the help of channel state information (CSI), antenna arrays can form a beam pointing at a receiver. Beamforming concentrates a limited amount of energy not only to increase a received signal power but also to extend signal reach.

In addition to spatial diversity for reliability enhancement, the multi-input multi-output (MIMO) technologies also offer a great potential for capacity boosting via spatial multiplexing. Spatial multiplexing refers to indepen-

dent and simultaneous transmission of multiple symbols over antenna arrays. Additional degrees of freedom provided by multiple antennas can be exploited with spatially multiplexed data traffic onto multiple paths [5]. Spatial multiplexing takes advantage of multipath propagation rather than battle with it and transforms impediment into advantage. Thus, MIMO channel capacity scales with a number of antennas. Spatial multiplexing can also be used to simultaneously support multiple users, which is referred to as multi-user MIMO (MU-MIMO). Compared against MU-MIMO, single user MIMO (SU-MIMO) refers to spatial multiplexing dedicated to a particular user.

MIMO technology has been adopted and deployed in many modern wireless communication systems. MIMO configurations can provide either improved reliability with diversity techniques or higher throughput with spatial multiplexing techniques or mixture of both. MIMO itself refers to multiple antennas at both a transmitter and a receiver; however, many times MIMO is interchangeably used to refer to spatial multiplexing. For this reason, a larger antenna arrays has been taken into consideration in base stations of cellular networks where heavy power consumption can be supported. For instance, Long Term Evolution (LTE) has been employing and improving MIMO technologies including MU-MIMO since day one [6]. In Wi-Fi standards, MIMO was first introduced in 802.11n [7], and later in 802.11ac [8], MU-MIMO was adopted.

#### 1.1.4 Millimeter Wave Frequency Bands

Millimeter wave (mmWave) is typically defined as radio waves whose frequency ranges from 30–300 GHz; thus, the corresponding wavelength is on the order millimeters. Even though high frequency radio signals are known to suffer high power attenuation over distance, mmWave frequency bands started to draw attention in cellular networks [9–12] due to heavy congestion in conventional low frequency bands. Put differently, mmWave offers a great opportunity to enjoy available abundant spectra that enable higher data rates and lower latency [13,14]. One of the evolutions in progression from LTE to 5G New Radio (5G NR) is adoption of mmWave which is coined as Frequency Range 2 (FR2). In late 2018 and early 2019, the Federal Communications Commission (FCC) conducted the first-ever mmWave spectrum auctions on 24 GHz and 28 GHz frequency bands [15]. These licensed spectra were made available primarily for 5G NR cellular networks, Internet of Things and other advanced spectrum based services.

In communication systems, mmWave and large-scale antenna arrays are complementary due to the compact size of antenna arrays and beamforming gain. Changes in the hardware architecture that mmWave brings about is due to its tiny wavelengths – a smaller form factor can accommodate antenna arrays. For example, a wavelength of 28 GHz frequency is about 10.7 mm. A half-wavelength uniform linear array (ULA) of 64 antennas will be around 34 cm long, which is diminutive compared with the current LTE remote radio head even with much more antennas. On the other hand, large-scale MIMO

is inevitable in adopting mmWave due to natural high path loss of the high frequency bands. In other words, high beamforming gain that can be achieved with large-scale antenna arrays is necessary to compensate a great deal of path loss.

### 1.1.5 Hybrid Analog and Digital Beamforming

In conventional MIMO communication systems, each antenna element comes with its own dedicated transmit and receive radio frequency (RF) chains which consist of power-hungry and costly components, e.g., power amplifiers (PAs) and high-resolution analog-to-digital converters (ADCs). Such a MIMO system is referred to as an all-digital beamforming architecture which provides high flexibility at the price of high power consumption and cost. The power consumption and cost of the RF front end generally scale with a number of antennas at a base station. There are two technologies to address excessive power consumption: hybrid beamforming architectures [16] and low-resolution ADCs [17].

A motivation of the hybrid analog and digital beamforming architectures is simple: to reduce the number of RF chains while maintaining the antenna array size. From hardware implementation perspectives, an intermediate stage needs to be designed to interconnect discrepant numbers of antenna elements and RF chains. It leads to decomposition of beamforming into two stages: baseband (BB) and RF beamforming stages, or equivalently, digital and analog domains. The benefits of the all-digital beamforming architec-

ture remain with the hybrid beamforming architecture. The order of spatial multiplexing is limited by the number of RF chains which is a system design parameter that can be properly selected. The beamforming gain is obtained by the antenna arrays which stays intact. Thus, the overall objective of the hybrid beamforming architecture is to maintain the performance of the all-digital beamforming while reducing the hardware complexity and power consumption [18, 19]. It was shown that the hybrid beamforming can achieve the spectral efficiency of the all-digital beamforming under certain conditions [20–22].

### **1.1.6 Low-resolution Analog-to-Digital Converter**

As mentioned earlier, in addition to the hybrid beamforming architecture, low-resolution ADC is another option for power consumption reduction. Power consumption of typical ADCs is linearly proportional to the sampling rate and grows exponentially in the number of quantization bits [23]. Low resolution is a result from employing a small number of quantization bits in ADCs, and the most extreme case is one-bit ADCs [24]. Meanwhile, coarser signal quantization also leads to larger non-linear quantization errors; e.g., a one-bit ADC outputs either positive or negative one, which violates assumptions of most existing widely used algorithms because of the severe quantization leading to quantization noise power that is significantly higher than the thermal noise power. Analyses of low-resolution ADC systems are complicated due to non-linearity of quantization, which usually results in non-existence of closed-form solutions. It leads to revision of system model and development of new

algorithms under coarse quantization conditions [25–29].

Performance loss due to coarse quantization is inevitable; however, there are some research that shows the loss is allowable. Channel capacity analysis shows that the loss due to 2-3 bit quantization in additive white Gaussian noise (AWGN) channel is small, especially in the low SNR regime [30]. A mixture of high and low resolution ADCs can be employed [31] and is shown to be able to achieve significant portion of full resolution ADC systems.

### 1.1.7 Channel Estimation

In communication systems, the importance of channel estimation lies in data detection. By understanding the channel characteristics through channel estimation, many techniques can be applied to the received signals in order to recover original messages encoded in the signals with much higher probability. In addition to data detection, acquisition of accurate CSI is also vital to MIMO technologies. The benefits of MIMO discussed earlier are based on a critical assumption – perfect CSI is known at both a transmitter and a receiver. It is worth mentioning that non-coherent detection based communication systems do not require channel estimator as the detection algorithms do not rely on the carrier phase. However, in this dissertation, I focus on the coherent detection since the main target systems are the cellular network where the coherent detection is employed.

CSI at a transmitter/receiver is often referred to as CSIT/CSIR separately as their usage and acquisition methods are generally different. CSIR is

naturally available through channel estimation at a receiver; however, CSIT may be not. There are two broad approaches for CSIT acquisition: the open-loop and closed-loop approaches. The open-loop approach relies on the channel reciprocity which is only available in TDD systems. Due to the channel reciprocity, the reverse-link channel information does not require further channel estimation or feedback from the other end. On the other hand, the closed-loop approach is generally based on digital feedback generated from the other end's CSIR and can be used in both TDD and FDD systems. In this dissertation, I investigate open-loop channel estimation.

Channel estimation techniques can be categorized into blind algorithms and pilot-aided algorithms. Blind algorithms take advantage of signal structures and statistical properties of channel. On the contrary, the pilot-aided algorithms explicitly exploits receiver-aware signals. For channel estimation purposes, pilot-aided algorithms are popular in wireless communication systems including Wi-Fi and cellular networks. For example, the pilots are embedded in the Long Training Field (LTF) in 802.11 Wi-Fi systems and various reference signals (e.g., cell specific reference signals and demodulation reference signals) in LTE.

Popular traditional channel estimation algorithms include least squares (LS) estimation and minimum mean square error (MMSE) estimator and their variants [32–34]. More advanced algorithms such as parametric model based [35–38] and iterative [39–42] channel estimation algorithms have also been developed.

The algorithms mentioned above were developed targeting the all-digital beamforming architecture and are not suitable for the hybrid beamforming architectures due to signal dimension compression caused by an analog beamforming stage. To overcome the challenges, new algorithms have been investigated and explored for channel estimation in hybrid beamforming systems, and the algorithms include compressed sensing [19, 43, 44], subspace estimation [45], and tensor decomposition [46].

### 1.1.8 Compressed Sensing

In the mathematical sense, compressed sensing (CS) aims to solve underdetermined linear systems. Put differently, CS is able to recover signals from fewer measurements under particular conditions. The conditions can be described by two words: sparsity and incoherence.

Signals are called sparse when significant components of a given signal can represent the signal and the number of such components are many fewer than the length of the signal. Sparsity of such signals generally can be observed through transformation; i.e., in the transform domain, the majority of coefficients are zero or close to zero. Incoherence is another important key enabler of compressed sensing as it is involved with signal recovery guarantees.

A traditional and powerful measure is the restricted isometry constant (RIC) that is characterized by the restricted isometry property (RIP). RIP is a condition under which a given matrix performs as an orthonormal transformation for sparse vectors, and was first introduced in [47, 48]. Due to its better

computational tractability, on the other hand, the mutual coherence [49, 50] also became a popular measure and is considered more practical in particular situations. There are more measures used for recovery guarantees; however, the RIP and the mutual coherence are the most popular ones.

CS was introduced in [51–53], and these seminal publications triggered and were followed by numerous developments. Due to the compressive feature of CS, it has drawn dramatic attention from various fields that include biomedical imaging, radar imaging, image processing and wireless communications [54–56]. I discuss compressive sensing algorithms for channel estimation next in Section 1.2.

## **1.2 Motivation**

### **1.2.1 Compressed Sensing based Channel Estimation**

Hybrid analog/digital beamforming requires revision to signal processing operations related to beamforming, e.g., codebook design and channel estimation. Hybrid beamforming architectures suffer from having fewer degrees of freedom compared with the all digital architecture. It is because (i) the BB beamforming has smaller dimension than that of RF beamforming and (ii) the RF beamforming stage has constraints due to analog components. The constraints in the RF stage varies depending on designs and components. Conventional channel estimation algorithms employed in all-digital beamforming architectures, e.g., least squares estimator, generally yield degraded performance for hybrid beamforming architectures. Therefore I consider compressed

sensing techniques to estimate millimeter wave channels when using hybrid analog and digital hybrid beamforming architectures.

This dissertation is composed of six chapters as follows. In Chapters 2 and 3, I design the deterministic analog and digital codebooks for CSI acquisition purposes under the hardware constraints for several different hybrid beamforming architectures. As precoding matrices are multiplied by a vector of pilot symbols in the transmitter, the precoding matrices and the pilot symbols must be known in terms of their codebooks in the receiver during channel estimation. The importance of well-designed codebooks lies in that they are involved in a sensing matrix which plays a key role in successful CS-algorithm performance. In addition to the codebook design efforts, I propose to reduce training overhead for CSI estimation by using deterministic approaches which reduce randomness. Then, in Chapter 4, I incorporate low-resolution ADCs (for significant power reduction) in the design of CS-based channel estimation methods. In this case, a random codebook is used, as is often used in CS-based channel estimation methods. CS techniques are useful not only in CSI estimation, but also in beam detection. In Chapter 5, I develop CS-based beam detection during initial access in 5G NR systems. The dissertation summary and future work are provided in the last chapter.

For point-to-point mmWave communications, I develop a deterministic hybrid beamforming codebook design framework for CS-based channel estimation. The framework is versatile in that it is applicable to various hybrid beamforming architectures including phase shifters, switches and RF lens. The

design approach is to configure analog and digital beamformers in the most favorable forms to CS techniques under the hybrid beamforming constraints. I decouple the joint transmit and receive optimization problem into two disjoint problems, which allows the precoder codebook in the transmitter and combiner codebook in the receiver to be designed separately. When one tries to reduce training and measurement overhead of CS-based channel estimation, extra randomness is often introduced in the form of random RF precoder permutation. I propose a computationally efficient algorithm to find a deterministic order of RF precoders that can reduce the overhead in half without significant performance loss.

Low-resolution ADC is another means for further power consumption reduction along with the hybrid beamforming architecture. However, the combination of a hybrid beamforming architecture and low-resolution quantizers makes channel estimation in such systems even more challenging. Adopting an extremely low (one-bit) ADC resolution in the hybrid beamforming communication systems, I develop a CS-based channel estimation algorithm that is suitable for one-bit quantization. Generalized approximate message passing (GAMP) and its variants have widely been used for channel estimation. One variant of GAMP is named one-bit GAMP and is specifically developed for measurements taken with one-bit quantizers. I modify the algorithm and make it better match a communication system model.

In developing 5G NR, a new challenge has been arisen: beam management. Since beamforming became an essential component in 5G, beam search

and detection are performed even in the initial access. The time-frequency resource allocated for initial access is limited, and a growing number of antennas is taken into consideration. It gives rise for the need of an efficient approach for beam detection. I investigate CS-based downlink beam detection for mmWave hybrid beamforming systems taking the 3GPP standard into account. With the exhaustive search being a benchmark, the CS approach is evaluated using the random and the DFT RF codebooks in terms of the beam detection probability. I also consider a larger number of antennas at the base station and explore two RF codebook options in the CS context.

### **1.3 Dissertation Summary**

In summary, I propose to improve channel estimation accuracy in hybrid analog and digital beamforming communication systems, which in turn increases beamforming accuracy and cell throughput.

#### **1.3.1 Thesis Statement**

In this dissertation, I defend the following statement:

*Compressed sensing is the key to exploit sparsity in angular mmWave channel responses in order to unlock fast and accurate channel estimation and beam selection algorithms for power efficient hybrid beamforming architectures.*

### 1.3.2 Overview of Contributions

The main contributions of this dissertation are summarized as follows:

1. **Versatile Compressive mmWave Hybrid Beamformer Codebook Design Framework:** A versatile deterministic hybrid beamforming codebook design framework is proposed for CS algorithms with coherence-based recovery guarantees. The versatile framework is developed in consideration of various promising hybrid beamforming architectures including phase shifting network based, switching network based, and RF lens based architectures. The design philosophy is to find the codebooks that minimize *total coherence* of the sensing matrix. The joint transmitter and receiver optimization problem is shown to be decomposable into two disjoint problems. The codebooks obtained by solving the disjoint optimization problems outperform the random codebooks in terms of estimation error and spectral efficiency.
2. **Hybrid Beamformer Codebook Design and Ordering for Compressive mmWave Channel Estimation:** A deterministic codebook and pilot symbol design is proposed for phase shifting network hybrid beamforming architecture. The codebook works for CS-based channel estimation algorithms that rely on low coherence for their recovery guarantees. The codebook is obtained by *mutual coherence* minimization of the sensing matrix. In addition, the precoder column ordering algorithm for the pilot symbol design is proposed for further determinism in CSI

acquisition process.

3. **Narrowband Channel Estimation for Hybrid Beamforming Millimeter Wave Communication Systems with One-bit Quantization:**

A narrowband CS-based channel estimation algorithm is investigated for hybrid beamforming architecture with one-bit quantization ADCs. The proposed CS algorithm is based on a variant generalized approximate message passing (GAMP) that outperforms least-squares estimator without quantization. The modified one-bit GAMP that I propose yields the lowest estimation error among the GAMP variants as well as least-squares estimator.

4. **Compressed-Sensing based Beam Detection in 5G New Radio**

**Initial Access:** A CS approach is investigated for downlink beam detection for 5G NR initial access under the context of a hybrid beamforming architecture. Abiding by the 3GPP standards with some simplification, the CS-based beam detection approach is shown to outperform exhaustive search in terms of beam detection accuracy. A massive number of antennas (e.g., 256), with which exhaustive search is not feasible, is also taken into account to show the importance of careful codebook design.

## 1.4 Conclusion

This section provides the background information that helps readers to understand the motivation of this dissertation. A brief summary of contribu-

tions are also provided. The following chapters discuss the technical details of my research starting from the deterministic codebook design framework for CS-based channel estimation.

## 1.5 Notation and Abbreviations

### 1.5.1 Notation

Bold uppercase  $\mathbf{A}$ , bold lower case  $\mathbf{a}$ , and calligraphic letter  $\mathcal{A}$  denote a matrix, a column vector, and a set, respectively.  $|\mathcal{A}|$  denotes the cardinality of a set  $\mathcal{A}$ . Superscripts  $\mathbf{A}^H$ ,  $\mathbf{A}^T$  and  $\mathbf{A}^*$  denote the conjugate transpose, the transpose, and the conjugate, respectively.  $\|\mathbf{A}\|_F$ ,  $\|\mathbf{a}\|_0$  and  $\|\mathbf{a}\|_2$  denotes the Frobenius,  $\mathcal{L}_0$ , and  $\mathcal{L}_2$  norms, respectively.  $\text{vec}(\mathbf{A})$  is a vector obtained through the vectorization of a matrix  $\mathbf{A}$ .  $\text{blkdiag}(\mathbf{A}_1, \dots, \mathbf{A}_N)$  denotes a block diagonal matrix whose diagonal block entries are  $\{\mathbf{A}_1, \dots, \mathbf{A}_N\}$ . A column vector that consists of the diagonal entries of  $\mathbf{A}$  is denoted by  $\text{diag}(\mathbf{A})$ . The  $n$ -th column and  $(m, n)$ -th entry of the matrix  $\mathbf{A}$  are denoted by  $\mathbf{A}(n)$  and  $\mathbf{A}(m, n)$ , respectively. The  $n$ -th element of the vector  $\mathbf{a}$  is denoted by  $\mathbf{a}_n$ .  $\text{Tr}(\mathbf{A})$  denotes the trace.  $\mathbf{A} \otimes \mathbf{B}$  is the Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$ . The  $N \times N$  identity matrix is denoted by  $\mathbf{I}_N$ .

### 1.5.2 Abbreviations

**3GPP** 3rd Generation Partnership Project

**5G** the fifth generation of mobile network

**ADC** analog-to-digital converter

**AoA** angle of arrival  
**AoD** angle of departure  
**AWGN** additive white Gaussian noise  
**BB** baseband  
**BER** bit error rate  
**BPDN** basis pursuit denoising  
**BS** base station  
**CAP-MIMO** continuous aperture phased multi-input multi-output  
**CE** channel estimation  
**CN** core network  
**CS** compressed sensing  
**CSI** channel state information  
**DAC** digital-to-analog converter  
**DFT** discrete Fourier transform  
**DLA** discrete lens array  
**DMRS** demodulation reference signal  
**EM** expectation maximization  
**ES** exhaustive search  
**FCC** Federal Communications Commission  
**FDD** frequency division duplex  
**FD-MIMO** full-dimension MIMO  
**FR** frequency range  
**GAMP** generalized approximate message passing

**GM** Gaussian mixture  
**gNB** next generation NodeB  
**HB** hybrid beamforming  
**IA** initial access  
**IS** iterative search  
**ISI** inter symbol interference  
**LAN** local area network  
**LS** least squares  
**LSE** least squares estimator  
**LTE** Long Term Evolution  
**MIMO** multi-input multi-output  
**MMSE** minimum mean square error  
**mmWave** millimeter wave  
**MPC** multi-path component  
**MTC** maximal total coherence  
**MU-MIMO** multi-user multi-input multi-output  
**NMSE** normalized mean square error  
**NR** New Radio  
**OFDM** orthogonal frequency division multiplexing  
**OMP** orthogonal matching pursuit  
**PA** power amplifier  
**PAN** personal area network  
**PSS** primary synchronization signal

**RAN** radio access network  
**RF** radio frequency  
**RIC** restricted isometry constant  
**RIP** restricted isometry property  
**SNR** signal-to-noise ratio  
**SSB** synchronization signal block  
**SU-MIMO** single-user MIMO  
**SW-OMP** simultaneous weighted orthogonal matching pursuit  
**TDD** time division duplex  
**UE** user equipment  
**ULA** uniform linear array

## Chapter 2

# Versatile Compressive mmWave Hybrid Beamformer Codebook Design Framework

In HB systems, channel estimation (CE) becomes challenging due to indirect access from the baseband processing to the communications channels and low SNR before beam alignment. Compressed sensing (CS) based algorithms have been adopted to address these challenges by leveraging the sparse nature of millimeter wave channels. In many CS algorithms for narrowband CE, hybrid beamformers are randomly configured. A random configuration does not always yield low-coherence sensing matrices desirable for those CE algorithms whose recovery guarantees rely on coherence.

In this chapter<sup>1</sup>, I propose a deterministic sensing matrix design for CS algorithms with coherence-based recovery guarantees to enhance CE accuracy. The contributions of this chapter include the design of deterministic beamformers that (1) minimizes total coherence of a sensing matrix and (2) provides the same total coherence for phase shifting, switching, and discrete

---

<sup>1</sup> Part of the work was published in J. Sung and B. L. Evans, “Versatile compressive mmWave hybrid beamformer codebook design framework”, in *Proceedings of IEEE International Conference on Computing, Networking and Communications (ICNC)*, Feb. 17–20, 2020, Big Island, HI, USA [57]. This work was supervised by Prof. Brian L. Evans.

lens array HB architectures. The codebooks obtained for deterministic beamformers are shown in simulations to outperform random configurations in terms of both CE error and spectral efficiency using orthogonal matching pursuit and basis pursuit denoising.

## 2.1 Introduction

Millimeter wave (mmWave) frequency bands have the large transmission bandwidths to enable the 10x increase in data rates to move from 4G to 5G cellular communications. These bands have already been adopted in the 5G New Radio standard [58]. Achieving large beamforming gains, and hence having a large number of antennas, is required to overcome the high propagation losses in these high frequency bands. Large-scale antenna arrays, thanks to the tiny wavelengths of mmWave bands, can be packed into a small form factor; however, in conventional all-digital beamforming systems, having as many radio frequency (RF) transceivers as antennas would bring about prohibitively high power consumption due to power-demanding components therein, e.g., high speed analog-to-digital converters (ADCs). For reduction of power consumption and cost, hybrid analog and digital beamforming (HB) architectures have drawn attention as they reduce the number of RF chains while retaining a large number of antennas.

In the course of reducing power consumption, channel estimation (CE) becomes challenging. Conventional CE algorithms require huge training overhead due to large-scale antenna arrays, and HB MIMO architectures prohibit

direct access and estimation of the communication channel matrix entries. In addition, low signal-to-noise ratio (SNR) is unavoidable without proper beamforming before aligning transmit and receive beams. Channel state information (CSI) which is acquired by CE is of importance in MIMO communication systems since the beamformer configuration is based on it. Especially for HB systems, CSI is critical for joint analog and digital beamformer configuration.

Through measurement campaigns, mmWave channels have been shown to be sparse in both time and angular dimensions [59,60] due to a small number of clusters in each dimension. The sparsity naturally takes into consideration the beamspace representation of MIMO channels, and is leveraged for communications including HB schemes and CE. It specifically raises effectiveness of compressed sensing (CS) algorithms in mmWave CE.

### **2.1.1 Prior Work**

For analog beamforming systems, closed-loop adaptive beam training/search approaches were proposed in [61–63] to avoid an exhaustive search that requires a significant amount of time for beam alignment. In communication systems considered in those publications, beamformers are implemented with a network of phase shifters connected to a single RF chain. Since the systems are constrained to have a single transceiver, they find the best transmit and receive beam pair on which only one data stream can be communicated. Assuming only a single data stream exists in the communication, the beamforming training in IEEE 802.11ad [64] could be used. Limitations of analog

beamforming are that 1) it only can support a single data stream [65] and 2) beam steering accuracy highly depends on resolution of quantized angles of the phase shifters [64, 65].

HB was conceived to reduce the number of transceivers in all-digital beamforming systems that are widely used for sub-6GHz bands [16, 18, 66]. Unlike all-digital beamforming systems in which each antenna has its own dedicated RF transceiver and all signal processing is performed in the baseband, HB architectures divide the beamforming processing between the analog and digital stages. It reduces the number of transceivers to be far fewer than the number of antennas and supports spatial multiplexing for multiple data streams. The analog stage of the HB architecture is flexible in its implementation with different analog devices such as phase shifters [18, 67, 68], switches [69] or lenses [70].

Using the HB architecture implemented with a network of variable phase shifters, adaptive CS was proposed to iteratively train beams in [19]. It was extended to cover the multi-user case in [71] while users therein were assumed to employ analog beamforming with a single transceiver.

Other approaches leveraging CS algorithms for sparse CE were proposed (1) for analog beamforming [72], (2) for digital beamforming [73], and (3) for HB [44, 69, 71, 74–76]. In those publications, transmit pilot symbols [73, 75] or angles of phase shifters in analog beamformers [44, 69, 71, 72, 74, 76] are randomly chosen from finite predefined sets satisfying hardware constraints. This random selection is expected to provide low mutual coherence sensing

matrices and satisfy the restricted isometry property (RIP) condition with high probability (see Section 1.1.8 for more on RIP). Authors in [43] instead proposed to design the pilot beam pattern by minimizing the total coherence of the equivalent dictionary and showed that the deterministic design achieved improvement in CE accuracy. The mutual coherence has been widely used to assess the dependence of columns in a given sensing matrix [49, 50]. Unlike mutual coherence, the total coherence can easily be used to derive an optimal beamformers. A qualitative comparison of CS-based CE methods including the proposed contribution is provided in Table 2.1 where PS and SW-OMP stand for phase shifter and simultaneous weighted OMP, respectively. Concerning the complexity evaluation given in the table, (1) codebook design is done offline and is not included; (2) the complexity is the runtime complexity of the channel estimation algorithm to evaluate the design; and (3) [44] has high complexity due to per-subcarrier OMP application without complexity reduction techniques.

Low-resolution ADCs can be combined with HB architectures to further reduce power consumption of high speed converters. Such systems were shown, in low and medium SNR range, to obtain comparable achievable rates to those for systems with full-precision converters [77]. Considering low-resolution ADCs with up to five bits in [78], CE errors obtained with the proposed CE algorithm therein are claimed to be acceptable, and more than three bits of resolution does not result in significant improvement. In [79], one-bit generalized approximate message passing (GAMP), which is a GAMP

variant, was used for CE in systems equipped with one-bit quantizers and shown to achieve better CE performance in normalized mean squared error than do other GAMP variants and the least-squares estimator (LSE).

Non-CS based CE techniques have been developed as well that make use of subspace estimation [45], low-rank tensor decomposition [46], and multiple frequency tones [80]. These publications have widely varying simulation setups, system models, configurations, and assumptions, which make a direct quantitative comparison difficult. Instead, Table 2.2 provides a qualitative comparison for these non-CS based CE techniques.

Table 2.1: Comparison among CS-based channel estimation algorithms

	[71]	[69]	[43]	[44]	[76]	Proposed
Hybrid	✓	✓	✓	✓	✓	✓
High resolution ADCs	✓	✓	✓	✓	✓	✓
Hybrid beamforming architecture	PS	PS switch	PS	PS	PS	PS switch DLA
Deterministic	×	×	✓	×	×	✓
Wideband	×	×	×	✓	✓	×
Algorithm	OMP	OMP	OMP	OMP	SW-OMP	OMP BPDN
Complexity	Med.	Med.	Med.	High	Med.	Med.
Performance	High	High	High	High	High	High

### 2.1.2 Contributions

In this chapter, I propose a deterministic design of hybrid precoders and combiners for CS based open-loop narrowband mmWave CE that is universally applicable to various HB architectures. In contrast with fully random beamformer configurations for narrowband CE [69, 71, 74, 75], I design a deterministic sensing matrix as in [43] and obtain codebooks for analog and digital beamformers. The deterministic design can lower the coherence of a sensing matrix and enhance the accuracy of CS algorithms that have coherence-based recovery guarantees. More accurate CSI in terms of normalized mean square error naturally results in greater spectral efficiency (SE) during the data transmission phase [81, 82]. The proposed sensing matrix design is ap-

Table 2.2: Brief comparison among non-CS based CE techniques

	[45]	[46]	[80]
Method	subspace estimation	low-rank tensor	multiple tones
Carriers	Single	Multi (OFDM)	Single
Users	Single	Single	Multi
Tx antennas and RF chains	32/4, 64/8, 128/16, 256/32	64/unknown	100/10
Rx antennas and RF chains	16/4, 32/8, 64/16, 128/32	32/1	16/1
Metrics	spectral efficiency	NMSE	spectral efficiency

plicable to phase shifting and switching network based HB architectures [69] and a discrete lens array (DLA) based HB architecture [70]. The sparse recovery formulation provided in this chapter allows various CS algorithms for CE. I adopt orthogonal matching pursuit (OMP) and basis pursuit denoising (BPDN) as representatives of CS algorithms with coherence-based recovery guarantees [83]. The contributions are summarized as follows:

- I propose a deterministic analog and digital precoder/combiner design that minimizes the total coherence of an equivalent dictionary and obtains codebooks by the proposed design. In sparse recovery formulation, the equivalent dictionary is a product of a sensing matrix and a given sparsifying dictionary, and the sensing matrix is involved with analog and digital precoder and combiner matrices. Prior publications [69, 71, 74, 75] randomly configure the beamformer components, i.e., angles of phase shifters and connections of switches, and apply CS based CE algorithms. However, I design deterministic analog and digital beamformers by taking into account the total coherence [43]. The deterministic beamformer codebooks are obtained by solving the total coherence minimization problem under hardware constraints due to the underlying analog beamformer hardware. When a system has enough antennas and uses full training, the proposed method can design deterministic codebooks with the minimum total coherence possible, subject to the hardware constraints. In simulations, I show that the employed CE algorithms (i.e., OMP and BPDN) achieve lower CE errors and higher

spectral efficiencies using deterministic codebook designs than random codebook designs.

- I show that, without any approximation techniques, the joint precoder and combiner optimization problem decomposes into two disjoint problems for a transmitter and a receiver, respectively. The precoder optimization problem starts with three sets of degrees of freedom (analog and digital precoders, and training symbol vectors), and I show that two of them are closely related with one-to-one mappings. Both the precoder and combiner, thus, have two sets of degrees of freedom that can be explored to minimize the total coherence. The resulting design is slightly different from what was proposed in [43] because I do not employ approximation techniques. In the derivation, I also provide a rationale behind matrix selection for analog precoding and combining matrices. To verify the proposed design method, I present two different derivations that lead to identical results.
- I show applicability of the design codebooks to the three most promising HB architectures, which are based on phase shifters, switches and a DLA. Seven types of HB architectures are discussed in this chapter: two with phase shifters, four with switches [69], and one with a DLA [70]. In formulation of the proposed beamformer design, no constraints on analog precoders and combiners are posed in the first place for versatility of the design. It leads to the following advantages: (1) universal applicability

to all seven types of hybrid architectures, (2) better performance than that of the random beamformer configuration, and (3) identical total coherence regardless of architectures.

## 2.2 System and Channel Models

I consider mmWave MIMO communication systems equipped with a general hybrid analog and digital beamforming architecture as shown in Fig. 2.1. Both the hybrid precoder and combiner are composed of digital and analog stages. The analog beamforming stage can be implemented with various devices (e.g., phase shifters, switches and a DLA), and when phase shifters are taken into account, they have quantized angles with the  $b_{\text{PS}}$ -bit quantization resolution. A transmitter and a receiver are equipped with  $N_t$  and  $N_r$  antennas, and  $L_t$  and  $L_r$  RF chains, respectively, and it is assumed that  $L_t \leq N_t$  and  $L_r \leq N_r$ . For frequency-flat channels, the discrete-time received signal in the  $m$ -th frame (or the  $m$ -th time instant) can be written as

$$\mathbf{y}_m = \sqrt{\rho} \mathbf{W}_m^H \mathbf{H} \mathbf{F}_m \mathbf{x}_m + \mathbf{W}_m^H \mathbf{n}_m \in \mathbb{C}^{L_r},$$

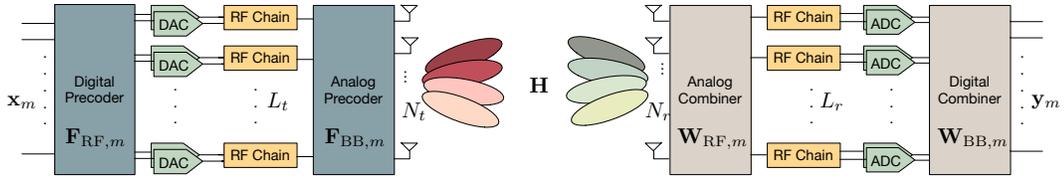


Figure 2.1: Block diagram of a general hybrid analog and digital beamforming architecture at both a transmitter and a receiver

where  $\rho$  is the transmit power in the training phase,  $\mathbf{W}_m = \mathbf{W}_{\text{RF},m} \mathbf{W}_{\text{BB},m} \in \mathbb{C}^{N_r \times L_r}$  is the hybrid combiner,  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  is the channel matrix,  $\mathbf{F}_m = \mathbf{F}_{\text{RF},m} \mathbf{F}_{\text{BB},m} \in \mathbb{C}^{N_t \times L_t}$  is the hybrid precoder,  $\mathbf{x}_m \in \mathbb{C}^{L_t}$  is the training symbols vector, and  $\mathbf{n}_m \in \mathbb{C}^{N_r} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$  is the additive noise vector. A hybrid combiner (precoder) in the  $m$ -th frame is composed of an RF combiner  $\mathbf{W}_{\text{RF},m} \in \mathbb{C}^{N_r \times L_r}$  (an RF precoder  $\mathbf{F}_{\text{RF},m} \in \mathbb{C}^{N_t \times L_t}$ ) and a baseband (BB) combiner  $\mathbf{W}_{\text{BB},m} \in \mathbb{C}^{L_r \times L_r}$  (a BB precoder  $\mathbf{F}_{\text{BB},m} \in \mathbb{C}^{L_t \times L_t}$ ). In order to keep a constant transmit power,  $\|\bar{\mathbf{x}}_m\|^2 = 1$  where  $\bar{\mathbf{x}}_m \triangleq \mathbf{F}_m \mathbf{x}_m$ . To exploit all available combinations of transmit and receive configurations, it is assumed that all available receive configurations are used for a given transmit configuration. With  $M_t$  and  $M_r$  denoting the number of different configurations of transmit and receive beamformers, respectively,  $M = M_t M_r$  is the maximum number of frames that can be used for CE. The received signal matrix  $\mathbf{Y}$  that contains  $M$  frames can be written as

$$\mathbf{Y} = [[\mathbf{y}_1^\top, \dots, \mathbf{y}_{M_r}^\top]^\top, \dots, [\mathbf{y}_{M-M_r+1}^\top, \dots, \mathbf{y}_M^\top]^\top] = \sqrt{\rho} \mathbf{W}^\text{H} \mathbf{H} \bar{\mathbf{X}} + \mathbf{N}, \quad (2.1)$$

where  $\mathbf{Y} \in \mathbb{C}^{L_r M_r \times M_t}$ ,  $\mathbf{W} \triangleq [\mathbf{W}_1, \dots, \mathbf{W}_{M_r}] \in \mathbb{C}^{N_r \times L_r M_r}$  is the receive configuration matrix,  $\bar{\mathbf{X}} \triangleq [\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_{M_t}] \in \mathbb{C}^{N_t \times M_t}$  is the transmit configuration matrix, and  $\mathbf{N} \in \mathbb{C}^{L_r M_r \times M_t}$  is the noise matrix which is expressed as

$$\mathbf{N} \triangleq \text{blkdiag}(\mathbf{W}_1, \dots, \mathbf{W}_{M_r})^\text{H} [[\mathbf{n}_1^\top, \dots, \mathbf{n}_{M_r}^\top]^\top, \dots, [\mathbf{n}_{M-M_r+1}^\top, \dots, \mathbf{n}_M^\top]^\top].$$

Adopting a geometric channel model [19, 43] and linear antenna arrays, each scatterer contributes a channel path associated with its own azimuth angle

of departure and arrival (AoD and AoA) denoted by  $\theta_{tl}$  and  $\theta_{rl}$ , respectively. Therefore, the channel matrix can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_p}} \sum_{l=0}^{N_p-1} \alpha_l \mathbf{a}_r(\theta_{rl}) \mathbf{a}_t^H(\theta_{tl}), \quad (2.2)$$

where  $N_p$  is the total number of paths (equivalent to the channel rank),  $\alpha_l \sim \mathcal{CN}(0, \sigma_\alpha^2)$  is the complex channel gain of the  $l$ -th path, and  $\mathbf{a}_t(\cdot) \in \mathbb{C}^{N_t}$  and  $\mathbf{a}_r(\cdot) \in \mathbb{C}^{N_r}$  are the transmit and receive array response vectors, respectively, evaluated for the given angles. Both  $\theta_{tl}$  and  $\theta_{rl}$  are random variables that follow the uniform distribution  $\mathcal{U}(0, 2\pi)$ . Assuming the transmit and receive antennas are in the form of uniform linear array (ULA) with a half wavelength antenna spacing, the array response vectors are given as

$$\begin{aligned} \mathbf{a}_t(\theta) &= \sqrt{\frac{1}{N_t}} [1, e^{-j\pi \cos(\theta)}, \dots, e^{-j\pi(N_t-1) \cos(\theta)}]^\top, \\ \mathbf{a}_r(\theta) &= \sqrt{\frac{1}{N_r}} [1, e^{-j\pi \cos(\theta)}, \dots, e^{-j\pi(N_r-1) \cos(\theta)}]^\top. \end{aligned}$$

Defining the array response matrices  $\mathbf{A}_t \triangleq [\mathbf{a}_t(\theta_{t0}), \mathbf{a}_t(\theta_{t1}), \dots, \mathbf{a}_t(\theta_{t(N_p-1)})]$  and  $\mathbf{A}_r \triangleq [\mathbf{a}_r(\theta_{r0}), \mathbf{a}_r(\theta_{r1}), \dots, \mathbf{a}_r(\theta_{r(N_p-1)})]$ , the channel matrix  $\mathbf{H}$  in (2.2) can be expressed in matrix form as

$$\mathbf{H} = \mathbf{A}_r \mathbf{H}_d \mathbf{A}_t^H, \quad (2.3)$$

where  $\mathbf{H}_d \in \mathbb{C}^{N_p \times N_p}$  is a square matrix with the scaled complex channel gains on the diagonal.

## 2.3 Sparse Formulation

For application of CS algorithms to sparse CE, the received signal matrix in (2.1) can be rewritten in vector form using the matrix equality,  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\text{vec}(\mathbf{B})$ , as

$$\mathbf{y} = \sqrt{\rho} (\bar{\mathbf{X}}^\top \otimes \mathbf{W}^\text{H}) \text{vec}(\mathbf{H}) + \mathbf{n}, \quad (2.4)$$

where  $\mathbf{y} \in \mathbb{C}^{ML_r}$  collects the  $M$  received frames in vector form, and  $\mathbf{n} \triangleq \text{vec}(\mathbf{N})$ . In addition, angle grids need to be selected based on which CS algorithms search pairs of an AoD and an AoA. I choose the angles to be uniformly distributed in the normalized discrete spatial angle domain, i.e.,  $[-1/2, 1/2)$ , such that  $\vartheta_n \in \Theta = \{\vartheta_n | \vartheta_n = \frac{n}{G} - \frac{1}{2}, n = 0, \dots, G-1\}$  where  $G \geq \max(N_t, N_r)$  is the grid size. Therefore, the higher a  $G$  value is, the finer the angle granularity becomes. The array response vectors with spatial angle  $\vartheta$  can be expressed as

$$\begin{aligned} \bar{\mathbf{a}}_t(\vartheta) &= \sqrt{\frac{1}{N_t}} [1, e^{-j2\pi\vartheta}, \dots, e^{-j2\pi(N_t-1)\vartheta}]^\top, \\ \bar{\mathbf{a}}_r(\vartheta) &= \sqrt{\frac{1}{N_r}} [1, e^{-j2\pi\vartheta}, \dots, e^{-j2\pi(N_r-1)\vartheta}]^\top. \end{aligned}$$

The angle grids are defined by the transmit and receive grid array response matrices  $\bar{\mathbf{A}}_t \triangleq [\bar{\mathbf{a}}_t(\vartheta_0), \dots, \bar{\mathbf{a}}_t(\vartheta_{G-1})] \in \mathbb{C}^{N_t \times G}$  and  $\bar{\mathbf{A}}_r \triangleq [\bar{\mathbf{a}}_r(\vartheta_0), \dots, \bar{\mathbf{a}}_r(\vartheta_{G-1})] \in \mathbb{C}^{N_r \times G}$ , respectively. Note that  $\bar{\mathbf{A}}_t \bar{\mathbf{A}}_t^\text{H} = \frac{G}{N_t} \mathbf{I}_{N_t}$  and  $\bar{\mathbf{A}}_r \bar{\mathbf{A}}_r^\text{H} = \frac{G}{N_r} \mathbf{I}_{N_r}$ . The channel matrix (2.3) can be redefined with such grid array response matrices as

$$\mathbf{H} = \bar{\mathbf{A}}_r \bar{\mathbf{H}}_d \bar{\mathbf{A}}_t^\text{H},$$

where  $\bar{\mathbf{H}}_d \in \mathbb{C}^{G \times G}$  is the new channel gain matrix. Ignoring the grid quantization errors,  $\bar{\mathbf{H}}_d$  is a sparse matrix with  $N_p$  non-zero entries being complex channel gains corresponding to a combination of each transmit and receive array response vector in  $\bar{\mathbf{A}}_t$  and  $\bar{\mathbf{A}}_r$ , respectively. Unlike  $\mathbf{H}_d$ , non-zero elements in  $\bar{\mathbf{H}}_d$  do not have to be on the diagonal. As  $\text{vec}(\mathbf{H}) = (\bar{\mathbf{A}}_t^* \otimes \bar{\mathbf{A}}_r) \text{vec}(\bar{\mathbf{H}}_d)$ , the received signal vector in (2.4) can be rewritten as

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho} (\bar{\mathbf{X}}^\top \otimes \mathbf{W}^\text{H}) (\bar{\mathbf{A}}_t^* \otimes \bar{\mathbf{A}}_r) \text{vec}(\bar{\mathbf{H}}_d) + \mathbf{n} = \sqrt{\rho} (\bar{\mathbf{X}}^\top \bar{\mathbf{A}}_t^* \otimes \mathbf{W}^\text{H} \bar{\mathbf{A}}_r) \mathbf{h} + \mathbf{n} \\ &= \sqrt{\rho} \bar{\Phi} \Psi \mathbf{h} + \mathbf{n} = \sqrt{\rho} \bar{\Phi} \mathbf{h} + \mathbf{n}, \end{aligned}$$

where  $\mathbf{h} \triangleq \text{vec}(\bar{\mathbf{H}}_d)$  is the  $N_p$ -sparse channel vector,  $\Psi \triangleq \bar{\mathbf{A}}_t^* \otimes \bar{\mathbf{A}}_r \in \mathbb{C}^{N_t N_r \times G^2}$  is the sparsifying dictionary,  $\bar{\Phi} \triangleq \bar{\mathbf{X}}^\top \otimes \mathbf{W}^\text{H} \in \mathbb{C}^{L_r M_t M_r \times N_t N_r}$  is the sensing matrix, and  $\bar{\Phi} \triangleq \bar{\Phi} \Psi$  is the equivalent dictionary.

## 2.4 Deterministic Sensing Matrix Design

Previous efforts have carefully designed a sensing matrix rather than using a random one in order to improve the performance of CS algorithms. Mutual coherence is popular measure used for recovery guarantees of CS algorithms. Mutual coherence is defined as the maximum absolute value of the cross correlation between the columns of a given matrix, or equivalently is defined by

$$\mu(\mathbf{A}) = \max_{1 \leq i \neq j \leq m} |\mathbf{a}_i^\text{H} \mathbf{a}_j|$$

where  $\mathbf{a}_i$  denotes the  $i$ -th column of the matrix  $\mathbf{A}$ . However, a significant drawback of using mutual coherence is the loss of differentiability. Finding op-

timal sensing matrices by minimizing mutual coherence becomes a “nonconvex nonsmooth minimization problem” [84].

Alternatives to the mutual coherence, thus, are considered in [85–87]. When it comes to CS based mmWave CE, in [44, 69, 76, 88], entries of the precoder and combiner are randomly drawn from a certain set that meets the hardware in use. In [43], on the other hand, a concept of the total coherence was taken into consideration for sensing matrix design. However, [43] only considered a subset of columns of the equivalent dictionary for optimization, and used approximation techniques to decompose into separate transmit and receive optimization problems.

Focusing on minimizing the total coherence, I provide two different approaches that lead to the same design methods. Sections 2.4.1 to 2.4.3 give the first approach, Section 2.4.4 gives an alternative approach, and Section 2.4.5 compares the proposed approach against other random selection approaches. Furthermore, Section 2.4.1 specifically gives conditions on which the proposed approach can be used in hybrid beamforming architectures. It is shown in Section 2.5 that various promising hybrid architectures meet the conditions.

### 2.4.1 Total Coherence

I define the total coherence of the equivalent dictionary  $\bar{\Phi}$  as

$$\mu^{\text{total}}(\bar{\Phi}) = \sum_m^{G^2} \sum_{n, n \neq m}^{G^2} |\bar{\Phi}(m)^H \bar{\Phi}(n)|^2$$

$$= \sum_m^{G^2} \left\{ \sum_n^{G^2} |\bar{\Phi}(m)^H \bar{\Phi}(n)|^2 - |\bar{\Phi}(m)^H \bar{\Phi}(m)|^2 \right\}.$$

As the total coherence is a sum of squared inner products of different columns in a matrix,  $\mu^{\text{total}}(\bar{\Phi})$  is the same as the squared Frobenius norm of  $\bar{\Phi}^H \bar{\Phi}$  without diagonal entries. Defining  $\tilde{\mathbf{X}} \triangleq \bar{\mathbf{X}}^T \bar{\mathbf{A}}_t^*$  and  $\tilde{\mathbf{W}} \triangleq \mathbf{W}^H \bar{\mathbf{A}}_r$ ,  $\mu^{\text{total}}(\bar{\Phi})$  can be expressed as

$$\begin{aligned} \mu^{\text{total}}(\bar{\Phi}) &= \sum_{m_1}^G \sum_{m_2}^G \left\{ \sum_{n_1}^G \sum_{n_2}^G \left| \left\{ \tilde{\mathbf{X}}(m_1) \otimes \tilde{\mathbf{W}}(m_2) \right\}^H \times \left\{ \tilde{\mathbf{X}}(n_1) \otimes \tilde{\mathbf{W}}(n_2) \right\} \right|^2 \right. \\ &\quad \left. - \left| \left\{ \tilde{\mathbf{X}}(m_1) \otimes \tilde{\mathbf{W}}(m_2) \right\}^H \left\{ \tilde{\mathbf{X}}(m_1) \otimes \tilde{\mathbf{W}}(m_2) \right\} \right|^2 \right\} \\ &= \sum_{m_1}^G \sum_{n_1}^G \left| \tilde{\mathbf{X}}(m_1)^H \tilde{\mathbf{X}}(n_1) \right|^2 \sum_{m_2}^G \sum_{n_2}^G \left| \tilde{\mathbf{W}}(m_2)^H \tilde{\mathbf{W}}(n_2) \right|^2 \\ &\quad - \sum_{m_1}^G \left| \tilde{\mathbf{X}}(m_1)^H \tilde{\mathbf{X}}(m_1) \right|^2 \sum_{m_2}^G \left| \tilde{\mathbf{W}}(m_2)^H \tilde{\mathbf{W}}(m_2) \right|^2. \end{aligned} \quad (2.5)$$

As  $\mu^{\text{total}}(\tilde{\mathbf{X}}) = \sum_m^G \sum_{n, n \neq m}^G |\tilde{\mathbf{X}}(m)^H \tilde{\mathbf{X}}(n)|^2$  and  $\mu^{\text{total}}(\tilde{\mathbf{W}}) = \sum_m^G \sum_{n, n \neq m}^G |\tilde{\mathbf{W}}(m)^H \tilde{\mathbf{W}}(n)|^2$ , (2.5) can be rewritten as

$$\mu^{\text{total}}(\bar{\Phi}) = \mu^{\text{total}}(\tilde{\mathbf{X}}) \mu^{\text{total}}(\tilde{\mathbf{W}}) + \mu^{\text{total}}(\tilde{\mathbf{X}}) \nu(\tilde{\mathbf{W}}) + \mu^{\text{total}}(\tilde{\mathbf{W}}) \nu(\tilde{\mathbf{X}}), \quad (2.6)$$

where  $\nu(\tilde{\mathbf{X}}) \triangleq \sum_m^G |\tilde{\mathbf{X}}(m)^H \tilde{\mathbf{X}}(m)|^2$ , and  $\nu(\tilde{\mathbf{W}})$  is similarly defined. As all terms in (2.6) are non-negative, minimizing  $\mu^{\text{total}}(\bar{\Phi})$  is split into four minimization problems for  $\mu^{\text{total}}(\tilde{\mathbf{X}})$ ,  $\mu^{\text{total}}(\tilde{\mathbf{W}})$ ,  $\nu(\tilde{\mathbf{X}})$  and  $\nu(\tilde{\mathbf{W}})$ . As  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{W}}$  are involved in the transmitter and receiver, respectively, the original optimization problem that seemingly jointly optimizes the transmitter and the receiver turns out to be disjoint. In the following subsections, I separately cover optimization problems for the transmitter and the receiver.

## 2.4.2 Transmitter Design

### 2.4.2.1 Minimization of Off-diagonal Terms

The total coherence of  $\tilde{\mathbf{X}}$  is identical to the squared Frobenius norm of  $\tilde{\mathbf{X}}^H \tilde{\mathbf{X}}$  excluding the diagonal. Thus,  $\min \mu^{\text{total}}(\tilde{\mathbf{X}})$  is equivalent to  $\min \|\tilde{\mathbf{X}}^H \tilde{\mathbf{X}} - \mathbf{I}_G\|_F^2$ . A coefficient can multiply the identity matrix to obtain zero diagonal entries in the difference expression, but this will not affect the results because the diagonal entries are fixed as discussed soon. The cost function can be expressed as

$$\begin{aligned} \left\| \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} - \mathbf{I}_G \right\|_F^2 &= \text{Tr} \left( \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} - 2\tilde{\mathbf{X}}^H \tilde{\mathbf{X}} + \mathbf{I}_G \right) \\ &= \text{Tr} \left( \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H - 2\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H + \mathbf{I}_{M_t} \right) + (G - M_t) \\ &= \left\| \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H - \mathbf{I}_{M_t} \right\|_F^2 + (G - M_t). \end{aligned} \quad (2.7)$$

Note that the lower bound on the cost function is  $G - M_t$  where  $G \geq \max(N_t, N_r)$ . By definition of  $\tilde{\mathbf{X}}$ ,  $\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H$  can be rewritten as  $\bar{\mathbf{X}}^T \bar{\mathbf{A}}_t^* \bar{\mathbf{A}}_t^T \bar{\mathbf{X}}^* = \frac{G}{N_t} \bar{\mathbf{X}}^T \bar{\mathbf{X}}^*$ . Due to the transmit power constraint  $\|\bar{\mathbf{x}}_m\|^2 = 1$ , the diagonal elements of  $\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H$  are fixed at  $G/N_t$ . They cannot be minimized without reducing the transmit power. Thus off-diagonal elements are of our interest, and each  $\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H(p, q)$  for  $p \neq q$  can be written as

$$\frac{G}{N_t} \mathbf{x}_p^T \mathbf{F}_p^T \mathbf{F}_q^* \mathbf{x}_q^* = \frac{G}{N_t} \mathbf{x}_p^T \mathbf{F}_{\text{BB},p}^T \mathbf{F}_{\text{RF},p}^T \mathbf{F}_{\text{RF},q}^* \mathbf{F}_{\text{BB},q}^* \mathbf{x}_q^*. \quad (2.8)$$

Note that  $p, q \leq M_t$  and that  $M_t$  is desired to be large to reduce the lower bound shown in (2.7). It leads to a larger matrix for which I would like to minimize the sum of squared off-diagonal elements. Here I have three sets of

degrees of freedom to control the elements: the analog and digital precoders, and the training symbols. I define  $\mathcal{F}_{\text{RF}}$  and  $\mathcal{F}_{\text{BB}}$  as codebooks for possible analog and digital precoders, respectively.

Assuming  $N_t$  is a multiple of  $L_t$ , I define the analog precoder codebook  $\mathcal{F}_{\text{RF}}$  as

$$\mathcal{F}_{\text{RF}} = \{\mathbf{F}_{\text{RF},p} \in \mathbb{C}^{N_t \times L_t} : \forall p, q \in \{1, \dots, N_t/L_t\}, \\ \mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},p} = \mathbf{I}_{L_t}, \mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},q} = \mathbf{0}_{L_t}, p \neq q\}.$$

The first condition of the codebook ( $\mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},p} = \mathbf{I}_{L_t}$ ) is to facilitate further optimization. The second ( $\mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},q} = \mathbf{0}_{L_t}$ ) is to make off-diagonal elements zeros when different analog precoders are used. One obvious example of such  $\mathcal{F}_{\text{RF}}$  is a set of submatrices of which each contains  $L_t$  columns of the  $N_t \times N_t$  discrete Fourier transform (DFT) matrix.

Even if both  $\mathbf{F}_{\text{RF},p}$  and  $\mathbf{F}_{\text{RF},q}$  use the same analog precoder which does not necessarily make  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H(p, q)$  zero, the digital precoder and the training symbol vector can be exploited to minimize such an element. Assuming  $\mathbf{F}_{\text{RF},p} = \mathbf{F}_{\text{RF},q} \in \mathcal{F}_{\text{RF}}$  and  $\mathbf{F}_{\text{BB},p} = \mathbf{U}\Sigma_p\mathbf{V}^H$  by the singular value decomposition (SVD) for all  $p$ , (2.8) can be written as

$$\frac{G}{N_t} \mathbf{x}_p^T \mathbf{F}_p^T \mathbf{F}_q^* \mathbf{x}_q^* = \frac{G}{N_t} \tilde{\mathbf{x}}_p^T \Sigma_p^T \Sigma_q^* \tilde{\mathbf{x}}_q^*, \quad (2.9)$$

where  $\tilde{\mathbf{x}}_p \triangleq \mathbf{V}^H \mathbf{x}_p$  is the transformed training symbol vector,  $\Sigma_p$  is a diagonal matrix with diagonal entries being non-negative real numbers, and  $\mathbf{U}$  and  $\mathbf{V}$  are  $L_t \times L_t$  arbitrary unitary matrices. To make (2.9) zero, I can choose one

of the vectors for the diagonal of  $\Sigma_p$  such that they do not overlap one another. Since the elements in  $\Sigma_p$  are restricted to be non-negative real numbers, a straightforward option is to choose  $L_t$ -dimensional standard bases so that  $\Sigma_p^T \Sigma_q^* = \mathbf{0}_{L_t}$ . The standard bases can be scaled; however, the digital precoders should meet the constraint  $\|\Sigma_p \tilde{\mathbf{x}}_p\|^2 = 1$ . The constraint comes from the transmit power constraint  $\|\mathbf{F}_{\text{RF},p} \mathbf{F}_{\text{BB},p} \mathbf{x}_p\|^2 = 1$  and the analog precoder codebook condition  $\mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},p} = \mathbf{I}_{L_t}$ . It leads  $\tilde{\mathbf{x}}_p$  to be the same standard basis that is used for  $\Sigma_p$  with a reciprocal of the scalar if scaled. It shows that the digital precoder and the training symbols have one-to-one correspondence, that the degrees of freedom I actually have is two, and that  $|\mathcal{F}_{\text{BB}}| = L_t$ . Thus, without loss of generality, I use the standard bases without scaling. Consequently the digital precoder codebook  $\mathcal{F}_{\text{BB}}$  can be expressed as

$$\mathcal{F}_{\text{BB}} = \{\mathbf{F}_{\text{BB},p} \in \mathbb{C}^{L_t \times L_t} : \forall p \in \{1, \dots, L_t\}, \mathbf{F}_{\text{BB},p} = \mathbf{U} \Sigma_p \mathbf{V}^H, \Sigma_p = \text{diag}(\mathbf{e}_p)\},$$

where  $\mathbf{e}_p$  denotes the  $L_t$ -dimensional standard basis with one in the  $p$ -th entry and zero for other elements. For the given  $\mathbf{F}_{\text{BB},p}$ , the corresponding training symbol vector is given as  $\mathbf{x}_p = \mathbf{V} \mathbf{e}_p$ .

As discussed in Section 2.1.2, I have two sets of degrees of freedom in configuring a hybrid precoder: an analog and a digital precoder. The maximum of  $M_t$  that makes all off-diagonal entries of  $\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H$  zeros is  $N_t$  since  $M_t = |\mathcal{F}_{\text{RF}}| |\mathcal{F}_{\text{BB}}| = N_t$ . In other words, there are  $N_t$  available  $\tilde{\mathbf{x}}_p$ 's such that  $\tilde{\mathbf{x}}_p^T \tilde{\mathbf{x}}_q^* = 0$  for all  $p \neq q$ . It is possible to have  $M_t$  greater than  $N_t$  by degrading the total coherence. However, when all combinations of analog and

### Minimization of off-diagonal terms Section 2.4.2.1:

To solve  $\min \mu^t(\tilde{\mathbf{X}}) \rightarrow \min \|\tilde{\mathbf{X}}^H \tilde{\mathbf{X}} - \mathbf{I}_G\|_F^2$

- |  |
|--|
| <p>1. Obtain RF precoders</p> <ul style="list-style-type: none"> <li>• Focusing on off-diagonal elements of <math>\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H</math> <ul style="list-style-type: none"> <li>• Due to transmit power constraint</li> </ul> </li> <li>• Off-diagonal element (<math>p \neq q</math>) <ul style="list-style-type: none"> <li>• <math>\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H(p, q) = G/N_t \mathbf{x}_p^T \mathbf{F}_{\text{BB},p}^T \mathbf{F}_{\text{RF},p}^T \mathbf{F}_{\text{RF},q}^* \mathbf{F}_{\text{BB},q}^* \mathbf{x}_q^*</math></li> </ul> </li> <li>• Obtained RF codebook that makes <math>\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H(p, q)</math> zero</li> </ul> $\mathcal{F}_{\text{RF}} = \{\mathbf{F}_{\text{RF},p} \in \mathbb{C}^{N_t \times L_t} : \forall p, q \in \{1, \dots, N_t/L_t\},$ $\mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},p} = \mathbf{I}_{L_t}, \mathbf{F}_{\text{RF},p}^H \mathbf{F}_{\text{RF},q} = \mathbf{0}_{L_t}, p \neq q\}.$   |
| <p>2. Obtain BB precoders and pilots</p> <ul style="list-style-type: none"> <li>• Focusing on the case where <math>\mathbf{F}_{\text{RF},p} = \mathbf{F}_{\text{RF},q}</math> (<math>p \neq q</math>) <ul style="list-style-type: none"> <li>• Non-zero <math>\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H(p, q)</math></li> </ul> </li> <li>• Off-diagonal element (<math>p \neq q</math>) <ul style="list-style-type: none"> <li>• <math>\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H(p, q) = G/N_t \tilde{\mathbf{x}}_p^T \Sigma_p^T \Sigma_q^* \tilde{\mathbf{x}}_q^*</math></li> </ul> </li> <li>• Obtained BB codebook that makes <math>\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H(p, q)</math> zero</li> </ul> $\mathcal{F}_{\text{BB}} = \{\mathbf{F}_{\text{BB},p} \in \mathbb{C}^{L_t \times L_t} : \forall p \in \{1, \dots, L_t\},$ $\mathbf{F}_{\text{BB},p} = \mathbf{U} \Sigma_p \mathbf{V}^H, \Sigma_p = \text{diag}(\mathbf{e}_p)\},$ <ul style="list-style-type: none"> <li>• Obtained pilots derived from BB codebook</li> </ul> $\mathbf{x}_p = \mathbf{V} \mathbf{e}_p$ |

Figure 2.2: Brief algorithm description of Section 2.4.2.1

digital codebooks are used for training, the number of measurements becomes  $ML_r = M_t M_r L_r = N_t N_r$  which coincides what the LSE requires for  $\Phi^H \Phi$  to be full rank. Explanation regarding  $M_r$  is provided in the following subsection. Using more training frames than LSE needs would be wasteful for CS-based algorithms; hence, I use CS-based algorithms to reduce the number of training frames.

By employing the analog and digital precoder codebooks I propose

**Minimization of diagonal terms Section 2.4.2.2:**

To solve  $\min \nu(\tilde{\mathbf{X}}) \rightarrow \min \sum_m^G |\tilde{\mathbf{X}}(m)^H \tilde{\mathbf{X}}(m)|$

**Derivation:**  $\sum_m^G |\tilde{\mathbf{X}}(m)^H \tilde{\mathbf{X}}(m)| = \text{Tr}(\tilde{\mathbf{X}}^H \tilde{\mathbf{X}}) = \text{Tr}(\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H)$

- $\text{Tr}(\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H)$  cannot be minimized due to transmit power constraint

Figure 2.3: Brief algorithm description of Section 2.4.2.2

above, it can be shown that all diagonal elements of  $\tilde{\mathbf{X}}^H \tilde{\mathbf{X}}$  are one, thus,  $\mu^{\text{total}}(\tilde{\mathbf{X}}) = \|\tilde{\mathbf{X}}^H \tilde{\mathbf{X}} - \mathbf{I}_G\|_F^2$ . Then the total coherence becomes

$$\mu^{\text{total}}(\tilde{\mathbf{X}}) = M_t \left( \frac{G}{N_t} - 1 \right)^2 + (G - M_t). \quad (2.10)$$

Since  $M_t = N_t$ , I have  $\mu^{\text{total}}(\tilde{\mathbf{X}}) = \frac{G}{N_t}(G - N_t)$ . From the equation, it can be seen that 1) the total coherence of  $\tilde{\mathbf{X}}$  quadratically increases with  $G$ , 2) zero total coherence can be achieved with  $G = N_t$ , and 3) more antennas are beneficial in lowering the total coherence not to mention sharp beamforming.

### 2.4.2.2 Minimization of Diagonal Terms

I now focus on minimizing  $\nu(\tilde{\mathbf{X}})$ . As all terms in summation in  $\nu(\tilde{\mathbf{X}})$  are non-negative, the problem  $\min \nu(\tilde{\mathbf{X}})$  is equivalent to  $\min \sum_m^G |\tilde{\mathbf{X}}(m)^H \tilde{\mathbf{X}}(m)|$ , and the cost function can be rewritten as  $\sum_m^G |\tilde{\mathbf{X}}(m)^H \tilde{\mathbf{X}}(m)| = \text{Tr}(\tilde{\mathbf{X}}^H \tilde{\mathbf{X}}) = \text{Tr}(\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H)$ . The cost function is a sum of the diagonal elements in  $\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H$  which, as shown in Section. 2.4.2.1, are constrained by the transmit power. It means that  $\nu(\tilde{\mathbf{X}})$  cannot be minimized without reducing the transmit power.

### 2.4.3 Receiver Design

#### 2.4.3.1 Minimization of Off-diagonal Terms

As with the minimization of  $\mu^{\text{total}}(\tilde{\mathbf{F}})$ ,  $\min \mu^{\text{total}}(\tilde{\mathbf{W}})$  is equivalent to  $\min \|\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} - \mathbf{I}_G\|_F^2$ . As with (2.7), I have the cost function given by  $\|\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} - \mathbf{I}_G\|_F^2 = \|\tilde{\mathbf{W}} \tilde{\mathbf{W}}^H - \mathbf{I}_{M_r L_r}\|_F^2 + (G - M_r L_r)$ , and  $\tilde{\mathbf{W}} \tilde{\mathbf{W}}^H$  can be expressed as

$$\tilde{\mathbf{W}} \tilde{\mathbf{W}}^H = \frac{G}{N_r} \begin{bmatrix} \mathbf{W}_1^H \mathbf{W}_1 & \mathbf{W}_1^H \mathbf{W}_2 & \cdots & \mathbf{W}_1^H \mathbf{W}_{M_r} \\ \mathbf{W}_2^H \mathbf{W}_1 & \mathbf{W}_2^H \mathbf{W}_2 & \cdots & \mathbf{W}_2^H \mathbf{W}_{M_r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{M_r}^H \mathbf{W}_1 & \mathbf{W}_{M_r}^H \mathbf{W}_2 & \cdots & \mathbf{W}_{M_r}^H \mathbf{W}_{M_r} \end{bmatrix}. \quad (2.11)$$

I assume that  $N_r$  is a multiple of  $L_r$ . Denoting  $\mathcal{W}_{\text{RF}}$  and  $\mathcal{W}_{\text{BB}}$  by codebooks for possible analog and digital combiners, respectively, I define the analog combiner codebook as

$$\mathcal{W}_{\text{RF}} = \left\{ \mathbf{W}_{\text{RF},p} \in \mathbb{C}^{N_r \times L_r} : \forall p, q \in \left\{1, \dots, \frac{N_r}{L_r}\right\}, \right. \\ \left. \mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},p} = \mathbf{I}_{L_r}, \mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},q} = \mathbf{0}_{L_r}, p \neq q \right\},$$

for the same reason as with analog precoders. Off-diagonal blocks in (2.11) become zero matrices due to the second condition. Focusing on the diagonal blocks in (2.11), it reduces to the following minimization problem:

$$\min \left\| \frac{G}{N_r} \mathbf{W}_{\text{BB},p}^H \mathbf{W}_{\text{BB},p} - \mathbf{I}_{L_r} \right\|_F^2.$$

With  $\mathbf{W}_{\text{BB},p} = \tilde{\mathbf{U}} \tilde{\Sigma}_p \tilde{\mathbf{V}}^H$  for all  $p$  by the SVD, the cost function can be expressed as

$$\left\| \frac{G}{N_r} \tilde{\mathbf{V}} \tilde{\Sigma}_p^H \tilde{\Sigma}_p \tilde{\mathbf{V}}^H - \mathbf{I}_{L_r} \right\|_F^2 = \left\| \frac{G}{N_r} \tilde{\Sigma}_p^H \tilde{\Sigma}_p - \mathbf{I}_{L_r} \right\|_F^2,$$

where  $\tilde{\Sigma}_p$  is a  $L_r \times L_r$  diagonal matrix with diagonal entries being non-negative real numbers, and  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$  are  $L_r \times L_r$  arbitrary unitary matrices. Note that  $\|\mathbf{W}_{\text{BB},p}\|_F^2$  is not constrained as it does not change receive SNR of the system.

With the constraint  $\tilde{\Sigma}_p$  being a diagonal matrix with non-negative entries, the only optimal solution that makes the cost function is the scaled identity matrix, in other words,  $\tilde{\Sigma}_p = \sqrt{N_r/G}\mathbf{I}_{L_r}$  for all  $p$ . Then the digital combiner codebook can be defined as

$$\mathcal{W}_{\text{BB}} = \left\{ \mathbf{W}_{\text{BB},p} \in \mathbb{C}^{L_r \times L_r} : \forall p \in \{1, \dots, L_r\}, \mathbf{W}_{\text{BB},p} = \sqrt{\frac{N_r}{G}} \tilde{\mathbf{U}} \tilde{\mathbf{V}}^H \right\}. \quad (2.12)$$

As  $\mathbf{W}_{\text{BB},p}$  is identical for all  $p$ , the only degree of freedom in making a hybrid combiner is the analog combiner. Therefore  $M_r = N_r/L_r$ .

With the proposed combiner codebooks, it can be shown that all diagonal elements of  $\tilde{\mathbf{W}}^H \tilde{\mathbf{W}}$  are  $N_r/G$  which is less than or equal to one. It leads to the following inequality:

$$\mu^{\text{total}}(\tilde{\mathbf{W}}) = \|\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} - \frac{N_r}{G} \mathbf{I}_G\|_F^2 \leq \|\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} - \mathbf{I}_G\|_F^2 = G - N_r,$$

where the equality holds if  $G = N_r$ , or equivalently,  $\mu^{\text{total}}(\tilde{\mathbf{W}}) = 0$ . As with the transmitter case, from the above inequality, it can be seen that 1) the upper bound of the total coherence of  $\tilde{\mathbf{W}}$  is proportional to  $G$ , 2) zero total coherence can be achieved when  $G = N_r$ , and 3) using more receive antennas helps to reduce the total coherence and allows finer angle grids.

*Remarks:* As pointed out in Section 2.4.2.1, a coefficient can also be multiplied by the identity matrix when converting the total coherence minimization problem into  $\min \|\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} - \mathbf{I}_G\|_F^2$ . If a coefficient other than one is

### Minimization of off-diagonal terms Section 2.4.3.1:

To solve  $\min \mu^t(\tilde{\mathbf{W}}) \rightarrow \min \|\tilde{\mathbf{W}}^H \tilde{\mathbf{W}} - \mathbf{I}_G\|_F^2$

1. Obtain RF combiners

- Focusing on off-diagonal elements of  $\tilde{\mathbf{W}}\tilde{\mathbf{W}}^H$ 
  - where

$$\tilde{\mathbf{W}}\tilde{\mathbf{W}}^H = \frac{G}{N_r} \begin{bmatrix} \mathbf{w}_1^H \mathbf{w}_1 & \mathbf{w}_1^H \mathbf{w}_2 & \cdots & \mathbf{w}_1^H \mathbf{w}_{M_r} \\ \mathbf{w}_2^H \mathbf{w}_1 & \mathbf{w}_2^H \mathbf{w}_2 & \cdots & \mathbf{w}_2^H \mathbf{w}_{M_r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}_{M_r}^H \mathbf{w}_1 & \mathbf{w}_{M_r}^H \mathbf{w}_2 & \cdots & \mathbf{w}_{M_r}^H \mathbf{w}_{M_r} \end{bmatrix}$$

- Obtained RF codebook that makes off-diagonal blocks zero

$$\mathcal{W}_{\text{RF}} = \{ \mathbf{W}_{\text{RF},p} \in \mathbb{C}^{N_r \times L_r} : \forall p, q \in \{1, \dots, \frac{N_r}{L_r}\} \}$$

$$\mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},p} = \mathbf{I}_{L_r}, \mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},q} = \mathbf{0}_{L_r}, p \neq q \}$$

2. Obtain BB combiners

- Focusing on the diagonal blocks of  $\tilde{\mathbf{W}}\tilde{\mathbf{W}}^H$ 
  - Solve  $\min \|G/N_r \mathbf{W}_{\text{BB},p}^H \mathbf{W}_{\text{BB},p} - \mathbf{I}_{L_r}\|_F^2$
- Obtained BB codebook that makes the diagonal blocks zero

$$\mathcal{W}_{\text{BB}} = \{ \mathbf{W}_{\text{BB},p} \in \mathbb{C}^{L_r \times L_r} : \forall p \in \{1, \dots, L_r\}, \}$$

$$\mathbf{W}_{\text{BB},p} = \sqrt{\frac{N_r}{G}} \tilde{\mathbf{U}} \tilde{\mathbf{V}}^H \}$$

Figure 2.4: Brief algorithm description of Section 2.4.3.1

used, it would be reflected in the scalar of  $\mathbf{W}_{\text{BB},p}$  in (2.12). However, since the different scaling does not affect either the receive SNR and the total coherence, an arbitrary non-zero coefficient instead of  $\sqrt{N_r/G}$  can be used.

### 2.4.3.2 Minimization of Diagonal Terms

As with Section 2.4.2.2,  $\min \nu(\tilde{\mathbf{W}})$  is equivalent to  $\min \sum_m^G |\tilde{\mathbf{W}}(m)^H \tilde{\mathbf{W}}(m)|$ , and its cost function can be expressed as  $\text{Tr}(\tilde{\mathbf{W}}\tilde{\mathbf{W}}^H)$ . Since the analog combiners have a condition  $\mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},p} = \mathbf{I}_{L_r}$  and the digital com-

**Minimization of diagonal terms Section 2.4.3.2:**

To solve  $\min \nu(\tilde{\mathbf{W}}) \rightarrow \min \sum_m^G |\tilde{\mathbf{W}}(m)^H \tilde{\mathbf{W}}(m)|$

**Derivation:**  $\sum_m^G |\tilde{\mathbf{W}}(m)^H \tilde{\mathbf{W}}(m)| = \text{Tr}(\tilde{\mathbf{W}}^H \tilde{\mathbf{W}}) = \text{Tr}(\tilde{\mathbf{W}} \tilde{\mathbf{W}}^H)$

- Solution obtained by posing a temporary constraint  
 $\mathbf{W}_{\text{BB},p} = \sqrt{\alpha/L_r} \tilde{\mathbf{U}} \tilde{\mathbf{V}}^H$
- Only difference from the BB codebook obtained in Section 2.4.3.1 is a coefficient
- An arbitrary non-zero coefficient is acceptable
- Thus, the obtained BB codebook in this section is also a valid solution for Section 2.4.3.1.

Figure 2.5: Brief algorithm description of Section 2.4.3.2

biner codebook has a single element,  $\text{Tr}(\tilde{\mathbf{W}} \tilde{\mathbf{W}}^H)$  reduces to  $G/N_r \text{Tr}(\mathbf{W}_p^H \mathbf{W}_p)$  which is  $G/N_r \text{Tr}(\tilde{\Sigma}_p^H \tilde{\Sigma}_p)$  assuming  $\mathbf{W}_{\text{BB},p} = \tilde{\mathbf{U}} \tilde{\Sigma}_p \tilde{\mathbf{V}}^H$ . A trivial solution to  $\min \text{Tr}(\tilde{\Sigma}_p^H \tilde{\Sigma}_p)$  is  $\tilde{\Sigma}_p = \mathbf{0}_{L_r}$ . Since a zero matrix does not make sense, I pose a temporary constraint such as  $\|\mathbf{W}_{\text{BB},p}\|_F^2 = \alpha > 0$ . Then the problem now becomes

$$\min \text{Tr}(\tilde{\Sigma}_p^H \tilde{\Sigma}_p) \text{ subject to } \|\tilde{\Sigma}_p\|_F^2 = \alpha > 0.$$

The optimal  $\tilde{\Sigma}_p$  can be found as  $\sqrt{\alpha/L_r} \mathbf{I}_{L_r}$ , which leads to  $\mathbf{W}_{\text{BB},p} = \sqrt{\alpha/L_r} \tilde{\mathbf{U}} \tilde{\mathbf{V}}^H$ . Considering the fact that an arbitrary scalar is acceptable, the digital combiner that minimizes  $\nu(\tilde{\mathbf{W}})$  also minimizes  $\mu^{\text{total}}(\tilde{\mathbf{W}})$ .

#### 2.4.4 Alternative Approach

There is an alternative approach that corroborates the arguments in the previous derivation. The total coherence of  $\bar{\Phi}$  can directly be minimized by solving  $\min \mu^{\text{total}}(\bar{\Phi})$ . Since the problem is equivalent to  $\min \|\bar{\Phi}^H \bar{\Phi} - \mathbf{I}_{G^2}\|_F^2$ , the cost function can be written as

$$\begin{aligned} \|\bar{\Phi}^H \bar{\Phi} - \mathbf{I}_{G^2}\|_F^2 &= \text{Tr} \left( \{ \bar{\Phi}^H \bar{\Phi} - \mathbf{I}_{G^2} \} \{ \bar{\Phi}^H \bar{\Phi} - \mathbf{I}_{G^2} \}^H \right) \\ &= \text{Tr} (\bar{\Phi}^H \bar{\Phi} \bar{\Phi}^H \bar{\Phi}) - 2\text{Tr} (\bar{\Phi}^H \bar{\Phi}) + \text{Tr} (\mathbf{I}_{G^2}). \end{aligned} \quad (2.13)$$

The first and the second terms in (2.13) can be rewritten as

$$\begin{aligned} \text{Tr} (\bar{\Phi}^H \bar{\Phi} \bar{\Phi}^H \bar{\Phi}) &= \text{Tr} \left( \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} \otimes \tilde{\mathbf{W}}^H \tilde{\mathbf{W}} \tilde{\mathbf{W}}^H \tilde{\mathbf{W}} \right) \\ &= \text{Tr} \left( \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \tilde{\mathbf{X}} \right) \cdot \text{Tr} \left( \tilde{\mathbf{W}}^H \tilde{\mathbf{W}} \tilde{\mathbf{W}}^H \tilde{\mathbf{W}} \right) \\ &= \left\| \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \right\|_F^2 \cdot \left\| \tilde{\mathbf{W}} \tilde{\mathbf{W}}^H \right\|_F^2, \end{aligned} \quad (2.14)$$

and

$$\text{Tr} (\bar{\Phi}^H \bar{\Phi}) = \left\| \tilde{\mathbf{X}} \right\|_F^2 \cdot \left\| \tilde{\mathbf{W}} \right\|_F^2, \quad (2.15)$$

respectively. We, again, have four non-negative terms to optimize, and two of them are involved with the transmitter and the rest are with the receiver.

For the transmitter, minimizing  $\|\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H\|_F^2$  in (2.14) becomes identical to Section 2.4.2.1 since the diagonal elements are constrained. With the optimal solution, the minimum of  $\|\tilde{\mathbf{X}} \tilde{\mathbf{X}}^H\|_F^2$  is  $(G/N_t)^2 M_t$ . Likewise,  $\|\tilde{\mathbf{X}}\|_F^2$  in (2.15) is fixed as described in Section 2.4.2.2, and  $\min \|\tilde{\mathbf{X}}\|_F^2 = (G/N_t) M_t$ .

For the receiver, (2.13) can be rewritten with the optimal transmitter as

$$\left(\frac{G}{N_t}\right)^2 M_t \left\| \tilde{\mathbf{W}} \tilde{\mathbf{W}}^H \right\|_F^2 - 2 \left(\frac{G}{N_t}\right) M_t \left\| \tilde{\mathbf{W}} \right\|_F^2 + G^2. \quad (2.16)$$

Due to the condition on the analog combiner  $\mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},q} = \mathbf{0}_{L_r}$  for  $p \neq q$ , the off-diagonal blocks in  $\tilde{\mathbf{W}}$  become zero matrices. Therefore, the first term in (2.16) can be written as

$$\begin{aligned} \left\| \tilde{\mathbf{W}} \tilde{\mathbf{W}}^H \right\|_F^2 &= \left(\frac{G}{N_r}\right)^2 \sum_{p=1}^{M_r} \left\| \mathbf{w}_p^H \mathbf{w}_p \right\|_F^2 = \left(\frac{G}{N_r}\right)^2 \sum_{p=1}^{M_r} \left\| \tilde{\Sigma}_p^H \tilde{\Sigma}_p \right\|_F^2 \\ &= \left(\frac{G}{N_r}\right)^2 M_r \sum_{n=1}^{L_r} \sigma_n^4, \end{aligned}$$

where the second equality holds due to  $\mathbf{W}_{\text{RF},p}^H \mathbf{W}_{\text{RF},p} = \mathbf{I}_{L_r}$  and  $\mathbf{W}_{\text{BB},p} = \tilde{\mathbf{U}} \tilde{\Sigma}_p \tilde{\mathbf{V}}^H$ , and  $\sigma_n$  is the  $n$ -th diagonal elements in  $\tilde{\Sigma}_p$ . As for the second term in (2.16),

$$\begin{aligned} \left\| \tilde{\mathbf{W}} \right\|_F^2 &= \frac{G}{N_r} \text{Tr}(\mathbf{W}^H \mathbf{W}) = \frac{G}{N_r} \sum_{p=1}^{M_r} \text{Tr}(\mathbf{w}_p^H \mathbf{w}_p) = \frac{G}{N_r} \sum_{p=1}^{M_r} \left\| \tilde{\Sigma}_p \right\|_F^2 \\ &= \frac{G}{N_r} M_r \sum_{n=1}^{L_r} \sigma_n^2. \end{aligned}$$

Finally, the problem I are solving becomes

$$\min_{\sigma_n^2} \left( \frac{G^4 M_t M_r}{N_t^2 N_r^2} \sum_{n=1}^{L_r} \sigma_n^4 - 2 \frac{G^2 M_t M_r}{N_t N_r} \sum_{n=1}^{L_r} \sigma_n^2 + G^2 \right). \quad (2.17)$$

Defining  $\alpha \triangleq \sum_{n=1}^{L_r} \sigma_n^2 = \|\mathbf{W}_{\text{BB},p}\|_F^2$  for all  $p$ ,  $\alpha$  must be a positive real number for the same reason in Section 2.4.3.2. Using the method of Lagrangian

multipliers, the minimum can be obtained with  $\sigma_n^2 = N_t N_r / G^2$  for all  $n$ . Thus the optimal singular value matrix of  $\mathbf{W}_{\text{BB},p}$  and the resulting  $\mathbf{W}_{\text{BB},p}$  are

$$\tilde{\Sigma}_p = \sqrt{\frac{N_t N_r}{G^2}} \mathbf{I}_{L_r} \text{ and } \mathbf{W}_{\text{BB},p} = \sqrt{\frac{N_t N_r}{G^2}} \tilde{\mathbf{U}} \tilde{\mathbf{V}}^H, \quad (2.18)$$

respectively. As discussed, the different scaling between (2.12) and (2.18) does not affect the receive SNR, the total coherence, and thus performance of CS CE with coherence-based recovery guarantees.

I show that, like the original approach, the alternative one also separates the joint optimization problem into two disjoint problems for the transmitter and the receiver. As for the transmitter, the two approaches converge at some point to yield the identical design whereas they do not for the digital combiner. Nevertheless the optimal solutions for the digital combiner in both approaches coincide except a scalar.

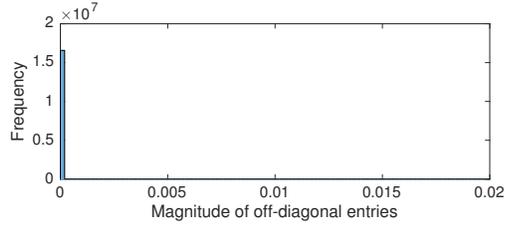
#### 2.4.5 Comparison of total coherence

I compare the the proposed sensing matrix with others [43, 69]. Considering the phase shifting network based hybrid architecture described in Section 2.5.1, histograms of the magnitudes of off-diagonal entries of  $\bar{\Phi}^H \bar{\Phi}$ , where all  $\bar{\Phi}$ 's are column-wise normalized for a fair comparison, are provided in Fig. 2.6. Note that a sum of squared magnitudes of off-diagonal entries composes the total coherence of a matrix under consideration. Subfigures are plotted with different horizontal and vertical ranges for a clear representation of distribution. When phase shifters are used, entries in  $\mathbf{F}_m$  and  $\mathbf{W}_m$  are chosen i.i.d. from a discrete uniform distribution with values  $\{\pm 1, \pm j\}$  in [69] or

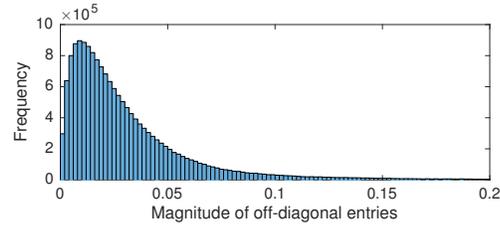
from a set of possible quantized angles in [44]. When with switches, columns of  $\mathbf{F}_m$  and  $\mathbf{W}_m$  are binary vectors with zeros and only one non-zero element, i.e., 1, with its position uniformly random distributed [69]. We, thus, refer to these as the random sensing matrix. Equivalent dictionaries obtained with random selection for a phase shifting network architecture in both [69] and [44] have very similar magnitude distribution and total coherence. Therefore, I only consider random sensing matrices in [69].

In [43], the beamformers are built to minimize the total coherence of the equivalent dictionary; however, it uses a slightly different definition of the total coherence from what I adopt in this chapter, and the derivation therein uses a couple of inequalities for approximation. Despite of the difference, the equivalent dictionary has the same total coherence and histogram as does the proposed one. Therefore the histograms for [43] are omitted from the figure.

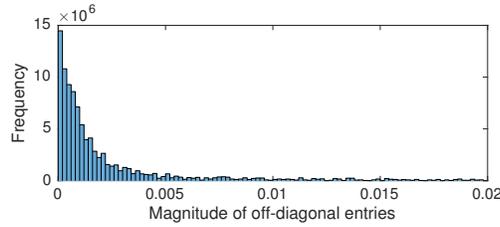
In Fig. 2.6, the proposed sensing matrix is compared to the random one with two different angle grid sizes. The analog precoder and combiner codebooks are built based on the DFT matrix. The histogram will be identical regardless of from which matrix the codebooks are generated, e.g., discrete cosine transform (DCT) and Hadamard matrices. The histograms in Figs. 2.6(a) and 2.6(c) are from the proposed matrix and the rests are from the random matrix in [69]. The grid size is 64 in Figs. 2.6(a) and 2.6(b) and 100 for the rest. Histograms in Figs. 2.6(b) and 2.6(d) have longer tails than do the other histograms. It implies that, regardless of grid angle granularity, the random sensing matrix has greater total coherence and mutual coherence than those



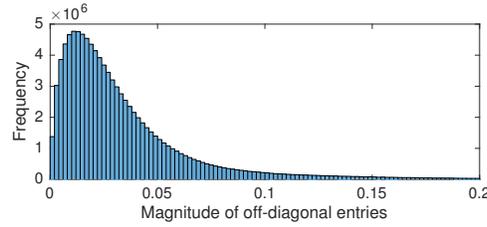
(a) With  $G = 64$  and the proposed sensing matrix



(b) With  $G = 64$  and the random sensing matrix



(c) With  $G = 100$  and the proposed sensing matrix



(d) With  $G = 100$  and the random sensing matrix

Figure 2.6: Histograms of the magnitudes of off-diagonal entries of column-wise normalized  $\bar{\Phi}^H \bar{\Phi}$ . Analog beamforming architectures are based on phase shifters (Section 2.5.1), and the following parameters are used:  $N_t = 64$ ,  $N_r = 16$ ,  $L_t = 8$ ,  $L_r = 4$ , and  $b_{\text{PS}} = 6$ . The proposed sensing matrix with (a)  $G = 64$  and (c)  $G = 100$ , and the random sensing matrix proposed in [69] with (b)  $G = 64$  and (d)  $G = 100$ . The DFT matrix is used in (a) and (c) for the analog precoder and combiner.

of the proposed in this chapter. It also means that the CS algorithm with coherence-based recovery guarantees are likely to perform better with the proposed deterministic sensing matrix than with the random one. Comparing Figs. 2.6(a) with 2.6(c) to conjecture effects of a grid size on performance of the CS algorithms with the proposed deterministic design, a higher  $G$  makes some off-diagonal entries have greater magnitudes. As discussed earlier, it is expected since the total coherence of the equivalent dictionary is shown to increase with  $G$ . The proposed sensing matrix has an outstanding tail in the histogram with  $G = 100$  compared to that with  $G = 64$ ; however, it still has entries more concentrated around zeros than do the random sensing matrices. With  $G = 64$ , the normalized total coherences, i.e.,  $\mu^{\text{total}}(\bar{\Phi})/G^2$ , are 7.2093 and 136.3428 for the proposed sensing matrix and the random one, respectively. With  $G = 100$ , they are 55.5094 and 361.5179, respectively.

It can be assumed that, from a perspective of the total coherence, the larger the grid size becomes the inferior the performance of the CS algorithms would have. However, it is not always true in CE problems such as one I tackle in this chapter since  $G$  represents granularity of angles in a sparsifying dictionary. Namely, large angle grids are required for fine angle resolution, but degrade estimation algorithm performance. This contradictory effects are numerically shown in Section. 2.6.2.

## 2.5 Hybrid Architectures

In this section, I introduce various hybrid architectures and show universal applicability of the proposed sensing matrix design to them. Codebook conditions I derived in the previous section with respect to the analog beamformers are as follows:

- $N_t$  ( $N_r$ ) is a multiple of  $L_t$  ( $L_r$ ).
- The cardinality of the analog precoder (combiner) codebook is  $N_t/L_t$  ( $N_r/L_r$ ).
- $\mathbf{F}_{\text{RF},m}^H \mathbf{F}_{\text{RF},m} = \mathbf{I}_{L_t}$  ( $\mathbf{W}_{\text{RF},m}^H \mathbf{W}_{\text{RF},m} = \mathbf{I}_{L_r}$ ) for all  $m$ .
- $\mathbf{F}_{\text{RF},m}^H \mathbf{F}_{\text{RF},n} = \mathbf{0}_{L_t}$  ( $\mathbf{W}_{\text{RF},m}^H \mathbf{W}_{\text{RF},n} = \mathbf{0}_{L_r}$ ) for all  $m \neq n$ .

In all, combining all conditions provides that  $\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} = \mathbf{I}_{N_t}$  where  $\mathbf{F}_{\text{RF}} \triangleq [\mathbf{F}_{\text{RF},1}, \dots, \mathbf{F}_{\text{RF},N_t/L_t}]$ . Any hybrid beamforming architecture whose analog precoders/combiners satisfy the conditions above can directly adopt the proposed deterministic configuration.

Promising architectures considered for hybrid beamforming are based on either networks of variable phase shifters, networks of switches or a DLA. I consider two architectures with phase shifters, four with switches [69], and one with a DLA which are illustrated in Fig. 2.7. In this section, I only discuss analog precoders, but it also similarly applies to analog combiners. I define a set of feasible analog precoding vectors  $\mathcal{F} \subset \mathbb{C}^{N_t \times 1}$  with different subscripts

for different architectures. In this section, DFT matrix is mainly discussed; however, other matrices such as DCT and Hadamard matrices that satisfy the unitarity condition will work as well.

### 2.5.1 Phase Shifting Network

In this architecture, each RF chain is connected to all antennas via a network of phase shifters as shown in Fig. 2.7(a). Each network is composed of  $N_t$  variable phase shifters, and there exists  $L_t$  such networks. In total,  $N_t L_t$  variable phase shifters are used in this architecture. Due to the hardware constraint of phase shifters, the set of feasible analog precoding vectors is given as

$$\mathcal{F}_1 = \left\{ \mathbf{f} \in \mathbb{C}^{N_t} : |\mathbf{f}_i| = \sqrt{\frac{1}{N_t}}, \angle \mathbf{f}_i \in \Theta \right\},$$

where  $\Theta \triangleq \{\theta : \theta = 2\pi n/2^{b_{\text{PS}}}, n = 0, \dots, 2^{b_{\text{PS}}} - 1\}$  is the possible quantized angle set. Columns of the normalized DFT matrix with a size of  $N_t \times N_t$  can be shown to be in  $\mathcal{F}_1$ . In this case, the cardinality  $|\mathcal{F}_1|$  is  $N_t$ , and  $\mathbf{F}_{\text{RF}}$  with columns being all  $\mathbf{f}$ 's  $\in \mathcal{F}_1$  satisfies  $\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} = \mathbf{I}_{N_t}$ . Note that angles of elements in the DFT matrix can be expressed with  $\Theta$  without quantization if  $b_{\text{PS}} \geq \log_2 N_t$ . Otherwise, angles are projected to the nearest possible angles, which should degrade CE performance.

### 2.5.2 Phase Shifting Network in Subsets

As illustrated in Fig. 2.7(b), a subset of antennas (i.e.,  $N_t/L_t$  antennas) is connected to an RF chain via a network of phase shifters, and there are  $L_t$

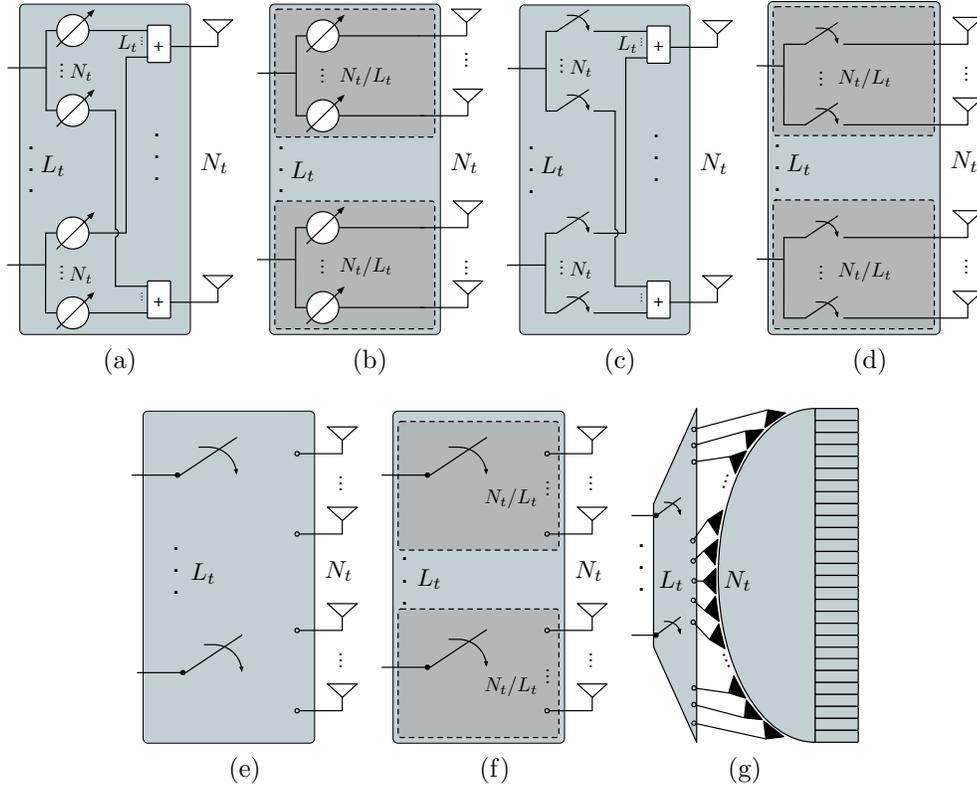


Figure 2.7: Seven types of analog precoders. Variable phase shifters are employed in (a) and (b), a DLA is used in (g), and the rests are based on switches. (a) Phase shifting network, (b) Phase shifting network in subsets, (c) Switching network with analog precoding, (d) Switching network in subsets with analog precoding, (e) Switching network, (f) Switching network in subsets, and (g) CAP-MIMO.

such subsets. The total number of phase shifters, therefore, is  $N_t$ . It can also be considered as duplication of an analog beamforming structure. As non-zero elements in precoding vectors for different subsets do not overlap with each other, columns in  $\mathbf{F}_{\text{RF},m}$  are chosen from  $\mathcal{F}_{2,1}$ ,  $\mathcal{F}_{2,2}$ ,  $\dots$  and  $\mathcal{F}_{2,L_t}$ , respectively, where

$$\mathcal{F}_{2,n} = \left\{ \mathbf{f} \in \mathbb{C}^{N_t} : \mathbf{f} = \left[ \mathbf{0}_{\frac{N_t}{L_t}(n-1)}^\top, \bar{\mathbf{f}}^\top, \mathbf{0}_{N_t - \frac{N_t}{L_t}n}^\top \right]^\top, \right. \\ \left. \bar{\mathbf{f}} \in \mathbb{C}^{N_t/L_t}, |\bar{\mathbf{f}}_i| = \sqrt{\frac{L_t}{N_t}}, \angle \bar{\mathbf{f}}_i \in \Theta \right\}.$$

As vectors are drawn from different sets, the vectors are orthonormal, and so is  $\mathbf{F}_{\text{RF},m}$ . For the same reason in Section 2.5.1, the columns of the normalized  $(N_t/L_t) \times (N_t/L_t)$  DFT matrix can be considered for  $\bar{\mathbf{f}}$ 's, which satisfies the condition  $\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} = \mathbf{I}_{N_t}$ . As with the phase shifting network, the angle quantization error is zero if the phase shifters have high enough quantization resolution.

### 2.5.3 Switching Network with Analog Precoding

This architecture is identical to the phase shifting network considered in Section 2.5.1, but the phase shifters are replaced with switches as shown in Fig. 2.7(c). Thus, the number of switches used in this architecture is  $N_t L_t$  as with the phase shifting network. As splitters in the analog beamformers are still in use, magnitudes of precoding vector elements are constrained. The feasible set of such vectors is given as

$$\mathcal{F}_3 = \left\{ \mathbf{f} \in \mathbb{R}^{N_t} : \mathbf{f} = \sqrt{\frac{1}{N_t}} \mathbf{b}, \mathbf{b} \in \mathcal{B}^{N_t} \right\}.$$

where  $\mathcal{B}$  is the binary set  $\{0,1\}$ . Unlike the previous phase shifter based architectures, the  $\mathcal{L}_2$  norm of a precoding vector  $\|\mathbf{f}\|_2$  is not fixed at 1, but varies from 0 to 1 depending on how many switches are on state due to the presence of splitters in this architecture. If a set of the standard bases of  $\mathbb{R}^{N_t}$  is considered for  $\mathbf{b}$ , then I have  $\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} = 1/N_t \mathbf{I}_{N_t}$ . It seemingly reduces the transmit power because of scaling; however, a corresponding digital precoder or a training symbol vector can make compensation to maintain the constant transmit power. Consequently, the switching network with analog precoding architecture satisfies the condition.

#### 2.5.4 Switching Network in Subsets with Analog Precoding

This architecture is illustrated in Fig. 2.7(d) and is analogous to the one considered in Section 2.5.2 with phase shifters replaced with  $N_t$  switches. Similarly the columns of  $\mathbf{F}_{\text{RF},m}$  are from  $\mathcal{F}_{4,1}$ ,  $\mathcal{F}_{4,2}$ ,  $\dots$  and  $\mathcal{F}_{4,L_t}$ , respectively, where each set is given as

$$\mathcal{F}_{4,n} = \left\{ \mathbf{f} \in \mathbb{R}^{N_t} : \mathbf{f} = \left[ \mathbf{0}_{\frac{N_t}{L_t}(n-1)}^\top, \sqrt{\frac{L_t}{N_t}} \bar{\mathbf{b}}^\top, \mathbf{0}_{N_t - \frac{N_t}{L_t}n}^\top \right]^\top, \bar{\mathbf{b}} \in \mathcal{B}^{N_t/L_t} \right\}.$$

The  $\mathcal{L}_2$  norm of  $\mathbf{f}$  also varies from 0 to 1. As with the case in Section 2.5.2, precoding vectors from different sets are orthogonal. Using a set of the standard bases of  $\mathbb{R}^{N_t/L_t}$  for  $\bar{\mathbf{b}}$ , I have  $\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} = L_t/N_t \mathbf{I}_{N_t}$ , and the scaling can also be compensated by baseband processing.

### 2.5.5 Switching Network

The switching network architecture gets rid of splitters and combiners to connect each RF directly to one of the antennas via a switch as shown in Fig. 2.7(e). Therefore, at a given moment, the number of active antennas is  $L_t$ . The feasible set of precoding vectors can be expressed as

$$\mathcal{F}_5 = \{ \mathbf{f} \in \mathcal{B}^{N_t} : \|\mathbf{f}\|_0 = 1 \}.$$

Due to the absence of splitters and combiners, the  $\mathcal{L}_2$  norm of  $\mathbf{f}$ 's  $\in \mathcal{F}_5$  is fixed at one. Without compensation from baseband processing, the conditions are satisfied with a set of standard bases of  $\mathbb{R}^{N_t}$  considered for the precoding vectors.

### 2.5.6 Switching Network in Subsets

This architecture is similar to the switching network; however, an RF chain can only be connected with one of the antennas in a corresponding subset through a switch. The architecture is illustrated in Fig. 2.7(f). Each subset consists of  $N_t/L_t$  antennas, and only one of them can be active at a given moment. Like the previous subset cases, an analog precoder draws columns from each of the following sets:

$$\mathcal{F}_{6,n} = \left\{ \mathbf{f} \in \mathcal{B}^{N_t} : \mathbf{f} = \left[ \mathbf{0}_{\frac{N_t}{L_t}(n-1)}^\top, \bar{\mathbf{b}}^\top, \mathbf{0}_{N_t - \frac{N_t}{L_t}n}^\top \right]^\top, \bar{\mathbf{b}} \in \mathcal{B}^{N_t/L_t}, \|\bar{\mathbf{b}}\|_0 = 1 \right\}.$$

Again, the  $\mathcal{L}_2$  norm of all precoding vectors is always one, and a set of standard bases of  $\mathbb{R}^{N_t/L_t}$  for  $\bar{\mathbf{b}}$  satisfies the conditions.

### 2.5.7 Continuous Aperture Phased MIMO

Continuous aperture phased MIMO (CAP-MIMO) that directly exploits beamspace MIMO communications is enabled by high resolution DLAs [70]. A transmitter architecture of a 1D DLA is illustrated in Fig. 2.7(g). Feed antennas can ideally generate mutually orthogonal beams, and  $L_t$  antennas out of  $N_t$  are activated by selecting desired antennas and feeding input streams. Namely, the analog precoding matrix  $\mathbf{F}_{\text{RF}}$  has mutually orthonormal columns, and  $L_t$  columns are used for analog beamforming. To this end, DLAs are designed in order to make the analog precoding matrix approximate the DFT matrix. Considering the ideal precoding matrix, the set of analog precoding vectors can be expressed as

$$\mathcal{F}_7 = \{\mathbf{f} \in \mathbb{C}^{N_t} : \mathbf{f} = \mathbf{U}_{dft}(i), i = 1, 2, \dots, N_t\}, \quad (2.19)$$

where  $\mathbf{U}_{dft}$  denotes the  $N_t$ -point DFT matrix. Since the DFT matrix is unitary, the analog precoding matrix obviously satisfies  $\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}} = \mathbf{I}_{N_t}$ .

### 2.5.8 Remarks

For all seven architectures I discussed, a set of analog precoding vectors that satisfies the unitarity condition can be found even though transmit power compensation from baseband processing is occasionally required. As long as the used analog precoders satisfy the condition, the resulting total coherences of  $\bar{\Phi}$  are identical, and so is the CE performance using certain CS algorithms regardless of hybrid beamforming architectures.

## 2.6 Numerical Results

In this section, performance of CE based on the proposed sensing matrix is evaluated using OMP and BPDN algorithms as representatives of CS algorithms with coherence-based recovery guarantees [83]. Both the normalized mean squared error (NMSE) and the achievable SE are used as performance metrics where NMSE is defined as  $\mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2 / \|\mathbf{H}\|_F^2]$ . I provide results obtained with the random sensing matrix in addition to the one I propose. For comparison purposes, the LSE and the oracle estimator are evaluated as well. The LSE evaluated in this section is based on the proposed sensing matrix. The oracle estimator refers to the LSE with actual AoAs and AoDs known at the receiver. Among many HB architectures, *phase shifting network* in Section 2.5.1, *switching network* in Section 2.5.5, and *DLA with adaptive selecting network* proposed in [75] are employed for point-to-point CE in this section. The system is equipped with  $N_t = 64$ ,  $N_r = 16$ ,  $L_t = 8$ ,  $L_r = 8$ ,  $G = 64$ ,  $N_p = 4$  and  $b_{\text{PS}} = 6$  for simulation unless otherwise specified. Table 2.3 pro-

Table 2.3: Simulation parameters

Notation	Description	Value
$N_t$	Number of transmit antennas	64
$N_r$	Number of receive antennas	16
$L_t$	Number of transmit RF chains	8
$L_r$	Number of receive RF chains	2, 8
$G$	Angle grid size	64, 80, 100, 120, 140, 160, 180
$N_p$	Number of channel MPC	1 to 15

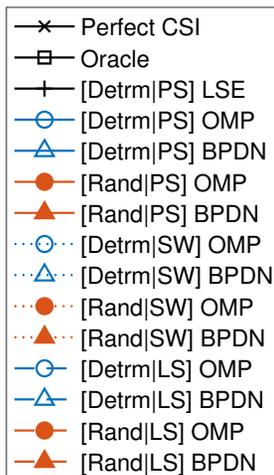


Figure 2.8: The above legend is for the plots in Fig. 2.9–2.12. In the HB architectures, PS, SW and LS refer to phase shifters, switches and lens, respectively. ‘Detrm’ and ‘Rand’ refer to deterministic and random designs, respectively.

vides a list of all simulation parameters used in this section. The channel has a bandwidth of 500 MHz, and the AoDs and AoAs of the multipath components are not constrained to be lied on the angle grids of the dictionary. In the following simulations, 500 channel realizations are used for each point, and the system parameters could be different for a clear comparison. Fig. 2.8 shows the legend for all figures henceforth in this section.

### 2.6.1 Receive RF Chains

Fig. 2.9 shows NMSE of channel estimates as a function of SNR obtained with the OMP and BPDN. Using both algorithms, the proposed and random sensing matrices are evaluated. The number of receive RF chains is either two or eight, and the grid size is either 64 or 180. In all subfigures, the

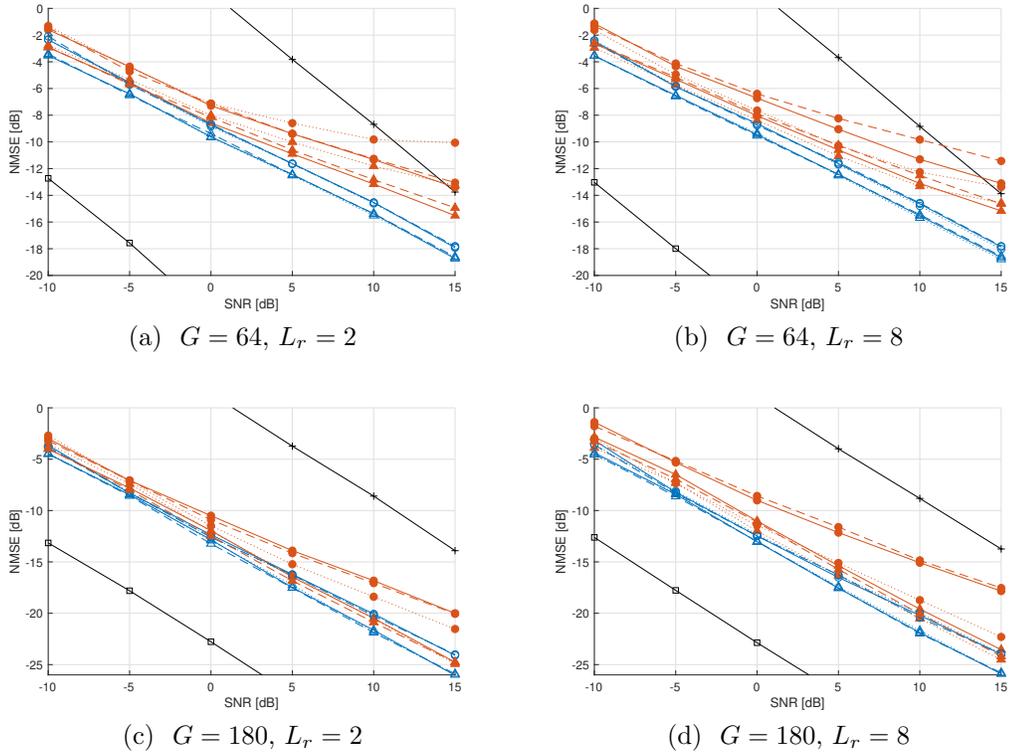


Figure 2.9: NMSE is evaluated with combinations of estimation algorithms and deterministic/random HB designs. Refer to Fig. 2.8 for the legend. OMP and BPDN are used as well as LSE and the oracle estimator. The grid size  $G$  is 64 for (a) and (b), and 180 for (c) and (d). The number of receive RF chains  $L_r$  is two for (a) and (c), and eight for (b) and (d).

oracle estimator outperforms the others and scales well with SNR. It can be seen that NMSE obtained with the LSE and the oracle estimator does not vary depending on the grid size. The reason is that the LSE only uses the sensing matrix  $\Phi$  that does not contain the array response matrices to directly estimate the vectorized channel matrix. As for the oracle estimator, the receiver is already given information regarding AoAs and AoDs and only estimates path

gains. The oracle estimator, therefore, serves as a lower-bound in terms of NMSE.

All plots in Fig. 2.9 show that the proposed deterministic design outperforms the random design for all considered combinations of  $G$  and  $L_r$  and for both OMP and BPDN across the entire SNR range. Performance of the OMP and BPDN with the random sensing matrix, in general, is between the proposed design and the LSE. It is expected because the equivalent dictionary has higher total coherence than the proposed one does. The proposed deterministic design makes the various HB architectures achieve very similar performance compared with the random design since they generate equivalent dictionaries with the identical total coherence as shown in Section 2.5. It can be seen by observing the maximum NMSE variations across the SNR range in systems considered in Fig. 2.9. The worst case of the both designs is with (OMP,  $G = 180$  and  $L_r = 8$ ): 6.52% and 55.81% variations for the deterministic and the random designs, respectively.

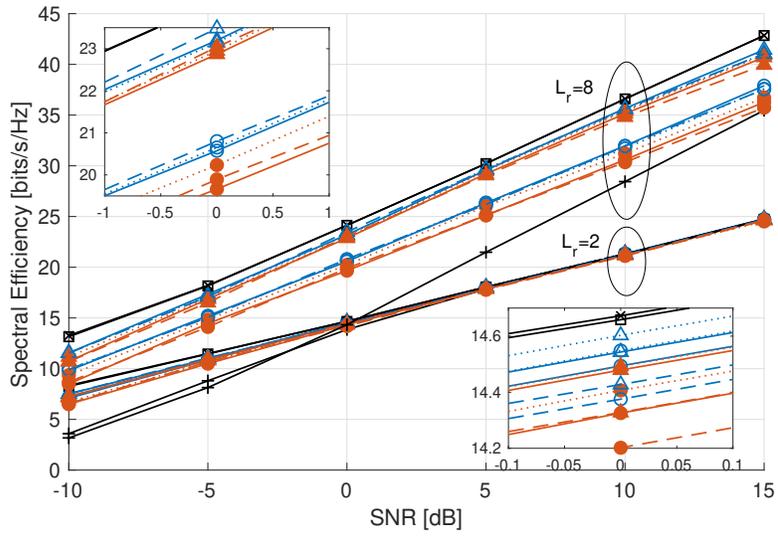
It is seen that performance is not affected by the number of the receive RF chains. It is due to the fact that the total number of frames  $M$  is defined as  $N_t N_r / L_r$ . Therefore with fewer number of receive RF chains,  $M$  increases to maintain the estimation performance, e.g.,  $M = 128$  for  $L_r = 8$  and  $M = 512$  for  $L_r = 2$ . For a fair comparison, the same numbers of frames are used for both deterministic and random designs. Employment of more receive RF chains is beneficial for the proposed deterministic design regardless of  $G$  values since doubling RF chains halves the number of frames while maintaining CE

performance. The case, however, is not the same for the random design, and the switching network HB architecture can only benefit from more RF chains. For  $G = 64$  (comparing Fig. 2.9(a) with 2.9(b)), only the switch-based beamformers can lower NMSE with both OMP and BPDN. For  $G = 180$  (comparing Fig. 2.9(c) with 2.9(d)), only the switch-based beamformers with OMP achieves lower NMSE, and the rests end up with similar or higher errors.

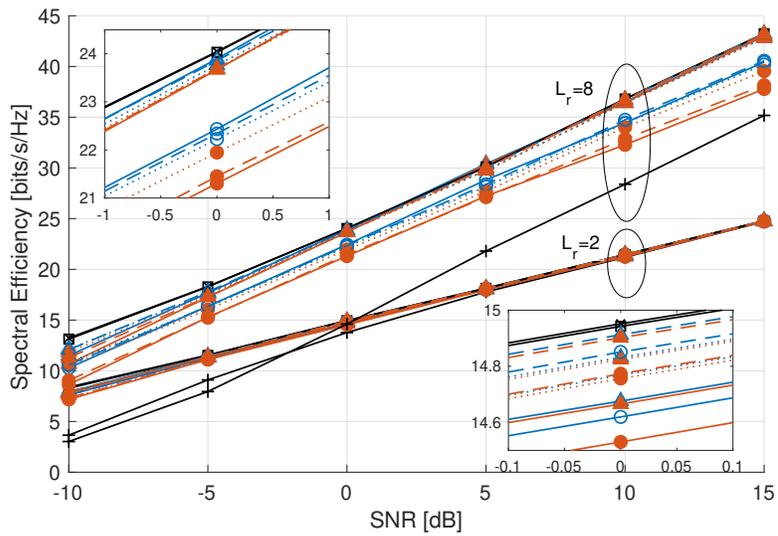
An increase in the grid size is favorable for both deterministic and random approaches since larger grids mean finer resolution in the angle search spaces. It can be seen by comparing Fig. 2.9(a) with 2.9(c) and 2.9(b) with 2.9(d) that all curves of the CS algorithms are shifted down. The performance gaps between the deterministic and random configurations become narrower with the larger grid size. The effects of the grid size are discussed in details in the following section.

Spectral efficiencies for various combinations of architectures and estimation algorithms are plotted in Fig. 2.10. The SE using the perfect channel state information (CSI) is provided in the figure as the performance upper-bound. Two zoom-in plots around 0 dB SNR are layered for better visibility. In both Figs. 2.10(a) and 2.10(b), the oracle estimator yields the indistinguishable SE from the perfect CSI since NMSE is very low across the SNR range as shown in Fig. 2.9. Therefore the perfect CSI and the oracle estimator serve as the upper-bound and the LSE as the lower-bound in terms of SE.

From the perspective of SE, the proposed deterministic design outperforms the random. On the systems equipped with two receive RF chains, SE



(a)  $G = 64$



(b)  $G = 180$

Figure 2.10: Spectral efficiencies as a function of SNR. The number of receive RF chains  $L_r$  is either four or eight, and the grid size  $G$  is (a) 64 or (b) 180. Estimated channels from LSE, OMP, BPDN and the oracle estimator are used as well as the perfect CSI. Refer to Fig. 2.8 for the legend.

curves are very close to each other in both subfigures in Fig. 2.10, and they can only be distinguished in the bottom right zoom-in plots. In those zoom-in plots, with tiny gaps, the proposed configuration achieves greater SE than the random one for all combinations of architectures and algorithms. With eight receive RF chains, the gaps between the two configurations are larger than with two RF chains. This different gaps are because the number of channel paths ( $N_p$ ) is fixed at four. For this reason, eight receive RF chains will be considered for the rest of the chapter for better visibility. Simulation results for various numbers of channel paths are provided in Section 2.6.3.

The spectral efficiencies also tend to converge for the various HB architectures by using the deterministic design. The trend holds regardless of the grid size, the number of receive RF chains and the CS algorithms. In terms of the maximum SE variations across the SNR range, the worst case of the deterministic design is 3.41% with (BPDN,  $G = 180$  and  $L_r = 8$ ), and that of the random design is 11.93% with (OMP,  $G = 180$  and  $L_r = 8$ ).

Comparing two subfigures in Fig. 2.10, the larger grids help CS algorithms to achieve higher SE as with Fig. 2.9, and the performance gaps between the deterministic and random configurations become smaller. Especially in Fig. 2.10(b) for  $L_r = 2$ , the random configuration almost catches up with the deterministic for most cases while there still are noticeable gaps for the case of  $L_r = 8$ .

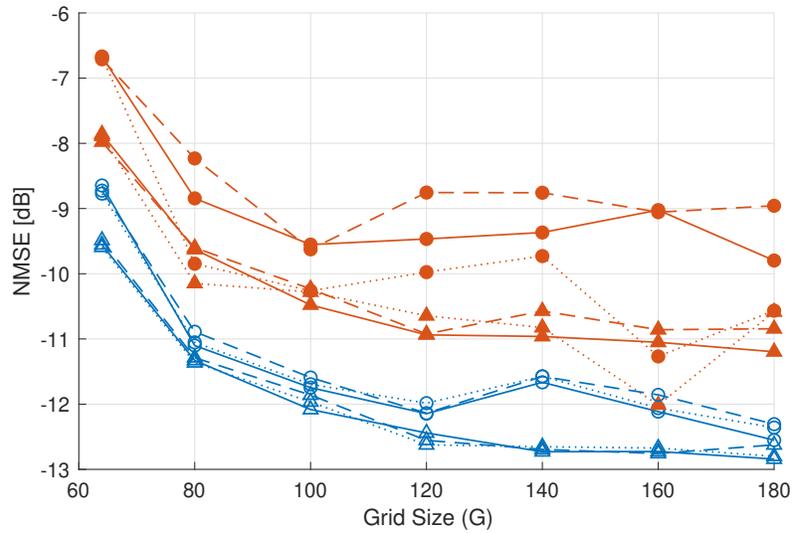
From comparison between Figs. 2.9 and 2.10, the performance order in NMSE does not always match the order in SE. For example, the order in

Fig. 2.9(b) is (1) BPDN with the deterministic, (2) OMP with the deterministic, (3) BPDN with the random and (4) OMP with the random. However, the performance order in Fig. 2.10(a) for  $L_r = 8$  is (1) BPDN with the deterministic, (2) BPDN with the random, (3) OMP with the deterministic and (4) OMP with the random. This mismatch implies that NMSE alone might not be the best metric and that SE should be considered as well. It also could be concluded that BPDN outperforms OMP; however, comparison of CE algorithms is beyond the scope of this chapter.

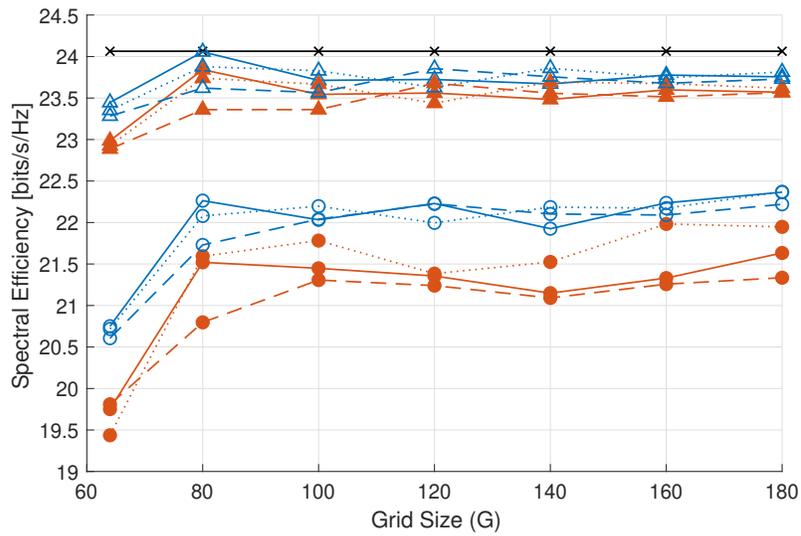
### 2.6.2 Granularity of Angle Grids in Dictionary

In the previous subsection, I take a glance at the effects of the grid size. It is natural to consider a dictionary matrix with finer angle granularity since the finer the angle grids becomes the more accurate estimated AoAs and AoDs are expected to be. At the same time, however, the grid size is desired to be low because computational complexity of CS algorithms is generally a polynomial function of  $G$ . For example, computational complexity of OMP and BPDN are  $O(ML_rG^2)$  and  $O(G^6)$ , respectively. Therefore a discussion in this section can be a guide to finding a good trade-off between computational complexity and performance.

Fig. 2.11 shows NMSE and achievable SE as a function of  $G$  obtained with combinations of two estimation algorithms, two beamformer configurations and three architectures. See Fig. 2.8 for the legend. SNR is fixed at 0 dB in the figures. The oracle estimator and LSE are excluded in the figure for



(a)



(b)

Figure 2.11: (a) NMSE and (b) SE as a function of the grid size ( $G$ ). Eight receive RF chains ( $L_r = 8$ ) is used, and SNR is fixed at 0 dB in this figure. Refer to Fig. 2.8 for the legend.

the sake of visibility of the considered algorithms.

From Fig. 2.11(a), NMSE for all combinations generally decreases with increase in  $G$  as expected and saturates with some fluctuation. Larger dictionary or finer angle granularity is beneficial for both random and deterministic configurations; however, performance gain obtained from the finer granularity in the angle grids is more stable for the proposed deterministic configuration. Even though the overall performance gains for both configurations are similar to each other, NMSE of the deterministic configuration is always lower than that of the random due to the lower starting points. To achieve the same NMSE, the deterministic is far more efficient in terms of computational complexity than the random since the deterministic configuration already achieves lower NMSE with  $G = 80$  than does the random configuration with  $G = 180$ .

SE curves in Fig. 2.11(b) also show similar trends that are observed in the NMSE plot. As  $G$  increases, spectral efficiencies of all combinations generally increase. The deterministic configurations achieve similar or greater spectral efficiencies with  $G = 80$  than do the random configurations with greater grid sizes.

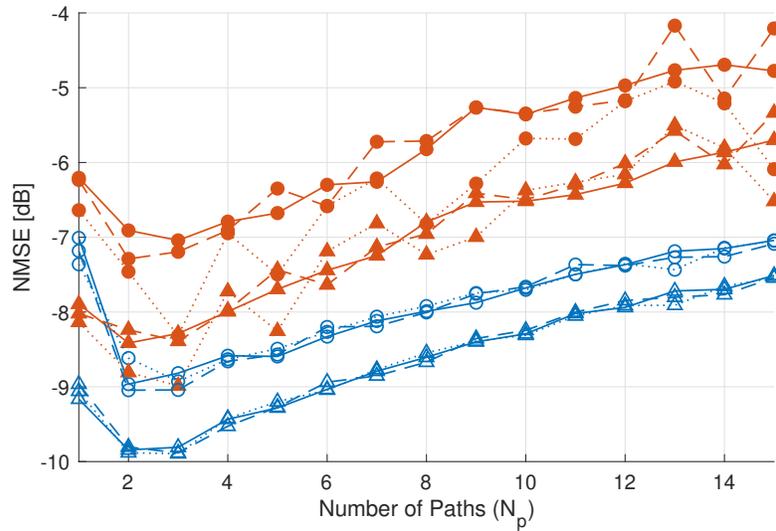
The mismatch of performance orders between NMSE and SE can also be observed as discussed in Section 2.6.1. Another mismatch that can be noticed is that performance improvement obtained with finer angle granularity is greater in NMSE than in SE. For all curves in Fig. 2.11(b), SE saturates around  $G = 80$  or  $100$ , and larger  $G$  values do not lead to significant improvement whereas NMSE decreases can be observed until  $G$  becomes  $140$  or

### 2.6.3 Channel Sparsity

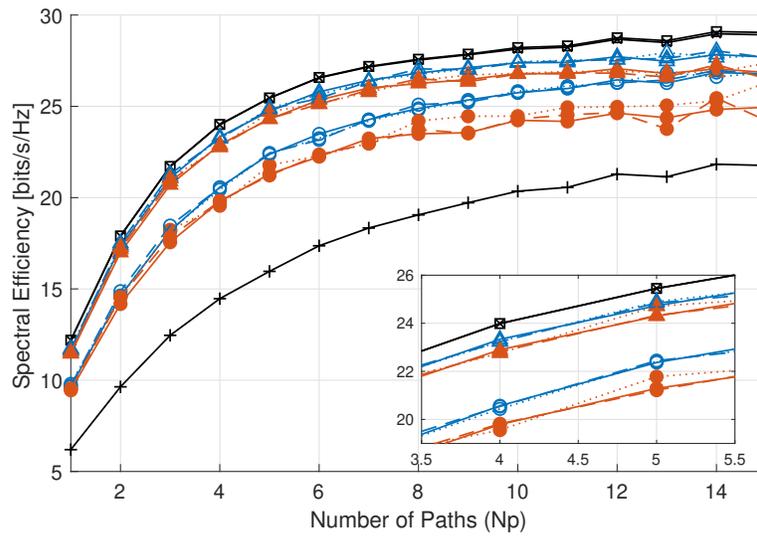
The number of channel paths – equal to rank of the channel – that can be recovered by CS algorithms is not unlimited. With noiseless measurements, it is bounded in terms of the mutual coherence of the sensing matrix as shown in [48]. With more channel paths to estimate, or equivalently, with less sparsity of a channel vector to estimate, estimation accuracy of CS algorithms tends to degrade.

Fig. 2.12(a) shows NMSE as a function of  $N_p$ . As more channel paths are involved, CE error of both OMP and BPDN algorithms generally increases except for the rank one channel case. NMSE with the deterministic configuration, however, is always lower than that with the random configuration across the entire  $N_p$  range. Furthermore, with the random configuration, more fluctuation is observed, and an NMSE increase rate is greater than the deterministic. From the CE error perspective, the deterministic configuration always achieves better performance.

SE curves are provided in Fig. 2.12(b) along with a zoom-in plot around  $N_p = 4$  and 5. For all curves in Fig. 2.12(b), SE increases with  $N_p$  unlike what NMSE curves imply. It is natural because greater  $N_p$  means greater channel rank and possibility of higher degree spatial multiplexing. Since there are eight RF chains at both transmitter and receiver ( $L_t = L_r = 8$ ), SE rises at a faster rate until the point of  $N_p = 8$  and at a slower rate after that.



(a)



(b)

Figure 2.12: (a) NMSE and (b) SE as a function of the number of channel paths ( $N_p$ ). Eight receive RF chain ( $L_r = 8$ ) is used, and SNR is fixed at 0 dB. Refer to Fig. 2.8 for the legend.

The deterministic configuration helps the CS algorithms achieve greater SE. In the zoom-in plot in Fig. 2.12(b), it is seen, regardless of architecture types and algorithms, that the proposed deterministic configuration achieves greater SE than does the random. It is noted that the performance order mismatch that I discuss in previous sections also happens in this section.

## 2.7 Conclusion

In this chapter, I proposed a deterministic design of hybrid beamformers for CS based CE in narrowband mmWave communication systems. The deterministic sensing matrix design I proposed works for a variety of hybrid beamforming architectures that are implemented with variable phase shifters, switches or a DLA. Our design approach is to configure analog and digital beamformers by minimizing the total coherence of the equivalent sparsifying dictionary. I decoupled the joint transmitter and receiver optimization problem into two disjoint problems. The analog and digital beamformer codebooks that are obtained by solving the optimization problems improve CE performance of the CS algorithms that rely on coherence to guarantee sparse recovery and were shown by simulation to outperform, in terms of both estimation error and SE, random configurations with various numbers of RF chains, grid sizes and channel ranks. The obtained codebooks also provides consistent CE performance across the HB architectures with up to 6.52% maximum NMSE difference compared to 55.81% for the random design. The proposed deterministic design is particularly beneficial when coarse angle grids are considered

for computation reduction.

In the following chapter, I tackle the codebook design problem again from a different perspective. In addition, non-full training is explicitly considered, and the determinism is further expanded.

## Chapter 3

# Hybrid Beamformer Codebook Design and Ordering for Compressive mmWave Channel Estimation

Separating all-digital beamforming into analog and digital stages can provide the large reduction in power consumption needed for practical implementations with a tradeoff of a small loss in spectral efficiency. Developing algorithms with this favorable tradeoff is challenging due to the additional degrees of freedom in the analog stage and its accompanying hardware constraints. In hybrid beamforming systems, for example, channel estimation algorithms do not directly observe the channels, face a high channel count, and operate at low SNR before transmit-receive beam alignment. Since mmWave channels are sparse in time and beam domains, many compressed sensing (CS) channel estimation algorithms have been developed that randomly configure the analog beamformers, digital beamformers, and/or pilot symbols.

For hybrid beamforming mmWave communication systems, this chapter<sup>1</sup> proposes a deterministic codebook design method for beamformers and

---

<sup>1</sup> This chapter is based on the work published in J. Sung and B. L. Evans, “Hybrid beamformer codebook design and ordering for compressive mmWave channel estimation”, in *Proceedings of IEEE International Conference on Computing, Networking and Commu-*

pilot symbols for open-loop channel estimation. The goal is to reduce the number of pilot symbols to reduce the training overhead and delay in estimating the channel state information (CSI) for the connection. The approach builds on the compressive sensing-based channel estimation framework and algorithms developed in Chapter 2. The CS-based algorithms rely on low coherence for their signal recovery guarantees. This chapter seeks to reduce the training overhead from the full training used in Chapter 2 by proposing a precoder column ordering to further remove randomness in the design. By reducing the transmitted and received training pilot symbols, the number of measurements for CS is also reduced. Thus both the training symbol transmission/reception time and the CS algorithms' processing time can be reduced. Simulation results show that our beamformer designs reduce channel estimation error over competing methods.

### 3.1 Introduction

Hybrid analog and digital beamforming architectures in millimeter wave (mmWave) communication systems have drawn a great amount of attention for multiple reasons. In practice, they can maintain achievable spectral efficiency as with all-digital MIMO architectures due to the sparse nature of mmWave channels [18, 22]. Reducing the number of RF chains leads to proportional power consumption reduction. Hybrid beamforming architectures,

---

*nications (ICNC)*, Feb. 17–20, 2020, Big Island, HI, USA [89]. This work was supervised by Prof. Brian L. Evans.

however, demand even more complicated signal processing [65] because an analog stage has fewer degrees of freedom compared with an all-digital MIMO architecture; e.g., phase shifters are constrained to have a discrete phase on the unit circle. As an example, the signal processing for channel estimation becomes significantly more complicated..

MmWave channel measurement campaigns have revealed that the channels are sparse in both time and angular dimensions [59,60]. The sparsity finds compressed sensing (CS) algorithms suitable for mmWave channel estimation. For phase shifter based hybrid beamforming architecture, the adaptive CS was proposed in [19] to narrow beams by iteratively adapting precoders and combiners. Another popular approach found in many publications is to formulate a sparse channel estimation problem and apply well-known or modified CS algorithms [44, 69, 74–76]. For example, orthogonal matching pursuit (OMP) or its variants are used in [44, 76], and [69] compares different channel estimation algorithms that include OMP as a representative of CS algorithms. The adaptive dictionary generation algorithm proposed in [90] is useful to deal with common issues regarding the dimension deficiency.

The random configuration for the phase shifters has been widely used in the literature because the generated sensing matrices have low mutual coherence and satisfy the restricted isometry property condition with high probability. In contrast with the random configuration, a deterministic codebook and pilot design was considered in [43]. The approach taken in [43] is to obtain codebooks that provide the minimal total coherence (MTC) of the sensing

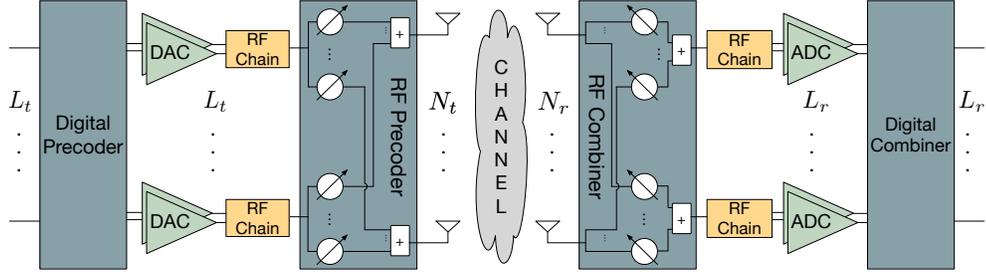


Figure 3.1: System block diagram of the phase shifter based hybrid beamforming architecture

matrix. However, it still contains randomness in beamformer column permutation.

In this chapter, I propose a completely deterministic beamformer codebook and pilot design method for CS based open-loop narrowband mmWave channel estimation. In practice, receivers should know about pilots and precoding schemes in order to properly perform channel estimation even with the random codebook. Therefore I design codebooks and pilots that can minimize mutual coherence of the resulting sensing matrix, which is of importance to CS algorithms. As suggested in [43], random beamformer column permutation plays a critical role to reduce measurement time and instances. I propose a greedy algorithm to find the best column permutation of the obtained RF beamformer codebook. Simulation results show that codebooks obtained by the proposed method outperforms the random codebook, the MTC codebook, and the adaptive CS in practical situations.

### 3.2 System Model

In the downlink, I assume a single base station (BS) and channel estimation is performed by each user. Both the BS and user equipment (UE) are equipped with the fully-connected phase shifter based hybrid beamforming architecture as shown in Fig 3.1. The BS has  $N_t$  transmit antennas and  $L_t$  transmit RF chains, and the UE has  $N_r$  receive antennas and  $L_r$  receive RF chains. The number of streams is assumed to be equal to  $L_t$ . Assuming the channel is narrowband, the received signal after RF and baseband combining at the time instance  $m$  can be expressed as

$$\mathbf{y}_m = \sqrt{\rho} \mathbf{W}_m^H \mathbf{H} \mathbf{F}_m \mathbf{x}_m + \mathbf{W}_m^H \mathbf{n}_m \in \mathbb{C}^{L_r}, \quad (3.1)$$

where  $\rho$  is the average transmit power,  $\mathbf{W}_m \in \mathbb{C}^{N_r \times L_r}$  and  $\mathbf{F}_m \in \mathbb{C}^{N_t \times L_t}$  denote the combiner and precoder matrices, respectively,  $\mathbf{x}_m \in \mathbb{C}^{L_t}$  denotes the pilot symbol vector,  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  is the channel matrix, and  $\mathbf{n}_m \in \mathbb{C}^{N_r} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$  denotes the additive noise.  $(\cdot)^H$  denotes the conjugate transpose. Both the precoder and combiner matrices are a product of RF and baseband ones, i.e.,  $\mathbf{F}_m = \mathbf{F}_{RF,m} \mathbf{F}_{BB,m}$  and  $\mathbf{W}_m = \mathbf{W}_{RF,m} \mathbf{W}_{BB,m}$ . By vectorizing the right hand side in (3.1), it can be rewritten as

$$\begin{aligned} \mathbf{y}_m &= \sqrt{\rho} (\mathbf{s}_m^T \otimes \mathbf{W}_m^H) \text{vec}(\mathbf{H}) + \mathbf{v}_m, \\ &= \sqrt{\rho} \mathbf{\Phi}_m \text{vec}(\mathbf{H}) + \mathbf{v}_m, \end{aligned}$$

where  $\mathbf{s}_m = \mathbf{F}_m \mathbf{x}_m$ ,  $\mathbf{v}_m = \mathbf{W}_m^H \mathbf{n}_m$ ,  $\mathbf{\Phi}_m = \mathbf{s}_m^T \otimes \mathbf{W}_m^H$ , and  $(\cdot)^T$  and  $\otimes$  denote the matrix transpose and the Kronecker product. By stacking  $M$  instances of

the received signal vectors, I can obtain

$$\mathbf{y} = \sqrt{\rho}\mathbf{\Phi}\text{vec}(\mathbf{H}) + \mathbf{v} \in \mathbb{C}^{ML_r}, \quad (3.2)$$

where  $\mathbf{y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots, \mathbf{y}_M^\top]^\top$  is the total received signal vector,  $\mathbf{\Phi} = [\mathbf{\Phi}_1^\top, \mathbf{\Phi}_2^\top, \dots, \mathbf{\Phi}_M^\top]^\top$  is the sensing matrix, and  $\mathbf{v} = [\mathbf{v}_1^\top, \mathbf{v}_2^\top, \dots, \mathbf{v}_M^\top]^\top$ .

For the narrowband channel, I adopt the geometric channel model. Assuming  $N_p$  clusters constitute the channel, the channel matrix can be given as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_p}} \sum_{l=0}^{N_p-1} \alpha_l \mathbf{a}_r(\theta_l) \mathbf{a}_t^H(\vartheta_l) \in \mathbb{C}^{N_r \times N_t}, \quad (3.3)$$

where  $\alpha_l \sim \mathcal{CN}(0, \sigma_\alpha^2)$  is the complex channel gain,  $\mathbf{a}_t(\cdot)$  and  $\mathbf{a}_r(\cdot)$  are, respectively, the transmit and receive array response vectors evaluated at the angles.  $\theta$  and  $\vartheta$  are the angle of arrival (AoA) and departure (AoD). In this chapter, the employed antenna arrays are assumed to be uniform linear array with a half wavelength spacing. The transmit and receive array response vectors are then given as

$$\begin{aligned} \mathbf{a}_t(\vartheta) &= \sqrt{\frac{1}{N_t}} [1, e^{-j\pi \cos \vartheta}, \dots, e^{-j\pi \cos(N_t-1)\vartheta}]^\top \in \mathbb{C}^{N_t}, \\ \mathbf{a}_r(\theta) &= \sqrt{\frac{1}{N_r}} [1, e^{-j\pi \cos \theta}, \dots, e^{-j\pi \cos(N_r-1)\theta}]^\top \in \mathbb{C}^{N_r}, \end{aligned}$$

and the corresponding transmit and receive array response matrices can be constructed as

$$\mathbf{A}_t = [\mathbf{a}_t(\vartheta_0), \mathbf{a}_t(\vartheta_1), \dots, \mathbf{a}_t(\vartheta_{N_p-1})] \in \mathbb{C}^{N_t \times N_p},$$

$$\mathbf{A}_r = [\mathbf{a}_r(\theta_0), \mathbf{a}_r(\theta_1), \dots, \mathbf{a}_r(\theta_{N_p-1})] \in \mathbb{C}^{N_r \times N_p}.$$

With the transmit and receive array response matrices, (3.3) can be rewritten as

$$\mathbf{H} = \mathbf{A}_r \mathbf{H}_d \mathbf{A}_t^H, \quad (3.4)$$

where  $\mathbf{H}_d \in \mathbb{C}^{N_p \times N_p}$  is a diagonal matrix with the scaled channel gains on its diagonal.

Now I define the angle dictionary matrix for sparse formulation. The transmit and receive angle dictionary matrix,  $\bar{\mathbf{A}}_t$ , is defined as

$$\bar{\mathbf{A}}_t = \sqrt{\frac{1}{N_t}} [\mathbf{a}_t(\nu_0), \mathbf{a}_t(\nu_1), \dots, \mathbf{a}_t(\nu_{G_t-1})] \in \mathbb{C}^{N_t \times G_t},$$

where  $G_t$  is the size of transmit angle grid, and  $\nu_n \in \Theta = \{\nu_n = \arccos \frac{\phi_n}{\pi} \mid \phi_n = \frac{n}{G_t} - \frac{1}{2}, n = 0, 1, \dots, G_t - 1\}$  where  $G_t \geq N_t$ . It is worth noting that  $\bar{\mathbf{A}}_t \bar{\mathbf{A}}_t^H = \frac{G}{N_t} \mathbf{I}_{N_t}$  where  $\mathbf{I}_{N_t}$  is the  $N_t \times N_t$  identity matrix. The receive angle dictionary matrix,  $\bar{\mathbf{A}}_r$ , is also similarly defined with  $G_r (\geq N_r)$ . Then the channel matrix in (3.4) can be rewritten as

$$\mathbf{H} = \bar{\mathbf{A}}_r \bar{\mathbf{H}}_d \bar{\mathbf{A}}_t^H, \quad (3.5)$$

where  $\bar{\mathbf{H}}_d \in \mathbb{C}^{G_r \times G_t}$  is the channel gain matrix which needs not be a diagonal matrix. Ignoring the grid quantization errors,  $\text{vec}(\bar{\mathbf{H}}_d)$  is an  $N_p$ -sparse vector that will be estimated.

By plugging (3.5) into (3.2), I can get

$$\mathbf{y} = \sqrt{\rho} \Phi (\bar{\mathbf{A}}_t^* \otimes \bar{\mathbf{A}}_r) \text{vec}(\bar{\mathbf{H}}_d) + \mathbf{v}$$

$$= \sqrt{\rho} \Phi \Psi \mathbf{h} + \mathbf{v},$$

where  $\Psi = \bar{\mathbf{A}}_t^* \otimes \bar{\mathbf{A}}_r \in \mathbb{C}^{N_t N_r \times G_t G_r}$  is the dictionary matrix, and  $\mathbf{h} = \text{vec}(\bar{\mathbf{H}}_d) \in \mathbb{C}^{G_t G_r}$  is the sparse channel gain vector. The product of the sensing matrix and the dictionary matrix is called the equivalent sensing matrix, i.e.,  $\Phi \Psi$ . I assume that the channel gain vector is estimated by using CS algorithms.

The major difference compared with Chapter 2 is that Chapter 2 assumes that the total beamformer configurations ( $M = M_t M_r$ ) are combinations of precoders ( $M_t$ ) and combiners ( $M_r$ ) in the system model whereas there is no such restriction in this chapter. Therefore the model in this chapter is more general. However, later in Section 3.3, the restriction is added as part of problem solving.

Another difference is the angle grid sizes. The grid size is common for both transmitter and receiver in Chapter 2, but they can be different in this chapter.

### 3.3 Coherence Minimizing Codebook

Recovery guarantees of CS algorithms can be assessed by mutual coherence of the equivalent sensing matrix, and the mutual coherence,  $\mu$ , is defined by

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2} = \max_{i \neq j} \frac{|(\mathbf{A}^H \mathbf{A})_{ij}|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2},$$

where  $\mathbf{a}_i$  denotes the  $i$ -th column vector in the matrix  $\mathbf{A}$ , and  $\mathbf{A}_{ij}$  denotes the element in the  $i$ -th row and  $j$ -th column of the matrix  $\mathbf{A}$ . Therefore

mutual coherence minimization can be achieved by minimizing all off-diagonal elements of  $\mathbf{A}^H \mathbf{A}$ , i.e.,  $\min \|\mathbf{A}^H \mathbf{A} - \mathbf{I}\|_F^2$ . Plugging the equivalent sensing matrix, the optimization problem is given by

$$\min \|\Psi^H \Phi^H \Phi \Psi - \mathbf{I}\|_F^2. \quad (3.6)$$

The objective function in (3.6) can be simplified as

$$\|\Psi^H \Phi^H \Phi \Psi - \mathbf{I}\|_F^2 = \|k\Phi^H \Phi - \mathbf{I}\|_F^2 = \left\| k \sum_{m=1}^M \Phi_m^H \Phi_m - \mathbf{I} \right\|_F^2,$$

where the first equality holds due to the fact that  $\Psi\Psi^H = k\mathbf{I}$  and  $k = \frac{G_t G_r}{N_t N_r}$  can be ignored since it scales all columns of the equivalent sensing matrix and does not affect mutual coherence. Our goal, thus, is to find the sets of the pilots, precoders and combiners that minimize the objective function. Considering the fact that  $\Phi_m$  is  $\mathbf{s}_m^T \otimes \mathbf{W}_m^H$ , the summation in the objective function can also be written as

$$\sum \Phi_m^H \Phi_m = \sum \mathbf{s}_m^* \mathbf{s}_m^T \otimes \mathbf{W}_m \mathbf{W}_m^H.$$

The above equation can readily be made the identity matrix if (i) the summation can be distributed over the Kronecker product and (ii)  $M$  is large enough. For illustration, let  $M$  be  $M = M_t M_r$  where  $M_t$  and  $M_r$  denote the number of configurations of transmitter and receiver, then I have

$$\begin{aligned} \sum_m^M \mathbf{s}_m^* \mathbf{s}_m^T \otimes \mathbf{W}_m \mathbf{W}_m^H &= \sum_{m_t}^{M_t} \sum_{m_r}^{M_r} \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^T \otimes \mathbf{W}_{m_r} \mathbf{W}_{m_r}^H \\ &= \sum_{m_t}^{M_t} \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^T \otimes \sum_{m_r}^{M_r} \mathbf{W}_{m_r} \mathbf{W}_{m_r}^H. \end{aligned} \quad (3.7)$$

Since  $\mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top$  is a rank one matrix, and  $\mathbf{W}_{m_r} \mathbf{W}_{m_r}^H$  can be up to a rank  $L_r$  matrix,  $M_t$  and  $M_r$  must be at least  $N_t$  and  $\frac{N_r}{L_r}$  in order to make both  $\sum \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top$  and  $\sum \mathbf{W}_{m_r} \mathbf{W}_{m_r}^H$  become a full rank matrix. Being full rank matrices is a critical requirement for it to be the identity matrix. Then the solutions for making both terms,  $\sum \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top$  and  $\sum \mathbf{W}_{m_r} \mathbf{W}_{m_r}^H$ , the identity matrix can easily be found. I now set this as the baseline, and start to decrease  $M$ .

As  $M$  is a product of  $M_t$  and  $M_r$ , either or both can be decreased to have a lower  $M$  value. A BS normally has far more antennas than UEs do ( $N_t \gg \frac{N_r}{L_r}$ ), I target  $M_t$  for  $M$  reduction. Assuming  $\sum \mathbf{W}_{m_r} \mathbf{W}_{m_r}^H = \mathbf{I}$ , I focus on  $\sum \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top$ . It can be rewritten as

$$\begin{aligned}
\sum_{m_t} \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top &= \sum_{m_t} \mathbf{F}_{m_t}^* \mathbf{x}_{m_t}^* \mathbf{x}_{m_t}^\top \mathbf{F}_{m_t}^\top \\
&= \sum_{m_f} \mathbf{F}_{m_f}^* \left( \sum_{m_x} \mathbf{x}_{m_x}^* \mathbf{x}_{m_x}^\top \right) \mathbf{F}_{m_f}^\top, \\
&= \sum_{m_f} \mathbf{F}_{m_f}^* \mathbf{X} \mathbf{F}_{m_f}^\top, \tag{3.8}
\end{aligned}$$

where  $\mathbf{X} = \sum \mathbf{x}_{m_x}^* \mathbf{x}_{m_x}^\top$ , the second equality holds if I use the same technique that is used for the transmitter and receiver separation, and  $M_t = M_f M_x$  where  $M_x$  and  $M_f$  denote the number of pilot symbol vectors and precoders, respectively. In the ideal case, it is desired that both  $\mathbf{X}$  and  $\sum \mathbf{F}_{m_f}^* \mathbf{F}_{m_f}^\top$  are to be the identity matrix to obtain  $\sum \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top = \mathbf{I}$ . Then  $M_x$  should be  $L_t$ , and accordingly  $M_f$  becomes  $\frac{N_t}{L_t}$ . Since I do not want to lose beamforming capability, I control  $M$  values by adjusting  $M_x$ . The value of  $M_x$

determines the rank of  $\mathbf{X}$ . Then it boils down to the low rank approximation which can be expressed as

$$\min_{\mathbf{X}} \|\mathbf{I} - \mathbf{X}\|_F \text{ subject to } \text{rank}(\mathbf{X}) \leq M_x,$$

and its analytic solution is to exploit truncated singular value decomposition (SVD). It implies that the smaller rank naturally leads  $\mathbf{X}$  to have greater deviation from the identity matrix. There exists infinite solutions; however, I have some criteria on choosing the solution. Firstly, all elements of all  $\mathbf{x}_{m_x}$  should not be zero to take advantage of all possible beams. Secondly, all elements of  $\mathbf{x}_{m_x}$  should have the identical magnitude to equally weight all beams. The  $L_t$ -point Discrete Fourier Transform (DFT) matrix satisfies these criteria, and any  $M_x$  columns can be chosen from the DFT matrix and conclude the pilot codebook. The pilot codebook  $\mathcal{X}$  can be expressed as

$$\mathcal{X} = \{\mathbf{x}_{m_x} : \forall m_x \in 1, 2, \dots, L_t, \mathbf{U}_{L_t} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{L_t}]\},$$

where  $\mathbf{U}_N \in \mathbb{C}^{N \times N}$  is the  $N$ -point DFT matrix. According to  $M_x$ ,  $M$  is determined by  $\frac{N_t N_r}{L_t L_r} M_x$ .

I have assumed that  $\sum \mathbf{F}_{m_t}^* \mathbf{F}_{m_t}^\top$  and  $\sum \mathbf{W}_{m_r} \mathbf{W}_{m_r}^H$  are the identity matrix. Taking into account the phase shifter based analog beamformers,  $M_f = \frac{N_t}{L_t}$  and  $M_r = \frac{N_r}{L_r}$ , the column partition of the proper size DFT matrices can compose the beamformer codebooks. Namely then can be expressed as

$$\mathcal{F} = \left\{ \mathbf{F}_{m_f} : \forall m_f \in \{1, 2, \dots, M_f = \frac{N_t}{L_t}\}, \mathbf{U}_{N_t} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{M_f}] \right\},$$

---

**Algorithm 1** Greedy Precoder Column Ordering

---

**Input:**  $L_t \times L_t$  pilot codebook  $\mathbf{X}$  and  $N_t$ -point DFT precoder matrix  $\mathbf{U}$ **Output:**  $N_t \times N_t$  ordered precoder matrix  $\mathbf{F}$ 

- 1: *Initialization:* Set  $\mathbf{F}$  to an empty matrix.
  - 2: **for**  $n = 1$  to  $N_t$  **do** ▷ Choose  $N_t$  columns
  - 3:     **for**  $m = 1$  to the number of columns left in  $\mathbf{U}$  **do**
  - 4:         ▷ Go through a pool of possible columns
  - 5:          $\mathbf{F}^m \leftarrow [\mathbf{F}, \mathbf{U}_{:,m}]$
  - 6:          $\mathbf{S}^m \leftarrow (\mathbf{F}^m)^* (\mathbf{I} \otimes \mathbf{X}) (\mathbf{F}^m)^\top$
  - 7:         Calculate  $\mu(\Phi)_m$  with  $\mathbf{S}^m$  by using (3.11).
  - 8:     **end for**
  - 9:      $m_{\min} = \arg \min_m \mu(\Phi)_m$
  - 10:      $\mathbf{F} \leftarrow [\mathbf{F}, \mathbf{U}_{:,m_{\min}}]$  ▷ Append the found vector to the codebook
  - 11:      $\mathbf{U} \leftarrow \mathbf{U}_{[:, \dots, m_{\min}-1, m_{\min}+1, \dots]}$  ▷ Remove the found vector from the pool
  - 12: **end for**
  - 13: **return**  $\mathbf{F}$
- 

$$\mathcal{W} = \left\{ \mathbf{W}_{m_r} : \forall m_r \in \{1, 2, \dots, M_r = \frac{N_r}{L_r}\}, \mathbf{U}_{N_r} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{M_r}] \right\}, \quad (3.9)$$

where  $\mathcal{F}$  and  $\mathcal{W}$  denote the precoder and combiner codebooks, respectively. Another advantage of a DFT matrix is that the required phase shifter resolution in bits is  $\log_2(\text{number of antennas})$ .

### 3.4 Precoder Column Permutation and Pilot Design

When the number of pilot symbol vectors,  $M_x$ , is less than the number of transmit RF chains,  $L_t$ ,  $\mathbf{X}$  is not the identity matrix, and neither is (3.8). In this case, the column order in  $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{M_f}]$  and  $\mathbf{x}_{m_t}$  selection result in changes not only in  $\mathbf{S} = \sum \mathbf{s}_{m_t}^* \mathbf{s}_{m_t}^\top$ , but also mutual coherence of  $\Phi$ . In [43],

the columns are proposed to be randomly permuted; however, I want to find a deterministic order so that the codebook can practically be used. Note that column permutation does not affect the fact that  $\sum \mathbf{F}_{m_t}^* \mathbf{F}_{m_t}^\top = \mathbf{I}$ .

Since  $\mathbf{F}$  has  $N_t$  columns and  $N_t$  is usually a large number, the exhaustive search for the best column order would not be feasible. With  $N_t = 64$ , the number of permutations is  $64! \approx 1.27 \times 10^{89}$ . Therefore by adopting a greedy algorithm, I iteratively seek one column vector from the DFT matrix that achieves the lowest mutual coherence with the pre-selected vectors for the codebook. At the same time,  $\Phi$  is a  $ML_t \times N_t N_r$  matrix, and calculating its mutual coherence directly from  $\Phi$  is computationally expensive. Thus I first obtain a simplified formulation for mutual coherence of the sensing matrix. Mutual coherence can also be defined as

$$\mu(\Phi) = \max_{i \neq j} \left| \left( \tilde{\Phi}^H \tilde{\Phi} \right)_{ij} \right|, \quad (3.10)$$

where  $\tilde{\Phi}$  is the column-wise normalized version of  $\Phi$  which can also be expressed as  $\tilde{\Phi} = \Phi \Sigma^{-\frac{1}{2}}$  where  $\Sigma = \mathbf{I}_{N_t N_r} \circ (\Phi^H \Phi)$  and  $\circ$  denotes the Hadamard product. Then I have

$$\begin{aligned} \tilde{\Phi}^H \tilde{\Phi} &= \left\{ \mathbf{I}_{N_t N_r} \circ (\Phi^H \Phi) \right\}^{-\frac{1}{2}} \Phi^H \Phi \left\{ \mathbf{I}_{N_t N_r} \circ (\Phi^H \Phi) \right\}^{-\frac{1}{2}} \\ &= \left\{ \mathbf{I}_{N_t N_r} \circ (\mathbf{S} \otimes \mathbf{I}_{N_r}) \right\}^{-\frac{1}{2}} (\mathbf{S} \otimes \mathbf{I}_{N_r}) \times \left\{ \mathbf{I}_{N_t N_r} \circ (\mathbf{S} \otimes \mathbf{I}_{N_r}) \right\}^{-\frac{1}{2}} \\ &= \left\{ (\mathbf{S} \circ \mathbf{I}_{N_t})^{-\frac{1}{2}} \otimes \mathbf{I}_{N_r} \right\} (\mathbf{S} \otimes \mathbf{I}_{N_r}) \times \left\{ (\mathbf{S} \circ \mathbf{I}_{N_t})^{-\frac{1}{2}} \otimes \mathbf{I}_{N_r} \right\} \\ &= \left\{ (\mathbf{S} \circ \mathbf{I}_{N_t})^{-\frac{1}{2}} \mathbf{S} (\mathbf{S} \circ \mathbf{I}_{N_t})^{-\frac{1}{2}} \right\} \otimes \mathbf{I}_{N_r}, \end{aligned}$$

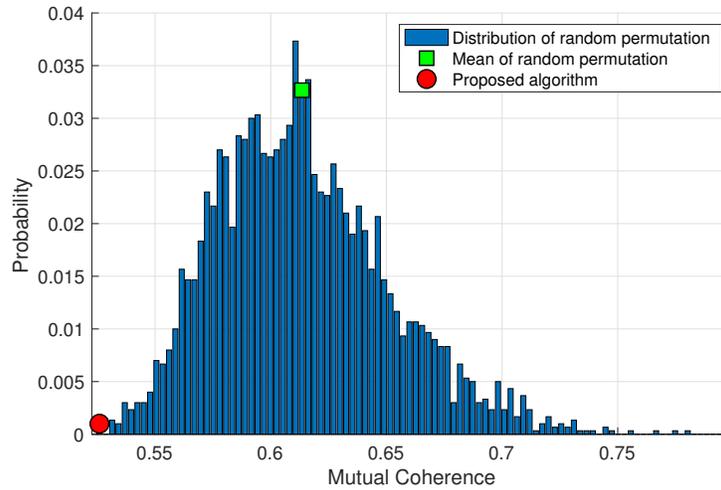
where the second equality comes from (3.7) and (3.9), and (3.10) becomes

$$\mu(\Phi) = \max_{i \neq j} \left| \left( (\mathbf{S} \circ \mathbf{I}_{N_t})^{-\frac{1}{2}} \mathbf{S} (\mathbf{S} \circ \mathbf{I}_{N_t})^{-\frac{1}{2}} \right)_{ij} \right|. \quad (3.11)$$

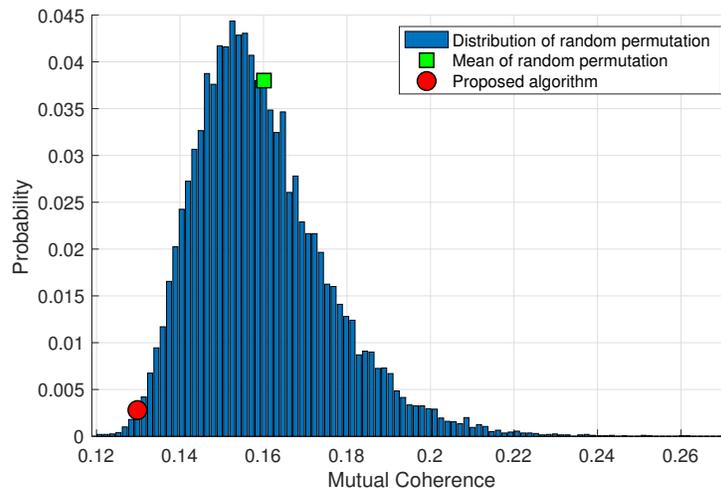
Here,  $\mathbf{S}$  can be written in matrix form as  $\mathbf{F}^* (\mathbf{I}_{\frac{N_t}{L_t}} \otimes \mathbf{X}) \mathbf{F}^\top$  by simplifying (3.8). In (3.11), the matrix size is reduced by a factor of  $N_r$  compared with (3.10) in both dimensions. The matrix inversion and square root are taken on a diagonal matrix.

For pilot codebook construction,  $M_x$  column vectors are to be chosen from the  $L_t$ -point DFT matrix. As  $L_t$  is usually not a large value, I perform exhaustive search for this selection. However, the first column of the DFT matrix should always be included in selection. This is because the elements in the first row of  $\mathbf{F}_m$  are identical, and a sum of elements in the non-first column of the DFT matrix is zero. It leads to the first elements of  $\mathbf{s}_m = \mathbf{F}_m \mathbf{x}_m$  being a zero and the first  $N_r$  columns in  $\Phi_m$  zeros. Not having the first column of the DFT matrix as a pilot vector eventually ends up with the first  $N_r$  columns in  $\Phi$  being zeros. It is also worth mentioning that the order of pilot vectors does not matter considering  $\mathbf{X}$  is a sum of outer products of each pilot vector.

Taking into account all that is mentioned above, the greedy column ordering algorithm is given in Algorithm 1.  $\mathbf{U}_{:,m}$  denotes the  $m$ -th column of the matrix  $\mathbf{U}$ . Since the complexity of the ordering algorithm is dominated by the matrix product in line 6 and the loops around it, the asymptotic complexity is  $O(N_t^5)$ . However, it can be performed in advance offline. As Algorithm 1 is for a given set of pilots, I perform the algorithm  $\binom{L_t-1}{M_x-1}$  times to find the best



(a)  $M_x = 2$



(b)  $M_x = 7$

Figure 3.2: Distribution of mutual coherence of the sensing matrix generated with 20,000 random precoder column permutation. The mean of the distribution and the mutual coherence of the proposed algorithm are marked as well. The parameters in this simulation is  $N_t = 64$ ,  $L_t = 8$ , and (a)  $M_x = 2$  and (b)  $M_x = 8$ .

pilot selection and precoder column order that achieves the lowest mutual coherence. As long as the predefined pilot and precoding vectors are used, selection and ordering can be done offline and be used online without additional processing.

The distribution of mutual coherence obtained by random column permutation and its mean value are provided in Fig. 3.2 along with one obtained by the proposed algorithm. For illustration, two different values (two and seven) are considered for  $M_x$ . As shown in the figure for both cases, the mutual coherence of the proposed algorithm is lower than the mean of random permutation and is near the lowest value that random permutation can achieve. This observation can also be seen with all possible  $M_x$  values, and the evaluated mutual coherence values from random permutation and the proposed algorithm are given in Table 3.1. For all possible  $M_x$  values, the proposed algorithm achieves lower mutual coherence than the mean of mutual coherence distribution of random permutation, and as  $M_x$  increases, mutual coherence declines. When  $M_x = L_t (= 8$  in this case), the mutual coherence converges to zero. It implies that higher  $M_x$  values make the CS algorithms perform more accurate channel estimation.

### 3.5 Numerical Results

In this section, I compare performance of the deterministic codebook obtained by the proposed algorithm with that of (i) the random phase shifter and pilot configuration, (ii) the MTC codebook with random precoder column

permutation [43] and (iii) the adaptive CS [19]. For evaluation of the codebooks, OMP is employed except for the adaptive CS. The system parameters for simulation are as follows unless otherwise specified:  $N_t = 64$ ,  $N_r = 16$ ,  $L_t = 8$ ,  $L_r = 4$ ,  $G_t = \gamma N_t$ ,  $G_r = \gamma N_r$ ,  $b_{PS} = 6$  and  $N_p = 6$  where  $\gamma = 1.5 (> 1)$  is the grid multiplier. A list of important parameters are given in Table 3.2 with brief descriptions and values. For the adaptive CS,  $G_t = G_r = 96$ , which results in  $M = 864$ . The normalized mean squared error (NMSE) is defined as  $\mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2 / \|\mathbf{H}\|_F^2]$  where  $\mathbf{H}$  and  $\hat{\mathbf{H}}$  are the true and the estimated channel matrix, respectively, SNR is defined as  $\rho/\sigma^2$ , and  $(\cdot)_F$  denotes the Frobenius norm of a matrix. Multipath components of the channels, in simulations, have AoDs and AoAs that are not necessarily aligned with the grids of the dictionary. The complete source code is available [91].

Fig. 3.3 shows NMSE as a function of the transmit SNR, and the three codebooks and the adaptive CS are considered. The first observation is that NMSE decreases with SNR and the number of snapshots ( $M$ ). With an increase in  $M$ , the estimator can take more measurements, which improves performance of CS-based estimator. At the same time, however, higher  $M$  also

Table 3.1: Mean total coherence of random permutation and total coherence of the proposed algorithm vs. number of pilot symbol vectors,  $M_x$

$M_x$	1	2	3	4	5	6	7	8
Permutation	0.86	0.62	0.48	0.38	0.30	0.23	0.16	0
Proposed	0.75	0.52	0.39	0.31	0.25	0.19	0.13	0

requires a longer measurement time and more computation. Thus I have a performance versus time and power consumption trade-off.

Three values of  $M$  ( $= 64, 128, 256$ ) are considered in Fig. 3.3, and performance ranking of the codebooks changes depending on the value. The proposed codebook achieves the lowest NMSE among the three when  $M$  is either 128 or 256. With  $M = 256$ , its performance is the best together with the MTC, and  $M = 64$  makes it the worst. It implies that the proposed algorithm works well when the number of snapshots is not too small. This is due to the fact that  $\mathbf{X}$  deviates from the identity matrix with small  $M_x$ . Different approaches may be explored to address this issue, which will be our future work.

The adaptive CS generally needs a large number of snapshots which is

Table 3.2: Simulation parameters

Notation	Description	Value
$N_t$	Number of transmit antennas	64
$N_r$	Number of receive antennas	16
$L_t$	Number of transmit RF chains	8
$L_r$	Number of receive RF chains	4
$G_t$	Transmit angle grid size	$\gamma N_t$
$G_r$	Receive angle grid size	$\gamma N_r$
$\gamma$	Grid multiplier	1.5
$M_x$	Number of pilot symbol vectors	2, 4, 8
$M_f$	Number of precoders ( $= N_t/L_t$ )	8
$M$	Number of frames ( $= M_f M_x N_r/L_r$ )	64, 128, 256
$N_p$	The number of multi-path components in channel	6

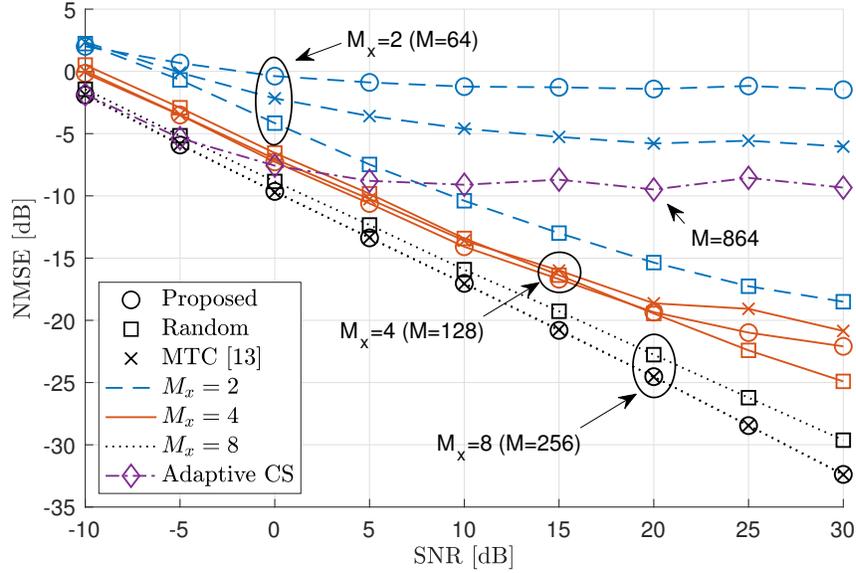
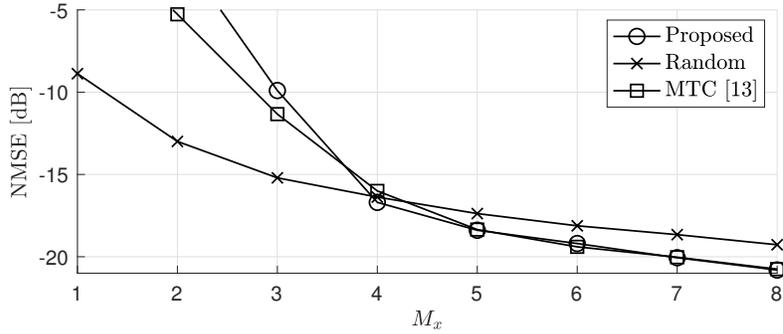


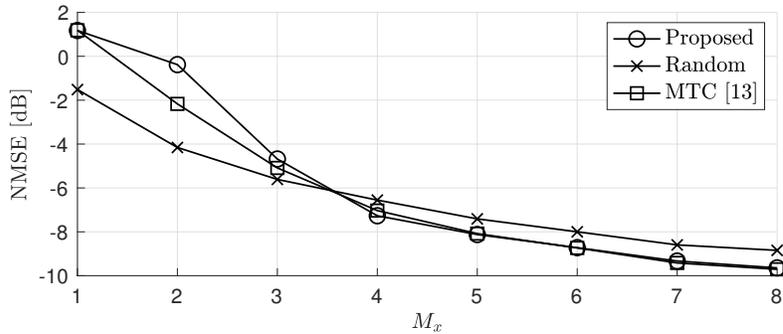
Figure 3.3: NMSE vs. SNR with the proposed, random and MTC codebooks using OMP [43] along with the adaptive CS channel estimation [19].

determined by the grid size. Even with more snapshots, estimation errors are greater than the considered codebooks including the proposed one. In Fig. 3.3, the adaptive CS has  $M = 864$ , but the NMSE is higher than the proposed codebook with  $M = 256$  across the SNR range and than that with  $M = 128$  in medium and high SNR regimes.

To illustrate the relationship between performance and  $M_x$ , Fig. 3.4 is provided where SNR is fixed at either 15 or 0 dB. The NMSE of the random codebook starts from a relatively low NMSE and gradually declines with  $M_x$ . The other two codebooks, on the other hand, start from high NMSE, drop at low  $M_x$ , and gradually decline in medium and high  $M_x$  regimes. Due to the



(a) SNR = 15 dB



(b) SNR = 0 dB

Figure 3.4: NMSE vs.  $M_x$  with SNR of (a) 15 dB and (b) 0 dB.

step drops in the low  $M_x$  regime, NMSE curves of the proposed and MTC codebooks cross that of random codebook. For 15 dB SNR, the crossover happens at  $M_x = 4$  with the proposed codebook which is earlier than  $M_x = 5$  with the MTC codebook. For 0 dB SNR, the it happens at  $M_x = 4$  for both codebooks. With system configurations such as one used in this simulation, thus, the proposed codebook is preferred when  $M_x$  is greater than three as (i) it achieves NMSE that is lower or equal to that of the MTC codebook and (ii) is lower than that of the random codebook when  $M_x > 3$  and (iii) the

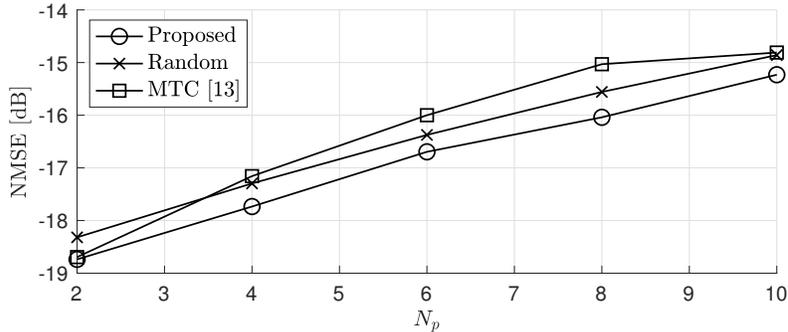


Figure 3.5: NMSE vs.  $N_p$  channel paths with 15 dB SNR and  $M_x = 4$

proposed codebook has no randomness.

The number of channel paths  $N_p$  affects channel estimation performance as well since it is directly related to the sparsity of the channel vector being estimated. In Fig. 3.5, SNR and  $M_x$  are fixed at 15 dB and four, respectively. In this figure, the proposed codebook yields the lowest estimation error across the  $N_p$  values. The order of the MTC and the random codebooks varies depending on  $N_p$ . Since  $N_p$  varies and is determined by the channel environment, the proposed codebook would be a good choice as it achieves the lowest channel estimation error regardless of the number of channel paths.

### 3.6 Conclusion

In this chapter, I proposed a codebook design method for mmWave channel estimation in hybrid beamforming communication systems based on mutual coherence minimization. I first obtained the criteria for codebooks of pilots, precoders and combiners, and proposed the pilot selection and precoder

column ordering algorithm for further mutual coherence reduction. The mutual coherence distribution of random column permutation was provided to show the proposed greedy precoder column ordering algorithm achieves lower mutual coherence than random permutation.

I also provided the channel estimation simulation results using OMP for performance comparison between the proposed, the random and the MTC codebooks. The proposed method provides the best tradeoff between channel estimation performance and measurement time, and has more accurate channel estimation vs. number of channel paths.

In the following chapter, I turn my focus from codebook designs to CS-based channel estimation algorithm development by taking extremely low resolution ADCs into consideration.

## Chapter 4

# Narrowband Channel Estimation for Hybrid Beamforming Millimeter Wave Communication Systems with One-bit ADCs

In Chapters 2 and 3, I propose the deterministic analog and digital codebook design frameworks for CSI acquisition under the CS context. The contributions of the two chapters focus on codebook designs in HB architectures. Deviating from the previous contributions, I turn my focus to investigation on channel estimation algorithm in more complicated systems.

In this chapter<sup>1</sup>, I consider a further power-saving hardware platform which is a hybrid beamforming architecture with one-bit quantizers and develop a CS algorithm that is suitable for such a platform. Traditionally each antenna is paired with a high-speed analog-to-digital converter (ADC), which results in high power consumption. A hybrid beamforming architecture and one-bit resolution ADCs have been proposed to reduce power consumption.

---

<sup>1</sup> This chapter is based on the work published in J. Sung, J. Choi, and B. L. Evans, “Narrowband channel estimation for hybrid beamforming millimeter wave communication systems with one-bit quantization”, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 15–20, 2018, Calgary, AB, Canada [79]. This work was supervised by Prof. Brian L. Evans, and valuable feedback and contributions from Dr. Jinseok Choi improved the quality of this work.

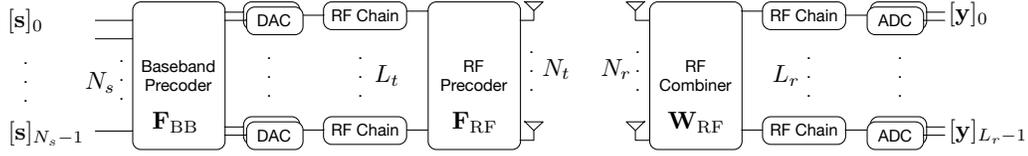


Figure 4.1: A block diagram of a MIMO hybrid beamforming communication system. A transmitter and a receiver in this system are equipped with  $N_t$  and  $N_r$  antennas, and  $L_t$  and  $L_r$  RF chains, respectively.

However, analog beamforming and one-bit quantization make channel estimation more challenging.

In particular, this chapter proposes a narrowband channel estimation algorithm for mmWave communication systems with hybrid beamforming and one-bit ADCs based on generalized approximate message passing (GAMP). I show through simulation that 1) GAMP variants with one-bit ADCs have better performance than do least-squares estimation methods without quantization; 2) the proposed one-bit GAMP algorithm achieves the lowest estimation error among the GAMP variants; and 3) exploiting more frames and RF chains enhances the channel estimation performance.

## 4.1 Introduction

With the adoption of millimeter wave (mmWave) bands in 5G cellular communications to use wider transmission bands, communication systems need a large number of antennas to compensate the high propagation losses in such frequency bands. In a traditional all-digital beamforming architecture, an RF chain with a high-speed ADC is paired with each antenna, which would

lead to excessively high power consumption for the large mmWave antenna arrays. Power consumption is proportional to the number of RF chains and exponential in the number of bits of resolution the ADCs. As an alternative, hybrid analog/digital beamforming architectures with low-resolution ADCs have been proposed. Hybrid architectures reduce the number of RF chains, and low-resolution ADCs gives exponential savings in the number of bits of resolution reduced.

Hybrid beamforming with low-resolution ADCs yields comparable achievable rates to those systems with high-resolution ADCs in the low and medium SNR regimes [92]. The hybrid and full-digital beamforming architectures with low-resolution quantizers are shown to alternatively achieve better spectral and energy efficiency trade-off depending on characteristics of system components [93]. Spectral and energy efficiencies can be further enhanced by employing resolution adaptive ADCs with bit allocation algorithms [94, 95].

The combination of hybrid beamforming architecture and low-resolution ADC intended to reduce power consumption, however, makes channel estimation in such systems more challenging. Prior work regarding channel estimation in the context of mmWave can be grouped into a MIMO hybrid beamforming architecture with perfect quantizers [19, 71, 74, 96] or a MIMO fully-digital architecture with low-resolution quantizers [25, 97, 98]. Very few publications concern a system equipped with hybrid beamforming and low-resolution ADCs [78, 97]. In [78], The modified expectation-maximization algorithm is shown to yield acceptable channel estimation errors with low-resolution ADCs. In

[97], generalized approximate message passing (GAMP) is proposed to use for wideband channel estimation, and four-bit ADCs are shown to achieve close performance to infinite bit ADCs at medium SNR. The channel estimation algorithms in [78, 97], however, are not specially designed for one-bit ADCs. Brief comparison of the mentioned prior work is given in Table 4.1. More detailed qualitative comparison is provided in Table 4.2 where the proposed contribution is compared with the selected prior work on low-resolution based channel estimation.

In this chapter, I propose a compressed sensing based channel estimation algorithm for a mmWave communication system equipped with one-bit ADCs and hybrid beamforming. GAMP and its variants have widely been used for channel estimation [25, 97, 99]. One variant of GAMP used in this chapter is named one-bit GAMP and is specifically developed for measurements taken with one-bit quantizers [100]. To make the algorithm better match a communication system model, I modify the algorithm to take the thermal noise into account. Simulation results show that GAMP variants considered in this chapter – GAMP [101], Expectation-Maximization Gaussian Mixture AMP (EM-GM-AMP) [102], and one-bit GAMP [100] – using one-bit ADCs perform better than does least-squares (LS) estimation without quantization. Among the variants, the modified one-bit GAMP achieves the lowest channel estimation error. The results also show that estimation performance can be enhanced by exploiting more frames and RF chains.

## 4.2 System Model

Consider a single-user MIMO hybrid beamforming mmWave system with  $N_t$  transmit and  $N_r$  receive antennas in the form of uniform linear arrays. The transmitter and receiver are equipped with  $L_t$  ( $\leq N_t$ ) and  $L_r$  ( $\leq N_r$ ) RF chains, respectively. The receiver employs quantizers that generate one-bit outputs.  $N_s$  data streams are transmitted over narrowband MIMO channels where  $N_s \leq \min(L_t, L_r)$ . The transceivers are assumed to have the same number of RF chains, i.e.  $L_t = L_r$ . The block diagram for the system is illustrated in Fig. 4.1. The baseband signal vector transmitted at  $N_t$  antennas in the  $m^{\text{th}}$  frame can be expressed as  $\mathbf{x}_m = \mathbf{F}_{\text{RF},m} \mathbf{F}_{\text{BB},m} \mathbf{s}_m$  where  $\mathbf{F}_{\text{RF},m} \in \mathbb{C}^{N_t \times L_t}$  is the RF precoder,  $\mathbf{F}_{\text{BB},m} \in \mathbb{C}^{L_t \times N_s}$  is the baseband precoder, and  $\mathbf{s}_m \in \mathbb{C}^{N_s \times 1}$  is the training symbol vector with the constraint  $\mathbb{E}[\mathbf{s}_m \mathbf{s}_m^*] = \frac{1}{N_s} \mathbf{I}_{N_s}$ . The RF precoder is assumed to be built using a network of analog phase shifters; therefore, all elements of  $\mathbf{F}_{\text{RF},m}$  should have the identical norm of  $\frac{1}{N_t}$ . In order to control the transmit power, the baseband precoder has a constraint

Table 4.1: Comparison of prior work

Ref.	Beamforming	ADC Resolution	Bandwidth
[19]	Hybrid	Infinite	Narrowband
[74]	Hybrid	Infinite	Narrowband
[96]	Hybrid	Infinite	Wideband
[71]	Hybrid	Infinite	Narrowband
[25]	Digital	Low (1 bit)	Narrowband
[98]	Digital	Low (1 bit)	Narrowband
[78]	Hybrid	Low (1–5 bits)	Narrowband
[97]	Digital	Low (1–4 bits)	Wideband

Table 4.2: Comparison among channel estimation algorithms for low-resolution ADCs

	[25]	[98]	[78]	[97]	Proposed
Hybrid	×	×	✓	×	✓
ADC resolution	1-bit	1-bit	1–5 bits	1–4 bits	1-bit
Algorithm	GAMP	Ellipsoid method	Modified EM	EM-GM-AMP	Modified one-bit GAMP
Complexity	Med.	Low	High	Med.	Med.
Performance	Med-low	Low	High	Med-low	Med-high

such that  $\|\mathbf{F}_{\text{RF},m}\mathbf{F}_{\text{BB},m}\|_F^2 = N_s$ . The quantized received baseband signal in the  $m^{\text{th}}$  frame can be expressed as

$$\mathbf{y}_m = Q\left(\sqrt{\rho}\mathbf{W}_{\text{RF},m}^H\mathbf{H}\mathbf{x}_m + \mathbf{W}_{\text{RF},m}^H\mathbf{n}_m\right), \quad (4.1)$$

where  $(\cdot)^H$  denotes the conjugate transpose,  $\rho$  denotes the average received power,  $\mathbf{W}_{\text{RF},m} \in \mathbb{C}^{N_r \times L_r}$  is the RF combiner,  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  is the flat-fading MIMO channel, and  $\mathbf{n}_m \in \mathbb{C}^{N_r \times 1} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$  is the noise vector. As with the RF precoder, the RF combiner has a constraint that all elements have the identical norm of  $\frac{1}{N_r}$ . Since quantization of the received signal in this system is performed with one-bit ADCs, the quantization operator  $Q(\cdot)$  extracts signs of real and imaginary components of a complex argument.

Using a geometric channel model, the normalized channel matrix  $\mathbf{H}$  can be constructed by summing up  $N_p$  paths. The azimuth angles of departure and arrival (AoD and AoA) associated with the  $l^{\text{th}}$  path are denoted as  $\theta_{dl}$  and

$\theta_{rl}$ , respectively. Both  $\theta_{tl}$  and  $\theta_{rl}$  are uniform random variables distributed over  $[0, 2\pi)$ . Therefore,  $\mathbf{H}$  can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{N_p}} \sum_{l=0}^{N_p-1} \alpha_l \mathbf{a}_r(\theta_{rl}) \mathbf{a}_t^H(\theta_{tl}),$$

where  $\alpha_l \sim \mathcal{CN}(0, \sigma_\alpha^2)$  is the complex channel gain of the  $l^{\text{th}}$  path,  $\mathbf{a}_t(\cdot) \in \mathbb{C}^{N_t \times 1}$  and  $\mathbf{a}_r(\cdot) \in \mathbb{C}^{N_r \times 1}$  are the transmit and receive array response vectors at the given AoD and AoA, respectively. The channel matrix  $\mathbf{H}$  is constrained to have  $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$  to maintain a constant channel power on average.  $\mathbf{H}$  can also be represented with the virtual channel representation as

$$\mathbf{H} = \mathbf{U}_r \mathbf{H}_v \mathbf{U}_t^H,$$

where  $\mathbf{U}_t$  and  $\mathbf{U}_r$  denote the normalized  $N_t$ -point and  $N_r$ -point unitary Discrete Fourier Transform (DFT) matrices, and  $\mathbf{H}_v \in \mathbb{C}^{N_r \times N_t}$  is the virtual channel matrix in the angular domain. As AoDs and AoAs are random in each channel realization, the array response vectors do not always align with columns of DFT matrices. This misalignment causes spectral leakage which degrades the performance of channel estimation algorithms [19]. The leakage effect can be seen by comparing two subfigures in Fig. 4.2. When AoDs and AoAs are perfectly aligned, or equivalently, antenna array response vectors can be expressed with the DFT matrix columns,  $\mathbf{H}_v$  has exactly  $N_p$  non-zero elements as seen in Fig. 4.2(a). Otherwise, each beam leaks into adjacent bins, which results in spreads around them as shown in Fig. 4.2(b). In Section 4.4, performance degradation due to the leakage effect is analyzed.

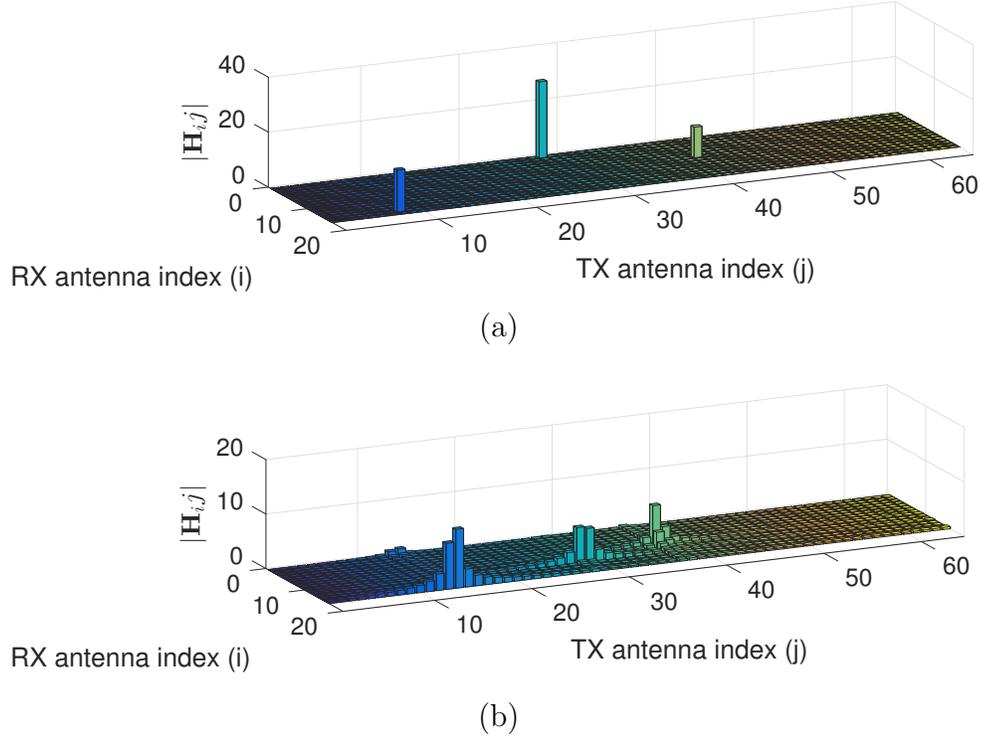


Figure 4.2: Magnitude of virtual channel matrix with  $N_p = 3$  in the angular domain. Spreads shown in (b) illustrate leakage into adjacent bins.

### 4.3 Compressed Sensing Channel Estimation

For estimation of the sparse virtual channel, the  $m^{\text{th}}$  received frame in (4.1) is reformulated into a vector form using  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ .

Then  $\mathbf{y}_m$  is rewritten as

$$\mathbf{y}_m = Q \left( \sqrt{\rho} (\mathbf{s}_m^T \mathbf{F}_{\text{BB},m}^T \mathbf{F}_{\text{RF},m}^T \otimes \mathbf{W}_{\text{RF},m}^H) (\mathbf{U}_t^* \otimes \mathbf{U}_r) \text{vec}(\mathbf{H}_v) + \tilde{\mathbf{n}}_m \right),$$

where  $(\cdot)^*$  denotes the complex conjugation,  $\otimes$  denotes the Kronecker product operator, and  $\tilde{\mathbf{n}}_m$  denotes  $\mathbf{W}_{\text{RF},m}^H \mathbf{n}_m$ . For simple notation, I define  $\Phi = [\Gamma_0^T, \Gamma_1^T, \dots, \Gamma_{M-1}^T]^T$ ,  $\Psi = \mathbf{U}_t^* \otimes \mathbf{U}_r$ , and  $\tilde{\mathbf{n}} = [\tilde{\mathbf{n}}_0^T, \tilde{\mathbf{n}}_1^T, \dots, \tilde{\mathbf{n}}_{M-1}^T]^T$  where

$\mathbf{\Gamma}_m = \mathbf{s}_m^\top \mathbf{F}_{\text{BB},m}^\top \mathbf{F}_{\text{RF},m}^\top \otimes \mathbf{W}_{\text{RF},m}^H$ . By stacking  $M$  received frame vectors,  $\mathbf{y}$  is obtained as

$$\mathbf{y} = [\mathbf{y}_0^\top, \mathbf{y}_1^\top, \dots, \mathbf{y}_{M-1}^\top]^\top = Q \left( \sqrt{\rho} \mathbf{\Phi} \mathbf{\Psi} \tilde{\mathbf{h}}_v + \tilde{\mathbf{n}} \right),$$

where  $\tilde{\mathbf{h}}_v$  denotes  $\text{vec}(\mathbf{H}_v)$ . I define the unquantized received vector  $\tilde{\mathbf{r}} = \tilde{\mathbf{W}} \tilde{\mathbf{h}}_v + \tilde{\mathbf{n}}$  where  $\tilde{\mathbf{W}} = \sqrt{\rho} \mathbf{\Phi} \mathbf{\Psi}$ . To separate in-phase and quadrature components, it is further reformulated as  $\mathbf{r} = \mathbf{W} \mathbf{h}_v + \mathbf{n}$  where

$$\mathbf{r} = \begin{bmatrix} \Re\{\tilde{\mathbf{r}}\} \\ \Im\{\tilde{\mathbf{r}}\} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \Re\{\tilde{\mathbf{W}}\} & -\Im\{\tilde{\mathbf{W}}\} \\ \Im\{\tilde{\mathbf{W}}\} & \Re\{\tilde{\mathbf{W}}\} \end{bmatrix},$$

$$\mathbf{h}_v = \begin{bmatrix} \Re\{\tilde{\mathbf{h}}_v\} \\ \Im\{\tilde{\mathbf{h}}_v\} \end{bmatrix}, \text{ and } \mathbf{n} = \begin{bmatrix} \Re\{\tilde{\mathbf{n}}\} \\ \Im\{\tilde{\mathbf{n}}\} \end{bmatrix}.$$

For sparse reconstruction of the virtual channel vector, I consider one-bit GAMP which is specifically developed for measurements taken with one-bit quantizers [100]. I modify the proposed algorithm to take into consideration the additive noise and set the quantization threshold to zero. The modified algorithm is described in Algorithm 2 where  $\bullet$  denotes the element-wise product and  $[\cdot]_i$  denotes the  $i^{\text{th}}$  element in a vector. The expected value and the variance in lines 6 and 7 are with respect to  $r \sim \mathcal{N}([\hat{\mathbf{p}}]_i, [\mathbf{v}_p^{t+1}]_i + \sigma_n^2)$ . Those in lines 12 and 13 are with respect to  $p_{h_v|[\hat{\mathbf{r}}^{t+1}]_i}(h_v|[\hat{\mathbf{r}}^{t+1}]_i)$  which is proportional to the product of  $p_{h_v}(h_v)$  and the Gaussian PDF with mean  $[\hat{\mathbf{r}}^{t+1}]_i$  and variance  $[\mathbf{v}_r^{t+1}]_i$ .

---

**Algorithm 2** One-bit GAMP
 

---

 1: **Initialize:**

$$t = 0, \hat{\mathbf{h}}_v^t = \mathbb{E}[\mathbf{h}_v], \mathbf{v}_{\mathbf{h}_v}^t = \text{Var}[\mathbf{h}_v], \hat{\mathbf{s}}^t = 0,$$

 2: **for**  $t = 1, \dots, T$  **do**

Measurement update:

3:  $\mathbf{v}_p^{t+1} = (\mathbf{W} \bullet \mathbf{W}) \mathbf{v}_{\mathbf{h}_v}^t,$

4:  $\hat{\mathbf{p}}^{t+1} = \mathbf{W} \mathbf{h}_v^t - \mathbf{v}_p^{t+1} \bullet \hat{\mathbf{s}}^t,$

 5: **for all**  $i$  **do**

6: 
$$[\hat{\mathbf{s}}^{t+1}]_i = \frac{1}{[\mathbf{v}_p^{t+1}]_i + \sigma_n^2} (\mathbb{E}[r | r \in Q^{-1}([\mathbf{y}]_i)] - [\hat{\mathbf{p}}^{t+1}]_i),$$

7: 
$$[\mathbf{v}_s^{t+1}]_i = \frac{1}{[\mathbf{v}_p^{t+1}]_i + \sigma_n^2} \left( 1 - \frac{\text{Var}[r | r \in Q^{-1}([\mathbf{y}]_i)]}{[\mathbf{v}_p^{t+1}]_i + \sigma_n^2} \right),$$

 8: **end for**

Estimation update:

9:  $\mathbf{v}_r^{t+1} = ((\mathbf{W} \bullet \mathbf{W})^\top \mathbf{v}_s^{t+1})^{-1},$

10:  $\hat{\mathbf{r}}^{t+1} = \hat{\mathbf{h}}_v^t + \mathbf{v}_r^{t+1} \bullet (\mathbf{W}^\top \hat{\mathbf{s}}^{t+1}),$

 11: **for all**  $i$  **do**

12: 
$$[\hat{\mathbf{h}}_v^{t+1}]_i = \mathbb{E}[h_v | [\hat{\mathbf{r}}^{t+1}]_i],$$

13: 
$$[\mathbf{v}_{\mathbf{h}_v}^{t+1}]_i = \text{Var}[h_v | [\hat{\mathbf{r}}^{t+1}]_i],$$

 14: **end for**

 15: **end for**

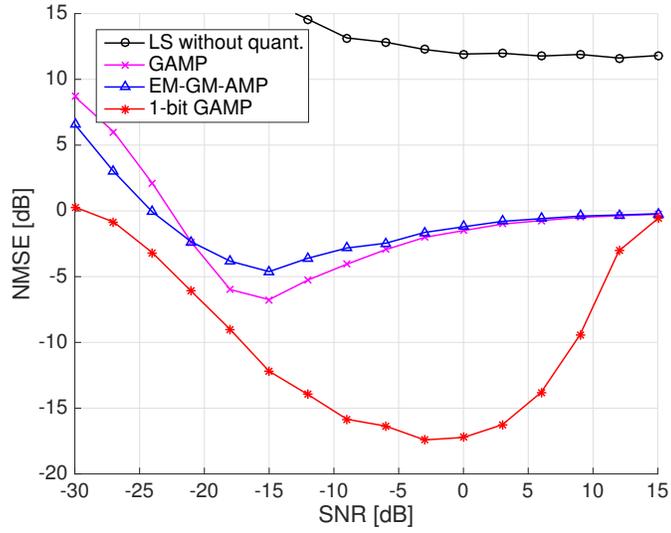

---

## 4.4 Numerical Results

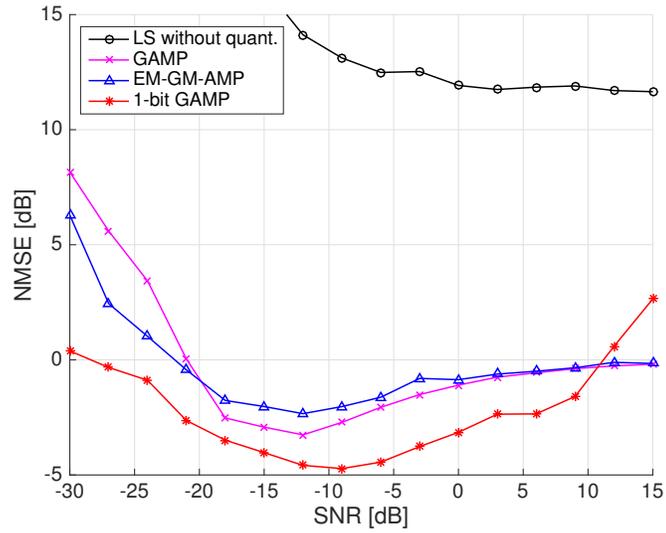
In this section, I evaluate channel estimation performance of the modified one-bit GAMP and compare it with other algorithms that include LS estimation, GAMP, and EM-GM-AMP [91]. The normalized mean squared error (NMSE) is used as a performance metric:

$$\text{NMSE} = \mathbb{E} \left[ \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2}{\|\mathbf{H}\|_F^2} \right].$$

For simulation, system parameters used in this section are as follows unless otherwise stated:  $N_t = 64$ ,  $N_r = 16$ ,  $L_t = 4$ ,  $L_r = 4$ ,  $N_p = 2$ , and



(a)



(b)

Figure 4.3: NMSE of four channel estimation algorithms: LS, GAMP, EM-GM-AMP and one-bit GAMP (proposed). The leakage effect is considered in (b) and degrades channel estimation performance.

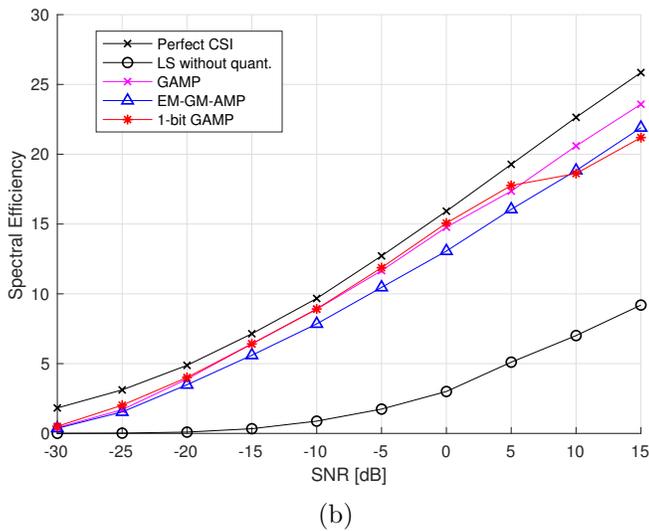
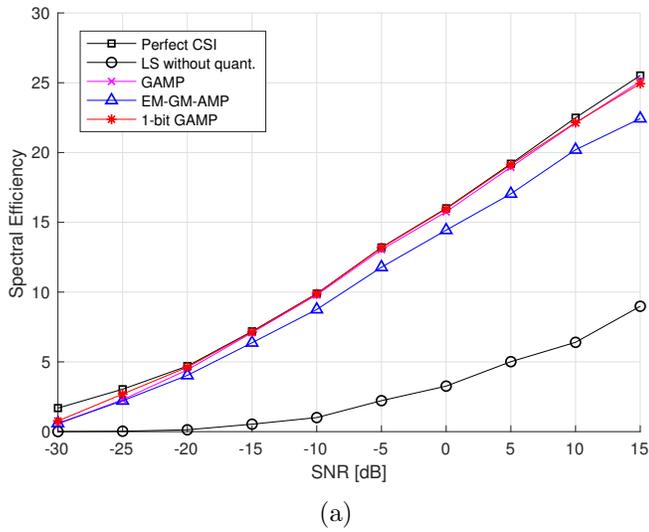


Figure 4.4: Spectral efficiency of four channel estimation algorithms: LS, GAMP, EM-GM-AMP and one-bit GAMP (proposed). Spectral efficiency with perfect CSI knowledge is also provided. The subfigures correspond to the ones in Fig. 4.3. As in Fig. 4.3, the leakage effect in (b) causes a slight degradation in spectral efficiency for the proposed algorithm.

$M = 64$ . Columns of a Hadamard matrix are used for training symbol vectors. The simulation parameters are provided in Table 4.3 as well.

Fig. 4.3 shows NMSE of four channel estimation algorithms. For comparison purposes, the LS estimator is without signal quantization while others use one-bit ADCs. Simulated channels for Fig. 4.3(a) are intentionally constructed to avoid the leakage effect whereas those for Fig. 4.3(b) do not have such constraint. As seen in both subfigures, the LS estimator without quantization achieves far worse performance than do GAMP variants with one-bit ADCs. One-bit GAMP yields the lowest estimation errors among the considered algorithms in both subfigures. Comparing two subfigures, I can see that all algorithms suffer performance degradation due to the leakage effect, which is described in Section 4.2. Figures henceforth are with one-bit GAMP and channels that experience the leakage effect.

Fig. 4.4 shows the spectral efficiency achieved by the four channel estimation algorithms when using the simulation parameters in Table 4.3. As with

Table 4.3: Simulation parameters

Notation	Description	Value
$N_t$	Number of transmit antennas	64
$N_r$	Number of receive antennas	16
$L_t$	Number of transmit RF chains	4
$L_r$	Number of receive RF chains	4
$N_p$	Number of channel MPC	4
$M$	Number of training frames	64

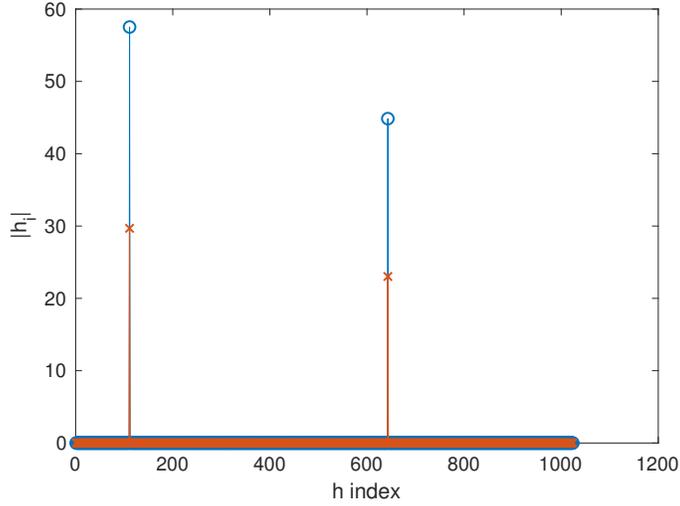


Figure 4.5: Comparison of true (blue) and estimated (red) path gain vectors at an SNR of 10 dB. The estimated path gain vector is denoted by  $\tilde{\mathbf{h}}_v$

Fig. 4.3, two subfigures are without and with the leakage effect, respectively. In terms of spectral efficiency, the proposed one-bit GAMP outperforms the other benchmark algorithms with an exception at a high SNR regime when there is leakage. An interesting observation by comparing the trend of spectral efficiency with that of NMSE is that the U-shape observed in the NMSE curves do not appear in the spectral efficiency curves. This is due to the fact that the channel subspace plays a greater role than do the path gains in spectral efficiency. In other words, the retrogressing NMSE in a high SNR regime is due to poor path gain estimation; however, beam angle estimation performance is of good quality. A sample from one of the channel realizations with 10 dB SNR is given in Fig. 4.5 where the true and estimated path gain vectors are compared. Note that indices of non-zero elements indicate transmit

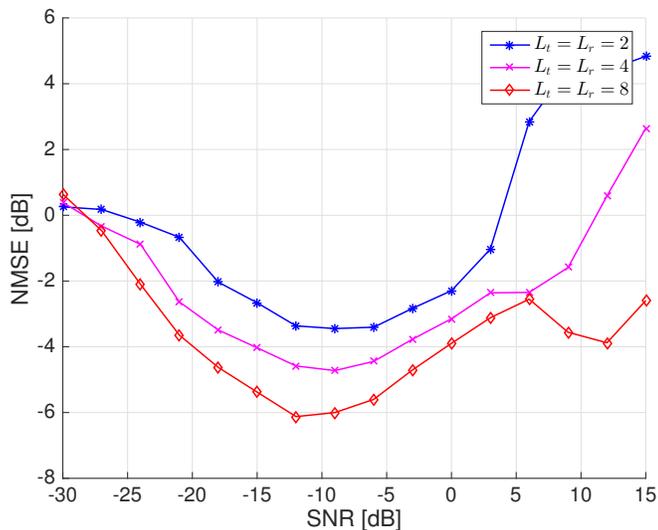


Figure 4.6: NMSE as a function of RF chains. More RF chains improve channel estimation performance.

and receive beam pairs. It can be seen that the true and estimated path gain vectors have the non-zero elements at the same indices, but their magnitudes are different. It means that the proposed one-bit GAMP performs great beam detection and poor path gain estimation.

The number of RF chains affects the channel estimation performance as shown in Fig. 4.6. Two, four and eight pairs of RF chains are plotted. More RF chains improve the performance across the considered SNR range. From a compressed sensing perspective, this is because more RF chains allow longer measurement vector.

Fig. 4.7 shows effects of the number of frames on channel estimation performance in various SNR regimes. As shown in Fig. 4.3(b), the minimum

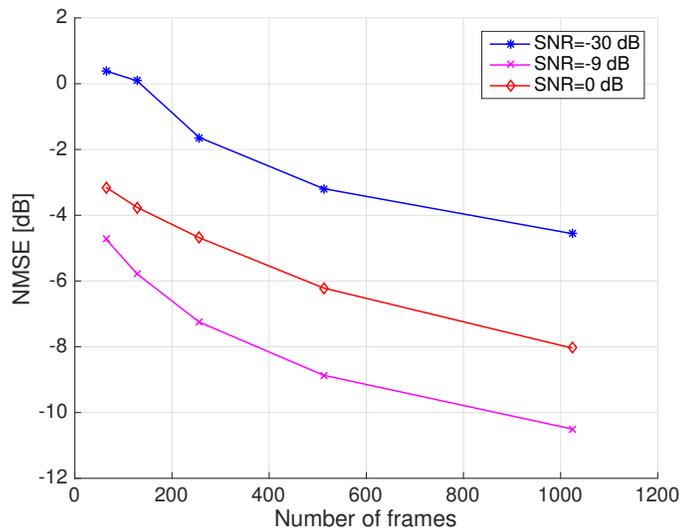


Figure 4.7: NMSE as a function of frames. For any SNR, estimation error decreases with increasing number of frames.

NMSE can be obtained around an SNR of -9 dB, which corresponds to the fact that the curve for -9 dB SNR in Fig. 4.7 has lower NMSE than the others. Regardless of SNR values, estimation error decreases as more frames are used for estimation. It is expected since mutual coherence of the measurement matrix decreases with the increasing number of frames for any SNR.

## 4.5 Conclusion

In this chapter, I proposed a channel estimation algorithm for a mmWave communication system with one-bit quantizers and hybrid beamforming based on GAMP. This one-bit GAMP method is specifically designed for measurements taken with one-bit ADCs, and I modified it to take into account the

thermal noise. Simulation results showed that GAMP variants with one-bit ADCs achieve better performance than LS without quantization, and that the proposed algorithm yields the lowest channel estimation error among the GAMP variants. Results also showed that the channel estimation performance can be enhanced by exploiting more frames and RF chains.

In the following chapter, I investigate yet another area where CS can show its efficiency – beam detection. 5G NR standard as well as multi-carrier communications are taken into account.

## Chapter 5

# Compressed-Sensing based Beam Detection in 5G New Radio Initial Access

To support millimeter wave (mmWave) frequency bands in cellular communications, both the base station and the mobile platform utilize large antenna arrays to steer narrow beams towards each other to compensate the path loss and improve communication performance. The time-frequency resource allocated for initial access, however, is limited, which gives rise to need for efficient approaches for beam detection. In this chapter<sup>1</sup>, for hybrid analog-digital beamforming (HB) architectures, which are used to reduce power consumption, I propose a compressed sensing (CS) based approach for 5G initial access beam detection that is for a HB architecture and that is compliant with the 3GPP standard. The CS-based approach is compared with the exhaustive search in terms of beam detection accuracy and by simulation is shown to outperform. Up to 256 antennas are considered, and the importance of a careful codebook design is reaffirmed.

---

<sup>1</sup> This chapter is based on the work published in J. Sung and B. L. Evans, “Compressed-sensing based beam detection in 5G NR initial access”, *in Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, May. 26–29, 2020 [103]. This work was supervised by Prof. Brian L. Evans.

## 5.1 Introduction

Directional transmission and reception link has become inevitable in 5G New Radio (NR) due to adoption of millimeter wave (mmWave) frequency bands. As a growing number of antennas are taken into account at base stations (gNB), many issues have been brought up to be addressed. A larger antenna array inherently comes with sharper beams, which in turn more beams are required to sweep and illuminate the entire cell coverage. From the beam management perspective, it implies more frequent beam losses and harder beam recovery. Especially in the initial access process, only a limited opportunity is given to gNB to sweep its possible beams in downlink, e.g., 64 synchronization signal blocks (SSB) in FR2 [104]. The limitation gives rise to need for an efficient approach to fully exploit the given resources.

Beam management at mmWave is extensively surveyed from the 3GPP perspectives in [105]. The hybrid beamforming (HB) architecture is shown to be a good trade-off in terms of reactivity and design complexity. However, the provided results only considers the exhaustive search (ES) for the sake of the standard compliance. Compressive sensing (CS) has been considered as a beam detection method in [106–109], but the literatures lack consideration of the standard and hybrid beamforming architectures. The iterative search (IS) is also a good candidate [110, 111], but is not considered in this chapter as (i) it is shown to exhibit higher misdetection probability in general in [112], (ii) it requires feedback from users at every iteration to narrow down the search space and (iii) it is not suitable for initial access due to the user specificity.

A qualitative comparison among the selected prior work and the proposed contribution is provided in Table 5.1 where CFO, PS and OH stand for carrier frequency offset, phase shifter and overhead, respectively. Note that exhaustive search (ES) beam detection has low complexity because it uses received beam power/SNR which requires a small amount of processing.

In this chapter, I propose a CS-based downlink beam detection approach for mmWave hybrid analog and digital beamforming communication systems that is compliant with the 5G NR 3GPP standard. The frequency

Table 5.1: Comparison among beam detection algorithms using compressive sensing (CS), exhaustive beam search (ES) and iterative search (IS)

	[110]	[106]	[107]	[111]	[108]	[109]	[105]	Proposed
Method	IS	CS	CS	2-way IS	CS	CS	ES	CS
Impairment	×	×	phase error	×	CFO	CFO	×	×
Hybrid	✓	×	×	×	×	×	✓	✓
Standard compliant	×	×	×	×	×	✓	✓	✓
Architecture	PS	N/A	N/A	N/A	N/A	N/A	PS	PS
Multi carrier	✓	×	×	×	×	✓	✓	✓
Deterministic	✓	✓	×	✓	△	×	✓	✓ & ×
Complexity	High OH	Med.	High	High OH	Med.	High	Low	Med.
Performance	High	High	High	High	High	High	Med.	High

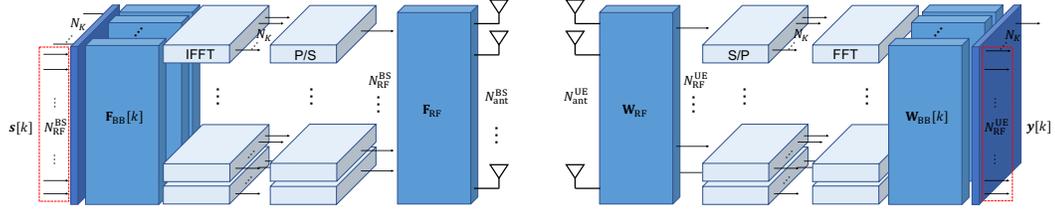


Figure 5.1: Block diagram of a general OFDM hybrid analog and digital beam-forming architecture at both a transmitter and a receiver.

and time structure of synchronization signal block is specifically taken into account in the design. The ES is considered as a baseline method which the CS approach is compared against using a beam pair detection probability as a performance metric.

## 5.2 System and Channel Models

In downlink, a gNB with  $N_{\text{ant}}^{\text{BS}}$  transmit antennas and  $N_{\text{RF}}^{\text{BS}}$  transmit RF chains and a UE with  $N_{\text{ant}}^{\text{UE}}$  receive antennas and  $N_{\text{RF}}^{\text{UE}}$  receive RF chains are considered. Since our focus is on HB architectures, the following inequalities hold:  $N_{\text{RF}}^{\text{BS}} < N_{\text{ant}}^{\text{BS}}$  and  $N_{\text{RF}}^{\text{UE}} < N_{\text{ant}}^{\text{UE}}$ . Then the multipath channel matrix can be expressed as

$$\begin{aligned} \mathbf{H}(\tau) &= \gamma \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \mathbf{H}_{i,l} \delta(\tau - \tau_{i,l}) \in \mathbb{C}^{N_{\text{ant}}^{\text{UE}} \times N_{\text{ant}}^{\text{BS}}} \\ &= \gamma \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{i,l} \delta(\tau - \tau_{i,l}) \mathbf{a}_{\text{UE}}(\theta_{i,l}) \mathbf{a}_{\text{BS}}^{\text{H}}(\phi_{i,l}), \end{aligned}$$

where  $\gamma = \sqrt{N_{\text{ant}}^{\text{UE}} N_{\text{ant}}^{\text{BS}} / N_{cl} N_{ray}}$  is the normalization factor,  $N_{cl}$  and  $N_{ray}$  denote the number of multipath clusters and the number of rays in each cluster,

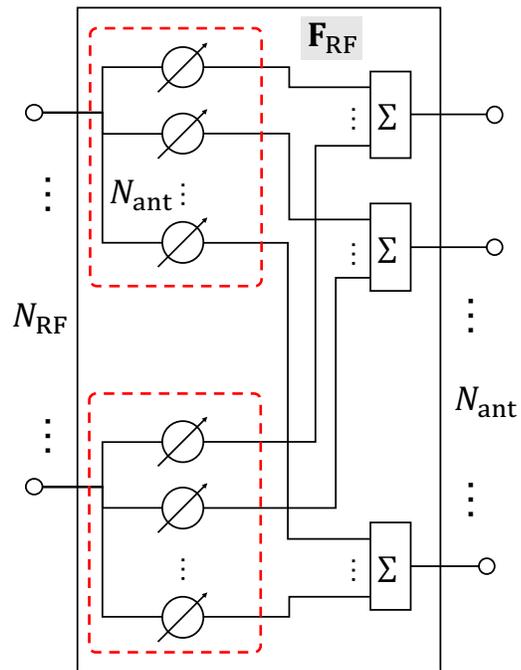


Figure 5.2: A diagram of the phase shifting RF stage of the precoder.

respectively,  $\alpha_{i,l} \sim \mathcal{CN}(0, \sigma_\alpha^2)$  is the  $(i, l)$ -th path complex gain with  $\sigma_\alpha^2 = 1$ ,  $\tau_{i,l}$  is the  $(i, l)$ -th path propagation delay, and  $\theta_{i,l}$  and  $\phi_{i,l}$  are the angles of arrival and departure (AoA and AoD) associated with the  $(i, l)$ -th path, respectively.  $\delta(\cdot)$  denotes the impulse function, and  $(\cdot)^H$  denotes the conjugate transpose matrix. The  $(i, l)$ -th path AoA and AoD are further defined as  $\theta_{i,l} = \theta_i + \Delta\theta_{i,l}$  and  $\phi_{i,l} = \phi_i + \Delta\phi_{i,l}$ , respectively.  $\mathbf{a}_{\text{UE}}(\theta_{i,l})$  and  $\mathbf{a}_{\text{BS}}(\phi_{i,l})$  are the receive and transmit array response vectors at the given angles. Assuming uniform linear arrays (ULAs) with half-wavelength spacing for both transmit and receive antennas, the transmit array response vector is given as

$$\mathbf{a}_{\text{BS}}(\phi) = \sqrt{\frac{1}{N_{\text{ant}}^{\text{BS}}}} [1, e^{j\pi \sin(\phi)}, \dots, e^{j\pi(N_t-1)\sin(\phi)}]^\top,$$

where  $(\cdot)^\top$  denotes the matrix transpose. The receive array response is similarly defined. By transforming into the frequency domain, the channel matrix for the  $k$ -th subcarrier is given by

$$\mathbf{H}[k] = \gamma \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{i,l} e^{-\frac{j2\pi f_s \tau_{i,l} k}{K}} \mathbf{a}_{\text{UE}}(\theta_{i,l}) \mathbf{a}_{\text{BS}}^H(\phi_{i,l}), \quad (5.1)$$

where  $f_s$  is the sample rate of ADCs at the UE and  $K$  is the total number of subcarriers in OFDM. For the sake of simplicity, it is assumed that the propagation delays of rays in a given cluster are identical, i.e.,  $\tau_i = \tau_{i,l}, \forall l$ .

The transmit array response matrix  $\mathbf{A}_{\text{BS}}$  can be constructed by collecting the transmit array responses evaluated at  $N_{cl}N_{ray}$  AoDs as

$$\mathbf{A}_{\text{BS}} = [\mathbf{a}_{\text{BS}}(\phi_{1,1}), \mathbf{a}_{\text{BS}}(\phi_{1,2}), \dots, \mathbf{a}_{\text{BS}}(\phi_{N_{cl}, N_{ray}})],$$

and the receive array response matrix  $\mathbf{A}_{\text{UE}}$  can also be similarly constructed. The channel matrix in (5.1) can be rewritten with the transmit and receive array response matrices as

$$\mathbf{H}[k] = \mathbf{A}_{\text{UE}}\mathbf{H}_d[k]\mathbf{A}_{\text{BS}}^H, \quad (5.2)$$

where  $\mathbf{H}_d[k] \in \mathbb{C}^{N_{cl}N_{ray} \times N_{cl}N_{ray}}$  is the diagonal matrix with the elements being the scaled complex multipath component (MPC) path gain associated with a pair of an AoD and an AoA.

Considering a single UE, the received signal in the  $m$ -th block and the  $k$ -th subcarrier is denoted by  $\mathbf{y}_m[k] \in \mathbb{C}^{N_{\text{RF}}^{\text{UE}} \times 1}$  and can be expressed as

$$\begin{aligned} \mathbf{y}_m[k] &= \sqrt{\rho}\mathbf{W}_m[k]^H\mathbf{H}[k]\mathbf{F}_m[k]\mathbf{s}_m[k] + \mathbf{W}_m[k]^H\mathbf{z}_m[k] \\ &= \sqrt{\rho}\mathbf{W}_m[k]^H\mathbf{H}[k]\mathbf{x}_m[k] + \mathbf{n}_m[k], \end{aligned}$$

where  $\rho$  is the transmit power of each subcarrier, and  $\mathbf{W}_m[k] \in \mathbb{C}^{N_{\text{ant}}^{\text{UE}} \times N_{\text{RF}}^{\text{UE}}}$  and  $\mathbf{F}_m[k] \in \mathbb{C}^{N_{\text{ant}}^{\text{BS}} \times N_{\text{RF}}^{\text{BS}}}$  are the receive combiner at the UE and the transmit precoder at the gNB, respectively. The combiner/precoder are a product of the RF and baseband (BB) combiner/precoder, in other words,  $\mathbf{W}_m[k] = \mathbf{W}_{\text{RF},m}\mathbf{W}_{\text{BB},m}[k]$  and  $\mathbf{F}_m[k] = \mathbf{F}_{\text{RF},m}\mathbf{F}_{\text{BB},m}[k]$ . Note that the RF combiner and precoder are frequency-flat whereas BB ones are frequency dependent.  $\mathbf{s}_m[k] \in \mathbb{C}^{N_{\text{RF}}^{\text{BS}}}$  is the transmit symbol vector, and  $\mathbf{x}_m[k]$  denotes  $\mathbf{F}_m[k]\mathbf{s}_m[k]$ .  $\mathbf{z}_m[k] \in \mathbb{C}^{N_{\text{ant}}^{\text{UE}}}$   $\sim \mathcal{CN}(0, \sigma_n^2\mathbf{I})$  is the noise vector, and  $\mathbf{n}_m[k]$  denotes  $\mathbf{W}_m[k]^H\mathbf{z}_m[k]$ .  $k$ ,  $m$  and  $n$  denote the index of the subcarrier, the SRS, the user, respectively. The transmit SNR is defined by  $\rho/\sigma_n^2$ . Fig. 5.1 shows a block diagram of a general

hybrid beamforming architectures by visualizing the signal flows in an OFDM communication system.

In this chapter, I consider that the RF stage of the precoder and the combiner is implemented with phase shifters and that the phase shifters have a  $b_{PS}$ -bit resolution. Therefore, column vectors of the RF precoder  $\mathbf{F}_{\text{RF},m}$  are selected from a set  $\mathcal{F} = \{\mathbf{f} \in \mathbb{C}^{N_{\text{ant}}^{\text{BS}}} : |\mathbf{f}_i| = \sqrt{1/N_{\text{ant}}^{\text{BS}}}, \angle \mathbf{f}_i \in \Theta\}$  where  $\Theta = \{\theta : 2\pi n/2^{b_{PS}}, n = 0, \dots, 2^{b_{PS}} - 1\}$  and  $\mathbf{f}_i$  denotes the  $i$ -th element of the vector  $\mathbf{f}$ . Fig. 5.2 illustrates a detailed view of phase shifting network based RF precoder.

### 5.3 Sparse Formulation

For initial access and beam detection purposes, it is assumed that gNB uses a precoding codebook that contains  $M_{\text{BS}}$  precoding matrices to illuminate its coverage. While the gNB sweeps the codebook, the UE keeps one combining matrix and then switch to the next one in the subsequent gNB codebook sweep. Assuming the combining codebook has  $M_{\text{UE}}$  combining matrices, the UE can collect up to  $M (= M_{\text{BS}}M_{\text{UE}})$  blocks. For simplicity, I suppose that one block contains a single OFDM symbol. Assuming that the channel remains constant over  $M$  blocks, the collected received signal vectors  $\mathbf{y}_{1,\dots,M}[k]$  at the  $k$ -th subcarrier can be combined into a matrix  $\mathbf{Y}[k] \in \mathbb{C}^{N_{\text{RF}}^{\text{UE}} M_{\text{UE}} \times M_{\text{BS}}}$  which can be expressed as

$$\mathbf{Y}[k] = [\bar{\mathbf{y}}_1[k], \dots, \bar{\mathbf{y}}_{M_{\text{BS}}}[k]]$$

$$\begin{aligned}
&= \sqrt{\rho} \mathbf{W}[k]^H \mathbf{H}[k] \mathbf{X}[k] + \mathbf{W}[k]^H \mathbf{Z}[k] \\
&= \sqrt{\rho} \mathbf{W}[k]^H \mathbf{H}[k] \mathbf{X}[k] + \mathbf{N}[k],
\end{aligned} \tag{5.3}$$

where  $\bar{\mathbf{y}}_i[k] = [\mathbf{y}_{(i-1)M_{\text{UE}}+1}[k]^T, \dots, \mathbf{y}_{iM_{\text{UE}}}[k]^T]^T$ ,  $\mathbf{W}[k] = [\mathbf{W}_1[k], \mathbf{W}_2[k], \dots, \mathbf{W}_{M_{\text{UE}}}[k]] \in \mathbb{C}^{N_{\text{ant}}^{\text{UE}} \times N_{\text{RF}}^{\text{UE}} M_{\text{UE}}}$ ,  $\mathbf{X}[k] = [\mathbf{x}_1[k], \mathbf{x}_2[k], \dots, \mathbf{x}_{M_{\text{BS}}}[k]] \in \mathbb{C}^{N_{\text{ant}}^{\text{BS}} \times M_{\text{BS}}}$ , and

$$\begin{aligned}
\mathbf{N}[k] &= \text{blkdiag}\{\mathbf{W}_1[k], \dots, \mathbf{W}_{M_{\text{BS}}}[k]\}^H \times \\
&\quad \begin{bmatrix} \mathbf{n}_1[k] & \mathbf{n}_{M_{\text{UE}}+1}[k] & \dots & \mathbf{n}_{(M_{\text{BS}}-1)M_{\text{UE}}+1}[k] \\ \mathbf{n}_2[k] & \mathbf{n}_{M_{\text{UE}}+2}[k] & \dots & \mathbf{n}_{(M_{\text{BS}}-1)M_{\text{UE}}+2}[k] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{n}_{M_{\text{UE}}}[k] & \mathbf{n}_{2M_{\text{UE}}}[k] & \dots & \mathbf{n}_{M_{\text{BS}}M_{\text{UE}}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{z}_1[k] & \mathbf{z}_{M_{\text{BS}}+1}[k] & \dots & \mathbf{z}_{M-M_{\text{BS}}+1}[k] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_{M_{\text{BS}}}[k] & \mathbf{z}_{2M_{\text{BS}}}[k] & \dots & \mathbf{z}_M[k] \end{bmatrix}.
\end{aligned}$$

By vectorizing both sides of (5.3), the following equation can be obtained:

$$\begin{aligned}
\mathbf{y}[k] &= \sqrt{\rho} (\mathbf{X}[k]^T \otimes \mathbf{W}[k]^H) \text{vec}(\mathbf{H}[k]) + \mathbf{n}[k] \\
&= \sqrt{\rho} \mathbf{\Phi}[k] \text{vec}(\mathbf{H}[k]) + \mathbf{n}[k] \in \mathbb{C}^{N_{\text{RF}}^{\text{UE}} M \times 1},
\end{aligned} \tag{5.4}$$

where  $\mathbf{\Phi}[k] = \mathbf{X}[k]^T \otimes \mathbf{W}[k]^H \in \mathbb{C}^{N_{\text{RF}}^{\text{UE}} M \times N_{\text{ant}}^{\text{UE}} N_{\text{ant}}^{\text{BS}}}$  is the sensing matrix for the  $k$ -th subcarrier, and  $\mathbf{n}[k] = \text{vec}(\mathbf{N}[k]) \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{M_{\text{UE}}} \otimes \mathbf{W}[k]^H \mathbf{W}[k])$ . The matrix operator  $\otimes$  denotes the Kronecker product, and  $\text{vec}(\cdot)$  is the vectorization operation.

I define the transmit array response grid matrix by

$$\bar{\mathbf{A}}_{\text{BS}} = [\mathbf{a}_{\text{UE}}(\vartheta_0), \mathbf{a}_{\text{UE}}(\vartheta_1), \dots, \mathbf{a}_{\text{UE}}(\vartheta_{G_{\text{BS}}-1})],$$

$\in \mathbb{C}^{N_{\text{ant}}^{\text{BS}} \times G_{\text{BS}}}$  where  $G_{\text{BS}} = N_{\text{ant}}^{\text{BS}} K_{\text{BS}}$  is the number of transmit angle grid bins,  $K_{\text{BS}}$  is the transmit angle grid multiplier, and  $\vartheta_i = \sin^{-1}(\frac{-i}{G_{\text{BS}}})$  is the transmit grid angle. The receive array response grid matrix  $\bar{\mathbf{A}}_{\text{UE}}$  is similarly defined. Ignoring errors that may be caused by angle grid quantization, the channel matrix in (5.2) can be rewritten as

$$\mathbf{H}[k] = \bar{\mathbf{A}}_{\text{UE}} \bar{\mathbf{H}}[k] \bar{\mathbf{A}}_{\text{BS}}^{\text{H}},$$

and  $\text{vec}(\mathbf{H}[k])$  can be expressed as

$$\begin{aligned} \text{vec}(\mathbf{H}[k]) &= (\bar{\mathbf{A}}_{\text{BS}}^* \otimes \bar{\mathbf{A}}_{\text{UE}}) \text{vec}(\bar{\mathbf{H}}[k]) \\ &= (\bar{\mathbf{A}}_{\text{UE}}^* \otimes \bar{\mathbf{A}}_{\text{BS}}) \mathbf{h}[k] \\ &= \mathbf{\Psi} \mathbf{h}[k], \end{aligned}$$

where  $\mathbf{\Psi} = \bar{\mathbf{A}}_{\text{BS}}^* \otimes \bar{\mathbf{A}}_{\text{UE}} \in \mathbb{C}^{N_{\text{ant}}^{\text{BS}} N_{\text{ant}}^{\text{UE}} \times G_{\text{BS}} G_{\text{UE}}}$  is the sparsifying dictionary matrix,  $\bar{\mathbf{H}}[k]$  is the modified MPC gain matrix associated with the dictionary matrix, and  $\mathbf{h}[k] = \text{vec}(\bar{\mathbf{H}}[k])$  is a sparse vector with  $N_{\text{cl}} N_{\text{ray}}$  non-zero elements. Then the received signal in the right hand side of (5.4) can be rewritten as

$$\begin{aligned} \mathbf{y}[k] &= \sqrt{\rho} \mathbf{\Phi}[k] \mathbf{\Psi} \mathbf{h}[k] + \mathbf{n}[k]. \\ &= \sqrt{\rho} \left( \tilde{\mathbf{X}}[k] \otimes \tilde{\mathbf{W}}[k] \right) \mathbf{h}[k] + \mathbf{n}[k]. \end{aligned} \quad (5.5)$$

Detection of pairs of transmit and receive beams can be achieved by finding non-zero elements in the sparse vector  $\mathbf{h}[k]$ , and various CS methods now can be employed for this end. The sparse formulation in (5.5) is for the  $k$ -th subcarrier and can directly be explored to detect beams. In that case,

beam detection should be performed  $N_{rs}$  times and combine the detected beam pairs where  $N_{rs}$  denotes the number of subcarriers that can be exploited for beam detection purposes.

Instead, the received signal vector  $\mathbf{y}[k]$  for  $k \in \{1, \dots, N_{rs}\}$  can be concatenated based on an assumption that the transmit and receive array responses, i.e,  $\mathbf{a}_{UE}(\theta)$  and  $\mathbf{a}_{BS}(\phi)$ , are common in all subcarriers. The support in  $\mathbf{h}[k]$  – which indicates angle grid points – is common for all subcarriers and is denoted by  $\mathbf{h}$  without the subcarrier index. This is a reasonable assumption considering (1) a relatively narrow frequency span of SSB and (2) angle estimation performed by only finding indices of a few largest support. Denoting  $[\mathbf{y}[1]^\top, \mathbf{y}[2]^\top, \dots, \mathbf{y}[K]^\top]^\top$  by  $\mathbf{y}$ , the following equation can be obtained:

$$\begin{aligned} \mathbf{y} &= [\mathbf{y}[1]^\top, \mathbf{y}[2]^\top, \dots, \mathbf{y}[N_{rs}]^\top]^\top \\ &= \sqrt{\rho} \begin{bmatrix} \Phi[1] \\ \vdots \\ \Phi[N_{rs}] \end{bmatrix} \Psi \mathbf{h} + \begin{bmatrix} \mathbf{n}[1] \\ \vdots \\ \mathbf{n}[N_{rs}] \end{bmatrix} \\ &= \sqrt{\rho} \Phi \Psi \mathbf{h} + \mathbf{n} \\ &= \sqrt{\rho} \bar{\Phi} \mathbf{h} + \mathbf{n}, \end{aligned}$$

where the sensing matrix  $\Phi = [\Phi[1]^\top, \dots, \Phi[N_{rs}]^\top]^\top$  is a stack of sensing matrices of subcarriers, and  $\mathbf{n} = [\mathbf{n}[1]^\top, \dots, \mathbf{n}[N_{rs}]^\top]^\top$  is the concatenated noise vectors.  $\tilde{\mathbf{X}}[k] = \mathbf{X}[k]^\top \bar{\mathbf{A}}_{UE}^*$  and  $\tilde{\mathbf{W}}[k] = \mathbf{W}[k]^\mathbf{H} \bar{\mathbf{A}}_{BS}$ . The objective is to solve the optimization problem

$$\min \|\mathbf{h}\|_1 \text{ such that } \|\mathbf{y} - \sqrt{\rho} \Phi \Psi \mathbf{h}\|_2 < \epsilon$$

in order to find the downlink beam pairs.

## 5.4 Initial Access

3GPP NR defines the concept of synchronization signal block (SSB) that gNB periodically transmits in a bursty manner for multiple purposes including the initial access (IA). One SSB is composed of the primary and secondary synchronization signals (PSS and SSS), the physical broadcast channel (PBCH) and the demodulation reference signal (DMRS). One SSB spans on four OFDM symbols and 240 subcarriers in time and frequency domain, respectively. The burst periodicity is configurable, i.e., 5, 10, 20, 40, 80, 160 ms, and is generally set to 20 ms for IA. The number of SSB in each burst is also configurable, i.e., 4, 8, 64 and determined by the frequency band. In this chapter, 64 SSB in a burst is assumed as I am targeting the mmWave frequencies. For more details, refer to [113, 114]. The IA procedure consists of four stages: beam sweeping, beam measurement, beam determination and beam reporting. During the beam sweeping stage, at a given time, one SSB is transmitted in a single beam toward a pre-specified direction. The subsequent SSB will be transmitted in another beam so that the gNB illuminate the cell coverage. During beam sweeping, multiple SSB's are transmitted in different pre-specified beams to illuminate the cell coverage. Single beam transmission is analogous to the analog beamforming. In the beam measurement stage, for a given transmitted SSB, each HB architecture UE collects received signal measurements from up to  $N_{\text{RF}}^{\text{UE}}$  directions. In the subsequent beam determination stage, various approaches can be applied to estimate the best beam.

In this chapter, I compare the CS approach against the ES in beam

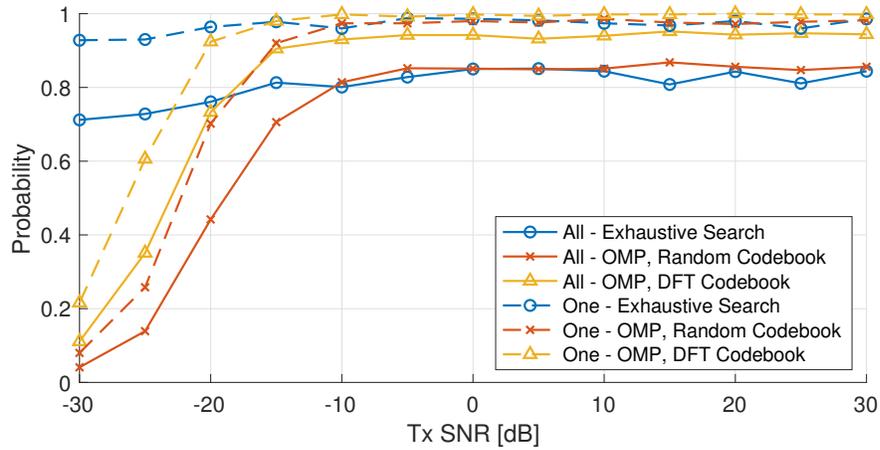


Figure 5.3: Beam detection probability as a function of the transmit SNR.

determination. For the ES, a receive SNR that can be obtained in the beam measurements stage is used as a metric in finding beams. For the CS approach, the sparse formulation derived in Sec. 5.3 is used and a CS algorithm is applied, and there are two precoding and combining options for this approach: the random and the deterministic. The random codebook randomly configures the phase shifters in the RF stage whereas the deterministic codebook makes use of a predetermined set of matrices for the phase shifters. In [57], it is shown that the DFT codebook is one of those that minimize the total coherence and can better estimate mmWave channels than the random codebook does.

## 5.5 Numerical Results

For simulation, the following parameters are used unless otherwise specified:  $N_{\text{ant}}^{\text{BS}} \in \{64, 128, 256\}$ ,  $N_{\text{RF}}^{\text{BS}} = 8$ ,  $N_{\text{ant}}^{\text{UE}} = 8$ ,  $N_{\text{RF}}^{\text{UE}} = 4$ ,  $b_{\text{PS}} = 6$ , 120 kHz

subcarrier spacing, 4096 FFT size, 400 MHz BW, 491.52 MS/s sample rate,  $N_{rs} = 10$  middle subcarriers,  $K_{BS} = K_{UE} = 3$ ,  $M_{BS} = 64$ ,  $M_{UE} = 2$ ,  $N_c = 2$ , and  $N_{ray} = 3$ . The path delay is uniformly distributed from 0 to 200 ns, the cluster means AoD and AoA are uniformly distributed in  $[-\pi/2, \pi/2]$ , and the rays have the Laplace distribution with a cluster mean and a  $2^\circ$  standard deviation. The simulation parameters are also provided in Table 5.2. For each figure, 500 channel realizations are generated. As a representative of the CS approach, orthogonal matching pursuit (OMP) is used. The all beam detection probability is the probability that the estimated beam pairs match the true pairs, and the single beam detection probability is the probability that at least one estimated beam pair matches to one of the true pairs.

Table 5.2: Simulation parameters

Notation	Description	Value
$N_{\text{ant}}^{\text{BS}}$	Number of BS antennas	64, 128, 256
$N_{\text{ant}}^{\text{UE}}$	Number of UE antennas	8
$N_{\text{RF}}^{\text{BS}}$	Number of BS RF chains	8
$N_{\text{RF}}^{\text{UE}}$	Number of UE RF chains	4
$N_{rs} = 10$	Number of reference symbol subcarriers	10
$K_{\text{BS}}$	Transmit angle grid multiplier	3
$K_{\text{UE}}$	Receive angle grid multiplier	3
$G_{\text{BS}} = N_{\text{ant}}^{\text{BS}} K_{\text{BS}}$	Transmit angle grid size	192, 384, 768
$G_{\text{UE}} = N_{\text{ant}}^{\text{UE}} K_{\text{UE}}$	Transmit angle grid size	24
$M_{\text{BS}}$	Precoding codebook configurations	64
$M_{\text{UE}}$	Combining codebook configurations	2
$b_{\text{PS}}$	Phase shifter resolution	6
$N_c$	Number of channel clusters	2
$N_{ray}$	Number of rays in a cluster	3

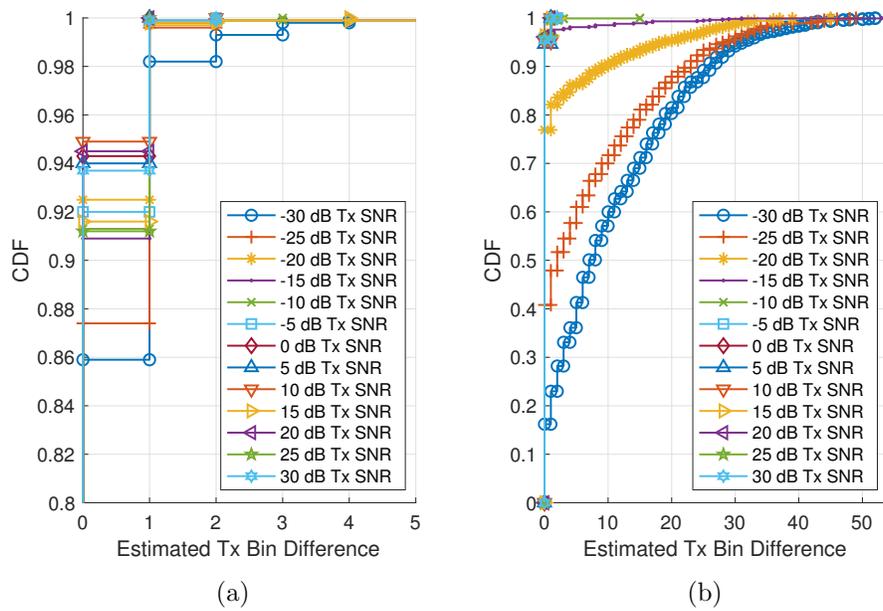


Figure 5.4: Empirical CDF of estimated transmit beam detection errors: (a) ES and (b) CS

Fig. 5.3 shows both the all and single beam detection probabilities with three different approaches: the ES, OMP with the random codebook and OMP with the DFT codebook. For the all beam detection probability, OMP with the DFT codebook achieves the highest detection probability in the medium and high SNR regimes. In low SNR, the ES has higher probability than the other approaches until  $-20$  dB SNR. OMP with the random codebook's probability is the lowest in the low SNR regime and is similar to that of the ES in the medium and high SNR regimes. The figure shows that the ES is more robust to a low SNR and that high SNR is more favorable to the CS approach to better estimate the beam pairs. In addition, a smart codebook choice is critical in achieving good performance when using a CS approach. A similar trend is observed in the single beam detection probability with smaller performance gaps. All considered approaches have a very high ( $> 0.95$ ) probability with an SNR greater than  $-10$  dB, and especially, OMP with the DFT codebook shows a consistently high probability.

I take a closer look at difference in detection probability between the ES and CS approach by plotting CDF curves in Fig. 5.4 and 5.5. They plot empirical CDF of beam detection errors in terms of beam index. Fig. 5.4 and 5.5 are for the transmit and receive beams, respectively, and the subfigures (a) and (b) are for the ES and the CS approach, respectively. For each figure, 13 CDF curves for a range of a transmit SNR from  $-30$  to  $30$  dB with an increment of  $5$  dB are provided.

The transmit beam detection errors in Fig. 5.4 show that the both ap-

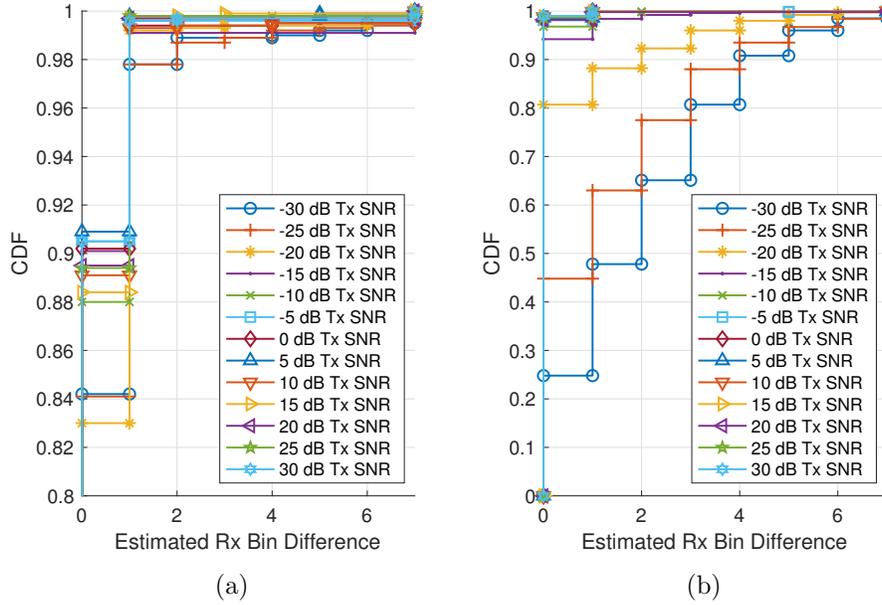
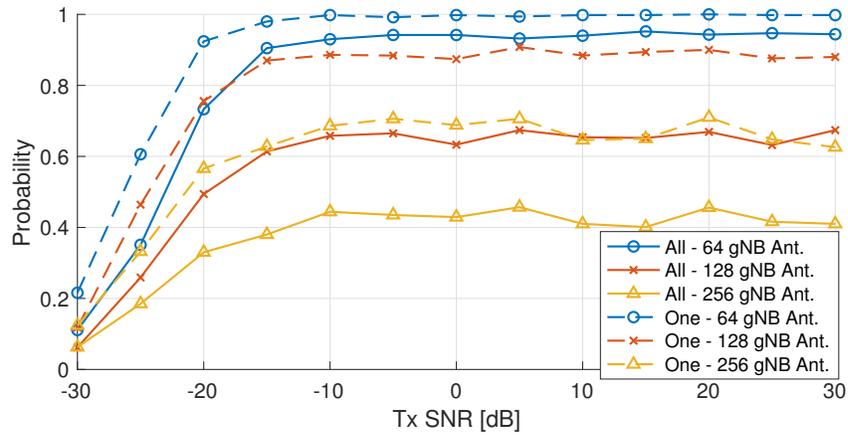


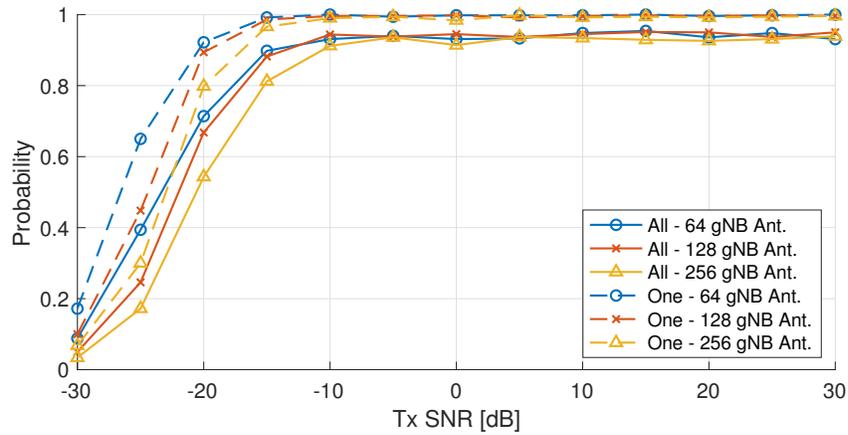
Figure 5.5: Empirical CDF of estimated receive beam detection errors: (a) ES and (b) CS

proaches can achieve similar high detection accuracy ( $\sim 95\%$  at zero difference) in the high and medium SNR regimes (e.g.,  $-15$  to  $30$  dB SNR); however, the receive beam detection accuracy from the ES is worse than the CS approach as observed in Fig. 5.5. It is a primary factor that causes the performance gap between the two approaches in Fig. 5.3. It implies that the CS-based approach is a better candidate if the system cannot tolerate any beam errors. In the low SNR regime (e.g.,  $-30$  to  $-20$  dB SNR), the CS approach yields lower accuracy than the ES in both transmit and receive beam detection. Thus, the ES may be considered for cell-edge users.

Now I take more than 64 gNB antennas into consideration: 128 and



(a)



(b)

Figure 5.6: Beam detection probability as a function of the transmit SNR: (a) DFT codebook and (b) codebook proposed in [57]

256, and the beam detection probability is provided in Fig. 5.6. Since the maximum number of SSB in a burst is 64 and the number of beams in the DFT codebook is the same as the transmit antennas, 128 and 256 antenna gNB cannot make use of the ES. Thus, in Fig. 5.6, I focus on the CS approach.

Each SSB is transmitted using either (i)  $N_{\text{ant}}^{\text{BS}}/L$  DFT beams blindly, e.g., 2 beams and 4 beams in 128 and 256 antenna systems (Fig. 5.6a) or (ii) all available beams using the carefully designed codebook [57] (Fig. 5.6b). In order for further performance improvement, the deterministic ordering [89] may be considered. The two codebooks achieves similar performance with 64 antennas; however, detection probabilities are degraded with a greater number of antennas in Fig. 5.6a whereas the probabilities remain almost unchanged in Fig. 5.6b. Comparison of the two subfigures suggests that careful codebook design is crucial in maintaining beam detection accuracy even with a growing number of antennas.

## 5.6 Conclusion

I proposed a CS-based downlink beam detection approach for mmWave hybrid beamforming systems taking the SSB structure in the 5G NR 3GPP standard into account. With the exhaustive search being a benchmark, the CS approach was evaluated using the random and the DFT RF codebooks in terms of the beam detection probability. The simulation results showed that (1) the detection probability increases with transmit SNR until a saturation point, (2) OMP with the DFT codebook achieves the highest detection probability after saturation and (3) the exhaustive search provides the least performance degradation due to low SNR. I further looked into the beam detection errors to figure out the receive detection error in the exhaustive search primarily contributes the lower performance. I also considered a larger number

of antennas at the gNB and explored two RF codebook options. It is observed that a smart choice of codebook is crucial in maintaining a beam detection capability with various number of antenna elements.

# Chapter 6

## Concluding Remarks

This chapter concludes the dissertation and provide a contribution summary and potential future work.

### 6.1 Summary

In this dissertation, I investigated the use of compressed sensing algorithms for mmWave CSI acquisition purposes in hybrid analog and digital beamforming architectures. To this end, I developed the analog and digital codebook design frameworks for HB architectures with infinite resolution ADCs and the CS algorithm for a HB architecture with one-bit resolution ADCs. In addition, I proposed and investigated a CS based approach for 5G initial access beam detection for a HB architecture that is compliant with the 3GPP standard.

The codebook design frameworks are discussed in Chapters 2 and 3. In those chapters, I developed a deterministic hybrid beamforming codebook design frameworks for CS algorithms. By minimizing the coherence of the sensing matrix, the analog and digital codebooks are obtained. It was shown that when with the full training the proposed codebook outperforms the ran-

dom codebook in all considered scenarios (i.e., various numbers of grid size, RF chains and channel sparsity) in terms of both the channel estimation error and the spectral efficiency. When using the partial training, the deterministic codebook achieves better performance in practical settings which implies the need of proper codebook selection.

The development of the CS-based channel estimation algorithm for a HB architecture equipped with one-bit quantizers is discussed in Chapter 4. General CS and conventional channel estimation algorithms hardly function due to combination of two sets of hardware constraints: HB and one-bit quantization. I modified one-bit GAMP to adapt to a communication system model and showed the modified one-bit GAMP outperforms other GAMP variants and a conventional algorithm.

The investigation of CS-based beam detection in 5G NR is discussed in Chapter 5. For the sake of compliance with the standard, I took the signal structure into account in the investigation. With the exhaustive search being a benchmark, the CS approach was evaluated to show its higher detection rate especially in medium and high SNR regimes. I took one step further and considered a larger number of antennas at the gNB to show a smart choice of codebook is crucial in maintaining a beam detection capability.

## 6.2 Future Work

- Wideband channel model

Except for the CS-based beam detection approach, other works in this

dissertation assume the narrowband channel model. Modern cellular networks exploit wideband channels, and they have become even wider by adopting mmWave bands. If OFDM is considered to address frequency selectivity of the wideband channels, it would be straightforward to extend the existing work in this dissertation as channels for each subcarrier in OFDM is assumed to be narrowband. If single carrier wideband channels are taken into account, it may become a bit more complicated and may require an overhaul of the problem formulation.

- Complexity reduction

Extension to the wideband channel model using OFDM may be deemed simple as the same algorithms/techniques can be applied on each subcarrier. However, it causes a linear computational complexity increase proportional to the number of subcarriers that contain a reference signal for CSI acquisition. One possible way to reduce the computational complexity is to exploit the fact that the angles of arrival and departure of each multipath component is shared among subcarriers. It already has been partially explored in the last work in this dissertation and can be extended for full CSI acquisition purposes.

- Hardware impairments

One common challenge of CSI acquisition based on parametric channel models is that it may break down due to some hardware impairments that are not taken into consideration in the model. The hardware impairments include uplink/downlink timing and frequency offset and drift

which are caused by a frequency mismatch between a transmitter and a receiver, imperfect antenna designs, antenna mutual coupling, imperfect calibration for channel reciprocity, etc. As not all of these may cause issues or some may have little impact on performance, it seems ideal to build a proof-of-concept system on flexible hardware such as software-defined radio so that quick rounds of implementation and testing is possible.

- Enhancement of the proposed contributions

In the codebook design and channel estimation algorithm development, knowledge of the noise covariance matrix can be exploited to perform de-noising in baseband receivers. For the channel estimation algorithm proposed in Chapter 4, effects of training sequences on channel estimation performance can be analyzed and optimized. In order to actively address the retrogression of NMSE in the high SNR regime, a dithering technique can be employed prior to one-bit ADCs [115].

Lastly, for all contributions, end-to-end communication system performance measure (e.g., bit error rate and block error rate) could be simulated and compared.

## Bibliography

- [1] Qualcomm, “Making 5G NR a commercial reality.” [Online]. Available: <https://www.qualcomm.com/documents/accelerating-5g-new-radio-nr-enhanced-mobile-broadband-and-beyond>
- [2] Y. Nam, M. S. Rahman, Y. Li, G. Xu, E. Onggosanusi, J. Zhang, and J. Seol, “Full dimension MIMO for LTE-Advanced and 5G,” in *Information Theory and Applications Workshop*, Feb. 2015, pp. 143–148.
- [3] Q. Nadeem, A. Kammoun, M. Debbah, and M. Alouini, “Design of 5G full dimension massive MIMO systems,” *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 726–740, 2018.
- [4] “3GPP Release 15,” <https://www.3gpp.org/release-15>, accessed: 2020-01-20.
- [5] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.
- [6] 3GPP TS 36.201, “Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer; General description (v8.3.0 Release 8),” Apr. 2009.
- [7] IEEE Std 802.11n-2009, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput,” Oct. 2009.

- [8] IEEE Std 802.11ac-2013, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications—Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz.” Dec. 2013.
- [9] 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.
- [10] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [11] 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.
- [12] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [13] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, “Millimeter-wave massive MIMO: the next wireless revolution?” *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

- [14] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.
- [15] M. Zhou, "FCC kicks off 5G spectrum auction to help make hype reality," *CNET*. [Online]. Available: <https://www.cnet.com/news/5g-craze-continues-the-5g-spectrum-auction-is-starting/>
- [16] 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.
- [17] J. Liu, Z. Luo, and X. Xiong, "Low-resolution ADCs for wireless communication: A comprehensive survey," *IEEE Access*, vol. 7, pp. 91 291–91 324, 2019.
- [18] 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.
- [19] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [20] 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 Int. Works. Signal Process. Advances in Wireless Commun.*, Jun. 2012, pp. 100–104.
- [21] A. Alkhateeb, G. Leus, and R. W. Heath, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [22] F. Sotiraki 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.
- [23] R. H. Walden, “Analog-to-digital converter survey and analysis,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [24] J. Mo and R. W. Heath, “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.
- [25] J. Mo, P. Schniter, N. González-Prelcic, and R. W. Heath, “Channel estimation in millimeter wave MIMO systems with one-bit quantization,” in *Proc. Asilomar Conf. Sig., Sys., and Comp.*, Nov. 2014, pp. 957–961.
- [26] J. Choi, J. Mo, and R. W. Heath, “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.

- [27] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, “Channel estimation and performance analysis of one-bit massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.
- [28] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, “Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs,” *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
- [29] S. Wang, Y. Li, and J. Wang, “Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.
- [30] J. Singh, O. Dabeer, and U. Madhow, “On the limits of communication with low-precision analog-to-digital conversion at the receiver,” *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [31] N. Liang and W. Zhang, “Mixed-ADC massive MIMO uplink in frequency-selective channels,” *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4652–4666, Nov. 2016.
- [32] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, “On channel estimation in OFDM systems,” in *Proc. Vehicular Tech. Conf.*, vol. 2, Jul. 1995, pp. 815–819 vol.2.
- [33] T. Hwang, C. Yang, G. Wu, S. Li, and G. Ye Li, “OFDM and its wireless applications: A survey,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp.

1673–1694, May 2009.

- [34] M. K. Ozdemir and H. Arslan, “Channel estimation for wireless OFDM systems,” *IEEE Commun. Surveys and Tutorials*, vol. 9, no. 2, pp. 18–48, 2007.
- [35] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, “Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling,” *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.
- [36] A. L. Swindlehurst, “Time delay and spatial signature estimation using known asynchronous signals,” *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 449–462, Feb. 1998.
- [37] I. C. Wong and B. L. Evans, “Sinusoidal modeling and adaptive channel prediction in mobile OFDM systems,” *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1601–1615, Apr. 2008.
- [38] M. R. Raghavendra, S. Bhashyam, and K. Giridhar, “Exploiting hopping pilots for parametric channel estimation in OFDM systems,” *IEEE Signal Process. Lett.*, vol. 12, no. 11, pp. 737–740, Nov. 2005.
- [39] F. Sanzi, S. Jelting, and J. Speidel, “A comparative study of iterative channel estimators for mobile OFDM systems,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 849–859, Sep. 2003.

- [40] J. Bonnet and G. Auer, “Optimized iterative channel estimation for OFDM,” in *Proc. IEEE Vehicular Techn. Conf.*, Sep. 2006, pp. 1–5.
- [41] Y. Lee, A. Ashikhmin, and J. Chen, “Impact of soft channel construction on iterative channel estimation and data decoding for multicarrier systems,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2762–2770, Jul. 2008.
- [42] J. Ylioinas and M. Juntti, “Iterative joint detection, decoding, and channel estimation in turbo-coded MIMO-OFDM,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1784–1796, May 2009.
- [43] J. Lee, G. T. Gil, and Y. H. Lee, “Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications,” *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [44] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, “Channel estimation for hybrid architecture-based wideband millimeter wave systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
- [45] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, “Subspace estimation and decomposition for large millimeter-wave MIMO systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

- [46] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, “Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.
- [47] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [48] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [49] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [50] 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.
- [51] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [52] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

- [53] E. J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [54] M. Rani, S. B. Dhok, and R. B. Deshmukh, “A systematic review of compressive sensing: Concepts, implementations and applications,” *IEEE Access*, vol. 6, pp. 4875–4894, 2018.
- [55] H. Huang, S. Misra, W. Tang, H. Barani, and H. Al-Azzawi, “Applications of compressed sensing in communications networks,” *arXiv:1305.3002*, 2013. [Online]. Available: <http://arxiv.org/abs/1305.3002>
- [56] M. Leinonen, M. Codreanu, and G. B. Giannakis, “Compressed sensing with applications in wireless networks,” *Foundations and Trends in Signal Processing*, vol. 13, no. 1-2, pp. 1–282, 2019.
- [57] J. Sung and B. L. Evans, “Versatile compressive mmWave hybrid beamformer codebook design framework,” in *Proc. IEEE Int. Conf. Computing, Networking and Commun.*, Feb. 2020.
- [58] 3GPP, *TS 38.104 V15.0.0 Base Station (BS) radio transmission and reception (Release 15)*, Dec. 2017.
- [59] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, “Wide-band 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.
- [60] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, “Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.
- [61] “IEEE standard for information technology–local and metropolitan area networks–specific requirements–Part 15.3: Amendment 2: Millimeter-wave-based alternative physical layer extension,” *IEEE Std 802.15.3c-2009 (Amendment to IEEE Std 802.15.3-2003)*, pp. 1–200, Oct. 2009.
- [62] J. Wang, “Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.
- [63] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [64] “IEEE standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements–Part 11: Wireless LAN medium access control

- (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band,” *IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012)*, pp. 1–628, Dec. 2012.
- [65] R. W. Heath, N. Gonzalez-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.
- [66] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, “MIMO precoding and combining solutions for millimeter-wave systems,” *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [67] 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.
- [68] V. Venkateswaran and A. J. van der Veen, “Analog beamforming in MIMO communications with phase shift networks and online channel estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.
- [69] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” *IEEE Access*, vol. 4, pp. 247–267, 2016.

- [70] 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.
- [71] A. Alkhateeb, G. Leus, and R. W. Heath, “Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?” in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Apr. 2015, pp. 2909–2913.
- [72] D. Ramasamy, S. Venkateswaran, and U. Madhow, “Compressive adaptation of large steerable arrays,” in *Proc. Information Theory and Applications Workshop*, Feb. 2012, pp. 234–239.
- [73] P. Schniter and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: The sparse way,” in *Proc. Asilomar Conf. Sign., Sys. and Comp.*, Nov. 2014, pp. 273–277.
- [74] S. Park and R. W. Heath, “Spatial channel covariance estimation for mmWave hybrid MIMO architecture,” in *Proc. Asilomar Conf. Sig., Sys. and Comp.*, Nov. 2016, pp. 1424–1428.
- [75] X. Gao, L. Dai, S. Han, C. L. I, and X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.

- [76] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, “Frequency-domain compressive channel estimation for frequency-selective hybrid mmWave MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [77] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, “Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.
- [78] J. Rodríguez-Fernández, N. González-Prelcic, and R. W. Heath, “Channel estimation in mixed hybrid-low resolution MIMO architectures for mmWave communication,” in *Proc. Asilomar Conf. Sig., Sys. and Comp.*, Nov. 2016, pp. 768–773.
- [79] J. Sung, J. Choi, and B. L. Evans, “Narrowband channel estimation for hybrid beamforming millimeter wave communication systems with one-bit quantization,” in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Apr. 2018.
- [80] L. Zhao, D. W. K. Ng, and J. Yuan, “Multi-user precoding and channel estimation for hybrid millimeter wave systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.
- [81] J. Huang and S. Signell, “Impact of channel estimation error on performance of adaptive MIMO systems,” in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Mar. 2008, pp. 2865–2868.

- [82] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2783–2794, Jun. 2011.
- [83] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
- [84] C. Lu, H. Li, and Z. Lin, "Optimized projections for compressed sensing via direct mutual coherence minimization," *arXiv:1508.03117*, 2015. [Online]. Available: <https://arxiv.org/abs/1508.03117>
- [85] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
- [86] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [87] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
- [88] P. Wang, M. Pajovic, P. V. Orlik, T. Koike-Akino, K. J. Kim, and J. Fang, "Sparse channel estimation in millimeter wave communications:

- Exploiting joint AoD-AoA angular spread,” in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.
- [89] J. Sung and B. L. Evans, “Hybrid beamformer codebook design and ordering for compressive mmWave channel estimation,” in *Proc. IEEE Int. Conf. Computing, Networking and Commun.*, Feb. 2020.
- [90] Y. Xiao, Y. Wang, and W. Xiang, “Dimension-deficient channel estimation of hybrid beamforming based on compressive sensing,” *IEEE Access*, vol. 7, pp. 13 791–13 798, Jan. 2019.
- [91] J. Sung and B. L. Evans, “Hybrid beamformer codebook design and ordering for compressive mmWave channel estimation,” Software Release, Jul. 26, 2019, <https://github.com/junmo-sung/hb-cb-design-ordering>.
- [92] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, “Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.
- [93] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, “Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8131–8146, Dec. 2017.
- [94] J. Choi, B. L. Evans, and A. Gatherer, “Resolution-adaptive hybrid MIMO architectures for millimeter wave communications,” *IEEE Trans.*

- Signal Process.*, vol. 65, no. 23, pp. 6201–6216, Dec. 2017.
- [95] —, “ADC bit allocation under a power constraint for mmWave massive MIMO communication receivers,” in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Mar. 2017, pp. 3494–3498.
- [96] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. González-Prelcic, “Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture,” in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Mar. 2017, pp. 6493–6497.
- [97] J. Mo, P. Schniter, and R. W. Heath, “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.
- [98] C. Rusu, R. Méndez-Rial, N. González-Prelcic, and R. W. Heath, “Adaptive one-bit compressive sensing with application to low-precision receivers at mmWave,” in *Proc. IEEE Global Commun. Conf.*, Dec. 2015, pp. 1–6.
- [99] J. García, J. Munir, K. Roth, and J. A. Nossek, “Channel estimation and data equalization in frequency-selective MIMO systems with one-bit quantization,” *arXiv:1609.04536*, 2016. [Online]. Available: <http://arxiv.org/abs/1609.04536>
- [100] U. S. Kamilov, A. Bourquard, A. Amini, and M. Unser, “One-bit measurements with adaptive thresholds,” *IEEE Signal Processing Letters*,

vol. 19, no. 10, pp. 607–610, Oct. 2012.

- [101] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in *Proc. IEEE International Symposium on Information Theory Proceedings*, Jul. 2011, pp. 2168–2172.
- [102] J. P. Vila and P. Schniter, “Expectation-maximization gaussian-mixture approximate message passing,” *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.
- [103] J. Sung and B. L. Evans, “Compressed-sensing based beam detection in 5G NR initial access,” in *Proc. IEEE Int. Work. Sig. Process. Adv. Wireless Commun.*, May 2020.
- [104] 3GPP, “5G; NR; Physical layer procedures for control – Rel. 15,” TS 38.213, 2018.
- [105] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, “A tutorial on beam management for 3GPP NR at mmWave frequencies,” *IEEE Commun. Surveys and Tutorials*, vol. 21, no. 1, pp. 173–196, First quarter 2019.
- [106] J. Choi, “Beam selection in mm-Wave multiuser MIMO systems using compressive sensing,” *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2936–2947, Aug. 2015.

- [107] H. Yan and D. Cabria, “Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems,” in *Proc. IEEE Global Conf. on Sig. and Inform. Process.*, Dec. 2016, pp. 620–624.
- [108] N. J. Myers and R. W. Heath, “Localized random sampling for robust compressive beam alignment,” in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, May 2019, pp. 4644–4648.
- [109] H. Yan and D. Cabric, “Compressive initial access and beamforming training for millimeter-wave cellular systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1151–1166, Sep. 2019.
- [110] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, “Initial beamforming for mmWave communications,” in *Proc. Asilomar Conf. Sign., Sys. and Comp.*, Nov. 2014, pp. 1926–1930.
- [111] L. Wei, Q. Li, and G. Wu, “Exhaustive, iterative and hybrid initial access techniques in mmWave communications,” in *Proc. IEEE Wireless Commun. and Networking Conf.*, Mar. 2017, pp. 1–6.
- [112] M. Giordani, M. Mezzavilla, and M. Zorzi, “Initial access in 5G mmWave cellular networks,” *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 40–47, Nov. 2016.
- [113] 3GPP, “5G; NR; Physical channels and modulation – Rel. 15,” TS 38.211, 2018.
- [114] —, “5G; NR; Overall description – Rel. 15,” TS 38.300, 2018.

- [115] Y. Cho and S. Hong, "One-bit successive-cancellation soft-output (OSS) detector for uplink MU-MIMO systems with one-bit ADCs," *IEEE Access*, vol. 7, pp. 27 172–27 182, 2019.

## Vita

Junmo Sung received BSEE degrees from Kyungpook National University (2008) and The University of Texas at Dallas (2010) through dual enrollment. He received his MSEE degree from The University of Texas at Dallas (2012). From 2012 to 2015, he worked for National Instruments in Seoul, South Korea, where he designed and implemented a real-time FD-MIMO receiver on an FPGA and where he conducted research on 5G cellular communication systems. He is currently a staff system engineer at Samsung Research America and also pursuing his PhD degree in Electrical and Computer Engineering at The University of Texas at Austin. His research interests include channel estimation and prediction in both sub- and above 6 GHz frequency bands considering MU-MIMO.

Permanent address: junmo.sung@utexas.edu

This dissertation was typeset with  $\text{\LaTeX}^\dagger$  by the author.

---

<sup>†</sup> $\text{\LaTeX}$  is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's  $\text{\TeX}$  Program.