Improved multipath channel estimation and data transmission through beamforming training using hierarchical codebook

by

Yi-Ming Sun
Bachelor of Engineering, University of Victoria, 2018

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Applied Science

in the Department of Electrical and Computer Engineering

Improved multipath channel estimation and data transmission through beamforming training using hierarchical codebook

by

Yi-Ming Sun
Bachelor of Engineering, University of Victoria, 2018

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical Engineering)

Dr. Wu-Sheng Lu, Departmental Member
(Department of Electrical Engineering)

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical Engineering)

Dr. Wu-Sheng Lu, Departmental Member
(Department of Electrical Engineering)

ABSTRACT

Multiple-input and multiple-output (MIMO) technology with antenna arrays is a vital solution to achieve the advertised features in the next generation wireless communication. Multiple antennas at the transmitter and receiver can achieve diversity as well as multiplexing gain during data transmission. In order to take advantage of the multiplexing gain of MIMO systems, two or more channel paths are required to send multiple signal streams simultaneously. Beamforming (BF) training using low resolution and high resolution array beams is already implemented in the IEEE 802.11ad standard, making hierarchical codebook design an attractive approach. In this thesis, our goal is to improve multi-path channel estimation and data transmission through BF training using hierarchical codebook design. Kaiser Window sector array design and restricted orthogonal projection are applied during the beam training phase. The pre-defined hybrid-implemented codewords selected after the BF training are used for data transmission directly. With these combined efforts, a 30% higher spectral efficiency compared to the reference design [1] is achieved.

Contents

List of Figures

Acronyms

A-BFT association beamforming training.

AGC automatic gain control.

ATI announcement transmission interval.

BF beamforming.

BHI beacon head interval.

BRP beacon refinement protocal.

BTI beacon transmission interval.

DTI data transmission interval.

MIMO multiple input multiple output.

RXSS receive sector sweep.

SLS sector-level sweep.

TXSS transmit sector sweep.

ACKNOWLEDGEMENTS

I would like to thank:

Supervisor Xiaodai Dong, for her mentoring, support, encouragement, and immense patience during my study.

Professor Wu-Sheng Lu, for his interests in my study and being my role model with his work ethic.

Fiona Matthews and Jean Haddow, for funding me with the James B. Haddow Scholarship.

Faculty of Graduate Studies, for awarding me with British Columbia Graduate Scholarship (BCGS) funded by the provincial government.

My father for being strict with me and reminding me the goal of doing a master study is being able to research and learn.

Boyfriend, Tim Hu, for encouraging me to finish and editing my interesting first draft.

Friend, Jia Xu, for always trusting in me without judgement, even when I doubt myself.

Last but not least, the cat, Rosie, for never failing to be an amazing cat.

DEDICATION

This thesis is dedicated to my mother, who values education and always supports me.

Chapter 1

Introduction

With fifth generation (5G) wireless communication currently deployed and sixth generation (6G) on the rise, it will not be uncommon to see the adaptation of 30 GHz and above millimeter wave (mmWave) frequencies. As such, opportunities shall arise from mmWave, including increased bandwidth from a broader spectrum; with such benefit, challenges will also present itself. The increased propagation loss, scattering, and atmosphere attenuation at high frequencies promotes the innovation of antenna design to compensate for the loss of wireless signal strength over distance.

The use of antenna arrays is one popular approach and a promising way to achieve desired communication range, while increasing capacity. The adaptation of antenna arrays at both transmitter and receiver is referred to as *multiple input and multiple output (MIMO)* technology. The promised benefits of MIMO beamforming (BF) in next generation wireless communication depends on how well antenna array inputs are designed for the wireless channel. Therefore, it is of vital importance to first understand mmWave channel in order to design an efficient MIMO system.

A mmWave MIMO channel distinguishes itself from a sub-6 GHz wireless channel through the exhibition of sparse multipath structure [8]. Older wireless communication systems at lower frequency ranges use omnidirectional or quasi-omnidirectional antennas, where signals coming from most directions can be received with no special angular information needed. Due to the nature of the wireless environment at the mmWave frequency, not all directions are equally suitable for data transmission. Thus, the distinctive challenge in mmWave wireless is to find suitable directions for signal transmission, and reception in a timely manner. The process of finding channel spatial information is part of the channel estimation process.

Traditionally, least square and minimum mean square error estimators are popular

approaches for channel estimation. However, due to the channel size and mmWave channel sparsity, new methods are needed to estimate non-zero elements of the channel. Compressed sensing is a mathematical tool to solve this type of problems [9], where a sparse signal can be estimated with a lot fewer samples required by Shannon-Nyquist theorem. Inspired by the compressed sensing approach, codebook-based channel estimation is proposed using a sparse set of pre-defined BF vectors to estimate channel information.

Codebook-based channel estimation has two main stages: BF training and data transmission. During the BF training stage, two devices try to establish communication link by transmitting known data sequences using predefined BF vectors, known as a codebook. At the end of the BF training stage, both the transmitter and receiver have established communication link and acquired channel information. The actual data is then transmitted and received. One consideration to be given is the ratio of time spent on BF training to data transmission. BF training may not be worthwhile for high mobility applications. For this reason, this research is limited to indoor stations with relatively low mobility.

In short, MIMO technology with antenna arrays is a vital solution to achieve the advertised features in the next generation wireless communication. The quality of MIMO communication link depends on how well the wireless channel is known at the transmitter and receiver. To gain channel knowledge, codebook-based channel sounding is proposed. In the next section, motivations for improving codebook-based multi-path channel estimation will be introduced.

1.1 Motivation

Multiple antennas at the transmitter and receiver can obtain diversity as well as multiplexing gain during data transmission. Diversity gain is achieved when the same symbol is sent over each transmit antenna assuming channel is known at the receiver for coherent combining. Multiplexing gain is achieved when more than one independent data streams N_s are transmitted and received over the channel simultaneously.

From real world measurements of an indoor 60-GHz channel, it is shown that line-of-sight (LOS) and the first-order reflected waves are main contributions to the multipath channel components [10]. Where strong reflectors exist (such as metal furniture or a blackboard), a multipath component demonstrates comparable power to the LOS [10]. Additionally, the NYUSIM channel model developed by Sun *et al.*

for outdoor urban environments displays 3 to 4 clusters in dense urban Non-Line-of-Sight environments [11]. The multipath nature of a mmWave channel means it can be decomposed in to R_H parallel independent channels. When $N_s \leq R_H$ data streams are transmitted and received simultaneously, the data rate increases N_s-fold. The fundamental design question in MIMO systems is to decide antennas be used for diversity gain, multiplexing gain, or both. Trade-offs between diversity and multiplexing gain are discussed in [12]. To take advantage of multiplexing gain, it is important to be able to discover more than one paths between two devices during the channel discovery phase. However, through literature research, it is found that less efforts were put into multi-path channel estimation and we aim to improve multi-path channel estimation.

In [13], multiple-user communication is studied on a network level where one AP establishes simultaneous communication links between multiple stations. In this thesis, the focus will not be on multiple links across multiple stations, but instead on a single pair of stations. When more than one good path exits between two stations, we want to take advantage of the multi-path environment, especially without line-of-sight (LOS). Our interest lies in finding two or more paths in the wireless channel for data transmission.

The IEEE 802.11ad standard specifies a two-phase BF training to determine and refine the receive and transmit antenna sectors between two stations. These being the sector level sweep (SLS), to select initial coarse sectors, and beam refinement phase (BRP), to fine-tune selected sectors. The adaptation of coarse and fine antenna array beams in 802.11ad encourages study in multi-resolution codebook-based channel estimation. Later in the introduction, details of BF training in the standard will be introduced to gain insight of the current implementation. The two-phase BF training proposed in the standard greatly reduces training time compared to an exhaustive search. However, since the standard only specifies a single-link communication between two stations, it does not take advantage of the multipath channel in mmWave.

Ramasamy *et al.* models the massive MIMO channel estimation problem as a compressed sensing problem and first introduces multi-path channel discovery using orthogonal projection [14]. By using orthogonal projection, the influence of the first found channel path is subtracted from the received signal. The second channel path is then found in the remaining received signal. Based on the orthogonal projection proposed in [14] and adopting multi-resolution codebook, Alkhateeb *et al.* proposes

multi-path channel estimation where orthogonal projection is performed on the received signal of each training stage [1]. By implementing the proposed multi-path algorithm in [1], we found that orthogonal projection at each stage of the hierarchical codebook training will introduce error. We propose a restricted orthogonal projection to mitigate such error and the detail will be discussed in Chapter 4.

In addition to BF training design, data transmission is another aspect we studied. For the case of large MIMO technology, with thousands of antenna elements potentially, a dedicated radio-frequency (RF) chain circuit for each array element is impractical due to its increased complexity, cost, and power consumption. Considering the hardware constraint, much work has been done on designing hybrid precoding algorithms that use reduced RF chains and analog phase shifters. Hybrid precoding algorithm works by assuming channel $\mathbf{H}$ is known at both the transmitter and receiver [15] [16]. These algorithms provide a great tool to adapt any BF vector into a hybrid architecture. However, hybrid precoding based on an estimated channel during data transmission in real time is computationally complex and difficult to implement into a low latency system. To solve this problem, pre-defined codewords can be used directly for data transmission followed by the the BF training process. The concern is whether codewords can fully take advantage of the multi-path channel. Later in Chapter 4, we will address this issue and demonstrate codewords are more efficient and accurate than hybrid precoding based on a constructed channel.

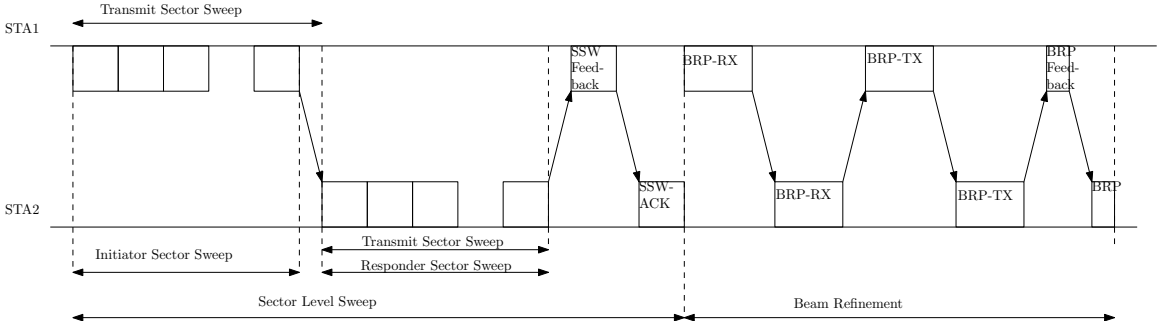


Figure 1.1: An example of beamforming training [2]

1.2 IEEE 802.11ad

In recent decades, mmWave standards have been developed to bring consumer products to the mmWave band, which include IEEE 802.11ad for wireless local area networks at 60GHz [2]. Much attention was also given to mobile network research into

the 28 GHz, 38 GHz, 60 GHz, 71 GHz and 81 GHz bands. The process to align antenna array beams between two devices is known as part of the beam management in the 3rd Generation Partnership Project New Radio or beamforming (BF) training in IEEE 802.11ad. IEEE 802.15.3c is an alternative standard for mmWave-based Wireless Personal Area Networks. However, it will be disregarded due to its inactivity.

Beamforming Training Concept in IEEE 802.11ad

During SLS, each of the two stations trains their transmit and receive antenna sector using coarse-grained antenna sectors. The station that transmits first is the initiator, the second is the respond. The SLS phase can have four sweep combinations: Transmit sector sweep (TXSS) at both initiator and responder, receiver sector sweep (RXSS) at both stations, initiator RXSS and responder TXSS, and initiator TXSS and responder RXSS. An example of TXSS at both stations is shown in left portion of Fig. 1.1.

As shown in Fig. 1.1, an initiator (the station that transmits first in a SLS) transmits multiple beacon frames, each containing different sectors to cover all possible directions, while the pairing station receives with a quasi-omnidirecitonal pattern. Each frame is marked with an identifier for the used antenna sector. The responder (the station that transmits second in a SLS) then transmits a series of sector sweep (SSW) frames while carrying the feedback of the best initiator sector in every frame as the initiator receives with a quasi-omnidirecitonal pattern. After all sectors are swept or given frames are used, the initiator then sends feedback of the optimum antenna configuration to the responder in a single SSW Feedback frame. The responder acknowledges SSW Feedback by sending an SSW-ACK frame. The last SSW-ACK frame is used in a following optional BRP to refine selected coarse-grained sectors, as shown in the right portion of Fig. 1.1.

Beamforming Protocol in IEEE 802.11ad

In lower frequency band, IEEE 802.11 controls the medium access by periodical beacon intervals initiated by a single beacon frame transmitted omnidirectionally from the access point (AP) or coordinating station. To replace the single beacon frame in lower frequency band, IEEE 802.11ad changes a beacon interval (BI) to include both beacon head interval (BHI) and Data Transmission Interval (DTI). The DTI

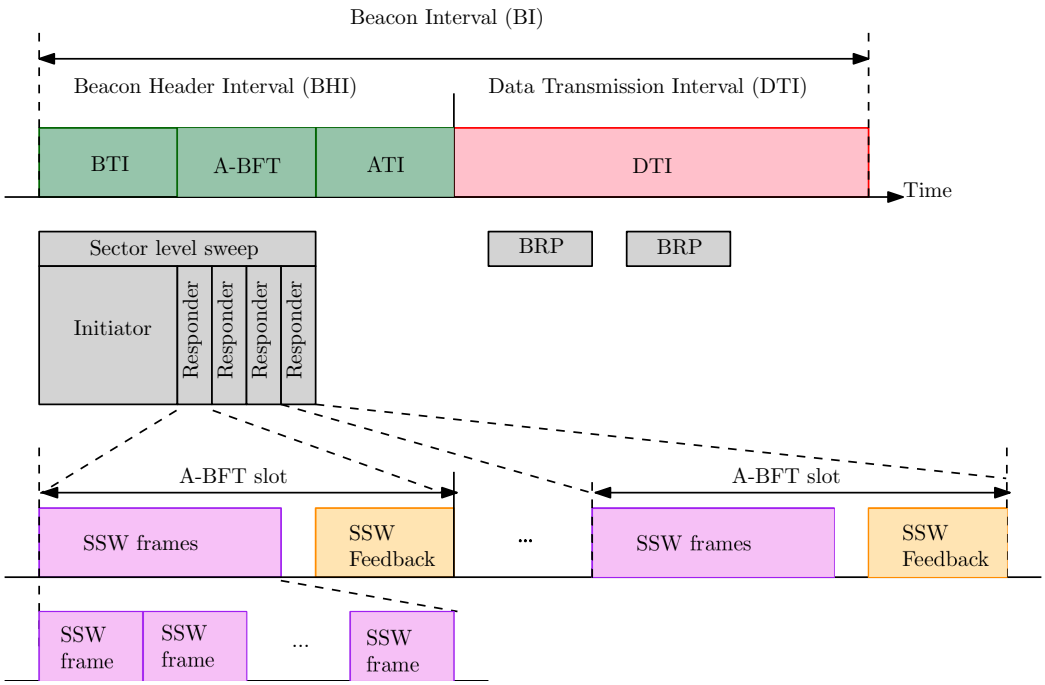


Figure 1.2: Association BF training proposed in IEEE 802.11ad [3].

can implement different types of medium access to provide different levels of quality of services.

BRP furthers refines sectors found in the SLS phase during the data transmission interval (DTI). Since the BRP happens after SLS, where a reliable frame exchange is secured, multiple antenna configurations can be tested during the same frame, thus reducing BF training overhead compared to SLS [17].

The challenges in BF training is that no coordination between stations can be relied upon prior to BF training. To overcome this issue, an AP can use beacon transmission interval (BTI) as an initiator sector sweep for all stations. The association beamforming training (A-BFT), as shown in Fig. 1.2, reserves channel time to allow multiple stations to respond to the initiator sector sweep without coordination. Note it is possible to split a beacon sweep into multiple BIs, with an increased association delay [3]. The following announcement transmission interval (ATI) serves associated nodes using trained beams with management data. More efficient modulation and coding schemes are used in directional transmitted frames in ATI.

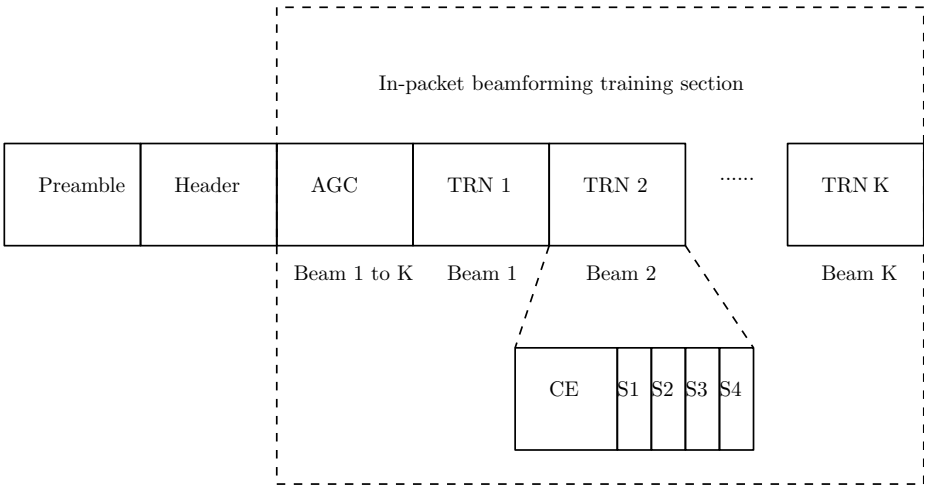


Figure 1.3: Packet structure for in-packet training in IEEE 802.11ad: automatic gain control (AGC) followed by four subfields to estimate phase and amplitude of channel taps [4].

Beam Training in the Data Transmission Period

Immediately after SLS, BRP during DTI typically follows as the initiator uses SSW ACK frame to request transmit or receive training. To improve BF training during the BRP in 802.11ad, Tsang *et al.* proposes to encode beams with Walsh code and

train multiple orthogonal antenna beams simultaneously [4]. By encoding orthogonal beams with Walsh code, the number of training packets required for exhaustive search is reduced while providing a flat power variation within a packet. Inspired by [4], the work of De Donno *et al.* reduces hierarchical codebook based BF training overhead using Walsh code-coded beams [18]. Implementation and trade-offs of Walsh code will be discussed in Chapter 2.

1.3 Contribution

Discovering multiple paths between two stations in a mmWave channel has been long neglected. A single communication link through BF only explores the diversity gain of a MIMO system. In order to take advantage of the multiplexing gain of a MIMO system, two or more signal streams are sent over the wireless channel simultaneously. Therefore, it is important to be able to use multiple channel paths for data transmission between two stations.

Drawing ideas from [19] and [1], the goal is to study multi-resolution codebook-based BF training with the focus on refining multi-path channel estimation. Instead of hybrid precoding followed by channel estimation, trained codewords from BF training are used directly for data transmission. The feasibility of such approach is verified theoretically via math derivation and simulation results. Lastly, some insights regarding BF training design trade-offs is provided.

The proposed multi-resolution codebook-based BF training makes the following contributions to the research community:

- improved sector level BF using Kaiser window weights with hybrid implementation adopting 2-bit phase shifters;
- increased multi-path BF training accuracy through restricted orthogonal projection;
- derived expression of spectral efficiency based on work from Telatar [20] using trained codewords directly for data transmission.
- demonstrated through spectral efficiency simulation results that trained beams can be used directly for data transmission, thus reducing time wasted during channel construction and hybrid precoding

- provided discussion on the influence of power allocation, codeword hybrid adaptation, codeword angular resolution, and restricted orthogonal projection on BF training accuracy.

Notation used through this thesis are: all vectors are column vectors expressed in bold small letters, such as $\mathbf{a}$ and $\mathbf{z}$. All matrices are expressed in bold capital letters, such as $\mathbf{A}$ and $\mathbf{H}$. Constant variables are in small letters, such as α and ω. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, $\mathbf{A}^+$ indicates transpose, Hermitian transpose, conjugate, inverse, pseudo-inverse of a matrix $\mathbf{A}$ respectively; $\square$, $\otimes$, and $\odot$ denotes Khatri-Rao product, Kronecker product and Hadamard product between two matrices; $vec(\mathbf{A})$ is the vectorization of a matrix $\mathbf{A}$ by stacking its column on top of each other.

1.4 Thesis Organization

Chapter 1 contains the introduction to the channel estimation problem in mmWave MIMO system and the motivation and contribution of this thesis. Chapter 1 also provides an overview of the thesis itself.

Chapter 2 describes in details the BF training system model using hierarchical codebook. Uniform linear arrays are used at both transmitters and receivers, although the proposed BF training algorithm will work on any array structure provided with codebook. A geometric-based channel model is used through out our research done. The mathematics derivation is presented to model the BF training problem as a compressed sensing problem. Coding the beams for multi-directional search is also discussed as its time-saving feature may not be as expected.

Chapter 3 gives the codebook design method, which has two main parts: codeword design using Kaiser window weights adapted to the hybrid structure using Alt-Min algorithm; multi-resolution codebook design. Power allocation methods are also introduced here.

Chapter 4 introduces codebook based BF training algorithm for single-path and multi-path channel estimation. Followed by successful channel sounding, found beams are directly used for data transmission. Data transmission spectral efficiency using codewords is derived and compared with hybrid precoding via channel construction and perfect channel estimation. Water-filling algorithm is applied during multi-stream data transmission for the first time and it is effects across received SNR is shown in this Chapter.

Chapter 5 evaluates and compares BF training accuracy using mean square error of channel constructed using with respect to the actual channel. Training complexity is presented here followed by discussion on effects of restricted orthogonal projection, codebook starting level, power allocation, hybrid adaptation and angular resolution on BF training performance. We show our BF training system design achieves higher spectral efficiency during data transmission compared to the reference design in [1].

Chapter 6 brings the conclusion of this dissertation. It also enumerates the avenues of future work for further development of the concept and its applications.

Chapter 2

Beam Training System Model

Sparsity in mmWave MIMO channels bring out a new spatial dimension to channel estimation. Our interest is in angle of departures (AoDs) and angle of arrivals (AoAs) of the best channel path in relationship to the transmitter and receiver. To take full advantage of the mmWave MIMO channel, we want to know channel state information at both base station and receiver prior to data transmission.

In order to verify the proposed beamforming training algorithm via simulations, a channel model and transceiver model representing real applications need to be chosen first. A geometric-based channel model is chosen based on uniform linear array response vector and hybrid transceivers are adopted throughout BF training and data transmission.

The beam training problem is formulated as a compressed sensing problem [1]. Lastly, training multiple beams at once can be achieved by coding beams with Walsh code. The limitations of this approach is discussed to help making design choices.

2.1 Hardware System Model

We consider a single user mmWave MIMO link composed of uniform linear arrays at both the transmitter and receiver. Depends on how many RF chains are implemented, a system can be fully digital, analog, or hybrid. The fully digital implementation of a transceiver has dedicated RF chain to each antenna. An RF chain allows passband signal at higher frequency to be processed in baseband [21]. The amplitude and phase of each antenna input is controlled independently via the dedicated RF chain. However, for mmWave frequencies over 30 GHz, implementing RF chains can be

costly and power hungry. In contrast, an analog transceiver consists of a single RF chain with certain number of phase shifters to send one data stream. The number of phase shifters can be equal to or less than the number of array elements depending on whether it's fully-connected or partially-connected.

Based on how the RF chain is connected to antenna array elements, a system can be fully-connected or partially-connected. A fully-connected transceiver has every RF chain connects to all elements via phase shifters; a partially-connected structure has every RF chain connects to certain elements, forming disjoint subarrays [22]. To strike a balance between performance and economic feasibility, a fully-connected hybrid system as shown in Fig. 2.1 is adopted, as shown in Fig. 2.1. N_s data streams are sent from the transmitter with N_t transmitting antennas and N_{RF}^t RF chains to the receiver with N_r receiving antennas and N_{RF}^r RF chains.

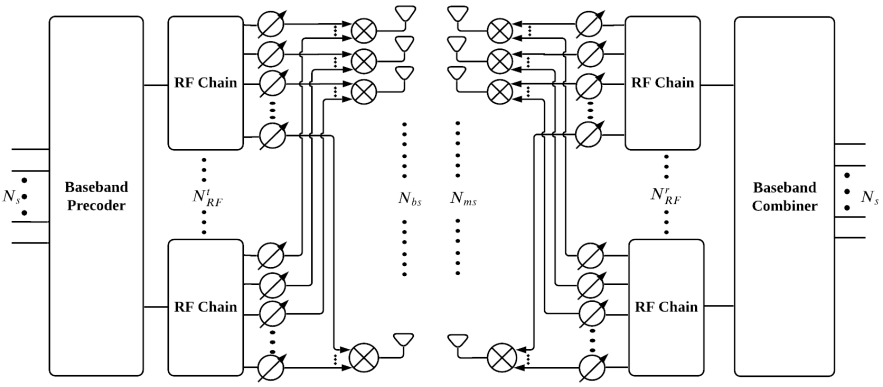


Figure 2.1: Simplified mmWave single user block diagram with fully-connected hybrid precoder and combiner

2.1.1 Uniform Linear Array Factor

A simple system model where uniform linear arrays are implemented on both the transmitter and receiver. A different antenna implementation will change the channel model and codebook design, but the beam-training algorithm can still apply. More sophisticated codebook design methods for distributed phased arrays are proposed in [23].

The operation principle of phased array is governed by radiation fields from elements in the array, interfere constructively for some directions (maximus) and destructively for other directions (nulls). Given a uniformly distributed linear array

located on the x-axis in a spherical coordinate system Fig. 2.2, the general form of the array factor at far field location (θ, ϕ) is

$$A(\theta, \phi) = \sum_{n=0}^{N-1} c_n e^{jnkd \sin\theta \cos\phi} \tag{2.1}$$

where $k = 2\pi/\lambda$ is the wavenumber, λ is the wavelength related to the frequency the antenna is designed for, $d = \lambda/2$ is the spacing between antenna elements, and $c_n = a_n \exp(j\omega_n)$ is the amplitude and phase excitation on each element.

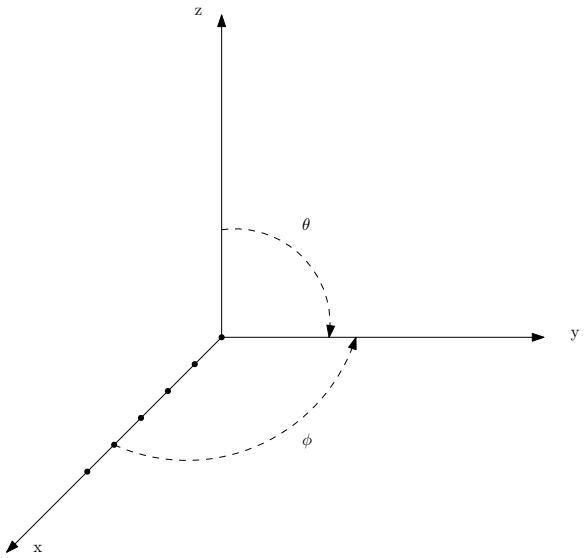


Figure 2.2: Spherical coordinate system adopted with a uniform linear array on the x-axis.

When the ULA(uniform linear array) with isotropic elements is placed along the x-axis, its broadside is at $\theta = \pi/2$ ($\sin(\pi/2) = 1$) and its array response is symmetric about the xz-plane. For arrays placed in user equipment, radiation behind the array is undesirable. Therefore, a ground screen is placed on the xz-plan to reflect radio wave and ϕ varies from 0 to π. Therefore, we are interested in array factor evaluated at $\theta = \frac{\pi}{2}$ and $\phi \in [0, \pi)$.

Now consider an ULA with a uniform excitation phase and amplitude, the array factor can be expressed as the superposition of all far field contributions from each element:

$$A(\phi) = 1 + e^{jkd\cos(\phi)} + e^{j2kd\cos(\phi)} + e^{j3kd\cos(\phi)} + e^{j4kd\cos(\phi)} + ...e^{j(N-1)kd\cos(\phi)}, \tag{2.2}$$

where $nd\cos(\phi)$ indicates the path difference from the n^{th} element to the origin in the far-field. When an excitation phase shift ω_n and amplitude a_n is applied to each element, the array factor can then be represented as:

$$\begin{aligned} A(\phi) = a_0 + a_1 e^{jkd\cos(\phi)+\omega_1} + a_2 e^{j2kd\cos(\phi)+\omega_2} + a_3 e^{j3kd\cos(\phi)+\omega_3} \\ + a_4 e^{j4kd\cos(\phi)+\omega_4} + ...a_{N-1} e^{j(N-1)kd\cos(\phi)+\omega_{N-1}} \end{aligned} \tag{2.3}$$

To form a main beam at direction ϕ_0, we need all the components to have their maximum at the desired ϕ_0, i.e. $nkd\cos(\phi_0) + \omega_n = 0$. We then have $\omega_n = -nkd\cos(\phi_0) = -n\omega_0$. By defining the spatial frequency as $\omega(\phi) = kd\cos\phi$, array factor in (2.3) can be expressed as

$$A(\omega) = \mathbf{a}^T \mathbf{z}(\omega), \tag{2.4}$$

where

$$\mathbf{z}(\omega) = (1, e^{j\omega}, e^{j2\omega}, ..., e^{j(N-1)\omega})^T \tag{2.5}$$

$$\mathbf{a} = (a_0, a_1 e^{j\omega_0}, a_2 e^{j2\omega_0}, ..., a_{(N-1)} e^{j(N-1)\omega_0})^T. \tag{2.6}$$

For example, to steer towards broadside ($\phi_0 = \frac{\pi}{2}$) corresponds to $\omega_0 = -kd\cos\frac{\pi}{2} = 0$, or $|A(\omega)|_{max} = |A(0)|$. Since $\mathbf{a}$ is the $N \times 1$ dimension weights applied to array elements, to design precoding or combining vectors is to design weights $\mathbf{a}$.

2.2 Channel Model

In order to evaluate the beam training algorithm, a good MIMO channel model that reflects the spatial diversity of the channel is needed. There are two main channel models: parametric channel and non-parametric channel [24]. A parametric channel model generates channel matrix based on key channel parameters (i.e., captured by ray-tracing technique), such as the statistical parametric channel model based on the measurements of mmWave outdoor cellular propagation at 28 GHz and 73 GHz is derived in [8]. The non-parametric method uses transmit and receive spatial correlation matrices to reproduce MIMO channel spatial correlation, such as the geometric-based channel model in [24]. Although parametric models are able to accurately reflect the actual wireless links in a MIMO system, non-parametric models are more suitable for theoretical analysis of correlated MIMO channels [24]. For this exact reason, [25] and [26] both adopts a geometric-based channel model with R_H channel paths. A

geometric-based channel model is therefore chosen for the design and performance analysis of beam training algorithms proposed in this study. The channel matrix $\mathbf{H}$ with ULA defined in the previous section at both the transmitter and receiver is given by

$$\mathbf{H} = \sum_{j=1}^{R_H} \alpha_j \mathbf{z}_r(\omega_{j,r}) \mathbf{z}_t^H(\omega_{j,t}), \tag{2.7}$$

where α_j is the complex fading gain of the j-th subpath modeled as a sequence of i.i.d complex Gaussian with equal variant on complex and real part, i.e., $\alpha_j \sim \mathcal{CN}(0, \sigma_\alpha^2)$. Spacial frequency $\omega_{j,t}$ and $\omega_{j,r}$, as defined in (2.4), contains the AoD and AoA respectively for the jth propagation path. $\mathbf{z}_t(\cdot)$ are the antenna array response vector at the receiver and transmitter respectively as defined in (2.5). By defining matrices $\mathbf{Z}_t$ and $\mathbf{Z}_r$ as

$$\mathbf{Z}_r = \frac{1}{\sqrt{N_r}} [\mathbf{z}_r(\omega_{1,r}), \mathbf{z}_r(\omega_{2,r}), ..., \mathbf{z}_r(\omega_{R_H,r})] \tag{2.8}$$

and

$$\mathbf{Z}_t = \frac{1}{\sqrt{N_t}} [\mathbf{z}_t(\omega_{1,t}), \mathbf{z}_t(\omega_{2,t}), ..., \mathbf{z}_t(\omega_{R_H,t})], \tag{2.9}$$

channel matrix in (2.7) can be expressed as

$$\mathbf{H} = \mathbf{Z}_r \mathbf{A} \mathbf{Z}_t^H, \tag{2.10}$$

where $\mathbf{A}$ is a diagonal matrix of size $R_H \times R_H$ with path gain

$$\boldsymbol{\alpha} = \sqrt{N_r N_t} [\alpha_1, \alpha_2, ...\alpha_{R_H}]^T \tag{2.11}$$

along its diagonal. Since transmitter and receiver can only transmit to quantized angles, the AoDs and AoAs in the channel matrix defined in (2.7) are chosen from a bin of discrete angles. Therefore, entries of $\mathbf{Z}_r$ and $\mathbf{Z}_t$ follows discrete uniform distribution in the angular bin.

2.3 Beam Training Problem Formulation

A hybrid precoder $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ $(N_t \times N_s)$ consists of digital base band precoder $\mathbf{F}_{BB}$ $(N_{RF}^t \times N_s)$ and RF phase shifter precoder $\mathbf{F}_{RF}$ $(N_t \times N_{RF}^t)$ with unit amplitude to control signal phase. A hybrid combiner $\mathbf{W} = \mathbf{W}_{RF}\mathbf{W}_{BB}$ $(N_r \times N_s)$ is implemented

in the same way. In order to keep equations more readable, we use $\mathbf{F}$ or one column of it $\mathbf{f}_i$ in following derivations. Though it is beneficial to always keep in mind the precoders and combiners used throughout our case study is the hybrid implementation that contains the digital baseband precoder and RF precoder unless otherwise noted. Details of the hybrid implementation are covered in Sec. 3.4.

When symbol s_i is transmitted at a specific time, the received signal y_i for a normal transmit beamforming vector $\mathbf{f}_i$ and a receive beamforming vector $\mathbf{w}_i$ is

$$y_i = \mathbf{w}_i^H \mathbf{H} \mathbf{f}_i s_i + \mathbf{w}_i^H \mathbf{n}, \tag{2.12}$$

where $\mathbf{n}$ is the complex Gaussian noise received on each receiver antenna element with zero mean and noise power σ_n^2. Assuming N_s data streams are transmitted over N_s directions across N_s time slots, given a precoder matrix at the transmitter $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_{N_s}]$ of size $N_t \times N_s$ and signal $\mathbf{S}$ of size $N_s \times N_s$ where signal s_i is along the diagonal, the transmitted symbols are

$$\mathbf{T} = \mathbf{F}\mathbf{S} = [\mathbf{f}_1 s_1, \mathbf{f}_2 s_2, ..., \mathbf{f}_{N_s} s_{N_s}], \tag{2.13}$$

where the symbol expected value satisfies $\mathcal{E}(s_i s_i^*) = \rho$ where ρ is the average power of each training symbol. Assuming the same symbol is used during the training phase, the transmitted signal is given by:

$$\mathbf{T} = \sqrt{\rho}\mathbf{F}. \tag{2.14}$$

The received signal at a receiver is then represented as

$$\mathbf{R} = \mathbf{H}\mathbf{T} + \mathbf{N}, \tag{2.15}$$

where $\mathbf{H}$ is the $N_r \times N_t$ channel matrix representing the wireless channel between the transmitter and receiver and $\mathbf{N}$ is the received noise vector of size $N_r \times N_s$ with independent complex values distributed as $\mathcal{CN}(0, \sigma_n^2)$. At the receiver, combiner $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, .., \mathbf{w}_{N_s}]$ of size $N_r \times N_s$ is used to process the received signal and results in the final received data

$$\mathbf{Y} = \mathbf{W}^H \mathbf{R} = \sqrt{\rho}\mathbf{W}^H \mathbf{H}\mathbf{F} + \mathbf{W}^H \mathbf{N}. \tag{2.16}$$

Knowing $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$ [6], the vectorized matrix $\mathbf{Y}$ then becomes

$$\mathbf{y}_v = vec(\mathbf{Y}) = \sqrt{\rho}(\mathbf{F}^T \otimes \mathbf{W}^H)vec(\mathbf{H}) + vec(\mathbf{W}^H\mathbf{N}) \tag{2.17}$$

Given channel model in (2.10) and knowing the property of Khatri-Rao product [27] the above equation can further be expressed as

$$\mathbf{y}_v = \sqrt{\rho}(\mathbf{F}^T \otimes \mathbf{W}^H)(\mathbf{Z}_t^* \square \mathbf{Z}_r)\boldsymbol{\alpha} + vec(\mathbf{W}^H\mathbf{N}), \tag{2.18}$$

where $\mathbf{Z}_t^* \square \mathbf{Z}_r$ is the Khatri-Rao product of size $N_t N_r \times R_H$. The result contains R_H columns where each column has the form $\mathbf{z}_t^*(\omega_{j,t}) \otimes \mathbf{z}_r(\omega_{j,r})$. In other words, each column of $\mathbf{Z}_t^* \square \mathbf{Z}_r$ is the Kronecker product of the transmitter and receiver array response vector associated with each path of the channel. Quantizing the interested azimuth region from 0 to π at both transmitter and receiver into D samples will lead to D array response vectors at each side, forming a dictionary $\mathbf{A}_D = \mathbf{A}_{D,t}^* \otimes \mathbf{A}_{D,r}$. Ignoring quantization error, the above equation can be replaced by

$$\mathbf{y}_v = \sqrt{\rho}(\mathbf{F}^T \otimes \mathbf{W}^H)\mathbf{A}_D\mathbf{x} + vec(\mathbf{W}^H\mathbf{N}), \tag{2.19}$$

where $\mathbf{A}_D$ is of size $N_t N_r \times D^2$ and $\mathbf{x}$ is a sparse vector contains values of $\boldsymbol{\alpha}$ at locations corresponding to channel AoAs and AoDs pairs. By defining $\mathbf{x} = \mathbf{x}_t \otimes \mathbf{x}_r$, the vectorized received signal $\mathbf{y}_v$ can be expressed as

$$\begin{aligned} \mathbf{y}_v &= \sqrt{\rho}(\mathbf{F}^T \otimes \mathbf{W}^H)(\mathbf{A}_{D,t}^* \otimes \mathbf{A}_{D,r})(\mathbf{x}_t \otimes \mathbf{x}_r) + vec(\mathbf{W}^H\mathbf{N}) \\ &= \sqrt{\rho}(\mathbf{F}^T \otimes \mathbf{W}^H)(\mathbf{A}_{D,t}^*\mathbf{x}_t \otimes \mathbf{A}_{D,r}\mathbf{x}_r) + vec(\mathbf{W}^H\mathbf{N}) \end{aligned} \tag{2.20}$$

where $\mathbf{A}_{D,t}$ $(N_t \times D)$ and $\mathbf{A}_{D,r}$ $(N_r \times D)$ represent dictionary at the transmitter and receiver respectively, and $\mathbf{x}_t$ and $\mathbf{x}_r$ are $D \times 1$ sparse selection vectors with R_H nonzero elements corresponding to channel AoAs and AoDs. The problem at hand is to recover the position and value of nonzero elements in $\mathbf{x}_t$ and $\mathbf{x}_r$ through beam training. Applying the Kronecker product property [27], the above equation can be written as

$$\mathbf{y}_v = \sqrt{\rho}(\mathbf{F}^T\mathbf{A}_{D,t}^*\mathbf{x}_t \otimes \mathbf{W}^H\mathbf{A}_{D,r}\mathbf{x}_r) + vec(\mathbf{W}^H\mathbf{N}). \tag{2.21}$$

By being able to express the received signal with quantized set of angles, application of orthogonal projection to remove the effect of the best path, and finding the second best path is possible. This will be discussed in detail in Chapter 4.

2.4 Coding the beams

Previously, the same training symbols are transmitted throughout the training phase and one symbol is transmitted at a time using one precoding vector. In order to train more than one beam at once, Walsh code can be used to code multiple beams and achieve simultaneous transmission across desired directions [4]. Assuming four beams are trained at once, four symbols are being sent over N_s time slots and the sent signal $\mathbf{S}_\perp$ $(4 \times N_s)$ is

$$\mathbf{S}_\perp = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & ... & \mathbf{s}_{N_s} \end{bmatrix}, \tag{2.22}$$

where each column is orthogonal to each other. These training sequences can be generated via Walsh code and are known to both the transmitter and receiver. The transmitted signal is then

$$\mathbf{T}_\perp = \mathbf{F}\mathbf{S}_\perp \tag{2.23}$$

The received signal after combining is

$$\mathbf{Y}_\perp = \mathbf{W}^H\mathbf{H}\mathbf{F}\mathbf{S}_\perp + \mathbf{W}^H\mathbf{N}, \tag{2.24}$$

where $\mathbf{Y}_\perp$ is $(4 \times N_s)$, $\mathbf{W}$ is $(N_r \times 4)$ combining matrix and $\mathbf{F}$ is $(N_t \times 4)$ precoding matrix and $\mathbf{N}$ is the received noise vector of size $N_r \times N_s$ as defined earlier. We know $\mathbf{I} = \frac{1}{N_s}\mathbf{S}_\perp\mathbf{S}_\perp^T$, where $\mathbf{I}$ is the identity matrix. To separate beams used to transmit and receive the Walsh code, the correlation between the received signal and the transmitted Walsh sequence is calculated using

$$\mathbf{Y} = \frac{1}{N_s}\mathbf{Y}_\perp\mathbf{S}_\perp^T \tag{2.25}$$

$$= \frac{1}{N_s}\mathbf{W}^H\mathbf{H}\mathbf{F}\mathbf{S}_\perp\mathbf{S}_\perp^T + \frac{1}{N_s}\mathbf{W}^H\mathbf{N}\mathbf{S}_\perp^T \tag{2.26}$$

$$= \mathbf{W}^H\mathbf{H}\mathbf{F}\mathbf{I} + \frac{1}{N_s}\mathbf{W}^H\mathbf{N}\mathbf{S}_\perp^T \tag{2.27}$$

When orthogonal beams are used, the beam coded with Walsh code will also enjoy reduced power fluctuation between the main lobe and sidelobes as discussed in [4]. The reduced power fluctuation observed from both receiving and transmitting beams will improve beamforming training accuracy.

Apart from benefits of using orthogonal beams, the limitation of encoding multiple

beams for transmission using Walsh code is that the number of packets required must be more than the number of beams. Additionally, the number of beams transmitted simultaneously must be to the power of two to generate Walsh code successfully. For instance, in order to distinguish four beams, at least four symbols are transmitted. This is due to the computation of the correlations to the received signal, we rely on following property of Walsh code: $\mathbf{I} = \frac{1}{N_s}\mathbf{S}_{\perp}\mathbf{S}_{\perp}^{T}$, where $\mathbf{I}$ is identity matrix and $\mathbf{S}_{\perp}$ is Walsh code matrix with orthogonal rows. However, if matrix $\mathbf{S}_{\perp}$ has more rows than columns (i.e., tall matrix) or not power of two rows or columns, the above equation does not hold. Additionally, in order to receive the signal, the four beams at the receiver must listen to four transmitted symbols one at a time. A total of 16 time slots are required to train 4 beams at the receiver and 4 beams at the transmitter. This observation renders the time-saving advantage of coding the beam to none compared to feedback based beam training. It is safe to say multiple beam training cannot reduce overhead.

In Chapter 2, we introduced the hybrid architecture adopted during the beam training phase, which implements ULA at both transmitters and receivers. A geometric-based channel model is used for its simplicity without losing generosity. More complicated and accurate channel models exist but the geometric-based model is widely adopted for beam training simulation. With the system setup, the beam training problem is formulated similar to compressed sensing problem where the goal is to find nonzero elements of $\mathbf{x}_t$ and $\mathbf{x}_r$ in (2.21). On top of conventional beam training approach, multi-beam training can be achieved via coding means with Walsh code. The benefits and limitations of multi-beam training are explained. In the next Chapter, we will focus on the codeword design for beam training and data transmission.

Chapter 3

Codebook Design

A practical codebook design largely depends on antenna design and hardware architecture implemented in the system. In 2019, a data-driven codebook design based on E-fields is proposed in [23] using distributed phased arrays; 3D codebook using uniform planar array is proposed in [28] for the unmanned aerial vehicle application. There are other researches coming out in this field adopting more sophisticated codebook design methods. Uniform linear arrays (ULA) with 2-dimensional beamforming are widely used when designing MIMO systems for its simplicity without losing generality, therefore we still adopt ULA codebook design. The proposed codebook architecture and beam training algorithm can be extended to any array configurations with hardware-specific codewords designed.

Two common codebook structures for beam training are: single-resolution and multi-resolution. In order to cover the entire range of angle of departures (AoDs) and angle of arrivals (AoAs) during the beam alignment stage, a single-resolution codebook contains certain number of pencil beam codewords with the same resolution, i.e., beamwidth. The number of codewords needed depends on the half power beam width (HPBW) that an antenna array can achieve, which in turn depends on the number of array elements in an antenna array. The approximate HPBW of a ULA is given in [29] and [5] as a function of number of array elements N, steering angle ϕ_0, and broadening factor f which is related to the choice of windowing weights $\mathbf{a}$ defined in (2.6):

$$\Delta\phi_{3dB} = \frac{0.886\lambda}{Nd\sin(\phi_0)}f, \tag{3.1}$$

where λ is the wavelength that the array is designed for and d is the array spacing. As shown in (3.1), the HPBW becomes smaller as the number of array elements

increases.

For multi-resolution codebook, the AoDs and AoAs at a base station and mobile station are first partitioned into coarse sectors. Instead of pencil beams focusing all energy at a certain angle, sector array beams are designed to radiate towards an angular domain. The next level beamforming vectors would then further divide the angular regions of interests. The angular resolution of sector beams increases as the codebook level increases. Since the codebook has multiple levels, it is also called hierarchical codebook. In this research, we will focus on the design of hierarchical codebook, thus codebook referred from this point on is hierarchical.

In this chapter, we first introduce the hierarchical codebook structure. Kaiser Window is then adopted to design the best sector beams. In order to reduce beam misalignment errors, power allocations at both codeword level and codebook level are proposed. Finally, to adopt the codebook into a hybrid hardware architecture, where a reduced number of RF chains are used due to cost and power constraints, the Alt-Min hybrid precoding method [7] with 2-bit phase shifters is adopted.

3.1 Hierarchical Codebook

The proposed hierarchical codebook contains M levels. Each level contains beamforming vectors of certain beamwidth to be used at corresponding stages in the proposed channel estimation algorithm. We define K as the number of beamforming vectors used in each stage of channel estimation. The interested angular range $[0, \pi)$ is first divided into K equal sectors in the first level codebook. Each sector in the first level codebook is further divided into K equal sectors for beam refinement. When $K = 2$, the path to find the best angular domain is a bi-section search. The desired AoAs and AoDs resolution π/D is another design parameter. Knowing the last level codebook contains D beamforming vectors, the number of codebook levels required is $M = \log_K D$.

In order to keep track of the trajectory taken in a hierarchical codebook, the concept of subset is introduced. We use B to denote the total number of subsets and b refers to the specific subset. Beamforming vectors in each codebook level are grouped into $B = K^{m-1}$ subsets, where m represents the codebook level. Each subset contains K beamforming vectors. For instance, in the example shown in Fig. 3.1, the third level codebook contains 2^2 subsets. In summary, the m_{th} codebook level contains K^{m-1} subsets and K^m beamforming vectors.

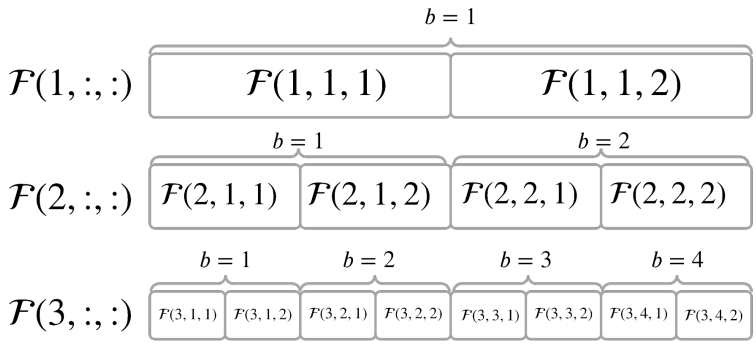


Figure 3.1: Example multi-resolution codebook design for angular resolution $D = 8$ and $K = 2$ beamforming vectors in each training stage.

Fig. 3.1 shows a sample hierarchical codebook at the base station with angular resolution $D = 8$, $K = 2$, and $M = \log_2 8 = 3$. A codeword at base station $\mathcal{F}(m, b, k)$ is a vector of size $N_{bs} \times 1$, where m refers to the codebook level, b is the subset, and k is the subdivision of the subset.

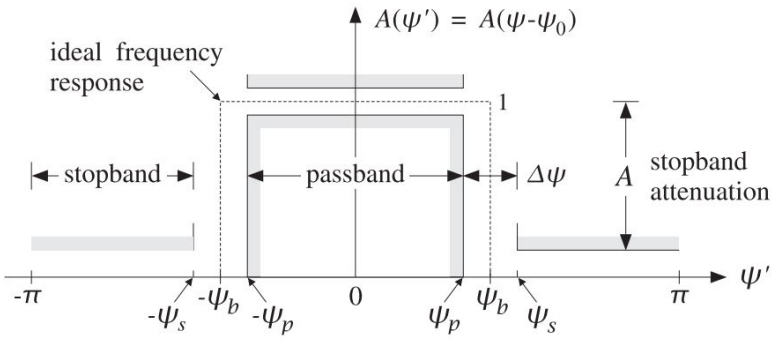


Figure 3.2: Specification of equivalent lowpass response [5]

3.2 Sector Beamforming

Notation: All vectors are column vectors expressed in bold small letters, such as **a** and **z**. All matrices are expressed in bold capital letters, such as **A** and **H**. Constant variables are in small letters, such as α and ω. $\mathbf{A}^T$ and $\mathbf{A}^H$ indicates transpose and Hermitian transpose of **A**.

Sector beamforming is concentration of the array's pattern into a desired angular sector. Similar to designing FIR digital filters in DSP, the sector beamforming is

based on inverse discrete-space Fourier transform of the array factor with windowing technique. The bandpass problem can be transferred to its equivalent lowpass representation as shown in Fig. 3.2.

3.2.1 Sector Beamforming Using Set of Quantized Angles

We take the design of transmitter beamforming vectors $\mathbf{F}$ as an example. The same principle can be applied to design receiver beamforming vectors. In [1], K beamforming vectors in level m subset b, expressed as $\mathbf{F}_b^m$ $(N_t \times K)$, are derived by the set of quantized angles in $\mathbf{A}_{D,t}$ $(N_t \times D)$ that $\mathbf{F}_b^m$ covers:

$$\mathbf{A}_{D,t}^H \mathbf{F}_b^m = {\mathbf{s_f}}^T \odot \mathbf{G}_b^m, \tag{3.2}$$

where $\mathbf{G}_b^m$ $(D \times K)$ are masks where desired angular ranges are set to one and otherwise zero; $\mathbf{s_f}$ is the scaling factor of size $(K \times 1)$ such that $\|\mathbf{F}(m,b)\|_F^2 = K$ and $\odot$ represents the element-wise multiplication where each column in $\mathbf{G}_b^m$ is multiplied by the corresponding columns of ${\mathbf{s_f}}^T$.

This method can adapt to any shapes of sector required given the mask $\mathbf{G}_b^m$. However, when the angular resolution D is too small compared to the number of antenna elements, i.e., the number of samples/columns in $\mathbf{A}_{D,t}$ is not large enough, the approximation can be very coarse and the beam shape becomes unacceptable. For the $N_t = 64$ and $N_{ms} = 32$ case, when angular resolution is coarse, such as $D = 81$ and the number of precoding vectors $K = 3$, the method presented in (3.2) fails to produce adequate beamforming vectors, resulting in inaccurate beam training. The angular resolution must be large enough compared to the number of array elements. For example, $N_t = 9$, $N_{ms} = 27$, $K = 3$, and $D = 81$ finds the desired angular frequency successfully.

Solving $\mathbf{Ax} = \mathbf{b}$

1. When $\mathbf{A}$ is square and invertible, $\mathbf{x} = \mathbf{A}^{-}\mathbf{b}$, where $\mathbf{A}^{-}$ is the inverse of A. In this case, $\mathbf{A}_{D,t}^H$ is not square and invertible.

2. When the number of rows in $\mathbf{A}$ is greater than the number of columns in $\mathbf{A}$, there are too many equations to expect a solution. If columns of $\mathbf{A}$ are independent and not too ill-conditioned, the least square solution $\hat{\mathbf{x}}$ can be found by solving

normal equations

$$\mathbf{A}^{\mathrm{H}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^{\mathrm{H}}\mathbf{b}, \tag{3.3}$$

where $\hat{\mathbf{x}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{b}$. Note $\mathbf{F}_b^m$ in (3.2) can be solved as such. The angular resolution D is the number of rows and the number of antennas N is the number of columns in $\mathbf{A}_{D,t}^H$. However, when D is not large enough, not enough sample points are given to approximate the solution and a large error can occur.

3. When matrix $\mathbf{A}$ is in bad condition, i.e., virtually dependent columns or the inverse of $\mathbf{A}$ has huge elements, Gram-Schmidt can be applied to orthogonalize the columns of $\mathbf{A}$. For more details of numeric linear algebra, please refer to [30].

3.2.2 Codeword Design Applying Kaiser Window

Instead of using the method in Sec. 3.2.1, we propose to use Kaiser window to design the sector beams as it is known to be the best and simplest windows that allows variable stop-band attenuation [5]. Kaiser window is one of the methods that use Fourier Series to design array weights. It is the simplest window that allow a variable choice for stopband attenuation. To represent the ideal filter response in Fig. 3.2 requires an infinite number of weights. Realistic coefficients will introduce unwanted ripples in the passband [5]. Passband ripples can be reduced using Kaiser window by specifying larger stopband attenuation at the expense of a wider transition region. The detailed derivation of Kaiser window can be found in Appendix A.

In each codebook level, there are total of B subsets. $B = 1$ for the first level codebook and number of codewords in the previous codebook level for subsequent levels. Each subset is then subdivided into K sectors and K beamforming vectors are needed to cover each subset. The goal at each stage of hierarchical channel discovery is to find the best subset to refine further in the next level. An angular domain of subdivision k at subset b and codebook level m has the lower angle ϕ_1 and upper angle ϕ_2 given by

$$\begin{aligned} \phi_{1,(m,b,k)} &= \frac{\pi}{K^m}(K(b-1)+k-1), \\ \phi_{2,(m,b,k)} &= \frac{\pi}{K^m}(K(b-1)+k), \end{aligned} \tag{3.4}$$

The window weights are applied to the pass-band array steering vector to achieve the desired shape. The resulting array weights are amplitude modified. The application of Kaiser window requires the fully digital beam steering. Here hybrid precoding

from [7] is applied and 2-bit phase shifters are incorporated to better approximate the real world application. With Kaiser window, we can achieve a relatively flat sector beam with a 20 dB stop band attenuation. This means the proposed codeword power variation between the main lobe and side lobe is 20 dB or more. This can mitigate miss-alignments caused by one beam's side lobe having higher directivity of another beam's main lobe.

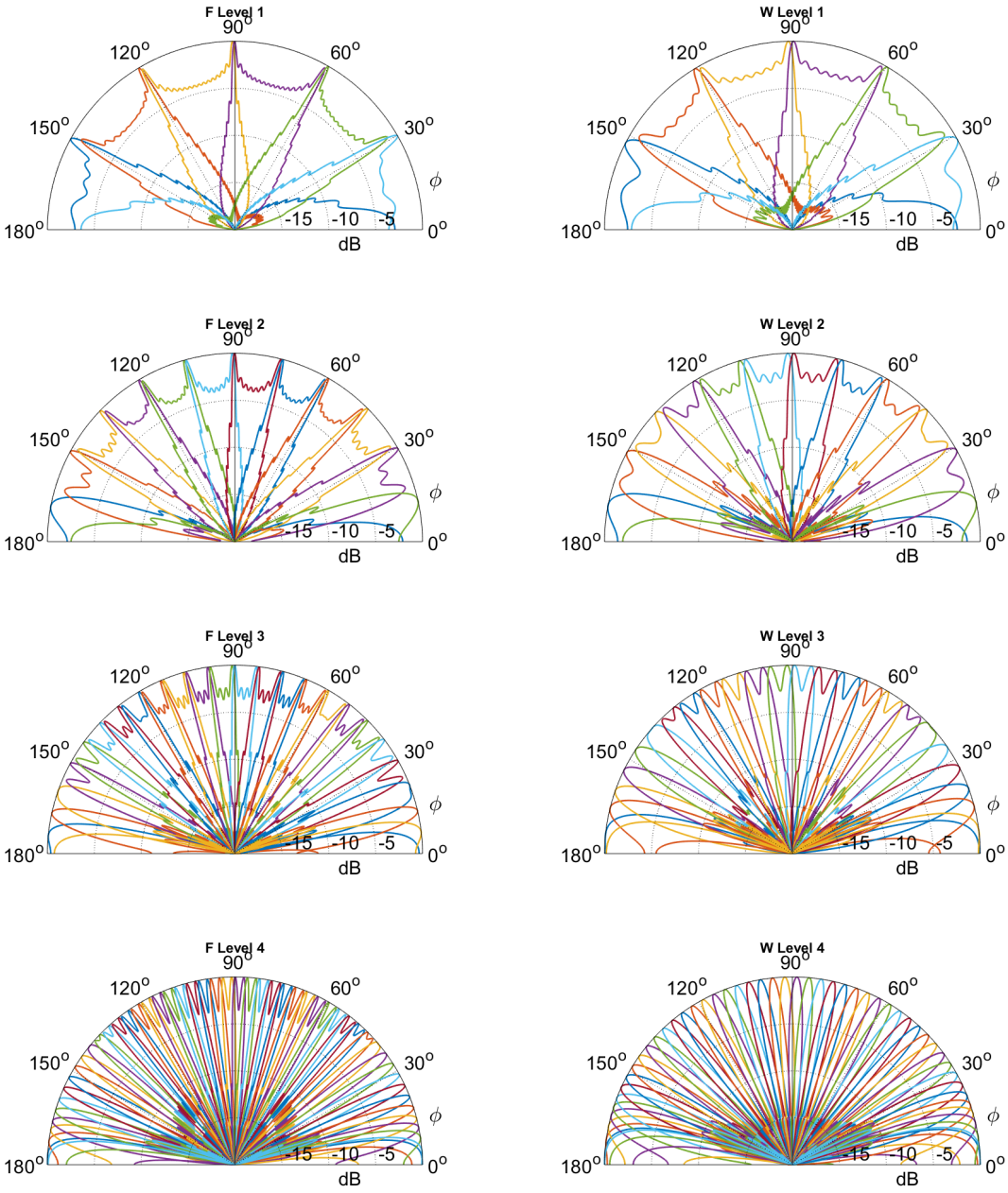


Figure 3.3: The resulting beam pattern of the beamforming vectors in the first four codebook levels. The codebook is designed with a set of quantized angles that a sector covers.

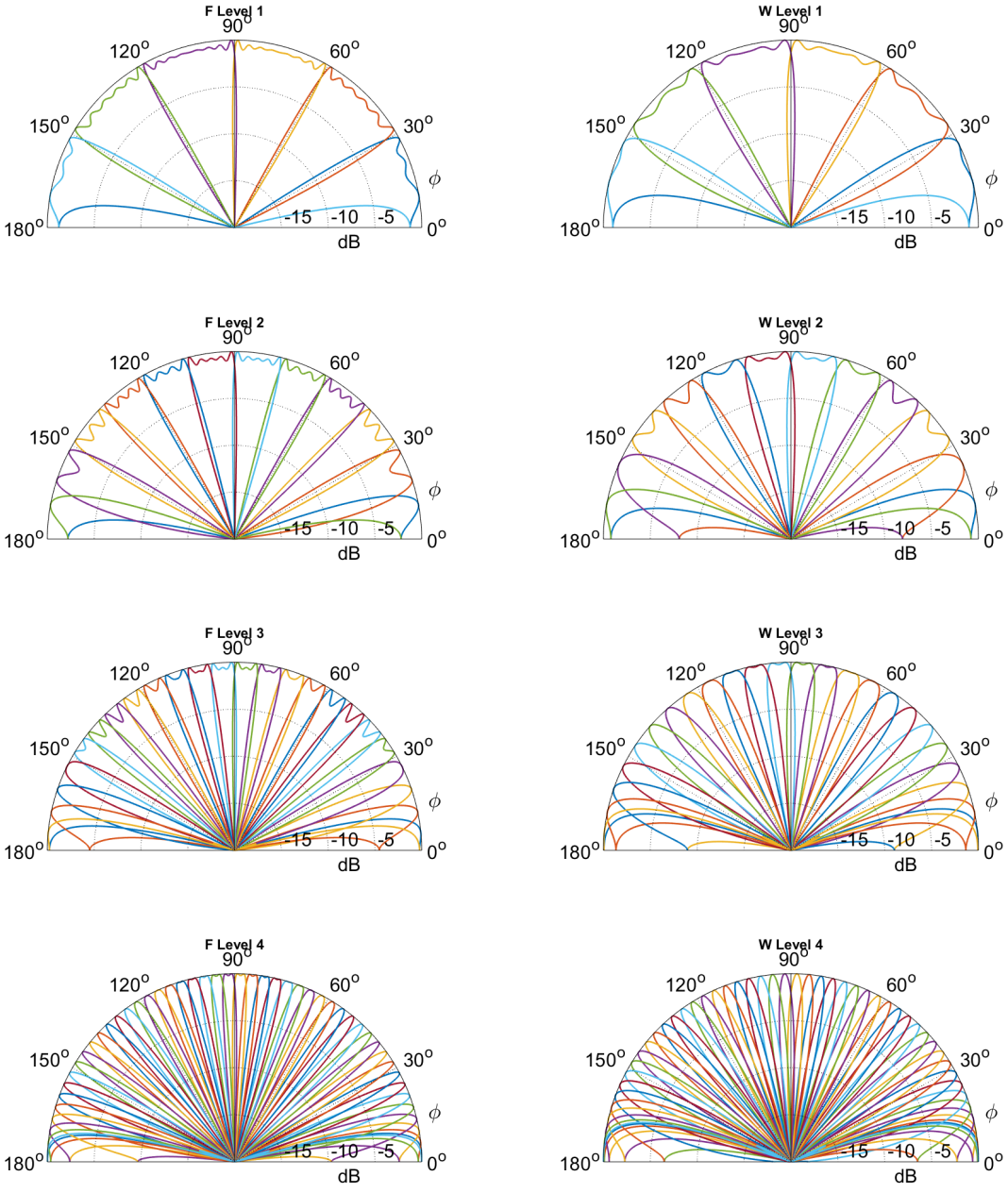


Figure 3.4: The resulting beam patterns of beamforming vectors in the first four levels of a codebook designed with Kaiser window

3.3 Power Allocation

Due to noise at the receiver, varying beam directivity at different resolutions (wide v.s. narrow) and angles (broadside v.s. end fire), beam misalignment or estimation error can occur. For these reasons, in order to improve beam training accuracy, there are two types of power allocation: across codewords within one level of a codebook and across different codebook levels.

At each codebook level, the optimal pair of beam training vectors is denoted by $\hat{i}_{opt}$ and defined as

$$\hat{i}_{opt} = \arg\max_{1 \le i \le K^2} |y[i]|^2,$$

where K is the number of beam training vectors used at each stage at the BS and MS, $y[i]$ corresponds to a received signal using a subset pair of ($\mathbf{W}$, $\mathbf{F}$) as defined in (2.12). From [31], assuming $\hat{i}_{opt}$ is uniformly distributed in $1, 2, ..., K^2$, the probability of misalignment is

$$P_{mis} = \frac{1}{K^2} \sum_{\hat{i}_{opt}=1}^{K^2} \text{Prob} \left(\bigcup_{i_{opt} \neq \hat{i}_{opt}}^{K^2} \{ |y[\hat{i}_{opt}]|^2 < |y[i_{opt}]|^2 \} \right). \tag{3.5}$$

Based on this general expression, an upper bound of beam training misalignment probability as a function of beamforming gains and codebook level power allocation is further derived in [1]. We define the forward to backward ratio $\beta_{m,d,k}$ as the ratio between the joint quantized AOA/AOD beamforming gain $G_{m,d}$ in direction d and the sidelobe gain of beamforming vector k in direction d.

For a desired resolution $\pi/2$, the average probability of misalignment $\overline{P}_{mis}$ is upper bounded by [1]

$$\overline{P}_{mis} \le \frac{K^2 - 1}{2D^2} \sum_{m=1}^{M} \sum_{d=1}^{D^2} \left(1 - \frac{1}{4} \frac{(1-X)Y}{\sqrt{1 + \frac{Y}{2}(1+X) + \frac{Y^2}{16}(1-X)^2}} \right) \tag{3.6}$$

where $X = \frac{1}{\overline{\beta}_{m,d}}$, $\overline{\beta}_{m,d} = \min_{\forall k=1,2,...,K^2, d \notin k} \beta_{m,d,k}$ is the worst forward to backward ratio in the direction d; and $Y = \mathbf{p}(m) G_{m,d} \overline{\gamma}$, where $\mathbf{p}(m)$ is the power allocation at codebook level m, $G_{m,d}$ is the actual beamforming gain in direction d, and $\overline{\gamma} = \frac{\text{Var}(\boldsymbol{\alpha})}{\rho \text{Var}(\mathbf{n})}$, is the average channel signal to noise ratio and Var$(\cdot)$ means random variable variance.

3.3.1 Codebook Power Allocation

The beamforming gain from both the precoder and combiner are $G_{m,d} = G_{m,d,t}G_{m,d,ms}$, where $G_{m,d,t}$ is the average codeword directivity in direction d at level m of a base station codebook. We define power at each stage of training $\mathbf{p}(m),\ m = 1, 2, 3, ..., M$, and M is the total number of levels in a codebook. When

$$\mathbf{p}(m) \geq \frac{\Gamma}{G_m}, \tag{3.7}$$

where

$$\Gamma = \frac{2}{\overline{\gamma}}(\frac{(K^2-1)M}{\delta} - 2), \tag{3.8}$$

$\overline{\gamma} = \frac{\mathrm{Var}(\boldsymbol{\alpha})}{\mathrm{Var}(\mathbf{n})}$ is the average channel signal to noise ratio and $\mathrm{Var}(\cdot)$ means random variable variance, and the beam training is guaranteed to be estimated with an average error probability $\overline{p} \leq \delta$ [1].

Based on (3.7), the total training power across all stages and subdivisions must satisfy $P_T \geq K^2\Gamma\sum_{m=1}^{M} 1/G_m$. Distributing the total training power P_T over all stages of channel discovery, we have

$$\mathbf{p}(m) = \frac{P_T}{K^2 G_m \sum_{m'=1}^{M} \frac{1}{G_{m'}}}, m = 1, 2, ..., M \tag{3.9}$$

3.3.2 Codeword Power Allocation

Codeword level power allocation guarantees an equal directivity across all codewords in the same level of a codebook. As shown in Fig. 3.5, codeword directivity varies within the same level of codebook and increases as the level increases. The variation in directivity among the same level will cause errors in beam training. In order to equalize codeword directivity, individual level power allocation is applied. The power allocated to each codeword x_i satisfies $\sum_{i=1}^{Q} x_i \leq P$, where Q is the total number of codewords in a level and P is the total power. For the example shown in Fig. 3.6, there are Q = 4 codewords within a level of a codebook. The directivity of each sector antenna array beams G_i and the total power $P = Q = 4$, x_i must satisfy the

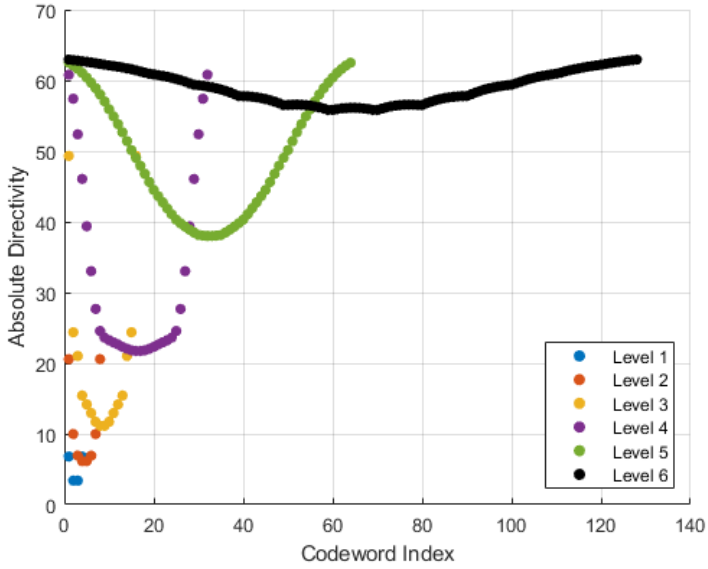
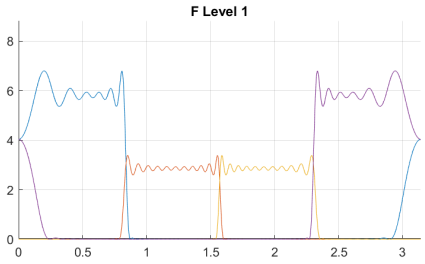


Figure 3.5: Absolute directivity of each codeword across all levels in a $L = 2$, $K = 2$, $D = 128$ codebook design before applying codeword power allocation

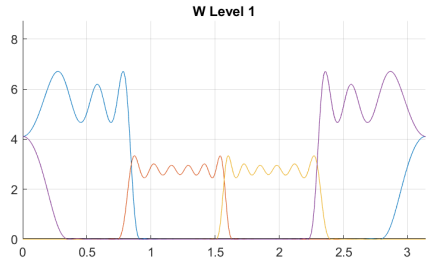
following conditions:

$$\begin{aligned} x_1 G_1 - x_2 G_2 &= 0 \\ x_2 G_2 - x_3 G_3 &= 0 \\ x_3 G_3 - x_4 G_4 &= 0 \\ x_1 + x_2 + x_3 + x_4 &= 4, \end{aligned} \tag{3.10}$$

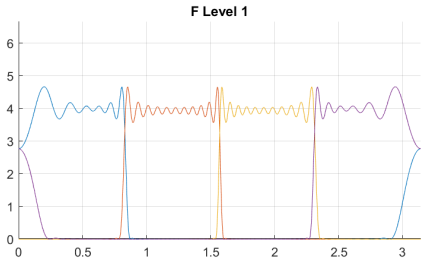
which expressed in the matrix form becomes $\mathbf{Ax} = \mathbf{b}$. Since $\mathbf{A}$ has independent columns and is invertible, $\mathbf{x}$ can be solved as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Fig. 3.6 compares absolute radiation patterns before and after codeword power allocation. It demonstrates the proposed method equalizes directivity within the same level of a codebook. The effect of power allocation among codewords has a bigger effect when applied to a codebook's lower levels.



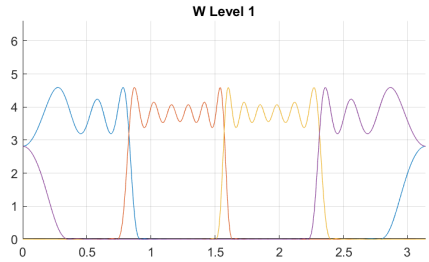
(a) Transmitter codewords before power allocation



(b) Receiver codewords before power allocation



(c) Transmitter codewords after power allocation



(d) Receiver codewords after power allocation

Figure 3.6: Absolute directivity before and after power allocation in level 1 of a codebook design

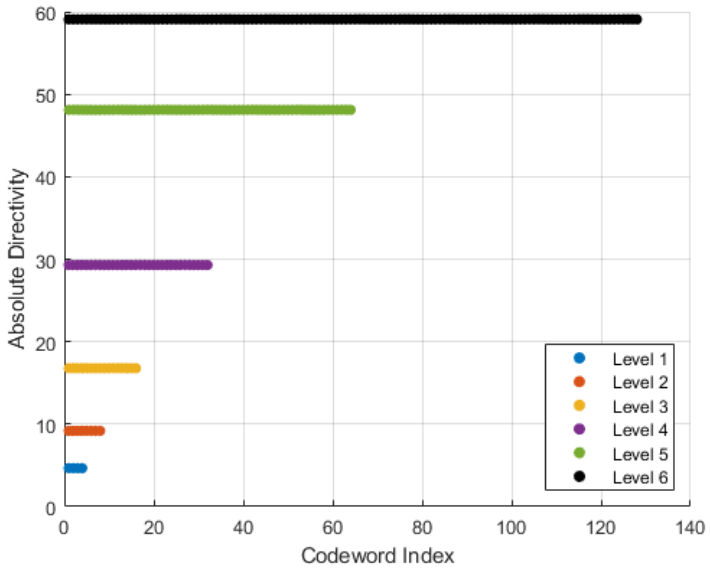


Figure 3.7: Absolute directivity of each codeword across all levels in a $L = 2$, $K = 2$, $D = 128$ codebook design after applying codeword power allocation

3.4 Hybrid Implementation of the Codebook

Manifold optimization based Alternating Minimization (MO-AltMin) algorithms proposed in [7] is adopted as the hybrid precoding algorithm in this study. To solve the hybrid precoding problem, orthogonal matching pursuit (OMP) algorithm -the most popular approach- is proposed in [16] where a RF precoder $\mathbf{F}_{RF}$ is found based on a dictionary of vectors. Instead of using a predefined dictionary like OMP, MO-AltMin proposed in [7] solves the hybrid precoding design problem under the condition that the magnitude of nonzero elements in $\mathbf{F}_{RF}$ is equal to one. By relaxing feasible RF precoding vectors, the performance of hybrid precoding is improved. The design of hybrid precoder and combiner can be separated as shown in [16]. The combiner can be designed with the same algorithm described below. The problem formulation from [7] is:

$$\underset{\mathbf{F}_{RF},\mathbf{F}_{BB}}{\text{minimize}} \quad \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}\|_F \tag{3.11}$$

$$\text{subject to} \quad \begin{cases} \mathbf{F}_{RF} \in \mathcal{A} \\ \|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_s, \end{cases} \tag{3.12}$$

where $\mathbf{F}_{opt}$ is the optimal fully digital precoder obtained from Kaiser window in our adaptation of MO-AltMin. $\mathbf{F}_{RF}$ $(N_t \times N_{RF}^t)$ and $\mathbf{F}_{BB}$ $(N_{RF}^t \times N_s)$ are analog and digital precoders to be designed. Dictionary $\mathcal{A}$ is the set of feasible analog precoders that has unit modulus, .i.e. $|(\mathbf{F}_{RF})_{i,j}| = 1$. The number of RF chains fall in the region where $N_s \leq N_{RF}^t \leq N_t$ and $N_s \leq N_{RF}^r \leq N_r$. The first step is to select a random $\mathbf{F}_{RF}$ and solve for $\mathbf{F}_{BB}$ as shown in the equation below

$$\underset{\mathbf{F}_{BB}}{\text{minimize}} \quad \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}\|_F. \tag{3.13}$$

The solution of (3.13) is given by

$$\mathbf{F}_{BB} = \mathbf{F}_{RF}^{+}\mathbf{F}_{opt}, \tag{3.14}$$

where $\mathbf{F}_{RF}^{+}$ is the pseudo-inverse of $\mathbf{F}_{RF}$. The key idea of alternating minimization is to solve for $\mathbf{F}_{BB}$ and $\mathbf{F}_{RF}$ alternatively while fixing the other. Thus, in the next step, $\mathbf{F}_{BB}$ is fixed in order to find $\mathbf{F}_{RF}$, which means to solve the following problem.

$$\underset{\mathbf{F}_{RF}}{\text{minimize}} \quad \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 \tag{3.15}$$

$$\text{subject to} \quad |(\mathbf{F}_{RF})_{i,j}| = 1, \forall i, j. \tag{3.16}$$

Since the unit module constraints in (3.16) are non-convex, a manifold optimization algorithm is proposed in [7] to find a near-optimal solution. Since the input of Manifold Optimization Based Hybrid Precoding (MO-AltMin) for a fully-connected structure take $\mathbf{F}_{opt}$ as the only input, we can adapt Kaiser window weights into a hybrid architecture using the proposed algorithm. The antenna array weights obtained applying Kaiser window are fully digital. In order to adopt the practical hybrid precoding architecture using the AltMin algorithm, the Kaiser window weights are used as the input optimal precoder $\mathbf{F}_{opt}$ and baseband precoder $\mathbf{F}_{BB}$ and RF precoder $\mathbf{F}_{RF}$ are output from the algorithm. In order to reduce implementation cost, 2-bit phase shifters are used. This means the available phases in $\mathbf{F}_{RF}$ are [1, -1, i, -i] and the approximation method is demonstrated in Fig. 3.8, where phases fall into one coloured sector are approximated to the corresponding 2-bit phase shifter value.

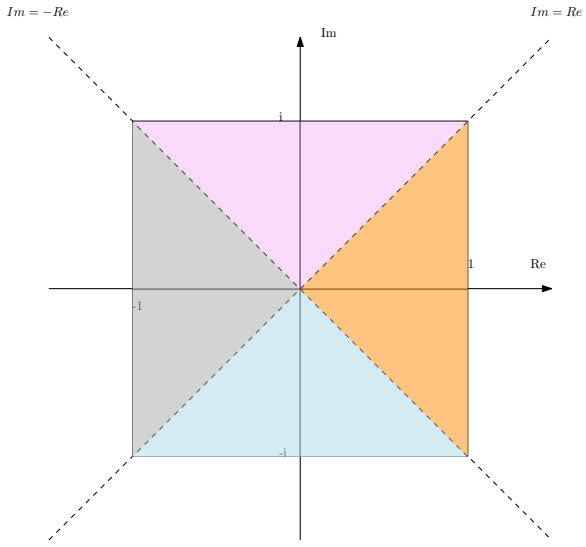


Figure 3.8: The 2-bit phase shifter design approach where points falling in the orange region will be approximated to 1 and etc.

3.5 Conclusion

In this chapter, a multi-resolution codebook design is proposed where angular domain is first divided into coarse sectors and later refined into desired resolution. Codewords for ULAs are designed using Kaiser Window sector array design algorithm. Note here codewords and codebooks are not to be confused with error correction codes in information theory. Kaiser Window generates sector beams with low attenuation across beamwidth compared to sector beamforming using quantized angles. A smooth sector beam will improve beam training accuracy. Power allocation methods are also introduced in order to optimize system performance and combat lower beam directivity at wider beamwidth. Lastly, for practical implementation, fully digital weights generated using Kaiser Window are adapted to the fully-connected hybrid architecture using MO-AltMin algorithm with 2-bit phase shifters. In the next chapter, codebook based beam training algorithm for single and multiple channel path discovery will be introduced. Followed by successful beam training, data transmission scheme using found codewords is proposed in Chapter 4 and compared with existing systems in Chapter 5.

Chapter 4

Codebook Based Beam Training Algorithm

To begin beam training, the transmitter first sends training packets with a set of predefined beams. Depending on the number of array elements and area to be searched, the number of beams required may differ. The IEEE 802.11ad standard recommends 32 beams for a 16-element ULA. Exhaustively searching through 32 packets is too long to be practical. Therefore, multi-resolution training scheme is proposed to reduce exhaustive search overhead [32] [2]. The multi-resolution search starts with sending Lower-Resolution (L-Re) beams by individual packets. L-Re beams cover a large sector space. Once the communication link is established and a coarse sector is selected, Higher-Resolution (H-Re) beams associated with the chosen L-Re beam are sent within one packet in the next step. While training time is greatly reduced, multi-resolution beam training tends to face problems finding multiple paths since two paths can fall in the same L-Re beams in earlier stages. One solution to improve multi-path BF training is to increase the number of L-Re beams at the cost of increasing PbP training overhead.

As mentioned, IEEE 802.11ad offers the ability for Packet-by-Packet (PbP) and in-packet beamforming (BF) training. A packet contains preamble, header, data, and optional BF training section for in-packet BF training. Since preamble and headers in a packet are used only for message identification, they do not contribute to BF training. Therefore in-packet BF training is proposed to reduce overheads in beam refinement. The in-packet training uses H-Re beams after sector level sweep is completed and an initial communication link is established between two devices. The

problem faced occurs with array beams applied during in-packet training sharing the same input power. This means in-packet training can suffer from power fluctuation from antenna directivity variation amongst different angles.

In [4], an improved in-packet BF training scheme is proposed which uses Walsh code to transmit multiple orthogonal beams at once while receiving using one beam at a time. The proposed method can provide a flat power variation within packets when orthogonal array beams are used. Overheads introduced by header and preamble sections in PbP training are also reduced at the same time.

In this chapter, we try to solve the beam channel estimation formulated in (2.21) of Chapter 2 based on the previous work in [8]. We first propose the algorithm to solve for a single path channel. Multi-path beam training algorithm is then proposed applying restricted orthogonal projection for the first time in the literature.

4.1 Single-path Algorithm

The first stage of hierarchical codebook based BF training is to find the best pair of BF vectors using codewords in the first level codebook. Training symbols are sent using one BF vector (also known as codeword) at the Tx for K_r times slots. The Rx switches between K_r beamforming vectors and measures the received signal. After K_r time slots, the Tx switches to the next beamforming vector in its first level codebook while the Rx measures received signal using K_r codewords, one in each time slot, in the first stage. After K_rK_t measurements, the strongest received signal represents the best BF vector pairs for further refinement is available at the Rx.

The Rx then sends the index of the chosen best Tx codeword back to the Tx. During the second stage BF training, the Rx measures K_rK_t times again, only codewords used at Tx and Rx are the subdivision of previous sectors.

At the last stage of the algorithm, the best beamforming vector pair at the Rx and Tx represents the best channel AoAs and AoDs respectively. In order to simplify the design, $K_r = K_t = K$ is assumed for now. Therefore, K^2 time slots are required in each stage for Rx to receive all pairs of BF vectors between Tx and Rx.

It is worth noting here that beam training is not reciprocal. As an example, using mobile downlink and uplink communication, transmitting and receiving beams for the downlink are not necessary the best receiving and transmitting beams for the uplink.

Algorithm 1: Single-Path Channel Estimation Algorithm

input : D, K, $\mathcal{F}$, $\mathcal{W}$

output: $\overline{\mathbf{f}}, \overline{\mathbf{w}}, \overline{\alpha}$

Initialization: $\mathbf{t_t}(1) = 1$, $\mathbf{t_r}(1) = 1$, $M = \log_K(D)$

for $m = m_s : M$ **do**

  $t_t = \mathbf{t_t}(m)$, $t_r = \mathbf{t_r}(m)$;

  $\mathbf{F} = \mathcal{F}(m, t_t, :)$;

  $\mathbf{W} = \mathcal{W}(m, t_r, :)$;

  /* measure received signal using each combination of channel sounding vectors at Tx and Rx */

  for *each column* k_t *in* $\mathbf{F}$ **do**

    for *each column* k_r *in* $\mathbf{W}$ **do**

      $\mathbf{Y}(k_r, k_t) = \sqrt{\mathbf{p}(m)}\mathbf{W}(:, k_r)^H \mathbf{H}\mathbf{F}(:, k_t) + \mathbf{n}$;

    end

  end

  $(\overline{k}_r, \overline{k}_t) = \underset{k_r, k_t = 1,2,\ldots K}{\operatorname{argmax}} [\mathbf{Y} \odot \mathbf{Y}^*]$;

  /* obtain the subset to refine in the next codebook level */

  $\mathbf{t_t}(m+1) = K * (t_t - 1) + \overline{k}_t$;

  $\mathbf{t_r}(m+1) = K * (t_r - 1) + \overline{k}_r$;

end

$\overline{\mathbf{f}} = \mathbf{F}(:, \overline{k}_t)$;

$\overline{\mathbf{w}} = \mathbf{W}(:, \overline{k}_r)$;

$\overline{\alpha} = \mathbf{Y}(\overline{k}_r, \overline{k}_t)/\sqrt{\mathbf{p}(m)}$;

4.2 Multi-path Algorithm

Most literature focuses on the code word design and finding the best path [18]. In order to take advantage of the multiplexing gain of MIMO channel during data transmission, two or more channel paths are required. In this research, we focus on improving multi-path channel estimation using hierarchical codebook design.

The multi-path codebook is built upon the single-path codebook as described in Chapter 3.1. The difference between a single-path and a multi-path codebook is in the starting level. Knowing K beamforming vectors are used at each subsequent codebook level, in order to find L independent channel paths, the angular range

$[0, \pi)$ is first divided into KL equal sectors in a multi-path codebook instead of K equal sectors in the single-path codebook. Each sector in the first level codebook is then divided into K sectors in the next level stage and so on. For example, the total number of beamforming vectors in the second level codebook is K^2L. Fig. 4.1 provides an example multi-path codebook.

In order to keep track of the trajectory taken to find the best channel path using a hierarchical codebook, the concept of subset is introduced. The codebook's first level contains one subset that spans the entire interested angular domain. Beamforming vectors in subsequent codebook levels are grouped into $K^{m-1}L$ subsets, where $m \geq 2$ represents codebook level. Each subset contains K beamforming vectors. The second level codebook contains KL subsets. Therefore, except for the starting codebook level, the m_{th} codebook level contains $K^{m-1}L$ subsets and K^mL beamforming vectors. Given the finest angular resolution as π/D, the total number of levels M in a codebook is then the integer rounding of $\log_K^{D/L}$. In other words, to find the codebook height M is to solve for M when $K^ML = D$.

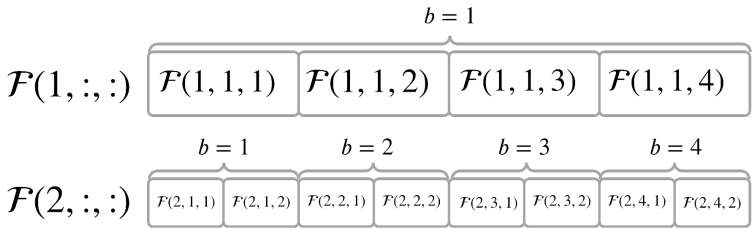


Figure 4.1: Example multi-resolution codebook design for $L = 2$ channel paths to be found, angular resolution of $D = 8$, and $K = 2$ beamforming vectors in each subsequent training stage.

The multi-path channel estimation algorithm first finds the best channel path following the procedure in Algorithm 1 with one difference at the initial stage. Instead of K^2 measurements, $(KL)^2$ measurements are required to search through all sector pairs in the first level multi-path codebook. The strongest received signal represents the best sector pair for further refinement by the Rx sending the chosen subset index back to the Tx. The subsequent BF training stage follows Algorithm 1 where the Rx has K^2 measurements based one the previously chosen subset. As the last stage of the BF training, the best beamforming vector pair represents the first dominant channel path.

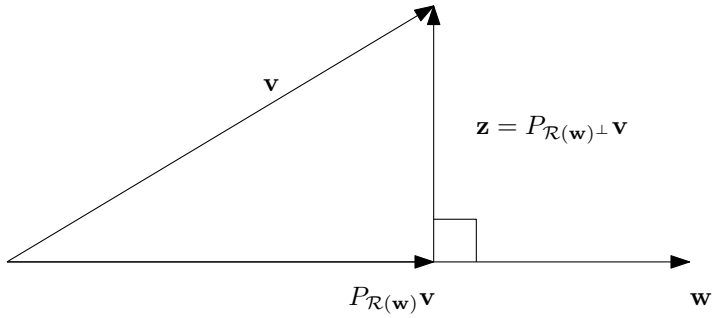


Figure 4.2: Orthogonal projection onto a vector [6]

Algorithm 2 takes the angular domain resolution D, the number of subdivisions K, the number of desired channel path L, Tx codebook $\mathcal{F}$, and Rx codebook $\mathcal{W}$ as inputs to find the best beamforming vectors reflecting AoDs and AoAs at Tx and Rx, i.e. $\overline{\mathbf{F}}$ and $\overline{\mathbf{W}}$ respectively; as well as channel path gain $\overline{\boldsymbol{\alpha}}$.

4.2.1 Restricted Orthogonal Projection

An example of orthogonal projection is illustrated in Fig. 4.2. Given $\mathbb{C}^n$ as the set of n-tuples of real numbers represented as column vectors, orthogonal projection $P_{\mathcal{R}(\mathbf{w})}$ of a vector $\mathbf{v} \in \mathbb{C}^n$ on to another nonzero vector $\mathbf{w} \in \mathbb{C}^n$ is

$$P_{\mathcal{R}(\mathbf{w})}\mathbf{y}_v = \mathbf{w}\mathbf{w}^+\mathbf{y}_v,$$

where $\mathbf{w}^+ = (\mathbf{w}^H\mathbf{w})^{-1}\mathbf{w}^H$ when $\mathbf{w} \neq 0$. The vector $\mathbf{z}$ orthogonal to $\mathbf{w}$ is given by

$$\mathbf{z} = \mathbf{v} - P_{\mathcal{R}(\mathbf{w})}\mathbf{v} = (\mathbf{I} - P_{\mathcal{R}(\mathbf{w})})\mathbf{v} = P_{\mathcal{R}(\mathbf{w})^\perp}\mathbf{v}.$$

The received signal are the combination of many channel paths. The influence of the best path $\mathbf{g}$ to the vectorized received signal $\mathbf{y}_v$ can be obtained based on (2.21) as

$$\mathbf{g} = \mathbf{F}^T\overline{\mathbf{f}}_1^* \otimes \mathbf{W}^H\overline{\mathbf{w}}_1, \tag{4.1}$$

where $\overline{\mathbf{f}}_1(N_t \times 1)$ and $\overline{\mathbf{w}}_1(N_r \times 1)$ are beamforming vectors representing previously found best channel AoD and AoA respectively, and $\mathbf{F}(N_t \times K)$ and $\mathbf{W}(N_r \times K)$ are beamforming vectors used in the current channel sounding stage.

In order to find the second best path, the influence of the best path $\mathbf{g}$ needs to be eliminated from the received signal. This process is operated on the vectorized

received signal $\mathbf{y}_v$ at the Rx with orthogonal decomposition with respect to $\mathbf{g}$.

Similar to the example shown in Fig. 4.2, the vectorized received signal $\mathbf{y}_v$ is equivalent to $\mathbf{v}$; The influence of first dominant path $\mathbf{g}$ is equivalent to $\mathbf{w}$; The remaining $\mathbf{y}_p$ to find the second best path is equivalent to $\mathbf{z}$. The received signal can be expressed as $\mathbf{y}_v = P_{\mathcal{R}(\mathbf{g})}\mathbf{y}_v + \mathbf{y}_p$, where the orthogonal projection of $\mathbf{y}_v$ onto $\mathbf{g}$ is

$$P_{\mathcal{R}(\mathbf{g})}\mathbf{y}_v = \mathbf{g}(\mathbf{g}^H\mathbf{g})^{-1}\mathbf{g}^H\mathbf{g}\mathbf{y}_v. \tag{4.2}$$

The second best path can then be found in $\mathbf{y}_p$ obtained via

$$\mathbf{y}_p = \mathbf{y}_v - P_{\mathcal{R}(\mathbf{g})}\mathbf{y}_v = \mathbf{y}_v - \mathbf{g}(\mathbf{g}^H\mathbf{g})^{-1}\mathbf{g}^H\mathbf{y}_v. \tag{4.3}$$

Previously, orthogonal projection of the received signal is performed during every stage of the beam training using multi-resolution codebook. During the simulation, we found that it is only necessary to perform orthogonal decomposition on the received signal in the first stage. Orthogonal decomposition with respect to $\mathbf{g}$ in subsequent stages will introduce errors since the received signal doesn't contain the best channel path. Therefore, a restricted orthogonal projection is proposed as shown in Algorithm 2.

One unavoidable problem of multi-path channel sounding using codebooks is that when multiple dominant channel paths lies within the same sector in the initial stage of the codebook. With the process of orthogonal projection, the sector containing the best path is eliminated together with the rest of the dominant channel paths. The solution to counter this effect is to have a finer sector resolution in the initial beam training stage but having to suffer from increased training overhead.

Algorithm 2: Multi-Path Channel Estimation Algorithm with Restricted Orthogonal Projection

input : D, K, L, $\mathcal{F}$, $\mathcal{W}$

output: $\overline{\mathbf{F}}, \overline{\mathbf{W}}, \overline{\boldsymbol{\alpha}}$

Initialization: $\mathbf{T_t}(:,1) = 1$, $\mathbf{T_r}(:,1) = 1$, $M = \log_K(D/L)$

for $l = 1 : L$ **do**

  for $m = m_s : M$ **do**

    $t_t = \mathbf{T_t}(l,m)$, $t_r = \mathbf{T_r}(l,m)$;

    $\mathbf{F} = \mathcal{F}(m, t_t, :)$;

    $\mathbf{W} = \mathcal{W}(m, t_r, :)$;

    /* measure received signal using each combination of channel sounding vectors at Tx and Rx */

    for *each column* k_t *in* $\mathbf{F}$ **do**

      for *each column* k_r *in* $\mathbf{W}$ **do**

        $\mathbf{Y}(k_r, k_t) = \sqrt{\mathbf{p}(m)}\mathbf{W}(:,k_r)^H \mathbf{H}\mathbf{F}(:,k_t) + \mathbf{n}$;

      end

    end

    if $m = m_s$ **then** // restrict OP to the starting codebook level

      for $p = 1 : l-1$ **do** // OP starts from the second channel path

        $\mathbf{g} = \mathbf{F}^T\overline{\mathbf{F}}^*(:,p) \otimes \mathbf{W}^H\overline{\mathbf{W}}(:,p)$;

        $\mathbf{y_v} = vec(\mathbf{Y})$;

        $\mathbf{y_v} = \mathbf{y_v} - \mathbf{g}(\mathbf{g^H g})^{-1}\mathbf{g^H y_v}$;

        return $\mathbf{y_v}$ into $\mathbf{Y}$;

      end

    end

    $(m_r^*, m_t^*) = \underset{m_t, m_r = 1,2,\ldots K}{\operatorname{argmax}} [\mathbf{Y} \odot \mathbf{Y}^*]$;

    /* obtain the subset to refine in the next codebook level */

    $\mathbf{T_t}(l, m+1) = K * (t_t - 1) + m_t^*$;

    $\mathbf{T_r}(l, m+1) = K * (t_r - 1) + m_r^*$;

  end

  $\overline{\mathbf{W}}(:,l) = \mathbf{W}(:, m_r^*)$;

  $\overline{\mathbf{F}}(:,l) = \mathbf{F}(:, m_t^*)$;

  $\overline{\boldsymbol{\alpha}}(l) = \mathbf{Y}(m_r^*, m_t^*)/\sqrt{\mathbf{p}(m)}$;

end

4.3 Data Transmission Design

The previous two sections cover beam training algorithms to find the desired channel paths' directions and gains at both the transmitter and receiver. *How will this information be used to design precoding and combining vectors?* For a MIMO system where both the transmitter and receiver are equipped with antenna arrays, multiplexing gain can be achieved. Such multiplexing gain is feasible as MIMO channel can be decomposed in to a number of parallel independent channels. When independent data streams are transmitted over independent channels, the gain in data rate is called multiplexing gain. Multiplexing gain is obtained by transmitter precoding and receiver combining.

As defined in Chapter 2, a MIMO channel is described by $N_r \times N_t$ channel gain matrix $\mathbf{H}$. Any matrix $\mathbf{H}$ has singular value decomposition (SVD) as

$$\text{svd}(\mathbf{H}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{4.4}$$

where $\mathbf{U}$ $(N_r \times N_r)$ and $\mathbf{V}$ $(N_t \times N_t)$ are unitary and where $\mathbf{\Sigma}$ is an $N_r \times N_t$ diagonal matrix of channel's singular values σ_i where R_H of these singular values are nonzero.

4.3.1 Spectral Efficiency Using Channel Construction

In [1], a hybrid precoding design is proposed by constructing the estimated channel matrix at the transmitter using the equation below

$$\overline{\mathbf{H}} = \bar{\mathbf{W}}\bar{\mathbf{\Sigma}}\bar{\mathbf{F}}^H, \tag{4.5}$$

where $\overline{\mathbf{W}}$ and $\overline{\mathbf{F}}$ are matrices containing beam training results from Algorithm 1 or 2, and $\overline{\mathbf{\Sigma}}$ is diagonal matrix with channel path gain $\overline{\boldsymbol{\alpha}}$ obtained from Algorithm 1 or 2 along its diagonal. The optimal beamforming vectors are then obtained from a subset of channel singular vectors

$$(\mathbf{U}_{N_s}, \mathbf{V}_{N_s}) = \text{svds}[\bar{\mathbf{H}}, N_s] \tag{4.6}$$

$$\mathbf{W}_{DATA}^{opt} = \mathbf{U}_{N_s} \tag{4.7}$$

$$\mathbf{F}_{DATA}^{opt} = \mathbf{V}_{N_s}, \tag{4.8}$$

where svds$[\bar{\mathbf{H}}, N_s]$ returns N_s largest left and right singular vectors. Applying spatially sparse precoding algorithm in [16], the hybrid precoders and combiners for data transmission $\mathbf{F}_{DATA}$ and $\mathbf{W}_{DATA}$ are then obtained using $\mathbf{F}^{opt}_{DATA}$ and $\mathbf{W}^{opt}_{DATA}$. The same procedure is also applied in [18], where estimated channel information obtained from beam training is used to design hybrid beam patterns for data transmission via channel construction at the transmitter and receiver. The received signal after combining is then

$$\hat{\mathbf{y}} = \mathbf{W}^H_{DATA}\mathbf{H}\mathbf{F}_{DATA}\mathbf{s} + \mathbf{W}^H_{DATA}\mathbf{n}, \tag{4.9}$$

where for clarity of the expression, $\mathbf{W}_{DATA}$ and $\mathbf{F}_{DATA}$ are hybrid beamforming vectors composed of

$$\mathbf{F}_{DATA} = \mathbf{F}^{RF}_{DATA}\mathbf{F}^{BB}_{DATA} \tag{4.10}$$

$$\mathbf{W}_{DATA} = \mathbf{W}^{RF}_{DATA}\mathbf{W}^{BB}_{DATA}. \tag{4.11}$$

With this expression of the received signal, the spectra efficiency during downlink data transmission expression is

$$\hat{R} = \log\det\left[\mathbf{I}_{N_s} + \frac{P}{N_s}\mathbf{R}^{+}_{n}\mathbf{W}^H_{DATA}\mathbf{H}\mathbf{F}_{DATA}\mathbf{F}^H_{DATA}\mathbf{H}^H\mathbf{W}_{DATA}\right], \tag{4.12}$$

where$\mathbf{R}_n = \sigma^2_n\mathbf{W}^H_{DATA}\mathbf{W}_{DATA}$ is the noise covariance matrix after combining [1].

We find this process of redesigning $\mathbf{F}_{DATA}$ and $\mathbf{W}_{DATA}$ based on estimated channel overly complicated. The excessive step of constructing estimated channel and hybrid precoding may adds significant overhead in data transmission yet low latency is required. Thus we propose to use trained codewords directly for data transmission. Later in Fig. 4.4, this direct approach is shown to have superior achievable spectral efficiency compared to channel construction. The reason is hybrid precoding has errors from solving the optimization problem, which does not guarantee 100% success.

4.3.2 Spectral Efficiency Using Codewords

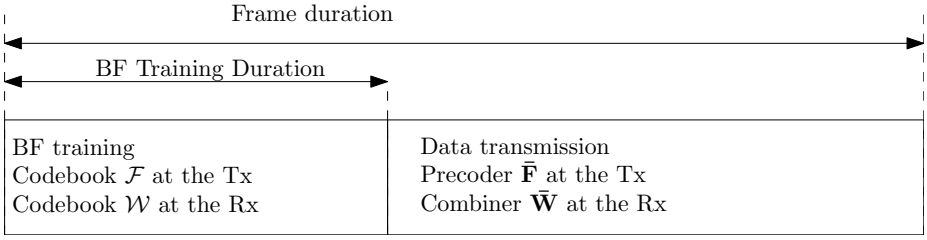


Figure 4.3: Frame structure including both the beam training phase and data transmission phase. The data transmission precoder and combiners are final codewords from beam training algorithms proposed in the previous section

The hardware block diagram shown in Fig. 2.1 is the single-user mmWave system adopted during the data transmission. Training results ($\overline{\mathbf{F}}$ and $\overline{\mathbf{W}}$) from Algorithm 1 or 2 are used directly as the precoding and combining vectors during data transmission as shown in Fig. 4.3. All hybrid transmitting and receiving vectors are implemented using the hybrid precoding algorithm as described in Sec. 3.4. The transmitter $\overline{\mathbf{F}}$ ($N_t \times N_s$) and receiver $\overline{\mathbf{W}}$ ($N_r \times N_s$) are designed as

$$\overline{\mathbf{F}} = \overline{\mathbf{F}}_{RF}\overline{\mathbf{F}}_{BB} \tag{4.13}$$

$$\overline{\mathbf{W}} = \overline{\mathbf{W}}_{RF}\overline{\mathbf{W}}_{BB}, \tag{4.14}$$

where $\overline{\mathbf{F}}_{RF}$ ($N_t \times N_{RF}^t$) is the RF precoder, $\overline{\mathbf{F}}_{BB}$ ($N_{RF}^t \times N_s$) is the baseband precoder, $\overline{\mathbf{W}}_{RF}$ ($N_r \times N_{RF}^r$) is the RF combiner, and $\overline{\mathbf{W}}_{BB}$ ($N_{RF}^r \times N_s$) is the baseband combiner. The number of RF chains satisfy $N_s \leq N_{RF}^t \leq N_t$ and $N_s \leq N_{RF}^r \leq N_r$. The symbol to be transmitted is $N_s \times 1$ vector $\mathbf{s} = [s_1, s_2, ..., s_{N_s}]^T$ during a specific time slot. Symbols are transmitted spontaneously over L directions where $L = N_s$. The transmitted signal is then

$$\mathbf{x} = \overline{\mathbf{F}}_{RF}\overline{\mathbf{F}}_{BB}\mathbf{s} \tag{4.15}$$

The received signal after combining is then

$$\overline{\mathbf{y}} = \overline{\mathbf{W}}^H\mathbf{H}\overline{\mathbf{F}}\mathbf{s} + \overline{\mathbf{W}}^H\mathbf{n}, \tag{4.16}$$

where each column in $\overline{\mathbf{F}}$ ($N_t \times L$) and $\overline{\mathbf{W}}$ ($N_r \times L$) is normalized so that $\|\overline{\mathbf{F}}(:,i)\| = \|\overline{\mathbf{W}}(:,i)\| = 1$. The noise vector $\mathbf{n}$ ($N_r \times 1$) is zero-mean unit variant complex Gaussian noise with independent, equal variant real and imaginary parts, i.e. the covariance

matrix of $\mathcal{E}[\mathbf{n}\mathbf{n}^H] = \sigma_n^2\mathbf{I}_{N_r}$. Since $\overline{\mathbf{W}}$ does not change the noise distribution, the noise after combining $\overline{\mathbf{n}} = \overline{\mathbf{W}}^H\mathbf{n}$ follows the same statistic model as $\mathbf{n}$, i.e., $\mathcal{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^H] = \sigma_n^2\bar{\mathbf{W}}^H\bar{\mathbf{W}}$. Based on the definitions in [12], the input symbol satisfies the power constraint where

$$\mathcal{E}[\mathbf{s}^H\mathbf{s}] \leq P, \tag{4.17}$$

and are circularly symmetric complex Gaussian random variables with covariance $\mathbf{Q}$, i.e.,

$$\mathcal{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{Q}. \tag{4.18}$$

$\mathcal{E}[\mathbf{x}]$ denotes expected value of random vector $\mathbf{x}$. A complex random vector $\mathbf{x}$ being circularly symmetric means it is a zero-mean and proper random vector.

In order to evaluate the quality of the beam training, we evaluate the spectral efficiency achieved during data transmission across different received signal-to-noise ratios. Since the final pencil beams are narrow, we assume no interference's occur between the received signal directions. In practical applications, a spatial distance threshold can be set so that if the received signal directions fall within the threshold, diversity instead of spatial multiplexing will be employed. Spectral efficiency is derived from the Shannon's capacity theorem. MIMO channel capacity based on Shannon's capacity theorem is discussed in [20] and [33]. Based on these two sources, the spectral efficiency is expressed as

$$\bar{R} = \log\det\left[\mathbf{I}_{N_s} + \frac{1}{\sigma_n^2}(\bar{\mathbf{W}}^H\bar{\mathbf{W}})^{+}\bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\mathbf{Q}\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}\right]. \tag{4.19}$$

This expression for spectral efficiency is widely cited but never derived fully in any publication. Here we provide a detailed derivation.

Proof. Given the received signal in (4.9), we have the covariance of the received signal y as

$$\mathcal{E}[\bar{\mathbf{y}}\bar{\mathbf{y}}^H] = \overline{\mathbf{W}}^H\mathbf{H}\overline{\mathbf{F}}\mathbf{Q}\overline{\mathbf{F}}^H\mathbf{H}^H\overline{\mathbf{W}} + \sigma_n^2\overline{\mathbf{W}}^H\bar{\mathbf{W}}.$$

Since by definition $\mathcal{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{Q}$, $\mathcal{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^H] = \sigma_n^2\bar{\mathbf{W}}^H\bar{\mathbf{W}}$, noise $\bar{\mathbf{n}}$ and

transmit symbol $\mathbf{s}$ are independent and having zero correlation, then

$$\begin{aligned}\mathcal{E}[\bar{\mathbf{y}}\bar{\mathbf{y}}^H] &= \mathcal{E}[(\bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\mathbf{s}+\bar{\mathbf{n}})(\mathbf{s}^H\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}+\bar{\mathbf{n}}^H)] \\ &= \bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\mathcal{E}[\mathbf{s}\mathbf{s}^H]\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}+\mathcal{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^H] \\ &= \bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\mathbf{Q}\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}+\sigma_n^2\bar{\mathbf{W}}^H\bar{\mathbf{W}}\end{aligned}$$

The spectral efficiency achieved can be modeled as mutual information $\mathcal{I}(\mathbf{s};\bar{\mathbf{y}})$ between the input and output signal given by

$$\mathcal{I}(\mathbf{s};\bar{\mathbf{y}}) = \mathcal{H}(\bar{\mathbf{y}}) - \mathcal{H}(\bar{\mathbf{y}}|\mathbf{s}) = \mathcal{H}(\bar{\mathbf{y}}) - \mathcal{H}(\bar{\mathbf{n}}),$$

where $\mathcal{H}$ represents the differential entropy of a complex Gaussian variable. Knowing that entropy contains the natural log of the determinant of covariance matrix, we then have the mutual information as

$$\begin{aligned}\mathcal{I}(\mathbf{s};\bar{\mathbf{y}}) &= \log\det\left[\bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\mathbf{Q}\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}+\sigma_n^2\bar{\mathbf{W}}^H\bar{\mathbf{W}}\right] \\ &\quad - \log\det\left[\sigma_n^2\bar{\mathbf{W}}^H\bar{\mathbf{W}}\right] \\ &= \log\det\left[\mathbf{I}_{N_s}+\frac{1}{\sigma_n^2}(\bar{\mathbf{W}}^H\bar{\mathbf{W}})^{+}\bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\mathbf{Q}\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}\right],\end{aligned}$$

where $\det[A]$ represents the determinant of the matrix $\mathbf{A}$. If each signal stream in $\mathbf{s}$ is assumed to have equal power, i.e., $\mathcal{E}[\mathbf{s}\mathbf{s}^H] = \frac{P}{N_s}\mathbf{I}_{N_s}$, the mutual information, in other words, spectral efficiency in (4.19) becomes

$$\bar{R} = \log\det\left[\mathbf{I}_{N_s}+\frac{P}{N_s\sigma_n^2}(\bar{\mathbf{W}}^H\bar{\mathbf{W}})^{+}\bar{\mathbf{W}}^H\mathbf{H}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{H}^H\bar{\mathbf{W}}\right], \tag{4.20}$$

which matches the expression in (4.12) from [16]. □

4.3.3 Spectral Efficiency Assuming Perfect Condition

The maximum capacity is achieved when transmit power is optimally allocated between independent parallel channels. We assume a flat fading channel, which means the channel remains approximately constant for a duration to allow reliable estimation of the channel state at the receiver and timely feedback of such information to the transmitter. Through beam-training algorithms proposed earlier, channel AoDs ($\overline{\mathbf{f}}_i$), AoAs ($\overline{\mathbf{w}}_i$), and channel dominant path power gains ($\overline{\boldsymbol{\alpha}}_i$) are known, which means

the channel state information (CSI) is known at both the transmitter and receiver. In order to simplify (4.19), we assume the perfect channel estimation where $\bar{\mathbf{F}} = \mathbf{V}$, $\bar{\mathbf{W}} = \mathbf{U}$ and $\text{diag}(\bar{\boldsymbol{\alpha}}) = \boldsymbol{\Sigma}$. Substituting the channel SVD in (4.4) into (4.19), we can express spectral efficiency as

$$\dot{R} = \sum_{i=1}^{N_s} \log\left(1 + \frac{P_i \bar{\alpha}_i^2}{\sigma_n^2}\right), \tag{4.21}$$

where P_i is the power allocated to the ith signal stream transmitted over the ith channel path and σ_n^2 is received noise power. The optimum power allocation is obtained from applying water-filling algorithm based on the estimated channel path gains. Given a transmit power constraint $\sum_{i=1}^{N_s} P_i \leq P$, the water-filling power allocation is obtained by maximizing the spectra efficiency under the total power constraint, as shown in the equation below [12]. In order to maximize mutual information, it is well-known to use the water-filling algorithm when channel is known at the transmitter.

Proof. We use a channel with two paths as an example, i.e., $R_H = 2$. The received signal is shown as

$$\begin{aligned}
\bar{\mathbf{y}} &= \begin{bmatrix} \mathbf{w}_1^H \\ \mathbf{w}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^H \\ \mathbf{v}_2^H \end{bmatrix} \begin{bmatrix} \bar{\mathbf{f}}_1 & \bar{\mathbf{f}}_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1^H \mathbf{n} \\ \mathbf{w}_2^H \mathbf{n} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{w}_1^H \\ \mathbf{w}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^H \\ \mathbf{v}_2^H \end{bmatrix} \begin{bmatrix} \bar{\mathbf{f}}_1 s_1 + \bar{\mathbf{f}}_2 s_2 \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1^H \mathbf{n} \\ \mathbf{w}_2^H \mathbf{n} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{w}_1^H \\ \mathbf{w}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^H \bar{\mathbf{f}}_1 s_1 + \mathbf{v}_1^H \bar{\mathbf{f}}_2 s_2 \\ \mathbf{v}_2^H \bar{\mathbf{f}}_1 s_1 + \mathbf{v}_2^H \bar{\mathbf{f}}_2 s_2 \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1^H \mathbf{n} \\ \mathbf{w}_2^H \mathbf{n} \end{bmatrix}.
\end{aligned}$$

Since right singular vectors in $\mathbf{V}$ are orthogonal to each other and we assumed $\bar{\mathbf{F}} = \mathbf{V}$, we have $\mathbf{v}_1^H \bar{\mathbf{f}}_2 = 0$, $\mathbf{v}_2^H \bar{\mathbf{f}}_1 = 0$, $\bar{\mathbf{w}}_1^H \mathbf{u}_2 = 0$, and $\bar{\mathbf{w}}_2^H \mathbf{u}_1 = 0$. Additionally, since the singular vectors and beam-training vectors are all

normal, we have $\mathbf{v}_1^H\bar{\mathbf{f}}_1 = 1$, $\mathbf{v}_2^H\bar{\mathbf{f}}_2 = 1$, $\bar{\mathbf{w}}_1^H\mathbf{u}_1 = 1$, and $\bar{\mathbf{w}}_2^H\mathbf{u}_2 = 1$. Then,

$$\begin{aligned}
\bar{\mathbf{y}} &= \begin{bmatrix}\mathbf{w}_1^H \\ \mathbf{w}_2^H\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{bmatrix}\begin{bmatrix}\mathbf{v}_1^H\bar{\mathbf{f}}_1 s_1 \\ \mathbf{v}_2^H\bar{\mathbf{f}}_2 s_2\end{bmatrix} + \begin{bmatrix}\mathbf{w}_1^H\mathbf{n} \\ \mathbf{w}_2^H\mathbf{n}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{w}_1^H \\ \mathbf{w}_2^H\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}\begin{bmatrix}\sigma_1\mathbf{v}_1^H\bar{\mathbf{f}}_1 s_1 \\ \sigma_2\mathbf{v}_2^H\bar{\mathbf{f}}_2 s_2\end{bmatrix} + \begin{bmatrix}\mathbf{w}_1^H\mathbf{n} \\ \mathbf{w}_2^H\mathbf{n}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{w}_1^H \\ \mathbf{w}_2^H\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\sigma_1\mathbf{v}_1^H\bar{\mathbf{f}}_1 s_1 + \mathbf{u}_2\sigma_2\mathbf{v}_2^H\bar{\mathbf{f}}_2 s_2\end{bmatrix} + \begin{bmatrix}\mathbf{w}_1^H\mathbf{n} \\ \mathbf{w}_2^H\mathbf{n}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{w}_1^H\mathbf{u}_1\sigma_1\mathbf{v}_1^H\bar{\mathbf{f}}_1 s_1 + \mathbf{w}_1^H\mathbf{n} \\ \mathbf{w}_2^H\mathbf{u}_2\sigma_2\mathbf{v}_2^H\bar{\mathbf{f}}_2 s_2 + \mathbf{w}_2^H\mathbf{n}\end{bmatrix} = \begin{bmatrix}\sigma_1 s_1 + \mathbf{w}_1^H\mathbf{n} \\ \sigma_2 s_2 + \mathbf{w}_2^H\mathbf{n}\end{bmatrix}.
\end{aligned}$$

With the input signal power constraint $\mathcal{E}[s_i s_i^*] = P_i$, the covariance of $\mathcal{E}[s_i s_j^*] = 0$ or $\mathbf{s}$ being proper random, the covariance of the received signal is given by

$$\begin{aligned}
\mathcal{E}[\bar{\mathbf{y}}\bar{\mathbf{y}}^H] &= \mathcal{E}\left[\begin{bmatrix}\sigma_1 s_1 + \mathbf{w}_1^H\mathbf{n} \\ \sigma_2 s_2 + \mathbf{w}_2^H\mathbf{n}\end{bmatrix}\begin{bmatrix}\sigma_1^* s_1^* + \mathbf{n}^H\mathbf{w}_1 & \sigma_2^* s_2^* + \mathbf{n}^H\mathbf{w}_2\end{bmatrix}\right] \\
&= \begin{bmatrix}|\sigma_1|^2 P_1 + \sigma_n^2\mathbf{w}_1^H\mathbf{w}_1 & 0 \\ 0 & |\sigma_2|^2 P_2 + \sigma_n^2\mathbf{w}_2^H\mathbf{w}_2\end{bmatrix} \\
\mathcal{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^H] &= \begin{bmatrix}\sigma_n^2\mathbf{w}_1^H\mathbf{w}_1 & 0 \\ 0 & \sigma_n^2\mathbf{w}_2^H\mathbf{w}_2\end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}(\mathbf{s};\bar{\mathbf{y}}) &= \log\det\left[\mathcal{E}[\bar{\mathbf{y}}\bar{\mathbf{y}}^H]\right] - \log\det\left[\mathcal{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^H]\right] \\
&= \log\left(\left(|\sigma_1|^2 P_1 + \sigma_n^2\mathbf{w}_1^H\mathbf{w}_1\right)\left(|\sigma_2|^2 P_2 + \sigma_n^2\mathbf{w}_2^H\mathbf{w}_2\right)\right) - \\
&\log\left(\left(\sigma_n^2\mathbf{w}_1^H\mathbf{w}_1\right)\left(\sigma_n^2\mathbf{w}_2^H\mathbf{w}_2\right)\right) \\
&= \log\left(|\sigma_1|^2 P_1 + \sigma_n^2\mathbf{w}_1^H\mathbf{w}_1\right) + \log\left(|\sigma_2|^2 P_2 + \sigma_n^2\mathbf{w}_2^H\mathbf{w}_2\right) - \\
&\log\left(\sigma_n^2\mathbf{w}_1^H\mathbf{w}_1\right) - \log\left(\sigma_n^2\mathbf{w}_2^H\mathbf{w}_2\right) \\
&= \log\left(1 + \frac{|\sigma_1|^2 P_1}{\sigma_n^2}\right) + \log\left(1 + \frac{|\sigma_2|^2 P_2}{\sigma_n^2}\right) \\
&= \sum_{i=1}^{N_s}\log\left(1 + \frac{P_i\sigma_i^2}{\sigma_n^2}\right) = \sum_{i=1}^{N_s}\log\left(1 + \frac{P_i\bar{\alpha}_i^2}{\sigma_n^2}\right).
\end{aligned}$$

□

The above derivation uses many assumptions. Realistically, $\bar{\mathbf{F}} = \mathbf{V}$ and $\bar{\mathbf{W}} = \mathbf{U}$ may never occur; the beam training vector product $\mathbf{w_1}^H \mathbf{w_2}$ is small but never exactly zero. Without taking channel SVD and applying water-filling algorithm, (4.21) can be expressed as

$$\dot{R} = \sum_{i=1}^{N_s} \log\left(1 + \frac{P}{N_s \sigma_n^2} \left|\mathbf{w}_i^H \mathbf{H} \mathbf{f}_i\right|^2\right), \tag{4.22}$$

where total power P is equally allocated to each transmitted symbol on each path.

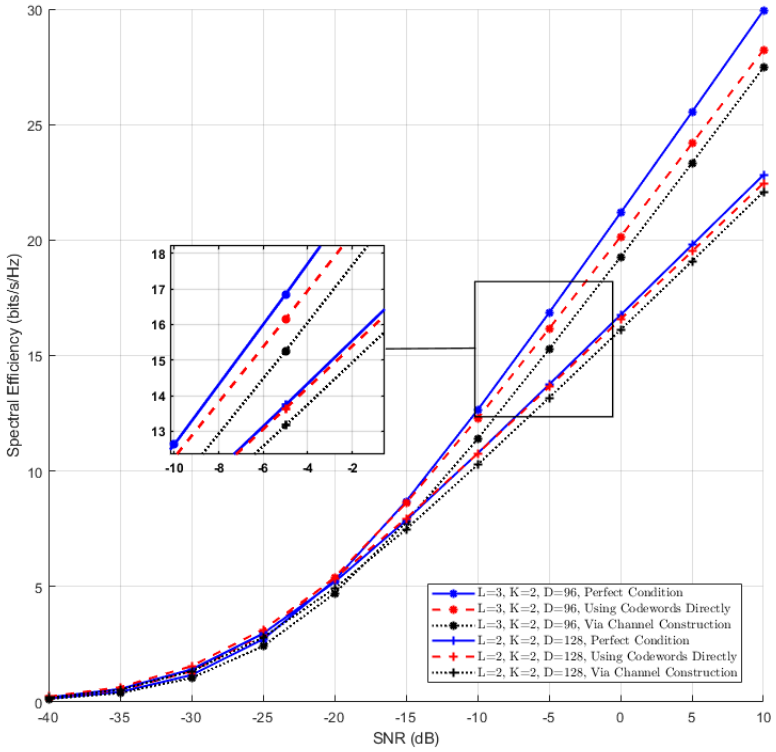


Figure 4.4: Comparison of spectral efficiency calculated by (4.22) under perfect estimation, (4.20) using codewords directly , and (4.12) via channel construction, all with equal power among transmitted symbols.

4.3.4 Water-filling Algorithm

In the previous section, we derived the expression of spectral efficiency when $\bar{\mathbf{F}}$ and $\bar{\mathbf{W}}$ match precisely to channel left and right singular vectors, under the assumption

of a known channel. However, these assumptions are not the case. For the water-filling algorithm, (4.21) offers a valuable insight to design power allocation among transmitted symbols over each independent channel paths.

The goal of water-filling is to maximize spectral efficiency under transmit symbol power constraint. The spectral efficiency is given by

$$R = \max_{P_i:\sum_i P_i \leq P} \sum_{i=1}^{N_s} \log(1 + \frac{\bar{\alpha}_i^2 P_i}{\sigma_n^2}) = \max_{P_i:\sum_i P_i \leq P} \sum_{i=1}^{N_s} \log(1 + \frac{\gamma_i P_i}{P}), \tag{4.23}$$

where $\gamma_i = \bar{\alpha}_i^2 P/\sigma_n^2$ is the received SNR associated with the ith channel, assuming it is allocated with all the power budget and $\bar{\alpha}_i^2$ is the ith channel path gain obtained from applying beam training algorithms. Solving the above optimization problem leads to a power allocation for MIMO channel as:

$$\frac{P_i}{P} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma_i}, & \gamma_i \geq \gamma_0 \\ 0, & \gamma_i < \gamma_0 \end{cases} \tag{4.24}$$

for some cutoff value γ_0 we need to find. By substituting (4.24) into the power constraint, γ_0 must satisfy

$$\sum_i^{N_s} \frac{P_i}{P} = \sum_i^{N_s} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma_i} \right) = 1. \tag{4.25}$$

The broad idea behind power allocation is to allocate more power to the stronger path, i.e., taking advantage of good channel conditions to send higher data rate over the channel. If SNR fall below the cutoff value, no power is allocated to the channel path and that path is not used.

By creating an intermediate matrix $\bar{\mathbf{Q}} = \text{diag}(\mathbf{P})$ where $\mathbf{P} = [\frac{P_1}{P}, \frac{P_2}{P}, ..., \frac{P_{N_s}}{P}]$, water-filling results can then be integrated into (4.19) as

$$\bar{R} = \log\det \left[\mathbf{I}_{N_s} + \frac{P}{\sigma_n^2} (\bar{\mathbf{W}}^H \bar{\mathbf{W}})^+ \bar{\mathbf{W}}^H \mathbf{H} \bar{\mathbf{F}} \bar{\mathbf{Q}} \bar{\mathbf{F}}^H \mathbf{H}^H \bar{\mathbf{W}} \right]. \tag{4.26}$$

Using the above equation and (4.20), spectral efficiency is simulated over 1000 random channel generations. The result in Fig. 4.6 shows the water-filling algorithm helps to achieve higher data rate when the received SNR is low and does not make a difference when received SNR is high. This observation is as expected from the discussion prior.

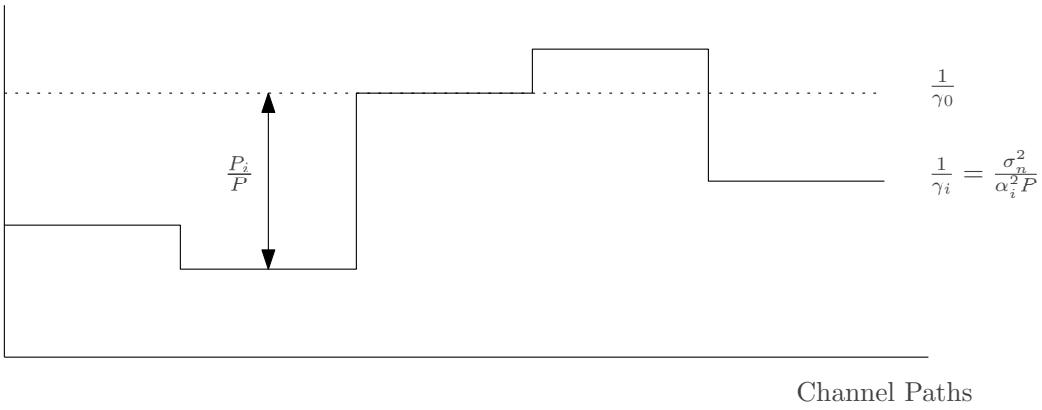


Figure 4.5: Water-filling across channel paths

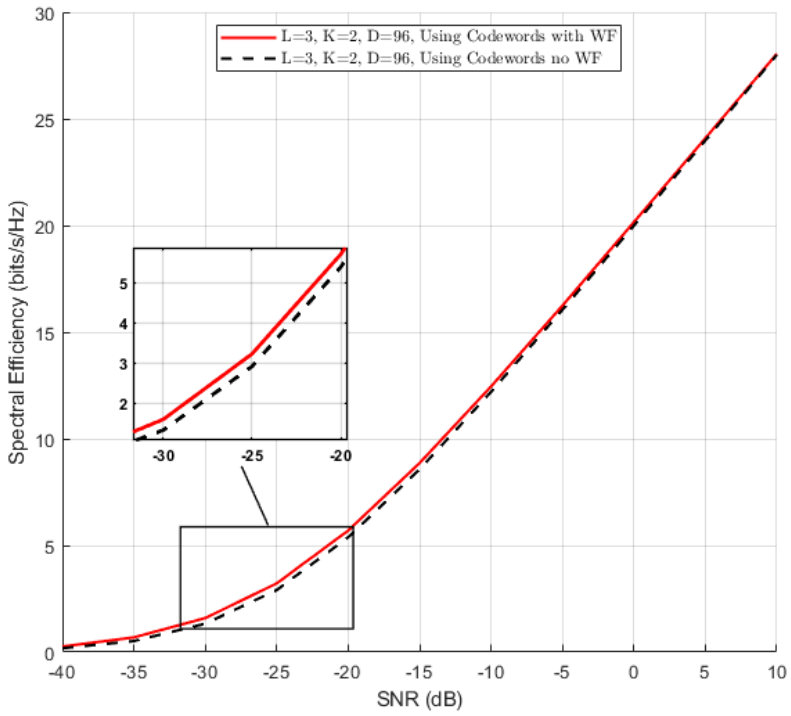


Figure 4.6: Compare effect of water-filling algorithm on spectral efficiency calculated using (4.20) and (4.26). Water-filling algorithm improves spectral efficiency especially when received SNR is poor by allocating more power towards stronger channel paths.

.

To conclude, the beam training algorithms for both single path and multiple path discovery using hierarchical casebooks are proposed. Restricted orthogonal projection is proposed in order to improve multipath estimation accuracy. Instead of channel contruction from beam training, a data transmission scheme, where trained codewords are used directly to transmit symbols is proposed. The simplified data transmission design is proven to be superior compared to traditional hybrid precoding design. In the next chapter, the proposed beam training algorithms and data transmission scheme will be evaluated against existing designs. Some key design parameters and trade-off they offer will be discussed to offer design insights for future implementation.

Chapter 5

Evaluation, Analysis, and Comparisons

In this chapter, we present numerical simulation results of the proposed codebook design and multi-path beam training algorithm. Throughout the simulation, the complex noise power level is fixed as one. For each simulation ran, the channel model is as described in (2.10), with complex channel path gain modeled as a zero-mean unit variance independent and identically distributed complex Gaussian, i.e., $\alpha_j \sim \mathcal{CN}(0,1)$. The number of channel paths is $R_H = 3$. The channel angle of departures (AoDs) at the transmitter and angle of arrivals (AoAs) at the receiver are assumed to take quantized values and uniformly distributed over $[0, \frac{\pi}{128}, ... \frac{127\pi}{128}]$, i.e., the angular resolution for channel paths is $D = 128$.

A hybrid system architecture is adopted as presented in Fig. 2.1. The transmitter has $N_t = 64$ antennas and $N_{RF}^t = 10$ RF chains, while the receiver has $N_r = 32$ antennas and $N_{RF}^r = 5$ RF chains. The antenna arrays are assumed to be uniform linear array with $\lambda/2$ element spacing. The RF phase shifters are assumed to have two bits, i.e., phase shifter values are from $[1, -1, j, -j]$. At the end of beam training, the chosen final level codewords represent the found channel paths and estimated channel path gain were fed back to the transmitter. The codeword of the finest angular range are then used for data transmission with water-filling power allocation at the transmitter. The spectral efficiency is calculated averaging the numerical data rate over 5000 randomly generated channel models defined in in (2.10). The precoding and combining vectors are beams trained applying algorithms proposed in Chapter 4 using the codebook defined in Chapter 3.

5.1 Mean Square Error

One way to evaluate the recovered channel path is to reconstruct the channel $\overline{\mathbf{H}}$ using

$$\overline{\mathbf{H}} = \bar{\mathbf{W}}\bar{\mathbf{\Sigma}}\bar{\mathbf{F}}^H, \tag{5.1}$$

where $\overline{\mathbf{W}}$ and $\overline{\mathbf{F}}$ are matrices with selected channel paths defined in (4.9) and $\overline{\mathbf{\Sigma}}$ is diagonal matrix with channel path gain $\overline{\boldsymbol{\alpha}}$ obtained from Algorithm 2 along its diagonal. The Normalized Mean Square Error (NMSE) of an estimated channel over an original channel is defined as

$$\text{NMSE} = \frac{\|\mathbf{H} - \overline{\mathbf{H}}\|_2^2}{\|\mathbf{H}\|_2^2}. \tag{5.2}$$

NMSE of the estimated $\overline{\mathbf{H}}$ using Algorithm 2 is compared to NMSE of the channel $\hat{\mathbf{H}}_L$ reconstructed using subsets of singular vectors and singular values.

For the single path beam training, $\hat{\mathbf{H}}_1$ is generated using the largest singular value of channel SVD and its corresponding left and right singular vectors, i.e. $\hat{\mathbf{H}}_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H$. Similarly, when $L \leq R_H$ independent channel paths are found, $\hat{\mathbf{H}}_L$ is defined as

$$\hat{\mathbf{H}}_L = \sum_{j=1}^{L} \sigma_j \mathbf{u}_j \mathbf{v}_j^H, \tag{5.3}$$

where singular values σ_j and its corresponding singular vectors are arranged in deceasing order.

Fig. 5.1 compares the NMSE of estimated channel using the proposed algorithm to the corresponding reconstructed channel using a subset of SVD. The test cases used here are the same as in Fig. 5.3. Algorithm 2 is used for simulations looking for three and two channel paths and Algorithm 1 is used for the single path scenario. From Fig. 5.1, we can see that Algorithm 2 is able to reconstruct the channel with NMSE similar to the SVD approach. Additionally, for the channel model with three paths, two paths has significantly smaller NMSE compared to a single path, but three paths are only marginally better compared to two paths. This shows the importance of finding multiple paths but also inspires us to rethink whether it's worth the effort to always try and find more paths.

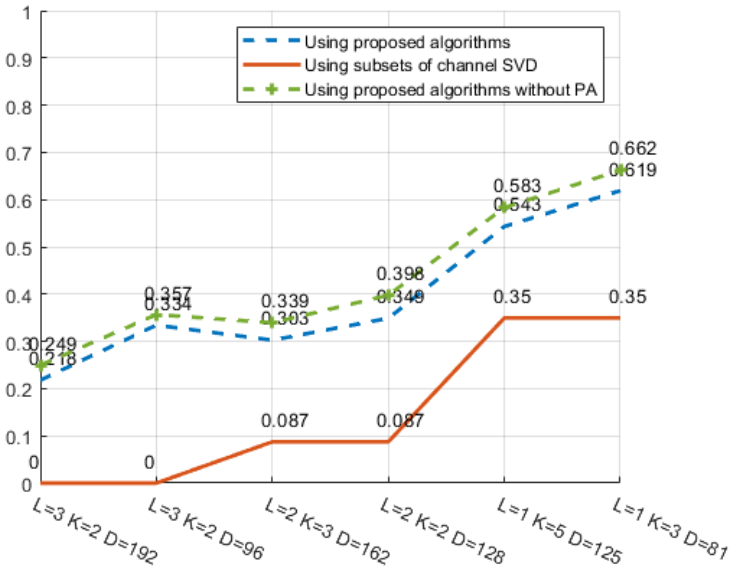


Figure 5.1: Average normalized mean square error over 5000 channel simulations.

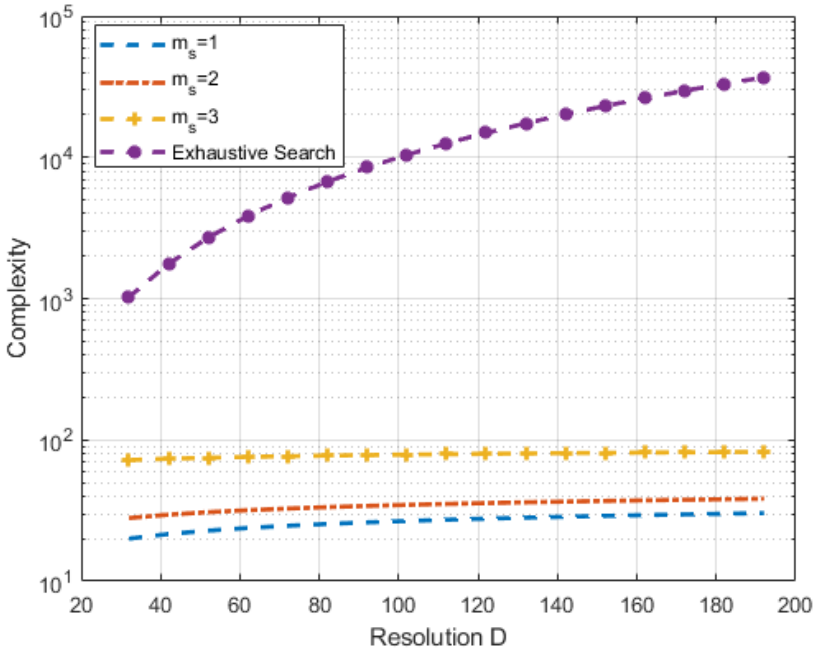


Figure 5.2: Complexity versus resolution for iterative search, hierarchical codebook starting at codebook level 1, 2, and 3.

5.2 Complexity

Since the hybrid codewords are designed ahead of time, computation time in real-time communication does not include codebook generation time. The complexity is calculated as

$$L[(LK^{m_s})^2 + K^2(M - m_s)], \tag{5.4}$$

where L is the number of desired channel paths, K is the number of directions to search at each stage which is the same for both transmitter and receiver, $M = \log_K(\frac{D}{L})$ is the hierarchical codebook height, D is the angular resolution, and m_s is codebook start level. The first term in (5.4) represents time spent on measurements in the first stage of beam training while the second term contains time spent on the remaining stages. Notice when $m_s = M$, the second term in (5.4) becomes zero and the beam training becomes an exhaustive search through all interested directions D.

Fig. 5.2 compares complexity of the proposed hierarchical codebook with exhaustive search when angular resolution varies. It is shown that hierarchical codebook significantly reduce complexity compared to the exhaustive search method. Starting at codebook level 3 adds significant computation time compared to level 2. In order to improve simulation accuracy, it can be worth while to start with codebook level 2 but going to codebook level 3 will significantly increase training overhead.

5.3 Discussion

Fig. 5.3 compares the performance of the proposed beam training and data transmission design with the reference design in [1]. The improved multi-path design with restricted orthogonal projection is shown to have improved spectral efficiency by nearly 30%. The reference design is showing really low spectral efficiency when $L = 1, K = 3, D = 81$. The reason is that the set of quantized angles with resolution $D = 81$ is too coarse to successfully generate sector beams using (3.2).

There are many variables in the system that influence design performance. This section aims to provide a summary on the trade-offs. The following sections will further discuss design parameters' influence on beam training.

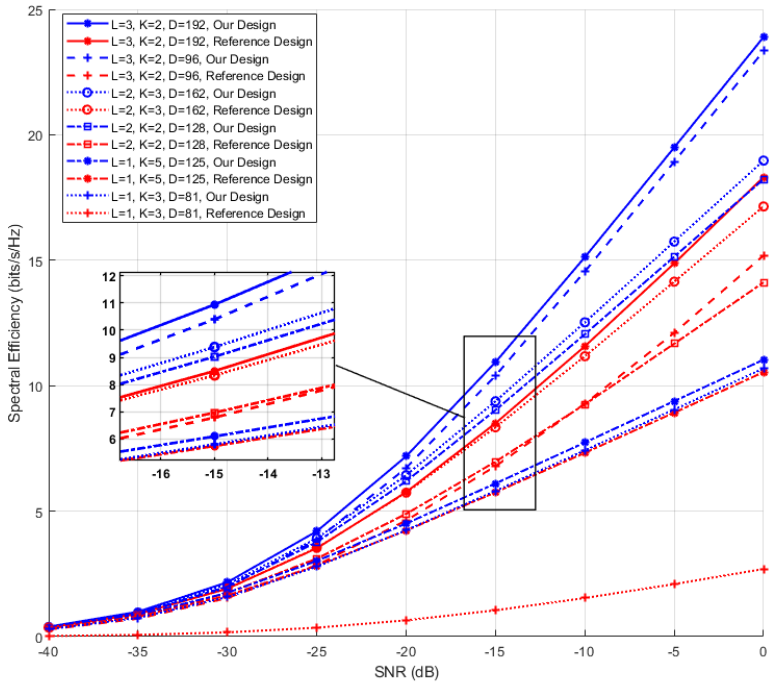


Figure 5.3: Comparison of spectral efficiency achieved with our design with the reference literature [1] over selected cases

5.3.1 Restricted Orthogonal Projection

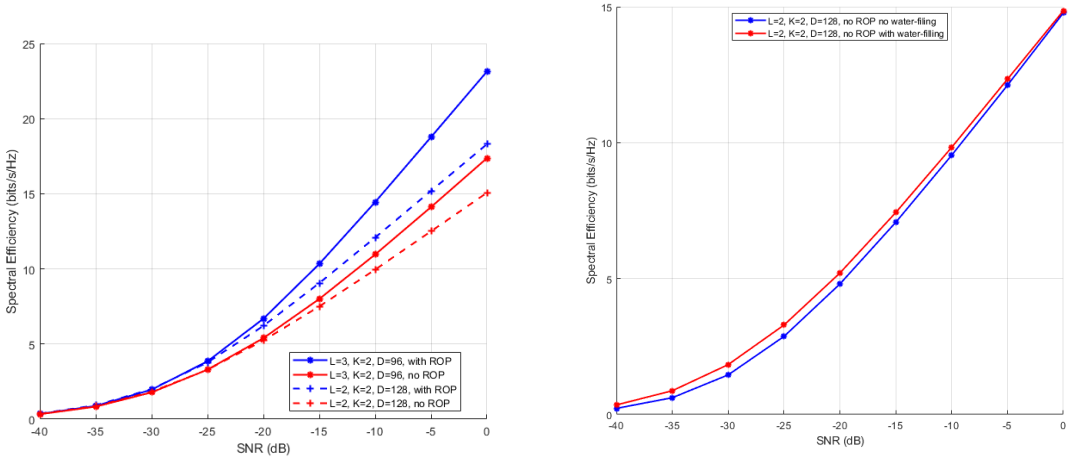
One important contribution of this study is the restricted orthogonal projection. During simulation, we observe that the subtraction of the influence of the found channel path from the received signal will introduce error when the received signal does not contain influence of the found channel path. In other words, using hierarchical codebook, when a searched sector does not contain the previously found path, no orthogonal projection should be performed. If a searched sector does not contain the previously found path, the orthogonal projection is not strictly zero and thus introduce an error in the detection. This observation has never been mentioned in the literature discussing multi-path beam training both in [1] and [19]. By restricting orthogonal projection to the first stage of beam training, multi-path beam training accuracy is improved which lead to significant improvement in data rate using multiple directions.

The examples used in Fig. 5.4 are $L = 2$, $K = 2$, and $D = 128$ for 2-path beam training and $L = 3$, $K = 2$, and $D = 96$ for 3-path beam training. The blue lines represent data rate using the beams selected applying ROP. Fig. 5.4 shows at 0 dB SNR, at most 30% increase in spectral efficiency can be seen compared to when no ROP is applied. Such improvement is because of the improved beam training alignment using ROP.

Fig. 5.4 also compares the effect of water-filling during data transmission. To create hybrid precoder and combiner like shown in [1], both transmitter and receiver needs to reconstruct channel using AoAs, AoDs, and channel path gains found from the beam training process. Such hybrid precoding process can take up significant computation resources and may lead to more transmission overhead. Instead, we propose to use trained beams directly for data transmission. To optimize data rate, water-filling algorithm is applied as described in Section 4.3.2. The result shows that the water-filling algorithm will improve data rate at the low SNR region by allocating energy to the strongest channel path.

5.3.2 Coding the Beams with Walsh Code

As discussed in Chapter 2, when orthogonal beams are used, the beam coded with Walsh code will enjoy a reduced power fluctuation between the main lobe and sidelobes as discussed in [4]. The reduced power fluctuation between the main lobe and sidelobes observed from both receiving and transmitting beams will improve beam-



(a) Influence of ROP

(b) Influence of Water-filling

Figure 5.4: Compare spectral efficiency achieved from applying Restricted Orthogonal Projection and water-filling algorithm

forming training accuracy.

In Fig. 5.5, since $K = 2$, a minimum two time slots are required to train two transmitting beams simultaneously. When 32 slots are allocated for beam training, the Walsh code method achieve higher spectral efficiency compared to without beam coding. When only 2 time slots are available, Walsh code only has a slight advantage.

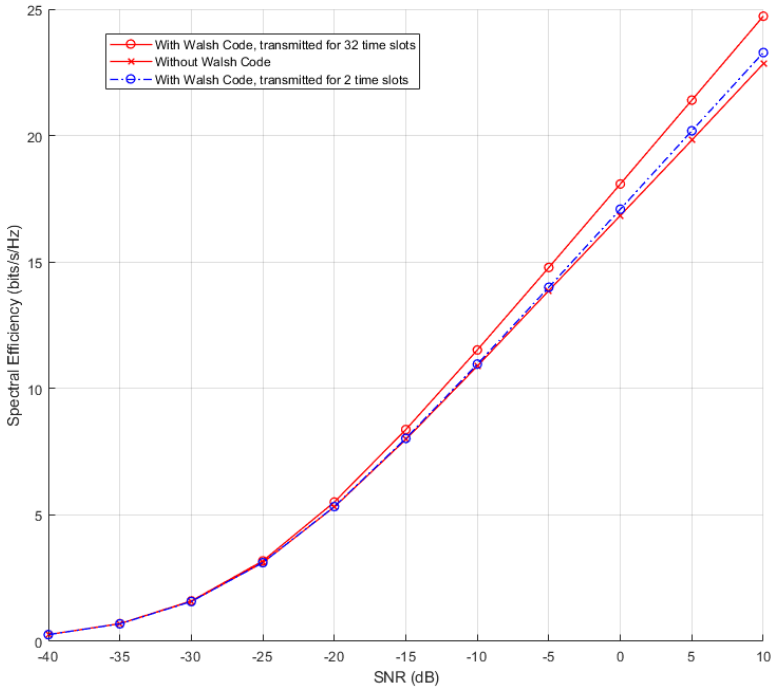


Figure 5.5: Spectral efficiency achieved via beam training with Walsh code coded beams. The simulation parameters are $L = 2$, $K = 2$, and $D = 128$.

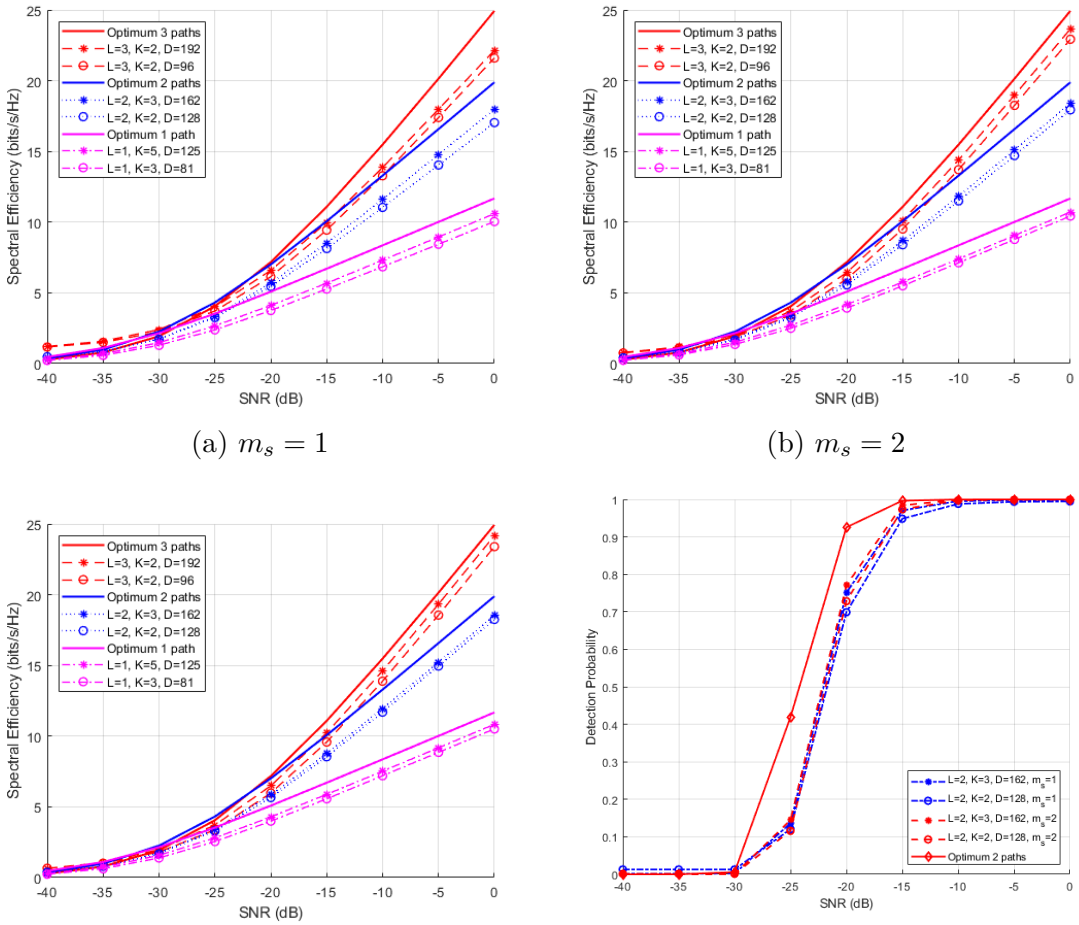
5.3.3 Codebook Design Starting Level

In [26], the author discussed the initial angular partition estimation accuracy being the bottleneck to improve beam training beam alignment. Since the first level codebook introduces the largest error, starting from level two in a codebook with finer angular resolution can improve detection accuracy but at the same time increase beam training time. It is a balancing act to choose the number of codewords in the initial level.

Note for Algorithm 2, the initial level codeword is divided into KL sectors instead of K in Algorithm 1. This design decision is based on the fact that codewords have

low resolutions at the beginning codeword level. If the dominant path and the second best path fall into the same codeword at the initial level, it becomes impossible to distinguish the two using ROP. Our study shows this potential drawback using ROP does not significantly influence beam training accuracy when the first level codebook resolution is increased. As shown in Fig. 5.6, when staring level m_s goes to three, the spectral efficiency is closest to the optimum situation. Given the pre-condition that the channel path spatial separation is known to the designer, he or she can decide whether to implement finer starting codewords in for the price of increased training time.

To assess detection performance, we also keep a record of number of times the received signal is above certain threshold during the Monte Carlo simulation. Based on the Neyman-Pearson detector, the threshold is set to $\gamma = -\sigma_n \ln P_{FA}$, where P_{FA} is false alarm probability and σ_n is noise standard deviation. The false alarm probability for each path estimation is set to $P_{FA} = 0.01$. The outage probability is compared in Fig. 5.6d.



(a) $m_s = 1$ (b) $m_s = 2$

(c) $m_s = 3$ (d) Detection Probability

Figure 5.6: Spectral efficiency achieved applying the proposed codebook and beam training algorithm with $R_H = 3$ and number of paths $L = 1, 2, 3$. The figure compares performance of different values of subdivision K, angular resolution D, and starting level m_s. Channel singular vectors are used as beamforming vectors to represent the optimum scenario.

5.3.4 Power Allocation in Codebook Design

Power variation in antenna beams over different beamwidths at different codebook levels will influence beam training accuracy [26]. In [1], a power allocation scheme is proposed to reduce this effect. The power allocation scheme allows more transmitting and receiving power towards beams with lower directivity. In Section 3.3, this is categorized as codebook power allocation. From our simulation, however, with the same total power duration beam training, more power allocated at coarse beams does not improve performance compared to even power allocation. This can be observed on Fig. 5.7.

Another type of power allocation is codeword power allocation as described in Section 3.3. By allocating power among beams with the same beamwidth, we can achieve even more uniform main lobe directivity across codewords in the same codebook level. This is also found having no significant improvement. The reason is that the variation among beams in the decibel level is very small. In the antenna design, 3 dB difference is the minimum to be identified as significant. Since power allocation in Fig. 5.7 combines both codebook and codeword PA, we can conclude PA is not a significant improvement in the current design.

5.3.5 Hybrid Adaptation in Codebook Design

We adopt hybrid precoding algorithm proposed in [7] with 2-bit phase shifters. As shown in Fig. 5.8a, hybrid beamforming vectors using 2-bit phase shifters are able to achieve near optimum data rate yet 1-bit phase shifters compromises performance by about 30%.

Fig. 5.8a compares 1-bit, 2-bit phase shifter and no phase shifter quantization. It shows that 2-bit phase shifters can achieve near optimum result whiling 30% more accurate compared to 1-bit phase shifters. For this reason, 2-bit phase shifters are implemented in the design.

5.3.6 Angular Resolution

The higher angular resolution in codebook design, the finer a beamwidth will go, leading to an increased beam directivity; too fine of a beamwidth will lead to increased training complexity. It is worth noting the narrowest achievable antenna array beamwidth is limited by its half-power beamwidth. Based on (3.1), we can

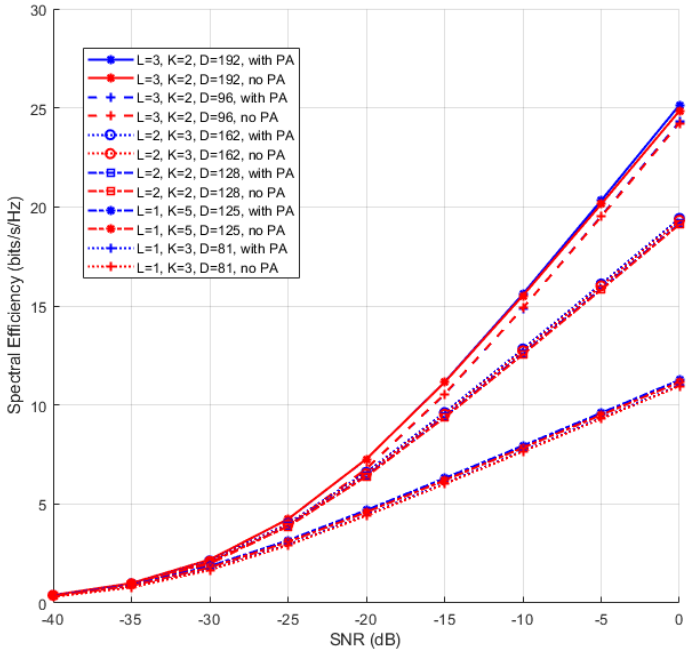
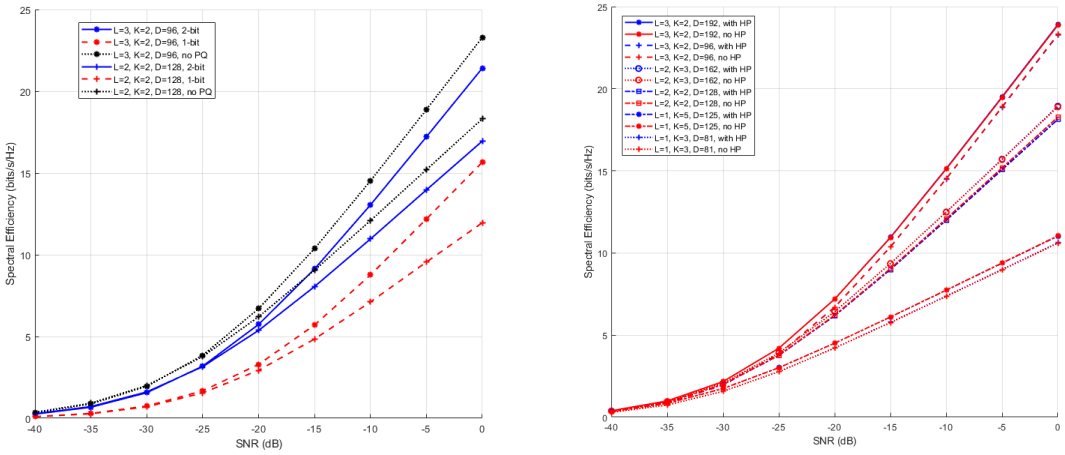


Figure 5.7: Compare spectral efficiency with or without power allocation at both codeword and codebook level



(a) Influence of Phase Shifter Quantization (b) Influence of Hybrid Precoding

Figure 5.8: Compare spectral efficiency achieved from hybrid precoding in [7] and Phase Shifter Quantization (PQ)

plot the relationship of beamwidth and steering angle. As shown in Fig. 5.9, the beamwidth decreases when the beam is steered away from broadside and deceases with a larger array size. For an 64-element array, the angular resolution must not be larger than $\frac{\pi}{\min \Delta\phi_{3dB}} = \frac{180\,\text{deg}}{1.6\,\text{deg}} = 112$. This mathematical proof also matches simulation results. As shown in Fig. 5.6, with the same number of paths, increase angular resolution D does not lead to significant improvement on the achievable spectral efficiency. In the design of codebooks, specific considerations much be given regarding the antenna implemented.

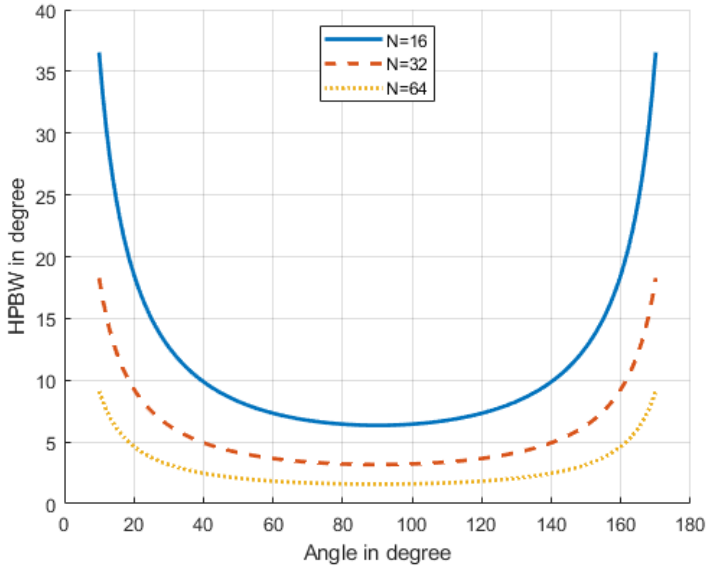


Figure 5.9: Array 3 dB beamwidth versus steering angle assuming uniform weights. N is the number of elements.

5.4 Conclusion

We have evaluated beam training performance based on mean square error, complexity, outage probability, and spectral efficiency. By using the proposed codebook, beam training algorithm, and data transmission design, data rate in simulation has been increased by nearly 30% compared with the reference design in [1]. The influence of restricted orthogonal projection, orthogonal beam coding, codebook starting level, power allocation, and angular resolution have been discussed with simulation results; design considerations have been given relating to these factors.

Chapter 6

Conclusions and Future Work

Compared to a machine learning approach, multi-resolution BF training is already implemented in the standard, thus making it more practical. A single communication link through beamforming only explores diversity gain of a MIMO system. In order to take advantage of the multiplexing gain of MIMO systems, two or more channel paths are required to send multiple signal streams simultaneously. It is therefore important to discover multiple channel paths for data transmission in the BF training stage, especially when the received SNR is high.

In this thesis, we have investigated codebook design, multi-path channel estimation and data transmission methods considering a single-user mmWave system. The hierarchical codebook design is improved by implementing a hybrid sector array design using Kaiser window [5] and an Alt-Min hybrid precoding algorithm [7].

The Restricted Orthogonal Projection based (ROP-based) algorithm is proposed to increase multi-path channel estimation. Orthogonal projection is the key mathematical tool to achieve multi-path channel estimation. The idea is to subtract the best channel path from the received signal in order to discovery the second-best channel path. From analyzing simulation results, we have found that errors are introduced when trying to subtract the best channel path from the received signal when the received signal does not contain the best channel path. To avoid this situation, restricting orthogonal projection to the first stage BF training is proposed to improve multi-path BF training accuracy.

To reduce the communication overhead, we have verified the feasibility of using trained codewords directly for data transmission. Since BF vectors in a codebook are defined and stored in the device prior to data transmission, no overhead is introduced by designing and optimizing BF vectors in a codebook ahead of time. Therefore, using

codewords for data transmission eliminates overhead introduced by channel construction and hybrid precoding, making the proposed BF training ideal for commercial implementation. The proposed codebook, beam training algorithm, and data transmission design improves data rate by nearly 30% compared to the reference design in [1].

For future work, the proposed multi-path BF training algorithm and data transmission scheme should be tested against parametric channel models based on real measurements to verify its effectiveness. Additionally, spectral efficiency in this study is based on calculation; empirical data rate applying Golay sequence as specified in the IEEE802.11ad standard will help implement the proposed BF training into actual engineering practices.

Appendix A

Kaiser Window Design Steps

1. Calculate the D-factor of the window from the user-specified attenuation A

$$D = \begin{cases} \frac{A-7.95}{14.36}, & \text{if } A > 21 \\ 0.922, & \text{if } A \leq 21 \end{cases}$$

and the parameter α that shapes the window is:

$$\alpha = \begin{cases} 0.1102(A - 8.7), & \text{if } A \geq 50 \\ 0.5842(A-21)^{0.4} + 0.07886(A-21), & \text{if } 21 < A < 50 \\ 0, & \text{if } A \leq 21 \end{cases}$$

2. Given the linear array width N, the transition width $\Delta\psi$ can be calculated by:

$$\Delta\psi = \frac{2\pi D}{N-1} \tag{A.1}$$

We can observe from this equation that the transition width reduces with larger N and increases with bigger A.

3. The Kaiser window weights are calculated as:

$$w(m) = \frac{I_0(\alpha\sqrt{1-m^2/M^2})}{I_0(\alpha)}, \quad m = 0, \pm 1, ..., \pm M, \tag{A.2}$$

where $I_0(x)$ is the modified Bessel function of first kind and zeroth order.

4. The cutoff frequency at the middle between the passband and stopband fre-

quencies is:

$$\psi_b = \frac{1}{2}(\psi_p + \psi_s) = \psi_p + \frac{1}{2}\Delta\psi \tag{A.3}$$

5. The final windowed array weights are calculated as:

$$a(m) = w(m)e^{-jm\psi_0}\frac{sin(\psi_b m)}{\pi m}, \quad m = 0, \pm 1, ..., \pm M, \tag{A.4}$$

Bibliography

[1] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[2] "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.

[3] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, "IEEE 802.11ad: Directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi [Invited Paper]," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, Dec. 2014.

[4] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the Beams: Improving Beamforming Training in mmWave Communication System," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec. 2011, pp. 1–6.

[5] S. Orfanidis, *Electromagnetic Waves and Antennas*. Sophocles J. Orfanidis, 2016.

[6] A. J. Laub, *Matrix Analysis For Scientists And Engineers*. USA: Society for Industrial and Applied Mathematics, 2004.

[7] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[8] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[9] K. Hassan, M. Masarra, M. Zwingelstein, and I. Dayoub, "Channel Estimation Techniques for Millimeter-Wave Communication Systems: Achievements and Challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1336–1363, 2020.

[10] H. Xu, V. Kukshya, and T. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 620–630, Apr. 2002.

[11] S. Sun, T. S. Rappaport, M. Shafi, P. Tang, J. Zhang, and P. J. Smith, "Propagation Models and Performance Evaluation for 5G Millimeter-Wave Bands," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8422–8439, Sep. 2018.

[12] A. Goldsmith, *Wireless Communications*, 1st ed. Cambridge University Press, Aug. 2005.

[13] H. Shokri-Ghadikolaei, L. Gkatzikis, and C. Fischione, "Beam-searching and transmission scheduling in millimeter wave communications," in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015, pp. 1292–1297.

[14] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *2012 Information Theory and Applications Workshop*, Feb. 2012, pp. 234–239.

[15] W. Ni, X. Dong, and W. Lu, "Near-Optimal Hybrid Processing for Massive MIMO Systems via Matrix Decomposition," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

[16] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[17] T. Baykas, C.-S. Sum, Z. Lan, J. Wang, M. A. Rahman, H. Harada, and S. Kato, "IEEE 802.15.3c: The first IEEE wireless standard for data rates over 1 Gb/s," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 114–121, Jul. 2011.

[18] D. De Donno, J. Palacios, and J. Widmer, "Millimeter-Wave Beam Training Acceleration Through Low-Complexity Hybrid Transceivers," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3646–3660, Jun. 2017.

[19] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *2012 Information Theory and Applications Workshop*, 2012, pp. 234–239.

[20] E. Telatar, "Capacity of Multi-antenna Gaussian Channels: Capacity of Multi-antenna Gaussian Channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[21] T. Gong, N. Shlezinger, S. S. Ioushua, M. Namer, Z. Yang, and Y. C. Eldar, "RF Chain Reduction for MIMO Systems: A Hardware Prototype," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5296–5307, Dec. 2020.

[22] X. Song, T. Kühne, and G. Caire, "Fully-Connected vs. Sub-Connected Hybrid Precoding Architectures for mmWave MU-MIMO," *arXiv:1810.13161 [cs, math]*, Oct. 2018.

[23] J. Mo, B. L. Ng, S. Chang, P. Huang, M. N. Kulkarni, A. Alammouri, J. C. Zhang, J. Lee, and W. Choi, "Beam Codebook Design for 5G mmWave Terminals," *IEEE Access*, vol. 7, pp. 98 387–98 404, 2019.

[24] A. Forenza, D. J. Love, and R. W. Heath, "Simplified Spatial Correlation Models for Clustered MIMO Channels With Different Array Configurations," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 1924–1934, Jul. 2007.

[25] A. M. Sayeed and V. Raghavan, "Maximizing MIMO Capacity in Sparse Multipath With Reconfigurable Antenna Arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 156–166, Jun. 2007.

[26] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[27] H. L. V. Trees, "Appendix A: Matrix Operations," in *Optimum Array Processing*. John Wiley & Sons, Ltd, 2002, pp. 1340–1406.

[28] Y. Wang, X. Wen, Y. Chen, W. Jing, and Q. Pan, "Joint 3D Codebook Design and Beam Training for UAV Millimeter-Wave Communications," in *2019 IEEE 30th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops)*, Sep. 2019, pp. 1–6.

[29] R. J. Mailloux, *Phased Array Antenna Handbook*, 2nd ed., ser. Artech House Antennas and Propagation Library. Boston: Artech House, 2005.

[30] G. Strang, "Computations with large matrices," in *Linear Algebra and Learning from Data*. Wellesley MA: Wellesley-Cambridge Press, 2019, ch. 2, pp. 113–134.

[31] 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 Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[32] J. Wang, Z. Lan, Chang-woo Pyo, T. Baykas, Chin-sean Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[33] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, Jun. 2003.