Beam Training Optimization for Millimeter Wave Communication

Evan R. Ding

A Thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2018

Reading Committee:

Sumit Roy, Chair

Sreeram Kannan

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Beam Training Optimization
for Millimeter Wave
Communication

Evan R. Ding

Chair of the Supervisory Committee:
Professor Sumit Roy
Electrical Engineering

Millimeter wave (mmWave) technology has emerged as a promising solution to the spectrum demands of 5G networks. To achieve the necessary link budget for high throughput data transfer, mmWave links employ beamforming using antenna arrays with a large number of elements. However, this introduces the need for *beam training*, the process of finding beam alignments that can support a robust data link. Due to the dynamic nature of 5G environments, beam training must be conducted quickly since devices move in and out of communication range over short time scales. In this thesis, the optimization of beam training is explored within this context. Specifically, we derive a complexity bound for beam training in the limited scope of an idealized Boolean model, and a class of algorithms which achieve the bound is presented. The performance of these algorithms in non-ideal contexts are evaluated through simulation, and improvements over the existing IEEE 802.11ad standard are demonstrated.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

TX: Device seeking to transmit application data

RX: Device seeking to receive application data

AP: Access Point

SNR: Signal-to-noise-ratio

DAC: Digital-to-Analog Converter

ADC: Analog-to-Digital Converter

IID: Independent and Identically Distributed

RMS: Root Mean Square

LOS: Line-of-Sight

NLOS: Non-Line-of-Sight

AOA: Angle of Arrival

AOD: Angle of Departure

CSMA/CA: Carrier Sensing Multiple Access (Collision Avoidance)

MIMO: Multiple-Input and Multiple-Output

SISO: Single-Input and Single-Output

ARV: Antenna Response Vector

AWV: Antenna Weight Vector

BPV: Beam Pattern Vector

BBPV: Binary Beam Pattern Vector

DSCM: Discretized Spatial Channel Matrix

BCM: Binary Channel Matrix

SLS: Sector Level Sweep

SSW: Sector Sweep

ACKNOWLEDGMENTS

Many thanks to Professor Sumit Roy for his guidance and support. I most certainly would still be chasing rabbits down bottomless holes without his advisement.

DEDICATION

To all the people who planned out their thesis a year in advance and wrote it incrementally.

For having more foresight than I did.

Chapter 1
INTRODUCTION

The onset of 5G networks has resulted in a scramble to identify new bands in the electromagnetic spectrum to support emerging services. A significant component of 5G networking consists of Device-to-Device (D2D) or Machine-to-Machine (M2M) communication, and more than 20 billion connected devices are expected to be in use by 2020 [1]. 5G spectrum will include frequencies above 24 GHz, where millimeter wave technology (mmWave) will be deployed. Specifically, the FCC is set to open up 4 GHz of spectrum in 28, 37, and 39 GHz bands for licensed operation, and another 7 GHz in the 64-71 GHz bands for unlicensed operation [5].

As of January 2018, no standards have been defined for these new bands. However, they are expected to be similar to IEEE 802.11ad, which regulates operation in the existing 60 GHz band. 802.11ad, itself, is undergoing an evolution to 802.11ay, incorporating greater channel bandwidth and support for up to four MIMO streams. As a result, improving the performance of mmWave networks is currently a topic of great interest.

1.1 Beamforming in mmWave

One of the main challenges of mmWave systems is the significant increase in free space attenuation resulting from increased carrier frequency, compared to Wi-Fi. From the standard Friis equation, the received power of a radio wave is proportional to the square of wavelength

[1] According to Gartner Inc. Report 2017 [4]

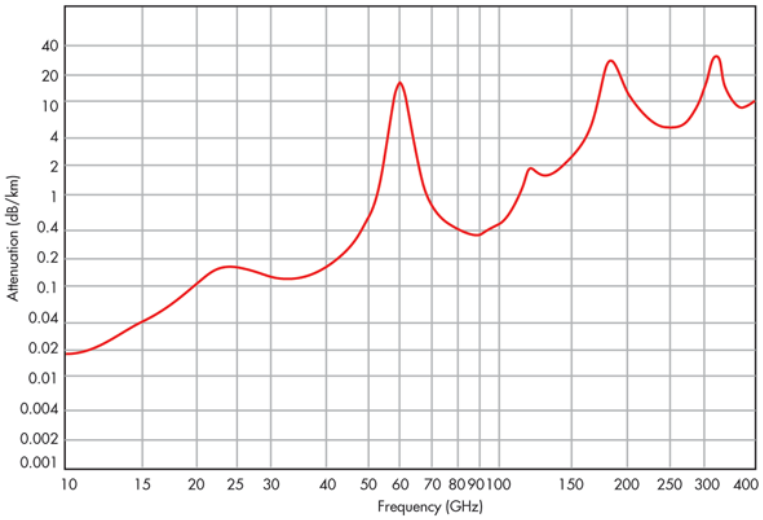


Figure 1.1: Attenuation of mmWave frequencies in the atmosphere [1]. Peaks reflect the increased absorption by atmospheric gases which add to the effects of Friis' law.

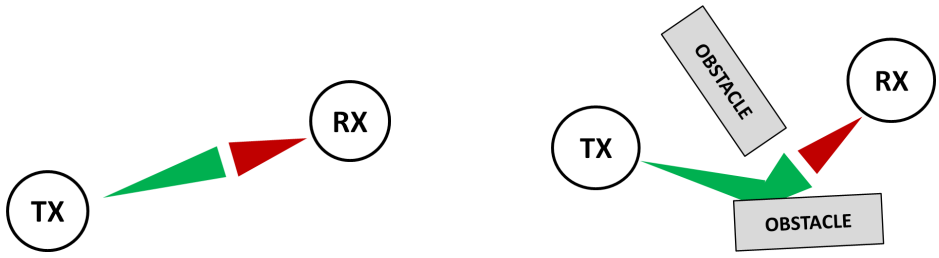
(λ):

$$P_{rx} = P_{tx}\frac{G_{tx}G_{rx}\lambda^2}{(4\pi d)^2} \tag{1.1}$$

When the wavelength is reduced from centimeters (ie. 2.4 GHz and 5 GHz carriers) to millimeters, the free space path loss increases by roughly two orders of magnitude. At select frequencies, such as 60GHz, mmWave also suffer from increased absorption by the atmosphere as shown in Figure 1.1. Such high attenuation would reduce the operating range of mmWave systems and render them impractical for many applications.

Beamforming has been proposed to mitigate the effects of path loss by increasing the antenna gains [2] at both the transmitter and receiver (G_{tx} and G_{rx} in Equation (1.1)). Beamforming utilizes multiple antenna elements to change the collective directivity of a device's radiation pattern, enabling increased antenna gain in particular spatial directions. At mmWave frequencies, arrays comprising of up to hundreds of antenna elements are

[2] Antenna gain is the ratio of the E-field amplitude relative to that of an isotropic antenna.



(a) (b)

Figure 1.2: Example of beam alignments: Line-of-Sight **(a)** and Non-Line-of-Sight **(b)**

feasible. This allows the formation of *narrow spot beams*. The antenna gain is high within the beam and falls off dramatically outside of the beam. The maximum achievable beamforming gain is proportional to the number of antennas used [6]. As a result, beamforming mmWave systems can achieve more than 20 dB improvements to the link budget, yielding an effective range competitive to that of Wi-Fi.

1.2 Beam Training Problem

While beamforming addresses the attenuation problem, it also introduces a different problem of beam alignment. To take full advantage of their beamforming gain, two devices must point their respective transmit and receive beams in the "optimal" direction. In free space, the optimal direction is simply the line-of-sight (LOS) path between the two devices. However, since mmWave carriers tend to reflect off of solid objects, it is often the case that the optimal beam alignment is not along the LOS direction (Figure 1.2).

Beam training is the process of discovering the best beam configuration from a state of little or no information about the channel. It is distinguished from **beam tracking**, which refers to the process of *maintaining* beam alignment in mobile situations. Beam training is usually conducted in the association phase, when devices are first establishing contact. In mobile 5G environments, periods of contact may be intermittent and brief. Thus, it is

important that beam training is conducted efficiently to enable data transfer before devices move out of range.

Fundamentally, beam training is a channel state estimation problem, where the spatial dependence of the channel is the attribute being estimated. Training sequences are used to probe the channel using beams pointed in different directions and the channel quality is estimated at the receiver. More time spent beam training yields better aligned beams and results in more efficient use of the channel. However, beam training also contributes to the link overhead, which reduces the amount of time available for data transfer. This creates a fundamental trade-off between the achieved beamforming gain and the beam training time. In this work, the **Beam Training Optimization Problem** refers to the design of a beam training strategy which maximizes beam training efficiency in the context of this trade-off. The expected time spent beam training is to be minimized while achieving a predetermined amount of beamforming gain. Solving the beam training optimization problem is crucial for evaluating the potential and limits of 5G mmWave networks.

1.3 Thesis Outline

In this thesis, we will introduce a novel Boolean model for tackling the algorithmic complexity of beam training. Within this model, a bounding relationship will be derived which describes the complexity of a beam training algorithm that is optimal with respect to time cost and hardware capabilities. We will show, via simulation, that such optimal algorithms can achieve performance gains over the existing standard in non-ideal environments. These results will open the door for future work in beam training using the proposed framework. The thesis outline is as follows:

1. **Chapter 1: Introduction** - A summary of the beam training problem for mmWave is presented.

2. **Chapter 2: Background** - The models, assumptions, and scope of the beam training problem being considered in this thesis are introduced.
3. **Chapter 3: Ideal Boolean Beam Training** - An ideal Boolean model for beam training is introduced, and the algorithmic optimization of beam training is examined within this context.
4. **Chapter 4: Impact of Non-Ideal Factors** - Some algorithms from Chapter 3 are evaluated through simulation, and the trade-offs induced by non-ideal factors are discussed.
5. **Chapter 5: Conclusion and Future Work** - The contributions of this thesis is summarized and possible directions for future work are presented.

Chapter 2
BACKGROUND

In this chapter, we present the background, assumptions, and models which define the scope of beam training discussed in this work. The terminology and models introduced in this chapter will be referenced in Chapter 3 and Chapter 4.

2.1 Beamforming Fundamentals

Beamforming is a process through which multiple antennas can change their collective radiation pattern by emitting the same signal with structured phase differences introduced between antennas. The phase differences produce interference patterns which become a highly directed radiation profile (i.e. a beam) in the far field. In transmit beamforming, the interference patterns are physically formed by the carrier waves (Figure 2.1). In receive beamforming, the interference patterns are virtual, representing the maximal combination of signal energy received in a particular direction. By changing the phase applied to individual antenna elements, a beam can be steered in different directions [6]. Depending on the beamforming architecture (see Section 2.1.2), the relative amplitudes of the transmitted signal can also be tuned to produce more refined beams with more uniform gain over the extent of the beam and better gain suppression elsewhere.

In mmWave systems, a large number of antenna elements can be placed on a single device due to the smaller footprint of each antenna. When equipped with beamforming capabilities, these antennas are referred to as a phased array. The simplest example of a phased array is a Uniform Linear Array (ULA), which consists of a line of equally-spaced,

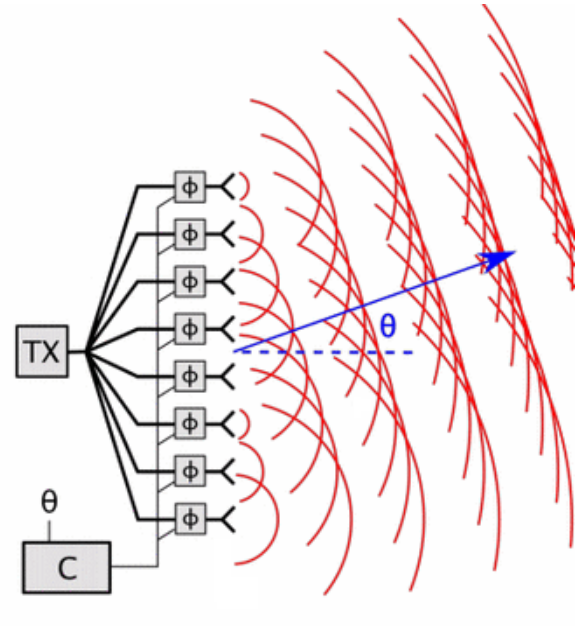


Figure 2.1: Transmitter beamforming using a ULA [2]. Curved lines depict carrier wave fronts. Maximum constructive interference occurs at a departure angle of θ.

isotropic antennas as illustrated in Figure 2.1. The distance between antenna elements influence the interference behavior of the array. A good beam training strategy should be applicable to all phased arrays, regardless if they are ULAs or have a more complicated physical structure. Beamforming should therefore be modeled without specific assumptions regarding array geometry.

2.1.1 Beamforming Mechanics

Consider a phased array of N_t identical isotropic antennas. For convenience, define a Cartesian coordinate system such that one antenna is located at the origin and the remaining antennas are positioned elsewhere on the X-Y plane, with relative separations measured in units of the carrier wavelength. Let $\vec{\mathbf{r}}_k$ to denote the position vector of the kth antenna in this coordinate system and define the 0-th antenna be the antenna located at the origin. Receive

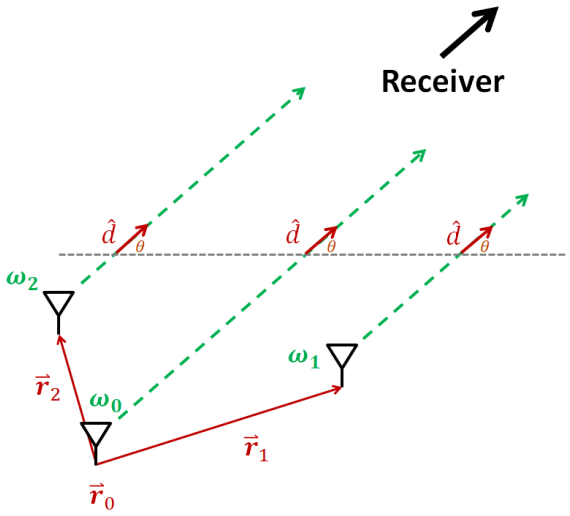


Figure 2.2: Beamforming geometry in a generic 2D array.

and transmit beamforming are identical due to reciprocity. Thus, without loss of generality, only transmit beamforming will be considered. Let $\hat{\mathbf{d}}$ denote the unit vector pointing in a particular direction in the far field. In 2-D space, $\hat{\mathbf{d}}$ is a function of the azimuth angle, θ, and can be expressed as $\hat{\mathbf{d}}(\theta) = [\cos\theta \sin\theta]^\intercal$. Let w_k be a complex number representing the phase and amplitude of the carrier signal transmitted by the k-th antenna relative to that transmitted by the 0th antenna. Figure 2.2 summarizes the geometry of this scenario.

From the linearity of wave propagation in free space, it can be shown that the amplitude and phase observed in the far field is given by:

$$Z(\hat{\mathbf{d}}(\theta)) = \tilde{Z}_0(\hat{\mathbf{d}}(\theta)) \sum_{k=0}^{N_t-1} w_k e^{j2\pi\left(\hat{\mathbf{d}}(\theta)^\intercal \vec{\mathbf{r}}_k\right)} \tag{2.1}$$

where $\tilde{Z}_0(\hat{\mathbf{d}}(\theta))$ is the amplitude and phase observed if only the 0-th antenna was transmitting.

It is sufficient to use $\frac{Z(\hat{\mathbf{d}}(\theta))}{\tilde{Z}_0(\hat{\mathbf{d}}(\theta))}$ to capture the *relative* changes in $Z(\hat{\mathbf{d}}(\theta))$ as a function of angular direction. The ratio $\frac{Z(\hat{\mathbf{d}}(\theta))}{\tilde{Z}_0(\hat{\mathbf{d}}(\theta))}$ is referred to as the **array factor** in antenna theory

and can be compactly expressed in the form of Equation (2.2).

$$AF(\mathbf{w}, \hat{\mathbf{d}}(\theta)) = \mathbf{X}(\theta)\mathbf{w} \tag{2.2}$$

$$\mathbf{X}(\theta) = \left[e^{j2\pi\left(\hat{\mathbf{d}}(\theta)^\intercal \vec{\mathbf{r}}_0\right)}\ e^{j2\pi\left(\hat{\mathbf{d}}(\theta)^\intercal \vec{\mathbf{r}}_1\right)} \cdots e^{j2\pi\left(\hat{\mathbf{d}}(\theta)^\intercal \vec{\mathbf{r}}_{N_t-1}\right)}\right] \tag{2.3}$$

$$\mathbf{w} = [w_0\ w_1 \cdots w_{N_t-1}]^\intercal \tag{2.4}$$

In this work, (2.3) will be referred to as the **Antenna Response Vector (ARV)** and (2.4) the **Antenna Weight Vector (AWV)**. The ARV captures the changes in amplitude and phase induced by the array *geometry* while the AWV captures changes induced by varying the amplitude and phase of the *transmitted signals* over each antenna.

In the context of beamforming, the ARV is pre-determined by hardware so the array factor is fully specified by the AWV. When the AWV is set to unit norm, the magnitude of the array factor corresponds to the far field antenna gain as a function of direction. This is typically referred to as the beam pattern of the array.

> **Definition 1.** The **Beam Pattern** corresponding to an AWV, $\mathbf{w}$, is a function mapping the observation angle (azimuth angle for 2-D space) to the far field antenna gain given by:
>
> $$G_{\mathbf{w}}(\theta) = \left|AF(\mathbf{w}, \hat{\mathbf{d}}(\theta))\right| \tag{2.5}$$
>
> where $AF(\mathbf{w}, \hat{\mathbf{d}}(\theta))$ is the array factor when $\mathbf{w}$ is applied to the array, and $\|\mathbf{w}\| = 1$

A beam pattern can be plotted to illustrate the radiation profile of the antenna array. Figure 2.3 shows an example where a beam has been formed to maximize gain at an azimuth angle of 345°. A beam pattern consists of several lobes. The lobe corresponding to the direction of maximum gain is referred to as the **main lobe** and lobes with much smaller gain are the **side lobes**. A beam pattern may also be designed with multiple main lobes

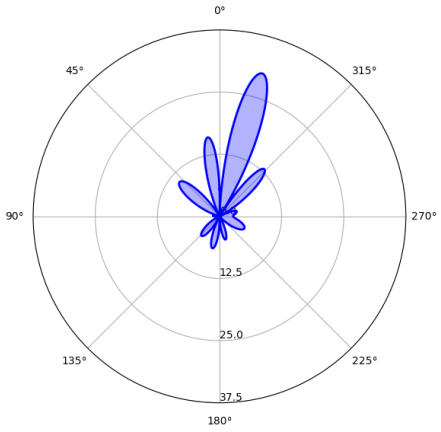


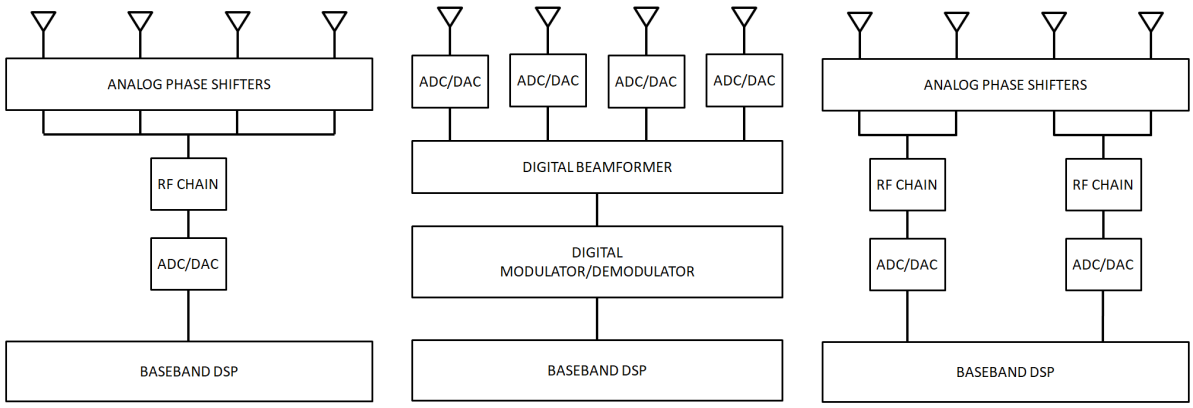
Figure 2.3: Example of a beam pattern.

and side-lobes interspersed between them. In this work, "beams" and "main lobes" will be used interchangeably.

2.1.2 Beamforming Architectures

Although beam training is ultimately concerned with beam patterns, it is important to consider the limitations of beamforming systems since these trade-offs impose constraints to the type of beam patterns that can be achieved. In this section, we will summarize the different architectures used for implementing AWVs in real systems.

In an analog beamforming architecture, RF phase shifters are inserted between the RF chain and the antenna feed (Figure 2.4a). These phase shifters are digitally controlled analog components that operate on raw signals fed to and from each antenna. Low power consumption and hardware simplicity and make analog architectures cost effective solutions [6]. However, the achievable phase shifts are restricted to only a few values and generally do



(a) (b) (c)

Figure 2.4: Beamforming Architectures: **(a)** Analog, **(b)** Digital, and **(c)** Hybrid

not support amplitude scaling. As a result, only a small set of achievable AWVs exist for an analog beamformer, and usually only one well-defined main lobe can be formed. Finally, analog beamforming systems can only transmit one baseband signal at any given time due to the use of a single RF chain.

In a digital beamforming architecture, DAC/ADC conversion happens immediately before/after the antenna (Figure 2.4b). Digital beamformers are true massive MIMO systems in the sense that the signal transmitted by every antenna can be individually tuned. As a result, AWVs are implemented through direct multiplication at baseband. This flexibility allows a digital beamformer to closely approximate any well-behaved beam pattern, including patterns with multiple main lobes in different directions. However, the superior performance of digital beamformers is held back by hardware complexity and increased power consumption [6]. A fully digital architecture is therefore infeasible in many systems due to power or form factor constraints.

The hybrid architecture has emerged as a middle ground between analog and digital. In a hybrid architecture, the full antenna array is partitioned into subarrays, and each subarray is

connected to an independent RF chain which makes use of analog beamforming (Figure 2.4c). In transmit beamforming, digital weights are applied to the signals sent to each subarray in a process known as **baseband precoding**. Likewise, receive beamforming can digitally fine tune how signals from different subarrays are summed in a process referred to as **baseband combining**. Baseband precoding and combining allow hybrid architectures to form different beams with each subarray, enabling spatial multiplexing. Furthermore, multiple subarrays can coordinate to form a single beam with more refined main lobes than analog beamformers.

2.2 Beamforming Model

In this work, we will adopt the codebook approach for modeling beamforming. In the codebook approach, a beam training algorithm defines a set of ideal beam patterns under the assumption that a **beamforming codebook** can be used to implement them. The beamforming codebook is a table of *pre-computed* AWVs which minimizes the mean squared error of achieved beam patterns relative to the desired ideal patterns [7]. This table can be stored in device memory to obviate any real time computation of antenna weights. The selection of AWVs for beamforming codebooks is an active area of research [8, 9]. However, we will assume any beam pattern which adheres to general beamforming constraints can be well-approximated.

The codebook approach has a couple advantages over the dynamic computation of AWVs. Since beam training is largely concerned with the spatial dependence of the channel, beam patterns with similar spatial profiles will produce links with similar qualities. Thus, the set of useful AWVs can be reduced to those with sufficiently different radiation profiles. In the codebook approach, these candidates are computed prior to deployment to reduce the search space that must be considered during training. Furthermore, the codebook approach isolates

the hardware details of beamforming from the algorithmic process of beam training. This allows the development of beam training strategies that are agnostic to array geometries. **For simplicity, we will only consider beam patterns in the 2-D azimuthal space. 2-D beam training strategies can be extended to 3-D space by generalizing the codebook accordingly.**

2.2.1 Discretized Beam Patterns

A 2-D beam pattern is a continuous function of the azimuth angle, θ. It is convenient to discretize the angle space so desired beam patterns can be specified as finite dimensional vectors.

Given some desired resolution R, an orthogonal basis set, $\mathcal{U}_R$, for the azimuth angle space can be defined as:

$$\begin{aligned} \mathcal{U}_R &= \{U_0, U_1, ..., U_{R-1}\} \\ U_k(\theta) &= \begin{cases} 1, & \frac{2\pi k}{R} \leq \theta < \frac{2\pi(k+1)}{R} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{2.6}$$

$\mathcal{U}_R$ can be interpreted as a set of ideal beam patterns which have a constant unity gain over one angular sector and zero gain elsewhere, where U_k is the pattern for the kth sector. These sector patterns are not necessarily be well-approximated in practice, but they serve as a basis for modeling purposes. Using $\mathcal{U}_R$, continuous beam patterns can be discretized into vectors of dimension R through linear projection.

Definition 2. The **Beam Pattern Vector (BPV)**, $\hat{\mathbf{b}}_{\mathbf{w}}$, corresponding to an antenna weight vector, $\mathbf{w}$, is defined as:

$$\begin{aligned} &\hat{\mathbf{b}}_{\mathbf{w}} \in \mathbb{R}^R \\ &\hat{\mathbf{b}}_{\mathbf{w}}[k] = \sqrt{\frac{R}{2\pi}\int_0^{2\pi} \left| X(\hat{\mathbf{d}}(\theta))\mathbf{w} \right|^2 U_k^2(\theta)\mathrm{d}\theta}, \quad k = 0, \ldots, R-1 \end{aligned} \tag{2.7}$$

The k-th element of a BPV is the RMS antenna gain over the k-th sector. So the square of each element corresponds to the the average power gain over the corresponding sector. A BPV approaches the continuous beam pattern in the limit as the discretization resolution, R, tends to infinity. However, since beams have finite width, it is sufficient to choose R large enough to resolve all of the main lobes of the beam patterns being modeled.

At the algorithmic level, it is more practical to model BPVs as binary vectors, where elements with value 1 correspond to sectors with high gain and elements with value 0 correspond to sectors with low gain.

Definition 3. The **Binary Beam Pattern Vector (BBPV)**, $\mathbf{b}$, corresponding to a BPV, $\hat{\mathbf{b}} \in \mathbb{R}^R$, is:

$$\begin{aligned} &\mathbf{b} \in \{0,1\}^R \\ &\mathbf{b}[k] = \begin{cases} 1, & \frac{\hat{\mathbf{b}}[k]}{b_{max}} \geq \frac{1}{\sqrt{2}} \\ 0, & \text{otherwise} \end{cases}, \quad k = 0, ..., R-1 \\ &b_{max} = \max_k \hat{\mathbf{b}}[k] \end{aligned} \tag{2.8}$$

The threshold of $\frac{1}{\sqrt{2}}$ corresponds to the cutoff of -3dB with respect to the main lobe

gain [6]. For well-defined beams, the BBPV representation is equivalent to assuming all sectors covered by a main lobe have unity gain and all other sectors have zero gain.

The quantization introduced in the BBPV removes information regarding the absolute antenna gain. This is not a problem for modeling beamforming at the algorithmic level as only the *relative* gain differences between spatial sectors are relevant. However, BBPVs are *not* sufficient for modeling signal reception over a noisy channel, since there is a clear trade-off between transmit power and channel estimation error.

Figure 2.5 illustrates the relationship between beam patterns, BPVs, and BBPVs for a resolution of $R = 4$. The top row corresponds to the true beam patterns being modeled, the middle row is the sectorized approximation represented by the BPV, and the bottom row is the quantized and sectorized approximation represented by the BBPV. Amplitudes have been intentionally normalized to highlight the differences in beam pattern shape. In this example, the four BBPVs correspond exactly to the basis beam patterns in $\mathcal{U}_4$. They are well approximated by the actual beam patterns whose main lobes are predominantly confined to each of the four spatial sectors.

A convenient property of BBPVs is that they are, to a degree, resolution invariant. A low resolution BBPV can be upsampled to a higher resolution without additional information. For example, the basis beam patterns in $\mathcal{U}_4$ are defined by BBPVs in $R = 4$ as:

$$\mathbf{b}_0 = [1\ 0\ 0\ 0]^\intercal$$

$$\mathbf{b}_1 = [0\ 1\ 0\ 0]^\intercal$$

$$\mathbf{b}_2 = [0\ 0\ 1\ 0]^\intercal$$

$$\mathbf{b}_3 = [0\ 0\ 0\ 1]^\intercal$$

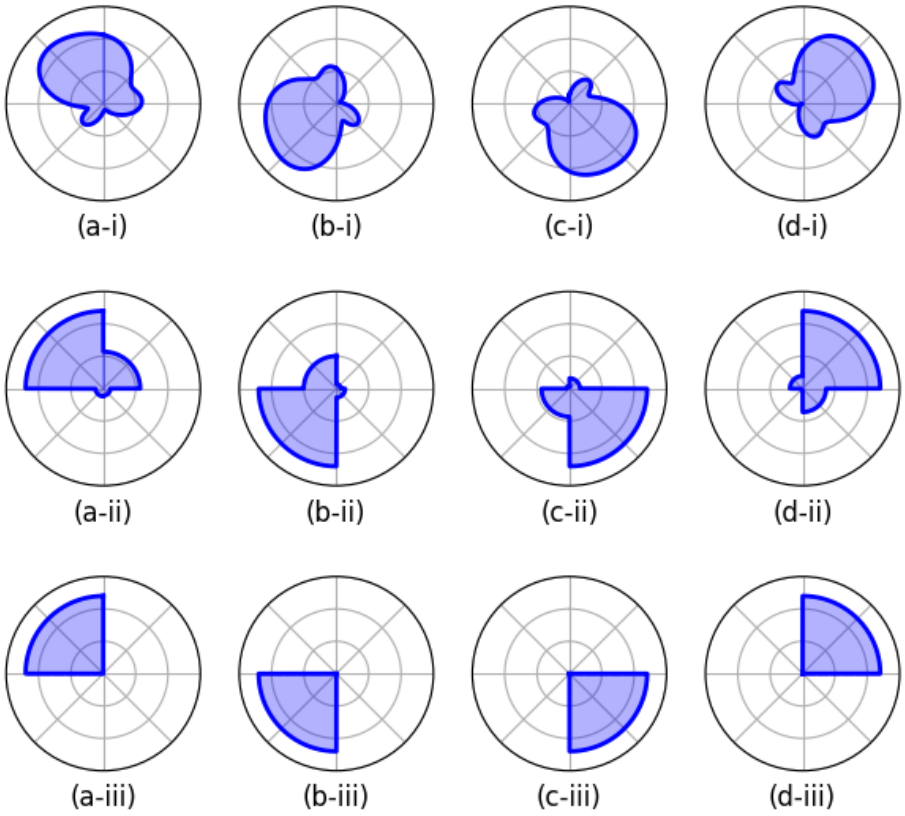


Figure 2.5: Discretization of **(i)** continuous beam patterns into **(ii)** beam pattern vectors and **(iii)** binary beam pattern vectors.

However, they can be equivalently written in $R = 8$ as:

$$\mathbf{b}_4 = [1\ 1\ 0\ 0\ 0\ 0\ 0\ 0]^\intercal$$

$$\mathbf{b}_5 = [0\ 0\ 1\ 1\ 0\ 0\ 0\ 0]^\intercal$$

$$\mathbf{b}_6 = [0\ 0\ 0\ 0\ 1\ 1\ 0\ 0]^\intercal$$

$$\mathbf{b}_7 = [0\ 0\ 0\ 0\ 0\ 0\ 1\ 1]^\intercal$$

The underlying ideal beam pattern represented by $\mathbf{b}_k$ and $\mathbf{b}_{k+4}$ are identical. Due to this property, low resolution BBPVs can be used as shorthands for describing certain BBPVs in a higher resolution space. This fluidity with respect to the discretization resolution will be used extensively in Chapter 3.

In this work, the use of finite dimensional vectors to represent spatial beam patterns will be referred to as the **sector model**. In Chapter 3, binary BBPVs will be used to model beam training at an algorithmic level. In Chapter 4, more general BPVs will be used to investigate the impact of non-ideal factors on beam training performance. $\hat{\mathbf{b}}$ will be used to denote general BPVs while $\mathbf{b}$ will be used to denote BBPVs. This choice of notation will be explained in Section 2.2.2.

2.2.2 Achievable Beam Patterns

The sector model provides a simple method for describing desired beam patterns. Namely, every achievable beam pattern can be mapped to a BBPV in $\{0,1\}^{R_{cb}}$ for some aptly chosen codebook resolution, R_{cb}. However, it is not the case that all vectors in $\{0,1\}^{R_{cb}}$ represent achievable beam patterns. The codebook should include only a subset, $\mathcal{B} \subset \{0,1\}^{R_{cb}}$, which correspond to BBPVs that can be implemented. Restrictions imposed on the elements of $\mathcal{B}$ must reflect the physical beamforming constraints that apply to phased arrays.

Beamforming Constraints

1. **Constraint 1** is on the number of beams that can be formed simultaneously. Due to the limitations of analog phase shifters, we assume analog beamforming arrays can only produce single beams. In other words, the number of simultaneous beams that can be formed is upper bounded by the number of digital chains (ie. subarrays) that are available.

2. **Constraint 2** is on beam width. When the number of antenna elements is large, the width of a main lobe scales inversely with the number of antenna elements forming the beam [6]. This can be explained intuitively by considering Equation (2.1) as the spatial equivalent of the Discrete Time Fourier Transform. The antenna weights form the sampled "time" signal and the "frequency" domain is roughly the cosine of the observation angle. Increasing the number of antennas is equivalent to increasing the number of taps. Since the widening of a signal in time results in the proportional narrowing of the signal in frequency, this implies the signal in the spatial domain (ie. beam pattern) becomes proportionally more narrow.

   In combination with Constraint 1, this induces a proportional relationship between the the number of distinct beams formed and the minimum width of each beam. To understand this relationship, consider the case where a device can form N beams simultaneously, each of minimum width W. By Constraint 1, this implies the device has N subarrays. Since the minimum beam width scales inversely with the number of antennas, each subarray must have $M \approx \frac{2\pi}{W}$ antenna elements. Now suppose instead of forming N separate beams, all the subarrays are used to form a single beam. Since a total of MN antennas are used to form the beam, the minimum beam width, W' must satisfy $MN \approx \frac{2\pi}{W'}$. Dividing the minimum beam width for the two cases yields $\frac{W}{W'} \approx \frac{MN}{M} = N$.

3. **Constraint 3** is imposed by the conservation of power. The average power gain achieved via beamforming should be proportional to the reciprocal of the total angle space covered by the beams. For example, a beam pattern with main lobe that spans a quarter of the entire angle space cannot have a power gain that exceeds four times that of an omnidirectional beam pattern. Likewise, a beam pattern with two main

lobes, each spanning an eighth of the angle space, is subject to the same upper bound. This is because beamforming does not add power to signals, but rather focuses existing power into particular directions. Therefore the integrated power gain over all angles should be unity.

BBPV Constraints

The constraints on beam count and minimum beam width translate to restrictions to the locations of 1s within valid BBPVs. The restrictions are best explained using the concept of "1-blocks".

Definition 4. A 1-block of a BBPV, $\mathbf{b} \in \{0,1\}^R$, is a set, $\mathcal{S}$, of consecutive indices (modulo R) such that $\mathbf{b}[k] = 1, \forall k \in \mathcal{S}$, and there exists no larger set with this property which completely contains $\mathcal{S}$.

To illustrate this definition, consider the BBPV below (for $R = 8$):

$$[1\,0\,1\,1\,0\,0\,1\,1]^\intercal$$

This BBPV has two 1-blocks. One is of size 2, spanning indices 2 through 3. The other is of size 3, spanning indices 6,7, and 0. The size 3 1-block wraps around the last index of the vector due to the rotational invariance of the angle space.

Each 1-block in a BBPV corresponds to a main lobe of the beam pattern being represented, and the size of the 1-block is proportional to the beam width. The beamforming constraints on beam count and beam width therefore translate to BBPVs as:

1. **The number of 1-blocks cannot exceed the number of digital chains used for beamforming.**

2. **The minimum size of each 1-block scales proportionally with the number of 1-blocks.**
3. **The minimum size of the 1-block in a BBPV with a single 1-block is equal to $\lceil \frac{R}{N_t} \rceil$, where R is the dimension of the BBPV, and N_t is the total number of antennas in the array.**

These constraints are best demonstrated with some examples. Suppose we wish to model beamforming on a device with a phased array of 4 antenna elements driven by analog phase shifters. The codebook resolution is chosen to be $R_{cb} = 12$.
Consider the following vectors:

$$\mathbf{b}_1 = [0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^\intercal$$
$$\mathbf{b}_2 = [1\ 1\ 0\ 0\ 1\ 1\ 0]^\intercal$$

Neither $\mathbf{b}_1$ nor $\mathbf{b}_2$ are valid codebook BBPVs for the following reasons:

- $\mathbf{b}_1$ is invalid because the beam pattern it represents has a main lobe that is too narrow. With a dimension of $R = 12$, the minimum 1-block size should be $\frac{12}{4} = 3$. However, $\mathbf{b}_1$ has a 1-block of size 1. Intuitively, $\mathbf{b}_1$ represents a beam pattern with a main lobe spanning a single sector of width $\frac{360^\circ}{12} = 30^\circ$, which is smaller than the minimum achievable beam width of 90°.
- $\mathbf{b}_2$ is invalid for two reasons. Firstly, the system only has a single baseband digital chain, so producing two well-defined beams is assumed to be impossible. Secondly, both 1-blocks in $\mathbf{b}_2$ are of size 2, but the minimum 1-block size is $\left(\frac{6}{4}\right)(2) = 3$ since there are two beams. The main lobes of the beam pattern represented by $\mathbf{b}_2$ are therefore too narrow.

An example of a valid BBPV for this device is:

$$\mathbf{b}_3 = [0\ 0\ 1]^\intercal$$

Here, the beam pattern being modeled corresponds to a single main lobe with width 90° centered at 300°. The equivalent BBPV stored in the codebook set, $\mathcal{B}$, would be:

$$\mathbf{b}_3 = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1]^\intercal$$

BPV Constraints

Since BBPVs do not contain information regarding the absolute gain of beam patterns, the power constraint on beam patterns is only applicable to BPVs. All valid BPVs must have a valid BBPV representation and unit Euclidean norm. The unit norm implies that a total of one unit power is applied to all antennas in the array and radiated out via the achieved beam pattern. This normalization is reflected by the hat accent used to distinguish BPVs ($\hat{\mathbf{b}}$) from BBPVs ($\mathbf{b}$).

2.3 Channel Model

In Section 2.2 we introduced the sector model which describes beam patterns in a discretized angular basis. In this section, we introduce a similar model which describes mmWave channels in a discretized joint angular basis.

2.3.1 Channel Assumptions

To simplify the channel model, we make two assumptions about the frequency and time dependence of mmWave channels:

1. **Slow Fading** - The channel response is static over the duration of beam training phase.

2. **Flat Fading** - The channel is narrowband and experiences no frequency selective fading (ie. the multipath delay spread is small with respect to the symbol period). This is generally *not* true for mmWave channels. However, the multipath arrival times are uncorrelated with the spatial dependence in mmWave channels [10]. The problem can therefore be addressed separately. For example, 802.11ad uses complementary Golay codes to estimate the full baseband channel impulse response (see Appendix A). The RMS gain of the baseband channel response is equivalent to a single-tap flat fading gain for beam training purposes.

2.3.2 Discretized Spatial Channel Matrix

The classical flat fading MIMO channel with N_{RX} receive and N_{TX} transmit antennas consist of a matrix, $\mathbf{H} \in \mathbb{C}^{N_{RX} \times N_{TX}}$, where each element corresponds to the complex gain associated with the signal propagating between a particular transmit-receive *antenna* pair. Given spatial resolutions of R_{RX} and R_{TX} at the receiver and transmitter, respectively, the goal is to define an equivalent channel matrix, $\tilde{\mathbf{H}} \in \mathbb{C}^{R_{RX} \times R_{TX}}$, such that each element of $\tilde{\mathbf{H}}$ corresponds to the complex gain for propagation paths associated with a particular transmit-receive *sector* pair.

Expressing a channel in the spatial domain is only reasonable if the gain associated with each propagation path is highly correlated with the **Angle of Departure (AOD)** and **Angle of Arrival (AOA)** of the path. This is indeed the case for mmWave channels [3]. Due to the optic-like properties of mmWave carriers, a mmWave channel is largely governed by geometric reflections with scattering concentrated over clusters of objects (Figure 2.6). Propagation paths with correlated AOD and AOA are highly likely to scatter from the same cluster and therefore result in the same path loss.

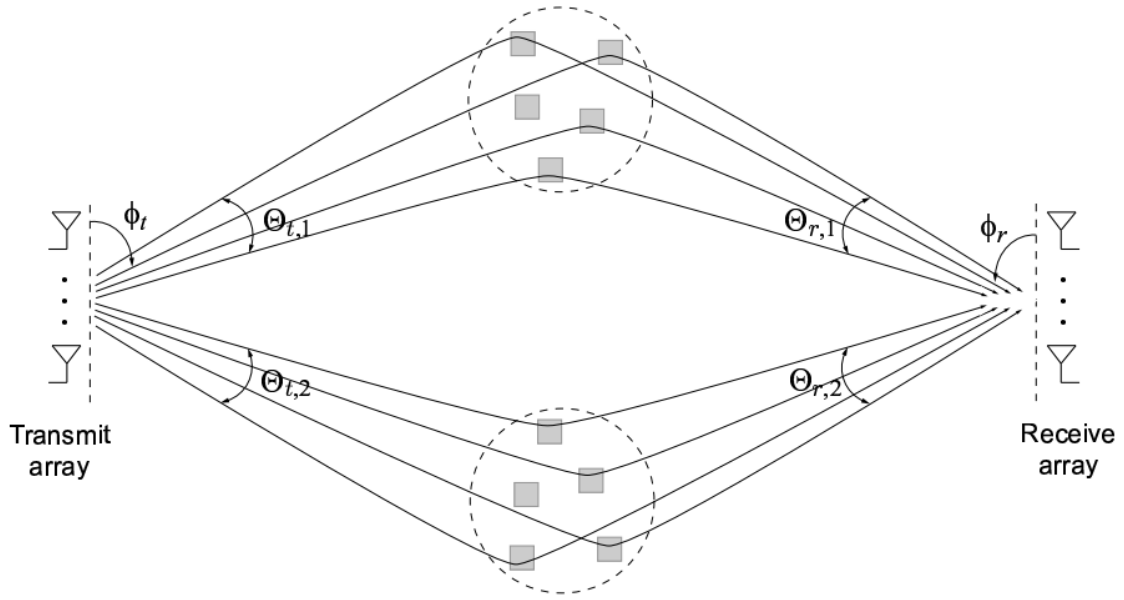


Figure 2.6: mmWave propagation is dominated by geometric reflections with clustered scattering [3].

To formally define a spatial channel matrix, we consider the stochastic MIMO channel matrix proposed by [11]. This model was developed from channel sounding conducted at 28 GHz and 73 GHz bands in New York City. The stochastic narrowband channel matrix is given by:

$$\mathbf{H} = \frac{1}{\sqrt{M}} \sum_{c=1}^{C} \sum_{m=1}^{M} \left(\alpha_{c,m} \mathbf{X}_{RX}^{H}(\theta_{c,m}) \mathbf{X}_{TX}(\phi_{c,m}) \right) \tag{2.9}$$

C is the number of spatial clusters. M is the number of paths in each cluster. $\alpha_{c,m} \sim \mathcal{CN}(0, \alpha_c 10^{-0.1PL})$ is the complex channel gain associated with the mth path in the cth cluster, where α_c is the relative power gain associated with the cth cluster and PL is the omnidirectional path loss in dB. X_{TX} and X_{RX} are the ARVs at the transmitter and receiver, and $\theta_{c,m}$ and $\phi_{c,m}$ are the AOAs and AODs associated with the mth path in the cth cluster. Details regarding the parameters of this channel matrix are provided in Chapter 4.

The channel can be written with normalized ARVs as:

$$\mathbf{H} = \sqrt{\frac{N_{RX}N_{TX}}{M}} \sum_{c=1}^{C} \sum_{m=1}^{M} \left(\alpha_{c,m} \hat{\mathbf{X}}_{RX}^{H}(\theta_{c,m}) \hat{\mathbf{X}}_{TX}(\phi_{c,m}) \right) \tag{2.10}$$

where $\hat{\mathbf{X}}$ denotes an ARV vector normalized to an Euclidean norm of 1. In this form, the outer product, $\hat{\mathbf{H}}(\theta_{c,m}, \phi_{c,m}) = \hat{\mathbf{X}}_{RX}^{H}(\theta_{c,m}) \hat{\mathbf{X}}_{TX}(\phi_{c,m})$, represents a unit basis matrix for the *joint* transmitter and receiver ARVs corresponding to a particular AOA and AOD. To conduct a change of basis, $\hat{\mathbf{H}}$ is converted to an equivalent representation in the joint AOA-AOD angle space.

For well-designed antenna arrays we expect the ARVs corresponding to angles that are far apart to be "roughly orthogonal". This is because arrays with ARVs that are highly correlated over angle space will yield beams with high side lobes due to the linearity of array factor equation (Equation (2.2)). Well-designed phased arrays are expected to have geometries which mitigate this problem. The roughly orthogonal property can be mathematically expressed as:

$$\left| AF(\hat{\mathbf{X}}^{H}(\theta_2), \theta_1) \right| = \left| \mathbf{X}(\theta_1) \hat{\mathbf{X}}^{H}(\theta_2) \right| \approx \begin{cases} \frac{1}{\sqrt{\Theta_{min}}}, & |\theta_1 - \theta_2| < \frac{\Theta_{min}}{2} \\ 0, & \text{otherwise} \end{cases} \tag{2.11}$$

where $\Theta_{min} \approx \frac{2\pi}{N_t}$ is the minimum achievable beam width of the array in radians. Using this assumption, $\hat{\mathbf{H}}(\theta_{c,m}, \phi_{c,m})$ can be approximated in the joint angle space by the function:

$$\hat{\mathbf{H}}_{(c,m)}(\theta, \phi) = \begin{cases} \frac{1}{2\pi}, & |\theta - \theta_{c,m}| < \frac{\pi}{N_{RX}}, |\phi - \phi_{c,m}| < \frac{\pi}{N_{TX}} \\ 0, & \text{otherwise} \end{cases} \tag{2.12}$$

Information regarding the phase components of each ARV is lost in this approximation.

However, since the gain associated with each path ($\alpha_{c,m}$) has a uniform phase distribution, this has no impact on the phase statistics of the channel.

Using $\hat{\mathbf{H}}_{(c,m)}(\theta, \phi)$, the MIMO channel matrix can be explicitly expressed in the joint AOA-AOD basis, and the desired spatial channel matrix is simply the discretized representation of (2.10) in the joint AOA-AOD basis.

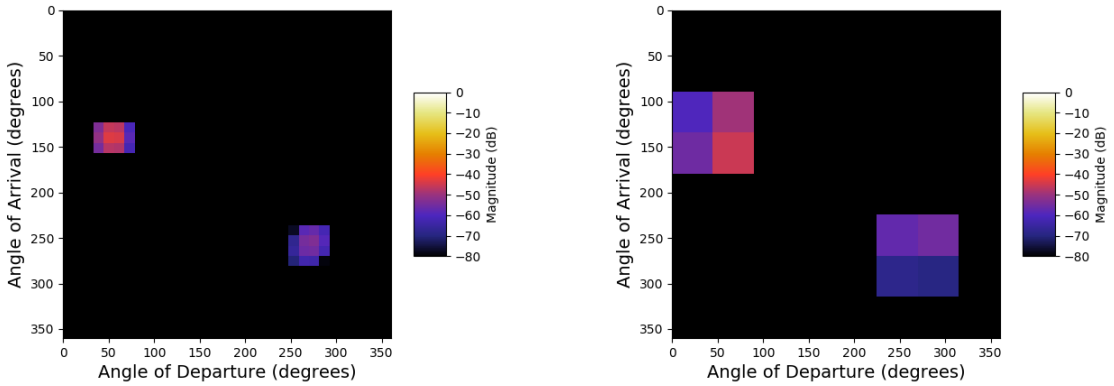
Definition 5. The **Discretized Spatial Channel Matrix (DSCM)**, $\tilde{\mathbf{H}} \in \mathbb{C}^{R_{RX} \times R_{TX}}$ is defined as:

$$\tilde{\mathbf{H}}[k,l] = \frac{\sqrt{R_{RX}R_{TX}}}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \sqrt{\frac{N_{RX}N_{TX}}{M}} \sum_{c=1}^{C} \sum_{m=1}^{M} \Big(\alpha_{c,m} \hat{\mathbf{H}}_{(c,m)}(\theta, \phi) U_k(\theta) U_l(\phi) \Big) \mathrm{d}\phi \mathrm{d}\theta$$

where $U_k(\theta) \in \mathcal{U}_{R_{RX}}$ and $U_l(\phi) \in \mathcal{U}_{R_{TX}}$ are the kth and lth basis beam patterns as defined in Equation (2.6).

Note that the integrand in the DSCM definition is piecewise continuous, allowing the DSCM to be directly computed. Figure 2.7 shows a sample channel with $N_{TX} = N_{RX} = 32$ represented by two DSCMs of resolutions 32×32 and 8×8. The heatmap illustrates the magnitude of the channel gains over different receive and transmit sector pairs. Two bright patches correspond to paths from two distinct clusters which have different AOAs and AODs centered roughly around $(140°, 60°)$ and $(275°, 275°)$. The lower resolution DSCM is a sufficient approximation to the high resolution DSCM when beam patterns with wider main lobes are used.

The visualization in Figure 2.7 illustrates how the spatial dependence of a mmWave channel becomes explicit in the DSCM. Correlations in channel multipaths manifest as proximity within the channel matrix and offer an intuitive perspective to the beam training problem. Beamforming at the transmitter and receiver allocates a certain proportion of



(a) (b)

Figure 2.7: DSCMs corresponding to the same channel with a discretization resolution of **(a)** $R_{RX} = R_{TX} = 32$ and **(b)** $R_{RX} = R_{TX} = 8$. Distinct spatial clusters are clearly visible in both cases.

power to each AOA-AOD sector pair in the DSCM. Each element of the DSCM scales the power allocated to the associated sector pair and the goal is to maximize the combined power which passes through the channel.

2.3.3 Idealized Channel

Since beam training is only concerned with the relative magnitudes of the elements in the DSCM, it is convenient to idealized DSCM as a binary matrix. An element of value 1 would indicate an AOA-AOD sector pair with high gain and an element with value 0 would indicate an AOA-AOD sector pair with low gain.

Definition 6. The **Binary Channel Matrix (BCM)**, $\mathcal{H}$, corresponding to a DSCM, $\tilde{\mathbf{H}} \in \mathbb{C}^{R_{RX} \times R_{TX}}$, is:

$$\mathcal{H} \in \{0,1\}^{R_{RX} \times R_{TX}}$$
$$\mathcal{H}[k,l] = \begin{cases} 1, & \frac{\left|\tilde{\mathbf{H}}[k,l]\right|}{h_{max}} \geq \frac{2}{\sqrt{R_{RX} R_{TX}}} \\ 0, & \text{otherwise} \end{cases}, \quad k = 0, ..., R_{RX} - 1, l = 0, ..., R_{TX} - 1 \tag{2.13}$$
$$h_{max} = \max_{k,l} \left|\tilde{\mathbf{H}}[k,l]\right|$$

This idealization can be interpreted as the channel equivalent of the BBPV. The ideal channel imposes unity gain on signals transmitted and received over certain sectors and infinite suppression on signals corresponding to all other sector combinations. The threshold is chosen so that sector pairs with value 0 contribute to less than half the total power gain when beam patterns are chosen to maximize the power transmitted over the sector pairs with value 1.

The BCM is a *highly simplistic* representation of a mmWave channel, but it captures the key sparsity properties which can be exploited in beam training. For example, the BCM associated with the DSCM given in Figure 2.7b would be:

$$\begin{bmatrix} 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \end{bmatrix}$$

The location of the single 1 marks the AOA-AOD sector pair centered at $(157.5°, 67.5°)$, which is a good estimate of the mean AOA and AOD angles for paths associated with the stronger of the two clusters.

In Chapter 3, the ideal BCM model will be used for beam training algorithm development. The non-ideal DSCM will be used in Chapter 4 to evaluate beam training performance in realistic scenarios.

2.4 System Model

In this work, focus will be placed on beam training conducted between a single pair of devices. The pairwise assumption simplifies analysis, but the restrictions imposed on the beam training format will allow generalization to parallel training (see Section 2.7.2). The **beam training initiator** is the device that is seeking to transmit data, and will be referred to as TX. The **beam training reciprocator** is the device that receives data from the initiator, and will be denoted as RX.

Identical hybrid phased arrays are assumed for TX and RX. Each array consists of N_t total antennas, where N_t is on the order of 16 to 256. The N_t elements are divided into N_c subarrays, each containing N_s elements (Figure 2.8). Each antenna element has an adjustable phase shifter with discrete phase increments and no analog amplification. From the discussion in Section 2.2.2, this implies each device can form beam patterns with up to N_c distinct main lobes, subject to the appropriate beam width constraints. N_c is assumed to be at most on the order of N_s, which means $N_c \leq \sqrt{N_t}$. This is a realistic assumption since $N_c = \sqrt{N_t}$ enables beamforming performance that is nearly comparable to that of a fully digital architecture [7].

Due to the symmetric hardware assumption, we will also assume the beamforming codebooks used by RX and TX are identical. From the sector model presented in Section 2.2, this means the set of beam patterns achievable by each device correspond to BBPVs which reside in some set, $\mathcal{B} \subset \{0,1\}^{R_{cb}}$. However, since R_{cb} is an abstract parameter that can be adjusted at the software level, algorithmic discussion will reference BBPVs using a generic resolution,

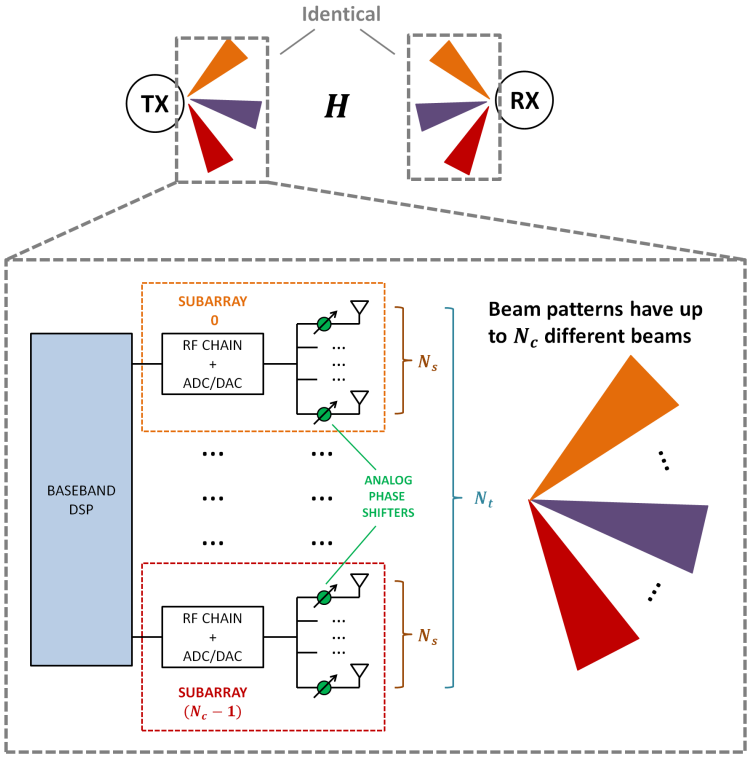


Figure 2.8: Hybrid beamforming hardware architecture assumed for both TX and RX.

R. This is because R_{cb} can be made large, and any R sufficiently smaller than R_{cb} can be upsampled as needed to the equivalent representation in $\{0,1\}^{R_{cb}}$.

2.5 Link Model

Modeling a beamformed communication link is straightforward with discretized beam patterns. Given $\hat{\mathbf{b}}_{TX}$ and $\hat{\mathbf{b}}_{RX}$ are the BPV used by TX and RX used to communicate, the spatial dependence of the channel collapses into a scalar **beamformed channel response** given by:

$$\begin{aligned} h_{(\hat{\mathbf{b}}_{RX},\hat{\mathbf{b}}_{TX})} &= \hat{\mathbf{b}}_{RX}^{\intercal}\tilde{\mathbf{H}}\hat{\mathbf{b}}_{TX} \\ &= \sum_{l=0}^{R}\sum_{k=0}^{R}\tilde{\mathbf{H}}[k,l]\hat{\mathbf{b}}_{RX}[k]\hat{\mathbf{b}}_{TX}[l] \end{aligned} \tag{2.14}$$

where $\tilde{\mathbf{H}}$ is the DSCM.

The beamformed channel response can be interpreted as a weighted sum of the elements of the DSCM, where the weights are given by the elements of the outer product, $\hat{\mathbf{b}}_{RX}\hat{\mathbf{b}}_{TX}^{\intercal}$. The magnitude of the beamformed channel response will be referred to as the **beamformed channel gain** and the squared magnitude of the beamformed channel response will be referred to as the **beamformed channel power gain**. The channel power gain is directly related to the achievable SNR over the beamformed link.

Using the beamformed channel response, a SISO link can be represented as:

$$\mathbf{y} = \sqrt{P_t}h_{(\hat{\mathbf{b}}_{RX},\hat{\mathbf{b}}_{TX})}\mathbf{x} + \mathbf{N_{ch}} \tag{2.15}$$

where $\mathbf{x} \in \mathbb{C}^L$ is a vector of complex baseband symbols that are sent by the transmitter, $\mathbf{y} \in \mathbb{C}^L$ is a vector of complex baseband symbols obtained at the receiver, P_t is the average transmit power, and $\mathbf{N_{ch}} \in \mathbb{C}^L$ is a vector of circularly Gaussian noise introduced by the channel.

Due to the linearity of the beamforming process, a multiple input channel consists of a

set of symbol streams sent in parallel via different transmit BPVs. In the multiple output scenario, different BPVs are also used by the receiver to yield separate received streams.

Suppose the transmitter sends K separate symbol streams, $\{\mathbf{x}_k\}_{k=0}^{K-1}$ and the receiver obtains M separate symbol streams, $\{\mathbf{y}_m\}_{m=0}^{M-1}$. Let $\hat{\mathbf{b}}_{TX,k}$ be the BPV used to transmit the kth stream and $\hat{\mathbf{b}}_{RX,m}$ be the BPV used to receive the mth stream. Then the transmission over the MIMO channel is modeled as:

$$\begin{aligned}
&\mathbf{Y} = \sqrt{\mathbf{P_t}}\bar{\mathbf{H}}\mathbf{X} + \mathbf{N_{ch}} \\
&\mathbf{X} = [\mathbf{x}_0 \cdots \mathbf{x}_{M-1}]^\intercal \\
&\mathbf{Y} = [\mathbf{y}_0 \cdots \mathbf{y}_{M-1}]^\intercal \\
&\bar{\mathbf{H}} \in \mathbb{C}^{M\times K} = \begin{bmatrix} h_{(\hat{\mathbf{b}}_{RX,0},\hat{\mathbf{b}}_{TX,0})} & \cdots & h_{(\hat{\mathbf{b}}_{RX,0},\hat{\mathbf{b}}_{TX,K-1})} \\ \vdots & \vdots & \vdots \\ h_{(\hat{\mathbf{b}}_{RX,M-1},\hat{\mathbf{b}}_{TX,0})} & \cdots & h_{(\hat{\mathbf{b}}_{RX,M-1},\hat{\mathbf{b}}_{TX,K-1})} \end{bmatrix} \\
&\sqrt{\mathbf{P_t}} = \text{diag}\left(\left[\sqrt{P_{t,0}} \cdots \sqrt{P_{t,K-1}}\right]\right)
\end{aligned} \tag{2.16}$$

where $\mathbf{N_{ch}} \in \mathbb{C}^{M\times L}$ is the circularly Gaussian noise from the channel, and $P_{t,k}$ is the average transmit power over the beam pattern, $\hat{\mathbf{b}}_{TX,k}$.

2.5.1 Channel Estimation

During beam training, **Channel State Information (CSI)** must be gathered by measuring the channel quality using specific beam patterns. Complete or partial CSI can be used to determine better beam patterns for communication. The gathering of CSI will be referred to in terms of channel probes.

Definition 7. A **Channel Probe** is the measurement of a beamformed channel power gain corresponding to a distinct beam pattern combination. The channel probe measurement corresponding to a pair of transmit and receive patterns, $(\hat{\mathbf{b}}_{RX}, \hat{\mathbf{b}}_{TX})$, is given by:

$$\gamma = \left| \sqrt{P_t} h_{(\hat{\mathbf{b}}_{RX}, \hat{\mathbf{b}}_{TX})} + \mathbf{N_{ms}} \right|^2 \tag{2.17}$$

where $h_{(\hat{\mathbf{b}}_{RX}, \hat{\mathbf{b}}_{TX})}$ is the beamformed channel response of interest, P_t is the average power used to probe the channel, and $\mathbf{N_{ms}}$ is noise of the measurement.

Channel probes are conducted through the transmission of training sequences over the channel. Let $\mathbf{A} \in \mathbb{C}^{K \times L}$ be a matrix representing a set of K orthogonal training sequences, each of norm $\sqrt{L}$. It is assumed that $\mathbf{A}$ is known by both TX and RX. Suppose TX transmits $\mathbf{A}$ to RX and RX listens with M streams. Using the same notation as in Equation (2.16), the signal received is:

$$\mathbf{Y} = \sqrt{\mathbf{P_t}} \bar{\mathbf{H}} \mathbf{A} + \mathbf{N_{ch}}$$

The m-th row of $\mathbf{Y}$ can be expressed as:

$$\mathbf{y}_m = \sum_{k=0}^{K-1} \left(\sqrt{P_{t,k}} \bar{\mathbf{H}}[m,k] \mathbf{a}_k \right) + \mathbf{N_{ch,m}}$$

where $\mathbf{a}_k \in \mathbb{C}^{1 \times L}$ is the kth row of $\mathbf{A}$ and $\mathbf{N_{ch,m}} \in \mathbb{C}^{1 \times L}$ is the mth row of the noise matrix, $\mathbf{N_{ch}}$.

To obtain channel estimates, RX computes $\mathbf{\Pi} = \frac{1}{L} \mathbf{Y} \mathbf{A}^H$. This projects each row of $\mathbf{Y}$ onto the dictionary of training sequences. From the orthogonality of training sequences we

have:

$$\forall k = 0, \ldots, K-1; m = 0, \ldots, M-1 :$$

$$\begin{aligned} \mathbf{\Pi}[m,k] &= \frac{1}{L}\mathbf{y}_m \mathbf{a}_k^H \\ &= \frac{1}{L}\sum_{l=0}^{K-1}\left(\sqrt{P_{t,k}}\bar{\mathbf{H}}[m,l]\mathbf{a}_l\mathbf{a}_k^H\right) + \frac{1}{L}\mathbf{N}_{\mathbf{ch,m}}\mathbf{a}_k^H \\ &= \frac{1}{L}\left(\sqrt{P_{t,k}}\bar{\mathbf{H}}[m,k]L\right) + \bar{\mathbf{N}}_{\mathbf{ch,m}} \\ &= \sqrt{P_{t,k}}h_{(\hat{\mathbf{b}}_{RX,m},\hat{\mathbf{b}}_{TX,k})} + \bar{\mathbf{N}}_{\mathbf{ch,m}} \end{aligned} \quad (2.18)$$

where $\bar{\mathbf{N}}_{\mathbf{ch,m}} = \frac{1}{L}\mathbf{N}_{\mathbf{ch,m}}\mathbf{a}_k^H$ is circularly Gaussian noise. Thus, $|\mathbf{\Pi}[m,k]|^2$ is the measurement for a channel probe corresponding to the beam patterns $\hat{\mathbf{b}}_{RX,m}$ and $\hat{\mathbf{b}}_{TX,k}$.

While the above assumes a slow, flat fading channel, the process can be generalized to frequency selective fading by substituting complementary Golay codes for every symbol in each training sequence (see Appendix A). In that case, each channel probe produces a vector of complex channel taps instead of a single complex scalar. The effective beamformed channel power gain would then be the average power over all the taps.

MIMO channel estimation allows multiple channel probes to be conducted in parallel, potentially reducing training time. However, the information provided by each probe can be equivalently gathered through sequential SISO. For this reason time cost of beam training will be evaluated only in terms of channel probes. The viability of MIMO parallelization can investigated separately as a general method for reducing training time.

2.6 Beam Training as a Search Problem

Under the sectored channel model, the goal of beam training is to estimate the locations of the elements of largest magnitude within the DSCM. This would allow RX and TX to determine the AOA and AOD sectors which correspond to paths of high gain, and optimal beam

patterns can be selected to maximize the channel power gain. Beam training is therefore equivalent to a search problem over the pairs of AOA-AOD sectors.

Problem 1. K-Best Beam Search Problem (K-BBS)

Given a DSCM, $\tilde{\mathbf{H}} \in \mathbb{C}^{R \times R}$, determine the set of indices $\mathcal{V} = \{(k_m, l_m)\}_{m=1}^{L}$ such that:

$$\left|\tilde{\mathbf{H}}[k_m, l_m]\right| \geq \left|\tilde{\mathbf{H}}[k, l]\right|, \quad \forall (k, l) \neq (k_j, l_j); j = 1, \ldots, m$$

if no initial information regarding the elements of $\tilde{\mathbf{H}}$ are provided.

K-BBS can only be solved by measuring the elements of the DSCM through channel probes. The probing process is associated with two different costs:

- Time - Each probe requires one or more frames to be sent over the channel. Furthermore, feedback frames are needed for RX to report CSI to TX for transmitter beam training.
- Power - Each probe requires TX and RX to allocate some amount of power to each spatial sector in the transmission and reception of the training sequence.

Since power constraints are largely dictated by hardware, we will assume the total power used by each channel probe is the same, and can be set to unity without loss of generality. Thus, beam training optimization is equivalent to solving K-BBS in the fewest number of channel probes.

Note that the DSCM is sparse due to spatial clustering. Multipaths reflect off different clusters which are spread randomly in space. Since it is unlikely that paths from well-separated clusters would experience comparable path loss, there is often a "dominant" cluster which has significantly lower path loss than the others. For example, in a LOS scenario the dominant cluster would simply be the LOS path. As a result, beam training is largely concerned with solving 1-BBS, which is equivalent to finding the AOA-AOD sector pair

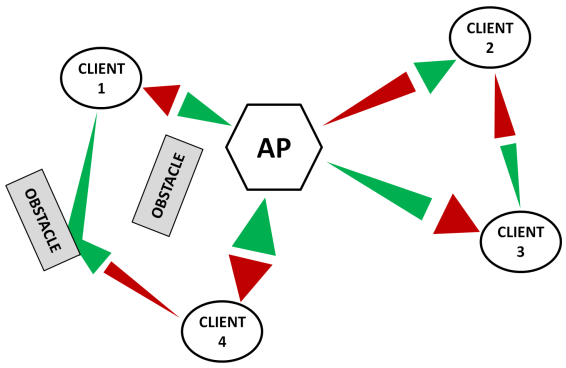


Figure 2.9: DMG operation in 802.11ad.

corresponding to the dominant cluster. This will be referred to as the **Dominant Cluster Assumption**.

2.7 Beam Training in 802.11ad

Before considering the optimization of beam training, we will establish the key constraints at the MAC level which limit how training can be conducted. IEEE 802.11ad is the current flagship standard for mmWave communication and will be used as the reference in this paper.

In 802.11ad, beam training is defined only for wireless networks operating in **Directed Multi-Gigabit Mode (DMG)**. Medium access in the DMG is summarized as follows [12]:

- Medium access is regulated by a central base station or **access point (AP)**.
- All nodes must beam train with the AP to gain access to the network.
- Two client nodes within the network must beam train prior to the transmission of application data between them.

DMG supports various methods for medium control. Different modes of medium access are scheduled over a periodic time interval known as a beacon interval (Figure 2.10). Each beacon interval begins with a **Beacon Header Interval (BHI)** during which the AP

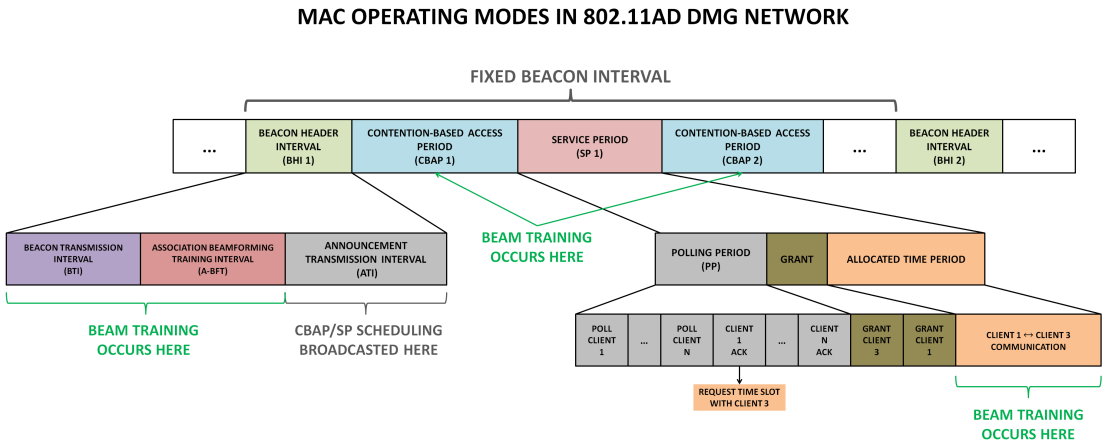


Figure 2.10: Different periods of medium acccess in 802.11ad DMG.

advertises the network, associates with new clients, and broadcasts the scheduling information for the remainder of the beacon interval. During the **Beacon Transmission Interval (BTI)** of the BHI, the AP broadcasts beacon frames which can be used for coarse beam training (see Section 2.7.1). During the **Association Beamforming Training Interval (A-BTI)**, nodes wishing to join the network contend for time slots to finish coarse beam training with the AP. The **Announcement Transmission Interval (ATI)** is used to broadcast scheduling and network management information to all the clients in the network.

After every BHI comes a series of time periods scheduled by the AP, each corresponding to a specific mode of medium access. In a **Contention-Based Access Period (CBAP)**, nodes compete for channel access via CSMA/CA. Minimal beam training is expected in CBAPs since channel access is sporadic, making beam training overhead extremely costly. In a **Service Period (SP)**, medium access is scheduled by the AP through polling. This access method is more ideal for beam training since the AP can reserve the medium for two nodes that wish to beam train, and interference from other nodes within the network is minimized as a result.

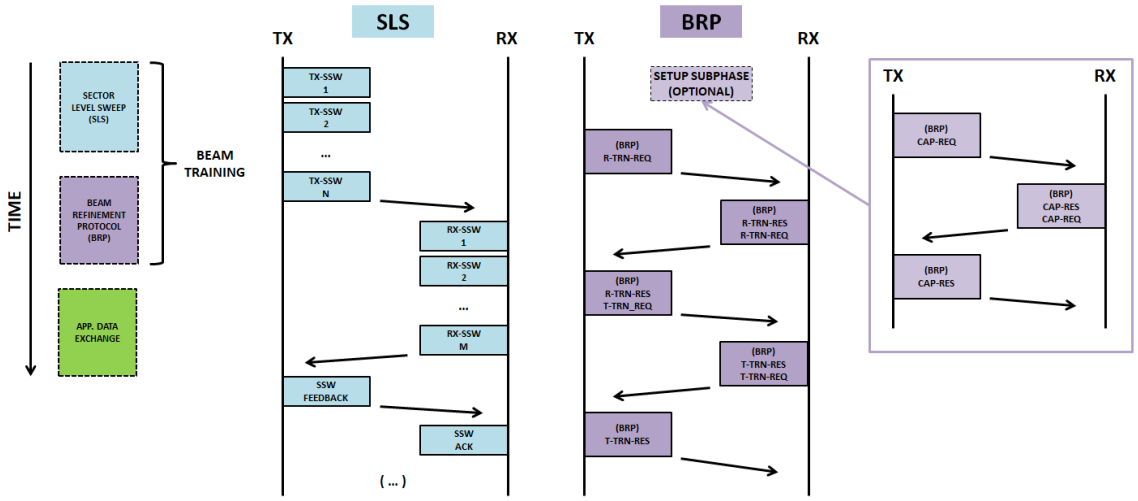


Figure 2.11: Beam training stages in 802.11ad.

Beam training, itself, is divided into two stages as shown in Figure 2.11. The **Sector Level Sweep (SLS)** stage is used to establish coarse beams that can be used to transmit control frames at relatively low rates, while the **Beam Refinement Protocol (BRP)** is used to further refine beams to establish a high-throughput communication link for application data [12].

SLS serves as a neighbor discovery and AP association mechanism in 802.11ad [12]. In SLS, separate frames are used to probe the channel and no prior exchange of overhead information is required. BRP can only be conducted after a control link has been established via SLS. This is because BRP requires the exchange of meta-information prior to any beam training. These overhead frames can be bundled with the SLS or conducted separately in a sub-phase immediately before BRP. In-frame training is used in BRP, where beam patterns are permitted to change within the transmission of a single frame. As a result, far fewer frames need to be transmitted to conduct the same number of channel probes. Coupled with the fact that some CSI has already been obtained by SLS, this makes BRP highly efficient.

In 5G mobile networks, nodes are expected to join and leave networks frequently. As a result, discovery and association training would make up the bulk of beam training overhead at the MAC layer. *For this reason and because BRP is much more efficient, we will only consider beam training in the SLS context.* It is assumed that the novel strategies developed for SLS can also be adapted for the less restrictive BRP context.

2.7.1 SLS Training

SLS treats transmitter and receiver beam training separately. Transmit training always happens first, and receive training may be optionally conducted afterwards. In both transmit and receive training, the strategy adopted is an exhaustive Sector Sweep (SSW) [12].

In transmit SSW, TX systematically points a beam over each of R spatial sectors. A separate training frame is sent over each sector. Beam training is a cross layer process. Channel estimation sequences are embedded in the PHY frame while a unique ID in the MAC frame identifies the sector over which the frame was transmitted. RX is listening omnidirectionally for the SSW frames and performs channel measurements at the PHY layer on every detected frame. The gathered CSI is passed to the MAC layer where the ID and estimated SNR corresponding to the sector with the best estimated channel is stored by RX. This information is sent as feedback in the MAC frame of a transmit SSW conducted by RX. TX adopts an omnidirectional pattern to receive the feedback frames. Once a feedback frame is received, TX uses the ID to identify the optimal sector for transmitting to RX.

In addition to processing the feedback, TX conducts measurements for RX's transmitter training. After all measurements have been made, TX provides the feedback for RX's transmit SSW by transmitting a single control frame over the best sector for transmitting to RX. RX listens omnidirectionally for this feedback, and uses it to determine the best sector for transmitting to TX. Figure 2.12 summarizes transmitter training in SLS.

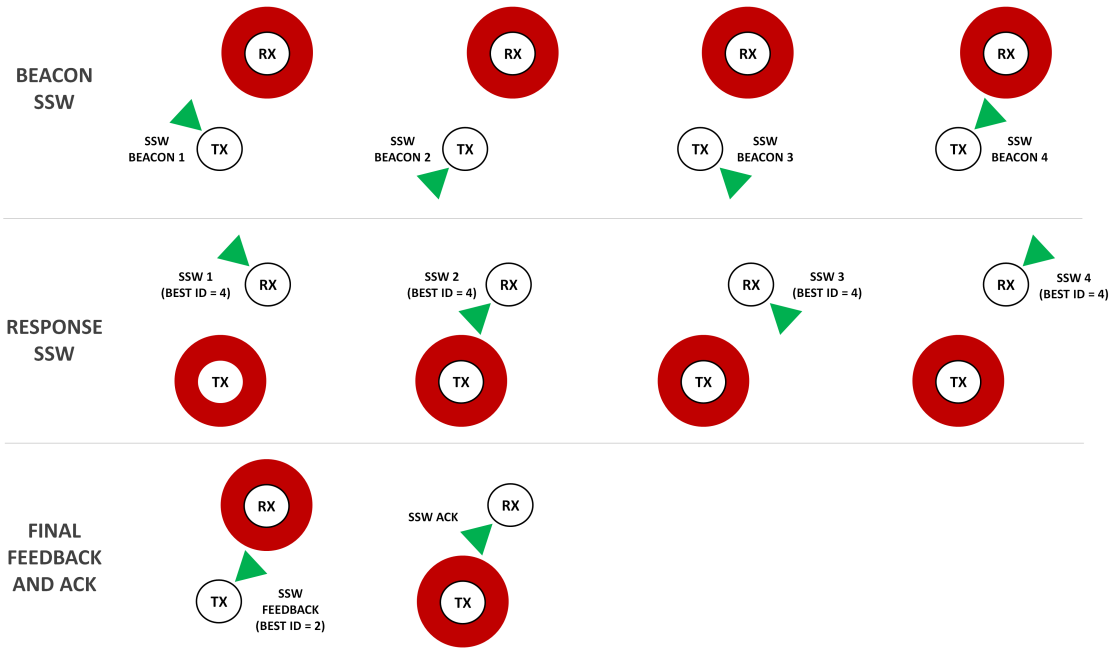


Figure 2.12: Example of SLS transmit training in 802.11ad with $R = 4$. This beam training format allows SLS to double as a neighbor discovery mechanism.

The first set of training frames sent by TX are beacon frames. This is because RX could be listening omnidirectionally by default and is originally unaware that it can associate with TX. The beacon frames therefore notify RX of the presence of TX. Likewise, TX may not be aware of RX's proximity when listening omnidirectionally for feedback. So the SSW feedback sent by RX is also a way to notify TX of the presence of RX. In this manner, SLS serves as a discovery mechanism in 802.11ad.

Receive SSW is similar to transmit SSW, but the processes are simplified since there is no need to send extensive feedback. Instead, TX sends R identical training frames using a beam pointed in a the best known direction while RX performs receive training by systematically scanning over the R receive sectors, one per every frame. The best receive sector can be

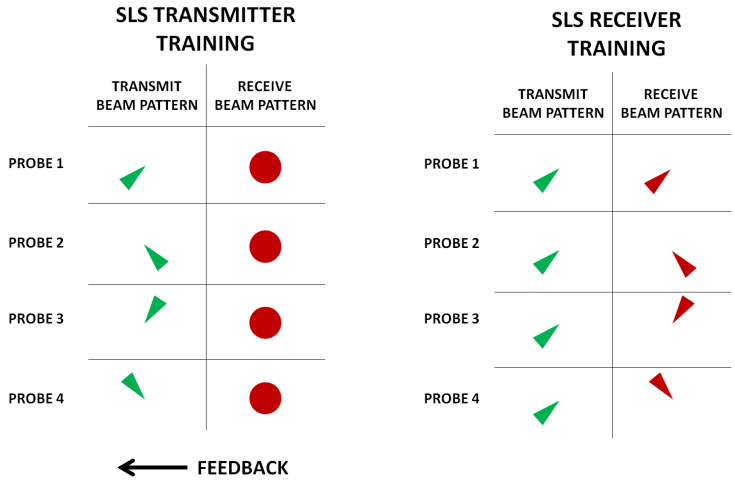


Figure 2.13: Transmit and receive training in SLS follow a symmetric structure.

determined from the channel measurements made for each frame. Figure 2.13 shows the symmetric structure of receive and transmit training in SLS.

At an algorithmic level, the main strategy employed by SLS is the separation of transmit and receive training. Instead of measuring each of the R^2 elements of the DSCM, SLS searches through the R sectors on each side, yielding a just $2R$ channel probes. This approach is adopted out of necessity since there is no guaranteed control link through which TX and RX can coordinate simultaneous transmit and receive training. Each channel probe performed by a transmit SSW is equivalent to measuring the *averaged* channel power gain over a column/row of the DSCM since the receiver/transmitter maintains a fixed allocation of power to every sector. In general, this strategy would not yield the globally optimal transmit beam pattern. However, in a sparse channel, the only columns/rows that have non-trivial average are those containing a non-trivial element. Under the dominant cluster assumption, the column and row for the maximum average must correspond to the AOD or

AOA sector for the dominant cluster. SLS is therefore an algorithm which solves 1-BBS in $2R$ channel probes.

Parallelized Training

One of the important properties of SLS training is the non-adaptive approach. Instead of adopting a "ping-pong" strategy in which feedback is provided after every channel probe, SLS performs all measurements in batch and makes the final beam pattern selection from the CSI provided in a single feedback.

This design choice may seem suboptimal since information from prior channel probes can be used to aid beam pattern selection for future channel probes. Indeed, there has been many proposals for improving SLS using such adaptive methods [13–15]. However, non-adaptive beam training has an advantage over adaptive approaches when beam training must be conducted in *parallel* over multiple devices.

The parallelization advantage can be seen in the implementation of SLS SSW for AP association in the BHI. During the BTI, the AP *broadcasts* its beacon SSW, which can be heard by any number of prospective client devices. To accommodate the response SSW from multiple devices, the AP sets time slots in the following A-BTI during which clients compete for access for their response SSW (Figure 2.14). This contention process is costly as the AP will only respond to at most one client per slot. When the number of clients is large, not all of the clients may be serviced in the given BTI. When the number of clients is small, some time slots in the A-BTI will be unused and wasted. However, due to the non-adaptive nature of SLS, only a single A-BTI is required and the time slots can be kept relatively short. This reduces the overhead for association beam training.

If instead an adaptive approach is used for association beam training, every channel probe after the first can no longer be broadcasted because they would have been tailored

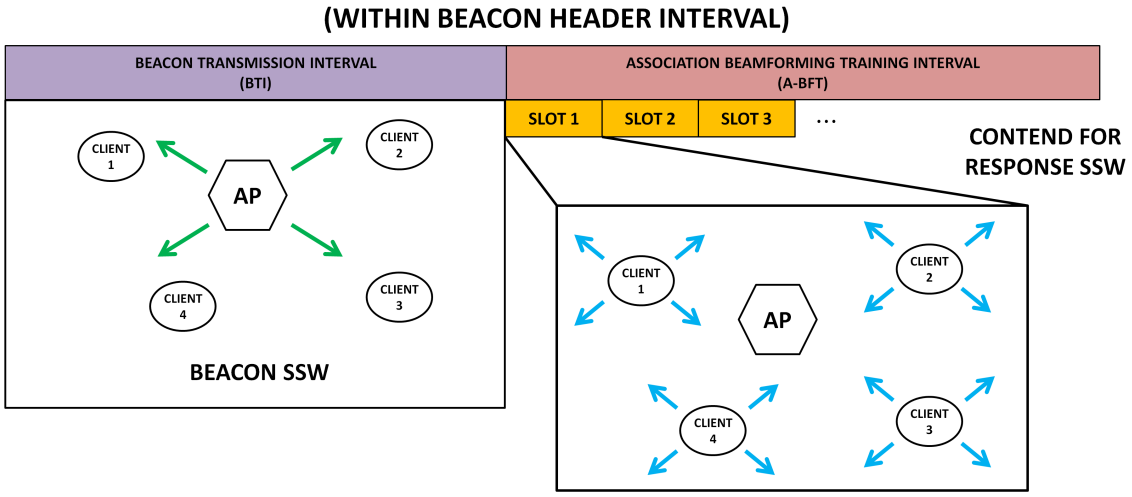


Figure 2.14: Parallelization of SLS SSW in AP association.

for a particular client device. As a result, the AP must either schedule multiple A-BTI sessions or extend the duration of every slot within the A-BTI to allow for dedicated beam training with each client. This additional time must be scheduled without any knowledge of the number of clients that may respond to the AP's beacon, incurring significantly more overhead than the non-adaptive approach.

2.7.2 Constraints on Beam Training Format

A beam training algorithm which seeks to replace SLS must satisfy the constraints that are essential to the role of SLS in the DMG MAC context. We identify three key constraints on beam training within the SLS context:

1. **Transmit and receive training must be conducted separately.**
2. **The amount of CSI transmitted in a feedback frame should be minimized.**
3. **Training is non-adaptive, and only a single feedback frame is sent for transmit training.**

Constraint 1 is due to the fact that SLS must serve as a neighbor discovery protocol. Without any prior knowledge of the transmitter, a receiving device must listen with an omnidirectional pattern to ensure no bias toward any spatial direction. This implies any transmit training must be conducted under the assumption that the receiver is using a fixed beam pattern. Receive beam training must therefore be conducted at a later time.

Constraint 2 is due to the fact that no reliable side channel exists for exchanging feedback information. A low rate modulation and coding scheme and extensive error correction must be adopted to ensure feedback information is received correctly. As a result, the amount of information included in each feedback should be minimized to reduce the time cost associated with a feedback frame. SLS achieves this by only sending information identifying the best sector instead of the full CSI gathered over all received training frames.

Constraint 3 is crucial if beam training is to be optimized for parallelization. As mentioned in Section 2.7.1, the non-adaptive approach allows the overhead for AP association to be minimized. While adaptive formats may be superior between single pairs of devices, they would incur more overhead when beam training with multiple devices is required. This is due to the inherent inefficiency associated with MAC scheduling.

2.8 Chapter Summary

In this chapter we presented a sector model for beamforming. In this model, beam patterns are represented by finite dimensional vectors, where each element corresponds to the average antenna gain over a particular angular sector. We showed that a mmWave channel can be similarly modeled using the joint angular sectors for the AOA and AOD of propagation paths. Using these models we defined the notion of channel probes, which are used to capture the time cost of beam training. The beam training problem was then stated as a search problem by invoking the sparsity of the spatial channel matrix. Finally, we examined how 802.11ad

approaches the beam training problem under the sector model, and the key implementation constraints for practical beam training were identified.

In the next chapter, the algorithmic optimization of beam training will be examined using the ideal BBPV and BCM models introduced in this chapter.

Chapter 3
IDEAL BOOLEAN BEAM TRAINING

In this chapter we will consider beam training using an ideal Boolean model. Within this model, we will be able to characterize the efficiency of a beam training algorithm, and therefore derive an optimum bound on the complexity of beam training. The tightness of the bound will be shown constructively. Furthermore, we will examine the implication of the bound on the scaling of training time cost with respect to hardware capabilities.

NOTE: All matrix indexing in this chapter will be be wrap around (i.e. modulo the dimension of the matrix).

3.1 Boolean Model

The channel estimation model developed in Section 2.5.1 can be simplified to a Boolean equivalent with the following assumptions:

- The BPVs used by TX and RX are BBPVs.
- The DSCM being measured is a BCM.
- The channel is noiseless.
- Boolean algebra is used, so scalar multiplication becomes logical AND and scalar addition becomes a logical OR operation. The truth tables for Boolean algebra are presented in Table 3.1.

a	b	$a \vee b$
0	0	0
0	1	1
1	0	1
1	1	1

a	b	$a \wedge b$
0	0	0
0	1	0
1	0	0
1	1	1

a	b	$a \oplus b$
0	0	0
0	1	1
1	0	1
1	1	0

a	$\neg a$
0	1
1	0

Table 3.1: Truth tables for Boolean algebra.

A channel probe in the Boolean model is equivalent to computing:

$$\bar{\gamma} \in \{0,1\} = \mathbf{b}_{RX}^{\intercal} \mathcal{H} \mathbf{b}_{TX} = \bigvee_{k=0}^{R-1} \bigvee_{l=0}^{R-1} (\mathbf{b}_{RX}[k] \wedge \mathbf{b}_{TX}[l] \wedge \mathcal{H}[k,l]) \tag{3.1}$$

where $\mathbf{b}_{RX}, \mathbf{b}_{TX} \in \{0,1\}^R$ are the BBPVs representing the beam patterns used by RX and TX to make the channel probe, and $\mathcal{H}$ is the BCM representation of the channel. In transmit training, $\bar{\gamma}$ would be the feedback provided by the receiving device.

The Boolean model adopts a conservative estimate of the maximum amount of information obtainable from each channel probe. A "large" beamformed channel power gain, is represented by $\bar{\gamma} = 1$ while a "small" beamformed channel power gain is represented by $\bar{\gamma} = 0$. A large gain is measured if and only if the main lobes of the beam pattern used by RX and TX cover the AOA and AOD sectors, respectively, corresponding to cluster paths with large gain.

This approximation is useful for a couple reasons. Firstly, as stated in Section 2.7.2, the amount of CSI conveyed in a feedback frame should ideally be minimized. In the Boolean model, each channel probe produces exactly one bit of CSI and requires only a single bit of feedback. Secondly, the channel is non-ideal in practice. Thermal noise and other sources of non-ideality would reduce the information that can be obtained from each channel probe. Thus, it is reasonable to assume some thresholding is necessary to quantify the quality of the channel.

3.2 Boolean Beam Training Optimization

From the SLS constraints given in Section 2.7.2, transmit and receive training are conducted separately. This means it is sufficient to focus on optimizing transmit training with the premise that the receiver is adopting a fixed beam pattern. The same algorithm can then be used for receive training at an equivalent or lower time cost.

Suppose Q channel probes are used in transmit training. Define $\mathbf{\Psi} \in \{0,1\}^{Q\times R}$ to be a **measurement matrix** such that the i-th row of $\mathbf{\Psi}$ is the BBPV used by TX for the i-th channel probe. The binary measurements procured by the channel probes can then be expressed as a vector, $\bar{\mathbf{\Gamma}} \in \{0,1\}^{Q}$:

$$\begin{aligned} \bar{\mathbf{\Gamma}} &= \mathbf{\Psi}\mathbf{v}_{AOD} \\ \mathbf{v}_{AOD}^{\intercal} &= \bigvee_{i=0}^{R-1} \mathcal{H}[i,:] \end{aligned} \tag{3.2}$$

where $\mathbf{v}_{AOD}$ is the binary **AOD channel vector** representing the effective channel at TX when RX is using a fixed omnidirectional beam pattern. Due to the clustering behavior in mmWave channels, the AOD channel vector is assumed to be sparse.

> **Definition 8.** A binary vector, $\mathbf{b}$, is said to be **K-Sparse** if $\|\mathbf{b}\|_0 \leq K$, i.e, it has at most K non-zero elements.

The objective of beam training is to identify all the non-zero elements of $\mathbf{v}_{AOD}$ using knowledge of $\bar{\mathbf{\Gamma}}$ and $\mathbf{\Psi}$. This is equivalent to fully recovering $\mathbf{v}_{AOD}$ from $\bar{\mathbf{\Gamma}}$ and $\mathbf{\Psi}$.

> **Definition 9.** The measurement matrix, $\mathbf{\Psi} \in \{0,1\}^{Q\times R}$, consisting of Q stacked BBPVs, is said to be **K-Recoverable** if any K-sparse $\mathbf{v}_{AOD}$ can be recovered from knowledge of $\bar{\mathbf{\Gamma}}$ and $\mathbf{\Psi}$.

Beam training optimization under the Boolean model is equivalent to constructing a K-recoverable matrix that minimizes Q for a given R.

This setup is reminiscent of a compressed sensing problem. In fact, tackling this problem for general values of K requires a deep foray into the field of group testing, which is a form of compressed sensing over Boolean algebra. A brief introduction to the group testing approach is provided in Appendix B, but the treatment is outside of the scope of this work. In the beam training context, we will only be considering the $K = 1$ case under the dominant cluster assumption. Even in this simple case, we will see the constraints imposed on the structure of the measurement matrix leads to non-trivial results that are not covered by existing literature.

3.2.1 Necessary and Sufficient Conditions for 1-Recoverability

Suppose $\mathbf{v}_{AOD}$ is 1-sparse. Let k^* be the index corresponding to the single 1 in $\mathbf{v}_{AOD}$. From Equation (3.2), $\bar{\mathbf{\Gamma}}$ is given by:

$$\bar{\mathbf{\Gamma}} = \bigvee_{j=0}^{R-1} \mathbf{\Psi}[:, j] \wedge \mathbf{v}_{AOD}[j] = \mathbf{\Psi}[:, k^*] \tag{3.3}$$

Thus, the vector of Q bits gathered from the channel probes is exactly the column of $\mathbf{\Psi}$ corresponding to the index of the optimal sector.

Each column of $\mathbf{\Psi}$ therefore *encodes* the index of the column into a binary vector of length Q. If $\mathbf{\Psi}$ is 1-recoverable, it must be possible to *decode* the column index from the column vector. Given $\mathbf{\Psi}$ is made available to the decoder, this is possible if and only if the columns of $\mathbf{\Psi}$ are *distinct*. In that case, the decoder can determine k^* by identifying the

column of $\mathbf{\Psi}$ that is the smallest hamming distance from $\bar{\mathbf{\Gamma}}$:

$$k^* = \underset{j=0,\ldots,R-1}{\operatorname{argmin}} \left\| \bar{\mathbf{\Gamma}} \oplus \mathbf{\Psi}[:, j] \right\|_h \tag{3.4}$$

Thus, to optimize beam training, a measurement matrix of dimension $Q \times R$ must be designed such that all R columns are distinct and Q is minimized.

3.2.2 Optimization Problem Statement

If there are no constraints on the structure of the measurement matrix, constructing an optimal 1-recoverable matrix is a trivial problem. From source coding theory, there can be a maximum of 2^Q distinct binary vectors of length Q. This means an optimal matrix can be constructed by simply choosing $Q \geq \log_2(R)$ and picking R distinct binary vectors of length Q as the columns of a matrix.

However, in the beam training context, each row of the measurement matrix must be a valid BBPV. The BBPV constraints described in Section 2.2.2 tie the structure of the measurement matrix to hardware requirements. These requirements must be accounted for when considering the optimality of a 1-recoverable matrix. To illustrate this nuance, consider the two measurement matrices given in (3.5) and (3.6).

$$\mathbf{\Psi}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{3.5}$$

$$\mathbf{\Psi}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{3.6}$$

It is easy to verify that both $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ have distinct columns. Furthermore, both are of dimension 3×2^3. Thus, in the absence of beam constraints, both would be equally optimal.

However, differences between the two are revealed if BBPV constraints are applied to the rows of each matrix.

First consider $\mathbf{\Psi}_1$. The first row of $\mathbf{\Psi}_1$ has a single beam with a minimum beam width of 4. From the beam constraints given in Section 2.2.2, this implies $N_c \geq 1$ and $N_t \geq R\left(\frac{1}{4}\right) = 2$. The second row of $\mathbf{\Psi}_1$ has two beams with a minimum beam width of 2. This implies $N_c \geq 2$ and $N_t \geq R\left(\frac{2}{2}\right) = 8$. The third row of $\mathbf{\Psi}_1$ has a four beams with a minimum beam width of 1. This implies $N_c \geq 4$ and $N_t \geq R\left(\frac{4}{1}\right) = 32$. In order to implement $\mathbf{\Psi}_1$, the BBPV constraints must be satisfied for *all* channel probes. Thus, the minimum hardware requirements for implementing $\mathbf{\Psi}_1$ is $N_c = 4$ and $N_t = 32$.

By applying the same process to $\mathbf{\Psi}_2$, we obtain the minimum hardware requirements for implementing $\mathbf{\Psi}_2$ is $N_c = 2$ and $N_t = 8$. Although both $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ achieve the same resolution at the same time cost, $\mathbf{\Psi}_2$ can do so using an antenna array with only a quarter the number of antenna elements and half the number of subarrays. Therefore, $\mathbf{\Psi}_2$ must be more efficient, and therefore more optimal, in the sense that hardware capabilities are better utilized when $\mathbf{\Psi}_2$ is implemented than when $\mathbf{\Psi}_1$ is implemented.

The underutilization of hardware capabilities can be captured with the concept of a resolution margin.

> **Definition 10.** Let B_i denote the number of 1-blocks in the i-th row of a measurement matrix, $\mathbf{\Psi}$. Let W_i denote the minimum 1-block size in the i-th row of the measurement matrix. The **Resolution Margin** for the measurement matrix is given by:
>
> $$M_r = \max_{i \in \{0,\ldots,Q-1\}} \frac{B_i}{W_i} \approx \frac{N_t}{R} \tag{3.7}$$

The resolution margin of a measurement matrix can be interpreted as the ratio of the beam angle uncertainty, after beam training, to the minimum achievable beam width of

the array. The former is inversely proportional to the training resolution while the latter is inversely proportional to the number of elements in the array. The resolution margin should ideally be minimized, which indicates that the beam training algorithm is able to achieve high angular precision with respect to the hardware capabilities of the array.

The resolution margin measures the cost of a beam training algorithm with respect to the minimum hardware required to implement it. Likewise, the number of channel probes measures the cost of a beam training algorithm with respect to the time required to implement it. An optimal beam training algorithm should minimize both costs.

Problem 2. 1-Sparse Single Feedback Optimization Problem (1-SFO)
For a given R, design a matrix, $\mathbf{\Psi} \in \{0,1\}^{Q\times R}$, with distinct columns, such that $M_r Q$ is minimized, where M_r is the resolution margin given in (3.7).

1-SFO is difficult to analyze due to the mutual dependency between Q and M_R. To make analysi more tractable, we will relax the problem by introducing a simple approximation and reframing it as a maximization of a single variable instead of a minimization over two.

In the first simplification step, we note that $W_i \geq 1$ in the definition of M_r. Thus, M_r is upper bounded by:

$$M_r \leq B_{max} = \max_{i\in\{0,\ldots,Q-1\}} B_i \tag{3.8}$$

Therefore, the cost function can be relaxed by replacing $M_r Q$ with $B_{max}Q$, where $B_{max} \leq N_c$ is simply the largest number of beams used simultaneously over all of the channel probes.

Next we note that R is not hardware constrained, so achieving a particular R is not necessary as much as achieving a value of R above some desired threshold. In fact, if $R+\epsilon$, $\epsilon > 0$, can be achieved at the same cost of achieving R, obtaining $R+\epsilon$ would be preferred. Thus, instead of minimizing $B_{max}Q$ for a given R, the problem can be stated as maximizing

R for any given (B_{max}, Q) pair. Thus, the relaxed optimization problem can be stated as the maximization of R with respect to B_{max} and Q.

> **Problem 3. 1-Sparse Relaxed Single Feedback Optimization Problem (1-RSFO)**
>
> Given B_{max} and Q, design a matrix, $\mathbf{\Psi} \in \{0,1\}^{Q \times R}$, with distinct columns, such that R is maximized.

The remainder of this chapter will discuss solutions to 1-RSFO.

3.3 Bound for 1-RSFO

In this section we derive the key result of this paper, which can be summarized by a theorem:

> **Theorem 1.** Given B_{max} and Q, if $\mathbf{\Psi} \in \{0,1\}^{Q \times R}$ has distinct columns, then R is tightly upper bounded by $2B_{max}Q$.

The bound will be derived in Section 3.3.2, and the tightness will be shown in Section 3.3.3.

3.3.1 Edge Matrix

To facilitate the discussion of matrix structure, we introduce the notion of an edge matrix.

> **Definition 11.** The **Edge Matrix** corresponding to a measurement matrix, $\mathbf{\Psi} \in \{0,1\}^{Q \times R}$, is given by:

$$\begin{aligned} &\mathbf{\Psi}' \in \{0,1\}^{Q \times R} \\ &\mathbf{\Psi}'[:,k] = \mathbf{\Psi}[:,k] \oplus \mathbf{\Psi}[:,k-1] \end{aligned} \tag{3.9}$$

As the name suggests, each row of an edge matrix indicates the location of the 1-

block edges within each row of the corresponding measurement matrix. An example of a measurement matrix and the corresponding edge matrix is given in (3.10).

$$\mathbf{\Psi} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad \mathbf{\Psi}' = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{3.10}$$

Edge matrices have some useful properties which will be formally enumerated for future reference.

> **Property 1.** A measurement matrix can be reconstructed from its edge matrix if any column of the measurement matrix is known.

Proof. Suppose $\mathbf{\Psi}[:, k]$ is known. By definition of the edge matrix, we have $\mathbf{\Psi}'[:, k+1] = \mathbf{\Psi}[:, k+1] \oplus \mathbf{\Psi}[:, k]$. Thus $\mathbf{\Psi}[:, k+1]$ can be determined by computing:

$$\begin{aligned} \mathbf{\Psi}'[:, k+1] \oplus \mathbf{\Psi}[:, k] &= (\mathbf{\Psi}[:, k+1] \oplus \mathbf{\Psi}[:, k]) \oplus \mathbf{\Psi}[:, k] \\ &= \mathbf{\Psi}[:, k+1] \oplus (\mathbf{\Psi}[:, k] \oplus \mathbf{\Psi}[:, k]) \\ &= \mathbf{\Psi}[:, k+1] \oplus \mathbf{0} \\ &= \mathbf{\Psi}[:, k+1] \end{aligned}$$

The above holds for all k. Thus, by induction, it is possible to determine all the columns of $\mathbf{\Psi}$ given some initial column $\mathbf{\Psi}[:, k_0]$ and the complete edge matrix, $\mathbf{\Psi}'$. □

Property 1 implies it is sufficient to define the structure of a measurement matrix by defining its edge matrix. This is because for any given edge matrix, a measurement matrix with the given properties can be constructed by choosing the elements of a particular column and using the edge matrix to construct the remaining columns.

Property 2. The i-th row of a measurement matrix contains B_i 1-blocks if and only if there are exactly $2B_i$ elements with value 1 in the corresponding row of the edge matrix. The only exception is in the case where the entire row is a single 1-block of size R.

Proof. By definition, each 1-block must consist of a contiguous set of indices with value 1 that is not a subset of a larger contiguous set of such indices. This implies adjacent 1-blocks must be separated by a 0-block, where a 0-block is a set of indices which would form a 1-block in the BBPV if every element were negated. The only exception is if the vector consists of a single 1-block of size R, in which case there are trivially no 0-blocks.

By construction, $\mathbf{\Psi}'[:,k] = 1$ if and only if $\mathbf{\Psi}[:,k] \oplus \mathbf{\Psi}[:,k-1] = 1$, or equivalently, $\mathbf{\Psi}[:,k] \neq \mathbf{\Psi}[:,k-1]$. Note that $\mathbf{\Psi}[:,k] \neq \mathbf{\Psi}[:,k-1]$ if and only if exactly one of $\mathbf{\Psi}[:,k]$ and $\mathbf{\Psi}[:,k-1]$ is part of a 1-block. Such a scenario will occur if and only if $k-1$ and k corresponds to the indices spanning a transition between a 1-block and 0-block. If all adjacents 1-blocks are separated by 0-blocks, there are exactly two such transitions per 1-block. Therefore, there are exactly $2B_i$ values of k for which $\mathbf{\Psi}'[:,k] = 1$ if and only if there are B_i 1-blocks in the corresponding BBPV. □

Intuitively, Property 2 is equivalent to the statement that all beams, with the exception of omnidirectional beams, have two edges.

Property 3. The k_1-th and k_2-th columns, where $k_2 > k_1$, of a measurement matrix are distinct if and only if the sub-matrix given by $\mathbf{\Psi}'[:,k_1:k_2]$ has at least one row which contain an odd number of 1s.

Proof. $\mathbf{\Psi}[:,k_1]$ and $\mathbf{\Psi}[:,k_2]$ are distinct if and only if $\mathbf{0} \neq \mathbf{\Psi}[:,k_1] \oplus \mathbf{\Psi}[:,k_2]$. Using the

associative property of XOR, we have:

$$\mathbf{0} \neq \mathbf{\Psi}[:, k_1] \oplus \mathbf{\Psi}[:, k_2] = \bigoplus_{i=k_1+1}^{k_2} (\mathbf{\Psi}[:, i] \oplus \mathbf{\Psi}[:, i-1])$$
$$\mathbf{0} \neq \bigoplus_{i=k_1}^{k_2-1} \mathbf{\Psi'}[:, i] \tag{3.11}$$

The RHS of (3.11) is precisely the XOR of all the the columns in the sub-matrix $\mathbf{\Psi'}[:, k_1 : k_2]$. Since the XOR of n bits is 1 if and only if there is an odd number of 1 bits, (3.11) only holds if there is an odd number of 1s in at least one row of $\mathbf{\Psi'}[:, k_1 : k_2]$. □

Property 3 is an important property of the edge matrix that makes it useful for enforcing distinct columns in the measurement matrix.

3.3.2 Bound Derivation

Suppose $\mathbf{\Psi} \in \{0, 1\}^{Q \times R}$ is a solution to 1-RSFO. By definition, the columns of $\mathbf{\Psi}$ must be distinct. Omnidirectional beam patterns transmit the same signal over all the sectors so they do not contribute to the distinction of columns and would not be part of an optimal measurement matrix. Thus, we can assume no row of $\mathbf{\Psi}$ corresponds to a vector of all 1s.

Since all the columns of $\mathbf{\Psi}$ are distinct, all adjacent columns of $\mathbf{\Psi}$ must be distinct. From Property 3 of edge matrices, this implies $\mathbf{\Psi'}[:, k : k+1]$ must contain at least one row with an odd number of 1s for all choices of k. Since $\mathbf{\Psi'}[:, k : k+1] = vb\Psi'[:, k]$ is just a column of the edge matrix, every column of the edge matrix must contain at least a single 1. Let $N_{(1)}$ denote the total number of 1s in the entire edge matrix, $\mathbf{\Psi'}$. Since every column contains at least a single 1, we have $N_{(1)} \geq R$.

The total number of 1s in the edge matrix is equal to the sum of the number of 1s in

each row over all of the rows. Using Property 2 of edge matrices, this implies:

$$N_{(1)} = \sum_{i=0}^{Q-1} 2B_i \leq 2B_{max}Q \tag{3.12}$$

Where B_i is the number of 1-blocks in the i-th row of $\mathbf{\Psi}$.

Thus, we have $R \leq N_{(1)} \leq 2B_{max}Q$, which means $R \leq 2B_{max}Q$ for any choice of B_{max} and Q, given $\mathbf{\Psi}$ has distinct columns.

This shows that $2B_{max}Q$ is *an* upper bound for R, but does not imply tightness. However, it does imply that any measurement matrix with Q rows and $2B_{max}Q$ distinct columns is optimal if there are exactly B_{max} 1-blocks in every row.

3.3.3 Bound Tightness

We will now show that the upper bound of $R = 2B_{max}Q$ is tight for all sufficiently large Q. The result can be summarized by the following theorem:

> **Theorem 2.** The upper bound of $R = 2B_{max}Q$ for 1-RSFO can be achieved for all $Q \geq \mathcal{P}(3B_{max})$, where $\mathcal{P}(n)$ denotes the smallest prime larger than n.

Since $\mathcal{P}(n) \leq 2n$ by Bertrand's Postulate [16], a looser, but more explicit bound would be $Q \geq 6B_{max}$. Constructions for $Q < \mathcal{P}(3B_{max})$ exists for at least for some B_{max}. but in this work, we will prove $Q \geq \mathcal{P}(3B_{max})$ is achievable for all B_{max}.

To prove bound tightness, we will define a general method for constructing matrices which achieve the bound. The matrices constructed via this method are not necessary unique in that there may be other solutions of the same dimension. However, their existence is sufficient to show tightness.

The proof is inductive and consists of two steps, which are summarized by the following lemmas:

> **Lemma 1.** For all $B_{max} > 0$, is is possible to construct a measurement matrix, $\mathbf{\Psi}$, of dimension $\mathcal{P}(3B_{max}) \times 2B_{max}\mathcal{P}(3B_{max})$ which has distinct columns and exactly B_{max} 1-blocks per row.

> **Lemma 2.** Given a 1-RSFO solution, $\tilde{\mathbf{\Psi}}$, with dimension $\tilde{Q} \times \tilde{R}$, where $\tilde{R} \geq 2B_{max}$, it is possible to construct another solution, $\mathbf{\Psi}$, with dimension $(\tilde{Q}+1) \times (\tilde{R}+2B_{max})$.

Lemma 1 establishes a base case by defining the existence of a solution with a particular Q_0 for all B_{max}. Lemma 2 is the induction step, which extends the base case constructions to all $Q > Q_0$.

3.3.4 Induction Step

Lemma 2 can be shown using edge matrices. Let $\tilde{\mathbf{\Psi}} \in \{0,1\}^{\tilde{Q} \times \tilde{R}}$ denote the existing solution where $\tilde{R} > 2B_{max}$, and let $\tilde{\mathbf{\Psi}}'$ denote the corresponding edge matrix. The goal is to construct a new matrix $\mathbf{\Psi} \in \{0,1\}^{(\tilde{Q}+1) \times (2B_{max}+\tilde{R})}$ with distinct columns and every row containing no more than B_{max} 1-blocks. The pseudocode for this construction is given in Algorithm 1 of Appendix C, but the process can be intuitively summarized as follows:

1. Compute $\tilde{\mathbf{\Psi}}'$ from $\tilde{\mathbf{\Psi}}$
2. Insert a zero row to the top of $\tilde{\mathbf{\Psi}}'$
3. Pick $2B_{max}$ columns of $\tilde{\mathbf{\Psi}}'$ and insert a zero column immediately after each.
4. In every zero column, set the top element to be 1. The resulting matrix will now be $\mathbf{\Psi}'$, the edge matrix for the new measurement matrix.

5. Use Property 1 of edge matrices to build a particular $\mathbf{\Psi}$ from $\mathbf{\Psi}'$ by setting $\mathbf{\Psi}[:,0]$ to be any binary vector.

   Claim 1. Given $\tilde{\mathbf{\Psi}}$ has distinct columns, a $\mathbf{\Psi}$ constructed in the manner given above must have distinct columns.

Proof. This can be proven by contradiction. Suppose there exists k_1 and $k_2 > k_1$ such that $\mathbf{\Psi}[:,k_1] = \mathbf{\Psi}[:,k_2]$. By Property 3 of edge matrices, this means every row of $\mathbf{\Psi}'[:,k_1:k_2]$ must contain an even number of 1s. Let $2n$ be the number of 1s contained in the top row of $\mathbf{\Psi}'[:,k_1:k_2]$. Let $\mathcal{Y}$ denote the set of column indices $\{i\}_{i=k_1}^{k_2-1}$ which correspond to the columns of $\mathbf{\Psi}'$ included in $\mathbf{\Psi}'[:,k_1:k_2]$. Let $\mathcal{Y}_\infty = i_{j}{}_{j=1}^{2n}$ denote the indices of the 1s contained in the top row of $\mathbf{\Psi}'[:,k_1:k_2]$, so $\mathcal{Y}_\infty \subset \mathcal{Y}$.

From the construction of $\mathbf{\Psi}'$, there does not exist l,m such that $i_l = i_m + 1$. Thus, we can define $\mathcal{J} = \mathcal{Y} - \mathcal{Y}_\infty \neq \varnothing$. Also from the construction of $\mathbf{\Psi}'$, the bottom Q elements of the columns indexed by $\mathcal{J}$ are a set of contiguous columns of $\tilde{\mathbf{\Psi}}'$ which can be treated as a sub-matrix, $\tilde{\mathbf{\Psi}}'[:,m_1:m_2]$, where $0 < m_2 - m_1 \leq k_2 - k_1$.

By assumption, every one of the bottom Q rows $\mathbf{\Psi}'[:,k_1:k_2]$ must contain an even number of 1s. Since the bottom Q elements of the columns in $\mathcal{Y}_\infty$ are all 0 by construction, the number of 1s in each of the bottom Q rows of $\mathbf{\Psi}'[:,k_1:k_2]$ is equal to the number of 1s in the corresponding row of $\tilde{\mathbf{\Psi}}'[:,m_1:m_2]$. This implies there exist $m_2 > m-1$ such that no row of $\tilde{\mathbf{\Psi}}'[:,m_1:m_2]$ contains an odd number of 1s. By Property 3 of edge matrices, there must be two columns in $\tilde{\mathbf{\Psi}}'[:m_1]$ that are identical, contradicting the premise that$\tilde{\mathbf{\Psi}}'$ has distinct columns. □

The induction step is best illustrated with an example. Consider the matrix of size 2×4 given in (3.13) with its edge matrix. This measurement matrix is trivially optimal for $B_{max} = 1$ since $R = 4 = 2(1)(2) = 2B_{max}Q$ and there is exactly one 1-block in each row.

The optimality can also be confirmed by noting each column of the edge matrix contains a single 1 and each row contains $2B_{max} = 2$ 1s.

$$\tilde{\mathbf{\Psi}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad \tilde{\mathbf{\Psi}}' = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \tag{3.13}$$

To construct the new 3×6 matrix, a zero row is added to $\tilde{\mathbf{\Psi}}'$ and $2B_{max} = 2$ zero columns are inserted after two distinct columns. Here, zero columns are inserted after columns 0 and 2 of $\tilde{\mathbf{\Psi}}'$. This yields the matrix:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

The top element of each zero column is set to 1, resulting in a new edge matrix:

$$\mathbf{\Psi}' = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

To produce the new measurement matrix, we can pick:

$$\mathbf{\Psi}[:, 0] = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

The next column, $\mathbf{\Psi}[:, 1]$, can be determined by computing:

$$\mathbf{\Psi}[:, 1] = \mathbf{\Psi}[:, 0] \oplus \mathbf{\Psi}'[:, 1] = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \oplus \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

This process is repeated iteratively to compute the remaining columns until the final $\mathbf{\Psi}$ is obtained:

$$\mathbf{\Psi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{3.14}$$

Note that the columns of $\mathbf{\Psi}$ are distinct and there is exactly one 1-block in each row. Its dimension also satisfy the optimality requirement for $B_{max} = 1$ since $R = 6 = 2(1)(3)$.

3.3.5 Base Cases

Lemma 1 can be shown by showing a solution to 1-RSFO with dimension $\mathcal{P}(3B_{max}) \times 2B_{max}\mathcal{P}(3B_{max})$ can be generated for all choices of $B_{max} > 1$ (The construction for $B_{max} = 1$ case is already provided in (3.14) since the given matrix has $Q = 3$ rows and $3 < \mathcal{P}(3) = 5$.). The pseudocode for generating the base cases is presented in Algorithm 2 of Appendix C. Here, we will provide the intuitive summary of the process and the proof of correctness.

Construction Process

The algorithm operates by constructing an auxiliary **generating vector**, $\mathbf{g} \in \{0, \ldots, Q-1\}^{B_{max}Q}$ ($Q = \mathcal{P}(3B_{max})$), which is then used to define the measurement matrix.
The construction of the generating vector can be decomposed into two steps:

1. Create a set of B_{max} vectors, $\{\mathbf{c}_m\}_{m=1}^{B_{max}}$, such that $\mathbf{c}_m \in \{0, \ldots, Q-1\}^Q$ and $\mathbf{c}_m[k] = mk \mod Q$ for all $m = 1, \ldots, B_{max}, k = 0, \ldots, Q-1$.
2. Concatenate the set $\{\mathbf{c}_m\}_{m=1}^{B_{max}}$ lengthwise to form the generating vector of dimension $B_{max}Q$:

$$\mathbf{g} = \left[\mathbf{c}_1^\intercal \cdots \mathbf{c}_{B_{max}}^\intercal\right]^\intercal \tag{3.15}$$

For example, if $B_{max} = 2$, then $Q = 7$, and $\mathbf{c}_1$ and $\mathbf{c}_2$ would be given by:

$$\mathbf{c}_1 = [0\ 1\ 2\ 3\ 4\ 5\ 6]^\intercal$$

$$\mathbf{c}_2 = [0\ 2\ 4\ 6\ 1\ 3\ 5]^\intercal$$

Concatenating the two yields the generating vector for $B_{max} = 2$:

$$\mathbf{g} = [0\ 1\ 2\ 3\ 4\ 5\ 6\ 0\ 2\ 4\ 6\ 1\ 3\ 5]^\intercal \tag{3.16}$$

Since $\mathbf{c}_i[0] = (\mathbf{c}_j[Q-1] + i) \mod Q = 0$ for all i, j, the generating vector can be defined more compactly by:

$$\begin{aligned} &\mathbf{g} \in \{0, \ldots, Q-1\}^{B_{max}Q} \\ &\mathbf{g}[0] = 0 \\ &\forall k = 0, \ldots, Q-1; m = 1, \ldots, B_{max} : \\ &\mathbf{g}[(m-1)Q + k] = (\mathbf{g}[(m-1)Q + k - 1] + m) \mod Q \end{aligned} \tag{3.17}$$

Where $Q = \mathcal{P}(3B_{max})$.

The generating vector can be used to generate an optimal measurement matrix by specifying the indices which are assigned a value 1 in each column of the matrix. The process of producing the measurement matrix from the generating vector can be summarized as follows:

1. Initialize all the elements of $\mathbf{\Psi} \in \{0, 1\}^{Q \times 2B_{max}Q}$ to 0.
2. Slide a window of size 2 incrementally over the generating vector (with wrap around).
3. At the j-th step, set the the q-th row of the j-th *even* column (ie. $2j$-th column) of $\mathbf{\Psi}$ to 1, for all q enclosed by the window.
4. Slide a window of size 3 incrementally over the generating vector (with wrap around).
5. At the j-th step, set the the q-th row of the j-th *odd* column (ie. $(2j+1)$-th column) of $\mathbf{\Psi}$ to 1, for all q enclosed by the window.

Figure 3.1 illustrates how a generating vector is related to the matrix it generates. In this example, the generating vector is $\mathbf{g} = [0\ 1\ 2\ 3]^\intercal$, corresponding to $Q = 4$ and $B_{max} = 1$. The same procedure can be applied to (3.16) to yield the 7×28 measurement matrix given

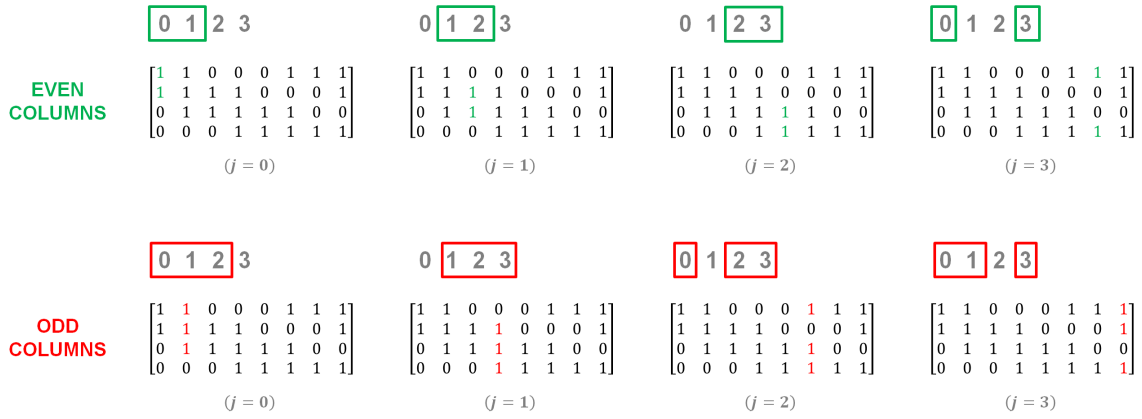


Figure 3.1: Example of the 2-generating process. A generating vector of length 4 is used to produce a measurement matrix of dimension 4×8.

in (3.18).

$$\begin{bmatrix} 1&1&0&0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&1&1&1 \\ 1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&1 \\ 0&1&1&1&1&1&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&0&0 \\ 0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&1&1&1&1&1&0&0&0&0&0 \end{bmatrix} \tag{3.18}$$

Note that each row of (3.18) contains exactly two 1-blocks of size 5, and all 28 columns are distinct.

For future reference, it is convenient to introduce a shorthand for this procedure of generating a measurement matrix from a generating vector.

Definition 12. Let $\mathcal{I}(\mathbf{v}) = \{j \text{ s.t. } \mathbf{v}[j] = 1\}$ denote the set of indices corresponding to the non-zero elements of a binary vector, $\mathbf{v}$. Let $\mathbf{v}$ denote the set of distinct elements in a (not necessarily binary) vector, $\mathbf{v}$.
$\mathbf{\Psi} \in \{0,1\}^{Q \times 2B_{max}Q}$ is said to be **L-Generated** by $\mathbf{g} \in \{0, \ldots, Q-1\}^{B_{max}Q}$ if:

$$\forall k = 0, \ldots, B_{max}Q :$$
$$\mathcal{I}(\mathbf{\Psi}[:, 2k]) = \{\mathbf{g}[k : k+L]\}$$
$$\mathcal{I}(\mathbf{\Psi}[:, 2k+1]) = \{\mathbf{g}[k : k+L+1]\}$$

Using this terminology, each base case matrix is said to be 2-generated from the generating vector defined by Equation (3.17).

Proof of Correctness

From the definition, every measurement matrix 2-generated from a generating vector defined by Equation (3.17) will have an optimal dimension of $Q \times 2B_{max}Q$. Thus, to show such a matrix is a solution to 1-RSFO, it is sufficient to prove two lemmas:

Lemma 3. A measurement matrix 2-generated from a generating vector defined by Equation (3.17) has exactly B_{max} 1-blocks in every row.

Lemma 4. A measurement matrix 2-generated from a generating vector defined by Equation (3.17) has distinct columns.

Proof of Lemma 3

From the definition of 2-generate, each element of the generating vector is used in the specification of exactly 5 contiguous columns of the measurement matrix. Namely, $\mathbf{g}[j]$ sets the $(\mathbf{g}[j])$-th row of the sub-matrix, $\mathbf{\Psi}[:, j-2 : j+3]$, to 1. Thus, it is sufficient to show

that there are exactly B_{max} occurrences of every element in $\{0, \ldots, Q-1\}$ within $\mathbf{g}$ and no two occurrences are within 2 indices each other.

> **Claim 2.** There are exactly B_{max} occurrences of every element in $\{0, \ldots, Q-1\}$ within any generating vector, $\mathbf{g}$, defined by Equation (3.17).

Proof. $\mathbf{g}$ can be expressed as the concatenation of B_{max} sub-vectors as defined in Equation (3.15). Because $Q = \mathcal{P}(3B_{max})$ is prime, each element of $\{0, \ldots, Q-1\}$ occurs exactly once in every sub-vector, $\mathbf{c}_i$. This is because $ik_1 \equiv ik_2 \mod Q$ implies $i(k_1 - k_2) \equiv 0 \mod Q$. Since $1 \leq i \leq B_{max} < Q$, it must be the case that $k_1 - k_2$ divides Q. This can only happen if $k_1 = k_2$ since $k_1 - k_2 < Q$ and Q is prime.
Because $\mathbf{g}$ is formed by the concatenation of B_{max} such sub-vectors, each element of $\{0, \ldots, Q-1\}$ must occur exactly B_{max} times $\mathbf{g}$. □

> **Claim 3.** There does not exist k_1, k_2 such that $k_1 < k_2 \leq k_1 + 2$ and $\mathbf{g}[k_1] = \mathbf{g}[k_2]$ for any $\mathbf{g}$ defined by Equation (3.17).

Proof. From Equation (3.17), we have:

$$\mathbf{g}[k_2] \equiv \mathbf{g}[k_1] + m_1 + m_2 \mod Q \tag{3.19}$$

Where $m_1, m_2 \leq B_{max}$, $m_1 \geq 1$, $m_2 \geq 0$, and $m_2 = 0$ if and only if $k_2 = k_1 + 1$.
If $\mathbf{g}[k_1] = \mathbf{g}[k_2]$, then $\mathbf{g}[k_1] \equiv \mathbf{g}[k_1] + m_1 + m_2 \mod Q$, which implies $0 \equiv m_1 + m_2 \mod Q$. However, since $1 \leq m_1 + m_2 \leq 2B_{max} < Q$, this is impossible. Therefore, such a k_1, k_2 pair cannot exist. □

Proof of Lemma 4

The j-th column of a measurement matrix is fully specified by the set of indices corresponding

to the non-zero elements of the column, denoted by $\mathcal{I}(\mathbf{\Psi}[:, j])$. From the definition of 2-generate, $\mathcal{I}(\mathbf{\Psi}[:, j])$ is fully specified by the set of elements corresponding to a sub-vector of length 2 or 3 within the generating vector, $\mathbf{g}$. Thus, Lemma 4 can be proven by showing the following:

- There does not exist $k_1 \neq k_2$ such that $\{\mathbf{g}[k_1 : k_1 + 2]\} = \{\mathbf{g}[k : k_2 + 2]\}$.
- There does not exist $k_1 \neq k_2$ such that $\{\mathbf{g}[k_1 : k_1 + 3]\} = \{\mathbf{g}[k : k_2 + 3]\}$.
- There does not exist k_1, k_2 such that $\{\mathbf{g}[k_1 : k_1 + 2]\} = \{\mathbf{g}k : k_2 + 3]\}$.

Claim 4. There does not exist $k_1 \neq k_2$ such that $\{\mathbf{g}[k_1 : k_1 + 2]\} = \{\mathbf{g}[k : k_2 + 2]\}$.

Proof. This can be proven by contradiction.

From Equation (3.17), we have:

$$\mathbf{g}[k_1 + 1] \equiv \mathbf{g}[k_1] + m_1 \mod Q \tag{3.20}$$

$$\mathbf{g}[k_2 + 1] \equiv \mathbf{g}[k_2] + m_2 \mod Q \tag{3.21}$$

Where $1 \leq m_1, m_2 \leq B_{max}$.

If $\{\mathbf{g}[k_1 : k_1 + 2]\} = \{\mathbf{g}[k_2], \mathbf{g}[k_2 + 2]\}$, it must be the case that either $\mathbf{g}[k_1] = \mathbf{g}[k_2]$ and $\mathbf{g}[k_1 + 1] = \mathbf{g}[k_2 + 1]$ or $\mathbf{g}[k_1 + 1] = \mathbf{g}[k_2]$ and $\mathbf{g}[k_1] = \mathbf{g}[k_2 + 1]$.

Suppose $\mathbf{g}[k_1 + 1] = \mathbf{g}[k_2]$ and $\mathbf{g}[k_1] = \mathbf{g}[k_2 + 1]$. Then adding (3.20) and (3.21) yields:

$$\mathbf{g}[k_1 + 1] + \mathbf{g}[k_1] \equiv \mathbf{g}[k_1 + 1] + \mathbf{g}[k_1]m_1 + m_2 \mod Q$$

$$0 \equiv m_1 + m_2 \mod Q$$

Since $2 \leq m_1 + m_2 \leq 2B_{max} < Q$, $0 \equiv m_1 + m_2 \mod Q$ is impossible.

Suppose $\mathbf{g}[k_1] = \mathbf{g}[k_2]$ and $\mathbf{g}[k_1 + 1] = \mathbf{g}[k_2 + 1]$. Then subtracting (3.20) from (3.21) yields:

$$0 \equiv m_1 - m_2 \mod Q$$

Since $1 \leq m_1, m_2 \leq B_{max} < Q$, it must be the case that $m_1 = m_2$. This implies that $\mathbf{g}[k_1]$ and $\mathbf{g}[k_2]$ are elements of the same sub-vector, $\mathbf{c}_{m_1}$ when $\mathbf{g}$ is expressed in the form in Equation 3.15. Since there is exactly one occurrence of each element in $\{0, \ldots, Q-1\}$ within the sub-vector (see proof of Claim 2), $\mathbf{g}[k_1] = \mathbf{g}[k_1]$ implies that $k_1 = k_2$, contradicting the premise that $k_1 \neq k_2$ □

Claim 5. There does not exist $k_1 \neq k_2$ such that $\{\mathbf{g}[k_1 : k_1 + 3]\} = \{\mathbf{g}[k : k_2 + 3]\}$.

Proof. First note that Claim 3 guarantees $\{\mathbf{g}[k : k + 3]\}$ contains three distinct elements for all k, since fewer than three distinct elements implies there exists $k \leq j_1, j_2 \leq k + 2$ such that $\mathbf{g}[j_1] = \mathbf{g}[j_2]$.

Consider the case where $\{k_1, k_1 + 1, k_1 + 2\} \cap \{k_2, k_2 + 1, k_2 + 2\} = \varnothing$. From Claim 4, the following conditions must hold:

- $\{\mathbf{g}[k_1], \mathbf{g}[k_1 + 1]\} \neq \{\mathbf{g}[k_2], \mathbf{g}[k_2 + 1]\}$
- $\{\mathbf{g}[k_1], \mathbf{g}[k_1 + 2]\} \neq \{\mathbf{g}[k_2], \mathbf{g}[k_2 + 2]\}$
- $\{\mathbf{g}[k_1 + 1], \mathbf{g}[k_1 + 2]\} \neq \{\mathbf{g}[k_2 + 1], \mathbf{g}[k_2 + 2]\}$

It is easy to verify, by listing out all six permutations of three distinct numbers, that $\{\mathbf{g}[k_1 : k_1 + 3]\} = \{\mathbf{g}[k : k_2 + 3]\}$ is impossible when these three conditions hold. Thus, $\{\mathbf{g}[k_1 : k_1 + 3]\} = \{\mathbf{g}[k : k_2 + 3]\}$ can only occur if $\{k_1, k_1 + 1, k_1 + 2\} \cap \{k_2, k_2 + 1, k_2 + 2\} \neq \varnothing$.

Without loss of generality, let $k_2 > k_1$. Since $k_1 \neq k_2$, it then follows that either $\{k_1, k_1 +$

$1, k_1 + 2\} \cap \{k_2, k_2 + 1, k_2 + 2\} = \{k_1 + 1 = k_2, k_1 + 2 = k_2 + 1\}$ or $\{k_1, k_1 + 1, k_1 + 2\} \cap \{k_2, k_2 + 1, k_2 + 2\} = \{k_1 + 2 = k_2\}$.

Suppose $\{k_1, k_1 + 1, k_1 + 2\} \cap \{k_2, k_2 + 1, k_2 + 2\} = \{k_1 + 1 = k_2, k_1 + 2 = k_2 + 1\}$ and $\{\mathbf{g}[k_1 : k_1 + 3]\} = \{\mathbf{g}[k : k_2 + 3]\}$. Then it must be that $\mathbf{g}[k_1] = \mathbf{g}[k_2 + 2] = \mathbf{g}[k_1 + 3]$.

Now suppose $\{k_1, k_1 + 1, k_1 + 2\} \cap \{k_2, k_2 + 1, k_2 + 2\} = \{k_1 + 2 = k_2\}$ and $\{\mathbf{g}[k_1 : k_1 + 3]\} = \{\mathbf{g}[k : k_2 + 3]\}$. From Claim 3, it cannot be the case that $\mathbf{g}[k_1 + 1] = \mathbf{g}[k_2 + 1]$ since $k_2 + 1 = (k_1 + 1) + 2$. This means $\mathbf{g}[k_1] = \mathbf{g}[k_2 + 1] = \mathbf{g}[k_1 + 3]$.

Thus, to prove the claim it is sufficient to show that there does not exist a k_1 such that $\mathbf{g}[k_1] = \mathbf{g}[k_1 + 3]$. This can be proven by contradiction. Suppose such a k_1 existed. From Equation (3.17), we have:

$$\mathbf{g}[k_1 + 3] \equiv \mathbf{g}[k_1] + m_1 + m_2 + m_3 \mod Q$$

Where $1 \leq m_1, m_2, m_3 \leq B_{max}$.

If $\mathbf{g}[k_1] = \mathbf{g}[k_1+3]$, then $0 \equiv m_1+m_2+m_3 \mod Q$. However, $3 \leq m_1+m_2+m_3 \leq 3B_{max} < Q$, so this is impossible. □

Claim 6. There does not exist k_1, k_2 such that $\{\mathbf{g}[k_1 : k_1 + 2]\} = \{\mathbf{g}k : k_2 + 3]\}$.

Proof. In the proof of Claim 4 and Claim 5, we have implicitly shown that $\{\mathbf{g}[k_1 : k_1 + 2]\}$ must contain 2 distinct elements and $\{\mathbf{g}[k_1 : k_1 + 3]\}$ must contain 3 distinct elements. Since sets of different sizes cannot be equal, the validity of this claim trivially follows. □

3.4 Time-Performance Trade-off

The tight bound of $R \leq 2B_{max}Q$ has offer insight into scaling complexity of the time cost with respect to hardware capabilities.

Recall from Definition 10 that the resolution margin captures how precisely an algorithm can identify the optimal direction to point a beam relative to the achievable beam width of the array. In a practical scenario, we expect the desired resolution margin to be fixed. That is, regardless of the size of an antenna array, the goal is to get within the same percentage of angular uncertainty with respect to the minimum beam width of the array. Suppose we wish to fix this resolution margin to a value of 1. Under this assumption, R can be approximated by N_t. Furthermore, since B_{max} is upper bounded by N_c, and N_c is expected to be no more than $\sqrt{N_t}$ (see Section 2.4), B_{max} can be replaced by $\sqrt{N_t}$. Making these substitutions into the bound yields a limiting value for Q that is completely determined by the number of elements in the array:

$$\begin{aligned} N_t &\leq 2\sqrt{N_t}Q \\ Q &\sim \mathcal{O}(\sqrt{N_t}) \end{aligned} \tag{3.22}$$

Thus, even at the theoretically *optimal* algorithmic efficiency, the time cost of beam training must at scale at least proportionally with the square root of the number of antenna elements. Making fewer than $\mathcal{O}(\sqrt{N_t})$ measurements will inevitably cause a drop in the algorithmic efficiency (within the context of the Boolean model).

3.5 Chapter Summary

In this chapter we considered beam training in a very ideal scenario where each beam training algorithm is represented by a binary measurement matrix consisting of the ideal beam patterns used to probe the channel. Beam pattern constraints were used to develop a metric which quantifies the efficiency of algorithms described by different measurement matrices. Using this metric, a bounding relationship was derived for the dimensions of an optimal matrix. A general method for constructing such optimal matrices were provided.

Finally, the bound is used to derive the scaling relationship between the number of array elements and the time cost.

In the next Chapter, we will explore the extent to which the optimality of algorithms developed for the ideal Boolean model holds in realistic, non-ideal environments.

Chapter 4
IMPACT OF NON-IDEAL FACTORS

In Chapter 3, an extremely ideal model was adopted to simplify beam training analysis. This enabled the development of a class of "optimal" beam training algorithms. In this Chapter, a few of these algorithms will be implemented through simulation to gauge their viability in realistic, non-ideal environments.

4.1 Tested Algorithms

In the Boolean model, optimal beam training algorithms were defined in the form of binary measurement matrices. We will compare two of such matrices with the existing SLS algorithm employed by 802.11ad.

The first matrix is optimal for the single beam case ($B_{max} = 1$) and will be referred to as the **1-Beam Optimal Algorithm (1-BOA)**. The second matrix is optimal for the two beam case ($B_{max} = 2$) and will be referred to as the **2-Beam Optimal Algorithm (2-BOA)**. The matrices are shown in (4.1) and (4.2), respectively. Both were derived in Chapter 3 as examples (See Equations (3.14) and (3.18)).

$$\mathbf{\Psi}_{1-BOA} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{4.1}$$

$$\mathbf{\Psi}_{2-BOA} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.2}$$

The SSW strategy used by SLS in 802.11ad will be used as a baseline for comparison. The measurement matrix corresponding to the SLS strategy is simply an identity matrix. To ensure a fair comparison, the number of channel probes (ie. rows in the measurement matrix) is kept fixed. Thus, 1-BOA will be compared with an identity matrix of dimension 3 while 2-BOA will be compared with an identity matrix of dimension 7. The two SLS matrices are explicitly shown in (4.3) and (4.4) for completeness.

$$\mathbf{\Psi}_{SLS(Q=3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.3}$$

$$\mathbf{\Psi}_{SLS(Q=7)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.4}$$

4.2 Models for Non-Ideal Factors

There are three main sources of non-ideality that must be addressed:

1. **Non-Ideal Channel:** The Boolean model assumes a channel that is well modeled by a BCM with a single non-zero element. This implies there is a single dominant cluster which subtends a single AOA-AOD sector pair. In practice, there can be situations where there are more than one cluster with high gain. Furthermore paths from the same cluster experience scattering and have finite AOA and AOD spread. Thus, the dominant cluster assumption is not always accurate.
2. **Non-Ideal Beam Patterns:** The Boolean model assumes ideal BBPVs, which do not capture the absolute power gains associated with each beam pattern. Furthermore, BBPVs do not account for side lobes, which can cause non-trivial amounts of power to be allocated to sectors which are not meant to be measured.
3. **Non-Ideal Channel Probes:** The ideal Boolean model assumes each channel probe gathers a single binary measurement in the complete absence of noise. In practice, measurements will be corrupted by additive channel noise; hence a binary metric for channel quality must be obtained by thresholding the noisy measurements, which will therefore degrade beam training performance.

4.2.1 Non-Ideal Channel

The DSCM model described in Definition 5 is used to capture the non-ideal characteristics of a mmWave channel. In this model, the channel is fully specified by a set of propagation paths. Each path is defined by a tuple, $(\theta_{c,m}, \phi_{c,m}, \alpha_{c,m})$, where $\theta_{c,m}$ is the AOA, $\phi_{c,m}$ is the AOD, and $\alpha_{c,m}$ is the complex gain associated with the path. The paths are grouped into C clusters, each containing M paths. Paths belonging to the same cluster have highly correlated parameters.

The method proposed by the NYU Wireless group[1] , with some simplification, is used to generate path parameters with the desired characteristics [11]. The associated variables are given in Table 4.1, and the procedure is summarized as follows:

1. Choose the number of clusters C from a Poisson distribution, and the number of paths per cluster, M, from a uniform distribution.
2. Choose the mean AOA and AOD angles, θ_c and ϕ_c, for each cluster from a uniform distribution, and choose the spread (ie. standard deviation) of the path angles, $\sigma_{c,AOA}$, $\sigma_{c,AOD}$, from an exponential distribution.
3. Choose the AOA and AOD for each path in each cluster from a normal distribution with the mean angle and spread chosen for the cluster.
4. To determine the relative power of each cluster, first choose auxiliary variables, U_c and Z_c, from uniform and Gaussian distributions, respectively. The relative power of the c-th cluster is then computed as $U_c^{r_t-1}10^{0.1Z_c}$, where r_t is a constant corresponding to the mean propagation delay. (This model is based on the observation that clusters have powers which decay exponentially with the mean propagation delay. [11] uses $U_c^{r_t-1}$ to capture the exponential relationship and $10^{0.1Z_c}$ to introduce log-normal variations.)
5. Normalize the total power over all the clusters to 1 and scale by the omnidirectional path loss to obtain the expected power of each cluster.
6. Choose the complex gain of each path from a circularly normal distribution with the expected power of the associated cluster.

The major simplification adopted in this work is the choice of a fixed omnidirectional path loss. In the model given by [11], the path loss (PL) is randomly generated to capture different operating scenarios. For example, LOS and NLOS scenarios yield significantly different path

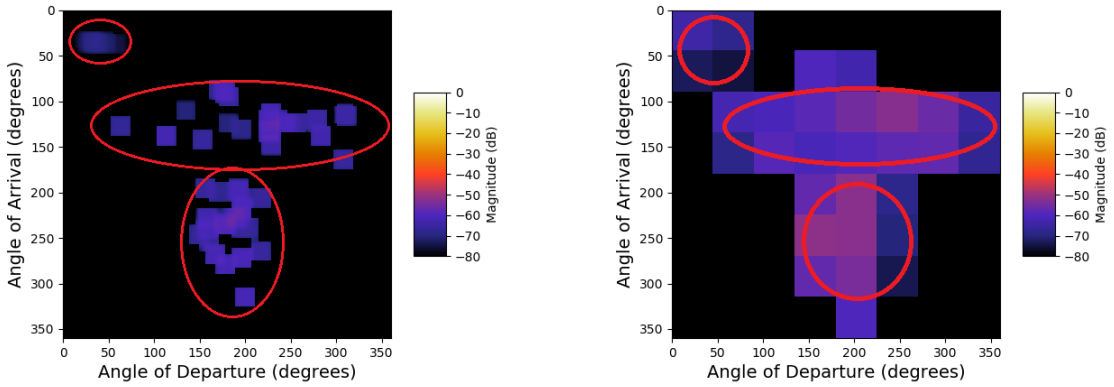
[1]Information regarding the channel modeling work of NYU Wireless can be found at: http://wireless.engineering.nyu.edu/mmwave-channel-modeling

Variable	**Value**	**Parameters**
Number of Clusters	$C \sim \text{Poisson}(\lambda)$	$\lambda = 1.8$
Number of Paths per Clusters	$M \sim \text{unif}(a, b)$	$(a, b) = (10, 40)$
AOD and AOA Angle Spread	$\sigma_{c,AOD}, \sigma_{c,AOA} \sim \exp(\lambda_h)$ $_{(c=1,\ldots,C)}$	$\lambda_h^{-1} = 10^\circ$
AOD and AOA Path Angles	$\phi_{c,m} \sim \mathcal{N}(\phi_c, \sigma^2_{c,AOD})$ $\theta_{c,m} \sim \mathcal{N}(\theta_c, \sigma^2_{c,AOA})$ $_{(c=1,\ldots,C;\quad m=1,\ldots,M)}$	$\phi_c, \theta_c \sim \text{unif}(0, 360^\circ)$
Omnidirectional Path Loss (dB)	$PL = 110$	
Cluster Power Fraction	$\alpha'_c = U_c^{r_t - 1} 10^{0.1 Z_c}$ $_{(c=1,\ldots,C)}$	$U_c \sim \text{unif}(0, 1)$ $Z_c \sim \mathcal{N}(0, \zeta^2)$ $r_t = 2.8, \zeta = 4.0$
Cluster Power	$\alpha_{c,m} \sim \mathcal{CN}(0, \alpha_c 10^{-0.1 PL})$	$\alpha_c = \frac{\alpha'_c}{\sum_{c=1}^{C}(\alpha'_c)}$

Table 4.1: Channel Model Parameters

loss in a mmWave channel. There could also be outage scenarios where there is no path with significant power. However, beam training is only concerned with the *relative* power differences between clusters, and training is futile when no link can be supported by the channel. Thus, variations in the path loss can be reasonably ignored when comparing beam training algorithms. A fixed path loss of 110 dB is chosen to be consistent with the average path loss measured by channel sounding at separation distances less than 100m [11].

Figure 4.1 shows an example of a non-ideal channel that can be generated using the proposed method. Each colored square in Figure 4.1a corresponds to a single propagation path. In this particular channel, there are three clusters. The cluster centered around



(a) (b)

Figure 4.1: Example of a non-ideal channel for TX and RX arrays with 32 antenna elements: **(a)** in "continuous" space and **(b)** as a DSCM with $R = 8$. The three clusters are circled in red.

$(40°, 30°)$ has a very small AOA and AOD spread. However, it is associated with extremely high path loss. The cluster centered around $(125°, 200°)$ is associated with less path loss, but has an extremely large AOD spread. The cluster centered around $(250°, 200°)$ has comparable path loss to the second cluster and a significant amount of spread in both AOA and AOD. These cluster characteristics reflect a possible NLOS operating scenario.

The dominant cluster assumption is not valid for this channel, as can be seen in the DSCM representation in Figure 4.1b. There are several sector pairs with comparable channel gain, so the proposed BOA algorithms would likely not perform well for this channel.

4.2.2 Non-Ideal Beam Patterns

BPVs described in Definition 2 model non-ideal characteristics that are present in real-world beam patterns. Thus, the mapping of ideal BBPVs to non-ideal BPVs is achieved by convolving with a length-3 template whose elements are randomly generated to represent

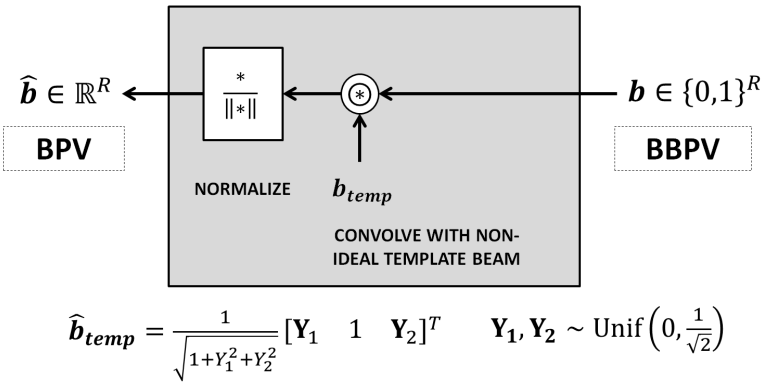


Figure 4.2: Conversion of ideal BBPVs into non-ideal BPVs.

sidelobes - as summarized by the block diagram in Figure 4.2. The procedure consists of two steps:

1. Each BBPV is circularly convolved with a template beam pattern that corresponds to a beam with a main lobe covering a single sector (magnitude 1 element) and non-zero side lobes in the adjacent sectors. The side lobe amplitudes are randomly chosen from a range of 0 to $\frac{1}{\sqrt{2}}$, relative to main lobe amplitude.
2. After side lobes are introduced, the resulting vector is normalized to unit Euclidean norm to satisfy the power constraint on BPVs.

Longer, more structured beam pattern templates can be used to better model side lobe patterns. However, for the purposes of this evaluation, a simple model is sufficient for capturing the impact of side lobes on beam training performance. Figure 4.3 illustrates how the beam pattern shapes are altered when the BBPVs for 1-BOA and 2-BOA are convolved with the given beam pattern template.

Beam pattern distortions caused by side lobes aggravates the effect of thermal noise since it introduces increased variations to the amplitude gain of each sector. Hence, the normalization process is expected to have the most impact on beam training performance,

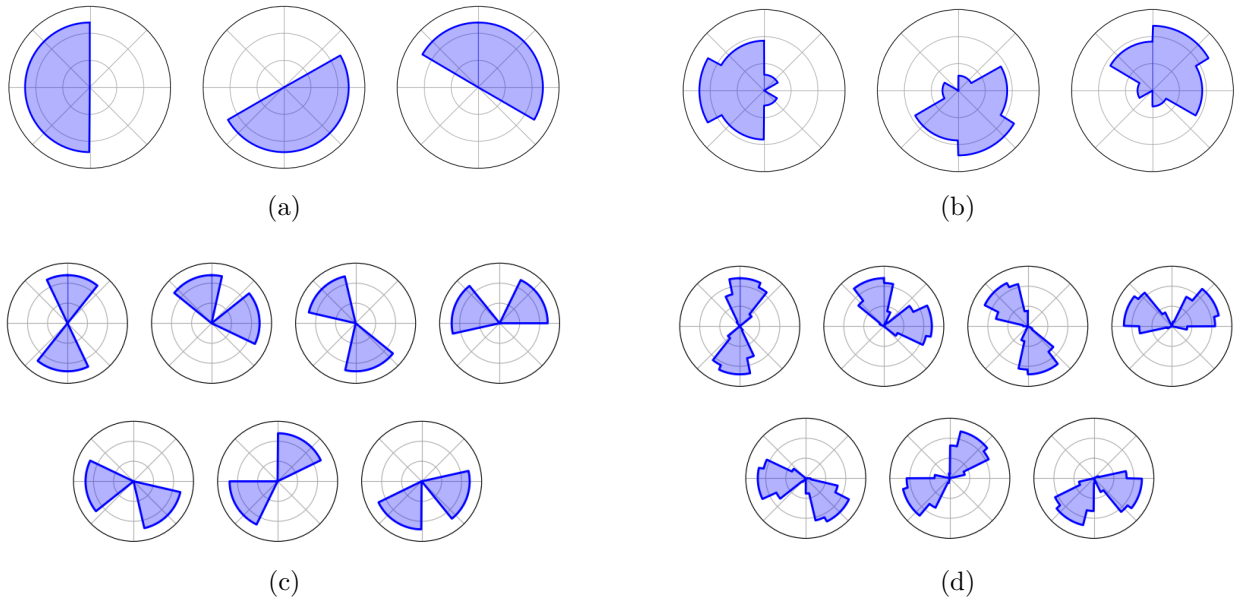


Figure 4.3: Beam patterns for ideal 1-BOA **(a)**, non-ideal 1-BOA **(b)**, ideal 2-BOA **(c)**, and non-ideal 2-BOA **(d)**.

since the reliable distinction of a 1 from a 0 in the Boolean model depends on the main lobe power gain of each beam pattern (see Section 4.2.3). In the absence of normalization, the main lobe power gain is always one. When the beam pattern is normalized, it is reduced by a factor roughly equal to the number of sectors covered by the main lobes.

4.2.3 Non-Ideal Channel Probes

Using BPVs and the DSCM, the measurements gathered from Q channel probes can be expressed as a real vector, $\mathbf{\Gamma} \in \mathbb{R}^Q$, where each element is a noisy measurement of a beamformed channel power gain as described in Definition 7. $\mathbf{\Gamma}$ must be mapped into the binary vector, $\bar{\mathbf{\Gamma}} \in \{0,1\}^Q$, which is accomplished by applying a threshold on the estimated

channel power gain:

$$\bar{\mathbf{\Gamma}}[q] = \begin{cases} 1, & \text{if } \mathbf{\Gamma}[q] \geq V_{thresh} \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

where V_{thresh} needs to be appropriately chosen.

Threshold Estimation

To facilitate analysis, we assume multiple measurements are taken per each channel probe, and the final metric used for thresholding is computed as the average of the measurements. This is a reasonable assumption since multiple training sequences are used for channel estimation in each PHY frame of 802.11ad [17]. By the Central Limit Theorem, each element in $\mathbf{\Gamma}$ can then be approximated as a random Gaussian variable with an unknown variance, σ^2. From the dominant cluster assumption, the mean can either be a low value, μ_-, or a high value, μ_+. A value of μ_+ corresponds to the probes in which the the sector of the dominant cluster is is covered by the main lobe of probing beam pattern. A value of μ_- corresponds to probes in which the sector of the dominant cluster is is not covered by the main lobe of the probing beam pattern. Note that μ_- can be non-zero since the beam patterns are non-ideal.

Under these assumptions, the information gathered from each channel probe will be modeled as one of two events:

$$\begin{aligned} &\text{E0: } \mathbf{\Gamma}[q] \sim \mathcal{N}(\mu_-, \sigma^2) \\ &\text{E1: } \mathbf{\Gamma}[q] \sim \mathcal{N}(\mu_+, \sigma^2) \end{aligned} \tag{4.6}$$

Ideally, all probes corresponding to event E1 should be converted to a binary value of 1 and all probes corresponding to E0 should be converted to a binary value of 0.

Let E_q denote the event that occurs for the q-th channel probe. Since the receiver has no prior information about the channel, the channel probes can be treated as IID with prior

probabilities of 0.5 for E1 and E0. The optimal threshold is given by:

$$V_{thresh} = \underset{d}{\text{argmin}}\ 0.5\left(\Pr\left\{\mathbf{\Gamma}[q] < d | E_q = \text{E1}\right\} + \Pr\left\{\mathbf{\Gamma}[q] > d | E_q = \text{E0}\right\}\right) \tag{4.7}$$

For any choice of q.

Due to the symmetry of the Gaussian distributions, the optimal threshold is simply $V_{thresh} = \frac{1}{2}(\mu_+ + \mu_-)$. However, μ_+ and μ_- are not known and need to be estimated. This can be accomplished via a bootstrap approach.

Since $\mu_+ > \mu_-$, and the noise variance is the same for the two possible distributions, we have:

$$\Pr\left\{\mathbf{\Gamma}[q] > k | E_q = \text{E1}\right\} > \Pr\left\{\mathbf{\Gamma}[q] > k | E_q = \text{E0}\right\}, \quad \forall k, q$$

This implies:

$$\begin{aligned} q_+ &= \underset{q}{\text{argmax}}\ \Pr\left\{E_q = \text{E1}\right\} = \underset{q}{\text{argmax}}\ \mathbf{\Gamma}[q] \\ q_- &= \underset{q}{\text{argmax}}\ \Pr\left\{E_q = \text{E0}\right\} = \underset{q}{\text{argmin}}\ \mathbf{\Gamma}[q] \end{aligned} \tag{4.8}$$

where q_+ and q_- are the indices of the channel probes most likely to correspond to events E1 an E0, respectively.

Thus, $\mathbf{\Gamma}[q_+]$ and $\mathbf{\Gamma}[q_-]$ can be used as initial estimates for μ_+ and μ_-, yielding an estimate of the optimal threshold as $V'_{thresh} = \frac{1}{2}(\mathbf{\Gamma}[q_+] + \mathbf{\Gamma}[q_-])$. A better estimate of the optimal

threshold can then be obtained in a recursive manner by computing:

$$\begin{aligned} V_{thresh} &= \frac{1}{2}\left(\frac{1}{|\mathcal{S}_+|}\sum_{j\in\mathcal{S}_+}\mathbf{\Gamma}[j] + \frac{1}{|\mathcal{S}_-|}\sum_{k\in\mathcal{S}_-}\mathbf{\Gamma}[k]\right) \\ \mathcal{S}_+ &= \{q \text{ s.t. } \mathbf{\Gamma}[q] \geq V'_{thresh}\} \\ \mathcal{S}_- &= \{q \text{ s.t. } \mathbf{\Gamma}[q] < V'_{thresh}\} \end{aligned} \tag{4.9}$$

Error Probability

Even with an optimal threshold and an ideal channel, there is a non-zero probability that the measurement from a noisy channel probe will be converted into an incorrect bit for the Boolean model. The proposed BOA algorithms have columns that are only Hamming distance $d_H = 1$ away from their nearest neighbors. Thus, the bit error probability is equivalent to the probability of misidentifying the sector corresponding to the dominant cluster, and will greatly impact beam training performance.

The bit error probability is given by:

$$p_{be} = 0.5\left(\Pr\left\{\mathbf{\Gamma}[q] < V_{thresh} | E_q = \text{E1}\right\} + \Pr\left\{\mathbf{\Gamma}[q] > V_{thresh} | E_q = \text{E0}\right\}\right) \tag{4.10}$$

From (4.6) and $V_{thresh} = \frac{1}{2}(\mu_+ + \mu_-)$, this can be evaluated as:

$$p_{be} = \mathcal{Q}\left(\frac{(\mu_+ - \mu_-)}{2\sigma}\right) \tag{4.11}$$

where $\mathcal{Q}$ is the single-sided Gaussian tail distribution function.

Under the dominant cluster assumption, μ_+ and μ_- are proportional to the main lobe and side lobe power gain of the beam pattern used for each channel probe. Since side lobes

cannot exceed half the power gain of a main lobe, this implies $(\mu_+ - \mu_-) \geq \frac{\mu_+}{2}$ and therefore:

$$p_{be} \leq \mathcal{Q}\left(\frac{\mu_+}{4\sigma}\right) \tag{4.12}$$

Thus, the bit error probability introduced by non-ideal channel probes is expected to increase when the main lobe power gain for each sector is decreased.

To quantify this effect, we can make a rough approximation of the Gaussian tail distribution function:

$$\mathcal{Q}(x) \approx \beta e^{-\alpha x^2} \tag{4.13}$$

where α and β are constants. (This approximation can be verified by noting $\frac{\mathrm{d}}{\mathrm{d}x}\left(1 - \beta e^{-\alpha x^2}\right) = 2\alpha\beta x e^{-\alpha x^2}$, which, for small positive x, matches the form of the standard Gaussian distribution.)

Using this approximation, the negative logarithm of the maximum bit error probability scales roughly with the the main lobe power gain. This implies if the the main lobe power gain is scaled by a factor of k, the maximum bit error probability would roughly scale by a power of k^2:

$$\mathcal{Q}\left(\frac{k\mu_+}{4\sigma}\right) \approx \mathcal{Q}\left(\frac{\mu_+}{4\sigma}\right)^{k^2} \tag{4.14}$$

4.3 Simulation

In Section 4.2, models for various non-ideal factors were introduced and and their potential implications on beam training performance were discussed. The effect of these factors will now be directly examined through simulation.

4.3.1 Simulation Setup

Figure 4.4 summarizes the simulation setup. At the TX, each BBPV (ie. row) of the binary measurement matrix, $\mathbf{\Psi}$, is "de-idealized" into BPVs using the process described in Section

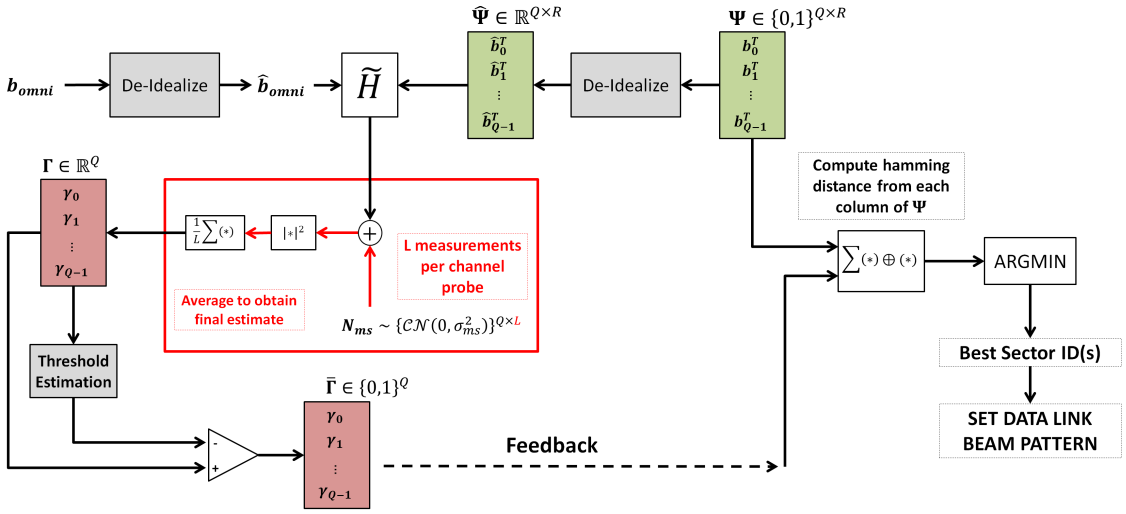


Figure 4.4: Simulation block diagram.

4.2.2. This produces a non-ideal measurement matrix, $\hat{\boldsymbol{\Psi}}$ composed of real-valued elements. At RX, an omnidirection BBPV (ie. vector of all 1s) undergoes the same process.

A non-ideal DSCM, $\tilde{\mathbf{H}}$, is generated using the method described in Section 4.2.1 and Definition 5. The number of antennas is set to $N_t = 64$, which is a representative size for 5G mmWave arrays [18, 19]. The discretization resolution, R, is determined by the number of columns in the measurement matrix being evaluated. Each channel probe measurement is computed as:

$$\gamma_q = \frac{1}{L}\sum_{l=1}^{L}\left|\hat{\boldsymbol{\Psi}}[\mathbf{q},:]\tilde{\mathbf{H}}^{\intercal}\hat{\mathbf{b}}_{omni} + N_{ms,l}\right|^2 \tag{4.15}$$

Where $N_{ms,l}$ is the IID circularly Gaussian noise added to the l-th measurement in the channel probe, and $\tilde{\mathbf{H}}^{\intercal}\hat{\mathbf{b}}_{omni}$ is the effective non-ideal AOD channel vector. The noise power is dynamically computed for each channel matrix to maintain the desired omnidirectional SNR (see Section 4.3.2). A total of $\mathbf{L = 5}$ measurements are conducted for each channel probe.

The estimated channel power gains for the Q channel probes are represented as a vector, $\mathbf{\Gamma} \in \mathbb{R}^Q$. $\mathbf{\Gamma}$ is converted into a vector of Q bits, $\bar{\mathbf{\Gamma}} \in \{0,1\}^Q$, via thresholding. Threshold estimation is conducted using the method described in Section 4.2.3.

From $\bar{\mathbf{\Gamma}}$, TX computes the best AOD sector by finding the column of $\mathbf{\Psi}$ that is closest, in hamming distance, from $\bar{\mathbf{\Gamma}}$. Due to noise, there may be more than one column which achieves the minimum hamming distance. In this case, it is assumed that there are more than one best sector. The beam pattern used for the data link is selected by choosing a BBPV with 1s located only at the best AOD sector indices.

To simulate receiver beam training, the same process is repeated except the measurement matrix is used by RX while TX adopts the data link beam pattern determined from transmit training. A total of 500 Monte Carlo trials are conducted for each algorithm to produce simulation results.

4.3.2 Evaluation Method

The goal of beam training is to achieve high channel power gain for a data link. Since omnidirectional beam patterns can be adopted without any beam training, the appropriate performance metric would therefore be the *increase* in beamformed channel power gain relative to the omnidirectional baseline.

Beam training performance is considered with respect to different levels of thermal noise. Since the impact of noise is dependent on channel and beam pattern characteristics, it is difficult to quantify directly. However, the relative significance of noise can be captured by the SNR of an omnidirectional measurement.

Let $\hat{\mathbf{b}}_{omni} \in \mathbb{R}^Q$ denote an omndirectional BPV. When $\hat{\mathbf{b}}_{omni}$ is used by both TX and RX

to make a channel measurement, the estimated beamformed channel response is given by:

$$\hat{h}_{omni} = \sqrt{P_t} h_{(\hat{\mathbf{b}}_{omni}, \hat{\mathbf{b}}_{omni})} + \mathbf{N_{ms}}$$

The SNR of this omnidirectional measurement is:

$$SNR_{omni} = \frac{P_t \left| h_{(\hat{\mathbf{b}}_{omni}, \hat{\mathbf{b}}_{omni})} \right|^2}{\sigma_{ms}^2} \tag{4.16}$$

For a general beam patterns, the measurement SNR is given by:

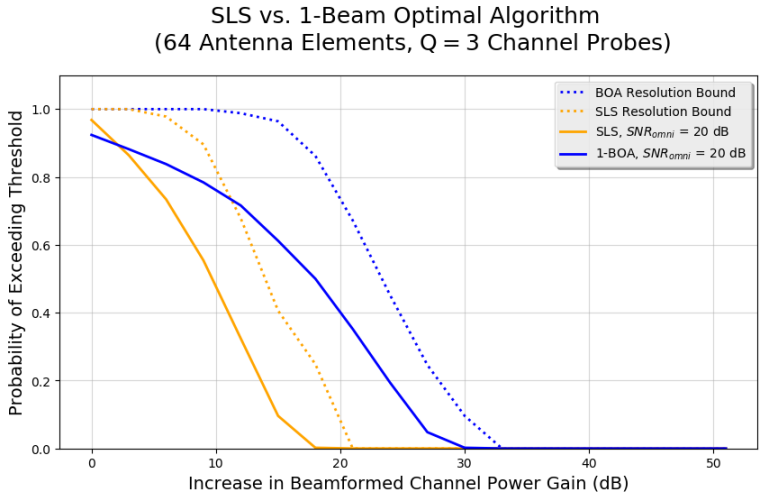
$$SNR_{ms} = \frac{P_t \left| h_{(\hat{\mathbf{b}}_{RX}, \hat{\mathbf{b}}_{TX})} \right|^2}{\sigma_{ms}^2} = \frac{\left| h_{(\hat{\mathbf{b}}_{RX}, \hat{\mathbf{b}}_{TX})} \right|^2}{\left| h_{(\hat{\mathbf{b}}_{omni}, \hat{\mathbf{b}}_{omni})} \right|^2} SNR_{omni} \tag{4.17}$$

Thus, the SNR associated with any estimated beamformed channel response is proportional to the baseline SNR of an omndirectional measurement. (Channel probe measurements are actually the squared magnitude of the estimated channel response the but the noise levels scale accordingly.)

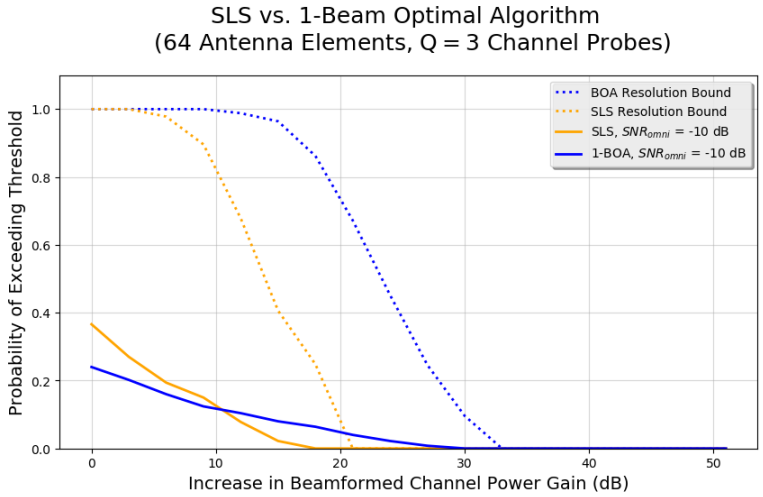
4.3.3 Simulation Results

Figure 4.5 summarizes the simulation results comparing 1-BOA with SLS using 3 channel probes. Figure 4.6 summarizes the simulation results comparing 2-BOA with SLS using 7 channel probes. The plots show the probability of achieving an increase in channel power gain greater than a particular threshold. A threshold of 0 dB indicates an achieved channel power gain at least equal to that of an omnidirectional configuration adopted at both TX and RX. Thus, achieving more than the 0 dB threshold implies beam training success while failing to reach the 0 dB threshold implies beam training failure. Two representative SNR

levels are shown to illustrate the impact of thermal noise. The dotted lines represent the theoretically achievable bound obtained by directly computing the largest singular value of the DSCM. The solid lines mark the performance of the algorithms in non-ideal conditions.

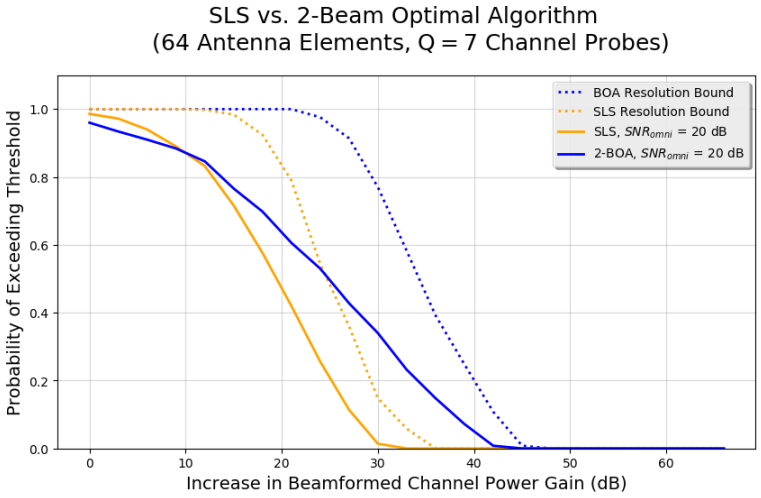


(a)

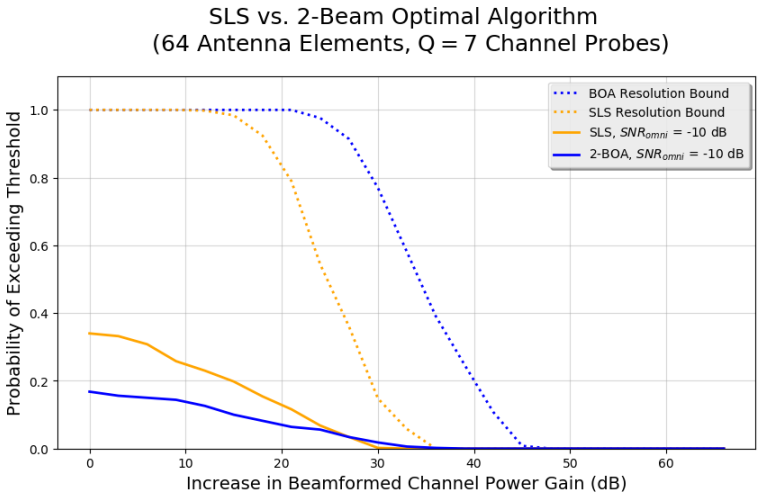


(b)

Figure 4.5: Simulation results for 1-BOA (blue) vs. SLS (orange): **(a)** high SNR ($SNR_{omni} = 20dB$), **(b)** low SNR ($SNR_{omni} = -10dB$) scenarios.



(a)



(b)

Figure 4.6: Simulation results for 2-BOA vs. SLS: **(a)** high SNR ($SNR_{omni} = 20dB$), **(b)** low SNR ($SNR_{omni} = -10dB$) scenarios.

4.4 Discussion of Results

It is clear that both BOA algorithms dominate SLS in high SNR scenarios, with much higher probability of achieving greater increase in beamforming gain. However, the improvements

are not as definitive in low SNR scenarios. This makes sense since the high SNR scenario is closer to the assumptions of the ideal model.

However, the improvements achieved by the BOA algorithms come at a cost: The probability of beam training success (ie. probability of exceeding 0 dB) for BOA is consistently lower than that of SLS. This is due to the power constraint on each channel probe.

The power used to perform each channel probe is fixed to one unit. From (3.14) and (4.3), the beam patterns used by 1-BOA are 1.5 times wider than those used by SLS. So the main lobes used by 1-BOA probes have roughly $\frac{2}{3}$ the power gain as those used by SLS. Likewise, the the beam patterns used by 2-BOA have roughly $\frac{2}{5}$ the power gain of SLS since more angular area is being covered by each beam. In Section 4.2.3, we showed that a reduced main lobe power gain translates into a greater probability of misidentifying the optimal sector, which, under the dominant cluster assumption, is equivalent to beam training failure.

The explanation can be validated by applying the scaling relationship given in Equation (4.14) to the low SNR [2] results. For 1-BOA, the probability of failure for SLS is roughly 0.6. The 1-BOA main lobe power gain is roughly 0.66 that of SLS, so the probability of failure for 1-BOA is upper bounded by $0.6^{(0.66^2)} = 0.8$, which agrees with the value in the simulation results. The probability of failure for 2-BOA can be similarly verified.

The trade-off between the probability of beam training success and the beamforming gain achieved from successful training is *fundamental*. To achieve higher beamforming gain, the beam training resolution must be increased. This means an increased number of sectors that must be distinguished. When the number of channel probes and the power used for each channel probe is fixed, this necessarily means less power, on average, is used to encode the channel information of each sector.

[2] The tail function approximation does not hold for high SNR scenarios

This trade-off is reflective of the so-called "no free lunch" principle in information theory. The process of obtaining CSI can be thought of as the removal of uncertainty from the channel. The total uncertainty removed by a given amount of work (ie. channel probes) is limited due to noise in the channel. Thus, either a smaller amount of uncertainty is removed from a larger number of sectors or a larger amount of uncertainty is removed from a smaller number of sectors. In noisy environments, the amount of uncertainty in each sector is high, so the penalty of reducing the work spent on each sector is greater.

Beam training optimality in non-ideal environments is therefore highly dependent on the noise level. Algorithms which perform well in high SNR environments necessarily do worse in when the noise levels are increased. Likewise, algorithms which perform well in low SNR environments necessarily become sub-optimal in high SNR environments. There is no globally optimal solution. However, if prior information about the channel SNR is known, training strategies may be chosen intelligently to obtain the best balance in the presence of this trade-off.

4.5 Chapter Summary

In this chapter, we investigated the impact of non-ideal factors on the real-world performance of algorithms developed using the Boolean model proposed in Chapter 3. Non-ideal models for the channel matrix, beam patterns, and channel measurements were used to simulate two algorithms that were deemed optimal in the ideal model. The simulation results showed that the proposed algorithms are superior to SLS in high SNR environments, but fare worse in low SNR scenarios. This is reflective of a fundamental trade-off in beam training performance resulting from the finite amount of power invested into channel probes.

Chapter 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this thesis, we considered the optimization of beam training within a simple Boolean model. In this model, a mmWave channel is treated as a binary vector containing a single non-zero element, and a beam training algorithms is a expressed as a binary matrix. Channel measurements consist of logical operations applied between the channel vector and the matrix. This abstraction enabled the derivation of an optimal bound on the algorithmic efficiency of beam training, and a general class of algorithms that achieve this bound was defined. A few of these algorithms where then evaluated through simulation to examine the extent to which optimality in the ideal model leads to performance gains in realistic environments. Simulation results showed that the proposed algorithms can outperform the existing 802.11ad standard when the measurement SNR is sufficiently high. However, there is a fundamental trade-off between the probability of beam training success and achieved beamforming gain. Thus, the 802.11ad standard becomes a better choice in low SNR environments.

5.2 Future Work

Future work should consider the extension of the Boolean model analysis to general K-sparse channels. This could potentially make use of results from group testing (see Appendix B. It is expected that the optimal bound on the training complexity will degrade when $K > 1$.

However, the resulting algorithms would be more robust to non-idealities in the channel due to increased hamming distance between columns.

More generally, optimization should also be extended to the higher MAC context. In this work, the scope of beam training was limited to the MAC architecture of DMG networks in 802.11ad. This led to specific constraints on the format in which channel probes must be conducted. If these constraints can be relaxed, it may be possible to bypass some of the fundamental trade-offs associated with non-adaptive training.

Finally, beam training algorithms developed in this work should be evaluated in real mmWave systems. While simulations can offer insight into the potential improvements offered by a proposed algorithm, the reliability of simulation results is highly dependent on the accuracy of the models used for simulation. Hardware implementations are the only way to gauge the true performance of theoretically "optimal" solutions.

BIBLIOGRAPHY

[1] L. Frenzel, "Millimeter waves will expand the wireless future," Online, March 2013, image. [Online]. Available: http://www.electronicdesign.com/communications/millimeter-waves-will-expand-wireless-future

[2] "Phased array animation with arrow," Online, Wikimedia Commons, 2016, image. [Online]. Available: https://commons.wikimedia.org/wiki/File:Phased_array_animation_with_arrow_10frames_371x400px_100ms.gif

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005, ch. 7.

[4] "Leading the IoT," Online, Gartner, 2017, e-Book. [Online]. Available: https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf

[5] F. C. Commission, "Use of spectrum bands above 24 ghz for mobile radio services," Federal Communications Commission, Tech. Rep., October 2017. [Online]. Available: https://docs.fcc.gov/public/attachments/DOC-347449A1.pdf

[6] V. Rabinovich and N. Alexandrov, *Antenna Arrays and Automotive Applications*. Springer, 2013, ch. 2,4.

[7] A. Molisch *et al.*, "Hybrid beamforming for massive mimo: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, September 2017.

[8] J. Zhang *et al.*, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4980–4995, November 2017.

[9] M. Cai, J. N. Laneman, and B. Hochwald, "Beamforming codebook compensation for beam squint with channel capacity constraint," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 76–80.

[10] M. Samimi and T. Rappaport, "3-d millimieter-wave statistical channel model for 5g wireless system design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2207–2224, July 2016.

[11] M. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164 – 1179, June 2014.

[12] *IEEE 802.11ad Amendment*, IEEE Std., July 2012.

[13] D. D. 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, June 2017.

[14] C. N. Manchn, E. de Carvalho, and J. B. Andersen, "Ping-pong beam training with hybrid digital-analog antenna arrays," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[15] J. Kim and A. F. Molisch, "Fast millimeter-wave beam training with receive beamforming," *Journal of Communications and Networks*, vol. 16, no. 5, pp. 512–522, Oct 2014.

[16] M.Aigner and G. M. Ziegler, *Proofs from the Book*, 4th ed. Springer, 2013, ch. 2.

[17] R. . Schwarz, "802.11ad - wlan at 60 ghz: A technology introduction," Rohde & Schwarz GmbH & Co. KG, Tech. Rep., 2017. [Online]. Available: https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_application/application_notes/1ma220/1MA220_3e_WLAN_11ad_WP.pdf

[18] A. Agrawal and A. Natarajan, "A scalable 28ghz coupled-pll in 65nm cmos with single-wire synchronization for large-scale 5g mm-wave arrays," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, January 2016, pp. 38–39.

[19] M. K. Ishfaq *et al.*, "8 × 8 phased series fed patch antenna array at 28 ghz for 5g mobile base station antennas," in *2017 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, September 2017, pp. 160–162.

[20] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, November 2009, pp. 1–6.

[21] C.-N. Hu *et al.*, "Design of a mm-wave microstrip antenna array," in *2015 International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM)*, November 2015, pp. 1–2.

[22] S. Hur *et al.*, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, October 2013.

[23] D. Malioutov and M. Malyutov, "Boolean compressed sensing: Lp relaxation for group testing," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3305–3308.

[24] C. Shangguan and G. Ge, "New bounds on the number of tests for disjunct matrices," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7518–7521, Dec 2016.

Appendix A

CHANNEL ESTIMATION WITH COMPLEMENTARY GOLAY CODES

IEEE 802.11ad utilizes Complementary Golay Pairs (CGP) for channel estimation [17]. A CGP consists two bipolar sequences with the property that the sum of their autocorrelations is a scaled discrete impulse function. Formally, if $(G_a^{(D)}, G_b^{(D)})$ is a CGP of order D, then:

$$\text{Corr}\left(G_a^{(D)}[n], G_a^{(D)}[n]\right)) + \text{Corr}\left(G_b^{(D)}[n], G_b^{(D)}[n]\right)) = 2D\delta[n] \tag{A.1}$$

Where Corr ()) is the discrete time correlation operator.

The autocorrelation property of CGPs make them useful for extracting the channel impulse response of frequency-selective channels. Due to the inverse relationship between bandwidth and time delay, a frequency-selective channel must have multipaths that extend over multiple symbol periods. Thus, the frequency-selective effects can be completely captured by a discrete time complex channel impulse response sampled at the baseband symbol rate.

Consider a noiseless baseband channel with discrete impulse response $h[n] \in C^{\tau}$, where τ is the channel delay spread in units of symbol periods. Suppose a CGP of order D is passed through this channel, the received signal would be:

$$y[n] = h[n] * \left(G_a^{(D)}[n] + G_b^{(D)}[n - D]\right)$$

where $*$ is the linear convolution operator. Correlating $y[n]$ with $G_a^{(D)}[n]$ and $G_b^{(D)}[n]$ yields:

$$\text{Corr}\left(G_a^{(D)}, y\right)) = y_a[n] = h[n] * \left(\text{Corr}\left(G_a^{(D)}, G_a^{(D)}\right)) + \delta[n-D] * \text{Corr}\left(G_a^{(D)}, G_b^{(D)}\right)\right)$$
$$\text{Corr}\left(G_b^{(D)}, y\right)) = y_b[n] = h[n] * \left(\text{Corr}\left(G_b^{(D)}, G_a^{(D)}\right)) + \delta[n-D] * \text{Corr}\left(G_b^{(D)}, G_b^{(D)}\right)\right)$$

If $D > \tau$, the channel estimate can then be computed as:

$$\frac{1}{2D}\left(y_a[0:D-1] + y_b[D:2D-1]\right) = h[n] \tag{A.2}$$

Thus, with a sufficiently long CGP pair, the full channel response can be recovered. In the absence of noise, the recovery is perfect.

In practice, there exists hardware implementations for the the above processes, making CGP channel estimation highly efficient [17].

Appendix B

BEAM TRAINING AS GROUP TESTING

Using the Boolean model, beam training optimization can be framed as a constrained **group testing** problem.

In group testing, the goal is to identify up to K faulty items out of N total through binary testing [23]. For each test, any subset of the N items can be chosen to be tested simultaneously, but the result only indicates if at least one of the tested item is faulty. In non-adaptive group testing, all of the tests must be predetermined and can be modeled as a binary matrix. Non-adaptive group testing can therefore be considered as a compressed sensing problem over Boolean algebra.

Problem 4. Non-Adaptive K-Sparse Group Testing Problem

Given $\mathbf{x} \in \{0,1\}^N$ is K-sparse, design a matrix, $\mathbf{A} \in \{0,1\}^{M \times N}$ such that is possible to recover $\mathbf{x}$ from $\mathbf{y} \in \{0,1\}^M$, where:

$$\mathbf{y} = \bigvee_{j=0}^{N-1} \mathbf{A}[:, j]\mathbf{x}[j], \quad \forall i = 0, \ldots, M-1 \tag{B.1}$$

and M is to be minimized with respect to N.

In the beam training context, $\mathbf{A}$ would be the measurement matrix, $\mathbf{x}$ would be the channel vector, and $\mathbf{y}$ would be the channel probe measurements.

Ideal solutions to non-adaptive group testing take the form of **disjunct** matrices [24].

Definition 13. A binary matrix, $\mathbf{A} \in \{0,1\}^{M \times N}$, is said to be **K-Disjunct** if the logical OR of any K columns do not contain any other column in the matrix. That is:

$$\begin{aligned} \nexists \quad & \mathcal{S} \subset \{0,\ldots,N-1\}, |\mathcal{S}| \leq K \text{ and } q \notin \mathcal{S} \text{ s.t:} \\ & \mathbf{A}[:,q] \vee \mathbf{u} = \mathbf{u} \\ & \mathbf{u} = \bigvee_{i \in \mathcal{S}} \mathbf{A}[:,i] \end{aligned} \tag{B.2}$$

This is because a K-disjunct $\mathbf{A}$ enables the efficient decoding of any K-sparse $\mathbf{x}$ from $\mathbf{y}$:

Theorem 3. If $\mathbf{A}$ is a K-disjunct group testing matrix and $\mathbf{x}$ is K-sparse, then $\mathbf{x}$ can be recovered from $\mathbf{y}$ as follows:

1. Initialize the $\mathbf{x}' = \{1\}^N$
2. Define the set $\mathcal{Z}_y = \{i \text{ st. } \mathbf{y}[i] = 0\}$
3. Set $\mathbf{x}'[n] = 0$ if $\exists m \in \mathcal{Z}_y$ s.t $\mathbf{A}[m,n] = 1$

Proof. It is sufficient to prove that $\mathbf{x}[n] = 0$ if and only if $\exists m \in \mathcal{Z}_y$ s.t $\mathbf{A}[m,n] = 1$.
If $\mathbf{x}[n] = 1$, then $\{m \text{ s.t. } A[m,n] = 1\} \cap \mathcal{Z}_y = \varnothing$. This follows directly from the definition of $\mathbf{y}[m]$. Thus, $\nexists n$ s.t $\mathbf{x}[n] = 1$ and $\exists m \in \mathcal{Z}_y$ s.t $\mathbf{A}[m,n] = 1$.
Let $\mathcal{S}_x = \{i \text{ st. } \mathbf{x}[i] = 1\}$. Since $\mathbf{A}$ is K-disjunct, there cannot exist $n \notin \mathcal{S}_x$ such that $\{m \text{ s.t. } \mathbf{A}[m,n] = 1\} \subseteq \{m \text{ s.t. } \mathbf{A}[m,k] = 1; k \in \mathcal{S}_x\}$. This implies $\exists m^* \notin \{m \text{ s.t. } \mathbf{A}[m,k] = 1; k \in \mathcal{S}_x\}$ s.t $\mathbf{A}[m^*,n] = 1, \forall n \notin \mathcal{S}_x$. Thus, $\mathbf{x}[n] = 0$ implies $\{m \text{ s.t. } A[m,n] = 1\} \cap \mathcal{Z}_y \neq \varnothing$. □

Theorem 3 is well-established in group testing theory and can be interpreted as the Boolean equivalent to the Restricted Isometry Property (RIP) of Euclidean compressed

sensing.

In the group testing framework, beam training optimization is equivalent to the design of a K-disjunct matrix with maximal column to row ratio subject to BBPV constraints applied to the rows of the matrix.

This is a difficult problem for general K, since there exists no efficient method of verifying the disjunctness of a matrix. Analysis is tractable for $K = 1$ since 1-disjunct is equivalent to having distinct columns. However, the number of conditions to check grows with $\binom{N}{K}$. Therefore, more sophisticated approaches need to be developed to consider disjunctness with respect to structural constraints in the measurement matrix.

Appendix C

ALGORITHMS

Algorithm 1 Inductive Construction of Optimal Measurement Matrix

Set $\tilde{\mathbf{\Psi}}' = \{0\}^{\tilde{Q}\times\tilde{R}}$
for $i = 0, \ldots, \tilde{R} - 1$ **do**
    Set $\tilde{\mathbf{\Psi}}'[:, i] = \tilde{\mathbf{\Psi}}[:, i] \oplus \tilde{\mathbf{\Psi}}[:, (i-1) \mod \tilde{R}$
end for
Set $R = 2N_c + \tilde{R}$
Set $Q = \tilde{Q} + 1$
Set $\mathbf{\Psi}' = \{0\}^{Q\times R}$
Choose $\mathcal{Y} \subseteq \{0, \ldots, \tilde{R} - 1\}, |\mathcal{Y}| = 2B_{max}$
Set $d = 0$
for $i = 0, \ldots, \tilde{R} - 1$ **do**
    Set $\mathbf{\Psi}'[1:, d] = \tilde{\mathbf{\Psi}}[:, i]$
    Set $d = d + 1$
    if $i \in \mathcal{Y}$ **then**
        Set $\mathbf{\Psi}'[0, d] = 1$
        Set $d = d + 1$
    end if
end for
Set $\mathbf{\Psi} = \{0\}^{Q\times R}$
Choose $\mathbf{\Psi}[:, 0] = \{0, 1\}^R$
for $i = 1, \ldots, R - 1$ **do**
    Set $\mathbf{\Psi}[:, i] = \mathbf{\Psi}[:, i-1] \oplus \mathbf{\Psi}'[:, i]$
end for
return $\mathbf{\Psi}$

Algorithm 2 Construction of Base Case Optimal Measurement Matrix

Compute $\mathcal{P}(3B_{max})$, the smallest prime larger than $3B_{max}$.
Set $Q = \mathcal{P}(3B_{max})$
Set $\mathbf{g} = \{0\}^{QB_{max}}$
for $i = 0, \ldots, QB_{max} - 1$ **do**
    Set $k = i \mod Q$
    Set $m = \frac{i-k}{Q}$
    Set $\mathbf{g}[i] = (\mathbf{g}[i-1] + m) \mod Q$
end for
Set $\mathbf{\Psi} = \{0\}^{Q\times 2QB_{max}}$
for $i = 0, \ldots, QB_{max} - 1$ **do**
    Set $\mathbf{\Psi}[q, 2i] = 1 \forall q \in \{\mathbf{g}[i : i+2]\}$
    Set $\mathbf{\Psi}[q, 2i+1] = 1 \forall q \in \{\mathbf{g}[i : i+3]\}$
end for
return $\mathbf{\Psi}$

VITA

Evan is a Master of Science student in the Department of Electrical Engineering at the University of Washington, Seattle. He welcomes your comments to `eding18@uw.edu`.