UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

School of Electronics and Computer Science

Intelligence-Aided Transceiver Design for Millimeter Wave Communications

by

K. Satyanarayana

Thesis for the degree of Doctor of Philosophy

Supervisors: *Dr. Mohammed El-Hajjar and Prof. Lajos Hanzo*

Next-Generation Wireless Group

January 2020

To the disembodied force that anchors me,
My Family and Sir Jagadesh Chandra Bose

University of Southampton

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
School of Electronics and Computer Science

Doctor of Philosophy

INTELLIGENCE-AIDED TRANSCEIVER DESIGN FOR MILLIMETER WAVE COMMUNICATIONS

by K. Satyanarayana

The millimeter wave (mmWave) frequency band spanning from 30 GHz to 300 GHz offers a tremendous potential frequency resource for future wireless communication systems in order to meet the ever-increasing capacity demand. However, an important challenge at mmWave frequencies is that they suffer from high propagation loss because of the oxygen absorption, rain-induced fading and foliage attenuation. In order to mitigate the propagation losses, directional transmission would be employed. Furthermore, owing to the high cost and high hardware complexity as well as power-hungry nature of analog-to-digital and digital-to-analog converters, a hybrid beamforming architecture relying on both analog beamforming and digital precoding is conceived. However, the impairments in the analog beamforming employed at the radio-frequency (RF) would limit the achievable rate of mmWave hybrid systems. Therefore, we enhance the achievable data rate of mmWave hybrid systems with the aid of diversity and beamforming schemes. More explicitly we conceive a dual-function hybrid beamforming transceiver, where both diversity and beamforming gains can be attained. This is achieved by separting the sub-arrays emerging from a full array by a sufficiently large distance so the correlation between the sub-arrays is minimum. Then to further enhance the data rate, we considered full-duplex communication at mmWave frequencies relying on hybrid beamforming, where we aimed for mitigating the self-interference (SI) by jointly designing the RF transmit and receive beamformer weights and the precoder as well as combiner matrices. The proposed solution preserves the signal's dimensionality, while mitigating the SI. We show that the proposed design is capable of reducing the SI by upto 30 dB, hence performing similarly to the hypothetical interference-free FD system.

However, we note that the aforementioned designs rely on the assumption of having perfect beam-alignment and channel impulse response knowledge. Therefore, in order to relax this assumption, we invoke machine learning tools for intelligent beam-alignment

based on the location of the user, where we demonstrate that our design performs similarly to beam-sweeping based beam-alignment relying on high-complexity exhaustive beam-search. Furthermore to counteract the channel aging phenomenon, we propose learning-assisted channel prediction. In order to reduce the signalling overhead and detector complexity at the receiver, we propose deep learning assisted semi-blind detection for index modulation mmWave MIMO systems. More particularity, we propose a detector for index modulation systems operating without relying on the explicit knowledge of the CSI at the receiver. We observe by our simulations that the number of computations required for learning-assisted soft-detection is four times lower than that of the conventional soft-detector.

Declaration of Authorship

I, K. Satyanarayana , declare that the thesis entitled *Intelligence-Aided Transceiver Design for Millimeter Wave Communications* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as shown in the list of publications.

Signed:

Date: .....

Publications

Accepted Papers

1. **K. Satyanarayana**, Mohammed El-Hajjar,Alain Mourad, Lajos Hanzo, *Deep Learning Aided Fingerprint Based Beam Alignment for mmWave Vehicular Communication* accepted in IEEE Trans. Veh. Technol., Aug'19.

2. **K. Satyanarayana**, Mohammed El-Hajjar, Alain Mourad, Lajos Hanzo, *Multi-User Full Duplex Transceiver Design for mmWave Systems Using Learning-Aided Channel Prediction* in IEEE Access May 2019.

3. **K. Satyanarayana**, Mohammed El-Hajjar, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Hybrid Beamforming Design for Full-Duplex Millimeter Wave Communication*, accepted in IEEE Trans. Veh. Technol., Nov'18.

4. **K. Satyanarayana**, Mohammed El-Hajjar, Alain Mourad, Lajos Hanzo, *Multi-User Hybrid Beamforming Relying on Learning-Aided Link-Adaptation for mmWave Systems*, IEEE Access, Feb. 2019.

5. **K. Satyanarayana**, Mohammed El-Hajjar, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Dual-Function Hybrid Beamforming and Transmit Diversity Aided Millimeter Wave Architectures*, in IEEE Trans. Veh. Technol., vol. 67, no. 3, pp. 2798 - 2803, 2017.

6. **K. Satyanarayana**, Mohammed El-Hajjar, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Adaptive Transceiver Design for C-RAN in mmWave Communications* in IEEE Access, pp. 16770 - 16782, 2018.

7. **K. Satyanarayana**, Teodor Ivanescu, Mohammed El-Hajjar, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Hybrid Beamforming Design for Dual-Polarized Millimeter Wave MIMO Systems*, in Electronics Letter, vol. 54, no. 22, 2018.

8. Ibrahim A. Hemadeh, **K. Satyanarayana**, Mohammed El-Hajjar, Lajos Hanzo, *Millimeter Wave Wireless Communications: Survey and Study of Radio Propagation Characteristics, Antenna Design and Channel Modeling* accepted in IEEE Commun. Surveys & Tuts., 2017

9. **K. Satyanarayana**, Mohammed El-Hajjar, Ping-Heng Kuo , Alain Mourad, Lajos Hanzo, *MBER Transmit Precoding for the Rank-Deficient MIMO-Aided Internet of Things*, accepted in IEEE PIMRC 2018.

10. Denisa Prisiceanum, **K. Satyanarayana**, Mohammed El-Hajjar, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Effects of Mutual Coupling on Lattice Reduction-Aided Millimeter Wave Hybrid Beamforming*, accepted in IEEE PIMRC 2018.

11. **K. Satyanarayana**, Mohammed El-Hajjar, Ping-Heng Kuo, Alain Mourad, Lajos Hanzo, *Millimeter Wave Hybrid Beamforming with DFT-MUB Aided Precoder Codebook Design*, in Proc. VTC, 2017.

12. Yichuan Li, Salman Ghafoor, **K. Satyanarayana**, Mohammed El-Hajjar, Lajos Hanzo, *Analogue Wireless Beamforming Exploiting the Fiber-Nonlinearity of Radio Over Fiber Based C-RANs* accepted in IEEE Transactions on Vehicular Technology, Jan'19.

Under Review

1. **K. Satyanarayana**, Mohammed El-Hajjar, Alain Mourad, Lajos Hanzo, *Deep Learning Assisted Semi-Blind Detection for Index Modulation Aided mmWave Systems*, submitted to TWC.

Under Preparation

1. **K. Satyanarayana**, Mohammed El-Hajjar, Alain Mourad, Lajos Hanzo, *Soft-Decoding for Multi-Set Space-Time Shift-Keying mmWave Systems: A Deep Learning Approach.*

Contents

Glossary

Mathematical Symbols and Notations

$\mathbf{A}$	Denotes matrix A
$\mathbf{a}$	Denotes vector a
$\sum_i x_i$	Denotes the sum of elements
$\|.\|_F$	The Frobenius Norm operator
$\mathrm{tr}(.)$	The trace operator
$\|.\|$	The modulus operator
$\|.\|_p$	The L_p norm operator
$\binom{n}{k}$	N Choose K
$(.)^H$	Hermitian transpose
$(.)^T$	Transpose operator
$\mathcal{O}(.)$	Complexity order
$\otimes$	Kronecker product operator
$\arg\max_{x\in\mathcal{X}}(.)$	Argument of $x_i \in \mathcal{X}$ that has maximum value
$\arg\min_{x\in\mathcal{X}}(.)$	Argument of $x_i \in \mathcal{X}$ that has minimum value
$P(Y/X)$	Probability of Y given X
$\log_2(.)$	Logarithm to base 2 operator
$\exp(.)$	Exponential function
$\mathcal{R}\{.\}$	Real part of a complex number
$\mathbb{R}^n$	n dimensional real number
$\mathcal{I}\{.\}$	Imaginary part of a complex number
π	$3.1415\ldots$
δ	Dirac delta function
$\mathcal{CN}$	Complex-valued normal distribution
σ^2	Variance of a random variable
i.i.d.	Independent and identically distributed
$\forall$	For all
$x \in \mathcal{X}$	x belongs to set $\mathcal{X}$

$\lfloor . \rfloor$	Flooring a real number to smallest integer
$[a, b]$	Set of real numbers between a and b

List of Symbols

$\mathbf{a}_r$	Rx array response vector
$\mathbf{a}_t$	Tx array response vector
bpcu	bits per channel use
B_c	Coherence bandwidth
C_{out}	Outage capacity
dB	Decibel
D	Distance
f	Frequency
f_d	Doppler frequency
f_p	Pilot density
$\mathbf{F_{BB}}$	Baseband precoder matrix
$\mathbf{F_{RF}}$	RF beamforming matrix
G_t	Gain of the Tx antenna
G_r	Gain of the Rx antenna
$\mathbf{H}$	Channel matrix
$\mathbf{H}_{\text{eff}}$	Effective channel matrix
$\mathbf{H}_i$	i^{th} sub-channel matrix
I_A	Apriori Information
I_E	Extrinsic Information
k (Ch.2)	Wave number
K (Ch.4)	Number of users
K (Ch.5)	Number of AEs on each AA at Tx
L	Number of AEs on each AA at Rx
m	meters
M	Size of AA
$\mathbf{n}$	Noise vector
N_0	Noise spectral density
N_t	Number of transmit antennas
N_r	Number of receive antennas
N_s	Number of symbols
N_{sub}	Number of sub-arrays
N_t^{RF}	Number of RF chains at Tx
N_r^{RF}	Number of RF chains at Rx
N_b	Number of beams
N_c	Number of clusters
N_d	Number of data symbols

N_f	Number of frames
N_p	Number of pilots
N_{ray}	Number of Rays
N_v	Number of vehicles
M_c	Constellation size
M_Q	Number of dispersion matrices
P_e	Probability of error
p_{out}	Outage probability
P_t	Transmit power
P_r	Received power
$\mathbf{s}$	Transmit symbol vector
T	Time Slots
T_c	Coherence Time
τ	Sample time
$\mathbf{U}$	Left Singular Matrix
$\mathbf{V}$	Right Singular Matrix
$\mathbf{W_{BB}}$	Baseband combiner matrix
$\mathbf{W_{RF}}$	RF combiner matrix
η	Efficiency
$\mathcal{S}_{\sigma_s}$	Shadowing factor
$\mathcal{S}$	Cardinality of the training samples
$\mathcal{CN}$	Complex-valued normal distribution
n	Number of sub-carriers
$\mathbf{y}$	Received signal vector
ψ	Wavenumber variable
ϕ_r	Angle of arrival at the Rx
ϕ_t	Angle of departure at the Tx
λ	Wavelength
ζ	Time correlation coefficient
σ_n^2	Noise variance
$\alpha_{n_c}^{n_{\text{ray}}}$	CIR of n_c^{th} cluster and n_{ray}^{th} ray

Abbreviations

1G	The First Generation
2G	The Second Generation
3G	The Third Generation
4G	The Forth Generation
5G	5^{th} Generation
3GPP	3^{rd} Generation Partnership Project
AA	Antenna Array

ABF	Analog Beamforming
ADC	Analog-to-Digital Converter
AE	Antenna Element
AF	Array Factor
AoA	Angle of Arrival
AoD	Angle of Departure
AMC	Adaptive Modulation and Coding
ANN	Artificial Neural Network
ASA	Array of Sub-Arrays
ASU	Antenna Selection Unit
AWGN	Additive White Gaussian Noise
BER	Bit Error Ratio
BF	Beamforming
BIM	Beam Index Modulation
BLAST	Bell-Labs Layered Space-Time
BP	Beam-Pair
bpcu	bits per channel use
BPSK	Binary Phase Shift Keying
BPTT	Back Propagation Through Time
BS	Base Station
CIR	Channel Impulse Response
C-RAN	Cloud Radio Access Network
CSI	Channel State Information
CU	Central Unit
DAC	Digital-to-Analog Converter
DCMC	Discrete-input Continuous-output Memoryless Channel
DFT	Discrete Fourier Transform
DL	Downlink
DM	Dispersion Matrix
EBF	Eigen Beamforming
ECR	Energy Consumption Ratio
FSL	Free Space Loss
eMBB	Enhanced Mobile BroadBand
EXIT	EXtrinsic Information Transfer
FD	Full Duplex
FDD	Frequency-Division Duplex
FP	Fingerprint
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HBF	Hybrid Beamforming
HPBW	Half-Power Beamwidth

KNN	K-Nearest Neighborhood
INR	Interference-to-Noise Ratio
LOS	Line-of-Sight
LLRs	Log-Likelihood Ratios
LO	Local Oscillator
LSSTC	Layered Steered Space-Time Coding
LTE	Long-Term Evolution
MF	Multi-Functional
MI	Multi-User Interference
MIMO	Multiple-Input Multiple Output
ML	Machine Learning
MMSE	Minimum Mean Squared Error
mmWave	Millimeter Wave
MS-STSK	Multi-Set STSK
MU	Multi-User
MUB	Mutually Unbiased Bases
NLOS	Non Line-of-Sight
NMSE	Normalized Mean Squared Error
NOMA	Non-Orthogonal Multiple Access
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
OMP	Orthogonal Matching Pursuit
OSTBC	Orthogonal STBC
PA	Power Amplifier
PDF	Probability Distribution Function
PL	Path Loss
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RRH	Remote Radio Heads
RSS	Received Signal Strength
Rx	Receiver
SCM	Spatial Channel Model
SI	Self-Interference
SIR	Signal-to-Interference Ratio
STBC	Space-Time Block Code
STSK	Space-Time Shift Keying
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
TDD	Time Division Duplex

TPC	Transmit Precoder
Tx	Transmitter
UE	User Equipment
UCA	Uniform Circular Array
UHF	Ultra High Frequency
ULA	Uniform Linear Array
UPA	Uniform Planar Array
V2V	Vehicular-to-Vehicular
ZF	Zero Forcing

Acknowledgements

My heartfelt gratitude to all my friends and peers in the wireless research community whose contributions — in big and little ways — to this fascinating field have laid the foundations of this thesis.

A pearl forming in an oyster needs the right chemicals, the right temperature and the right flow of the current to thrive. One can also say this about a student actively seeking knowledge. His/her hunger for knowledge is sated when met with: the good patronage, the right ambience and the necessary tools to flourish. These elements fortuitously came together in my pursuit of knowledge in wireless communications. I profusely thank InterDigital without whose generous patronage this feat would not have been possible. I also extend my thanks to the University of Southampton for providing me an edifying atmosphere and tools to carry out my research under the supervision of Prof. Lajos Hanzo and Dr. Mohammed El-Hajjar, who are beyond all that I can say.

The many ways, Lajos and Mohammed, changed and expanded my knowledge in wireless communications and are continuing to do so is the greatest of any and all gifts.

Very decidedly, the most cogent counsel came from Lajos, who is undoubtedly an aficionado and also, in my opinion, an aesthete. Lajos devotion to his profession is unwavering. My discussions with Lajos have enormously helped me navigate through the choppy waters of research smoothly. He 'being there' for me at every stage, whether do with research, or any piece of advice, was a boon that I have not recovered from. And despite his busy routine, he went through every draft with me meticulously working paragraph by paragraph to make sure I had every part true and right as I could.

I am eternally indebted to Mohammed, for not a day has gone without a knock on his door. My discussions with him hovered around 4Ps: Publications, Patents, Politics and, of late, Probability. I consider my sessions with him a truly blessed windfall of this entire project. It also bears stressing that he has been a constant and encouraging force who nurtured my understanding and provided succor when things went awry. Mohammed is astonishingly patient with me, unflinchingly listening to all my questions and complaints, for I have all the bad habits of a curmudgeonly young curious mind!

Thanks are also due to Alain Mourad from InterDigital and Ping-Heng Kuo (Wallace) who *was* also at InterDigital. Our weekly discussions with both Alain and Wallace have been instrumental and helped me build the nexus between industry and academia. More

particularly, these discussions have planted the seed of curiosity in knowing the modus operandi of wireless industry.

There are also scores of frontline people to whom I extend my gratitude, especially to Prof. Lie-Liang Yang, Prof. Sheng Chen, Dr. Soon (Michael) Ng, Dr. Rober Maunder. Equally, I thank my friends during my stay at the University of Southampton — Ibrahim Hemadeh, Yichuan Li, Siyao Lu, Ateeq Rehman, Abbas Ahmed, Daryus Chandra, Teodor Ivanescu, Chao Xu, Zeynep Kaykac Egilmez, Shruti Gupta, Karunakar — for their immense support.

A special mention to Charan Vala for engaging conversations spanning from existential to non-existential elements of realism. Due to close proximity to his house from the University, Charan has been a source of my nourishment by plying me with the delicious hand-cooked food! I also take this opportunity to thank my friends outside the University of Southampton, for whom the phrase '*a friend in need is a friend indeed*' becomes a byword: Abhinov Balagoni, Sana Sudheer, Murali Krishna.

Finally, I would like to express my warmest gratitude to my mother Lavanya Katla, to my father Kanaka Bhushanam Katla and to my sister Swapna Katla. I dedicate my thesis to the disembodied force that anchors me, to my family and to Sir Jagadish Chandra Bose.

No precedent in the history has seen the unprecedented success story as the wireless technology in grabbing billions of users — neither television nor the personal computer, or the washing machine and not even the omnipotent Internet!

Chapter 1

Introduction

The first transatlantic cable was seen as a folly and a hoax. But it heralded the greatest communications revolution in history — The Legend.

1.1 Introduction

Ever since Jagadeesh Chandra Bose has demonstrated the feasibility of electromagnetic wave transmission, wireless researchers have endeavored to fulfill the dream of flawless telepresence. Fig. 1.1 shows the evolution of wireless communication. The first commercial cellular phone has been introduced in early 1980s, which is commonly referred to as the first generation (1G) of mobile phones, where typically frequency modulation was used in the analog domain, as shown in Fig. 1.1. Following the success of 1G phones as well as the digital revolution, the wireless industry has moved to the 2G system known as Global System of Mobile Communications (GSM). The GSM was designed by international consortium of the European Union, but owing to its popularity, it rapidly spread across the entire globe. Developing the 3G systems was the result of a broader global collaboration within the third-generation partnership project (3GPP). This global standardization body has then also developed and ratified the 4G and 5G systems.

Nonetheless, given the explosive increase in the data rate demands as shown in Fig. 1.1, and the dearth of spectral resources in the sub-6 GHz band, harnessing millimeter wave (mmWave) frequencies for 5G and beyond has the benefit of large bandwidths resources in support of high data rates [1]. However, an important challenge in harnessing mmWave frequencies is that they suffer from high propagation losses because of the attenuation imposed by atmospheric absorption, foliage density and rain-induced fading [1]. To mitigate the propagation losses, typically directional transmission is employed, where large antenna arrays are used for attaining beamforming (BF) gains [2]. Conventionally, directional transmission is achieved by invoking digital signal processing s relying on

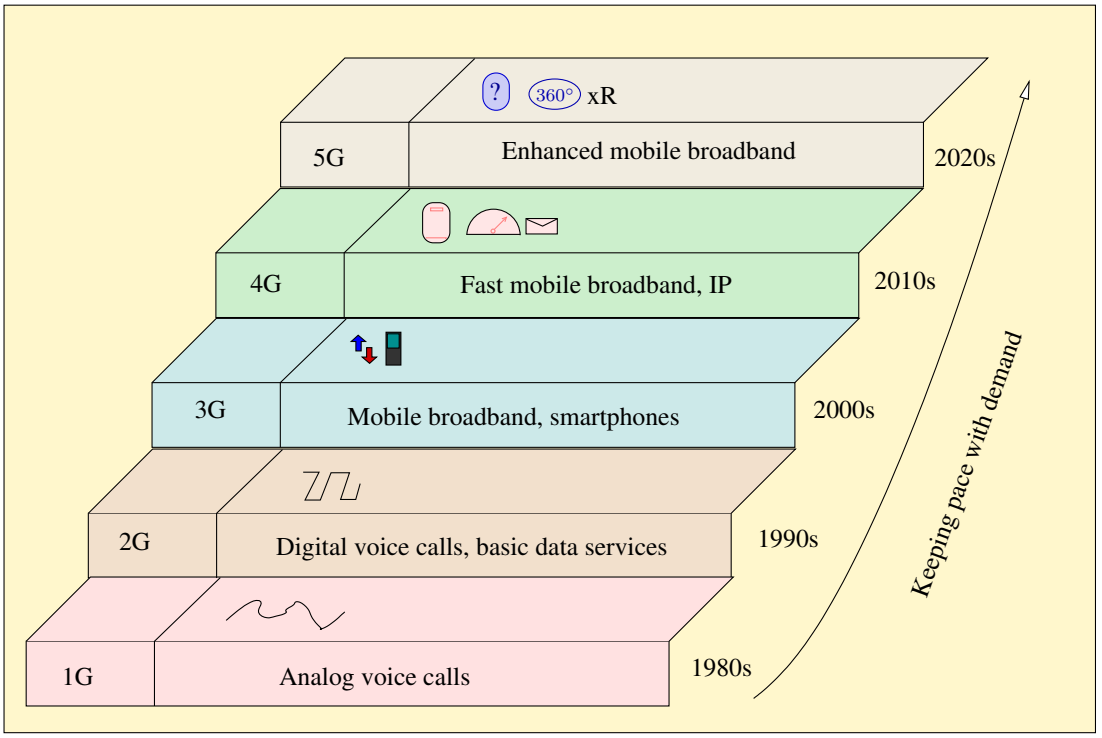


Figure 1.1: Evolution of wireless communication in keeping pace with the demand.

analog-to-digital/digital-to-analog converters (ADCs/DACs) for each of the RF chains. However, since large antenna arrays have to be employed at mmWave frequencies for attaining high BF gain, dedicating an ADC/DAC to each of the RF chains would impose high cost, complexity and power consumption. Hence, to circumvent the need for a large number of power hungry ADCs/DACs and to reduce the hardware complexity, a HBF design is conceived, where the signals generated by digital signal processing in the baseband relying on a few RF chains are fed to analog phase shifters in the RF stage before transmission from the antennas [3–5].

On the other hand, it is generally assumed that bidirectional radio communication in the same frequency band is not possible because of the high self-interference that results from the transmitter's own transmission at the local receiver which drowns out the low-power received signal. However, the research conducted in [6,7] proposed a solution to this challenge, where simultaneous transmission and reception in the same frequency were designed, which is termed as full-duplex (FD) wireless communication. Henceforth, the FD communication philosophy has attracted the attention of the wireless communication researchers, given its potential to double the spectral efficiency. However, the significant amount of self-interference (SI) at the local receiver tends to prevent FD communication from attaining satisfactory performance gains. Typically, the SI cancellation at the local receiver is carried out by a combination of passive and active methods [6,8]. The passive methods aim for increasing the path loss of the interfering signal through antenna isolation so as to reduce the SI power [9]. By contrast, the active methods rely

on the knowledge of the transmitted signal and they are generally carried out relying on three different approaches [10], namely radio-frequency (RF) cancellation, antenna cancellation, and digital cancellation. In the RF cancellation, the transmitted signal at known the local transmitter is used as a reference signal in the RF chain to reconstruct and then to subtract the SI. In the antenna cancellation, two replicas of the transmitted signal having opposite phases are generated using multiple transmit and receive antennas and the SI is removed by adding the two opposite-polarity replicas [11]. In contrast to RF and antenna cancellation, digital cancellation is typically employed together with the RF or antenna cancellation to further suppress the SI in the baseband [11, 12]. However, the benefits obtained by cascading RF/digital cancellation to remove the SI may remain limited because of the distortions at various stages, such as the power amplifier non-linearity, I/Q imbalance, and phase noise [3]. This is especially more pronounced in mmWave communications because of the high non-linearities of the ADCs/DACs at mmWave frequencies [13]. Hence, considering these non-linearities caused by the ADCs/DACs, power amplifiers, I/Q imbalance and phase noise, the residual SI would be quite high. In the state-of-the-art contributions on FD, most of the work has been focused on SI cancellation for MIMO relay systems [14–18], while some other related work includes dynamic resource allocation for FD systems [19]. More recently, Wang *et al.* [20] has derived a closed-form expression for achievable rates of FD MIMO relay systems communicating over Rician fading, when linear receivers are employed. Considering the limitations of the input circuitry, Day *et al.* derived upper and lowers bounds for the achievable sum rate of FD MIMO systems [21]. Everett *et al.* [22] studied the performance of passive self-interference suppression designed for FD, where the authors demonstrate that as high as 70 dB of SI suppression is possible in certain environments.

More recently, machine learning aided wireless transmission has gained attention owing to its more accurate predictions and superior performance over conventional methods dispensing with learning [23]. For example, the state-of-the-art designs in localization literature are predominantly focused on learning [24–26]. More particularly, learning based approaches in the localization have been shown to be capable of minimizing the location error. Wang *et al.* [27] demonstrated experimentally that a channel state information (CSI) based fingerprint relying on deep learning achieves a better performance than its counterparts. Capone *et al.* [28] employed a rudimentary learning technique for obstacle avoidance aided cell discovery during beam-alignment. An experimental study on indoor localization conducted by Chen *et al.* [29] attributed its superior performance to learning, where the authors invoke a convolutional neural network assisted learning scheme. Learning assisted algorithms have also been conceived for finding the most appropriate adaptive modulation and coding (AMC) mode. To increase the accuracy of link-adaptation, Daniels *et al.* [30] conceived a framework for overcoming the limitations of AMC aided MIMO-OFDM relying on supervised learning algorithms [31], such as the K-Nearest Neighbor (KNN) techinque. In addition to KNN, Daniels *et al.* [32] also presented an online AMC learning method, where support vector machines were employed.

In [33], link adaptation has been proposed for single carrier frequency domain equalization, again using the KNN algorithm. More recently, a broader class of machine learning algorithms, namely deep learning methods have been applied in both the context of indoor localization [34] as well as in detection [35].

Fig. 1.2 shows a typical mmWave transceiver chain illustrating the different processing stages spanning from baseband to the RF stage before transmission. In this thesis, we focus our attention mainly on baseband precoding, combining and RF beamforming, as shown in Fig. 1.2.

Furthermore, Fig. 1.3 shows the genealogy of the thesis. More particularly, it provides the map of the mmWave transceiver design pertaining to baseband precoding and analog RF beamforming. The focus of this thesis is on mmWave hybrid beamforming architectures and designs, as shown in Fig. 1.3. By invoking both conventional and sophisticated machine learning tools, we aim for designing mmWave transceivers that are both energy- and spectral-efficient. To elaborate further, in Chapters 3 and 4, we first propose hybrid beamforming designs for conventional mmWave hybrid systems, where we focus primarily on the blocks highlighted in Fig. 1.2. Then subsequently, in Chapters 5 and 6 we invoke machine learning tools to enhance the throughput of the system by circumventing channel estimation and detection.

In the next Chapter, we provide the a survey of both mmWave hybrid architectures and hybrid precoding designs, as well as self-interference cancellation methods for full duplex communication systems. Furthermore, as a pre-requisite, we present a short tutorial on machine learning followed by state-of-the-art wireless communication designs amalgamated with machine learning tools.

1.2 Contributions

The novel contributions of this thesis are summarized below. This work resulted in following publications [1, 4, 5, 36–38, 38–41].

1. A new hybrid BF architecture, where an antenna array is *partitioned* into two or more sub-arrays is conceived. In this architecture, the sub-arrays emerging from a full array are separated by a *sufficiently large* distance that minimizes the correlation between sub-arrays, say for example $\geq 5\lambda$, where λ is the wavelength, so that the channel of the sub-arrays becomes uncorrelated from one another. The rationale behind the sub-array separation is to attain diversity gains in addition to the BF gains provided by the sub-arrays.

2. We analyze the capacity of the proposed architecture and conclude that using *two sub-arrays* strikes an attractive compromise, since the diversity gain obtained from

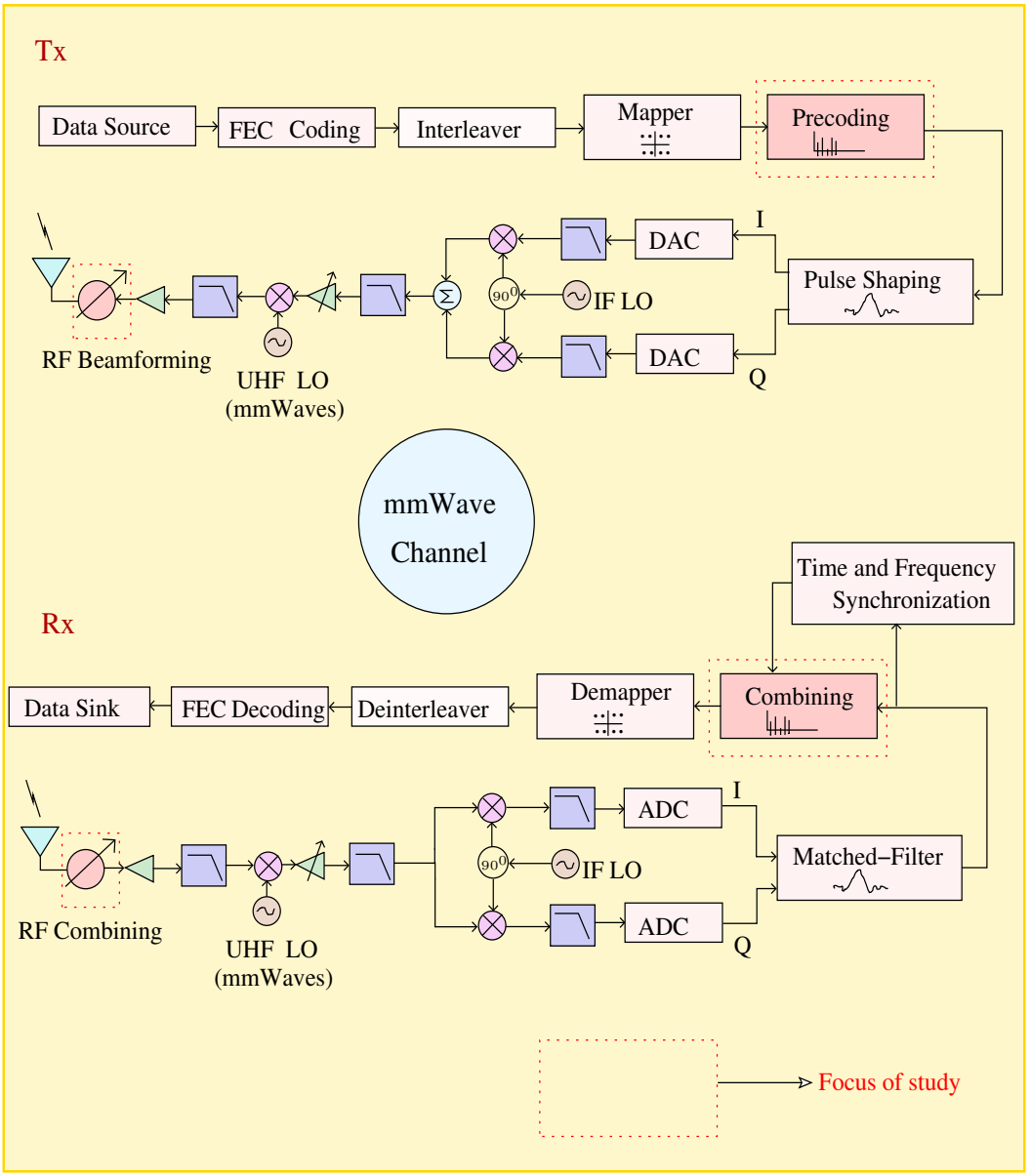


Figure 1.2: Block diagram of a transmitter and receiver.

the two sub-arrays out-weighs the loss in the BF gain imposed by splitting the array. By contrast, dividing the array into more than two sub-arrays would result in a reduction of BF gain, which the additional diversity gain attained cannot compensate for. We then analyze a codebook based on mutually unbiased bases (MUBs) amalgamated with DFT-based analog BF in the RF stage for the aforementioned mmWave MIMO systems.

3. We develop an adaptive array design for hybrid beamforming in mmWave communication that adapts the hybrid beamforming architecture as well as the digital precoding depending on whether the channel is of line-of-sight (LOS) or non-line-of-sight (NLOS) nature. Explicitly, we design an architecture relying on an adaptive array, where the array is fully-connected in LOS channel with an antenna spacing

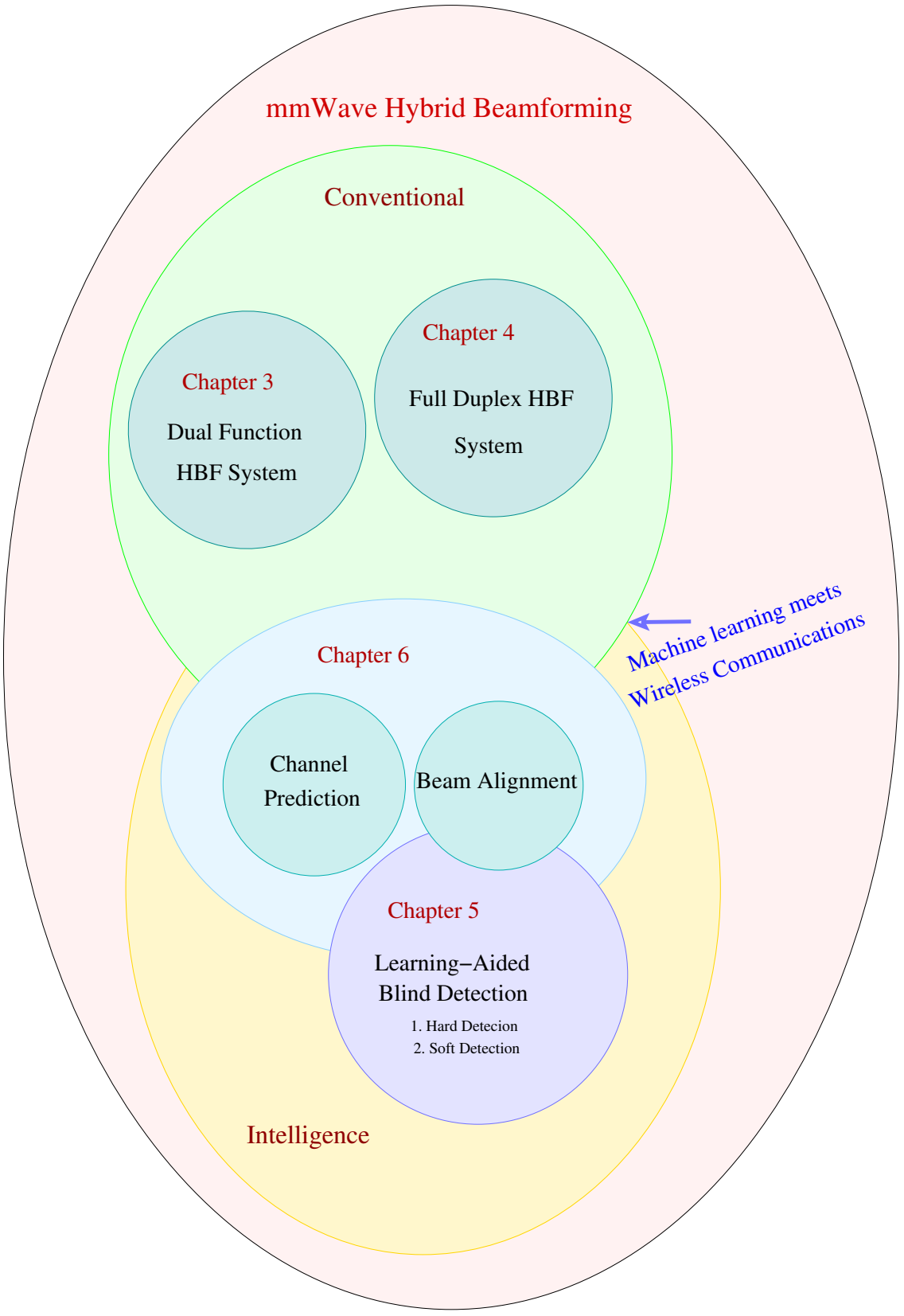


Figure 1.3: The map of this thesis.

of $\lambda/2$. By contrast, it is sub-array-connected where the antenna array is partitioned into sub-arrays having a sufficiently large separation so that each sub-array experiences independent fading in NLOS channel conditions.

4. We then utilize the adaptive array to propose adaptive virtual cell formation in the PHY layer of Cloud-Radio Access Networks (C-RAN) relying on mmWave hybrid beamforming, since the dynamic selection of remote radio heads (RRHs) is similar to a mmWave transmitter that adaptively switch between the fully-connected and array of sub-array (ASA) designs. In virtual cell formation, the user may maintain its connection with one or two RRH(s) depending on our proposed design criterion. We first present an algorithm conceived for virtual cell formation in support for a single mobile user, where the RRHs connected to the user experience different channel conditions (LOS/NLOS). Then we extend our algorithm to a multi-user scenario, where the mobile users are distributed randomly and experience different channel conditions (LOS/NLOS).

5. To further enhance the spectral efficiency, we consider FD communication at mmWave frequencies relying on hybrid beamforming, where we aim for mitigating the SI by jointly designing the transmit and receive RF beamformer weights and the precoder as well as combiner matrices. To design the beamformer, precoder and combiner matrices, we first obtain the fully-digital solution, where we employ an iterative algorithm relying on the idealized simplifying assumption of having perfect CSI. Then we derive the HBF solution from the digital solution obtained using least-squares approximation [42]. The proposed solution preserves the signal's dimensionality[1], while mitigating the SI. Based on our simulation results, we show that the proposed design is capable of reducing the SI by upto 30 dB, hence performing similarly to the interference-free FD system.

6. We then extend our FD hybrid beamforming design to K-user frequency selective interference channels, where the precoder and combiner are designed to minimize both the SI and multi-user interference (MI) in mmWave systems using beamforming. In this design, we aim for preserving the signal dimension, while mitigating both the SI and MI. Finally, we develop an iterative matrix decomposition for hybrid precoding aided OFDM systems, where the digital TPC weights are employed in the OFDM scheme's frequency-domain, while the analog RF beamformer weights are applied to the time-domain signal.

7. We propose a learning assisted adaptive transceiver regime based on the near-instantaneous post-processed SNR, where the adaptation switches between multiplexing versus diversity oriented transmission modes as well as by appropriately

[1] In contrast to spatial suppression [43], where the signal is projected into the null space of the interference, which may result in a reduced signal dimension, our design preserves the dimensionality of the signal subspace where the rank of the received signal is equal to the number of signal streams transmitted.

configuring the modulation employed so as to facilitate both low-BER and high-rate operation. The receiver relies on the instantaneous post-processed SNR to decide on the transmitter's multiplexing versus diversity aided transmission mode and on the choice of the specific modulation scheme to be employed with the aid of supervised learning relying on the feed-back information forwarded to the BS. In this paper, we invoke the KNN classification technique at the receiver for decision making, as a design example. We show through simulations that at a target BER of 10^{-3}, the learning-assisted adaptation scheme achieves a significantly higher throughput corresponding to an SNR gain of about 5 dB, while maintaining the required target BER compared to that of conventional link-adaptation carried out based on hard threshold values.

8. We propose deep learning assisted semi-blind detection for index modulation mmWave MIMO systems. More particularity, we propose a detector for multi-set space-time shift-keying (MS-STSK) without relying on the explicit knowledge of the CSI at the receiver. We also extend our design to a transmission scheme, where the MS-STSK is coupled with beam index modulation (BIM), whilst dispensing with CSI knowledge for detection. This philosophy makes our design spectral-efficient, since it eliminates the need for pilot-assisted channel estimation. We demonstrate by simulations that our proposed design detects the MS-STSK information with a high integrity while circumventing the CSI estimation. We also show by simulations that our design outperforms the ML-aided detection in the face of channel impairments introduced during the CSI estimation. Furthermore, we demonstrate that the net Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of the ML-aided detection is lower than that of our proposed learning-assisted detection. In other words, the DCMC capacity is significantly reduced by the pilot overhead.

9. To address the beam-alignment challenge in mmWave vehicular networks, we propose a multi-fingerprint based database, where the fingerprints are collected for different vehicular traffic densities in a given location. Then, the BS intelligently chooses the fingerprint based on the traffic density and location information, where we invoke deep learning for the selection of the fingerprint. Thereafter, the BS starts the training process to select the specific beam-pair from the fingerprint, which meets the received signal power target. The mobile station feeds back the index of the beam-pair from the selected fingerprint, provided it meets the threshold. This significantly reduces the search complexity involved. We demonstrate by our simulations that having multiple fingerprint-based beam-alignment provides a superior performance over the single fingerprint based beam-alignment. Furthermore, we show that our learning-aided multiple fingerprint design provides better fidelity than that of the scheme employing multiple fingerprints but dispensing with learning. Additionally, our proposed learning-aided reduced-complexity beam-alignment design performs similarly to beam-sweeping based beam-alignment relying on high-complexity exhaustive beam-search.

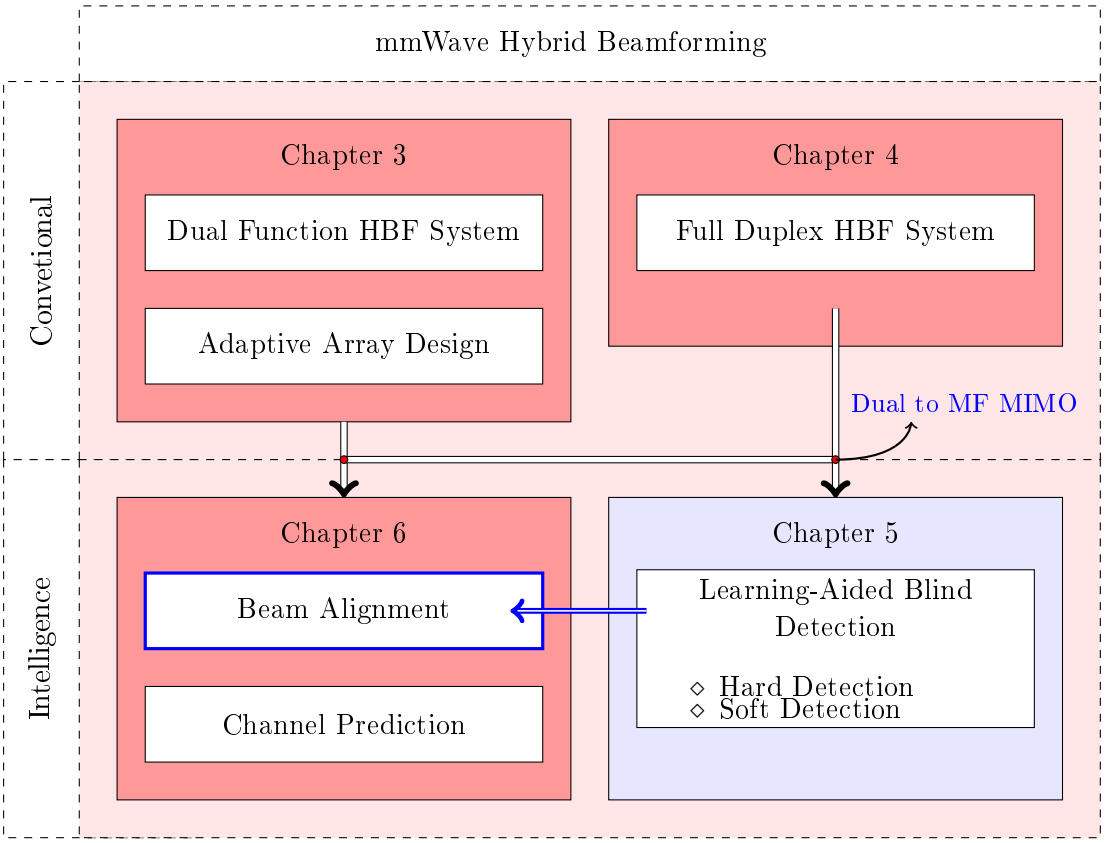


Figure 1.4: Structure of this thesis.

10. Finally, to counteract the effects of channel aging, we propose a learning-assisted channel prediction technique relying on a radial basis function neural network, where we show by simulation that upon employing sufficient training the neural network-aided channel prediction can faithfully reproduce the current channel. We empirically show that depending on the Doppler spread, the neural network used for channel predictions has to be periodically retrained. In other words, we demonstrate by simulations that the overhead dedicated to the CSI feedback for faithful reproduction of the actual channel at the transmitter is low for low Doppler spreads.

1.3 Thesis Outline

The structure of thesis is presented in Fig. 1.3 and Fig. 1.4. More particularly, Fig. 1.3 shows the genesis of this thesis under the umbrella of mmWave hybrid beamforming systems. To elaborate further, this thesis first embarks on *conventional* transceiver design for both half- and full-duplex hybrid beamforming systems. Then our conventional transceiver designs are enhanced with the aid of the machine learning techniques in Chapters 5 and 6. Explicitly, machine learning tools are invoked for channel prediction, beam alignment and symbol detection. On the other hand, Fig. 1.4 shows the flow of

the thesis structure, where Chapters 3 and 4 focus on dual-function hybrid beamforming system design, while Chapter 5 discusses multi-functional hybrid beamforming designs relying on index modulation systems. Both the dual-function and multi-function hybrid beamforming designs rely on having channel knowledge, an issue which will be detailed in Chapter 6.

More explicitly, the structure of the thesis is organised as follows.

Chapter 2: Literature Survey

In Sec. 2.1.1 we commence our discourse of this thesis by studying the radio wave propagation effects in mmWave systems. Then in Sec. 2.2, we analyze the array gain required for compensating the propagation losses, as well as beamsteering, beamwidth and antenna types, namely patch and horn antennas. Subsequently in Sec. 2.3 we provide a link-budget analysis discussing the number of AEs required for achieving a given spectral efficiency for a given transmit power. Furthermore, while Sec. 2.4 presents the state-of-the-art in mmWave hybrid architectures and beamforming solutions followed by Sec. 2.5, which considers energy-efficient hybrid beamforming designed for mmWave channels. Additionally, in Sec. 2.6 we survey the state-of-the-art in the context of full duplex designs, Finally, an exposition of machine learning for wireless communications is presented in Sec. 2.7.

Chapter 3: Dual-Function Hybrid Beamforming Transceiver Design

In Sec. 3.2, we conceive a dual-function hybrid beamforming transceiver design, where both diversity and beamforming gains are obtained. The capacity analysis of this design is presented in Sec. 3.2.1. Then, Sec. 3.3 we develop a hybrid precoding technique relying on the DFT and MUB matrices. Subsequently in Sec. 3.4, we propose a dual-stage adaptation technique, where the adaptation takes place in the architecture front as well as in the hybrid precoding, depending on whether the channel is LOS or NLOS nature. Then in Sec. 3.4.2, we extend our adaptive design to a C-RAN scenario relying on virtual cell formation as an application. Finally, our conclusions of the chapter are presented in Sec. 3.5.

Chapter 4: Hybrid Beamforming for Full Duplex mmWave Systems

In this chapter, we advocate the employment of beamforming in full duplex mmWave systems for mitigating the self-interference. More particularly, in Sec. 4.2 we propose a beamforming solution for both transmitter and receiver nodes in a single-user scenario. The convergence of the solution and its overall performance are discussed in Sec. 4.2.2 and Sec. 4.2.5, respectively. Then in Sec. 4.3, we extend our proposed design to a multiple user scenario, where a solution is derived for mitigating both the self-interference and multiuser interference. We provide our performance results characterizing this scenario in Sec. 4.3.3. Finally, the Chapter conclusions are presented in Sec. 4.4.

Chapter 5: Learning-Aided Transceiver Design for Multi-functional mmWave Hybrid Systems

In this chapter, we invoke machine learning for enhancing the performance of multifunctional mmWave hybrid systems. We commence our discourse by presenting a learning-aided adaptive transceiver design in Sec. 5.2, where the mmWave transmitter adapts between the transmission modes of multiplexing and diversity gains and discussion on the complexity of the design and results are presented in Sec. 5.2.3 and Sec. 5.2.4, respectively. Furthermore, in Sec. 5.3 we invoke machine learning tools, such as a neural network, for detection in our multifunctional MIMO system. More particularly, in Sec. 5.3.3 we propose a neural network design for MS-STSK operating system both with and without BIM. Our discussions on the receiver complexity and performance results are presented in Sec. 5.3.4 and Sec. 5.3.5, respectively. Additionally, we extend our proposed neural network design to soft-detection in Sec. 5.4. The corresponding performance versus complexity results are analyzed in Sec. 5.4.1 and Sec. 5.4.2. Finally, our conclusions are presented in Sec. 5.5.

Chapter 6: Deep Learning-Aided CSI Acquisition for mmWave Systems

This chapter's focus is mainly on beam-alignment and small-scale fading channel coefficient prediction. In Sec. 6.2, we propose a multi-fingerprint based beam-alignment technique proposed for vehicular communications, where we invoke learning for intelligent beam-alignment in Sec. 6.2.2. Our performance results and discussions on the complexity imposed are presented in Sec. 6.2.4 and Sec. 6.2.5, respectively. Subsequently, to combat the channel aging phenomenon of wireless communication, Sec. 6.3 discusses neural

network assisted channel prediction. Our performance results and complexity analysis are provided in Sec. 6.3.3 and Sec. 6.3.2. Finally, the chapter conclusions are presented in Sec. 6.4.

Chapter 7: Conclusions and Future Work

In this chapter, we provide a summary of our findings and future scope of the work highlighting some salient future research ideas.

Chapter 2

Literature Survey

A man's feet should be planted in his country, but his eyes should survey the world

— George Santayana

2.1 Introduction

Given the escalating mobile data rate demands, the world is pacing toward the next generation of wireless communication aiming for increased spectral efficiency. Recently, mmWave communication has gained much attention as one of the key disruptive technologies promising further increased data rates by harnessing the wide bandwidths available at mmWave frequencies [1].

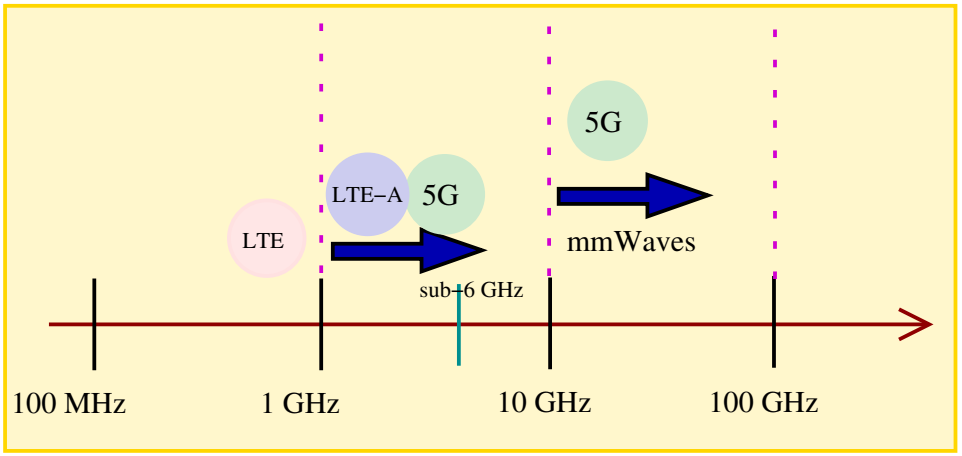


Figure 2.1: Carrier frequencies of next-generation wireless communications.

Fig. 2.1 shows the frequencies of the next-generation wireless communication, where the frequencies above $10 \sim 30$ GHz are classified as belonging to the mmWave band, since the wavelength of those frequencies is on the order of millimeters.

Although the mmWave band has gained attention and reached a certain prominence at the time of writing, the first mmWave communication system dates back to the late 1890s, when Sir Jagadish Chandra Bose pioneered the development of mmWave components: coherer, spark transmitter, horn antenna, polarizer, dielectric lens and cylindrical diffraction grating. By relying on these components, Bose demonstrated the feasibility of transmission and reception of electromagnetic waves at 60 GHz, over 23 meters through two intervening walls by remotely ringing a bell and detonating gunpowder. This was the first mmWave communication system in the world.

In the next section, we study the key large-scale propagation channel effects at mmWave frequencies.

2.1.1 Radio Wave Propagation Effects for mmWave

In this section, we study the mmWave propagation characteristics, namely, the atmospheric attenuation, path loss model, Doppler spread and coherence bandwidth.

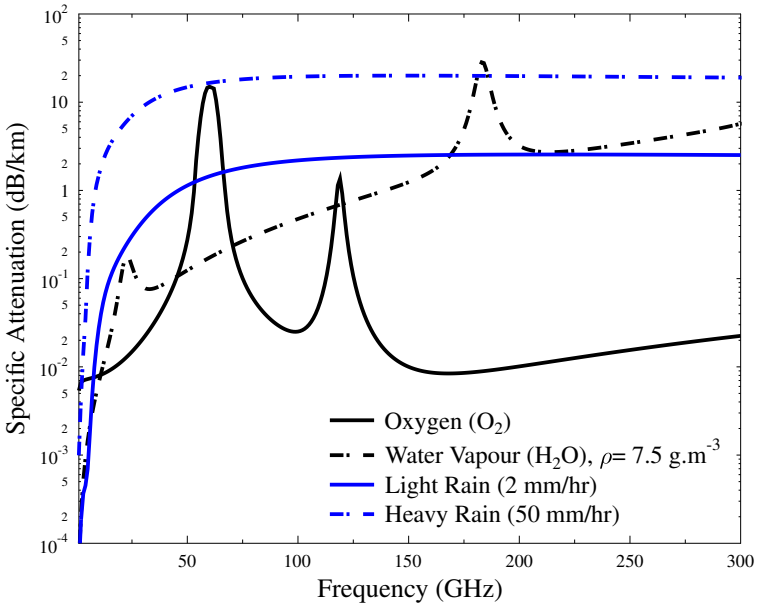
Atmospheric Effects

Some of the key challenges involved in mmWave communication are due to its high propagation losses imposed by the attenuation due to foliage, atmospheric absorption and rain-induced fading. Fig. 2.2 shows the attenuation caused by oxygen absorption and foliage. More particularly, Fig. 2.2(a) depicts the propagation loss attributed to rain-induced fading, water vapour, and oxygen absorption; while Fig. 2.2(b) shows the attenuation due to foliage at different frequencies. Because of the aforementioned losses, the signal-to-noise ratio (SNR) would be typically low at the receiver [1,44].

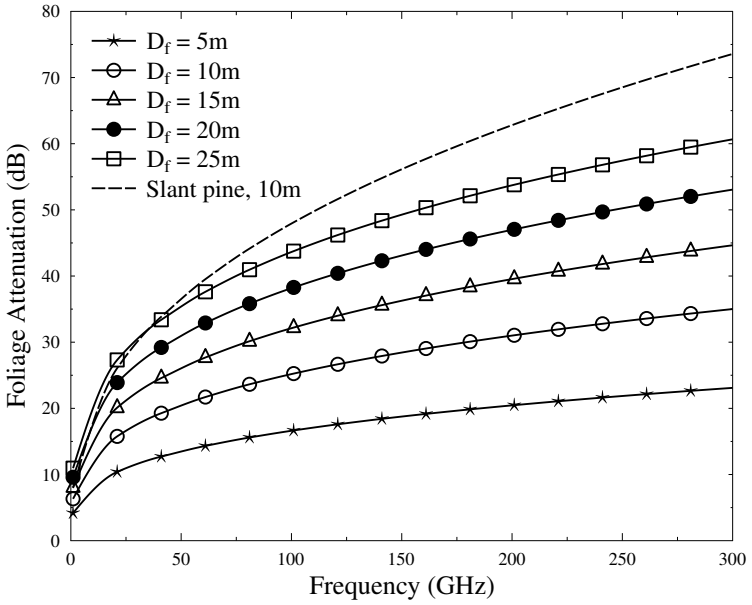
Therefore, to mitigate the propagation losses, high-gain directional transmissions have to be employed using large antenna arrays [2]. Given that the wavelength at mmWave frequencies is on the order of a few millimeters, large antenna arrays can be packed into a compact area for attaining high BF gains, which will be discussed later in Sec. 2.2.1.

Path Loss Model

Having briefly discussed the atmospheric effects at mmWave frequencies, we now focus our attention on the received power. The free space loss (FSL) of a signal for a unity



(a) Propagation loss due to rain-induced fading, water vapour, and oxygen absorption.



(b) Attenuation due to foliage.

Figure 2.2: Atmospheric effects observed at mmWave frequencies.

gain transmit and receive antenna is given by [45]

$$P_r = \frac{P_t}{K}\left(\frac{\lambda}{4\pi d}\right)^2, \tag{2.1}$$

where d is the propagation distance, λ is the operating wavelength, K is a dimensionless constant and P_t is the transmit power.

However, the FSL of (2.1) does not always hold in practice. It may only be accurate when there is only a LOS path and when the antennas are perfectly aligned on boresight. A modified model that best fits the path loss measurements is given by the log-distance model formulated as [44]

$$P_r(d) = P_t K_{fs}\left(\frac{d_o}{d}\right)^{\alpha}, \forall d \geq d_o, \tag{2.2}$$

where $d_o >> \lambda$ is a close-in FSL reference distance in the far-field, while the constant K_{fs} and the path loss exponent α are adjusted accordingly to fit the measurements. It is convenient to express (2.2) on a logarithmic scale, since the received power significantly changes even for small distances, when considered in linear scale. Thus we have,

$$P_r[\text{dBm}] = P_t[\text{dBm}] + G_r + G_t + 10\log K_{fs} - 10\alpha\log\left(\frac{d}{d_o}\right) \forall d \geq d_o, \tag{2.3}$$

where G_r and G_t are the gains of the transmit and receive antennas, respectively. Note that the pathloss exponent α may be different for different environments. Furthermore, by accounting for the atmospheric losses in addition to FSL, we can express (2.3) as

$$P_r[\text{dBm}] = P_t[\text{dBm}] + G_r + G_t + \text{PL}(d)\,\forall d \geq d_o, \tag{2.4}$$

where $\text{PL}(d)$ includes both the FSL as well as the atmospheric losses.

Diffraction

Diffraction is a phenomenon observed in radio signals (in general any wave) when they impinge on an obstacle. Diffraction is defined as the ability to propagation around an object; in other words, resulting in the bending of waves around the corners of the obstacle, as shown in Fig. 2.3, which is more dominant below 6 GHz. However, diffraction at mmWave frequencies is weak, hence cannot be relied upon for mmWave propagation [1]. Indeed, the received power may fall by more than 40 dB when the receiver is moved by just a few centimeters at the corner of an obstacle [44, 46].

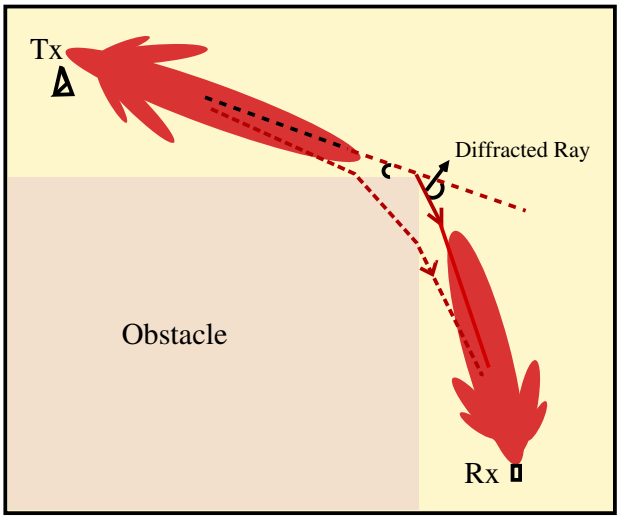


Figure 2.3: Illustration of diffraction in radio signals.

Reflection

While mmWave frequencies may have limited diffraction, they rather tend to be reflected. Empirically evidenced [47] that mmWave signals at frequencies 28, 38, 60, and 70 GHz are strongly reflected by human beings, buildings, walls, lampposts, trees [47]. Therefore, reflective properties may be invoked to mitigate the lack of diffraction by electronically steering the antenna array to find the reflected and paths scattered by the nearby obstacles.

Doppler Effects

For a user moving at a velocity v towards a transmitter whose velocity is v_0, the frequency of the signal observed at the receiver is given by [48]

$$f' = \frac{c + v}{c - v_0} f, \tag{2.5}$$

where f is the frequency of the signal from the transmitter and c is the velocity of electromagnetic signal.

Thus, the shift in the frequency, which is referred to as Doppler shift, observed at the receiver can be expressed as

$$f_d = f' - f \tag{2.6}$$

$$= \frac{v - v_0}{c - v_0} f. \tag{2.7}$$

For a given transmitter and receiver velocity, we can say from (2.6) that the Doppler shift is proportional to the transmitter frequency. While the conventional Doppler spread models developed by Clarke [49] and Gans [50], are suitable for omnidirectional antennas, the Doppler spread model of mmWave signals has to be based on directional transmission, which is contingent on the specific AoA [44]. Therefore, the Rayleigh fading distribution assumption may not hold for directional transmission. More explicitly, the fading and coherence time of mmWave signals depend on the particular beamwidth, velocity, frequency, and the specific number of *finite* multipath signals. Although there are no exact Doppler spread models for mmWave signals at the time of writing, they tend to experience more rapid fading upon increasing the Doppler frequency [51]. Since the coherence time expressed in (2.8) is inversely proportional to Doppler spread [51], the mmWave channels coherence time typically becomes shorter with high Doppler spreads.

$$T_c \approx \frac{1}{f_d}. \tag{2.8}$$

Delay Spread

Since the transmission of mmWave signals is highly directional while also being sparse, the delay spread tends to be lower than that of sub-6 GHz frequencies. Therefore, the coherence bandwidth [44,52] is generally high.

$$B_c \approx \frac{1}{5\tau_{\text{RMS}}}. \tag{2.9}$$

Following this rudimentary introduction to the radio wave propagation effects on mmWave, in the next section, we briefly highlight, how the AA gains may be achieved for compensating the propagation losses discussed.

2.2 Antenna Array Gain

In this section, we touch upon the design of AAs including their directionality and beamwidth obtained. We focus our attention on linear AAs. Consider a wavefront crossing a linear AA having N elements at an angle θ, as shown in Figure 2.4. The wave arriving at antenna element (AE) 1 travels an extra distance of $d\cos\theta$ with respect to the second element (AE 2), thus resulting in a phase difference of $\frac{2\pi}{\lambda}d\cos\theta$ between the two. Similarly, the phase difference between the first AE and the N^{th} element is $(N-1)\frac{2\pi}{\lambda}d\cos\theta$. The response vector and the AF for such an array configuration is given by [53],

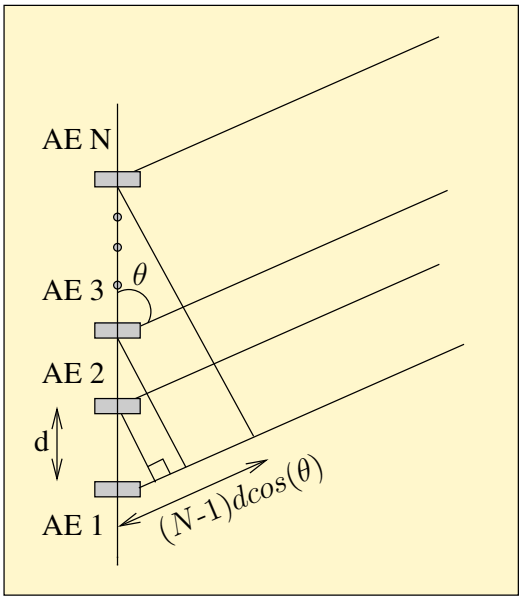


Figure 2.4: Signal passing through the array with N antenna elements (AE) at an angle θ.

$$\mathbf{a}_r = [1 \; e^{j\psi} \; e^{j2\psi} \ldots e^{j(N-1)\psi}] \tag{2.10}$$

$$\text{AF} = a_0 + a_1 e^{j\psi} + \ldots + a_{N-1} e^{j(N-1)\psi} \tag{2.11}$$

$$\psi = (2\pi/\lambda) d \cos\theta + \beta = kd\cos\theta + \beta, \tag{2.12}$$

where $a_0, a_1, \ldots, a_{N-1}$ are feed coefficients or voltage levels (amplitudes) of the array [54], d is the spacing between the AEs and β is excitation phase of the array, which is induced by the source because of phase changes involved the currents passing through the array [54], [53]. k is the wavenumber equal to $\frac{2\pi}{\lambda}$, it is a measure of the number of cycles of wave present in one meter of distance and ψ is often termed as wavenumber variable as it depends on the wave number k.

For uniform linear array, $a_0 = a_1 = \ldots a_{N-1} = 1$ and the array factor for such array is given by [53]:

$$\begin{aligned} \text{AF} &= 1 + e^{j\psi} + \ldots + e^{j(N-1)\psi} \\ &= e^{j\frac{(N-1)}{2}\psi} \left[\frac{e^{j\frac{N}{2}\psi} - e^{-j\frac{N}{2}\psi}}{e^{j\frac{N}{2}\psi} - e^{-j\frac{N}{2}\psi}} \right] \\ &= e^{j\frac{(N-1)}{2}\psi} \left[\frac{\sin(\frac{N}{2}\psi)}{\sin(\frac{1}{2}\psi)} \right], \end{aligned} \tag{2.13}$$

assuming the center of the AA as the reference point, we get $\text{AF} = \left[\frac{\sin(\frac{N}{2}\psi)}{\sin(\frac{1}{2}\psi)} \right]$.

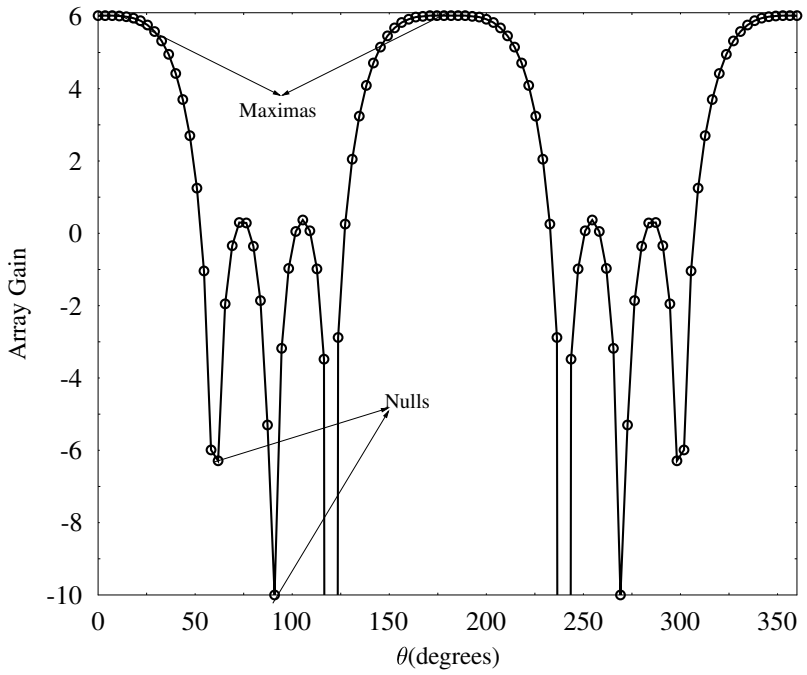


Figure 2.5: Array gain obtained at different angular directions, the first null (zero gain) is seen at 60^o while first maximum is seen at 0^o.

The direction, where the radiation or the gain of the AA is zero, is referred as a null. The nulls of the array occur at $\theta_n = \cos^{-1}\left(\frac{\lambda}{2\pi d}(-\beta \pm \frac{2n}{N}\pi)\right)$, where $n = 1, 2, 3, ...; \forall n \neq N, 2N, ...$ Maximum gain of the antenna array occurs at $\theta_m = \cos^{-1}\left(\frac{\lambda}{2\pi d}(-\beta \pm 2m\pi)\right)$, where $m = 0, 1, 2, ...$ The AF for 4 AEs at $\lambda/2$ spacing with zero excitation phase is shown in Figure 2.5. The first null in Figure 2.5 occurs at 60^o while the first maximum occurs at 0^o. Having obtained the array factor, the gain of the total array is defined as the product of the gain of one AE and the AF [54] as follows:

$$G_{\text{total}} = G_{\text{element}} \mid AF \mid^2 . \tag{2.14}$$

The directivity of an AA is defined as the radiation intensity of an array in the direction of its strongest emission with respect to isotropic AA, which transmits uniformly in all

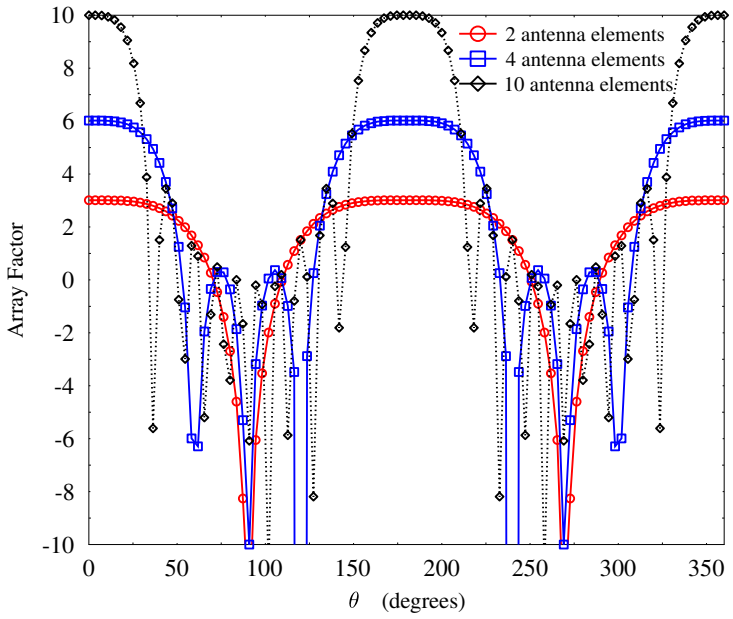


Figure 2.6: Array gain with N=2, 4, and 10 AEs

directions. Mathematically, it is expressed as [53]:

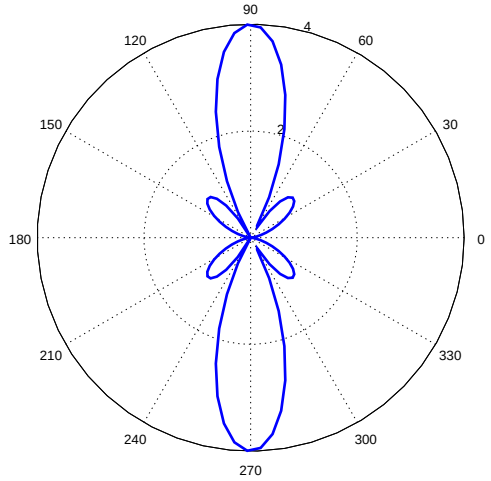
$$\begin{aligned} \text{Directivity}(D) &= \frac{G_{\max}}{G_{\text{iso}}}, \\ \text{and for } d \ll \lambda, \quad G_{\text{iso}} &= \frac{\lambda}{2Nd} \\ G_{\max} &= G_{\text{element}} \max_{0^0 \le \theta \le 360^o} \mid AF \mid^2 = G_{\text{element}} N^2. \end{aligned} \tag{2.15}$$

In practice, it is not possible to derive complete array gain in particular direction like a pencil beam, where there is always some leakage of radiation in the undesired direction. So the antenna gain with respect to isotropic antenna with directivity D is given by,

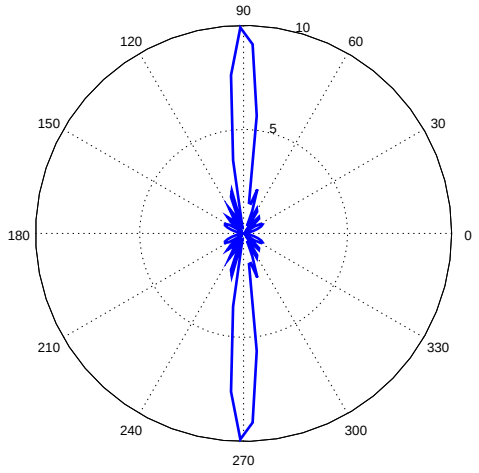
$$G_{\text{antenna}} = \eta D = \eta \frac{2Nd}{\lambda} G_{\text{element}}, \tag{2.16}$$

where η is the efficiency of the array in the desired direction.

As the number of AEs increase, narrower beams are observed, which increases the directionality of the array. The array gain for different number of AEs at $\lambda/2$ spacing is shown in the Figure 2.6. It is interesting to see from Figure 2.6 that the mainbeam lobe of the array with 2 AEs is wider and spans from 140^o to 230^o than the mainbeam lobe of the array with 4 AEs which spans from 150^o to 205^o. Similarly, the mainbeam lobe of 10 element AA is narrower than 2 and 4 element antenna arrays as shown in Figure 2.6.



(a) Beamwidth with 4 AEs.



(b) Beamwidth with 10 AEs.

Figure 2.7: Polar plots representing the beamwidth of 4 and 10 AEs.

For better understanding of the beam patterns, polar plots of the radiation pattern is shown in Figure 2.7. It is clear from Figure 2.7 that the radiation pattern of the mainbeam lobe becomes narrow as the number of AEs increase from 4 to 10. Figure 2.7(a) and Figure 2.7(b) correspond to the array gains obtained in Figure 2.6 for 4 and 10 AEs. The mainbeam lobe width for different number of AEs and its variation with the angle of departure/arrival of the signal will be discussed in the Section 2.2.1 and Section 2.2.2.

Another important factor that affects the mainbeam lobe is the spacing between the antennas. As the spacing between the AEs increase from $\lambda/2$ to λ, grating lobes are observed as shown in the Figure 2.8 for AE spacing of λ. Grating lobes are the mainbeam

lobes in the undesirable direction, which are formed because of the over-sampling and the repetition of the AF values in the region $-kd \leq \psi \leq kd$, when $kd > \pi$ [54]. Grating lobes are unnecessary for the BF in cellular communications and can be avoided by maintaining the space between the antenna $d < \lambda/2$. Also in Figure 2.8, when the spacing between the AEs decrease from $\lambda/2$ to $\lambda/16$, the beam becomes wider resulting in loss of the directionality of the AA. The mainbeam lobe in Figure 2.8 at $\lambda/16$ distance spreads with almost equal gain in all angular directions, that is from 0^o to 360^o, with a small deflection in 150^o to 230^o directions.

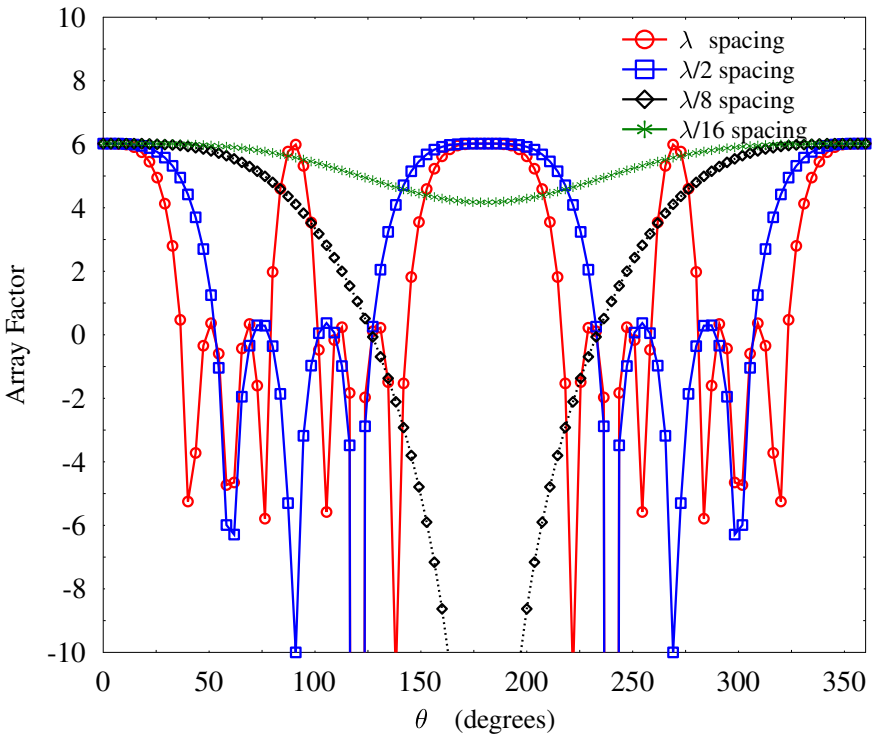
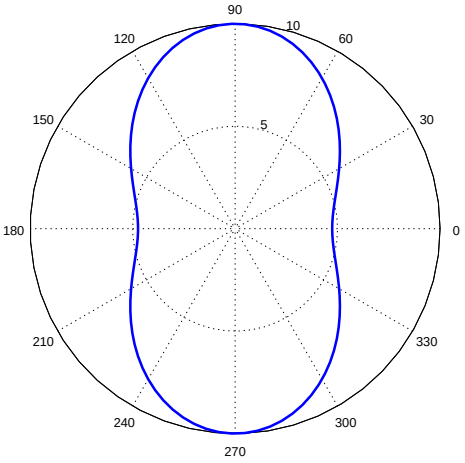


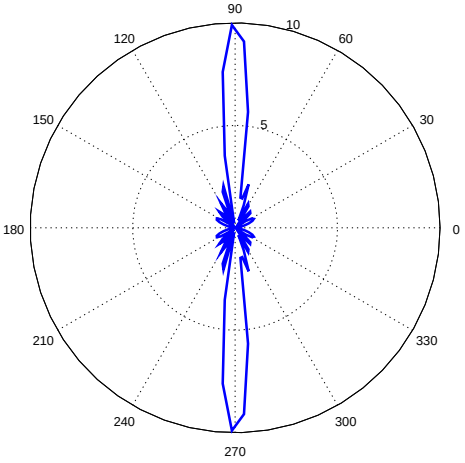
Figure 2.8: Array gain with $\lambda/16, \lambda/8, \lambda/2,$ and λ spacing between the antennas.

The polar plots at different spacing between the AEs is shown in Figure 2.9. The radiation pattern of the mainbeam lobe is widespread with $d = \lambda/16$ as shown in the Figure 2.9(a), which corresponds to the array gain in Figure 2.8 at $\lambda/16$ spacing, on the other hand with $d = \lambda$, a grating lobe or sidelobe in the undesired direction is observed as shown in the Figure 2.9(c), which corresponds to the array gain obtained in Figure 2.9 at λ spacing between AEs. The grating lobes can be eliminated completely by maintaining $d \leq \lambda/2$. Therefore narrower beams without grating lobes can be achieved by maintaining $d = \lambda/2$ as shown in Figure 2.9(b).

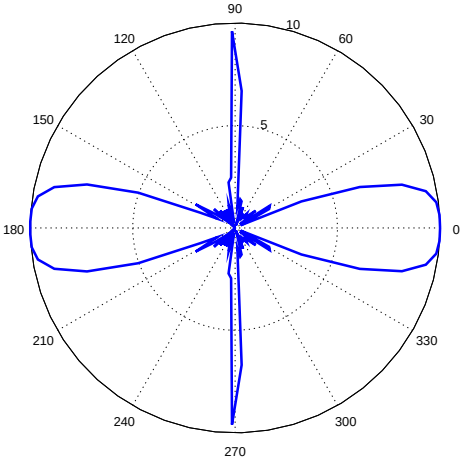
In mmWave communications, since the directional transmission is considered, steering the mainlobe beam in the direction of the UEs is crucial. In the following section, we



(a)

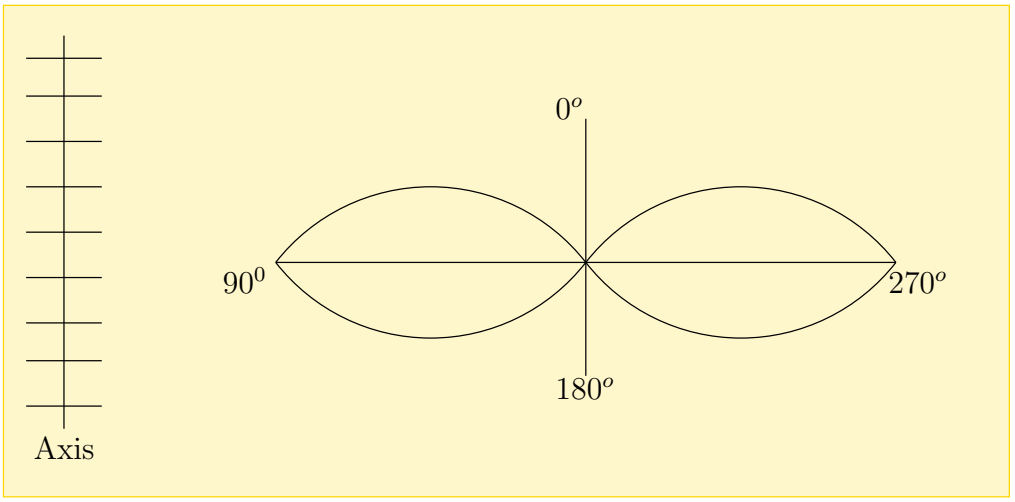


(b)

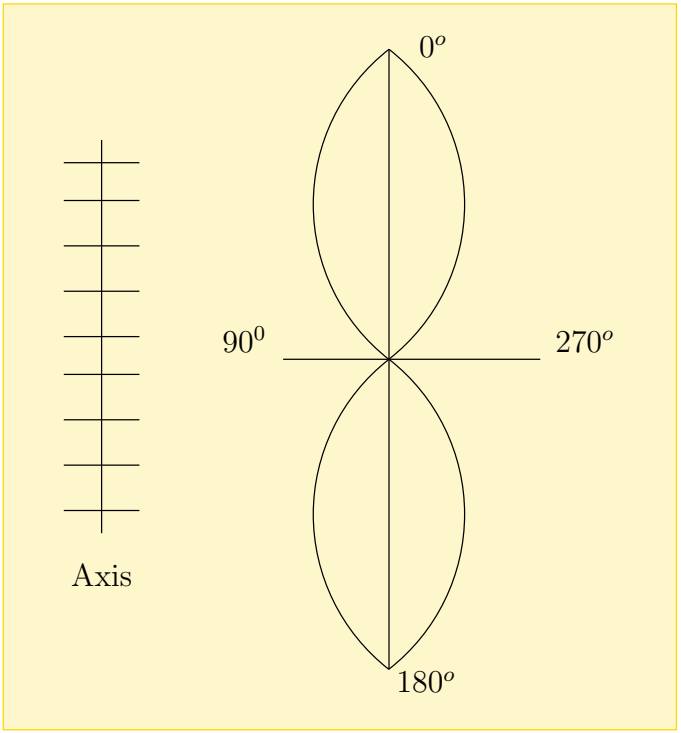


(c)

Figure 2.9: Polar plots representing the beam patterns at (a) $\lambda/16$, (b) $\lambda/2$ (c) λ spacing between AEs.



(a) Broad side array



(b) End fire array.

Figure 2.10: Mainbeam lobe direction at broad side and end fire arrays.

describe how a beam is guided in the desired direction.

2.2.1 Beam Steering

An AA is normally designed such that the maximum array gain occurs at broad side. Broad side array shown in Figure 2.10(a) refers to the directivity or the maximum array gain in the direction normal (90^o) to the axis of the array. Figure 2.10(a) shows the mainbeam lobe at 90^o and 270^o away from the axis of the array, whereas an end fire array as shown in Figure 2.10(b), refers to the directivity or the maximum array gain in the direction parallel to the axis of the array. Figure 2.10(b) shows the mainbeam lobe at 0^o and 180^0, which is along the axis of the array. So the mainbeam lobe in an array occurs at broad side when $\psi = kd\cos\theta + \beta = 0$. To steer the beam electronically in a direction say, θ_0, the excitation angle β is set to $-kd\cos\theta_0$. This translates the beam in the desired direction θ_o with wavenumber variable $\psi = kd(\cos\theta - \cos\theta_o)$. Figure 2.11 depicts the steering of a beam from 90^o to 60^o by setting $\beta = kd\cos 60^o$.

It is worth noticing in Figure 2.11, the increase in the beamwidth of the lobe as the beam direction is steered to 60^o from 90^o. In the next subsection, the beamwidth at different steering angles for different number of AEs is studied.

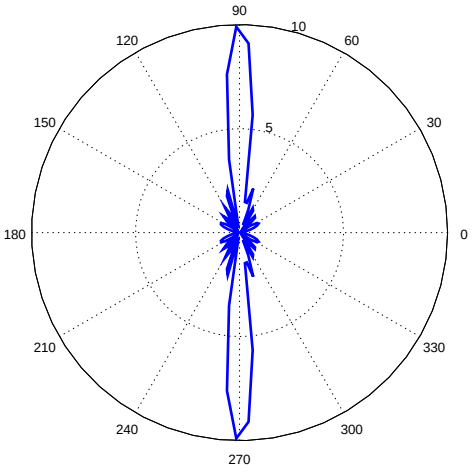
2.2.2 Beamwidth

Half power beamwidth is defined as the angular separation of radiation from the main lobe to half the power of the main lobe, i.e -3 dB from the main lobe. Double the half power beamwidth is referred as 3 dB beamwidth as shown in Figure 2.12. The beamwidth, for example, in Figure 2.12, is around 40^o, and the sidelobe attenuation level is around 13 dB below the mainlobe. The beamwidth for a uniformly linear antenna array towards the desired direction θ_o is obtained by linearizing the wavenumber variable ψ about $\theta = \theta_0$, and is given by [54],

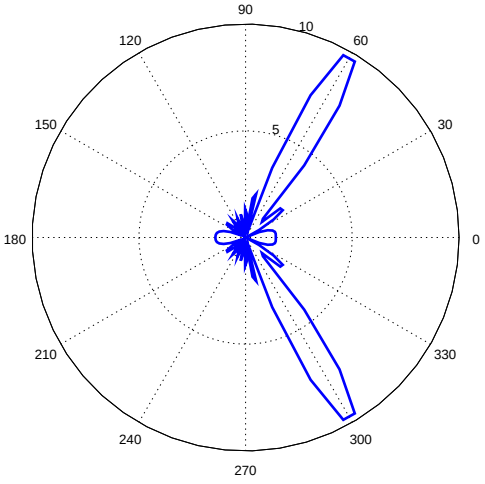
$$\begin{aligned} \triangle\phi_{3\text{dB}} &= \frac{1}{kd\sin\theta_o}\triangle\psi_{3\text{dB}} \text{ for } 0^o < \theta < 180^o \\ &= 2\sqrt{\frac{\triangle\psi_{3\text{dB}}}{kd}} \text{ for } \theta_0 = 0^o \text{and } 180^o, \end{aligned} \tag{2.17}$$

where $\triangle\psi_{3\text{dB}}$ is the beamwidth at broadside array, and it is equal to $0.886\frac{2\pi b}{N}$. 'b' is the broadening factor, and it depends on the type of the window and the attenuation of the sidelobe level. Larger attenuation of the sidelobe results in a larger broadening factor. Broadening factors for different windows are given in Table 2.1, where R and R_a are the sidelobe attenuation levels in dB and linear scale.

In Figure 2.13, beamwidth for different number of AEs at different directions of transmission is shown. As shown in Figure 2.13, as the beam is steered from 0^o to 90^o, the



(a) Beam at 90^0 away from the axis of the array



(b) Beam guided to 60^0 away from the axis of the array

Figure 2.11: Beam steering through wave number translation from 90^o to 60^o.

beamwidth decreases. The beamwidth is not constant with the angle at which the signal hits the array. It changes with the direction of transmission/reception [53]. Also, with increase in the number of AEs in an array, the beamwidth decreases in the desired direction. In other words, the beamwidth of an array depends on the length of the array, and the steering angle.

Depending on the space and gain requirements of the base station and the mobile station, it is important to study the type of antennas to be employed. In the next subsection we discuss two promising antenna types for mmWave communications.

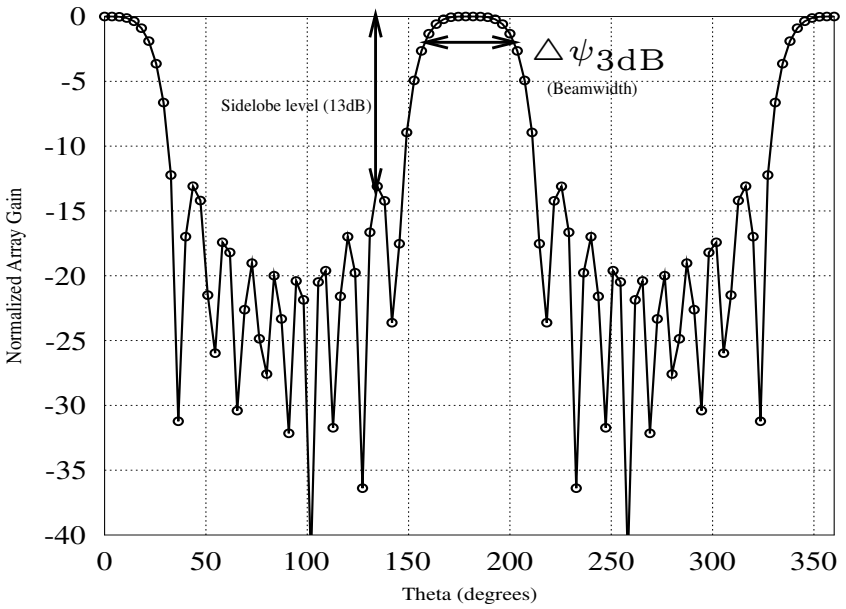


Figure 2.12: 3dB beamwidth of a uniformly linear AA.

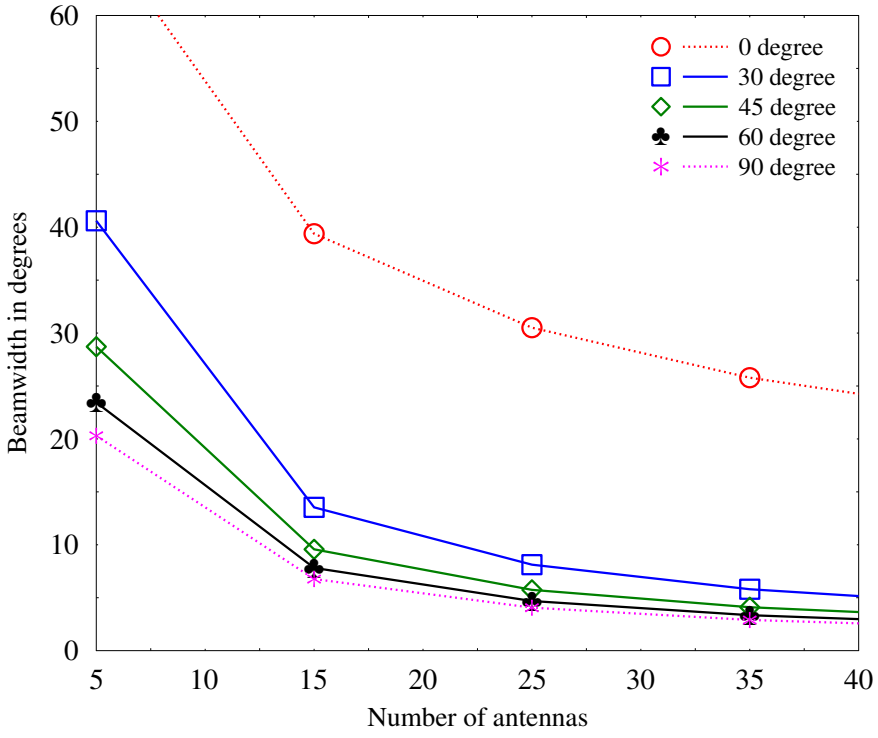


Figure 2.13: Beamwidth vs number of AEs at different desired angle of departures.

Rectangular	b=1,($R = 13$ dB)
Hamming	b=2, ($R = 40$ dB)
Taylor-Kalser	b=$\frac{6(R+12)}{155}$
Dolph-Chebyshev	b=$1 + 0.636 \left[\frac{2}{R_a} \cosh\left(\sqrt{a \cosh^2(R_a) - \pi^2}\right)\right]^2$

Table 2.1: Broadening Factors for Different Windows [54].

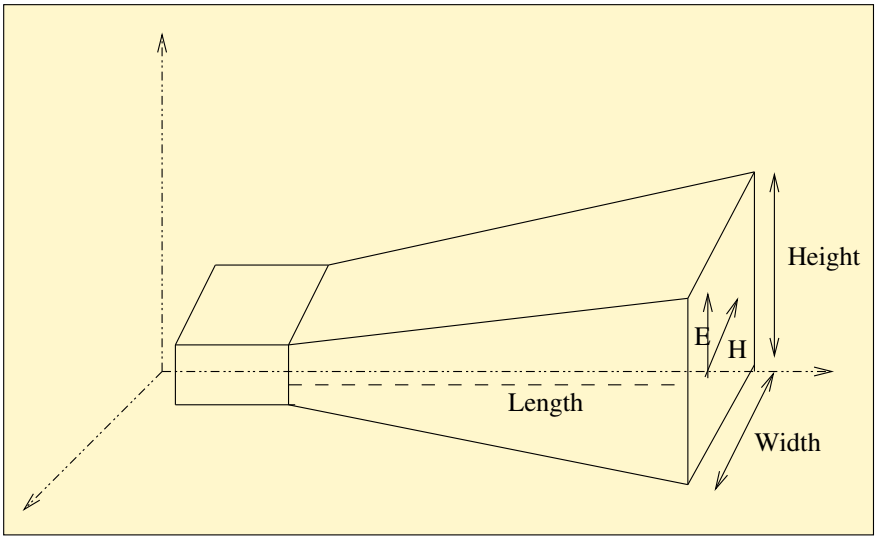


Figure 2.14: A typical horn antenna.

2.2.3 Type of Antennas for mmWave

In cellular communications, it is equally important to study the type of antennas employed at the base station and the mobile station. In this subsection, we discuss two promising antenna types for mmWave communications: horn antennas, and patch antennas. Horn antennas are widely used in ultra-high-frequency (UHF) band where they offer relatively high gains and directionality compared to other antennas. Horn antennas can provide gain as high as 25dBi. However, it is normally operated at low gains for higher steerability [55]. Compared to other antennas, horn antenna have minimal loss resulting in the high directivity with maximal efficiency which is equal to its gain. A typical horn antenna is shown in the Figure 2.14. E and H represents the direction in which the electric field is flared where the field distribution at the aperture of the horn antenna causes the radiation pattern. The radiation pattern and the gain can be altered by changing the dimensions of the horn antenna. Increasing the length of the antenna results in higher gains and directivity.

On the other hand, patch antennas are very flexible to integrate at the mobile station as they can be printed on the circuit board [56]. A typical patch antenna has a directivity gain around 5-7dBi and the steerability of a patch antenna is $< 180^o$ in azimuthal domain. However to achieve the degree of freedom up to 360^o in azimuthal domain, two sets of

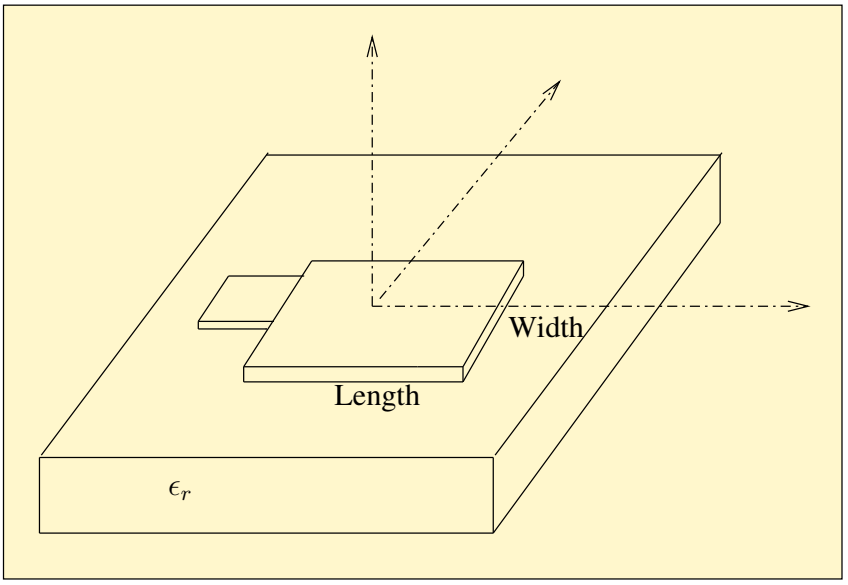


Figure 2.15: A typical patch antenna mounted on a substrate with permitivity ϵ_r.

patch antennas are used [55] . A typical patch antenna is shown in the Figure 2.15. The patch antenna in Figure 2.15 is mounted on the substrate with permitivity ϵ_r. The substrate's permitivity depends on the material of the circuit board. A patch antenna with larger widths support larger bandwidths.

Table 2.2: Link budget for a typical mmWave communication system

Link Budget	**Downlink**	**Uplink**
Transmit power (dBm)	40.00	20.00
Transmit antenna gain (dBi)	**25.00**	**12.00**
Carrier frequency (GHz)	28.00	28.00
Distance (km)	0.5	0.5
Free space propagation loss (dB)	115.32	115.32
Other losses (shadowing, fading)	20.00	20.00
Receive antenna gain (dB)	**12.00**	25.00
Received power (dBm)	-59.01	-75.88
Bandwidth (GHz)	0.50	0.5
Thermal noise (PSD) (dBm/Hz)	-174.00	-174.00
Noise figure	7	7
Thermal noise (dBm)	-80.01	-80.01
SNR (dB)	20.9-	4.13
Implementation loss (dB)	3.00	3.00
Spectral efficiency	6.00	1.2
Data rate (Gbps)	**3.00**	0.6

Having discussed the fundamentals of the antenna design, we now aim at analyzing the link budget for mm-wave communications. But prior to this, it is important to understand the path loss model that best fits for mmWave communication. In the next section, some preliminaries on the different path loss models in the literature are discussed.

2.3 Link Budget Analysis

To establish a communication link between a transmitter and a receiver, it is important to understand the power consumption at the transmitter and the power required at the receiver to obtain the desired data rate for a given channel with the thermal noise at the receiver. This analysis for establishing a link of a communication system is usually referred to as link budget. The link budget of a communication system depends mainly on the base station and the mobile station requirements, such as transmit power, antenna gains, signal-to-noise ratio (SNR) and the required throughput of the link through the channel.

A link budget for a typical uplink and downlink in mmWave communications system is shown in the Table 2.2 [57]. For the downlink communication as shown in Table 2.2, a 28GHz carrier is chosen with a base station transmit power of 40dBm [57]. Using the path loss model in (2.3), which best fits for the 28 GHz communication system as described in (2.3) the propagation loss for over 0.5 km distance using (2.3) in LOS environment ($n_p = 2$) with average shadowing loss of 20 dB is 135.32 dB. Our aim is to achieve a data rate of 3Gbps with 0.5GHz bandwidth, [58] has demonstrated using system level simulation that the data rate as high as 7.5 Gbps can also be achieved. For successful establishment of the link with 3 Gbps throughput, using Shannon's channel capacity [59], it requires 17.9 dB of SNR. Also, the efficiency of the RF components at such high frequencies could possibly be low [60]. So owing to the other system losses such as power drop in the circuitry, ADC/DAC, impedance mismatch, an additional 3 dB loss is considered, hence a total of 20.9 dB SNR is required to establish the link. We assume a noise figure of 7 dB [61], and the thermal noise[1] of -174 dBm/Hz. We also assume 12 dB of receive antenna gain. The reason for assuming 12 dB gain at the mobile station is because of the space constraints to pack large AAs, where 12 dB is a plausible assumption as it can be derived with 4 AE array with 6 dBi gain of each individual element using patch antennas. With these values in 2.2 and using (2.3) in $P_t = G_t + G_r + PL$, the base station is required to have an antenna gain of around 25 dBi. We assume the gain of an AE at the base station to be 10 dBi [57] as a low gain antenna would provide higher steering capability and minimal power leakage to side lobes [55] . So now we would need 15 dBi additional gain to achieve a rate of 3 Gbps. From (2.16), the total gain of the AA for $d = \lambda/2$ spacing is $G_{\text{antenna}} = \eta G_{\text{element}} N$. In dBi scale, we can view the gain of G_{antenna} as the sum of the gains obtained by the array and the element as:

$$G_{\text{antenna}}(\text{dBi}) = 10\log\eta + G_{\text{element}}(\text{dBi}) + \text{Array Gain(dB)} \tag{2.18}$$

Therefore, an additional 15 dBi is derived from the array. To obtain 15 dBi gain from the array with $\eta = 1$, we would need $10^{15/10} \approx 32$ AEs. However, in practice $\eta \neq 1$,

[1] Thermal noise is $10\log(1000KT)$ dBm, where $K = 1.38064852 \times 10^{-23}$ is the Boltzmann constant and T is the room temperature in kelvin = 300.

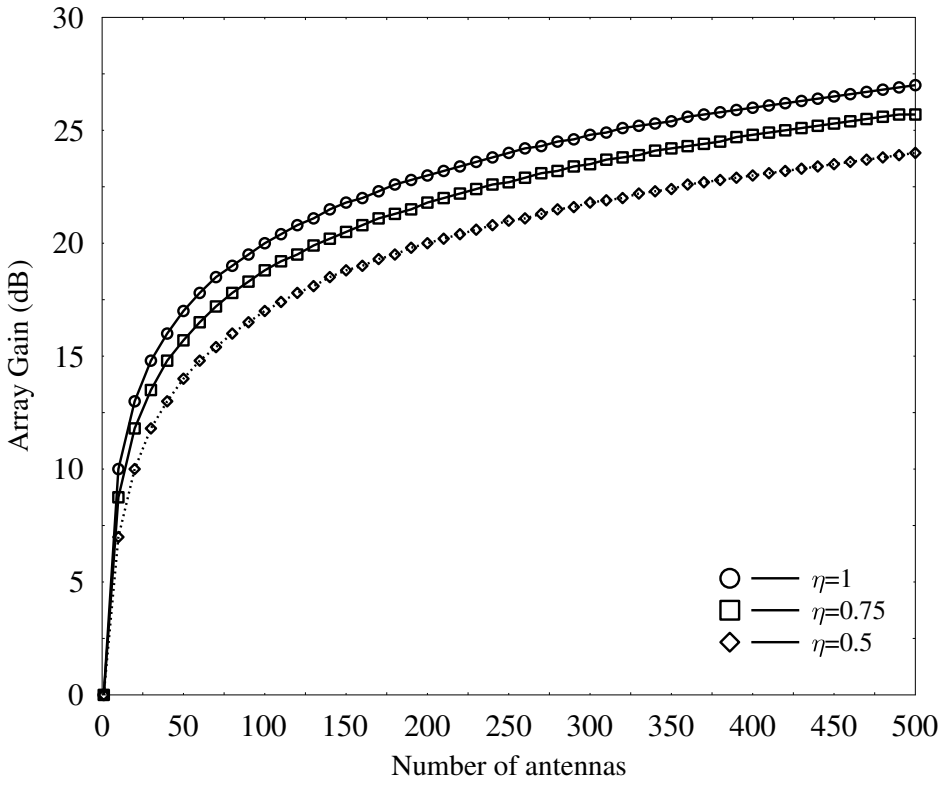


Figure 2.16: Array gain for different number of AEs at $\eta = 1, 0.75, \text{and}, 0.5$.

since there would be leakage of power (radiation) in other directions, particularly during steering. In other words, it is highly impractical to have a beamwidth that is very narrow like a pencil beam so as to derive full gain from AAs. So for a successful connection of the link, a little more than 32 AEs are required depending on the efficiency or the directivity of the array. Figure 2.16 shows the array gain at three different efficiencies. For $\eta = 0.75$, around 40 AEs are required to obtain an array gain of 15 dBi and for $\eta = 0.5$, as many as 60 AEs are required.

Similarly, the receive antenna gain of 12 dBi in Table 2.2 at the mobile station can be derived using 4 AE array with an element gain of 6dBi. *At the base station, horn antennas can be used as they provide relatively high gain with maximum efficiency compared to other antennas [55], as discussed in the subsection 2.2.3. And at the mobile station, considering the space and flexibility, patch antennas can be integrated. However, patch antennas azimuthal coverage is usually less than* 180^0*, while using 2D AA, would sweep entire* 360^0.

As we go up the frequency ladder from the microwave frequencies to the millimeter wave frequencies, the availability of bandwidth to meet the data requirements of the mobile

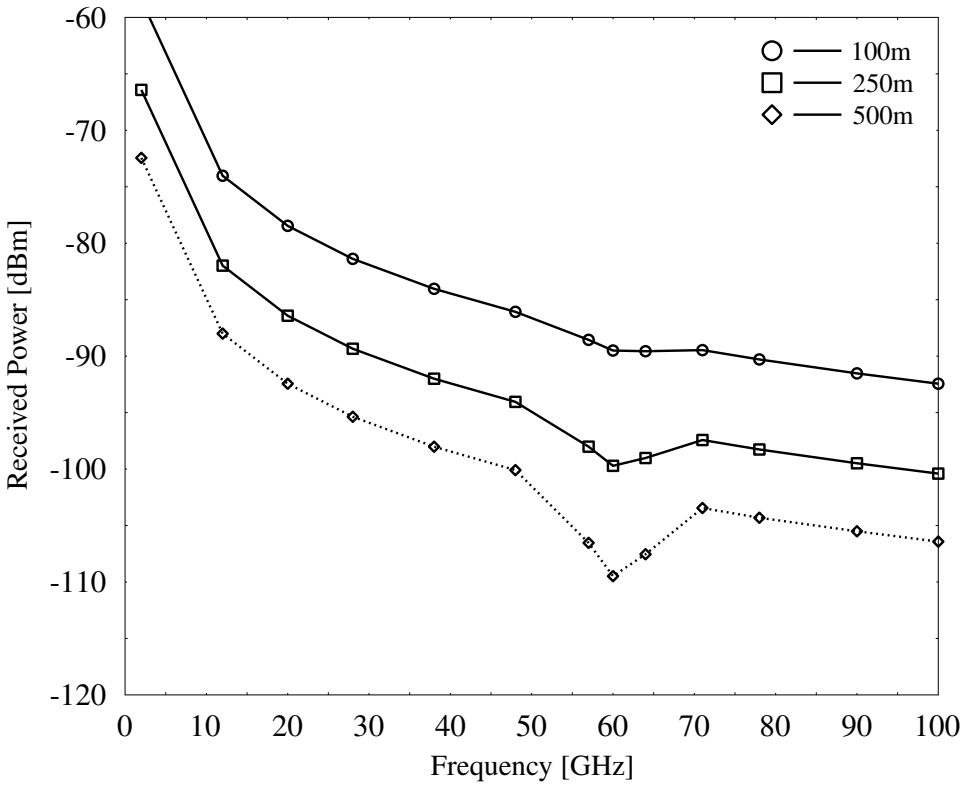


Figure 2.17: Received power at different frequencies with transmit power =40 dBm.

users increases significantly. However, as we move up along the frequency ladder, with significant increase in bandwidth, there is a significant propagation loss associated with the frequencies as shown in Figure 2.17. With double the frequency, there is a 6 dB loss in the propagation. The dip at 60GHz is because of the atmospheric attenuation and it increase with the propagation distance. Therefore, 60 GHz frequency is not suitable for outdoor environment.

One way to overcome the propagation loss is by using BF with AAs as discussed in the above section. Owing to the fact that at millimeter wave frequencies the spacing between the antennas grow smaller and smaller, large number of AEs can be employed at the base station to derive high gains using AAs.
To have a better understanding of how large is the size of the array or how many number of antennas are required at the base station, we consider again the link budget in Table 2.2, for a spectral efficiency of 6 bps/Hz at different frequencies. Figure 2.18 shows the number of AEs required to compensate the path loss at different frequencies. These values at different frequencies are obtained using the similar analysis discussed for 28 GHz. Because of the heavy propagation losses at high frequencies, there is an exponential

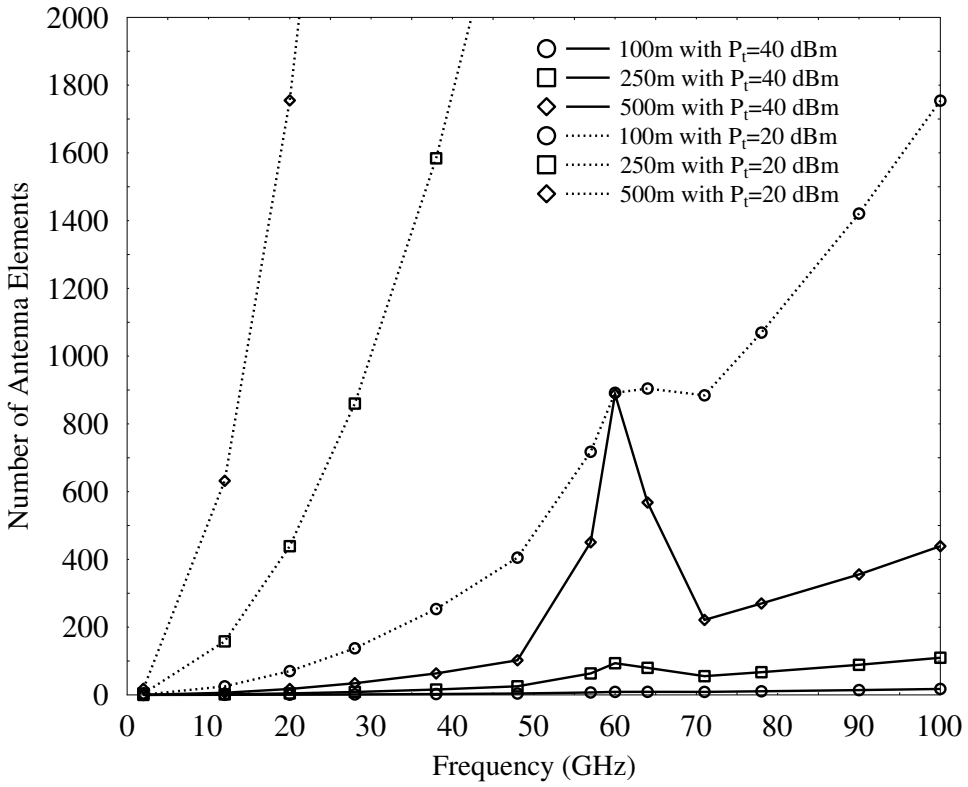


Figure 2.18: Number of AEs required to compensate for the path loss versus the carrier frequency. the frequency ladder.

growth in the number of AEs required. Also, for example, at 78 GHz, for 500 meters of distance, few hundreds of antennas are needed to compensate the path loss.

Besides shadowing, mmWave communications also suffer from other losses due to atmosphere absorption, which increases with the distance, foliage, and rain attenuation.

Received power at the mobile station at different distances for various frequencies is shown in Figure 2.19. It is interesting to investigate the path loss associated with the distance so as to know the coverage area and the number of AEs required for BF to increase the cell area. From Figure 2.19, we see that 57, 60, and 64 GHz have relatively high attenuation due to atmospheric absorption and they are not suitable for outdoor environment. 78GHz, too, has relatively high propagation loss than 28, and 38GHz. However, the deployment of a carrier frequency also depends on the BF gain required to compensate the path loss. Figure 2.20 shows the number of AEs required to achieve spectral efficiency of 6bps/Hz at various frequencies with different cell radius.

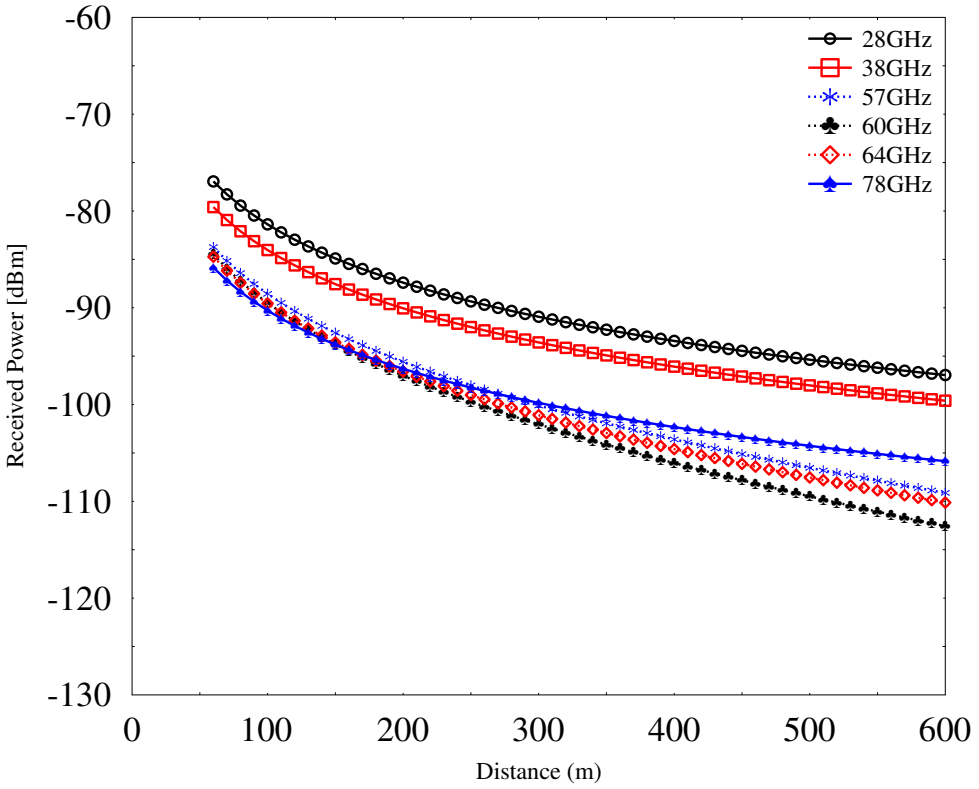


Figure 2.19: Downlink received power with different cell radius at frequencies 28, 38, 57, 60, 64, and 78 GHz. Transmit power is as in Table 2.2

It is evident from Figure 2.20 that for 60GHz several hundreds of antennas are required to compensate the losses due to attenuation using BF. Frequencies 57, 60 and 64 GHz, are more suitable only for indoor environments, as is employed in IEEE802.11ad, whereas for 28, and 38 GHz with less than a hundred antennas, a communication link up to a cell radius of 600 meters can be made. Although 78GHz demands few hundreds of antennas to cover the cell radius of 600 meters, given the flexibility and the size of the antennas, it is easy to integrate that many number of antennas at the base station at such frequencies.

In the foregoing sections, mmWave channel characteristics, antenna design and propagation loss as well as link budget analysis for mmWave communications is studied. Particularly, we have expounded that high path loss at mmWave frequencies entails high array gains, which can be harnessed with the employment of AAs. In the next section, we delineate on the architectures that can be implemented for achieving array gain (also referred as BF gain).

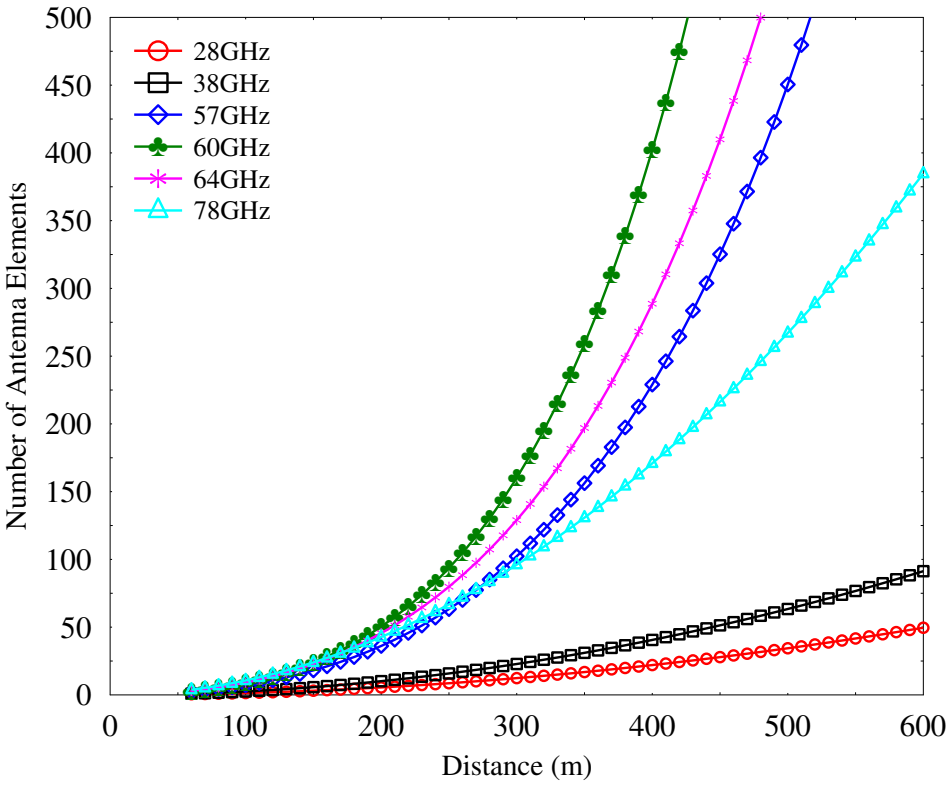


Figure 2.20: Number of AEs for different cell radii to achieve spectral efficiency of 6bps/Hz at 28, 38, 57, 60, 64, 78GHz.

2.4 Beamforming Architectures

We have seen in Sec. 2.3 that exploiting the high bandwidth of mmWave carriers requires employing a large number of antennas to provide a high array gain/BF gain. Fundamentally, BF is an angular spatial filtering carried out relying on an array of phase shifters. In contrast to omnidirectional transmission, BF is an operation invoked for steering the signal in a specific direction, as shown in Fig. 2.21. The benefit of BF is readily observed in terms of the SNR gain. This improvement in the SNR is attributed to the BF gain, which is some times referred as array gain, as discussed in Sec. 2.3.

Conventionally, BF is carried out by a pair of popular techniques, namely analog and digital BF. Considering the limitations of ABF and the complexity involved in digital BF, a hybrid solution may be preferred. In other words, hybrid BF relying on the amalgamation of both digital and ABF may be proposed for combining their benefits. Hence in the next section, we expound on analog, digital, and the state-of-the-art hybrid BF architectures.

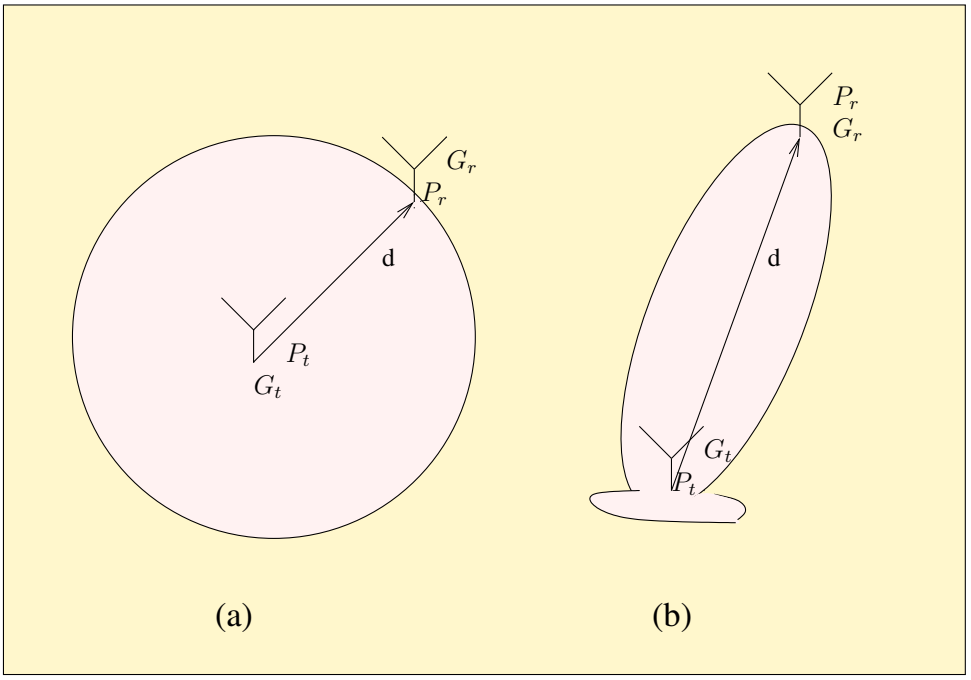
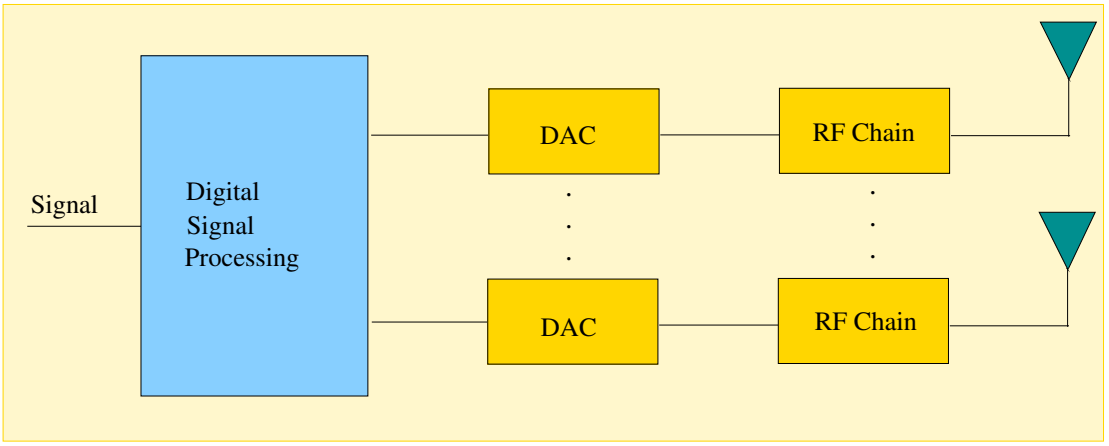


Figure 2.21: Transmission (a) Omnidirectional (b) Directional (BF).

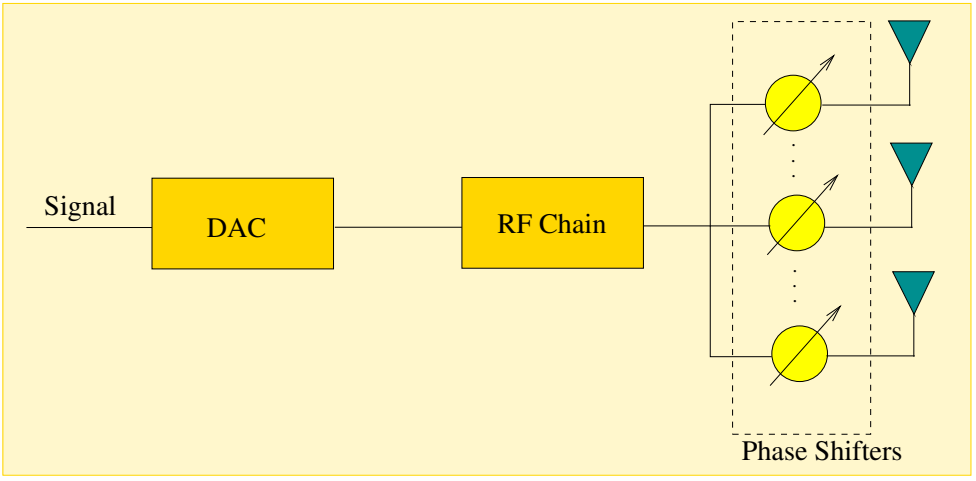
2.4.1 Conventional Beamforming

- **Digital Beamforming**: Figure 2.22 (a) shows the architecture of DBF, where the signal is processed using purely digital processing elements, with the specific array weights to adjust both the amplitude and phase of the signal. The resultant signal is then fed to the DAC for analog conversion, which is then upconverted at the RF stage and transmitted. Note that in Fig. 2.22 (a), a DAC and RF chain is dedicated to each and every AE. Furthermore, we have seen in Sec. 2.3 that for a spectral efficiency of 6 bps/Hz, a few hundred antennas are needed to achieve sufficiently high BF gain to compensate for the path loss. In other words, DBF having a few hundred antennas entails the same number of DACs at the transmitter and ADCs at the receiver. Additionally, we note that the sampling rate of the DACs/ADCs has to be increased proportionally with the bandwidth which in turn makes DACs/ADCs power hungry [62]. Given that large bandwidths are harnessed at mmWave frequencies, employing pure DBF would markedly increase both the hardware complexity and the power consumption — the power required for DACs/ADCs grows exponentially with the number of bits sample (resolution) — which scales with the bandwidth requirements. Nonetheless, conventional multi-input multi-output (MIMO) designed signal processing for frequencies below sub-6 GHz relies heavily on DACs/ADCs, because the associated bandwidth is lower than in mmWave systems.

- **ABF**: In contrast to the DBF scheme of Figure 2.22 (a), in the ABF arrangement shown in Fig. 2.22 (b), the signal is only processed in the analog domain. In



(a) DBF.



(b) ABF.

Figure 2.22: Conventional beamforming architectures: digital beamforming and ABF.

other words, the signal is directly fed to the DAC for analog conversion and then upconverted to the RF stage. The upconverted signal is then phase shifted using an array of phase shifters and then transmitted. In contrast to DBF, the ABF scheme of Fig. 2.22 (b) only has a single DAC and RF chain, which is connected to each and every AE through analog phase shifters. Thus this design exhibits high energy efficiency because of the absence of power-hungry DACs. However, the angular resolution is typically inaccurate in ABF [63], because of the quantized angular values used. Furthermore, the amplitude constraints imposed on the phase shifter would further erode the performance of the system in terms of its SNR gain. These constraints makes it a challenge to steer the nulls, adjust the sidelobes and to form multiple beams. While the angular resolution can be increased upon increasing the size of the AA [64], a hidden advantage of having constant-amplitude phase shifters is the absence of power imbalance, which is often caused because of the different average SNRs of the antennas. The performance may degrade if the average SNR is not identical across the AEs, since it may affect the diversity gain of the system.

Therefore, power imbalance is an undesirable phenomenon observed in wireless communication systems [65].

2.4.2 Hybrid Beamforming Architectures and Design

As discussed in the previous section, having DBF entails having an RF chain dedicated to each and every AE, which imposes both high cost and high complexity. On the other hand, employing analog only beamforming would introduce inaccuracies in terms of the angular resolution. Therefore, a hybrid beamforming [66] based on the amalgamation of analog and digital beamformers is expected to strike an attractive compromise. In this section, we commence our discussion by the state-of-the-art hybrid beamforming designs, namely the fully-connected design [4,5] and the sub-array connected design and then hybrid precoding techniques conceived for the proposed architectures.

2.4.2.1 Hybrid Architecture

- **Fully-connected Architecture**: In the fully-connected architecture [4], the signal in the baseband is precoded using a linear TPC. The precoded signal is then up-converted and fed to the analog phase shifters, which then is sent to the transmit antennas for transmission. However, in the fully-connected design, a full AA, where the AEs are separated by $\lambda/2$ distance, is connected to all the RF chains using phase shifters, as shown in the Fig. 2.23. The channel 'seen' by this design is spatially correlated across the antennas with an angular spread, of say δ for a given cluster. Quantitatively, the adjacent antennas which receive signals have a phase-difference of 180°. In this way, directional transmission relying on a beamforming order of N_t can be achieved at the transmitter.

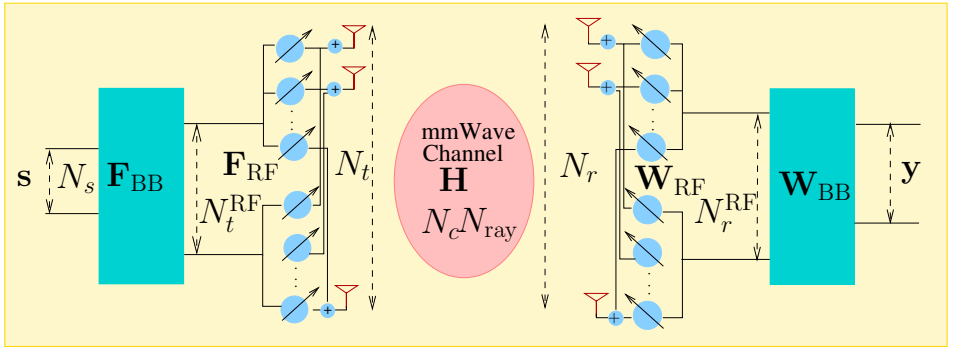


Figure 2.23: Hybrid architecture using fully-connected design.

Consider the single-user mmWave MIMO system shown in Fig. 2.23, where the transmitter is equipped with N_t transmit antennas and the receiver with N_r receive antennas. Then the received signal vector after RF and baseband combining is

given by:

$$\mathbf{y} = \sqrt{P}\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{n}, \tag{2.19}$$

where $\mathbf{H}$ is the statistical channel model expressed as:

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{N_c N_{\mathrm{ray}}}} \sum_{n_c=1}^{N_c} \sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}} \alpha_{n_c}^{n_{\mathrm{ray}}} \mathbf{a}_r(\phi_{n_c}^{n_{\mathrm{ray}}}) \mathbf{a}_t^{T}(\phi_{n_c}^{n_{\mathrm{ray}}}), \tag{2.20}$$

and the transmitter is equipped with N_t^{RF} chains and the receiver with N_r^{RF} chains, where $\mathbf{F}_{\mathrm{RF}}$ is the RF beamformer matrix of size $N_t \times N_t^{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$ is the baseband TPC matrix of size $N_t^{\mathrm{RF}} \times N_s$, $\mathbf{W}_{\mathrm{RF}}^{H}$ is the RF combiner matrix of size $N_r^{\mathrm{RF}} \times N_r$, $\mathbf{W}_{\mathrm{BB}}^{H}$ is the baseband combiner of size $N_s \times N_r^{\mathrm{RF}}$, while $\mathbf{y}$ is the $N_s \times 1$ received vector, $\mathbf{s}$ is the symbol vector of size N_s, where $N_s < N_t^{\mathrm{RF}}$ and $\mathbf{n}$ is the noise vector of identical and independent distributed entries with distribution $\mathcal{CN}(0, \sigma^2\mathbf{I})$. Furthermore, $\mathbf{H}$ is the statistical channel matrix of size $N_r \times N_t$ so that $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$, while $\alpha_{n_c}^{n_{\mathrm{ray}}} \sim \mathcal{CN}(0,1)$ is a complex-valued Gaussian random variable, whose amplitude and phase are Rayleigh and uniform distributed, respectively. For a uniform linear array (ULA) with N_r and N_t AEs the response vectors $\mathbf{a}_r$ and $\mathbf{a}_t$ are expressed as:

$$\mathbf{a}_r(\phi_r) = [1 \; e^{j\frac{2\pi}{\lambda}d\cos(\phi_r)} \ldots e^{j\frac{2\pi}{\lambda}(N_r-1)d\cos(\phi_r)}]^T, \tag{2.21}$$

$$\mathbf{a}_t(\phi_t) = [1 \; e^{j\frac{2\pi}{\lambda}d\cos(\phi_t)} \ldots e^{j\frac{2\pi}{\lambda}(N_t-1)d\cos(\phi_t)}]^T. \tag{2.22}$$

Finally, ϕ_t and ϕ_r, N_c and N_{ray} are the AoD and arrival, and the number of clusters and rays, respectively. To conceive 3D beamforming, the response vectors $\mathbf{a}_r(\phi_r), \mathbf{a}_t(\phi_t)$ can be extended to uniform planar arrays (UPA) as a function of the elevation angle.

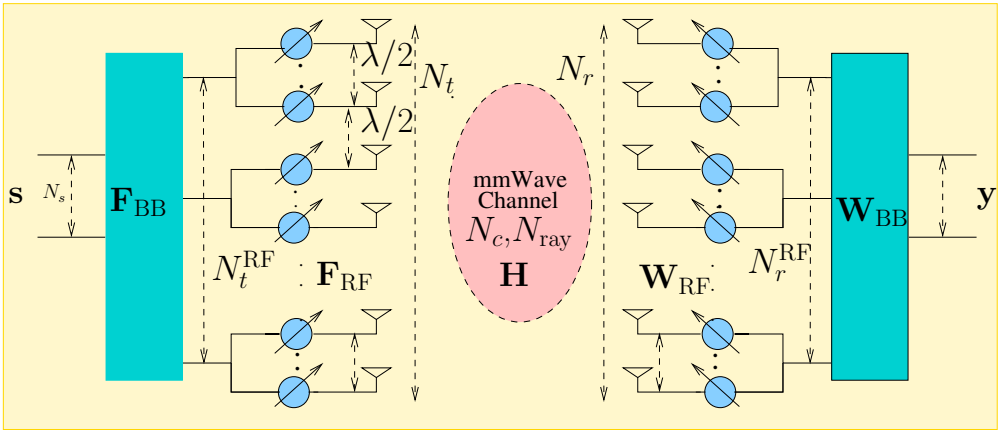


Figure 2.24: State-of-the-art sub-array-connected design.

- **Sub-Array-Connected Architecture:** In the arrays of the sub-array-connected design (ASA) of Fig. 2.24, the signal is precoded in the baseband using a linear

TPC. However, in contrast to the fully-connected design of Fig. 2.23, the precoded signals are fed to the phase-shifters of a sub-array before transmission. In contrast to the fully-connected design of Fig. 2.23, in ASA the AA is split into sub-arrays, where each sub-array is connected to an RF chain for digital processing, as shown in Fig. 2.24. The ASA reported in the state-of-the-art literature is constituted by *sub-arrays separated by a distance of* $\lambda/2$ from each other as shown in Fig. 2.24, where the channel is *spatially correlated across all the antennas*. The correlation between the sub-arrays depends on the angle of departure of the signal at the transmitter, which can be seen in the response vector ($\mathbf{a}_t^T$) in (2.20) [64]. More explicitly, $\mathbf{a}_t^T$ in (2.22) is a function of the AoD[2]. However, since the sub-array spacing is small (which is $\lambda/2$), the angular spread of the AoD from one sub-array to another sub-array would be very small, which results in increased correlation of the signals.

If the channel 'seen' by the sub-array i of the transmitter is $\mathbf{H}_i$, then the channel matrix of the design can be expressed as

$$\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_{N_{\mathrm{RF}}}],$$

where each of the sub-channels $\mathbf{H}_i$ is correlated with the other sub-channels H_j, $\forall i \neq j$. It is instructive to note that the number of phase shifters required by the ASA architecture is lower than that required by the fully-connected design. For example, with N_t transmit antennas and N_t^{RF} RF chains, the fully-connected regime would require $N_t N_t^{\mathrm{RF}}$ phase shifters, while the ASA design would only need N_t phase shifters.

In the next section, we discuss the state-of-the-art hybrid precoding designs for the aforementioned hybrid architectures.

2.4.2.2 Hybrid Precoding: A Survey

In the foregoing section, we discussed a pair of prominent hybrid beamforming architectures. We now focus our attention on the design of hybrid precoding techniques for the above-mentioned architectures.

In contrast to the DBF, where a single digital TPC is employed, in HBF, a digital TPC $\mathbf{F}_{\mathrm{BB}}$ is applied in the baseband and an analog beamformer $\mathbf{F}_{\mathrm{RF}}$ in the RF stage, as shown in Fig. 2.23.

There is a vast body of work in the literature on designing both the digital TPC and the analog beamformer matrix. The design of both the $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{F}_{\mathrm{RF}}$ matrices is typically

[2]When the AoDs are similar, the rank of the channel matrix $\mathbf{H}$ in (2) is less than the full rank, which means that the channel is correlated.

carried out jointly as a hybrid product that mimics the optimal digital solution, which is the right singular matrix of the channel matrix $\mathbf{H}$. Mathematically, this is formulated as

$$\min_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}} \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2. \tag{2.23}$$

$$s.t. |\mathbf{F}_{\mathrm{RF}}(m,n)|^2 = 1. \tag{2.24}$$

It is important to note that decomposing the matrix $\mathbf{F}_{\mathrm{opt}}^{(j)}$ into the product of $\mathbf{F}_{\mathrm{RF}}^{(j)}$ and $\mathbf{F}_{\mathrm{BB}}^{(j)}$ is not straightforward because of the constraints imposed on the $\mathbf{F}_{\mathrm{RF}}^{(j)}$ matrix, where the entries in $\mathbf{F}_{\mathrm{RF}}^{(j)}$ must have constant modulus to avoid having power imbalance in the system. Therefore, the state-of-the-art designs mainly focus on appropriately designing the matrices $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ which closely approximates the optimal precoder $\mathbf{F}_{\mathrm{opt}}$.

Early designs of hybrid precoding relied heavily on compressive sensing based algorithms [67–69]. More explicitly, Ayach *et al.* [67] proposed a technique for jointly designing the TPC and analog beamformer by exploiting the sparsity of the mmWave channel. To evaluate the hybrid product, the authors invoked Basis pursuit method, where the entries of the ABF matrix constitute the columns of the array response vectors. As a further solution, Alkhateeb [68] *et al.* developed an iterative solution by employing the basis pursuit variant relying on the partial knowledge of the channel response. Similarly, Kwon [69] *et al.* designed the ABF matrix by selecting the N_{RF}^t array response vectors relying on the basis pursuit, while the digital TPC is formed by an optimal combination of array response vectors. For a planar array, a subspace construction algorithm is proposed in [70], where the subspace is constructed by the Kronecker product of the horizontal and vertical spatial features of the array. In this design, the authors advocated the spatial covariance matrix based ABF, while ZF precoding is used as the digital TPC.

Other compressive sensing based hybrid precoding techniques are proposed in [71–74]. More particularly, by exploiting the sparse nature of the mmWave CIR in the angular domain, the ABF matrix of (2.23) can be reformulated as a sparse recovery problem, where the classic orthogonal matching pursuit (OMP) algorithm [75] relying on a codebook constituted by array response vectors having quantized angular values is employed. In other words, the columns of the ABF matrix are formed from the codebook of array response vectors. By exploiting the angular-domain structured sparsity of mmWave CIRs, Gao [76] *et al.* proposed a grid matching pursuit strategy associated with adaptive measurement matrix. The authors of [67] invoke the Basis pursuit algorithm to obtain the ABF matrix. However, in contrast to [67], the authors of [76] design the digital TPC for ensuring that it minimizes the sum-MSE of all data stream; thereby circumventing the step of deriving the fully-digital solution.

Sohrabi [77] *et al.* discussed a heuristic algorithm, where the analog beamformer matrix is determined iteratively by employing the gradient descent algorithm, while the digital TPC is obtained as the classic water-filling solution. By formulating (2.23) as a manifold optimization, Xiang [78] *et al.* developed a low-complexity alternating minimization method for determining the analog beamformer matrix. More particularly, the authors proposed the phase extraction method of [78] relying on semidefinite relaxation. On the other hand, the classic least squares solution is invoked to obtain the digital precoder, which minimizes the Euclidean distance between optimal and hybrid precoder for a fixed analog beamformer. In [79], the authors transformed (2.23) into a simultaneous matrix orthogonalization problem, where they employ the Jacobi algorithm [80] to arrive at the ABF matrix. Then the unconstrained digital precoder is obtained from the effective channel by minimizing the Euclidean distance.

In [81], Zhang *et al.* also exploited the sparsity of the mmWave CIR by utilizing the truncated higher order SVD [82] for evaluating the analog beamformer matrix. The digital TPC is then obtained by applying the truncated SVD to the effective channel. A Gram-Schmidt orthogonalization based greedy algorithm has been proposed by Alkhateeb *et al.* [83], where the ABF vectors were iteratively selected from a given codebook by maximizing the mutual information. Then the SVD of the effective channel is carried out to derive the digital TPC. While the authors of [84] proposed a minimum mean squared error based solution for both the digital and analog BF matrices, Rusu *et al.* [85] minimized the mean squared error between the optimal digital solution and the analog precoder for a fixed digital TPC, where the digital TPC is obtained by reformulating (2.23) as an orthonormal Procrustes problem [86]. In [87], Cai *et al.* relied on traditional methods such as ZF/MRT, where the authors set the magnitude of the ABF matrix to unity and then selected the phase as that of the ZF precoder of the channel matrix. The digital precoder is then obtained again by applying the ZF technique to the effective channel matrix.

In frequency-division duplex (FDD) systems, the receiver typically has to feed back the channel estimate to the transmitter for the TPC to compensate for the channel-effects about to be encountered during the next transmission. However, feeding back the channel coefficients to the transmitter would result in a high overhead. An appealing practical solution for feeding back the channel information to the transmitter is based on a carefully designed quantized TPC weight codebook, which is known both to the transmitter and the receiver. Explicitly, based upon the CIR estimate the receiver selects the best TPC matrix from its codebook, and reports the selected index to the transmitter over the feedback channel. This technique is referred to as limited-feedback design.

Song *et al.* [88] proposed a codebook design for mmWave systems using the classic OMP algorithm. In [89], Alkhateeb and Heath proposed efficient RF codebook designs for limited feedback mmWave systems, where the ABF matrix is designed to maximize the received signal power, while the digital TPC is obtained by minimizing the interference

between the users. By exploiting the channels' reciprocity, a heuristic codebook design was advocated by Eltayeb *et al.* [90]. In [91] the authors employed the semidefinite relaxation based greedy algorithm for analog codebook design, while linear precoding is applied on the effective channel for digital TPC. In [92–95], the authors discussed the classic DFT based ABF, while ZF precoder and gradient descent algorithm are invoked in [92, 93] and [94], respectively. The rationale for choosing the DFT codebook is that when the channel is spatially correlated, the codewords of the DFT fall in the statistical distribution of the optimal precoder solution. A greedy hybrid precoding algorithm relying on the Gram-Schmidt orthonormalization for limited feedback frequency selective systems is proposed in [96], where the codebook that minimizes the average mutual information is designed. Table 2.3 shows the summary of the state-of-the-art hybrid precoding designs reported in the literature, where the solution of each design aims to decompose the hybrid product of (2.23). Although there is a plethora of hybrid precoding designs discussed in the literature, the aforementioned designs impose a high complexity in terms of computations. Moreover, the performance of a practical codebook design depends on the choice of the codebook entries, as detailed by Clerckx *et al.* [97], which will be detailed in Chapter 3. In other words, *the efficiency of a practical hybrid precoding design critically hinges on its performance versus complexity trade-off*, which will also be discussed in Chapter 3.

Table 2.3: Summary of the state-of-the-art hybrid precoding designs, where each design aims for decomposing the hybrid product of (2.23).

Paper	**Solution Method**	**Precoder Design**	**Codebook Design**
[67–69]	Basis pursuit	✓	
[70]	Subspace construction	✓	
[71–74]	OMP	✓	
[76]	Grid matching pursuit	✓	
[77]	Heuristic algorithm	✓	
[78]	Alt. minimizing algorithm	✓	
[79]	Matrix orthogonalization	✓	
[81]	Truncated higher order SVD	✓	
[83]	Gram-Schmidt based iterative		✓
[84, 85]	MMSE	✓	
[87]	ZF/MRT	✓	
[88]	OMP		✓
[89]	Maximizing signal power		✓
[90]	Heuristic		✓
[91]	Semidefinite relaxation		✓
[92, 93]	DFT		✓
[95]	DFT + MRT		✓
[94]	DFT + Gradient Descent		✓
[96]	Min. mutual information		✓

In Sec. 2.3 we have analyzed the number of AEs required to compensate for the path loss in order to achieve the desired spectral efficiency, while in and Sec. 2.4, we have discussed

the state-of-the-art hybrid beamforming solutions. In the next section, we provide an energy-efficient HBF design, where the number of AEs per user is judiciously activated based on the his/her propagation loss. Then, we also present switching off the digital precoder when the channel exhibits line-of-sight (LOS).

2.5 Improved Energy-Efficient HBF

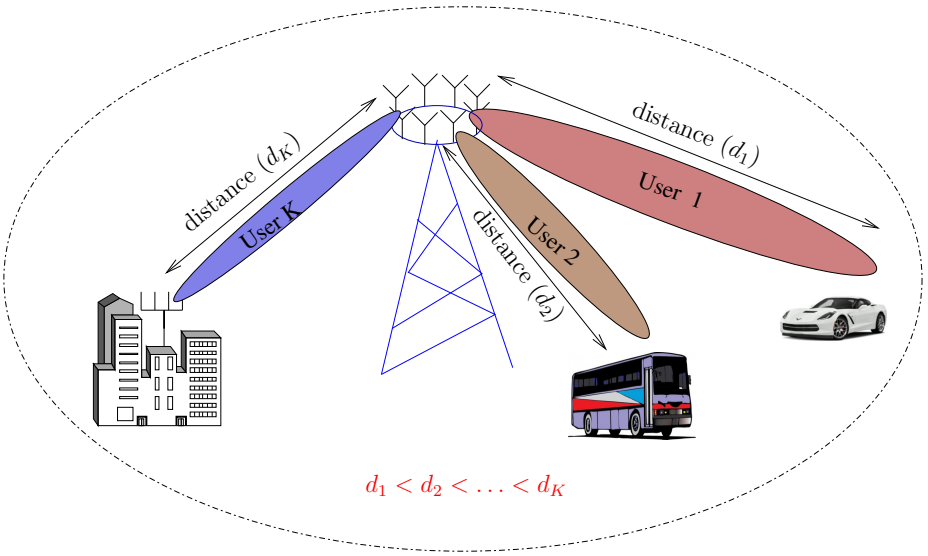


Figure 2.25: System Model.

Let us consider the BS communicating with K users, each equipped with N_u antennas and N_u^{RF} RF chains, where the users may be geographically separated from each other, as shown in Fig. 2.25. In this design, the BS is equipped with N_t antennas and N_t^{RF} chains, where the BS processes the signal digitally using N_t^{RF} chains in the baseband and then the processed signal is phase shifted using $N_t^{\mathrm{RF}} N_t$ phase shifters in the RF stage before its transmission from the N_t antennas, as shown in Fig. 2.26. This design is referred to as fully-connected hybrid beamforming, where every RF chain of the design in Fig. 2.26 is connected to N_t AEs using N_t phase shifters. It is also important to emphasize that the attainable BF gain is dictated by the number of active AEs, which is equivalent to the number of active phase shifters, since the output of the phase shifters is fed to the AEs as shown in Fig. 2.26.

It is instructive to note that the aim of the BF in mmWave systems is to compensate for the propagation loss involved. Therefore, in our design of Fig. 2.26 the number of phase shifters active at any given time is distance-dependent. This is because the BF gain required to compensate for the propagation loss for each user may be different, since it is dependent on the user's distance from the BS. Let us consider Fig. 2.25 again as a 'toy' example. In this figure, user 1 and user 2 are located at distances of d_1 and d_2, respectively, from the BS, where $d_1 < d_2$. Since user 1 is closer to the BS, the propagation

loss[3] experienced by the user 1 is lower than that of user 2, which is farther from the BS. As a result, the BF gain required to compensate for the loss is higher for user 2 than that of user 1. Therefore, the number of active phase shifters required at the BS for user 1 in order to compensate for the path loss is lower than the number of active phase shifters needed for user 2. This philosophy makes our design more energy-efficient than the conventional design where all the phase shifters remain active at a given time. Furthermore, in the scenario where there are more than one beams for a given user, BIM for that user may be employed to further enhance the throughput of the system.

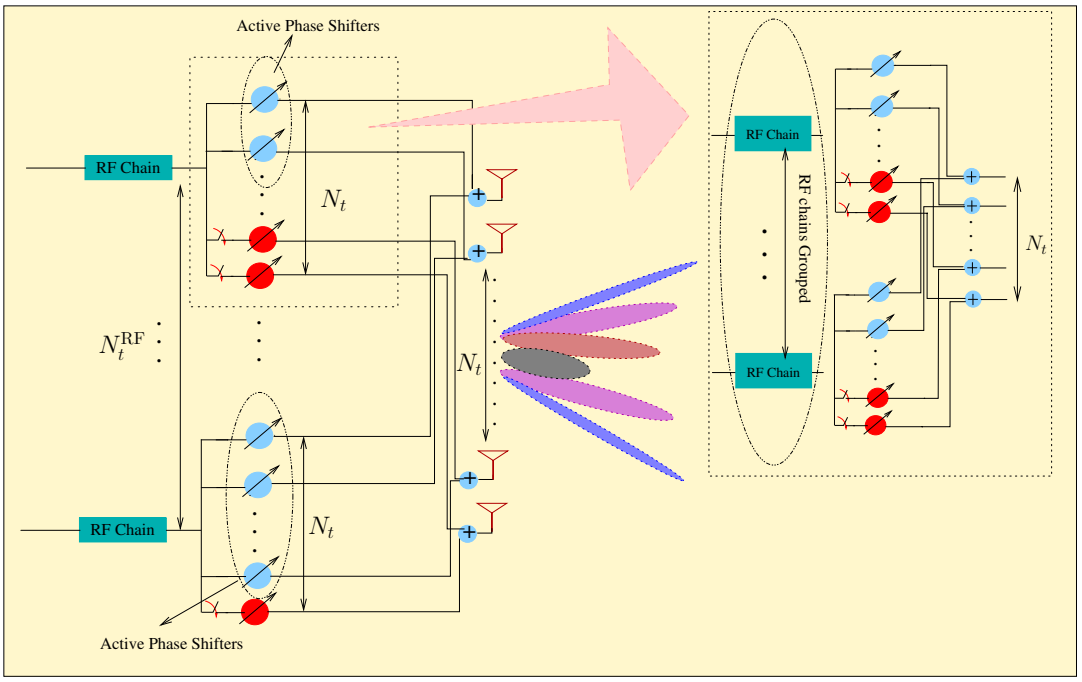
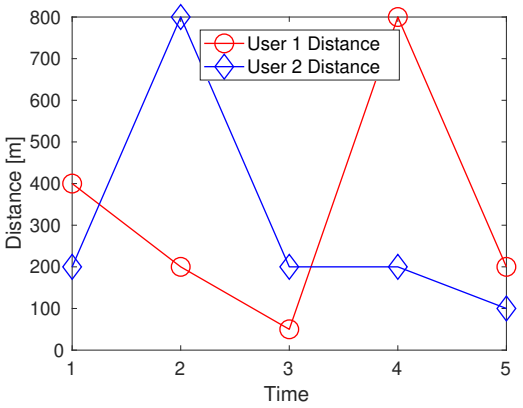


Figure 2.26: Pictorial illustration of the number of active phase shifters at any time. Note that the design on the left side of the figure shows a single RF chain serving a user by employing a specific number of active phase shifters. However, there may be a group of RF chains that may serve multiple users that share similar channel conditions. The design on the right side of the figure shows the grouping of RF chains to serve a specific group of users. It can also be interpreted as a set of fully-connected RF chains with the same number of active phase shifters to serve a user .

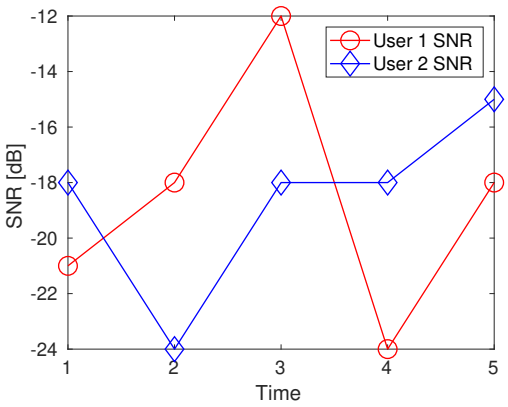
We also note that the beam of user 2 is narrower than that of user 1 as a benefit of having a higher-gain radiation pattern due to having more active phase shifters for user 2 so as to compensate for his/her higher propagation loss [1]. In other words, a high BF gain is achieved by activating a large number of phase shifters.

To expound further, let us consider Fig. 2.26, which shows the active phase shifters at the BS at a given point of time. Note that the BS seen in the figure shows a fully-connected

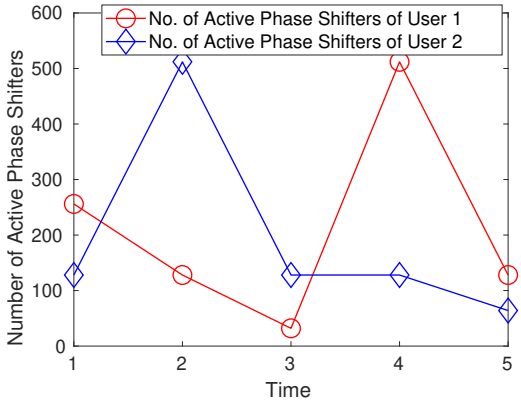
[3]The propagation loss considered in this paper is free-space loss; however, in practice, other large-scale fading factors such as foliage density, attenuation due to rain-induced fading, and shadowing should also be considered. Our design will still work when considering all these factors, but we opted to focus on the free space loss for the sake of simplifying the discussion.



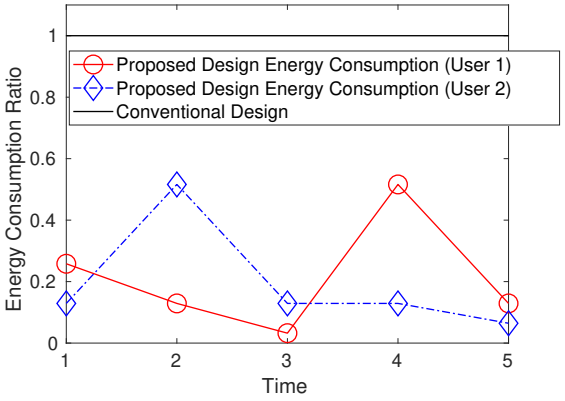
(a) Distance of users' from the transmitter.



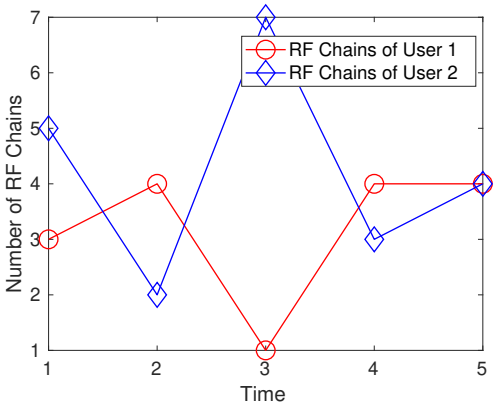
(b) SNR of users' versus time.



(c) Number of phase shifters activated per user versus time.



(d) Energy consumption ratio of the proposed design with respect to the conventional design, where all the phase shifters remain active all the time, in terms of the usage of phase shifters.



(e) Number of RF chains grouped to serve each user versus time.

Figure 2.27: Graphical illustration of the variation in the number of active phase shifters depending on the user's distance and SNR, and of the the number of RF chains grouped to serve each user as a function of time.

design, where all the phase shifters are connected to all the transmit antennas and each RF chain is connected to all the phase shifters. It can be seen in the figure that the number of active phase shifters shown in blue color of the first and the last RF chains are different, since the BF gain required for the respective users is different.

Having discussed the number of active phase shifters, we now focus our attention on the specific allocation of the RF chains, where more than one RF chain may be connected to the same number of active phase shifters[4] in a fully-connected fashion as shown at the right side of Fig. 2.26 [4,98]. In other words, a plurality of RF chains grouped together may either serve a single user or a group of users. It is important to emphasize that first we assign a single RF chain per user at the BS and only the remaining RF chains will be distributed accordingly. More explicitly, the assignment of multiple RF chains to each user at the BS depends on the availability of beams. This is because the number of RF chains grouped together to serve a user is equal to the number of beams available for transmission to that user. More explicitly, if the user's channel is capable of supporting a single beam, then the number of RF chains in the group cannot be more than one. Hence, for the sake of fairness, the total number of RF chains, which is N_t^{RF}, at the BS is grouped in such a way that more RF chains are grouped for serving users having more possible beam directions in order to grant them additional degrees of freedom in the angular domain. However, the maximum number of RF chains in a group is limited, which will be discussed later in the paper. Note that this design can also be readily extended to sub-array-connected design [4,37].

To elaborate a little further, Fig. 2.27(a) illustrates the distance of users 1 and 2 from the BS terminal versus time. In this illustration, we assume that the BS is equipped with 1024 AEs and 32 RF chains, where each RF chain is connected to 1024 phase shifters[5]. It can be seen from Fig. 2.27(b) that as the users' distance of Fig. 2.27(a) increases from the BS, the SNR of each user decreases because of the propagation loss. This loss can be compensated by BF gain, where the number of activated phase shifters required to perform BF increases with the propagation loss. This becomes evident from Fig. 2.27(c), which illustrates the number of active phase shifters for the users of Fig. 2.27(a). As an example, let us consider user 1 of Fig. 2.27 (a) at time instant 4, where the user is 800 meters (m) away from the BS, while the SNR observed is -24 dB, as shown in Fig. 2.27(b). Accordingly, the number of active phase shifters needed to compensate for the path loss and to achieve an SNR of 3 dB is 512[6], as presented in Fig. 2.27(c). In other words, 512 of the 1024 phase shifters connected to a RF chain are required to achieve a BF gain of $10\log(512) \approx 27$ dB [1]. Similarly, observe for user 2, who is at a distance of 200 m, only 128 of the 1024 phase shifters have to be activated for achieving same SNR of 3 dB. In contrast to the conventional design, where all 1024 phase shifters are

[4]Note that the rest of the phase shifters are switched off.

[5]The total number of phase shifters is equal to 1024*32= 32768.

[6]The mathematical relationship between the number of AEs required and the BF gain value is given by Eq. (13) of [1].

activated regardless of the users' distance, our design becomes more energy-efficient by appropriately adapting the number of active phase shifters, which will become explicit from Fig. 2.27(d).

It can be seen from Fig. 2.27(d) that the energy consumed by the users of our design is markedly lower than that of the conventional design, where all the phase shifters remain active all the time. More explicitly, in the conventional design, all phase shifters are used without considering the BF gain requirements, hence wasting energy owing to its higher-than-necessary BF gain. By contrast, our design activates exactly the required number of phase shifters, while attaining exactly the required BF gain. For example, we have seen that user 1 at time instant 4 would utilize 512 out of 1024, while user 2 utilizes only 128 out of 1024 phase shifters, which corresponds to 50% and 19%, respectively, of the energy consumed by the conventional design. The energy consumption ratio (ECR) is calculated as the ratio of the number of active phase shifters to the total number of phase shifters, given by

$$\text{ECR} = \frac{\text{Number of active phase shifters}}{\text{Total number of phase shifters}}. \tag{2.25}$$

Observe in Fig. 2.27(d) that energy consumption ratio of the conventional design is 1, while it varies for our design depending on the number of activated phase shifters.

Having judiciously activated the required number of phase shifters, the RF chains are arranged as groups to serve users depending on the channel conditions of the user. Let us again consider Fig. 2.27(e), where at time instant 4, the number of RF chains serving user 1 is 4 out of the available 32, while 3 out of the available 32 RF chains invoked for serving user 2. The choice in the number of RF chains that form a group is decided by the number of potential beam directions of the users. However, the maximum number of RF chains serving a user depends on the number of RF chains allocated to the preceding group. For example, let us assume there are 8 users and each user has 10 potential beam directions. Let us furthermore assume that there are 32 RF chains. In this scenario, for the sake of fairness, every user is served by 4 RF chains. On the other hand, for example, user 2 has 3 beams, which means that it can be served by a maximum of 3 RF chains, because having additional RF chains would be redundant[7]. So in this setting, user 4 may be served by 5 RF chains since the user has 6 more additional beams available for data transmission.

The flowchart of the proposed design is presented in Fig. 2.29. First the BF gain required for each of the K users is calculated, as detailed in [1]. Then we compute the number of AEs necessary to compensate for the propagation loss. Having obtained the number of active AEs required, the RF chains are appropriately distributed at the BS to serve individual users. However, first we assign a single RF chain per user at the BS and only then will the remaining RF chains be distributed, where the number of beam-pairs

[7] The number of RF chains is equal to the number of beams for transmission.

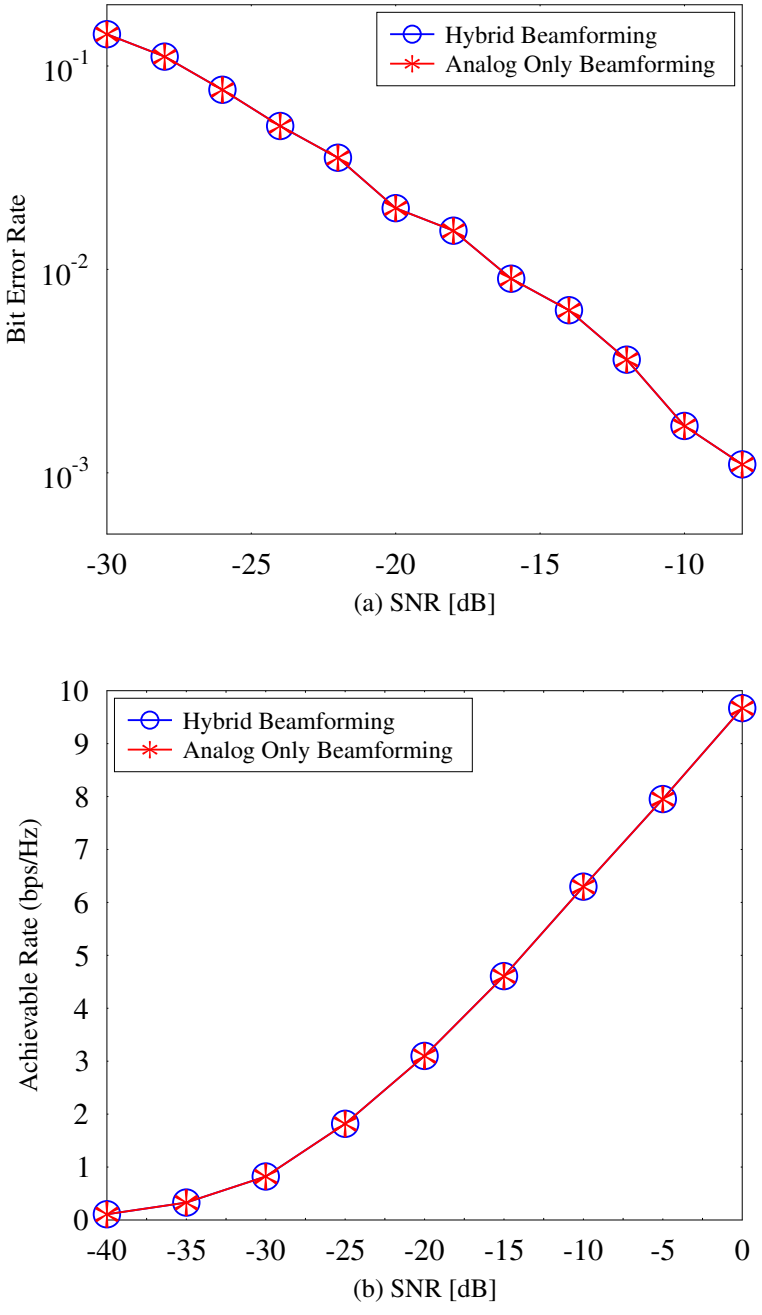


Figure 2.28: BER and achievable rate with HBF (analog and digital) and using only analog BF, when the channel has only one dominant path.

available for communication is decided. Having allocated the RF chains, then link-adaptation may also be carried out for every user, which we will discuss in Chapter 5.

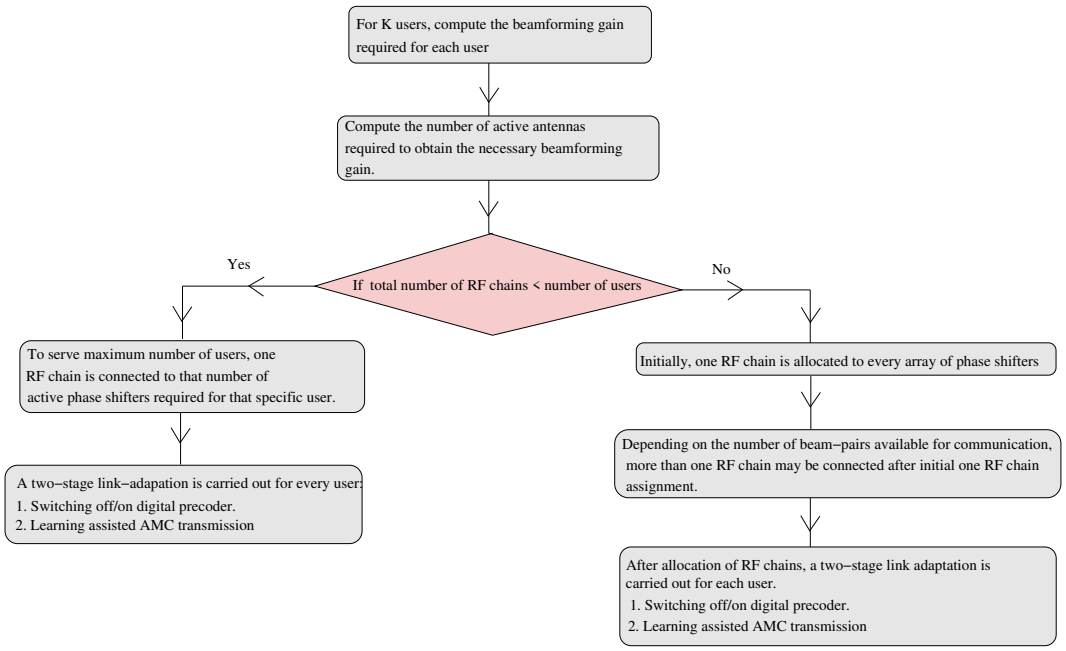


Figure 2.29: Flowchart of the proposed design.

Although at mmWave carrier frequencies can have a high contiguous bandwidths, the efficient use of spectral resources is crucial to meet the escalating data rate demands. Therefore, in the next section, we focus our attention the employment of full duplex communication at mmWave frequencies, which would further improve the attainable spectral efficiency.

2.6 Full Duplex Communications

It is generally assumed that bidirectional radio communication in the same frequency band is not possible because of the high self-interference that results from the transmitter's own transmission at the local receiver. However, the research conducted in [6, 7] proposed a solution to this challenge, where simultaneous transmission and reception using the same carrier frequency was demonstrated, which is termed as full-duplex (FD) wireless communication. Henceforth, the FD communication philosophy has attracted the attention of the wireless communication researchers, given its potential to double the spectral efficiency. However, the significant amount of self-interference (SI) at the local receiver tends to prevent FD communication from attaining satisfactory performance gains. Fig. 2.30 shows the interference levels at various stages of the full duplex transceiver. To elaborate more, for a transmitted signal power of 20 dBm as shown in

the Fig. 2.30, the SI signal constitutes mainly of three components: the linear component of 20 dBm strength, which is the main signal; the non-linear components of -10 dBm strength [3]; and the transmitter noise of -40 dBm strength [3].

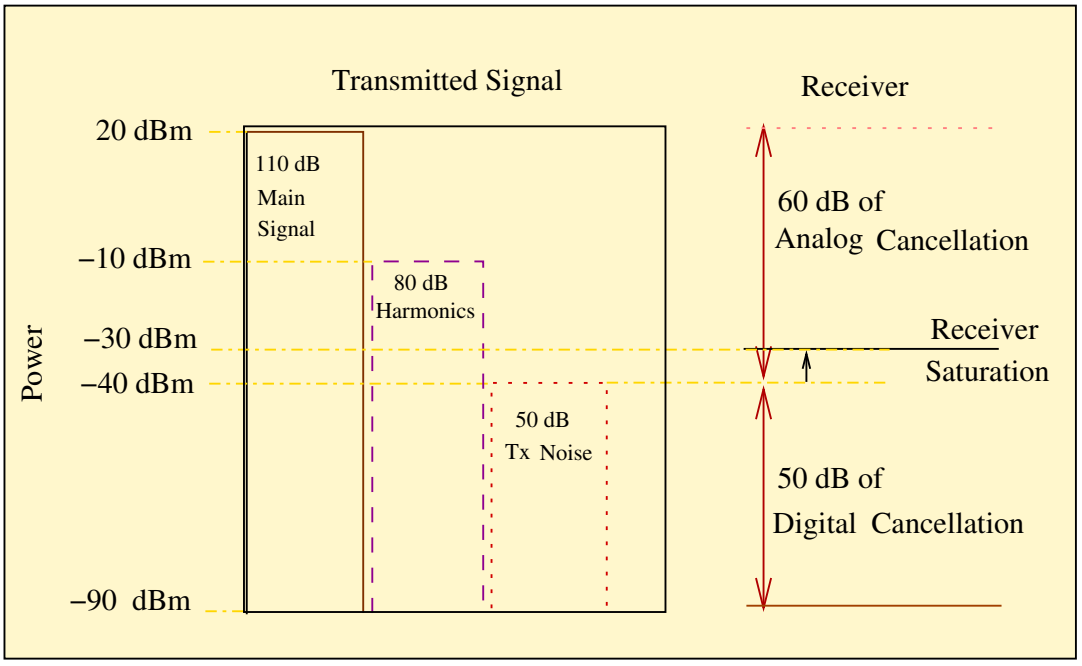


Figure 2.30: Interference levels at various stages of the full duplex transceiver.

Typically, the SI cancellation at the local receiver is carried out by a combination of passive and active methods [6, 8]. The passive methods aim for increasing the path loss of the interfering signal through antenna isolation so as to reduce the SI power [9]. By contrast, the active methods rely on the knowledge of the transmitted signal and they tend to rely on three different approaches [10], namely RF cancellation, antenna cancellation and digital cancellation, as shown at the right of Fig. 2.30. In the RF cancellation, the known transmitted signal at the local transmitter is used as a reference signal in the RF chain to reconstruct and then to subtract the SI. This can be achieved by estimating the leakage-path transfer function, which is required for the reconstruction of the leaked signal in support of its cancellation. In the antenna cancellation, two replicas of the transmitted signal having opposite phases are generated using multiple transmit and receive antennas and the SI is removed by adding the pair of opposite replicas [11]. In addition to RF and antenna cancellation, digital cancellation is typically employed together with the RF or antenna cancellation to further suppress the SI in the baseband [11, 12]. Fig. 2.31 depicts the various processing stages involved during the SI cancellation in a FD receiver. In Fig. 2.31 SI cancellation levels of about 60 dB in the analog domain and of about 50 dB in the digital domain are desirable. However, the harmonics arising due to the presence of nonlinear components in the power amplifier degrade the quality of SI cancellation, resulting in a significant residual SI.

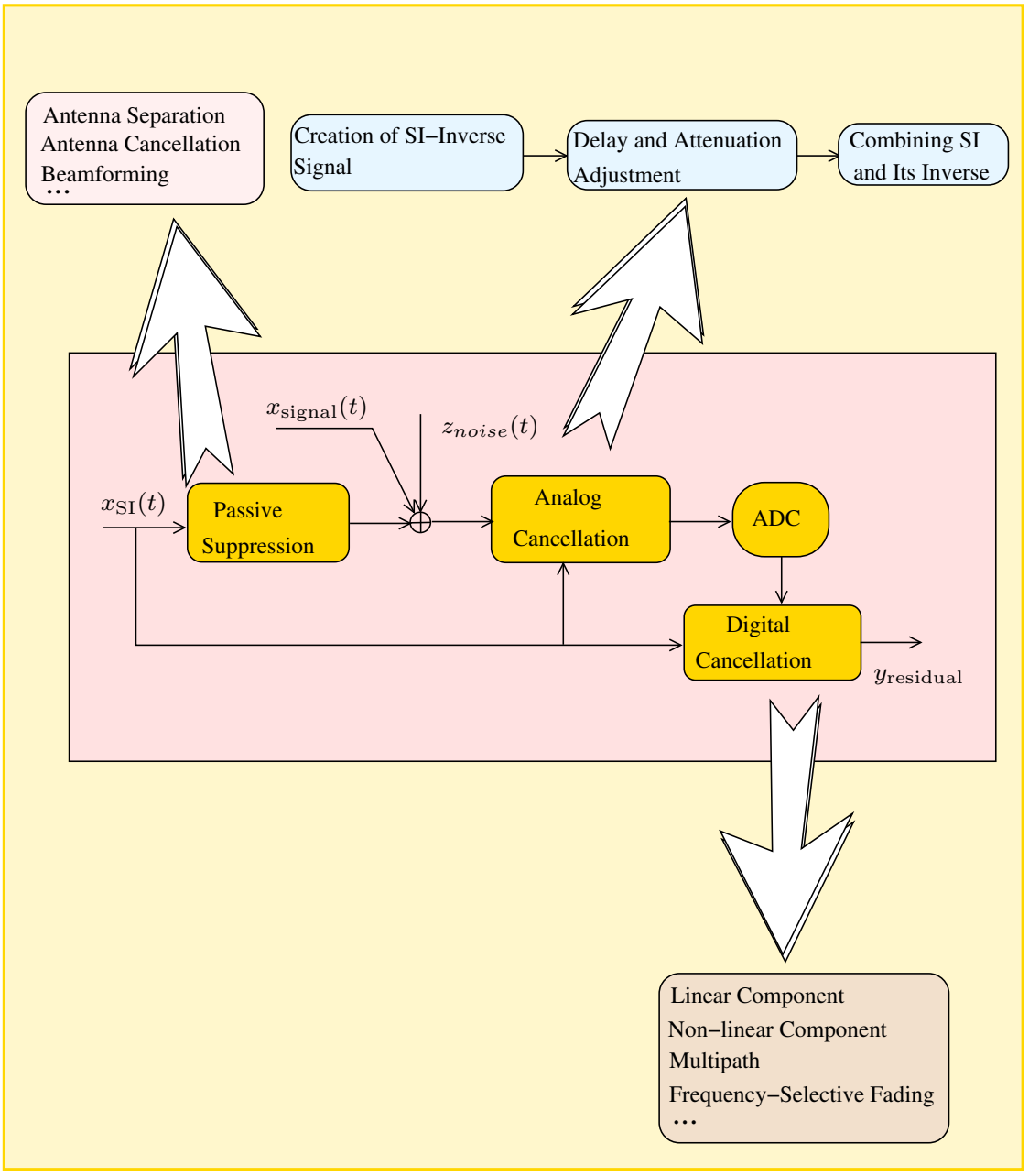


Figure 2.31: Stages of SI cancellation in a typical full duplex receiver.

However, the benefits obtained by cascading RF/digital cancellation for mitigating the SI may remain limited because of the distortions imposed at the various stages, such as the power amplifier non-linearity, I/Q imbalance and phase noise [3]. The SI levels may further increase in the face of the imperfections in the analog- or (and) digital-domain cancellation(s). These impairments become particularly pronounced in mmWave communications because of the high-level non-linearities of the ADCs/DACs at mmWave frequencies [13]. Hence, when considering these realistic non-linearities caused by the ADCs/DACs, power amplifiers, I/Q imbalance and phase noise, the residual SI become quite high. In the state-of-the-art contributions on the FD systems, most of the work

has been focused on SI cancellation designed for MIMO-aided relaying systems [14–18], while some other related work includes dynamic resource allocation conceived for FD systems [19]. More particularly, Riihonen *et al.* [14] analyzed a broad range of MIMO mitigation schemes, such as natural isolation, time-domain cancellation, and spatial suppression. Furthermore, the authors of [14,99] observed that employing MMSE filtering in conjunction with the above-mentioned schemes efficiently mitigates the SI. A FD cooperative network relying on a pair of two amplify-and-forward relays is considered in [15], where the authors proposed distributed space-time coding to mitigate the SI. More explicitly, the dispersive SI channel was interpreted as the action of a convolutional FEC code at each relay node and then the signals from pair of relay nodes was deemed to form a space-time code. Ju *et al.* [100] designed the precoder and decoder matrices relying on the simplifying assumption of having perfect CSI. As a further solution, Lin *et al.* [18] proposed a joint MMSE TPC constituted by the pair of precoders at the relay and source in order to reduce the residual SI level. The authors conceived an iterative solution by splitting the original problem into two subproblems. In other words, the authors designed the precoder at the source and relay separately by formulating into a pair of dedicated convex optimization problems.

A similar approach has been reported by Nguyen *et al.* [101, 102], where the authors proposed designing the TPC matrix by maximizing the spectral-efficiency, while also accounting for the residual SI associated with the FD MIMO systems. In this design, the authors invoked a sequential convex approximation method for designing the TPC matrix. As an extension, Nguyen also presented a TPC solution, which maximizes the energy-efficiency by formulating and solving the resultant concave-convex fractional program. In [103], Mohammadi *et al.* proposed a beamforming solution for FD relay systems, where maximum ratio transmission (MRT) and maximum ratio combining (MRC) reception is carried out at the relay. More specifically, the authors aim for designing the weight vectors for both the MRT and MRC schemes by maximizing the instantaneous throughput of the system. The optimization problem is solved by reformulating it into semidefinite relaxation to arrive at the sub-optimum solution; by contrast, Cirik *et al.* [104] designed the TPC and receiver combiner matrices that maximize the harmonic-sum of the SINRs. The authors relied on an iterative distributed algorithm for finding a local maximum.

In addition to the joint TPC and detection techniques presented in the state-of-the-art designs, Suraweera *et al.* [105] proposed antenna selection for minimizing the SI, where the specific antenna combination that provides maximum SINR between the source-to-relay and relay-to-destination links is selected. Furthermore, to eliminate the zero-diversity behavior observed in FD systems, the authors advocated a simple power allocation design at the relay. There is a vast body literature that focused extensively on the transmit power allocation optimization of FD systems. More particularly, in FD MIMO relay

systems, the transmit power allocation at the relay node is typically designed by exploiting the observation that the end-to-end performance is limited by the weakest link. Therefore, by noting that the end-to-end performance can be improved by decreasing the transmit power at the relay node, Riihonen *et al.* [106, 107] proposed a transmit power adaptation scheme for mitigating the residual SI level. To further enhance the spectral efficiency, the authors of [106, 107] conceived an opportunistic adaptation between half-duplex (HD) and FD modes of transmission based on the instantaneous CSI. The effects of the transmit power level on FD MIMO relay systems in terms of the coverage range attained is studied in [108]

Additionally, to combat the inter-cell interference in FD MIMO systems sophisticated interference management techniques have been proposed in [109]. The main idea is to bin the transmissions over complimentary channel realizations such that the inter-node interference occurring at the mobile receivers during one time slot can be completely negated by the inter-node interference occurring at another time slot. In [110], interference alignment is invoked to suppress both the SI and the inter-cell interference.

More recently, Wang *et al.* [20] has derived a closed-form expression for the achievable rates of FD MIMO relaying systems communicating over Rician fading channels, when linear receivers are employed. Considering the limitations of the input circuitry, Day *et al.* derived upper and lower bounds on the achievable sum rate for FD MIMO in [21]. Everett *et al.* [22] studied the performance of passive self-interference suppression for FD, where the authors demonstrate that as high as 70 dB of SI suppression is possible in certain environments.

As far as FD communication is concerned, there is a paucity of literature on mmWave communications relying on FD techniques. Additionally, employing beamforming would further mitigate the SI. More recently, Xiao *et al.* [10] proposed a beamforming based FD scheme for mmWave communication, where the authors designed the transmit and receive beamforming weights to cancel the SI. In [111], Snow *et al.* demonstrated that SI cancellation of upto 40 dB using DBF can be achieved relying on multiple coordinated transceivers. Despite the application of both the analog and digital cancellations, there is still a significant residual SI. Therefore, using SI cancellation relying on beamforming techniques is crucial at mmWave frequencies in addition to the conventional analog and digital SI suppression methods in the face of the strongly non-ideal nature of the ADCs/DACs and power amplifiers, which result in high residual SI [13]. Unfortunately, the designs in [10, 111] cannot be directly extended to HBF systems in mmWave communications, given that they are proposed for a single RF chain, hence their extension to multiple RF chains becomes mathematically intractable. Table 2.4 summarizes the state-of-the-art designs that aim for mitigating the residual SI relying on diverse digital signal processing techniques.

Table 2.4: Summary of popular state-of-the-art designs that aim for mitigating the residual SI relying on digital signal processing techniques.

Papers	Residual SI Cancellation Type	Solution
[14, 18, 99]	Digital signal processing	MMSE Filtering
[15]	Digital signal processing	Convolutional space-time coding
[101]	Digital signal processing	Sequential convex approximation method
[105]	Antenna selection	Maximum SINR
[106, 107]	Power allocation	Maximum throughput
[108]	Power allocation	Maximum coverage range
[109]	Interference management	Interference alignment
[103]	DBF	Semidefinite relaxation
[104]	Digital precoding	Distributed iterative algorithm
[111]	DBF	Implementation
[10]	DBF (mmWave)	Maximizing sum rate

2.7 Machine Learning: Primer

Beyond its powerful prediction and analytical capability, machine learning can be a major driver of intelligent data driven wireless communications and networks. Therefore, in this section, we focus our attention on the family of machine learning tools that may become the driving force for next-generation communications systems.

2.7.1 Classification of Machine Learning Techniques

Broadly speaking, machine learning is a computational technique that relies on experience to improve the performance of a system or to make accurate predictions. The *experience* refers to previously encountered data available for learning. By integrating sophisticated tools from computer science, statistics and probability, machine learning analyses the available data and conceives efficient algorithms for accurate predictions [112]. Since it is an experimental method, the success or accuracy of predictions is heavily dependent on the data used during learning. Furthermore, given its dependency on data, machine learning imposes data-volume-dependent complexity storage requirements.

Although there is a plethora of learning strategies, machine learning conceived for wireless communications can be commonly classified into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [112]. The key differences among the different learning methods is the order and the choice of the method by which training and testing data are used. We now briefly describe each learning method. We use the terminology 'learner' to describe a machine which learns.

1. **Supervised learning** [112]: In this method, the data samples used during training are labeled[8]. In other words, known data are used for training. Having trained the

[8]In machine learning labels refer to a particular class or category. For example, the training samples may consists of data that makes distinction between classes A and B.

learner, predictions are applied to unclassified data, where the learner classifies the hitherto unseen, unclassified data. Common examples associated with supervised learning are regression, classification, and ranking problems.

2. **Unsupervised learning** [112]: It is a self-organizing based learning method, where the training samples are not labeled, i.e., they are unclassified or uncategorized, which is in contrast to unlike supervised learning. The absence of labeled data in unsupervised learning makes it difficult to quantitatively evaluate the performance of a learner. Common examples associated with this type of learning are clustering and dimensionality reduction.

3. **Semi-supervised learning** [112]: In this class of learning, the training samples comprise both labeled and unlabeled data — typically more unlabeled data than labeled. The philosophy is that supplying the learner with a small amount of labeled data in conjunction with a large amount of unlabeled data can produce predictions with increased accuracy. Common examples associated with this semi-supervised learning includes classification, regression and also ranking related problems.

4. **Reinforcement learning** [112]: It relies on iterating between training and testing phases. More explicitly, the learner acts as an agent by interacting with the environment and reaping immediate rewards for each beneficial action. The aim of the agent is to maximize its reward over a period of interactions with the environment. Because of its broad appeal, reinforcement learning is generally studied in fields such as game theory, control theory, etc. where the agent, i.e., the learner strikes a balance between exploration and exploitation. In other words, the learner explores the uncharted territory by exploiting the hitherto accumulated knowledge.

Let us commence by briefly highlighting the key challenges involved during learning.

Let us define a linear regression as [113]

$$\hat{\mathbf{y}} = \mathbf{w}^T \mathbf{x}, \tag{2.26}$$

where $\mathbf{w}$ is the parameter that controls the system's behavior, while $\mathbf{x}$ and $\hat{\mathbf{y}}$ are the input and predicted vectors, respectively.

Now the learner makes predictions based on the input vector $\mathbf{x}$ by adjusting the weight $\mathbf{w}$, so that the training error defined as

$$\mathrm{E}_{\mathrm{train}} = \frac{1}{m} \sum_i (\hat{\mathbf{y}}_{\mathrm{train}} - \mathbf{y}_{\mathrm{train}})_i^2, \tag{2.27}$$

is minimized, where $\mathbf{y}$ is the known vector during training of the i^{th} sample, $\hat{\mathbf{y}}$ is the predicted vector of the i^{th} sample, and m is the total number of training samples.

Having designed the parameter $\mathbf{w}$ which minimizes the error during training, then testing data is used for evaluating its performance on m_{test} testing samples, which is expressed as

$$\mathrm{E}_{\text{test}} = \frac{1}{m_{\text{test}}} \sum_i (\hat{\mathbf{y}}_{\text{test}} - \mathbf{y}_{\text{test}})_i^2. \tag{2.28}$$

We note that the regression line (2.26) always passes through the origin. However, in general linear regression represents more sophisticated model, which also takes into account an additional parameter 'b'. This parameter is commonly referred to as an intercept, and thus the model of (2.26) represents an affine transformation expressed as

$$\hat{\mathbf{y}} = \mathbf{w}^T \mathbf{x} + \mathbf{b}. \tag{2.29}$$

In machine learning, it is desirable to achieve as low a testing and training error as possible, while keeping the difference between the testing error and the training error low. There are two undesirable scenarios that arise when these requirements are not met: underfitting and overfitting. To elaborate a little further, underfitting occurs when the training error based on the training set is not sufficiently low, i.e. the predictions made based on the training set are inaccurate. On the other hand, overfitting occurs when the training error is low, but the testing error is high. This generally happens when the training dataset is too small; in other words, the difference between the training error and test error is large. Fig. 2.32 illustrates the model fitting using machine learning, where the underfitting is shown on the left and the overfitting on the right, while the most appropriate fitting is presented between the two.

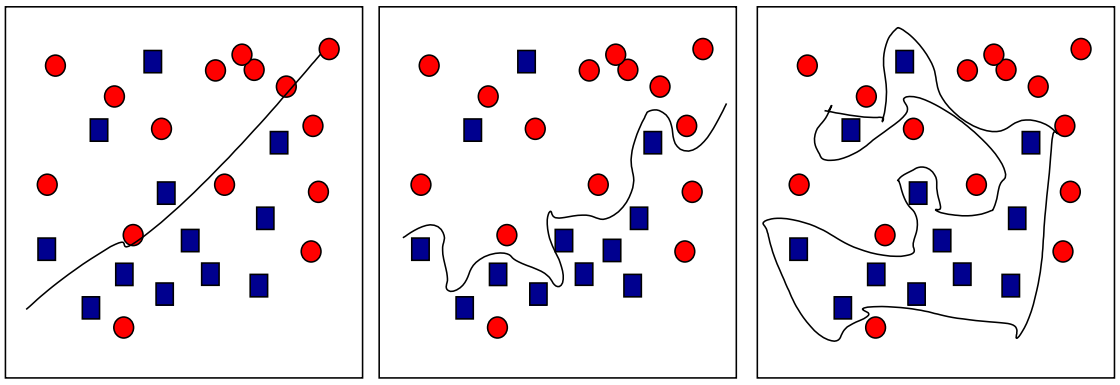


Figure 2.32: Model fitting using machine learning.

To avoid both the underfitting and overfitting scenarios, the optimum design of parameters is paramount. Diverse learning frameworks can be formed in the literature but

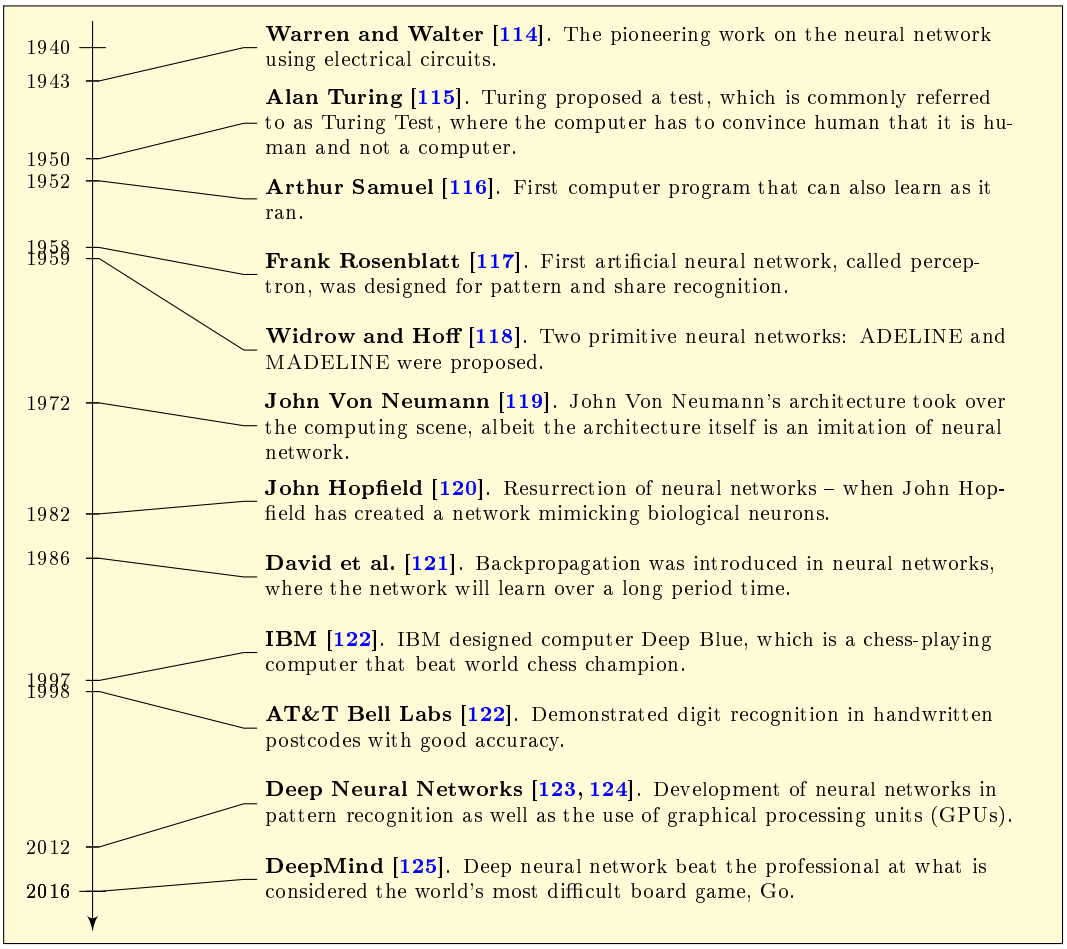


Figure 2.33: Timeline of important milestones in the area of neural networks.

artificial neural networks (ANN) constitute the most popular machine learning tools because of their ability to mimic human intelligence. ANNs are inspired by the structural and functional aspects of biological neural networks. ANN, given their ability to learn and generalize, they have become one of the major pillars of machine learning. Fig. 2.33 shows the timeline of important milestones in the area of neural networks. Following the major developments in neural networks from as early as 1943, we find that the first ANN was designed in 1958 by Frank Rosenblatt.

The operation of an ANN is akin to that of the human brain in terms of organizing neurons to evaluate certain operation. A biological neuron typically consists of the nucleus, axons and dendrites, as shown in Fig. 2.34. The node which connects two or more neurons is known as the synapse. Axons serve as a conduit between two neurons, i.e. the signal traverses from one neuron to the next through the axons, as shown in Fig, 2.34. As the signal travels from one neuron to the next, the membrane potential of the latter is altered. If the change in the potential exceeds a certain threshold, the neuron would send signals to all other neurons connected to it [127]. In this way, the signal propagates through the human nervous system. Analogous to the biological neural network, an ANN is comprised of several interconnected nodes. Furthermore, like the

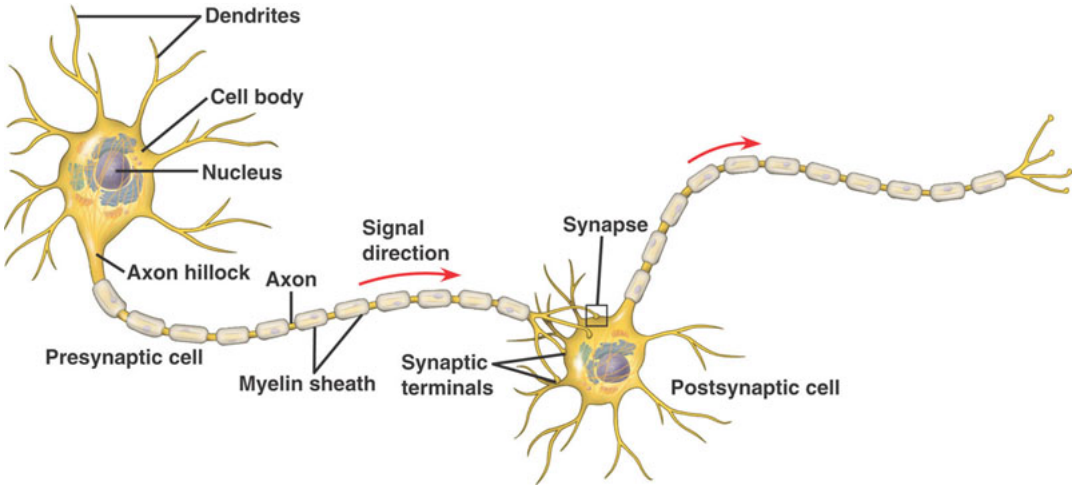


Figure 2.34: Human biological neuron [126].

connections emerging from dendrites and converging at the synapse, an ANN has a number of incoming and outgoing connections, as shown in Fig. 2.35. Furthermore, similar to the neuron's membrane potential, an activation value (score) is computed for each neuron.

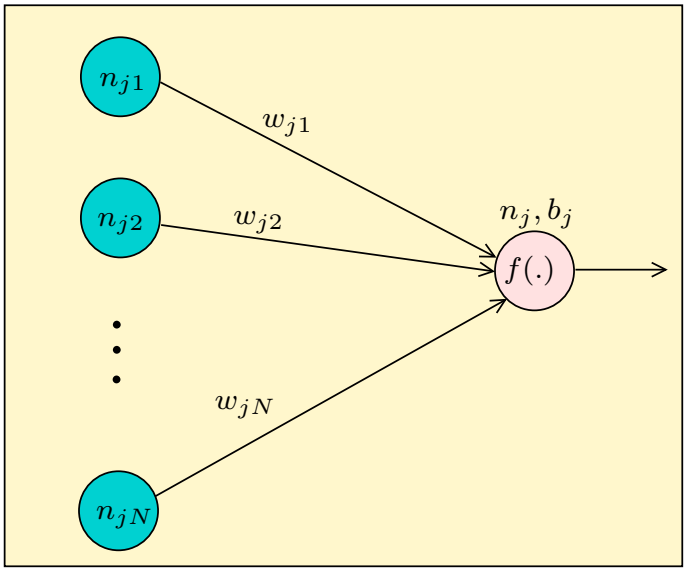
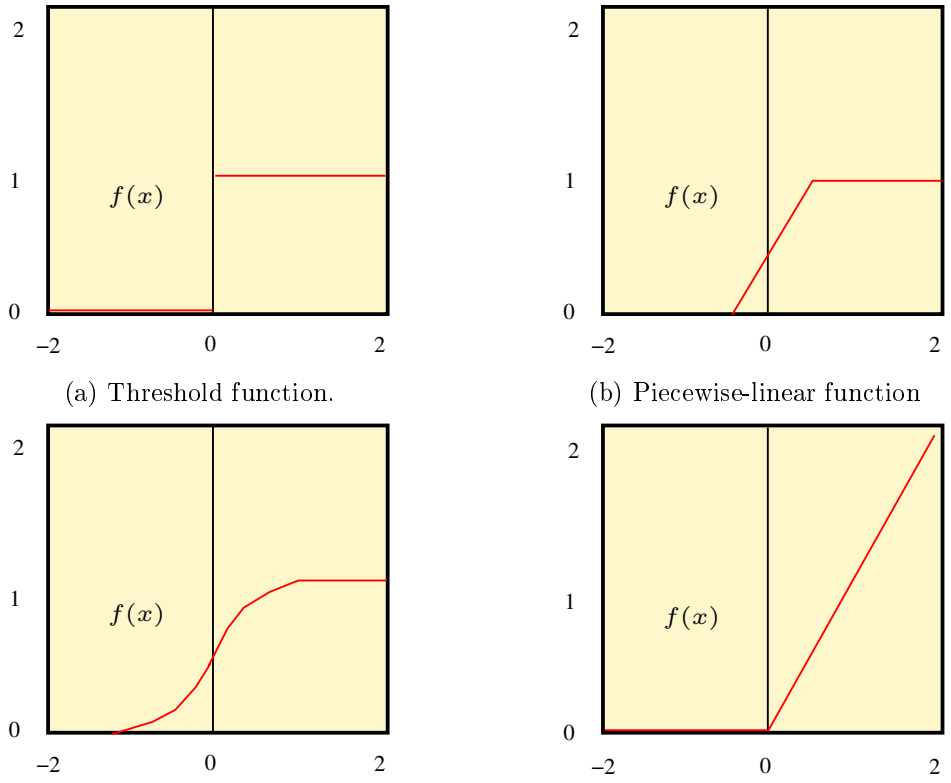


Figure 2.35: ANN model analogous to biological neural network.

The connection strength between two neurons of biological neural networks is represented by a weight value in ANNs, as shown in Fig. 2.34. Having captured the connection strengths from all the incoming signals to a particular neuron n_j, the activation value o_j, which is analogous to a membrane's potential, is expressed as

$$o_j\left(\mathbf{w}_j, b_j, \mathbf{n}_j\right) = f\left(b_j + \sum_{k=1}^{N} n_{jk} w_{jk}\right), \tag{2.30}$$

where n_{jk} and w_{jk} are the input signal and the connection strength from the k^{th} neuron to j^{th} neuron, N is the total number of connections to the neuron n_j, while b_j is the bias associated with the neuron n_j. It is instructive to note that the main objective of an ANN in machine learning is to infer the unknown functional mapping between the input and the output connections. As previously mentioned, this functionality entails modification of the synaptic weights between the neuron connections, so that the both the training error and testing error are minimized.



(a) Threshold function.

(b) Piecewise-linear function

(c) Sigmoid function.

(d) ReLu function.

Figure 2.36: Types of activation function.

In ANNs, the choice of the activation function is of salient importance, since it is the driving force of the entire neural network. The key properties of an activation function are: (a) the function has to be bounded for a bounded input; (b) the activation function should be differentiable on $\mathbb{R}$. Some examples of activation functions are: (a) threshold function; (b) piecewise-linear function; (c) Sigmoid function; (d) ReLu function, which are shown in Fig. 2.36.

1. **Threshold function** [128]: Mathematically, the threshold function depicted in Fig. 2.36(a) is given by [128]

$$f(x) = \begin{cases} 1, & \text{if} \quad x \geq 0. \\ 0, & \text{if} \quad x < 0. \end{cases} \tag{2.31}$$

   The threshold function's output is either 1 or 0. Therefore, this may be employed in binary classification problems, where the desired output is either 1 (yes) or 0 (no). However, it is unsuitable for non-binary classifications.

2. **Piecewise-linear function**: The piecewise-linear activation function of Fig. 2.36(b) is given by [128]

$$f(x) = \begin{cases} 1, & \text{if} \quad x \geq 0.5. \\ x, & \text{if} \quad -0.5 < x < 0.5. \\ 0, & \text{if} \quad x \leq -0.5. \end{cases} \tag{2.32}$$

   Like the threshold function, this function may also be used for binary classification, since it outputs 0 for values below -0.5 and 1 for values above 0.5. Additionally, it can also support non-binary classification problems, since it follows a steep gradient between -0.5 and 0.5. Since the gradient is steep, any small changes in the input of the activation function would affect the output significantly — which means that this function has a tendency to bring the output to either end of the curve. Furthermore, the function is bounded between 0 and 1, which means that it avoids the unbounded escalation of the activation function.

3. **Sigmoid function**: This is the most commonly used activation function in the ANN because of its smoothness and asymptotic properties. It is sometimes also referred to as the squashing function, since it limits the amplitude range, as shown in Fig. 2.36(c) and expressed as [128]

$$f(x) = \frac{1}{1 + \exp(-wx)}, \tag{2.33}$$

   where w is the slope parameter. This is similar to piecewise-linear function, however, the difference is in the gradient between -0.5 and 0.5. The gradient in this region is small and changes slowly. This may give rise to vanishing gradients. For example, any small change in the input between the values -0.5 and 0.5 of the activation function may not see any change in the output. In other words, the network may refuse to learn or change from the previous state.

4. **Rectified Linear Unit (ReLu) function** [128]: Mathematically, the ReLu function is defined as [129, 130]

$$f(x) = x^+ = \max(0, x). \tag{2.34}$$

ReLu is a type of piece-wise linear function. However, unlike the bounded piecewise linear function of Fig. 2.36(b), this may saturate the activation function Nonetheless, the advantage of ReLu activation function over the Sigmoid and Tanh functions is that it always has a non-zero gradient for x^+. Furthermore, the function does not compute exponentials or division, which makes the ReLu function simpler to evaluate.

Although the sigmoid/ReLu functions have a better appeal amongst the other activation functions, the choice of activation function may also depend on the objectives such as analytical tractability and the nature of the output signal. For example, to analyze an ANN in terms of its performance versus complexity, a piecewise-linear function or a threshold function may be invoked.

Remark: If we observe the activation functions illustrated in Fig. 2.36, their common property is non-linearity. If the activation is linear, then we may not need any hidden layers, simply because a function of linear functions remains always linear, which can be attained without any hidden layers in the network.

Fig. 2.35 shows that the neuron n_j is connected to other multiple neurons; in general an ANN is comprised of a layers of neurons connected to multiple neurons of multiple layers. A classic ANN commonly referred to as the feedforward neural network is shown in Fig. 2.37. The first and the last layers of an ANN are referred to as the input and the output layers, respectively, while the layers between them are hidden layers. If the number of hidden layers is more than one, then the neural network is referred to as a deep neural network (DNN). The input signal is transmitted to the network through the first layer and the neurons in the output layer represent the output signal, while the hidden layers mimic the human intelligence with the aid of activation functions. There exists a plethora of ANNs depending on the connection type between the neuron. For example, in Fig. 2.37, the incoming connections of a layer are from the previous layer and the output connections are to the subsequent layer. However, there are ANNs, where the output connections of a layer may also be connected to its own layer and other layers. Fig. 2.38 shows different types of ANNs depending on the type of connection between the neurons. Let us now expound on these networks in little more detail.

2.7.2 Feedforward Neural Networks

A neural network is said to be a feedforward neural network, if the incoming connections of a neuron are from the immediately preceding layer, while the outgoing connections are

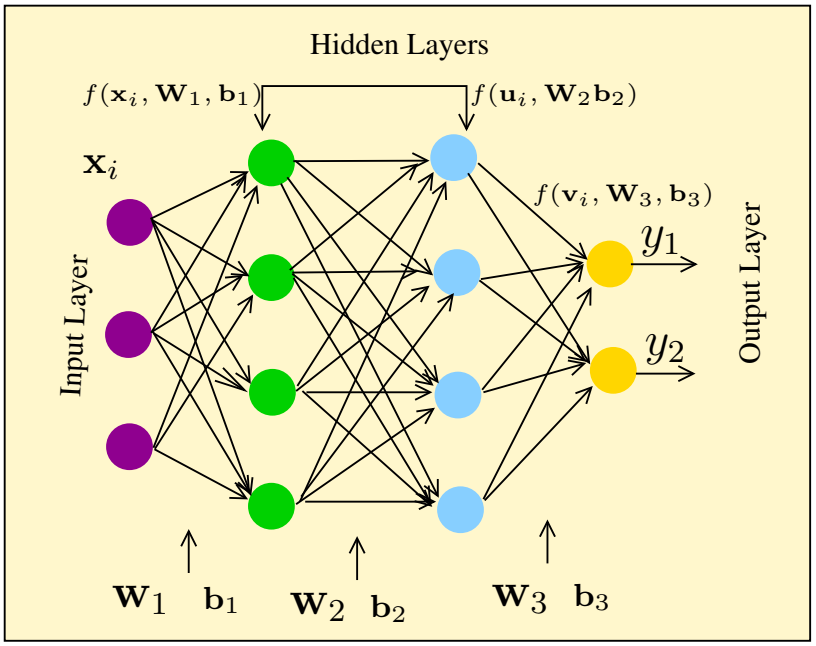


Figure 2.37: Feedforward neural network.

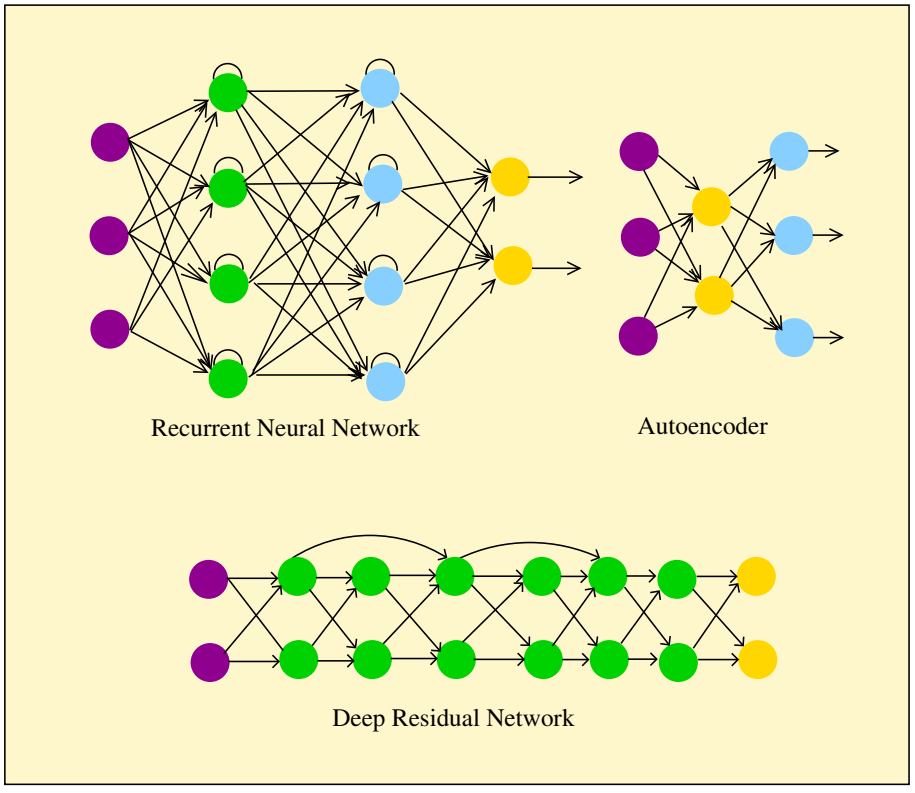


Figure 2.38: Types of ANNs depending on the connection between the layers.

to the layer succeeding immediately to it. For brevity, let us consider the feedforward neural network of Fig. 2.37 with only a single layer — the training in the supervised learning is carried out by minimizing the error between the desired output and the predicted output. This is typically quantified by a so-called loss function. The most commonly used loss function is the mean squared error, which is given by [112,131]

$$\mathrm{LF} = \frac{1}{\mathcal{S}} \underbrace{\sum_{i=1}^{\mathcal{S}} \left\| \mathbf{y}_i - \mathbf{y}_i^d \right\|}_{mean\ squared\ error} + \underbrace{\rho \|\mathbf{W}_r\|_2^2}_{regularization\ to\ avoid\ over\ fitting}, \tag{2.35}$$

where $\mathcal{S}$ is the total number of training samples, $\mathbf{y}$ is the predicted output, $\mathbf{y}^d$ is the desired output, while ρ is the regularization factor used to avoid over-fitting [131], as illustrated in Fig. 2.32. Having introduced the loss function, the weights are then designed for minimizing the loss function. For the output layer, the weights may be designed by adopting the gradient descent algorithm approach. For a neuron j in the output layer, the weights w_{jk} and bias b_k are updated by applying the gradient descent algorithm on the loss function, i.e., by taking derivative of the loss function with respect to each w_{jk}, yielding

$$w_{jk} = w_{jk} - \gamma \frac{\partial \mathrm{LF}(\mathbf{w}_j, b_j)}{\partial w_{jk}}, \tag{2.36}$$

$$b_j = b_j - \gamma \frac{\partial \mathrm{LF}(\mathbf{w}_j, b_j)}{\partial b_j}, \tag{2.37}$$

where γ is the learning rate of the gradient descent algorithm. Thus, the weights in the outer layer are updated according to (2.36). Note that these weights are only for the output layer of ANN. For the layer preceding it, the weights are updated by using back-propagation. In back-propagation, the gradient of the loss function with respect to the weights and biases is computed for each layer. We will now briefly discuss the back propagation for the layer preceding the output layer. However, similar procedure can be followed for any number of layers. A more detailed discussion on back-propagation is presented by Chauvin *et al.* in [132]. For neuron i in the layer preceding neuron j, the error propagation value δ_i is formulated as

$$\delta_i = \frac{\partial \mathrm{LF}(\mathbf{W}_i, b)}{\delta o_i} \frac{\delta o_i}{\delta net_i}, \tag{2.38}$$

where o_i is the output of neuron i, while net_i is the input signal of neuron i. Moreover, the value δ_i in (2.38) depends on the activation function used. For example, if the activation function is the sigmoid, then δ_i is given by

$$\delta_i = \begin{cases} (o_i - o_d) o_i (1 - o_i), & \text{if i an output neuron,} \\ \left(\sum_{l \in L} \delta_l w_{li} \right) o_i (1 - o_i), & \text{if} i \text{is an inner neuron,} \end{cases}, \tag{2.39}$$

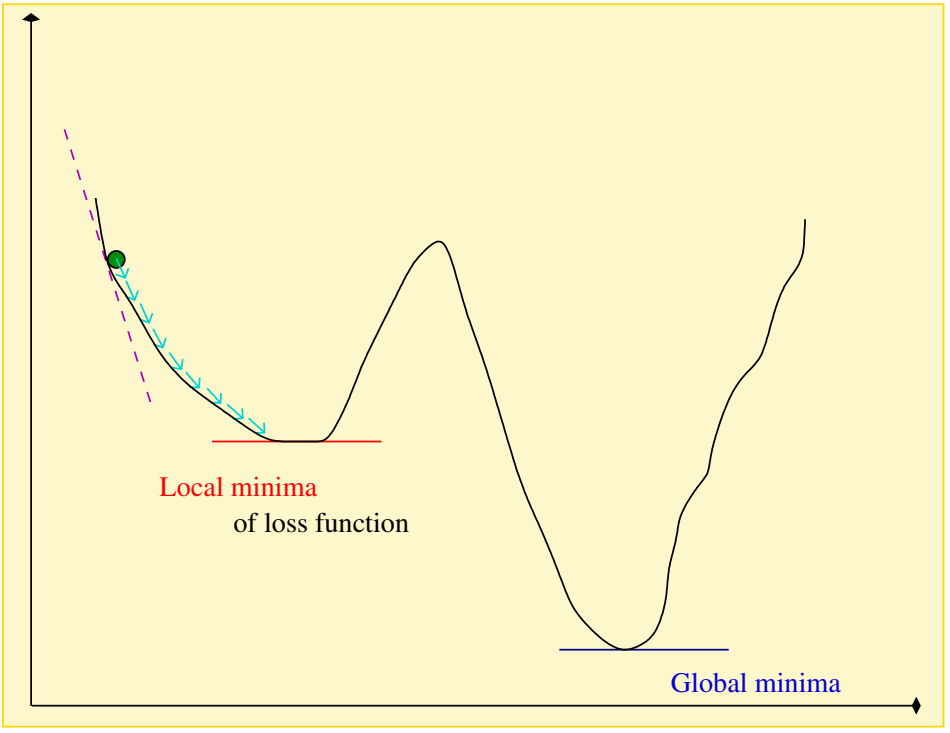


Figure 2.39: Illustration of gradient descent optimization.

where l is the index of the layer with respect to the output layer, o_i is the output of neuron i, and o_D is the desired output. Then, the weights are updated as

$$w_{ij} = w_{ij} - \gamma \delta_i o_j. \tag{2.40}$$

The weights are continued to be updated until convergence is reached. It is instructive to note that the weights obtained may not necessarily be optimal, since the gradient descent algorithm can rapidly converge to a local optimum, as shown in Fig. 2.39.

Furthermore, the rate of convergence of the algorithm also depends on the learning rate γ. A large γ value may miss the optimum solution, while a small γ will lead to slower convergence. To circumvent this problem, other gradient descent algorithms which are independent of the choice of the learning rate have also been proposed, such as Adagrad and AdaDelta [133]. It is instructive to note that the regularization term in (2.35) ensures that the neural network provides appropriate fitting, i.e., it avoids overfitting. However, the choice of regularization function used in the loss function is critical for producing accurate predictions. The most commonly used regularization functions are the L_1 norm and L_2 norm.

- L_1 *-Norm*: Let us consider the loss function in (2.35) by replacing the regularization term by the L_1-norm of the weight matrix, by $\|\mathbf{W}\|$. Then, our loss function can

be written as [112, 131, 134]

$$\text{LF} = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \left\| \mathbf{y}_i - \mathbf{y}_i^d \right\| + \rho \|\mathbf{W}\|. \tag{2.41}$$

This leads to a convex optimization problem, since the quadratic term in the above objective function is convex and $\|.\|$ is convex for all norms. Thus, this objective function be equivalently written as

$$\text{LF} = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \left\| \mathbf{y}_i - \mathbf{y}_i^d \right\| \tag{2.42}$$

$$\text{s.t.} \ \|\mathbf{W}\| \leq t, \tag{2.43}$$

where t is a positive parameter.

We observe that the objective function in (2.42) is quadratic and its contours are ellipsoids. The left side of Fig. 2.40 shows the ellipsoid as well as the area corresponding to the L_1-norm of the radius t. We know that the solution is given by the intersection of the contours with the L_1-norm. Observe from the figure that it can be where some coordinates are zero, hence giving us a sparse solution [134]. The physical significance of L_1-norm based regularization is that its solution may have a few non-zero values, which means that it may remove some features from the solution. Therefore, L_1-norm based regularization may be the best choice for feature selection problems.

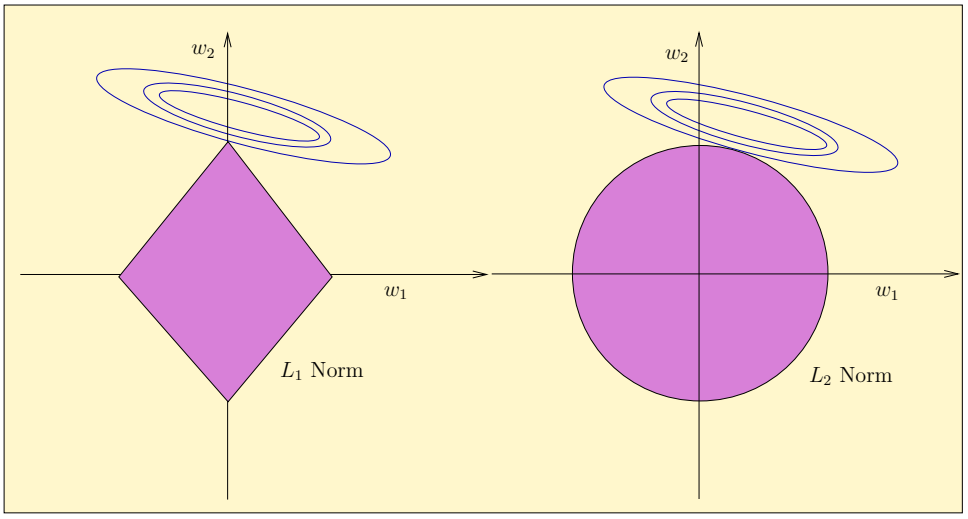


Figure 2.40: L_1 and L_2 norm regularization in neural networks.

- *L_2-Norm*: Let us now consider the loss function in (2.35) by using the L_2-norm of the weight matrix, $\|\mathbf{W}\|_2^2$. Then, our loss function can be written as [112, 131, 134]

$$\text{LF} = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \left\| \mathbf{y}_i - \mathbf{y}_i^d \right\| + \rho \|\mathbf{W}\|_2^2. \tag{2.44}$$

  This also leads to a convex optimization problem, since the quadratic term in the above equation is convex and $\|.\|$ is convex for all norms. Thus, this objective can further be written as

$$\text{LF} = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \left\| \mathbf{y}_i - \mathbf{y}_i^d \right\| \tag{2.45}$$

$$\text{s.t.} \ \|\mathbf{W}\|^2 \leq t, \tag{2.46}$$

  where t is a positive parameter.

  As in the L_1-norm, the objective function in (2.45) is of quadratic form and its contours are ellipsoids [134]. The right side of Fig. 2.40 shows the ellipsoid as well as the area corresponding to the L_2-norm of radius t. Again, the solution is given by the intersection of the contours with the L_2-norm. Observe from the figure that it can typically happen, where none of the coordinates are zero. In other words it returns non-zero values.

However, in terms of complexity the L_1-norm is more efficient, since it returns a sparse solution. In other words, some of the weights of the neural network may be driven to zeros, which reduces the number of computations required for evaluating the output.

In the foregoing paragraph concerning the objective function minimization, we have mentioned that the gradient descent algorithm may get trapped in local optima. In other words, since the minimization of (2.35) relies on an iterative solution, the convergence of the neural network's solution is heavily dependent on the weight and bias initializations. Let us assume that there are two hidden layers connected to the same input, which rely on the same activation function. Then having different initial parameters would ensure the best performance of the network, because having the same initial parameters for each unit would update these units in the same way. In other words, some input patterns may be lost in the null space during the forward-propagation, while some gradient patterns may be lost in the null space during the back-propagation. So to ensure that neither the input and nor the gradient patterns fall in the null space, different initial parameter are used, which motivates us to randomly initialize the weight parameters. Furthermore, the value of the weight parameters is chosen in such a way that they are large enough for propagating the information with integrity, but small enough to avoid saturating the activation function, since this may lead to a low or zero gradient. However, the bias parameter may be chosen heuristically, or may be assigned zero.

It may also be promising to invoke machine learning to initialize the parameters, where the parameters may be learned by an unsupervised technique applied to the same inputs. Furthermore, any algorithm which provides superior performance for one class of problems may perform poorly for for another class.

While Fig. 2.38 shows few examples of different neuron connection topologies, Fig. 2.41 shows the classification of different ANNs tailored for wireless communications. More explicitly, the ANNs classified in Fig. 2.41 may be invoked for solving wireless communication problems. The neural networks are typically classified into three categories [135]: feedforward neural network, deep neural network and recurrent neural networks; albeit there other relatively uncommon neural networks such as spiking neural network [135] , physical neural network and generative adversarial neural network. While Fig. 2.41 shows neural networks that are frequently used in communication systems, it is important to emphasize that there is no universal machine learning algorithm that fits every task [136]. We have to jointly optimize our machine learning (neural network) algorithm with the objective function.

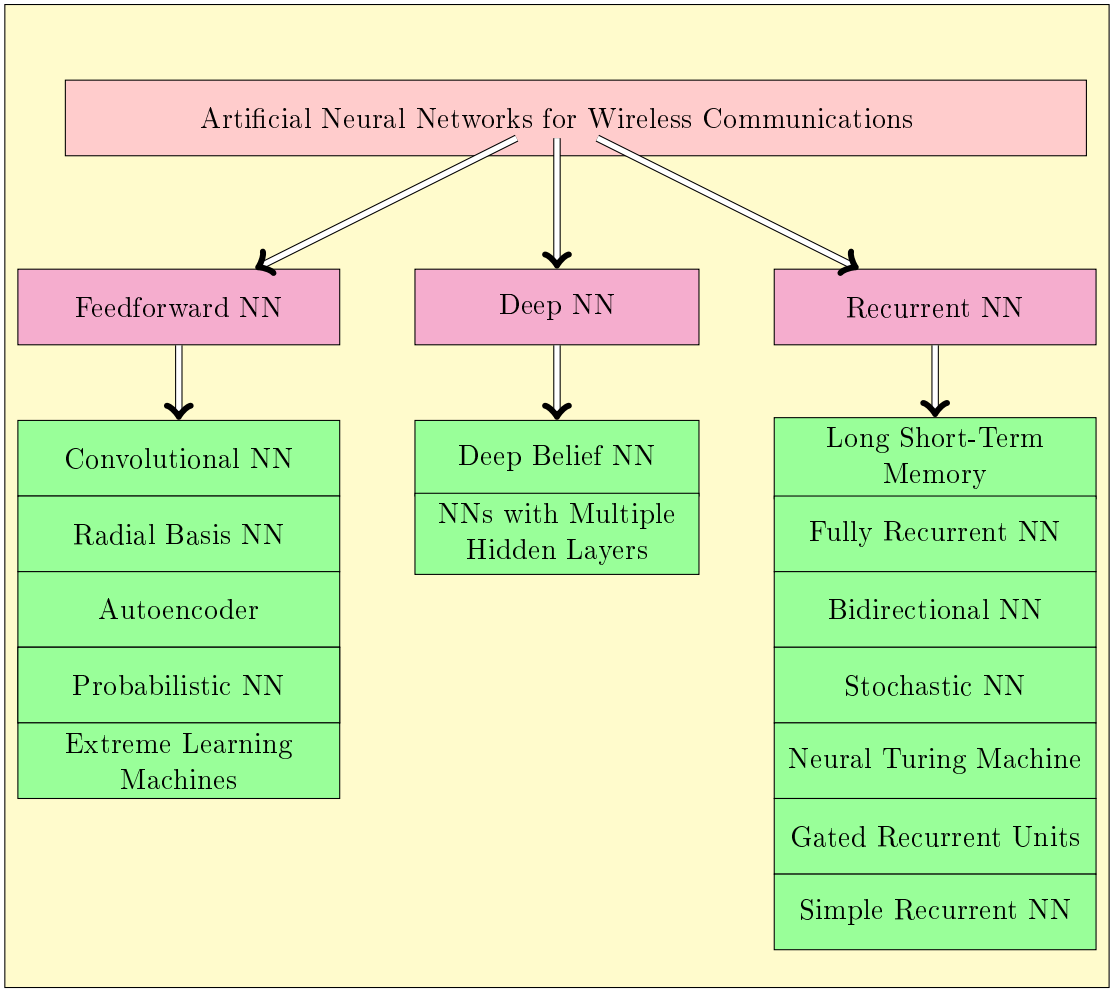


Figure 2.41: Classification of ANNs [127].

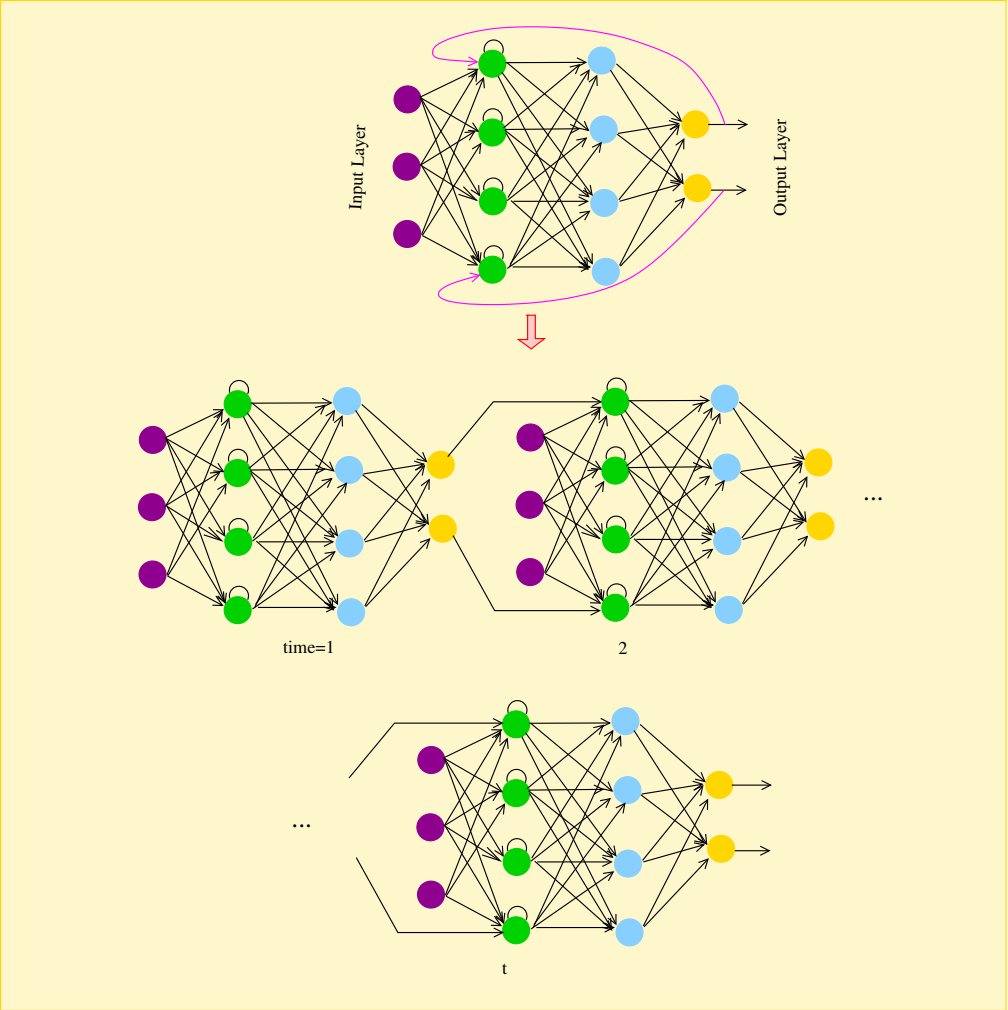


Figure 2.42: Recurrent Neural Network

2.7.3 Recurrent Neural Networks

Having studied the various parameters that affect the performance of neural networks, we now portray the basic philosophy of an other class of neural networks: the recurrent neural network. In the feedforward neural network of Subsection 2.7.2 the inputs and output are only interlinked in a feedforward manner. By contrast, in the recurrent neural networks, there is a more complex relationship between the input and the output. More explicitly, as shown in Fig. 2.42, contrast to the feedforward neural network, in recurrent networks a neuron from one layer may be connected to neurons in previous layers [112]. For example, in Fig. 2.42, the output neuron is connected to the neuron in the first hidden layer. This is particularly useful for sequential information as well as to exploiting the temporal behavior of the output. In other words, the recurrent neural network is an ANN that has memory which allows us to store historical information. The recurrent neural networks are helpful in scenarios such as user mobility prediction, where the objective is to find the future location of a user based on his/her previous patterns/locations recorded

in the past. In contrast to the feedforward networks, the recurrent networks are capable of handling time-related tasks, since the activation function value depends on the data stored in memory.

Although there are subtle differences between the traditional ANN and the recurrent neural network, the associated parameter optimization follows the principles conceived for feedforward neural networks. However, since there is a connection between the input and the output as seen in Fig. 2.42, the conventional back-propagation algorithm cannot be used. This motivates the NN community to conceive a modified backpropagation method which is capable of unfolding the recurrent neural network in time by stacking identical copies of the network. This method is commonly referred to as back-propagation through time, which is detailed in [137]. Despite its powerful capabilities, the back-propagation through time may get trapped in numerous sub-optimal local minimum than that of the gradient descent algorithm. This is because of the loop connection in the network. As an alternative solution, Williams *et al.* [138] proposed a real-time recurrent learning algorithm for accurately evaluating the error gradient at every time step. Fig. 2.41 shows different types of recursive neural networks [127], namely the long short-term memory, fully recurrent network, bidirectional network, stochastic network, neural turing machine, gated recurrent units, and simple recurrent neural network. Depending on the task at hand, an appropriate recurrent neural network may be chosen for gleaming accurate predictions. For example, the long short-term memory neural network may be invoked for time-series prediction problems, while stochastic neural network is employed to introduce random variations into the network so as to escape from the local minima.

2.7.4 Deep Neural Networks

All ANNs having more than one hidden layer are considered deep neural networks. The state-of-the-art deep neural networks include convolutional networks [127], deep recurrent neural networks [112], deep feedforward networks [112], deep belief networks [139], long-short term memory network [140], and deep Q-learning [141]. Although the history of deep ANNs dates back to as early as 1940s, it gained prominence recently because of the reasons summarized below [112]

- *Increased computational power*: The recent advances in the signal processing capabilities, particularly in the graphical processing units (GPUs) have expedited the execution of machine learning algorithms. To elaborate further, the advances in GPUs have facilitated faster and more parallelized computation.
- *Increased data*: In today's information age, data have become the new oil, which however cannot really be used if remains unrefined.

- *Improved algorithms*: Over the decades sophisticated training algorithms have been developed. The computational efficiency of the neural networks typically depends on the activation function [129, 130]. Although there is still substantial on-going research on the specific type of the activation function to be used in neural networks, it was found that employing ReLu speeds up the training process because of the reasons discussed in the context of Fig. 2.36.

A compelling advantage of using deep neural networks is that they can approximate any arbitrary function, which means that for every possible input x the neural network has the output $f(x)$. This is referred to as function approximation.

Universal Approximation Theorem [142]: It states that a feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of R^n .
Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a non constant, bounded, and continuous function (called the activation function). Let I_m, denote the m-dimensional unit hypercube $[0, 1]^m$ and the space of real-valued continuous functions on I_m be denoted by $C(I_m)$. Then, given any $\varepsilon > 0$ and any function $f \in C(I_m)$, there exists an integer N, real-valued constants $v_i, b_i \in \mathbb{R}$ and real vectors $w_i \in \mathbb{R}^m$ for $i = 1, \ldots, N$ so that $F(x) = \sum_{i=1}^{N} v_i \varphi\left(w_i^T x + b_i\right)$. Then, we can have $|F(x) - f(x)| < \epsilon, \forall x \in \mathbf{I}_m$.

2.7.5 Machine Learning in Wireless Communications: State-of-the-art Designs

As mentioned at the beginning of Section 2.7 that beyond its powerful prediction and analytical capabilities, machine learning can be a major driver of intelligent data driven wireless communications and networks. More recently, ANN-aided wireless transmission has gained substantial attention owing to its accurate predictions, which is often superior to conventional methods dispensing with learning [135]. With the aid of machine learning, a low complexity transceiver design can be conceived [127]. More particularly, it can be employed for perform intelligently allocating physical resources, scheduling users, managing traffic congestion, link-adaptation and SI cancellation.

In the literature of wireless communication, machine learning assisted algorithms have been studied in the context of adaptive modulation and coding (AMC) [143]. To increase the accuracy of link-adaptation, Daniels *et al.* [30] conceived a framework for overcoming the limitations of AMC aided MIMO-OFDM relying on supervised learning algorithms [31], such as the K-Nearest Neighborhood (KNN). In this design, without relying on the lookup table, link-adaptation is carried out by KNN algorithm which selects the specific AMC scheme for attaining both a high-rate and high-reliability. The KNN algorithm will be discussed later in this treatise. In addition to the KNN, Daniels *et*

al. [32] also conceived an online AMC, where support vector machines were employed. In [33], link adaptation has been proposed for single-carrier frequency domain equalization, again using the KNN algorithm. More recently, a broader class of machine learning algorithms, namely deep learning methods have been applied in both the context of indoor localization as well as in detection [35]. In [35], Samuel *et al.* proposed the deep neural network assisted detection of MIMO signals. However, this proposed scheme is not flexible enough for accommodating diverse modulated constellations.

Some other state-of-the-art designs focused on learning for localization [24–26]. More particularly, learning based approaches used in localization have been shown to be more efficient in terms of minimizing the location error. More explicitly, Wang *et al.* [34] proposed a fingerprint based localization algorithm relying on the classic feedforward neural networks, where the CSI acts a fingerprints. To elaborate, the aim of the neural network is to predict the CSI of the device, based on which the location is obtained. An experimental study of indoor localization conducted by Chen *et al.* [29] also attributed the superior performance to attained learning, where the authors invoke a convolutional neural network assisted learning scheme. An autoencoder based deep learning is proposed by Khatab *et al.* [26] for indoor localization. In this method, the received signal strength (RSS) is used for training the neural network, where the RSS database is regularly updated to take into account the changes in the environment for updating the weights designed. To a further development, Capone *et al.* [28] employed a rudimentary learning technique for obstacle avoidance aided cell discovery for beam-alignment.

Furthermore, machine learning has also been applied recently in the context of channel decoding and detection. Dorner *et al.* [144] demonstrated the feasibility of point-to-point communications using unsynchronized off-the-shelf software-defined radios relying on NNs. In this work, the authors aimed for designing an end-to-end MIMO system dispensing with the conventional model-based of transmission. More explicitly, Dornel *et al.* invoked learning for an end-to-end communications system by interpreting it as an autoencoder. Furthermore, Jin [145] *et al.* proposed sub-optimal deep learning based detection in the context of MIMO relay channels. Nachmani [146] *et al.* demonstrated that NN-aided belief propagation used for channel decoding improves the performance at a reduced complexity, where the authors invoked recurrent neural networks, for shrinking the search space of the conventional belief propagation technique. Liang *et al.* [147] analyzed the performance of the NN based iterative belief propagation. More explicitly, they [147] concatenated a convolutional neural network to the output of the belief propagation network, where the convolutional network is used for mitigating the estimation errors of the belief propagation. A theoretical model was conceived for feature extraction based on autoencoders by Yan [148] *et al.* In this work, the autoencoder network is amalgamated with an extreme learning machine for signal detection in MIMO-OFDM systems. In [149], Gui *et al.* proposed a deep learning network for non-orthogonal multiple access, while in [150] Huang *et al.* proposed channel estimation and DOA estimation

Table 2.5: Summary of NN applications in wireless communications.

Papers	Machine Learning Method	Contribution
[30]	KNN	Adaptive modulation and coding
[32]	SVM	Adaptive modulation and coding
[33]	KNN	Link adaptation
[24–26]	Classic feedforward NN	Localization
[29]	Convolutional NN	Indoor localization
[144]	Autoencoder NN	End-to-end communication
[145]	Classic feedforward NN	Detection in relay systems
[146]	Recurrent NN	Decoding of linear codes
[147]	Convolutional NN	Channel decoding
[148]	Autoencoders	Symbol detection
[151,152]	Classic feedforward NN	SI cancellation in FD
[153]	Classic feedforward NN	Channel estimation
[149,150]	Long short-term NN	Data detection
[154]	Assorted NNs	Data sequence detection

for massive MIMO systems. However, in the latter case, the channel is assumed to be time-invariant during the DOA estimation. A technique termed as sliding bidirectional recurrent neural network is proposed for data sequence detection in both classical and molecular communication systems.

In full duplex communications, machine learning may also be invoked for modeling the nonlinearities of the power amplifiers. Furthermore, since the transmitter and receiver have different local oscillators, there may be a mismatch of gain and phase between the transmitted and received symbols, which is termed as IQ imbalance. The IQ imbalance tends to be more pronounced at mmWave frequencies. Thus both linear distortions and IQ imbalance may significantly affect the success of SI cancellation. In this scenario, machine learning comes to rescue our aid for modeling these impairments and cancelling them in the digital domain. Very recently Alexios [151] applied neural networks to model the p^{th} order harmonics, which arise due to the nonlinear components in the power amplifiers. More explicitly, he demonstrated experimentally using a hardware testbed that a simple neural network based non-linear canceler is capable of matching the performance of the conventional non-linear cancellation at a significantly lower computational complexity. As a further advance, Hanqing *et al.* [152] demonstrated the feasibility of real-time non-linear SI cancellation using a software defined radio relying on deep neural networks, demonstrating that a SI reduction of 17 dB can be achieved in the digital domain. The above-mentioned contributions detailing the benefits of machine learning techniques in wireless communications are summarized at a glance in Table 2.5.

Chapter 3

Dual-Function Hybrid Beamforming Transceiver Design

Everyone is playing roles. Mother, friend, caretaker, friend, boss, etcetera. Every role comes with values and intentions. Identify all the roles you are playing. What are the values and intentions of those roles

—Akiroq Brost

3.1 Introduction

RECENT studies show a further escalation of the capacity demand [155], especially in the microwave band [156]. Hence next generation systems aim for additionally harnessing mmWave frequencies [63], as a benefit of the abundant availability of bandwidth. However, mmWave frequencies suffer from a high pathloss because of the attenuation due to water vapour, oxygen absorption and the foliage. As a result, the signal-to-noise (SNR) ratio would be typically low [63] at the receiver. To compensate the propagation losses, directional BF is considered to be promising [2]. Furthermore, since the $\lambda/2$ spacing between the antennas would be on the order of millimeters at mmWave frequencies, large number of antennas can be accommodated in a compact space for achieving high BF gains. However, employing DBF having large number of antennas, where both the amplitude and phase are controlled digitally using ADCs/DACs, would result in a high cost and hardware complexity imposed by the RF chain dedicated to every AE and owing to the high power consumption of the ADCs [62]. Although employing ABF, where the AA is connected to an RF chain through amplifiers and controllable analog phase shifters is capable of reducing the complexity imposed, the angular resolution of ABF is typically inaccurate [63]. Hence hybrid beamforming combining analog and DBF for mmWave systems was conceived in [157], where the ensuing digital processing is capable of correcting

the analog imperfections [158]. In hybrid solutions, ABF is carried out at the RF front end, where the digitally precoded signals are phase-shifted and then fed to the transmit antennas. Similarly, at the receiver RF front end the received signals are phase-shifted and then fed to the digital combiner operating in the baseband. Two different hybrid beamforming designs have been reported in [159], namely a so-called full-array-connected design and, sub-array-connected design [159]. In the full-array-connected design the digital precoded signals are phase-shifted and then fed to all the transmit antennas. By contrast, in the sub-array-connected design, after phase shifting the digitally precoded signals are fed only to a subset of the antennas. Hence, the sub-array-connected implementation is more energy efficient than the full-array-connected design [159].

In FDD systems, the receiver has to feed back the channel estimation to the transmitter for the TPC to compensate the channel-effects about to be encountered. However, feeding back the channel coefficients to the transmitter would result in an overhead [160]. An appealing practical solution for feeding back the channel information to the transmitter is based on a carefully designed TPC weight codebook, which is known both to the transmitter and the receiver. Explicitly, based upon the CIR estimate the receiver selects the best TPC matrix from its codebook, and relays its index to the transmitter over the feedback channel. This technique is referred to as, limited-feedbacks design.

Song *et al.* [88] proposed a codebook design for mmWave systems based upon a technique referred to as orthogonal matching pursuit. As a further solution, Rajashekar and Hanzo [94] discussed a codebook design using vector quantization conceived for finite modulation alphabets. In [96], Alkhateeb and Heath proposed RF codebook designs for limited feedback mmWave systems. By exploiting the channels' reciprocity, a heuristic codebook design was advocated by Eltayeb *et al.* [90]. Naturally, the performance of a practical codebook design depends on the codebook entries as detailed by Clerckx *et al.* [97].

On the other hand, the sub-array-connected design, constitutes to a large extent an open problem owing to the mathematical intractabilities and other challenges [159]. Early work on codebook design for sub-array-connected solution was reported in [161], where the analog beamformer weights are obtained using beam-search algorithms, which however impose high complexity, while [162] discussed a precoding design inspired by the successive interference cancellation.

Against this background, the novelty and rationale of this chapter are as follows:

1. *A new hybrid BF architecture, where an AA is partitioned into two or more sub-arrays is conceived. In this architecture, the sub-arrays emerging from a full array are separated by a sufficiently large distance that minimizes the correlation between sub-arrays, say for example* $\geq 5\lambda$*, where* λ *is the wavelength, so that the channel of the sub-arrays becomes uncorrelated from one another. The rationale behind the sub-array separation is to attain diversity gains in addition to the BF gains derived from the sub-arrays.*

2. *We analyze the capacity of the proposed architecture and conclude that using two sub-arrays strikes an attractive compromise, since the diversity gain obtained from the two sub-arrays out-weighs the loss in the BF gain imposed by splitting the array. By contrast, dividing the array into more than two sub-arrays would only result in a reduction of loss of BF gain which the additional diversity gain cannot compensate.*

3. *We analyze a codebook based on mutually unbiased bases (MUBs) amalgamated with DFT-based analog BF in the RF stage for the aforementioned mmWave MIMO systems. Early work on MUBs was reported in [163,164] for sub-6 GHz systems, where the benefit of employing MUBs is their lower complexity than that of Grasmannian scheme of [165], while performing close to a system relying on the simplifying assumption of having perfect CSI.*

4. *Then, we investigate the performance of hybrid beamforming system in the face of mutual coupling relying on a practical codebook using DFT and MUB precoders of [98].*

5. *Furthermore, we propose an adaptive array design for hybrid beamforming in mmWave communication that adapts the architecture as well as the digital precoding depending on whether the channel is LOS or NLOS. Explicitly, we design an architecture with an adaptive array, where the array is full-array-connected in LOS channel with antenna spacing $\lambda/2$, and sub-array-connected, where the AA is partitioned into sub-arrays with separation sufficiently large enough so that each sub-array experiences independent fading, in NLOS channel conditions.*

6. *Finally, we utilize the novel adaptive array design to propose an adaptive virtual cell formation in the PHY layer of Cloud-Radio Access Network (C-RAN) for mmWave hybrid beamforming aided systems, since the dynamic selection of RRHs is similar to a mmWave transmitter that adaptively switch between the fully-connected and ASA designs. In virtual cell formation, the user may maintain its connection with one or two RRH(s) depending on our proposed design criterion. We first present an algorithm for virtual cell formation for a single mobile user, where the RRHs connected to the user experience different channel conditions (LOS/NLOS). Then we extend our algorithm to a multi-user scenario, where the mobile users are distributed randomly and experience different channel conditions (LOS/NLOS) with the RRHs. Moreover, given the fronthaul constraints in C-RAN, for multi-user setting, we consider the maximum number of users that can be supported by the fronthaul link.*

The remainder of this chapter is organized as shown in the Fig. 3.1. In Sec. 3.2 we present our hybrid architecture conceived, followed by the capacity analysis of the proposed design. Subsequently, in Sec. 3.3, we discuss our hybrid precoder design, where we select the specific TPC matrix from the codebook designed for analog RF and digital baseband.

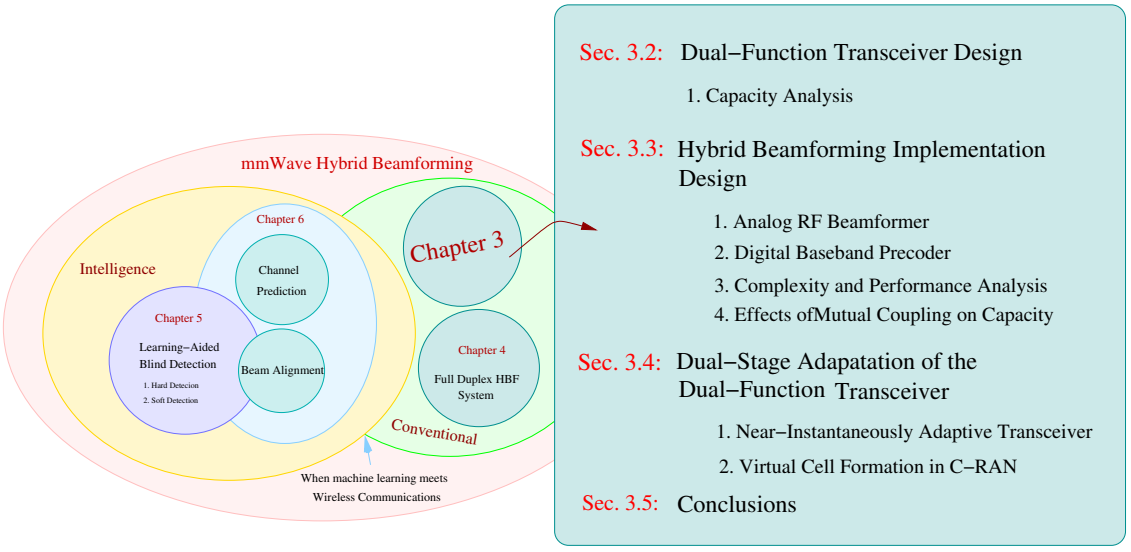


Figure 3.1: The organization of this chapter.

Then in Sec. 3.4, we propose a dual stage adaptation based on the characteristics of the mmWave channel. Finally, we provide the chapter conclusions in Sec. 3.5.

3.2 Dual-Function Transceiver Design

In Chapter 2, we have considered a suite of state-of-the-art hybrid precoding designs, with an emphasis on the fully-connected and on the sub-array-connected architecture. We discussed that the fully-connected design requires more phase shifters than the sub-array-connected design. Recall that in the sub-array-connected design discussed in Sec. 2.23, the AA is partitioned into sub-arrays, where each sub-array is positioned in the immediate vicinity of the subsequent array. However, the splitting of the array into sub arrays leads to undesired beamforming gain reduction. Furthermore, the tight close-placement of the sub-arrays result in the loss of diversity gain, since they do not experience independent fading. This motivates us to conceive a dual-function transceiver for mmWave systems.

In contrast to the state-of-the art, let us now consider an ASA design, where the AA is partitioned into an array of sub-arrays and *each sub-array is sufficiently separated* by a distance from the other sub-array, with one or more RF chain(s) dedicated to each sub-array, as shown in the Fig. 3.2. In contrast to conventional ASA, in our proposed architecture, the angular spread of the departure angles from one sub-array to another would be wide enough to experience independent fading, since the sub-arrays are placed at a sufficiently large distance. Hence, because of the wider angular spread in our proposed design the channel of one sub-array is *uncorrelated* with that of the other sub-arrays, while the channel within a given sub-array is correlated. Thus, this design is capable of providing a diversity gain as determined by the number of sub-arrays that are spatially separated, in addition to the BF gain provided by each sub-array.

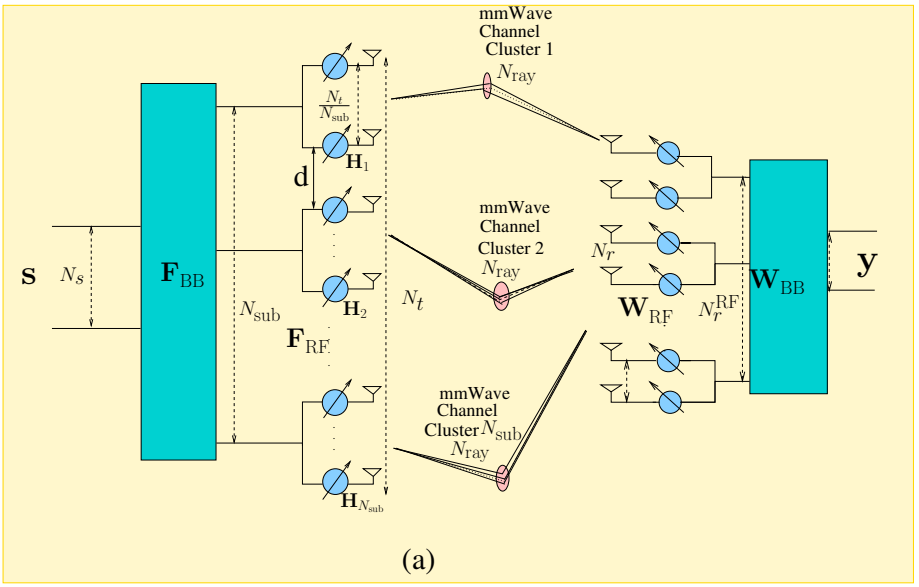


Figure 3.2: (a) Proposed arrays of sub-array-connected architecture, where the separation d between two adjacent sub-arrays is sufficiently large enough so that each sub-array experiences independent fading (b) 2-sub-array-connected with multiple RF chains in each sub-array.

Let us now consider the single-user mmWave MIMO system shown in Fig. 3.2, where the transmitter is equipped with N_t antennas and the receiver with N_r antennas. The received signal vector after RF and baseband combining at the receiver is given by

$$\mathbf{y} = \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{BB}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{n}, \tag{3.1}$$

where

$$\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \dots \mathbf{H}_{N_{\text{sub}}}]$$

is the channel matrix of size $N_r \times N_t$ so that

$$\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r,$$

and each sub-channel matrix $\mathbf{H}_i$ of size $N_r \times N_t/N_{\text{sub}}$ is modeled as in (2.20). Furthermore the transmitter is equipped with N_{sub} RF chains, where $\mathbf{F}_{\text{RF}}$ is the RF beamformer matrix of size $N_t \times N_{\text{sub}}$, which is expressed as

$$\mathbf{F}_{\text{RF}} = \text{diag}(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{N_{\text{sub}}}),$$

where $\mathbf{f}_i$ is a vector of size N_t with only N_{sub} non-zero entries, while $\mathbf{s}$ and $\mathbf{y}$ are the transmitted and received vectors of size N_s, respectively.

In the proposed ASA architecture of Fig. 3.2, a cluster of rays having an angle of departure ϕ^i impinges on each sub-array. Hence, the channel seen by the sub-array is spatially correlated across the antennas of the sub-array. Furthermore, since every sub-array is separated from another sub-array by a sufficiently large distance, the channel of a specific sub-array is uncorrelated with that of another sub-array, hence the channel matrix becomes less correlated than that of the fully-connected architecture. For a given AA, splitting it into two or more widely separated sub-arrays would result in an uncorrelated channel. As a result, the degrees of freedom in the spatial domain increases, hence increasing the diversity order attained. However, the additional diversity comes at the cost of degrading the BF gain of the AA. Hence, there is a trade-off between the diversity and the BF gain attained. In the next section we discuss the achievable rate of the proposed architecture.

Note that the receiver design can have the fully-connected or sub-array-connected architectures as shown in Fig. 3.2 (a) and Fig. 3.2 (b), depending on the specific requirements, such as space constraints, power consumption, BF and diversity gains required, while each sub-array can have multiple RF chains associated with a fully-connected design at the transmitter, as shown in Fig. 3.2 (b) for the 2-sub-array-connected architecture. Fig. 3.2 (a) shows a single RF chain in each sub-array as in state-of-the-art ASA, while Fig. 3.2 (b) is more generic, where each sub-array has multiple RF chains. More explicitly, Fig. 3.2 (b) shows the proposed sub-array architecture, where each sub-array may be fully-connected.

3.2.1 Capacity Analysis of the Proposed Architecture

In this section we analyze the system model of (2.19) in terms of its Shannon capacity and outage capacity.

3.2.1.1 Shannon Capacity

With the aid of the system model given in (2.19), the achievable rate (Shannon capacity) before combining is formulated as [166]:

$$C = \log_2 \det \left(\mathbf{I}_{N_s} + \frac{P}{N_s N_\circ} \mathbf{F}_{\mathrm{BB}}^H \mathbf{F}_{\mathrm{RF}}^H \mathbf{H}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \right), \tag{3.2}$$

where C is the number of bits transmitted per second per channel use (bps/cu) and P is the total transmission power, det(.) is the determinant, $\mathbf{I}_{Ns}$ is the identity matrix of size $N_s \times N_s$ and $N_\circ$ is the noise power.

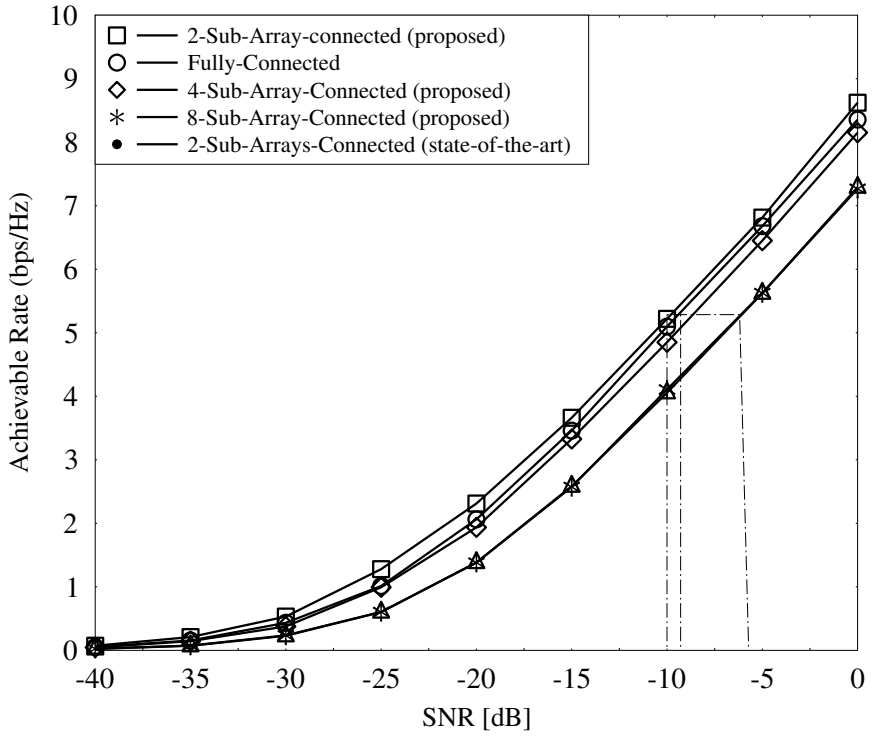


Figure 3.3: Achievable rate of the fully-connected design, and of the ASA designs for a 64x16 MIMO with $N_s = 1$, $N_t^{\mathrm{RF}} = 1$ in each sub-array, while $N_t^{\mathrm{RF}} = 2$ in fully-connected.

In Fig. 3.3 we provide a comparison of the achievable rate of the fully-connected design, of the state-of-the-art ASA design using 2 sub-arrays, and of the proposed 2, 4, 8 sub-array-connected architecture of Fig. 3.2. As shown in Fig. 3.3, the state-of-the-art ASA design's performance is inferior to the fully-connected design by about 5dB. On the other hand, the proposed 2-sub-array-connected design outperforms the fully-connected design with a gain of around 1dB, because in the case of our 2-sub-array-connected design, both

the diversity and the BF gains are obtained. This is because in our proposed 2-sub-array-connected architecture the BF gain reduction imposed by splitting the full AA into 2 sub-arrays is lower than the diversity gain obtained. However, it is worth observing from the figure that the extra diversity gain offers only diminishing returns upon further increasing the number of sub-arrays, which does not compensate for the BF gain reduction due to splitting the array into more sub-arrays. Hence, *the most appropriate number of sub-arrays that yields both high BF and diversity gains is 2.* This design can be referred to as a *dual-functional* MIMO [167] equipped both with BF and diversity gains.The RF beamformer $\mathbf{F}_{\mathrm{RF}}$ and the baseband TPC $\mathbf{F}_{\mathrm{BB}}$ can be obtained from [67, 88] or from the DFT and MUB, which will be described later in Sec 3.3. However, owing to the assumptions of an infinite block length with a probability of error approaching zero, the Shannon capacity does not provide a generic performance measure for practical systems, where finite block lengths associated with finite error probability are encountered. Thus, to analyze the performance of the proposed system we opt for the to outage capacity metric.

3.2.1.2 Outage Capacity

Owing to the assumptions of infinite block length with the probability of error approaching zero, the classic Shannon capacity does not provide a good performance measure for practical systems, where finite block lengths with finite error probability are designed. Thus, to analyze the practical performance of the proposed system we continue with its outage capacity analysis. The system is said to be in outage, when the rate falls below a minimum, which is referred to as the outage capacity [168]. To better understand the dual-functionality of the transmitter in the system conceived, we analyze the outage capacity when the receiver is equipped with a single antenna and the transmitter is equipped with N_t antennas and N_{sub} sub-arrays. This analysis can be extended to N_r receive antennas. An outage event occurs whenever the Shannon capacity calculated in (3.12) falls below a threshold rate C_{out}. Thus, the outage capacity associated with the outage probability p_{out} is given by

$$p_{\mathrm{out}} = \mathbb{P}\left(\log_2\left(1 + \|\mathbf{h}\|^2 \mathrm{SNR}\right) < C_{\mathrm{out}}\right), \tag{3.3}$$

where $\mathbb{P}$ represents the probability, $\mathbf{h}$ is the Rayleigh fading distribution and $\|\mathbf{h}\|^2$ is the sum of squares of random variables, which is Chi-square distributed with $2N_{\mathrm{sub}}$ degrees of freedom. Then (3.3) can be expressed as [168],

$$p_{\mathrm{out}} = \mathbb{P}\left(\|\mathbf{h}\|^2 < \frac{2^{C_{\mathrm{out}}} - 1}{\mathrm{SNR}}\right). \tag{3.4}$$

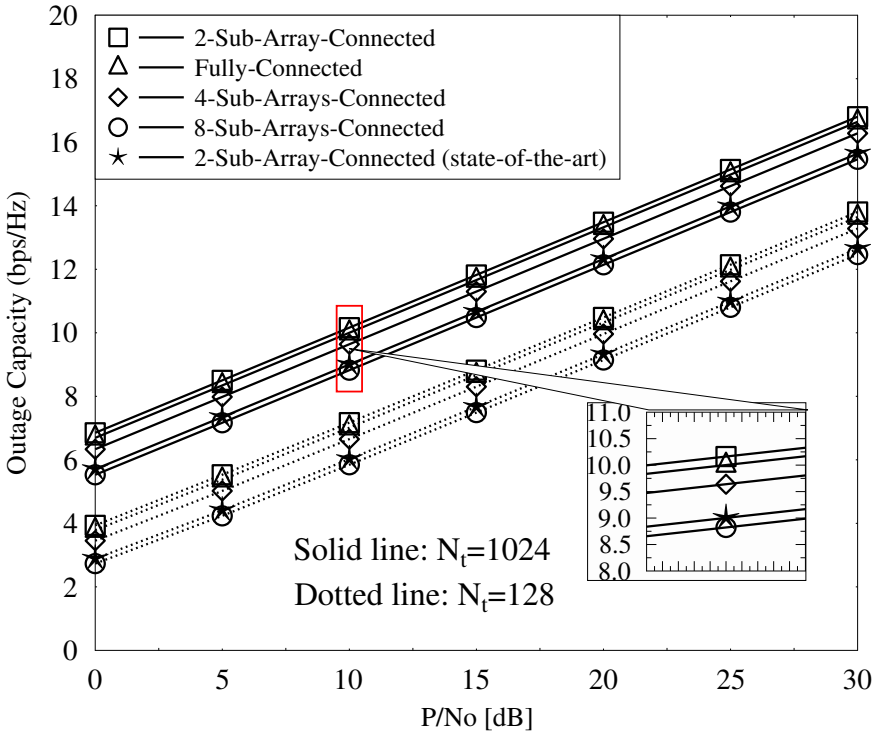


Figure 3.4: $p_{\text{out}} = 1\%$ Outage capacity for fully-connected array and sub-array-connected for $N_t = 128, 1024$ antennas.

The p.d.f. of $\|\mathbf{h}\|^2$ is $\frac{1}{(N_{\text{sub}}-1)!}h^{N_{\text{sub}}-1}e^{-h}, h \geq 0$. At high SNRs, we obtain [64],

$$p_{out} = \frac{(2^{C_{\text{out}}}-1)^{N_{\mathbf{sub}}}}{N_{\text{sub}}!\text{SNR}^{N_{\text{sub}}}}, \tag{3.5}$$

and $\text{SNR} = \frac{N_t P}{N_{\text{sub}}^2 N_{\text{o}}}$, where $\frac{N_t}{N_{\text{sub}}}$ accounts for beamforming, while $\frac{1}{N_{\text{sub}}}$ ensures equal power allocation for all sub-branches.

Fig. 3.4 shows the outage capacity for a fixed outage probability of 1%. Observe that splitting the array into two sub-arrays provides a better outage capacity than the fully-connected array. Regardless of the number of AEs, the performance starts degrading upon further partitioning of the array, as seen in Fig. 3.4.

3.2.1.3 Optimum Number of RF Chains:

It is instructive to note that, Sohrabi and Yu [77] derived a theoretical bound for the optimum number of RF chains required for the fully-connected architecture to achieve the maximum attainable rates of unconstrained digital BF. It was shown in [77] that to

achieve the rates of a digital BF design, the number of RF chains in the fully-connected design should be at least twice the number of spatial streams. This holds for our proposed design as well. Explicitly, in our proposed design, each sub-array can have multiple RF chains for supporting multiple streams, that is, each sub-array can behave as the fully-connected design as shown in Fig. 3.2(b). This implies that:

$$\mathbf{F}_{\mathrm{RF}} = \mathrm{diag}\left(\mathbf{F}_1,\ \mathbf{F}_2\right), \tag{3.6}$$

while $\mathbf{F}_1$and $\mathbf{F}_2$ are expressed as:

$$\mathbf{F}_1 = \left[\mathbf{f}_1^{(1,1)}, \mathbf{f}_1^{(1,2)}, \ldots, \mathbf{f}_1^{(1,N_{\mathrm{RF}}/2)}\right], \tag{3.7}$$

$$\mathbf{F}_2 = \left[\mathbf{f}_2^{(2,1)}, \mathbf{f}_2^{(2,2)}, \ldots, \mathbf{f}_2^{(2,N_{\mathrm{RF}}/2)}\right], \tag{3.8}$$

where $\mathbf{f}_m^{(m,n)}$ represents the vector in the n^{th} column of the m^{th} sub-array. In this paper, for brevity, we consider $N_{\mathrm{RF}} = 2$, which leads to the architecture of Fig. 3.2 associated with two sub-arrays. In contrast to the fully-connected architecture of Fig. 2.23, this design requires less phase shifters, which makes it more energy efficient. For example, for N_t transmit antennas and N_{RF} RF chains, the fully-connected architecture would need $N_t N_{\mathrm{RF}}$ phase shifters, while the proposed ASA design would need $N_t N_{\mathrm{RF}}/2$.

Furthermore, since the transmit antenna diversity offers diminishing returns upon increasing in the number of transmit antennas, it is essential to understand the transmit antenna diversity gains obtained and the loss incurred in the BF gains, when partitioning a given AA into more sub-arrays. Hence, to have a better understanding of the optimum number of sub-arrays, we embark upon investigating it by analyzing the symbol error probability in terms of the diversity order and the BF gain.

Let us consider a MISO system having L sub-arrays and N/L AAs for each sub-array, where N is the total number of AEs. If $\gamma_i = \frac{P}{L\sigma^2}|\alpha_i|^2$ is the SNR of the branch between the i^{th} sub-array of the transmitter and the receiver, where α_i is a complex-valued Gaussian random variable and N/L is the gain derived from BF using the sub-array i as shown in the Fig. 3.5, then the total SNR in the branch is given by $\frac{N}{L}\gamma_i$. For QPSK modulation, the symbol error probability of the MISO system conditioned on α is given by [64]

$$P_e = 2Q\left(\sqrt{\frac{2N}{L}\Gamma}\right), \tag{3.9}$$

where $\Gamma = (\gamma_1 + \gamma_2 + \ldots + \gamma_L)$ is the total SNR of the system.

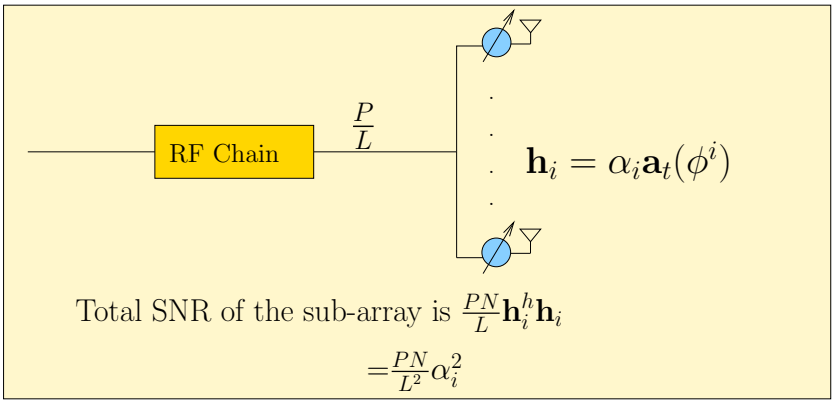


Figure 3.5: The SNR of a sub-array

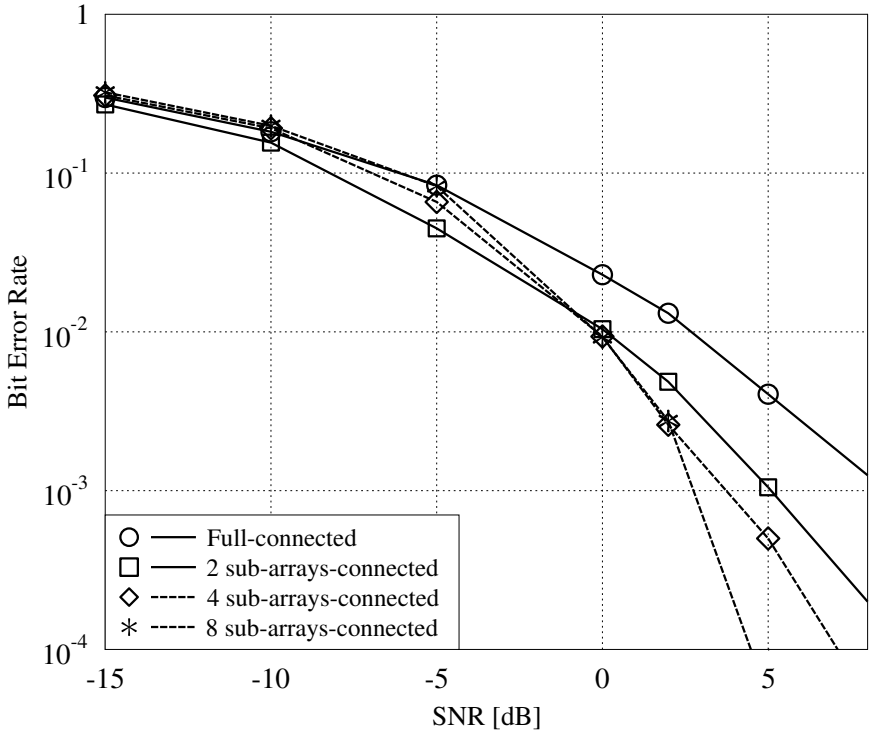


Figure 3.6: Probability of bit error rate of fully-connected design and of different number of sub-arrays-connected design for a (16×1) MIMO

By substituting $\Gamma = \gamma_1 + \gamma_2 + \ldots + \gamma_L$, we have

$$P_e = 2Q\left(\sqrt{\frac{2NP}{L^2\sigma^2}\|\alpha\|^2}\right), \tag{3.10}$$

where $\|\alpha\|^2$ is a chi-squared distribution with $2L$ degrees of freedom. Upon integrating over the distribution, we obtain

$$P_e \approx \binom{2L-1}{L} \frac{2}{\left(\frac{4NP}{L^2\sigma^2}\right)^L}. \tag{3.11}$$

At high SNRs, we have,

$$P_e \approx \frac{1}{L!} \frac{2}{\left(\frac{NP}{L^2\sigma^2}\right)^L}, \tag{3.12}$$

i.e.,

$$P_e \propto \frac{1}{L!} \left(\frac{L^2\sigma^2}{NP}\right)^L. \tag{3.13}$$

Having arrived at the symbol error probability (P_e), which is related to the bit error probability (P_e^b), it is evident from (3.13) that for a given P and σ^2, P_e is inversely proportional to the BF gain (N/L) and, proportional to the number of sub-arrays by a factor of $\frac{L^L}{L!}$.

To better understand the diversity and the BF gains achieved in splitting of an antenna-array into sub-arrays, we considered a (16×1) MISO system. Fig. 3.6 plots the bit error probability (P_e^b) of the fully-connected design and of the 2, 4, 8 sub-array-connected design. As expected, the curves in Fig. 3.6 tally with the P_e derived in (3.13) for different configurations. It is seen from Fig. 3.6 that at high SNRs, the 8 sub-array-connected design achieves better error rate than that of the fully-connected design, and the 2, 4 sub-array-connected designs. This is because of the diversity order, which is seen in the denominator of (3.13). It is conducive to say that the diversity offers better gains than the BF in the case of the bit-error rate. However, this behavior is not reflected in the capacity plot seen in Fig. 3.3. This is because the diversity gain erosion is only partially compensated by the beamforming gain when the array is partitioned into more than two sub-arrays.

3.3 Hybrid Beamforming Implementation Design

In this section, we expound on the signal processing aspects of mmWave systems. More explicitly, we focus our attention on the HBF, where we first present our analog RF beamformer design followed by the baseband precoder.

3.3.1 Analog RF Beamformer

In this section we present the design of the ABF in mmWave systems. In this treatise, we consider the DFT-aided codebook for RF beamforming, since the codewords in the DFT based beamforming weight-vectors match with the statistical distribution of the optimal beamforming weight vectors [169]. Furthermore, it satisfies the constant modulus constraint posed by the phase shifters. Fig. 3.7 shows the achievable rate for different spatial correlations using on DFT-based RF beamforming scheme. It is seen in Fig. 3.7 that with the increase of spatial correlation, the achievable rate of the DFT-aided RF codebook approaches the rate of the digital TPC relying on the idealized simplifying assumption of perfect CSI. Since the mmWave channel exhibits high spatial correlation, the choice of the DFT as our analog RF beamformer enhances the system throughput.

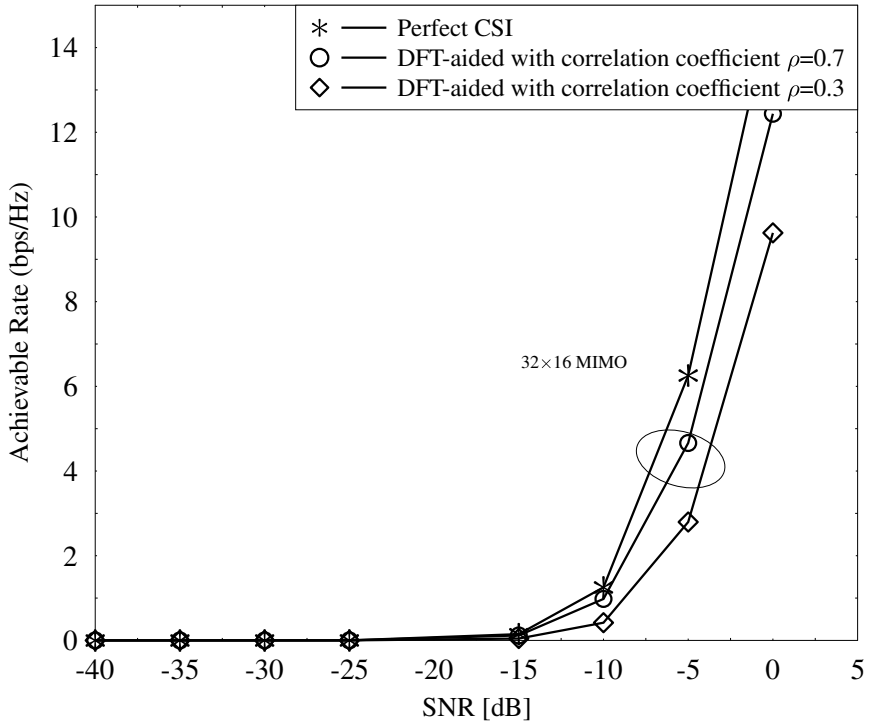


Figure 3.7: Rate performance of a DFT-aided RF beamforming for a 32×16 MIMO in a mmWave channel with correlation coefficients $\rho = 0.3, 0.7$

Motivated by this, we employ a DFT-based codebook for analog RF beamforming using phase shifters, since the DFT matrix satisfies the constraints to be met by the phase shifters, such as having constant modulus entries from a finite alphabet. Furthermore, for the baseband TPC matrix, we invoke a codebook design constructed based upon

MUBs as a benefit of its efficacy and low complexity, which will be detailed later in this section.

For the fully-connected design, the N_t^{RF} columns of the analog RF beamformer matrix $\mathbf{F}_{\mathrm{RF}}$ are selected from the DFT-matrix, by picking those which exhibit maximum correlation with the columns of the right singular matrix of the channel $\mathbf{H}$=$\mathbf{U}\Sigma\mathbf{V}$, i.e. we have

$$\mathbf{DFT}_{N_t} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N_t-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N_t-1)} \\ 1 & \omega^3 & \omega^6 & \dots & \omega^{3(N_t-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N_t-1} & \omega^{2(N_t-1)} & \dots & \omega^{(N_t-1)(N_t-1)} \end{bmatrix},$$

where $\omega = \exp\left(-j\frac{2\pi}{N_t}\right)$ and

$$\mathbf{F}_{\mathrm{RF}}(:,i) = \max_i < \mathbf{DFT}_t(:,\ i),\ \mathbf{V} >, \tag{3.14}$$

where $\mathbf{V}$ is the right singular matrix of the channel. Similarly, the N_r^{RF} columns of $\mathbf{W}_{\mathrm{RF}}$ are selected from the DFT-matrix, by picking those which exhibits maximum correlation with the columns of the left singular matrix of the channel $\mathbf{H}$, i.e., we have

$$\mathbf{W}_{\mathrm{RF}}(:,i) = \max_i < \mathbf{DFT}_r(:,\ i), \mathbf{U} >, \tag{3.15}$$

where $\mathbf{U}$ is the left singular matrix of the channel.

For the sub-array-connected design, the RF weights at the transmitter for each sub-array are selected from the DFT matrix of size $(N_{\mathrm{sub}} \times N_{\mathrm{sub}})$, which exhibits maximum correlation with the dominant right singular vector of the sub-channel. Similarly, the RF combiners weights are calculated at the receiver, depending on whether the fully-connected or the sub-array-connected design are preferred.

Having carried out analog RF beamforming, the channel seen by the baseband TPC after RF beamforming is the effective channel given by:

$$\mathbf{H}_{\mathrm{eff}} = \mathbf{W}_{\mathrm{RF}}^{*}\mathbf{H}\mathbf{F}_{\mathrm{RF}},$$

where the digital baseband precoder is applied to the effective channel matrix $\mathbf{H}_{\mathrm{eff}}$.

In the next section, we characterize the digital baseband precoder by considering a range of practical constraints.

3.3.2 Digital Baseband Precoder

In the 3GPP long term evolution (LTE), the digital baseband precoder is employed by selecting one of the codewords from the codebook entries derived from a finite alphabet, such as QPSK $\{1, -1, i, -i\}$. To feed back the best available precoder matrix from the codebook, the receiver has to evaluate the achievable throughput for every legitimate precoding matrix in the codebook for every rank, and relay the best precoder matrix index to the transmitter over the feedback channel. However, in practice the choice of the codebook design mainly hinges on the performance versus complexity trade-off. As a solution for designing the codebook, some key properties of a practical codebook, such as in LTE, are outlined as follows.

1. **The Codebook Entries:** The choice of a typical codebook is governed by the entries of the codebook [97], since the complexity of a codebook depends on the number of complex multiplications and complex additions required for evaluating the performance metric of a given precoder [170]. Thus the complexity and power consumption are commensurate with the number of complex operations involved in the precoder. Hence it is important to specifically design the codebook entries for mitigating the complexity.

2. **The Nested Property:** Having a nested structure for the codebook reduces the number of calculations required, if the transmitter has the capability for adapting its rank of transmission depending on the channel quality. That is if the transmitter switches from lower rank transmission to higher rank based on channel information knowledge provided by the receiver over the feedback channel, then calculations carried out for lower rank transmission can be used for higher rank transmission, since the lower rank precoder is also a part of the precoder used for a higher rank [97].

3. **Power Amplifier Imbalance:** The power amplifier imbalance across the antennas is highly undesirable at the transmitter [171]. Due to its power imbalance at the transmitter, the average signal-to-noise ratio per physical antenna would not be the same across all antennas [65, 171]. This would result in the loss of diversity gains as one of the antennas which has a lower signal-to-noise ratio will not be able to contribute much to the final signal-to-noise ratio [65]. Hence, power imbalance would cause the system to perform badly in terms of diversity gain as in correlation. So one way to avoid power imbalance is by using the constant modulus or constant gain codebook.

In the next section, we employ the baseband digital precoder selected from the specific codebook derived from the mutually unbiased bases (MUBs). The rationale for choosing MUBs is that they exhibit all the aforementioned properties.

3.3.2.1 Mutually Unbiased Bases for Codebook Design

In this section, we begin by discussing what mutually unbiased bases (MUBs) are and how they are generated. Later we focus our attention on the motivation behind designing a codebook based upon mutually unbiased bases.

The concept of MUBs was first introduced in quantum mechanics in [172]. In wireless communication, MUB based precoding codebook in baseband is proposed in [163, 164].

Let $\mathbf{B}$ and $\mathbf{B}'$ be two orthogonal bases of a N dimensional Hilbert space, then $\mathbf{B}$ and $\mathbf{B}'$ are said to be mutually unbiased if $|< \mathbf{b}, \mathbf{b}' >|^2 = \frac{1}{N}$, $\forall\ \mathbf{b} \in \mathbf{B}$ and $\mathbf{b}' \in \mathbf{B}'$.

However, the number of mutually unbiased bases (MUBs) for a given dimension is limited. The cardinality $\mathcal{C}$ of pairwise MUBs for a given dimension N is $\mathcal{C}(N) \leq N+1$ [173–175]. In particular, if the dimension of interest is a power of prime then the cardinality $\mathcal{C}(N)$ of pairwise MUBs is $N+1$. For dimension N which has at least two prime factors has no exact value of $\mathcal{C}(N)$.

Lemma 3.1. ***(Andreas and Martin)*** *[176], Suppose the factorization of N into distinct primes p_i is $N = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_r^{\alpha_r}$. Then the cardinality*

$$\mathcal{C}(N) \geq \min\{\mathcal{C}(p_1^{\alpha_1}), \mathcal{C}(p_2^{\alpha_2}), \dots, \mathcal{C}(p_r^{\alpha_r})\} \tag{3.16}$$

We refer to [177] for proof.

We shall now discuss the generation of MUBs for power of primes. We divide the primes into two categories, odd prime and even prime, where we generate MUBs separately for each case.

Odd Prime Power

Lemma 3.2. ***(Andreas and Martin)*** *[176], Let $\mathcal{F}_n$ be the finite field with n elements and of odd characteristic p and $\omega_p = \exp(2\pi i/p)$ is the p^{th} root of unity. Denote the standard basis and sets $B_a = \{v_{a,b} \mid a, b \in \mathcal{F}_n\}$ then the set of vectors given by*

$$v_{a,b} = \sqrt{n} \left(\omega_p^{tr(ax^2+bx)} \right)_{x \in \mathcal{F}_n}, \tag{3.17}$$

form an extremal set of $n+1$ MUBs of dimension n.

We refer to [176] for proof.

Example: For $N = 3$, the construction yields MUBs as

$$B_0 = \{v_{0,0}, v_{0,1}, v_{0,2}\} = \{3^{-1/2}(1,1,1),\ 3^{-1/2}(1,\omega_3,\omega_3^2),\ 3^{-1/2}(1,\omega_3^2,\omega_3)\}$$
$$B_1 = \{v_{1,0}, v_{1,1}, v_{1,2}\} = \{3^{-1/2}(1,\omega_3,\omega_3),\ 3^{-1/2}(1,\omega_3^2,1),\ 3^{-1/2}(1,1,\omega_3^2)\}$$
$$B_2 = \{v_{2,0}, v_{2,1}, v_{2,2}\} = \{3^{-1/2}(1,\omega_3^2,\omega_3^2),\ 3^{-1/2}(1,\omega_3,1),\ 3^{-1/2}(1,1,\omega_3)\}$$

Even Prime Power

For even prime powers, as in [163], we construct MUBs by using Hadamard matrix. For a dimension $N = 2^n$, where $(n = 1, 2, \ldots)$, we start by an Hadamard matrix $\mathbf{A}_{2^n}$, which satisfies the property of a MUB. Thus the Hadamard matrix is the easiest existing MUB. Using the Hadamard matrix $\mathbf{A}$, we construct other basis by finding a column vector $\mathbf{v}$ such that all the elements in $\mathbf{A}^H\mathbf{v}$ have magnitude equal to the $(1/2^n)$. And we obtain the basis $\mathbf{B}$ by element wise multiplication[1] of $\mathbf{v}$ with $\mathbf{A}$ i.e., if $\mathbf{A} = [\mathbf{a_1}\ \mathbf{a_2} \ldots \mathbf{a_{2^n}}]$ then $\mathbf{B} = \mathbf{v} \otimes \mathbf{A}_8 = \mathbf{va_1}\ \mathbf{va_2} \ldots \mathbf{va_{2^n}}]$. We continue this procedure until all the $2^n + 1$ basis are obtained.

Example: For $N = 2^3$ i.e., $n = 3$, the construction of MUBs is illustrated below:

Step 1: First obtain a Hadamard matrix $\mathbf{A}_8$,

$$\mathbf{A}_8 = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Step 2: Find a column vector $\mathbf{v}_i$ such that all the elements of $\mathbf{A}^H\mathbf{v}_i$ have magnitude equal to the $(1/2^n)$. For $N = 8$ dimension, the columns vectors $\mathbf{v}_i$ are:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ -i \\ i \\ -1 \\ 1 \\ -i \\ -i \end{bmatrix},\ \mathbf{v}_2 = \begin{bmatrix} 1 \\ i \\ -1 \\ -i \\ -1 \\ -i \\ -1 \\ -i \end{bmatrix},\ \mathbf{v}_3 = \begin{bmatrix} 1 \\ -i \\ 1 \\ i \\ -i \\ -1 \\ -i \\ 1 \end{bmatrix},\ \mathbf{v}_4 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ -i \\ -i \\ -i \\ i \end{bmatrix},\ \mathbf{v}_5 = \begin{bmatrix} 1 \\ -i \\ -i \\ -1 \\ i \\ -1 \\ -1 \\ -i \end{bmatrix},\ \mathbf{v}_6 = \begin{bmatrix} 1 \\ -1 \\ -i \\ i \\ i \\ i \\ 1 \\ 1 \end{bmatrix},\ \mathbf{v}_7 = \begin{bmatrix} 1 \\ -i \\ -i \\ 1 \\ 1 \\ i \\ -i \\ -1 \end{bmatrix}$$

[1] $\otimes$ denote the element wise operation

Step 3: Now the remaining $N-1$ bases are obtained by element wise multiplication of the vector $\mathbf{v}_i$ with $\mathbf{A}_8$. We have,

$$
\begin{aligned}
B &= \mathbf{v}_1 \otimes A, \\
C &= \mathbf{v}_2 \otimes A, \\
D &= \mathbf{v}_3 \otimes A, \\
E &= \mathbf{v}_4 \otimes A, \\
F &= \mathbf{v}_5 \otimes A, \\
G &= \mathbf{v}_6 \otimes A, \\
H &= \mathbf{v}_7 \otimes A,
\end{aligned}
$$

Therefore with identity matrix $\mathbf{I}_8$, we have obtained $N+1=9$ MUBs. There are also other ways to generate MUBs. MUBs generation using Galois ring is discussed in [176].

It is worth noticing that when the dimension is a power of 2, the elements in the MUBs are either $-1, 1, -i$ or i. In other words, the entries of the MUBs are the powers of i. Thus it can be seen that as if the MUBs are constructed from a finite alphabet drawn from the set of points $\mathcal{X} = \{-1,\ -i,\ 1,\ i\}$. We shall now discuss the motivation behind the choice of using MUBs as a codebook for baseband precoding.

These factors motivates us to employ MUBs. In a communication system, normally, the number of RF chains would be a power of 2. This gravitates us to look up for MUBs when the dimension is a power of 2 [163, 164]. It is interesting to see that for an even prime power, the entities of MUBs are the powers of i .

Also, it satisfies all the properties of a codebook, which shall be shown in this section with an example. Also, the entries are drawn from a finite alphabet $\mathcal{X} = \{-1,\ -i,\ 1,\ i\}$, the matrix has the nested property and all the entries are of constant modulus.

Example: We shall now illustrate the codebook construction from MUBs for 4 transmit antennas.

Following the steps discussed, MUBs for dimension $N=4$ are:

$$
\mathbf{A} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \mathbf{B} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -i & i & i & -i \\ -i & i & -i & i \end{bmatrix},
$$

$$\mathbf{C} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -i & -i & i & i \\ -i & i & i & -i \\ -1 & 1 & -1 & 1 \end{bmatrix}, \mathbf{D} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ i & i & -i & -i \\ 1 & -1 & -1 & 1 \\ -i & i & -i & i \end{bmatrix}$$

Having obtained the MUBs for $N = 4$, we now construct the codebook $\mathcal{F}$ of cardinality 16 (4-bit feedback).

Left circular shift of each of these matrices by p *modulo* 4 places, where $p = 1, 2, 3, 4$, gives the codebook of cardinality $\mathcal{C} = 16$ with 4-bit feedback.

Having obtained the precoder matrices of the codebook
$\mathcal{F} = \{\mathbf{A}_{p \text{ modulo } 4}, \mathbf{B}_{p \text{ modulo } 4}, \mathbf{C}_{p \text{ modulo } 4}, \mathbf{D}_{p \text{ modulo } 4}\} \quad \forall p = 1, 2, 3, 4,$

i.e., after circular shift to left by p *modulo* 4 places for each matrix, we get codebook

$\mathcal{F} = \{\mathbf{A}_0,\ \mathbf{A}_1,\ \mathbf{A}_2,\ \mathbf{A}_3,\ \mathbf{B}_0,\ \mathbf{B}_1,\ \mathbf{B}_2,\ \mathbf{B}_3,\ \mathbf{C}_0,\ \mathbf{C}_1,\ \mathbf{C}_2,\ \mathbf{C}_3,\ \mathbf{D}_0,\ \mathbf{D}_1,\ \mathbf{D}_2,\ \mathbf{D}_3\}$. For instance,

$$\mathbf{A}_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}, \mathbf{A}_2 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix},$$

$\mathbf{A}_3$ is obtained by shifting $\mathbf{A}$ circularly to left by 3 modulo 4 places.

Thus we have obtained a codebook of cardinality 16 which satisfies the properties such as constant modulus, restricted alphabet. Having obtained the TPC matrices of the codebook $\mathcal{F} = \{\mathbf{A}_{p \text{ modulo } 4}, \mathbf{B}_{p \text{ modulo } 4}, \mathbf{C}_{p \text{ modulo } 4}, \mathbf{D}_{p \text{ modulo } 4} \quad \forall p = 1, 2, 3, 4\}$, after the above-mentioned circular shift to left by p *modulo* 4 places for each matrix, we arrive at the codebook

$$\mathcal{F} = \{\mathbf{A}_0,\ \mathbf{A}_1,\ \mathbf{A}_2,\ \mathbf{A}_3,\ \mathbf{B}_0,\ \mathbf{B}_1,\ \mathbf{B}_2,\ \mathbf{B}_3, \mathbf{C}_0,\ \mathbf{C}_1,\ \mathbf{C}_2,\ \mathbf{C}_3,\ \mathbf{D}_0,\ \mathbf{D}_1,\ \mathbf{D}_2,\ \mathbf{D}_3\}.$$

We select the TPC matrix from the specific codebook, which maximizes the minimum SNR for the effective channel using a zero forcing (ZF) receiver [89].

The SNR of the k^{th} stream of (6.1) using a ZF receiver is given by [89]:

$$\text{SNR} = \frac{P}{N_s N_o \left(\mathbf{F}_{\text{BB}}^* \mathbf{H}_{\text{eff}}^* \mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}}\right)_{\text{kk}}^{-1}}, \tag{3.18}$$

where $\mathbf{H}_{\text{eff}} = \mathbf{W}_{\text{RF}}^* \mathbf{H} \mathbf{F}_{\text{RF}}$. Thus, the TPC which maximizes the minimum SNR is found by solving the following optimization problem

$$\mathbf{F}_{\text{BB}} = \arg \max_{\mathbf{F}_{\text{BB}} \in \mathcal{F}} \lambda_{\min}\{\mathbf{H}_{\text{eff}} \mathbf{F}_{\text{BB}}\}, \tag{3.19}$$

Algorithm 3.1 Proposed RF and Baseband Codebook Design

Acquire codebook $\mathcal{F}$ and store in memory

1: Obtain vectors $\mathbf{v}_i$ from Hadamard matrix $\mathbf{A}_{N_t^{\mathrm{RF}}}$ such that magnitude of all elements of $\mathbf{A}^H\mathbf{v}_i$ is 1;
2: Obtain MUBs by element-wise multiplication of $\mathbf{v}_i$ with $\mathbf{A}$, $\forall\ 1 \leq i \leq N_t^{\mathrm{RF}} - 1$;
3: Obtain the codebook $\mathcal{F}$ by left circular shift of each MUB to p places, where $p = 1, \ldots, N_t^{\mathrm{RF}} - 1$;

Selection of Digital TPC from the codebook

1: We have $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$;
2: Obtain columns of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ from $\mathbf{DFT}_{N_t}$ such that $\mathbf{F}_{\mathrm{RF}} = \max < \mathbf{DFT}_{N_t}(:, i), \mathbf{V} >,\ 1 \leq i \leq N_t^{\mathrm{RF}}$, and $\mathbf{W}_{\mathrm{RF}} = \max < \mathbf{DFT}_{N_r}(:, i), \mathbf{U} >,\ 1 \leq i \leq N_r^{\mathrm{RF}}$;
3: Set the effective channel to $\mathbf{H}_{\mathrm{eff}} = \mathbf{W}_{\mathrm{RF}}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}$;
4: Select the TPC from the codebook which satisfies: $\mathbf{F}_{\mathrm{BB}} = \arg\max_{\mathbf{F}_{\mathrm{BB}} \in \mathcal{F}} \Lambda_{\min}\{\mathbf{H}_{\mathrm{eff}}\mathbf{F}_{\mathrm{BB}}\}$;
5: $\mathbf{W}_{\mathrm{BB}}$ is chosen as ZF combiner.

where $\lambda_{\min}$ is the minimum singular value of $\{\mathbf{H}_{\mathrm{eff}}\mathbf{F}_{\mathrm{BB}}\}$.

Since the mmWave channel of the adjacent AEs is correlated because of their $\lambda/2$ spacing, using MUB baseband TPC amalgamated with DFT-aided RF beamforming significantly reduces the complexity, while having an improved performance in spatially correlated channels.
The algorithm 3.1 shows the pseudo code of the proposed design.

3.3.3 Complexity and Performance Analysis of the MUB Codebook

The codebook search complexity is quantified in terms of the number of computations required for finding the best TPC matrix. This requires $N_s N_r^{\mathrm{RF}}(N_t^{\mathrm{RF}} - 1)$ complex-valued additions for evaluating $\mathbf{H}_{\mathrm{eff}}\mathbf{F}_{\mathrm{BB}}$ [164], but no complex multiplications are required as multiplication with $\{\pm 1, \pm i\}$ are trivial operations [178]. In other words, multiplication of matrix with $\{\pm 1\}$ is sign change operation, while multiplication with $\{\pm i\}$ is a sign change operation with swapping real and imaginary parts. In practical digital systems, changing sign or swapping numbers is a trivial operation. On the other hand, the digital TPC operating in the baseband using the same number of RF chains, such as [67], would need $N_s N_r^{\mathrm{RF}}(N_t^{\mathrm{RF}} - 1)$ complex additions plus $N_s N_r^{\mathrm{RF}} N_t^{\mathrm{RF}}$ complex-valued multiplications [164]. Thus, the MUB based codebook reduces the overall search complexity by an order of $\mathcal{O}(N_s N_r^{\mathrm{RF}} N_t^{\mathrm{RF}})$.

The entries of the codebook are drawn from a finite alphabet of say, $\mathcal{X} = \{i,\ -i,\ 1,\ -1\}$, which meets the constant modulus constraint and hence avoids any potential power imbalance for the system. Furthermore, the codebook exhibits a nested structure, which means that the computations that are performed for higher-rank transmissions can be

reused for lower-rank transmissions. For example, for a rank 4 MUB matrix, the computations made for rank 2 transmission can be used for rank 3 transmission and computations made for rank 3 transmission can be used for rank 4 transmission.

Let us now characterize the performance and achievable rate, when the DFT-MUB based RF and the baseband precoding are invoked for the fully-connected architecture. We performed Monte Carlo simulations for investigating the performance gap between the digital baseband TPC based on SVD of the channel matrix and the hybrid TPC relying on limited-feedback. The system configuration and parameters, such as the number of RF chains, antennas and modulation schemes are shown in Table 3.1.

Parameters	*Values*
Modulation	4 QAM
N_c	4
N_{ray}	6
N_t	8, 32, 64
N_r	8, 16, 32
N_s	1, 2
N_t^{RF}	2, 4
N_r^{RF}	2, 4
N_{sub}	1, 2
$\phi_{n_c}^{n_{\text{ray}}}$	$\sim \mathcal{U}[0, 2\pi)$

Table 3.1: System parameters.

Fig. 3.8 plots the achievable rates of unconstrained digital precoding using SVD , of orthogonal matching pursuit [67], of DFT-MUB and DFT-Identity, where the identity matrix is chosen as the baseband TPC, using (3.12) for a 8×8 and 32×16 MIMO systems. In this figure, the DFT-MUB based codebook design with 4-bit feedback is investigated. In this system, two spatial streams were transmitted using 4 RF chains both at the transmitter and at the receiver. It is seen from Fig. 3.8 that the performance gap between the SVD based TPC and the DFT-MUB based codebook is around 2.5dB for 32×16 and as low as 0.8 dB for 8×8 MIMO configurations considered. Furthermore, there is a gap of around 1dB against the orthogonal matching pursuit aided precoding design for 32×16 MIMO, whereas for 8×8 MIMO the performance of DFT-MUB matches with the orthogonal matching pursuit. We also compared our results against DFT-Identity based system, where the TPC is constructed by selecting first N_s columns of the *identity matrix*. Against DFT-Identity aided system, there is around 0.5 dB gain when DFT-MUB is performed.

We also performed Monte Carlo simulations to investigate the attainable uncoded bit error rate (BER) performance of a 8×8 MIMO system using 4 QAM. In this configuration, two spatial streams using 4 QAM were transmitted over $N_t^{\text{RF}} = 4$ and $N_r^{\text{RF}} = 4$ RF chains. Fig. 3.9 shows the comparison between the BERs of the DFT-MUB aided codebook design using 4-bit feedback, of the SVD based TPC, of the DFT-Identity and

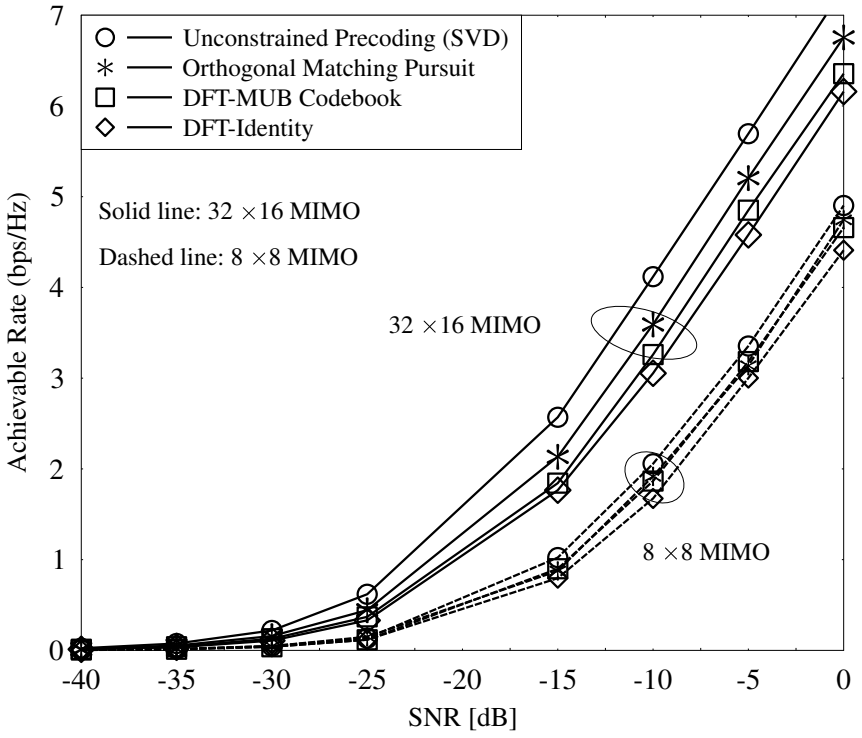


Figure 3.8: Achievable rate using the DFT-MUB based codebook design with a 4-bit feedback and digital precoding with perfect CSI i.e., with unconstrained precoding for 32×16 and 8×8 fully-connected MIMO configurations. In this configuration, $N_s = 2$ and $N_t^{\mathrm{RF}} = N_r^{\mathrm{RF}} = 4$.

of orthogonal matching pursuit. It is seen that at low SNRs the DFT-MUB codebook design outperforms the SVD based TPC. It is instructive to note that SVD based TPC is sub-optimal in terms of BER. Furthermore, at the BER of 10^{-3} the performance gap between the DFT-MUB based design and the SVD design is around 2.5 dB, despite the DFT-MUB scheme's constraint of constant modulus and significantly lower complexity. DFT-MUB also outperforms the DFT-Identity and orthogonal matching pursuit designs.

Additionally, it is important to note the complexity of different designs considered. In the system considered, for $N_t^{\mathrm{RF}} = 4$ and $N_r^{\mathrm{RF}} = 4$, only 24 *complex additions* are needed when using MUB. On the other hand, the orthogonal matching pursuit [67] would need 24 *complex additions* plus 32 *complex multiplications*. Furthermore, the unconstrained digital TPC using SVD relies completely on digital processing elements, where DACs and automatic gain controls (AGCs) are needed per AE, which imposes high complexity and power consumption.

In this subsection, we have analyzed the hybrid precoder without the effects of mutual

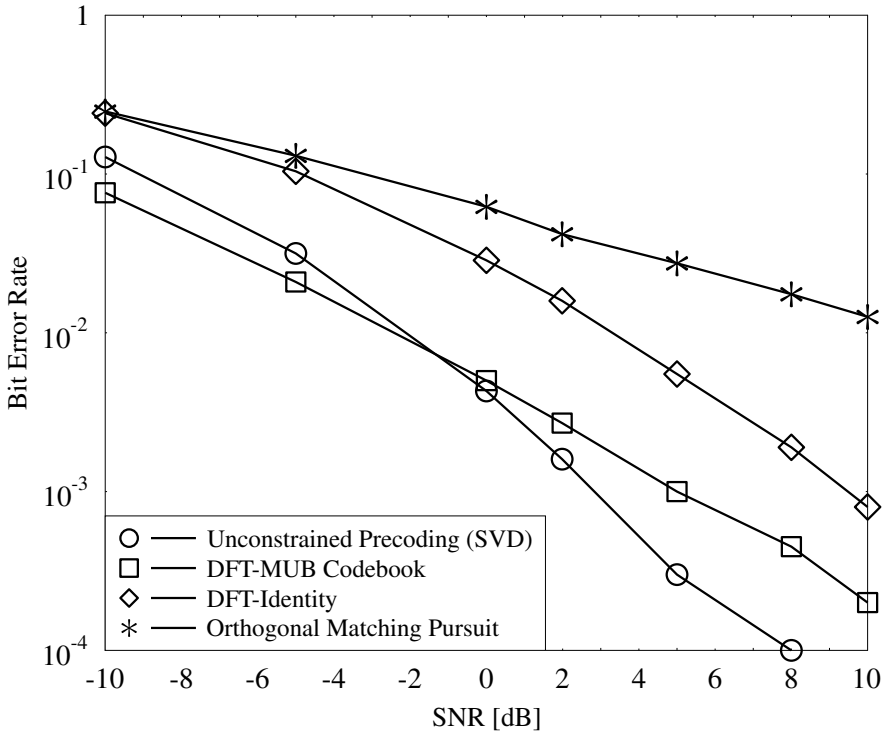


Figure 3.9: Uncoded bit error rate (BER) of a 8×8 fully-connected MIMO DFT-MUB based codebook design with 4-bit feedback and digital TPC with perfect CSI. In this configuration, we have $N_s = 2$ and $N_t^{\mathrm{RF}} = N_r^{\mathrm{RF}} = 4$.

coupling on AEs. Mutual coupling, which we will discuss in the next subsection, can be either detrimental or beneficial depending on the AEs placement. In the next subsection, we focus our attention on the mutual coupling effects on achievable rate.

3.3.4 Effects of Mutual Coupling on Achievable Rate

When the AEs are closely spaced, the radiation of each AE affects the radiation pattern of the other antennas as shown in Fig. 3.10, which in turn affects the input impedance of the AEs. This phenomenon is termed as mutual coupling. Mutual coupling may limit the system's performance in practice, since the radiation efficiency of the antennas is reduced because of the radiation from closely-spaced AEs.

Furthermore, since the AEs are closely spaced, the radiation of one AE affects the

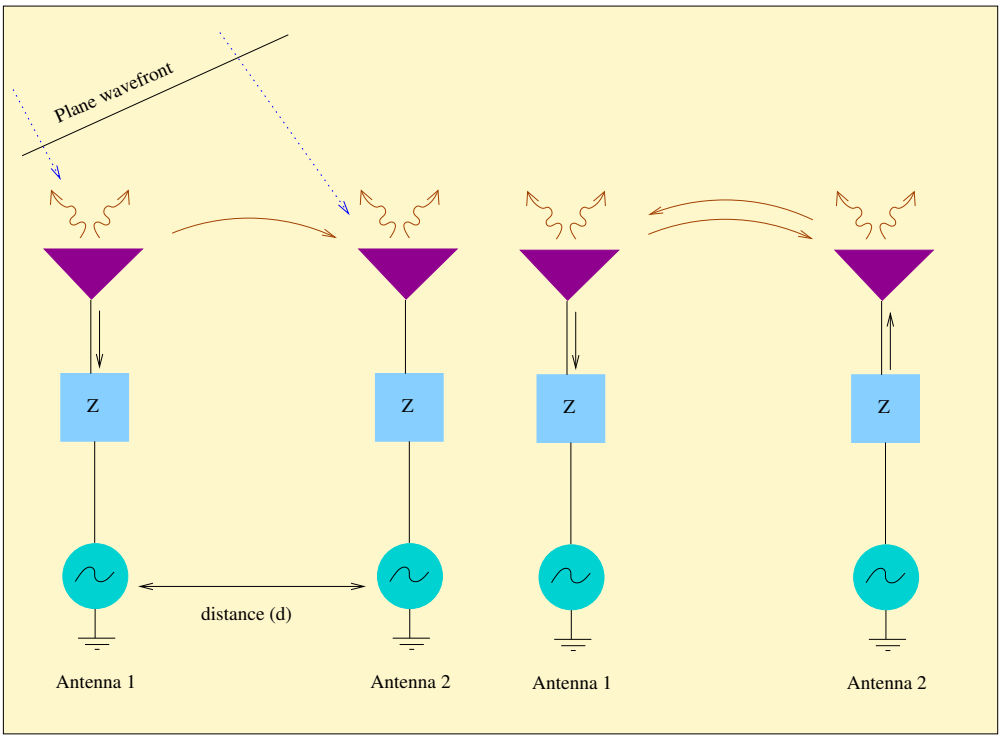


Figure 3.10: Mutual coupling between AEs.

impedance of other elements, which results in mutual coupling [179]. The mutual coupling matrix $\mathbf{C}$ of an AA is given by [179]:

$$\mathbf{C} = (Z_A + Z_T)\left(\mathbf{Z} + Z_T\mathbf{I}_{N_t}\right)^{-1},$$

where Z_T is the load impedance and Z_A is the antenna impedance. The mutual impedance matrix $\mathbf{Z}$ is given by [179]:

$$\begin{pmatrix} Z_A + Z_T & Z_{12} & \dots & Z_{1N_t} \\ Z_{21} & Z_A + Z_T & \dots & Z_{2N_t} \\ \vdots & \vdots & \dots & \vdots \\ Z_{N_t1} & Z_{N_t2} & \dots & Z_{N_tN_T} \end{pmatrix}.$$

For a side-by-side wire dipoles with length l, the expression for Z_{mn} is given by [179]:

$$\begin{aligned} Z_{mn} &= 30[2\mathcal{C}_i(u_\circ) - \mathcal{C}_i(u_1) - \mathcal{C}_i(u_2)] \\ &\quad -j[30(2\mathcal{S}_i(u_\circ) - \mathcal{S}_i(u_1) - \mathcal{S}_i(u_2))], \end{aligned}$$

where $u_\circ, u_1$, and u_2 are calculated as

$$\begin{aligned} u_\circ &= \kappa d_h \\ u_1 &= \kappa(\sqrt{d_h^2 + l^2} + l) \\ u_2 &= \kappa(\sqrt{d_h^2 + l^2} - 1), \end{aligned}$$

where κ is the wavenumber variable equal to $2\pi/\lambda$, while $\mathcal{C}_i$ and $\mathcal{S}_i$ are cosine and sine integrals given by [179]

$$\begin{aligned} \mathcal{C}_i &= \int_\infty^u (\frac{\cos(x)}{x})dx \\ \mathcal{S}_i &= \int_0^\infty (\frac{\sin(x)}{x})dx. \end{aligned}$$

The value of the antenna impedance Z_A for a half-wavelength dipole is equal to $73 + j42.5$ [Ω], while the load impedance Z_T is typically set to the conjugate of the antenna impedance so as to obtain the impedance match, i.e. $Z_T = Z_A^*$ [179].

For the system model considered in (6.1), the achievable rate, R, is given by

$$R = \log_2 \det \left(\mathbf{I}_{N_s} + \frac{P}{N_s} \mathbf{R}_n^{-1} \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{C} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{C}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}} \right), \tag{3.20}$$

where $\mathbf{R}_n = \sigma^2 \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}$ and P is the signal power. Fig. 3.12 shows the achievable rate performance of the DFT-MUB hybrid precoding [98] and of the optimally unconstrained precoding, where the precoder matrix is chosen as the right singular vectors and the combiner as the left singular vectors of the channel matrix $\mathbf{H}$ relying on singular value decomposition (SVD). Furthermore, the unconstrained precoding is fully-digital precoder relying on a simplifying assumption of having perfect channel state information. In Fig. 3.12, a 64×32 MIMO is used, where two spatial streams are transmitted using four RF chains. It can be seen in Fig. 3.12 (a) and (b) that the mutual coupling degrades the achievable rate of the system employing antenna spacing of 2λ and $\lambda/2$. On the other hand, when the spacing between the adjacent AEs is $\lambda/4$ as in Fig. 3.12 (c), the achievable rate of the system considering mutual coupling becomes better than that without mutual coupling. The capacity gain attained by reducing the antenna spacing to $\lambda/4$ is due to the de-correlation effect introduced by the mutual coupling. It is instructive to note that mutual coupling also affects the received power, where it has been shown in [180] that the achievable rate is more influenced by the de-correlation effect for small antenna spacings.

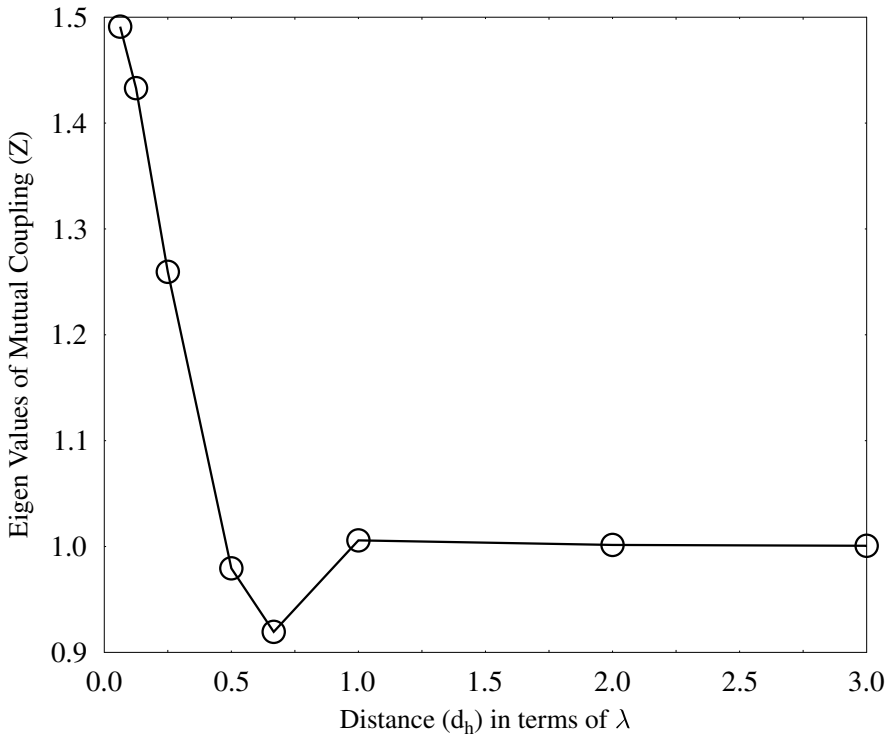


Figure 3.11: Eigenvalues of the mutual coupling matrix $\mathbf{C}$ versus the AE separation distance.

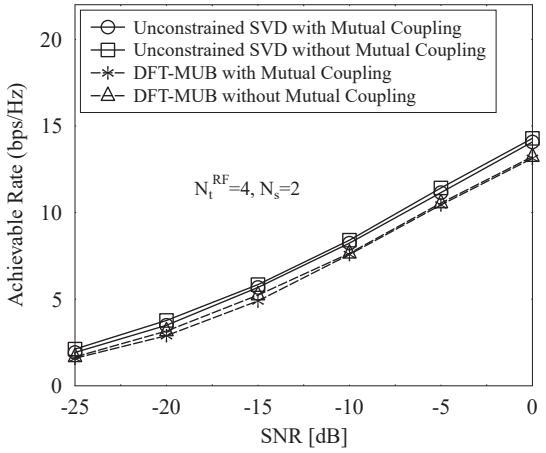
To elaborate further, let us denote $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} = \mathbf{F}$ and $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}} = \mathbf{W}$. Then, the achievable rate equation of (3.20) at high SNR can be written as [4]:

$$R = \log_2 \det \left(\frac{P}{N_s} \mathbf{R}_n^{-1} \mathbf{W}^H \mathbf{H} \mathbf{C} \mathbf{F} \mathbf{F}^H \mathbf{C}^H \mathbf{H}^H \mathbf{W} \right), \tag{3.21}$$

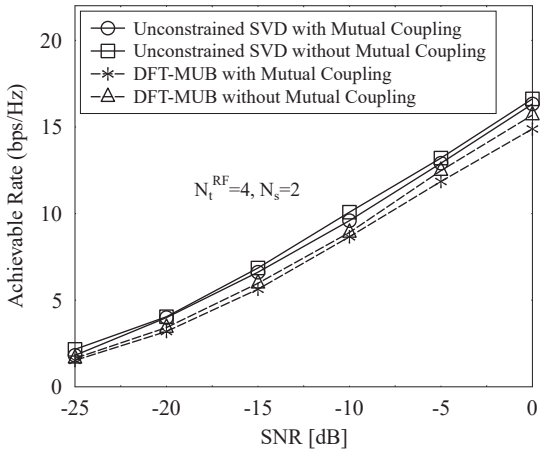
which can be written in terms of the rate without mutual coupling R_{nc} plus the difference due to the mutual coupling as:

$$R = R_{\mathrm{nc}} + \log_2 \det \left(\mathbf{C}\mathbf{C}^H \right). \tag{3.22}$$

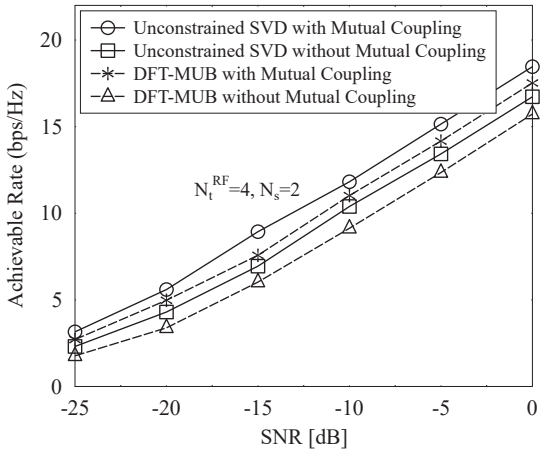
Fig. 3.12 shows the achievable rate performance of the DFT-MUB hybrid precoding [98] and of the optimally unconstrained precoding, where the precoder matrix is chosen as the right singular vectors and the combiner as the left singular vectors of the channel matrix $\mathbf{H}$ relying on SVD. Furthermore, the unconstrained precoding is fully-digital precoder relying on a simplifying assumption of having perfect channel state information. In Fig. 3.12, a 64×32 MIMO is used, where two spatial streams are transmitted using four RF chains. It can be seen in Fig. 3.12 (a) and (b) that the mutual coupling degrades the



(a)



(b)



(c)

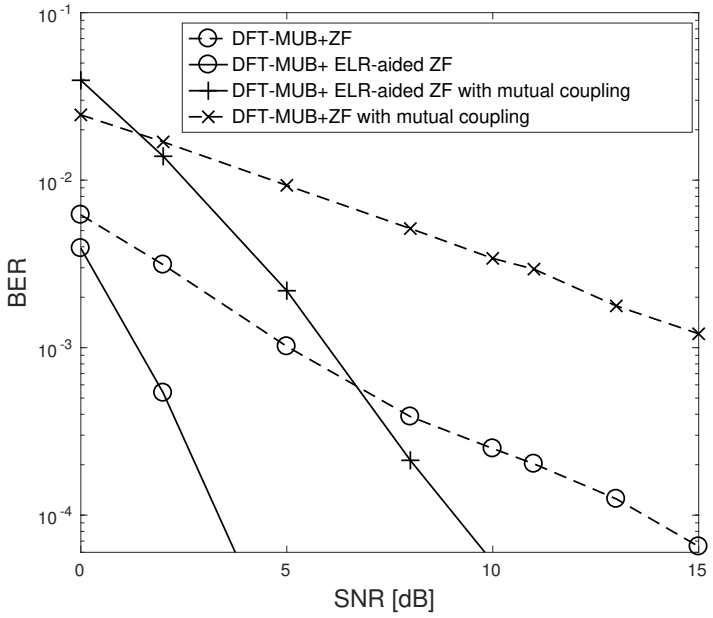
Figure 3.12: Achievable rate of the unconstrained precoding and of the DFT-MUB hybrid precoding both with and without mutual coupling for a 64×32 MIMO system. In this setting, two spatial streams are transmitted using four RF chains when (a) the spacing between the adjacent antenna is set to 2λ, (b) the spacing between the adjacent antenna is set to $\lambda/2$, (c) the spacing between the adjacent antenna is set to $\lambda/4$.

achievable rate of the system employing antenna spacing of 2λ and $\lambda/2$. On the other hand, when the spacing between the adjacent AEs is $\lambda/4$ as in Fig. 3.12 (c), the achievable rate of the system considering mutual coupling becomes better than that without mutual coupling. The capacity gain attained by reducing the antenna spacing to $\lambda/4$ is due to the de-correlation effect introduced by the mutual coupling. It is instructive to note that mutual coupling also affects the received power, where it has been shown in [180] that the achievable rate is more influenced by the de-correlation effect for small antenna spacings.

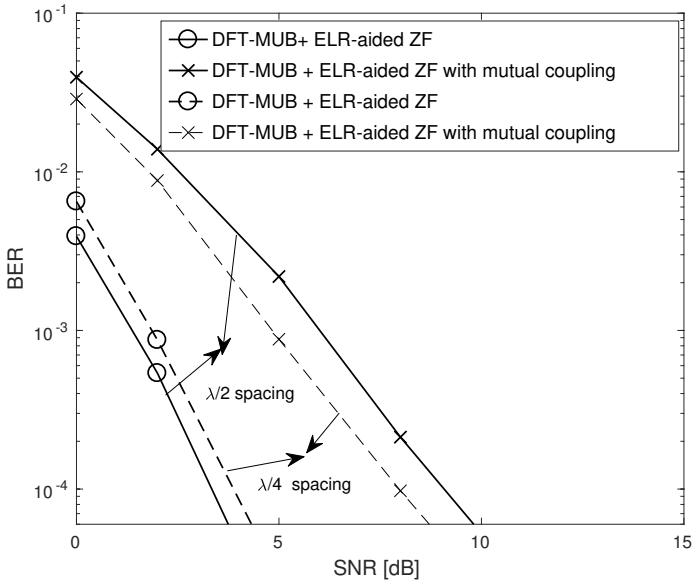
Furthermore, since $\det\left(\mathbf{C}\mathbf{C}^H\right) = \prod_{i=1}^{N_t} c_i$, where c_i represent the eigenvalues of the mutual coupling matrix ($\mathbf{C}$), the achievable rate depends on the eigenvalues of $\mathbf{C}$. Fig. 3.11 shows the eigenvalues of the mutual coupling matrix $\mathbf{C}$ versus the antenna separation distance, where it is shown that for distances less than $\lambda/2$ the eigenvalues are greater than 1, which means that $\log_2 \det\left(\mathbf{C}\mathbf{C}^H\right) > 0$ and hence the rate achieved with mutual coupling is greater than that of without mutual coupling as seen in Fig. 3.12 (c). On the other hand for distances between $\lambda/2$-to-λ, the eigenvalues are less than one, which means that the rate achieved with mutual coupling is less than that of without mutual coupling, as seen in Fig. 3.12 (a) and (b).

Finally, in order to understand the effect of mutual coupling on the system performance, we simulated the BER performance of the system employing the ELR-aided ZF decoding. Fig. 3.13 shows the BER performance for a 8×8 MIMO, where two 4-QAM symbols are transmitted using four RF chains. More explicitly, Fig. 3.13 (a) shows the performance of the system employing DFT-MUB with zero-forcing (ZF) and ELR-aided ZF decoder. It can be seen from the Fig. 3.13 that the ELR-aided ZF decoding significantly improves the system performance compared with the ZF decoder. There is around 5 dB gain at the BER of 10^{-4}. Furthermore, when the mutual coupling is considered, the BER performance of the system employing both decoders degrade, when the spacing between the adjacent AEs is $\lambda/2$, as shown in Fig. 3.13(a). In Fig. 3.13(b) we plot the BER performance of the system employing ELR-aided ZF with different antenna spacing in the presence of mutual coupling. As shown in Fig. 3.13 (b), the performance of the system employing $\lambda/4$ antenna spacing is worse than that employing $\lambda/2$ spacing when no mutual coupling is considered. This is expected due to the higher channel spatial correlation at $\lambda/4$ separation. On the other hand, when considering mutual coupling, we observe in Fig. 3.13 (b) that the performance of the system employing $\lambda/4$ antenna separation is now better than that for the system with $\lambda/2$, which is due to the decorrelation effect of the mutual coupling, which can be observed from Fig. 3.11.

Having studied the metrics such as the capacity, complexity of the proposed design, we now delve into a dual-stage adaptation using our proposed architecture, where the adaptation takes place base on the channel characteristics.



(a)



(b)

Figure 3.13: BER performance of the DFT-MUB with ZF and of DFT-MUB with ELR-aided ZF decoder for a 8×8 MIMO system. In this setting, two spatial streams are transmitted using four RF chains. (a) shows a comparison of the BER performance with ZF and ELR-aided ZF decoders when the spacing between the adjacent antennas is set to $\lambda/2$ and (b) shows a comparison of the BER performance for the system employing ELR-aided ZF decoding when the antenna spacing is varied from $\lambda/2$ to $\lambda/4$.

3.4 Dual-Stage Adaptation of the Dual-Function Transceiver

In this section, we present an adaptive array design for HBF in mmWave communication that adapts the architecture as well as the digital precoding depending on whether the channel is of LOS or NLOS nature. Explicitly, we design an adaptive array based architecture, where the array is fully-connected in LOS channel with an antenna spacing of $\lambda/2$, where λ is the wavelength. By contrast, in NLOS channels the sub-array-connected architecture is employed, where the AA is partitioned into sub-arrays with a sufficiently large separation so that each sub-array experiences independent fading to attain both beamforming and diversity gains.

3.4.1 Near-Instantaneously Adaptive Transceiver Design

In this section we present the architectural and hybrid precoding based adaptation depending on the nature of the channel. More explicitly, the proposed system can be adapted at two levels. The first level of the adaptation focuses on transmitter hybrid architecture design which is based on whether the channel is of LOS or NLOS nature, and the second level of the adaptation is based on the hybrid precoding employed.

We note that the LOS channels exhibit a single dominant path, hence they have a single degree of freedom for signal transmission. Therefore, the channel is devoid of any multiplexing/diversity gains that can be obtained at the transmitter/receiver. On the other hand, the NLOS channels offer multiple degrees of freedom due to the rich scattering environment. As a benefit in NLOS channel conditions, the transmitter can achieve multiplexing/diversity gains. We use this philosophy for our dual-stage adaptation.

In this section, the adaptation is discussed at two levels: one in array design, and the other in hybrid precoding. In the first level of adaptation, we switch between the two array designs, namely fully-connected and 2-sub-array-connected. The rationale for choosing only these two designs is that fully-connected design provides full BF gain, while the 2-sub-array-connected design provides both diversity and BF gains. We have shown in Sec. 3.2 that having more than 2-sub-array-connected design degrades the performance in terms of achievable rate, because of the diminishing returns of the diversity gain, which hardly compensates for the reduction in the BF gain caused by splitting the AA. Therefore, we have decided to adapt between these two specific architectures.

First level of adaptation: The first level of adaptation is between the two choices of fully-connected and 2-sub-array-connected designs. Fig. 3.14 shows the adaptive array design equipped with switches A, B and C in order to facilitate both fully-connected and 2-sub-array-connected designs. It is instructive to note that although the transmitter in Fig. 3.14 is equipped with $3N_t/2$ AEs to facilitate both designs, only N_t AEs are used at any given time. In other words, the design of Fig. 3.14 is an amalgamation of both

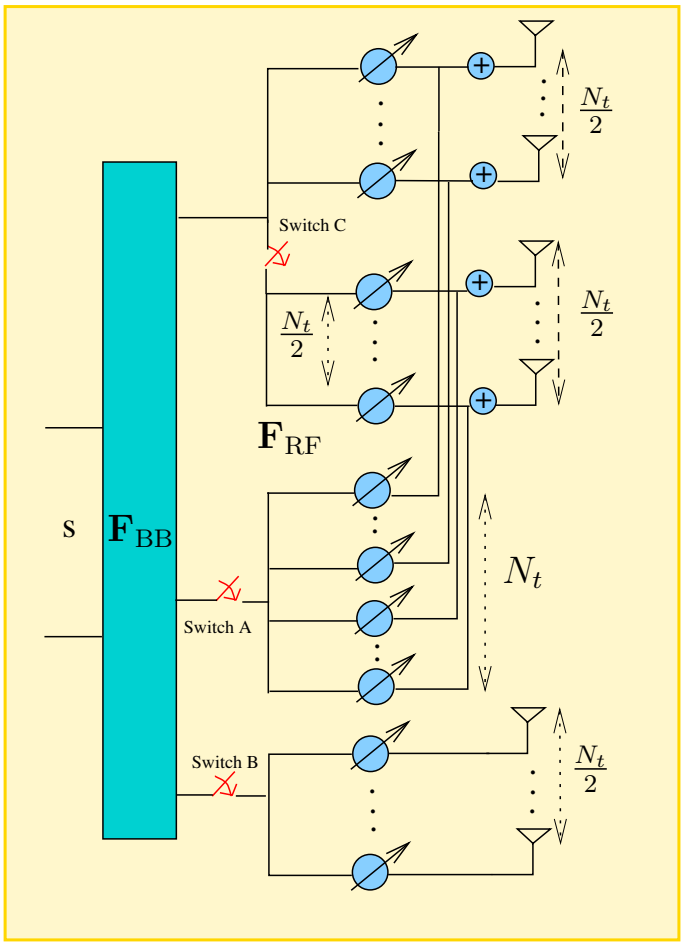


Figure 3.14: Adaptive array design, where switches A and C are activated to employ the fully-connected array in LOS channel conditions, while only switch B is activated to employ the 2-sub-array-connected array in NLOS channel conditions so as to attain both diversity and BF gains.

Fig. 2.23 and Fig. 3.2. Therefore, to employ the fully-connected design, switches A and C are activated to form a full AA having N_t AEs. By contrast, only switch B is activated to employ the 2-sub-array-connected design, where the top sub-array associated with $N_t/2$ AEs and the bottom sub-array with $N_t/2$ AEs are the constituent components of the design. More explicitly, to activate the fully-connected design, switches A and C are turned on while switch B is off. On the other hand, to activate the 2-sub-array-connected design, switch B is turned on while switches A and C are off.

To illustrate Fig. 3.14 in detail, let us consider an example, where the channel is of NLOS nature. In this setting, it was shown in [4] that having a 2-sub-array-connected design would provide better performance in terms of the achievable rate, because this design is capable of exploiting the diversity gain of the spatial degrees of freedom available from the channel as well as the BF gain of $N_t/2$ derived from the AA within the sub-array. This is achieved in Fig. 3.14 by activating switch B only, where the transmitter now behaves as a 2-sub-array-connected scheme with $N_t/2$ AEs in each sub-array. By contrast, when the channel is of LOS in nature, there is only an angular degree of freedom, since the signal

arriving from the sub-arrays impinge at the receiver at different angles of arrival, which is deprived of the additional spatial diversity gain obtained in NLOS channel conditions. Hence, we employ the fully-connected design, where we aim for achieving the full BF gain of N_t. This is achieved in Fig. 3.14 by activating the switches A and C while turning the switch B off.

Therefore, upon determining the channel conditions relying on Kurtosis-based LOS/N-LOS identification method of [181], the receiver relays the information concerning the LOS/NLOS flag to the transmitter, which in turn then adapts to the array design according to the channel conditions. It is instructive to note that this adaptive design finds applications where the users are distributed in a way that some of them experience LOS and others NLOS channel. The transmitter adapts its array accordingly to improve the throughput of the LOS/NLOS users, as it will be shown later in Fig. 3.17.

Second level of adaptation: The second level of adaptation is employed in the hybrid precoding. In LOS channels, using ABF at the RF stage is sufficient to capture the dominant LOS paths, as it will be shown in Fig. 3.16 and any digital precoder in the baseband would be redundant, since no multiplexing gain can be achieved in LOS channels. Hence, when the array is fully-connected in LOS channels, only ABF is employed and the baseband TPC is chosen as the identity matrix, where virtually no signal processing is carried out.

By contrast, when the 2-sub-array-connected array is used in NLOS channels, employing analog RF beamforming and baseband precoding is preferred, as it will be shown in Fig. 3.15. This is because the NLOS channel provides multiplexing/diversity gains, which ABF alone fails to capture effectively owing to the limitations of the analog phase shifters. Hence, baseband precoding is applied in addition to the analog RF beamforming to achieve the multiplexing/diversity gains in NLOS channels.

To characterize the performance of the constituent array designs in HBF, we employ the following hybrid precoding[2]

1. *DFT-MUB:* In this configuration, we employ a discrete Fourier transform (DFT) aided beamformer for analog RF beamforming (ABF) discussed in Sec. 3.3, where the beamformer weights are chosen from the DFT matrix. More explicitly, the columns of the ABF matrix $\mathbf{F}_{\mathrm{RF}}$ are chosen from the DFT matrix which exhibit maximum correlation with the right singular vectors of the channel matrix $\mathbf{H}$, and the columns of the ABF matrix $\mathbf{W}_{\mathrm{RF}}$ are chosen from the DFT matrix which exhibit maximum correlation with the left singular vectors of the channel matrix $\mathbf{H}$.

   On the other hand, the digital TPC matrix $\mathbf{F}_{\mathrm{BB}}$ is chosen from a codebook constructed from the mutually unbiased bases (MUB). More explicitly, the specific

[2]Note that any hybrid precoder combination that performs efficiently can be employed.

TPC matrix is selected from the codebook which maximize the minimum SNR of the effective channel matrix $\mathbf{H}_{\text{eff}} = \mathbf{W}_{\text{RF}}^{H}\mathbf{H}\mathbf{F}_{\text{RF}}$. The digital combiner matrix $\mathbf{W}_{\text{BB}}$ is chosen as the linear minimum mean squared error (LMMSE) solution.

2. *DFT-Identity:* Similar to the DFT-MUB hybrid precoding, we employ DFT assisted beamforming in the RF stage for both ABF matrices. However, for the baseband digital TPC an identity matrix is selected, where the first N_s columns of the identity matrix are chosen as the TPC weights. In other words, in this scenario there is virtually no digital processing of the signals in the baseband.

We will first analyze all the possible combinations of the array design with the aforementioned beamforming methods both in LOS and NLOS channel conditions. All the combinations of the array designs, the hybrid precoding and channel conditions are summarized in Table 3.2.

Table 3.2: *Combinations of array configuration and hybrid precoding designs.*

Array Design	*Channel*	*Hybrid Precoding*
Fully-connected	NLOS	DFT-MUB (ABF-TPC)
Fully-connected	NLOS	DFT-Identity (ABF-Identity)
Fully-connected	LOS	DFT-MUB (ABF-TPC)
Fully-connected	LOS	DFT-Identity (ABF-Identity)
2-sub-array-connected	NLOS	DFT-MUB (ABF-TPC)
2-sub-array-connected	NLOS	DFT-Identity (ABF-Identity)
2-sub-array-connected	LOS	DFT-MUB (ABF-TPC)
2-sub-array-connected	LOS	DFT-Identity (ABF-Identity)

The performance and achievable rate of all the combinations are shown in Fig. 3.15 and Fig. 3.16 for NLOS and LOS channels, respectively. The system configuration and parameters, such as the number of antennas, number of RF chains and modulation scheme are shown in Table 3.3.

Table 3.3: *System parameters.*

Parameters	*Values*
Modulation	4 QAM
N_t	64
N_r	16
N_t^{RF}	2
N_r^{RF}	2
N_s	1
N_{sub}	2
$\phi_{n_c}^{n_{\text{ray}}}$	$\mathcal{U}[0, 2\pi)$

Fig. 3.15 shows the achievable rate of a 64×16 MIMO in NLOS channel, when the DFT-MUB and DFT-Identity schemes are employed for the fully-connected as well as

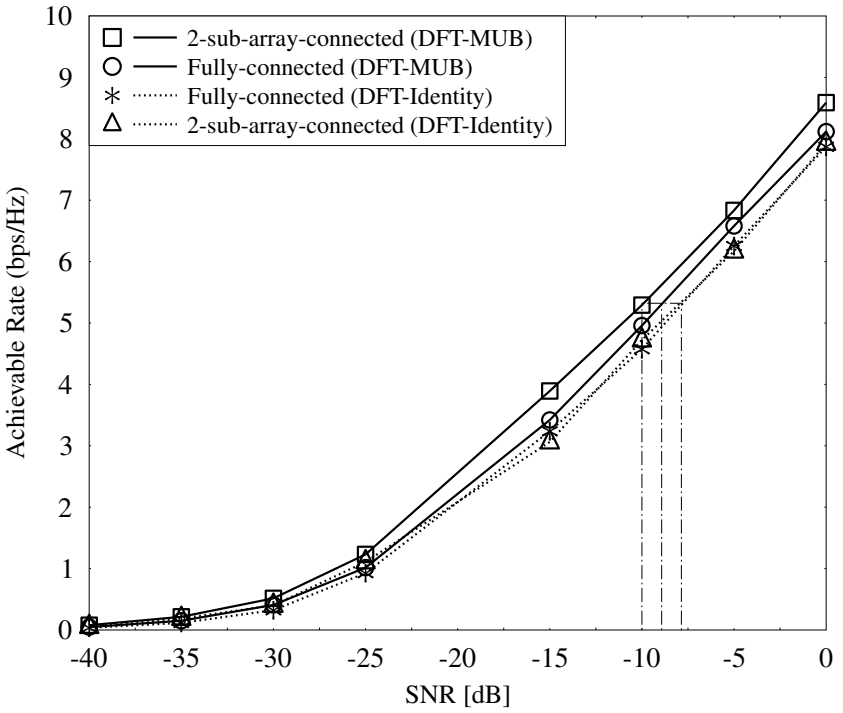


Figure 3.15: Achievable rate of 64×16 MIMO for the two designs with DFT-MUB (ABF-TPC) and DFT-Identity (ABF-Identity) in NLOS channel conditions, when $N_t^{\mathrm{RF}} = 4$ and $N_r^{\mathrm{RF}} = 4$.

for the 2-sub-array-connected architectures. Fig. 3.15 shows that for both fully-connected and 2-sub-array-connected designs, the DFT-MUB outperforms the DFT-Identity based TPC in NLOS channel. This is because in NLOS conditions, where the channel is rich in spatial diversity, the ABF in the RF stage fails to capture some of the spatial degrees of freedom, while the TPC can efficiently capture them. Furthermore, the 2-sub-array-connected design provides a better achievable rate than the fully-connected by a gain of around 1 dB in NLOS channels. This is due to the additional spatial diversity gain obtained from the spatially separated sub-arrays, which exceeds the loss in the beamforming gain [4].

Fig. 3.16 shows the achievable rate of a 64×16 MIMO in LOS channel condition when the DFT-MUB and DFT-Identity are employed for the two hybrid precoding architectures. It is seen in Fig. 3.15 that for both the fully-connected and 2-sub-array-connected designs, the achievable rate of DFT-MUB is similar to that of the DFT-Identity. This is because the ABF used in the RF stage can efficiently exploit the dominant path. Hence, employing TPC does not give any added advantage. Therefore, *any digital processing in the baseband is redundant in LOS channels*. Furthermore, it is also seen in Fig. 3.16 that in LOS channels, the fully-connected design is capable of providing a better rate than the 2-sub-array-connected design by about 2.5 dB gain. This is because we have no spatial

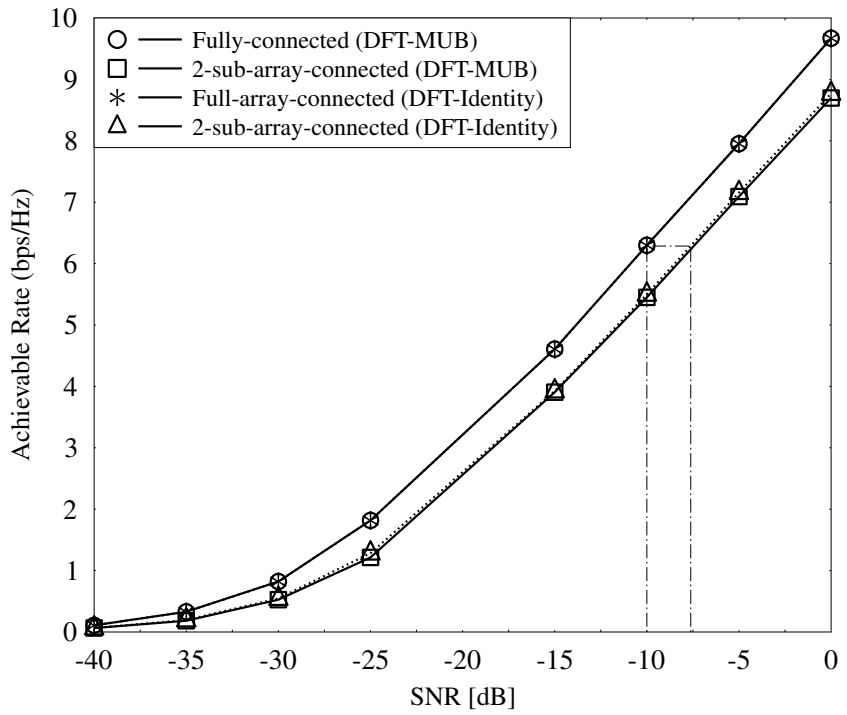


Figure 3.16: Achievable rate of 64×16 MIMO for the two designs with DFT-MUB (ABF-TPC) and DFT-Identity (ABF-Identity) in LOS channel conditions, when $N_t^{\mathrm{RF}} = 4$ and $N_r^{\mathrm{RF}} = 4$.

Table 3.4: *Adaptive array configuration and hybrid precoding designs.*

Channel Condition	Array Design	Hybrid Precoding
LOS	Fully-connected	ABF-Identity
NLOS	2-sub-array-connected	ABF-TPC

diversity in LOS channels and thus it is more profitable to employ full-beamforming than partitioning the array into sub-arrays.

Based on the results shown in Fig. 3.15 and 3.16, of all the combinations listed in Table 5.5, our adaptive design is configured to be fully-connected employing only ABF in LOS channels, and 2-sub-array-connected employing both ABF and TPC in NLOS channel conditions. In other words, in the scenario where the channel is of LOS nature, we employ carefully designed weights for ABF and choose the weights for the baseband TPC from an identity matrix. This is because the ABF alone captures the dominant LOS path. Hence, further digital processing of the signals using a TPC does not benefit us with any performance gain. Then, in the scenario where the channel is NLOS nature, we employ ABF-TPC, since the TPC is capable of capturing the degrees of freedom more efficiently. Table 3.4 shows the adaptation of the array design based on the channel conditions.

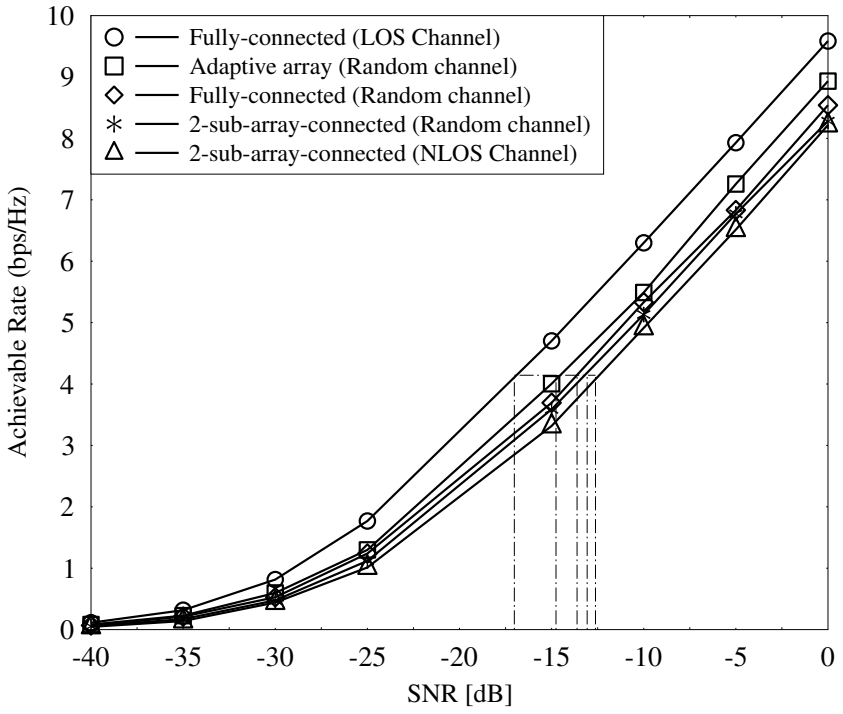


Figure 3.17: Achievable rate of 64×16 MIMO with adaptive design, where ABF-TPC is applied in NLOS and ABF-Identity is applied in LOS conditions. The channel is random since it varies between LOS and NLOS with probability 0.5.

Thus, to leverage the diversity and beamforming gains accordingly to the channel conditions, the receiver determines the channel conditions relying on the Kurtosis-based LOS/NLOS identification method of [181] and feeds back the information flag to the transmitter over the feedback channel. Upon receiving this information, the transmitter adapts its array design and hybrid precoding accordingly.

Adaptive system results: Here we present simulation results for characterizing the performance of the fully-connected design, of the 2-sub-array connected design and of the adaptive design in both LOS and NLOS channels, where channel switches from LOS to NLOS with a probability of 0.5. We performed Monte Carlo simulations for analyzing the performance of the designs, when the ABF-TPC and ABF-Identity schemes are employed using the capacity equation in (3.12). The system configuration and parameters, such as the number of antennas, number of RF chains and modulation schemes are shown in Table 3.3.

Fig. 3.17 shows the achievable rate of a 64×16 MIMO relying on the adaptive design, where the array switches to fully-connected relying on ABF-Identity precoding in LOS and to 2-sub-array-connected relying on ABF-TPC in NLOS conditions. The plot in

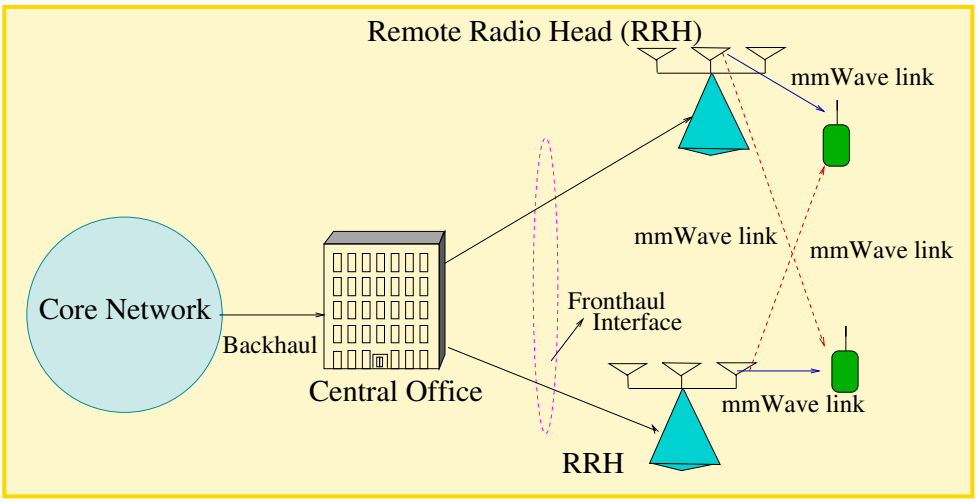


Figure 3.18: C-RAN architecture.

Fig. 3.17 shows the achievable rate and performance benefits of adaptive design, of fully-connected design in all channel conditions, and of 2-sub-array-connected in all channel conditions, where the random channel label refers to either LOS/NLOS. It is seen in Fig. 3.17 that the adaptive design achieves 1.5 dB gain against the fully-connected design and about 2 dB gain against the 2-sub-array-connected design, when the channel varies between LOS and NLOS.

We extend this analysis in the context of Cloud-Radio Access Network (C-RAN), where a user can be connected to one remote radio heads (RRHs) using fully-connected design or to two RRHs forming a 2-sub-array-connected design. The adaptation between these two design depends on the channel condition between the user and RRHs. A detailed explanation of the adaptive virtual cell formation in C-RAN is presented in the next section.

3.4.2 Virtual Cell Formation in mmWave C-RAN: A Use Case

The main idea of C-RAN is to split the BSs into low power and low complexity RRHs coordinated by a central unit (CU) located at the central office (CO) [182] as shown in Fig. 3.18.

This relies on using fronthaul links that connect the RRHs to the CU, with the aid of diverse transmission media that includes fiber-optic cable, digital subscriber lines (DSL), or wireless mmWave channels, as shown in Fig. 3.18. In C-RAN, the centralized/joint signal processing of RRHs is carried out at the CO by the CU, which makes the architecture both cost-effective and energy-efficient [182]. This makes the C-RAN an attractive choice for network densification [183]. Furthermore, the distributed nature of C-RAN can be seen as an extension of the ASA, i.e. when only one RRH is used, it is analogous

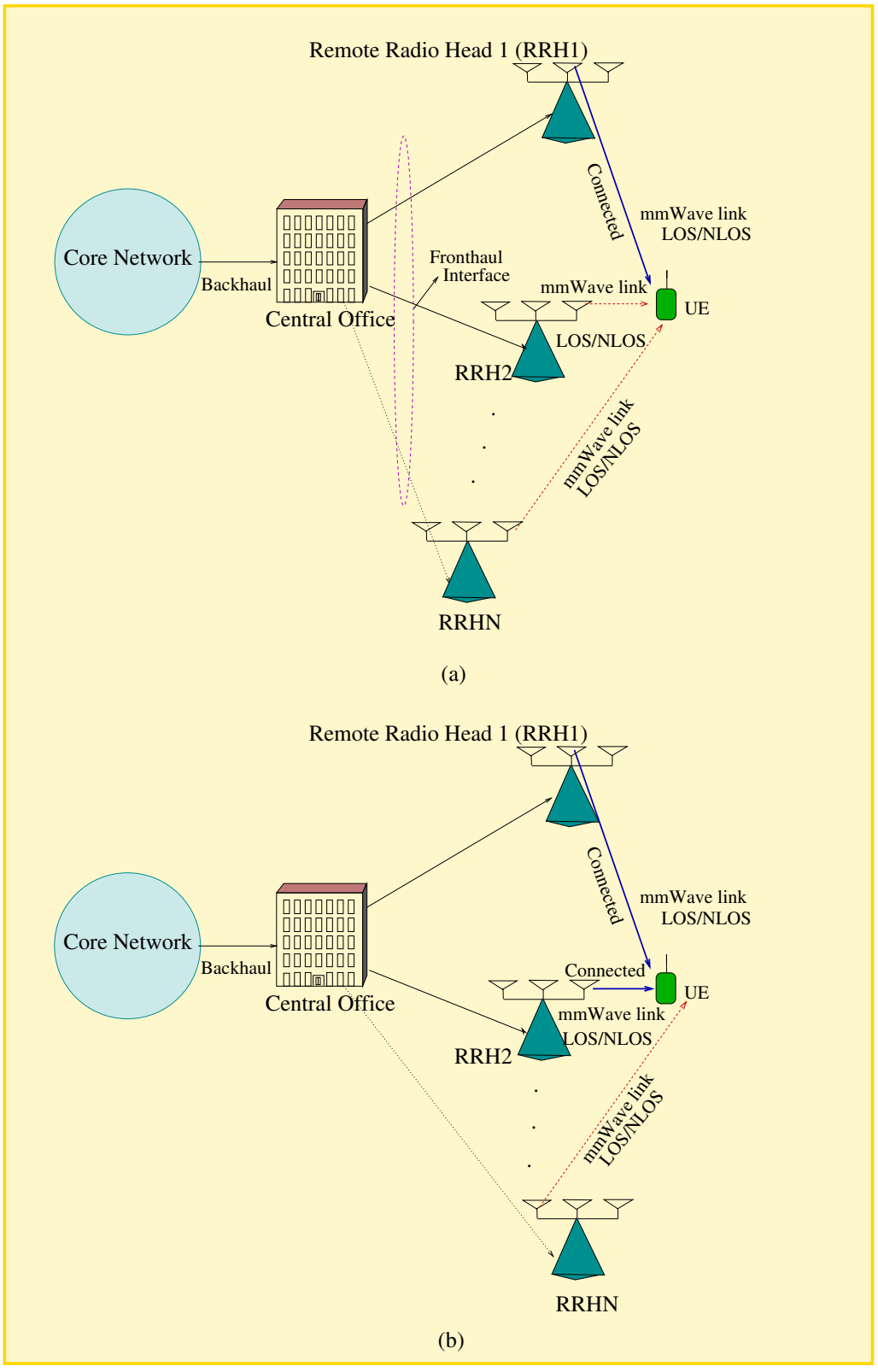


Figure 3.19: (a) RRH1 connected to user equipment (UE) using fully-connected design with N_t AEs. (b) Two RRHs connected to UE using fully-connected design with $N_t/2$ AEs on each to form a 2-sub-array-connected design.

to fully-connected design, and when more than one RRH is used, it is analogous to the ASA design, where each sub-array maybe fully-connected.

In this section, we utilize the adaptive array design for the physical layer design of C-RAN, where we assume that the total user load within the limited capacity of the fronthaul, in order to meet the finite capacity constraint imposed on the fronthaul [184]. Furthermore, the virtual cell formation here refers to the process where a mobile user connects to one or two RRH(s). As discussed in the previous section, when the user is in the LOS channel condition, the transmitter employs fully-connected design, while when the user in the NLOS channel condition, the transmitter employs 2-sub-array connected design. In this paper, we relate this to C-RAN, where one or two RRH(s) are used to form fully-connected or 2-sub-array-connected designs. The rationale for strictly limiting the user connections with two RRHs is because of the performance degradation in terms of achievable rate when more than two RRHs are used for the user association with the same transmit power. A more detailed mathematical analysis is provided in [4].

In the context of the proposed architectures, a single RRH can connect with a user using the fully-connected architecture as shown in Fig. 3.19. In Fig. 3.19 (a), the user equipment may connect to RRH1, RRH2, ..., RRHN according to the specific design criteria to be described in the next paragraph. Observe that RRH1 has been selected as the preferred RRH, where a fully-connected architecture is employed. Additionally, it is possible that two RRHs can form a virtual sub-array connected architecture as shown in Fig. 3.19 (b). Explicitly, as shown in Fig. 3.19 (b) RRH1 is coordinated with RRH2 to form a 2-sub-array-connected with $N_t/2$ AEs[3] in each RRH and forms a virtual cell with the UE as shown in Fig. 3.19 (b) with 2 bold lines from RRH1 and RRH2. More particularly, we propose a virtual cell formation algorithm for a single user, where the RRHs experience different channel conditions (LOS/NLOS), and the adaptive array design of the Sec. 3.4.1 is used while forming virtual cells. Then we extend our work to a multi-user environment, where the users are randomly distributed.

By acknowledging the challenges imposed on the fronthaul in C-RAN, such as its bandwidth, strict latency and jitter as well as the need for low cost transport network [185], in our proposed design we consider the maximum user load that can be supported by the fronthaul link. This corresponds to the total user load within the finite capacity of the fronthaul. However, to alleviate the fronthaul constraints, more effective signal processing techniques can be invoked to optimize the spectral efficiency, energy efficiency and delay, as in [186, 187]. Furthermore, to tackle the limited fronthaul capacity under heavy data traffic, techniques relying on sophisticated signal quantization and compression can be employed [15, 188]. More particularly, we focus our attention on the virtual cell formation, where a mobile user can simultaneously connect to one or more than

[3]Note that the RRH always is fully-connected, however, the number of AEs to form the fully-connected design may vary in order to meet the transmit power constraint.

one RRH(s) [189]. Furthermore, in the case of multi-user scenarios, we consider the maximum number of users that can be supported in the fronthaul link.

It is instructive to note that, in our work, the digital TPC weight matrix $\mathbf{F}_{\mathrm{BB}}$ discussed in Sec. 3.3 will be carried out at the CU, while the ABF weight matrix $\mathbf{F}_{\mathrm{RF}}$ will be employed at the RRH, although the possible locations (RRH/CU) for computing the ABF and digital TPC matrices can be subsumed into four categories, which are: i) both ABF and TPC at the CU; ii) TPC at the CU and ABF at the RRH; iii) ABF at the CU and TPC at the RRH iv) both ABF and TPC at the RRH. However, in this paper, ABF is carried out at the RRH owing to the large overhead in relaying the channel matrix to the CU, especially in mmWave communications, since large AAs would be employed. On the other hand, the digital TPC is designed relying on the effective channel matrix $\mathbf{H}_{\mathrm{eff}}$ as discussed in Sec. 3.3. Hence, overhead in relaying the effective channel matrix from the RRH to the CU would be low. Furthermore, since the baseband digital TPC is a joint signal processing for both RRHs, computing the TPC at the CU is an attractive solution.

1. **Single User Cell Formation:**

   Consider a single user that can be served by multiple RRHs, where the channel conditions between the user and each RRH can be classified as LOS and NLOS, as shown in Fig. 3.20.

   In Fig. 3.20 each RRH is equipped with N_t AEs and the architecture of RRH can be either only fully-connected with N_t AEs or 2-sub-array-connected using two RRHs with each RRH using $N_t/2$ AEs, depending on the channel conditions. For the fully-connected design, all the N_t AEs are activated, while in the 2-sub-array-connected design only $N_t/2$ AEs are used in each RRH. It is instructive to note that N_t AEs can be employed for achieving better performance at the expense of a higher energy consumption. The algorithm of single user virtual cell formation is given in Algorithm 1 and the detailed exposition of the algorithm is presented below.

2. **Multi-User Cell Formation:**

   In this section, we discuss a multi-user scenario, where the users are randomly distributed and the channel between the RRHs and users can be either LOS or NLOS. Furthermore, each user can be associated with multiple RRHs either in LOS or NLOS and each RRH can be associated with multiple users, unlike in the single user case. However, as in the single user case, in the multi-user scenario, the CU computes the capacity of the links associated with the first two highest SNRs in both LOS and NLOS for each user and then forms the virtual cell for that user using the link which gives the highest capacity. This is repeated for all the users as shown in algorithm 2. It is instructive to note that given the finite

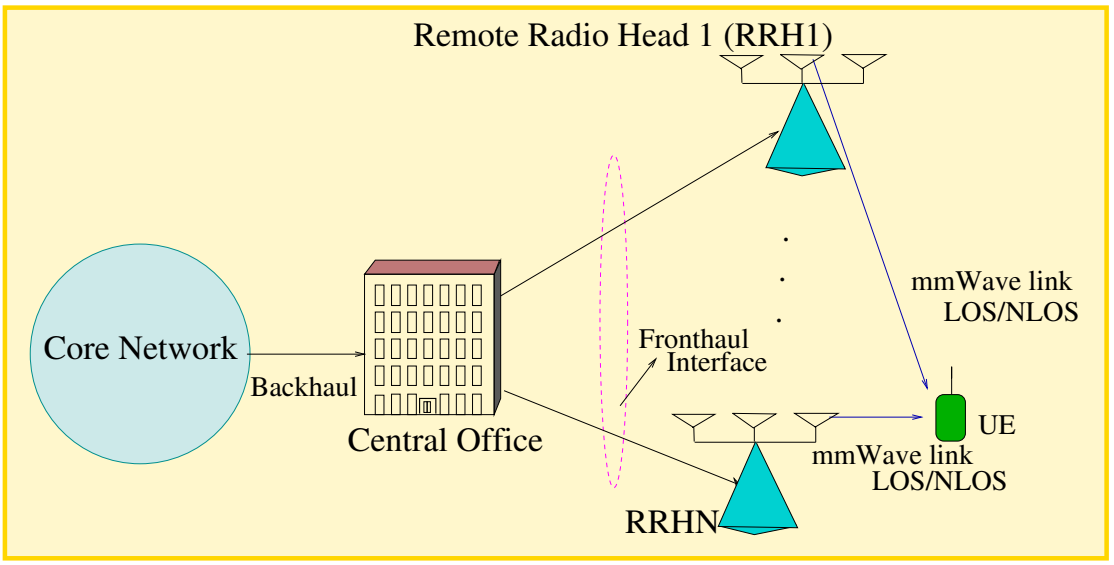


Figure 3.20: Single user served by two RRHs, where the channel between the user and each RRH can be either LOS or NLOS.

capacity constraints in the fronthaul, we consider the maximum user load that can be supported. In other words, we allow the maximum number of users to communicate depending on the fronthaul capacity. Furthermore, the digital TPC weight matrix $\mathbf{F}_{\mathrm{BB}}$ is applied at the CU, while the digitally precoded signals are phase shifted at the RRHs using $\mathbf{F}_{\mathrm{RF}}$.

In a scenario, where two or more users having different channel conditions (LOS/N-LOS) are associated with the same RRH, the transmitter activates fully-connected design with N_t AEs so as to satisfy all the users. For better illustration, let us consider an example shown in Fig. 3.21, where there are 2 RRHs as well as 2 users and let us consider the scenario where user 1 experiences LOS conditions associated with RRH1 and user 2 experiences NLOS propagation with both RRH1 and RRH2. In this setting, both RRHs will activate the fully-connected design. However, the RRH1 and the RRH2 transmit the signal to user 2 at half power. In other words, user 1 is served by RRH1 activating the fully-connected design to attain full BF gain, while user 2 is served both by RRH1 and RRH2, where each RRH uses fully-connected design and forms a 2-sub-array-connected design and each RRH transmits at half power. When combining the signal from the two RRHs, the receiver sees it as if it is transmitted from a transmitter associated with the 2-sub-array-connected design. Let us consider the example seen in Fig. 3.21, where the RRH1 employs the ABF $\mathbf{F}_{\mathrm{RF}}^{1}$ and the RRH2 employs $\mathbf{F}_{\mathrm{RF}}^{2}$, while the combined baseband digital TPC at the CU is $\mathbf{F}_{\mathbf{BB}}$. Then the signal vector received at UE2 in the downlink before combining is given by

$$\mathbf{y}_{\mathrm{UE2}} = \underbrace{\mathbf{H}_2^1\mathbf{F}_{\mathrm{RF}}^1\mathbf{F}_{\mathbf{BB}}\mathbf{x}_2 + \mathbf{H}_2^2\mathbf{F}_{\mathrm{RF}}^2\mathbf{F}_{\mathbf{BB}}\mathbf{x}_2}_{\text{desired signal}} + \underbrace{\mathbf{H}_2^1\mathbf{F}_{\mathrm{RF}}^1\mathbf{F}_{\mathbf{BB}}\mathbf{x}_1}_{\text{co-channel interference}} + \underbrace{\mathbf{n}}_{\text{noise}}. \tag{3.23}$$

Algorithm 3.2 Virtual Cell Formation for Single User

Categorize the RRHs into LOS and NLOS groups
Let LOS group has A RRHs and NLOS group has B DUs
For LOS group:
    Calculate the SNR for each link and sort the links in decreasing order as $\mathrm{SNR}_{\mathrm{L}_1} > \ldots > \mathrm{SNR}_{\mathrm{L}_\mathrm{A}}$
For NLOS group:
    Calculate the SNR for each link and sort the links in decreasing order as $\mathrm{SNR}_{\mathrm{NL}_1} > \ldots > \mathrm{SNR}_{\mathrm{NL}_\mathrm{B}}$
if $A > 0$ **then**
    if $\mathrm{SNR}_{\mathrm{NL}_1} - \mathrm{SNR}_{\mathrm{L}_1} <$ Threshold T **then**
        Form virtual cell by the RRH with associating $\mathrm{SNR}_{\mathrm{L}_1}$
    else
        Compute Capacities:
        C_1: Single RRH associated with $\mathrm{SNR}_{\mathrm{NL}_1}$
        C_2: Single RRH associated with $\mathrm{SNR}_{\mathrm{L}_1}$
        C_3: Dual-RRHs associated with $\mathrm{SNR}_{\mathrm{L}_1}$ and $\mathrm{SNR}_{\mathrm{NL}_1}$
        C_4: Dual-RRHs associated with $\mathrm{SNR}_{\mathrm{NL}_1}$ and $\mathrm{SNR}_{\mathrm{NL}_2}$
        C_5: Dual-RRHs associated with $\mathrm{SNR}_{\mathrm{L}_1}$ and SNR_{L_2}
        Form virtual cell with link that has highest capacity i.e., $\max\{C_1, C_2, C_3, C_4, C_5\}$
    end if
else
    Compute Capacities:
    C_1: Single RRH associated with $\mathrm{SNR}_{\mathrm{NL}_1}$
    C_2: Dual-RRHs associated with $\mathrm{SNR}_{\mathrm{NL}_1}$ and $\mathrm{SNR}_{\mathrm{NL}_2}$
    Form virtual cell with link that has highest capacity, i.e. $\max\{C_1, C_2\}$
end if

In this way, an RRH can serve multiple users. The multi-user cell formation technique is given in Algorithm 3.3. The exposition of Algorithm 3.3 is as follows: consider an N user system, and for each user the classification of the possible cases for virtual cell formation is carried out, as detailed for a single user. Furthermore, if the RRH has more than one user for cell formation, i.e. if $N_{\mathrm{RRH}} > 1$, then the RRH activates the fully-connected design for supporting multiple users. Otherwise, the RRH serves the users as required by the single user.

3.5 Conclusions

In this chapter, we proposed dual-function HBF concept and investigated the based baseband precoder relying on a codebook and the analog RF beamformer. Furthermore,

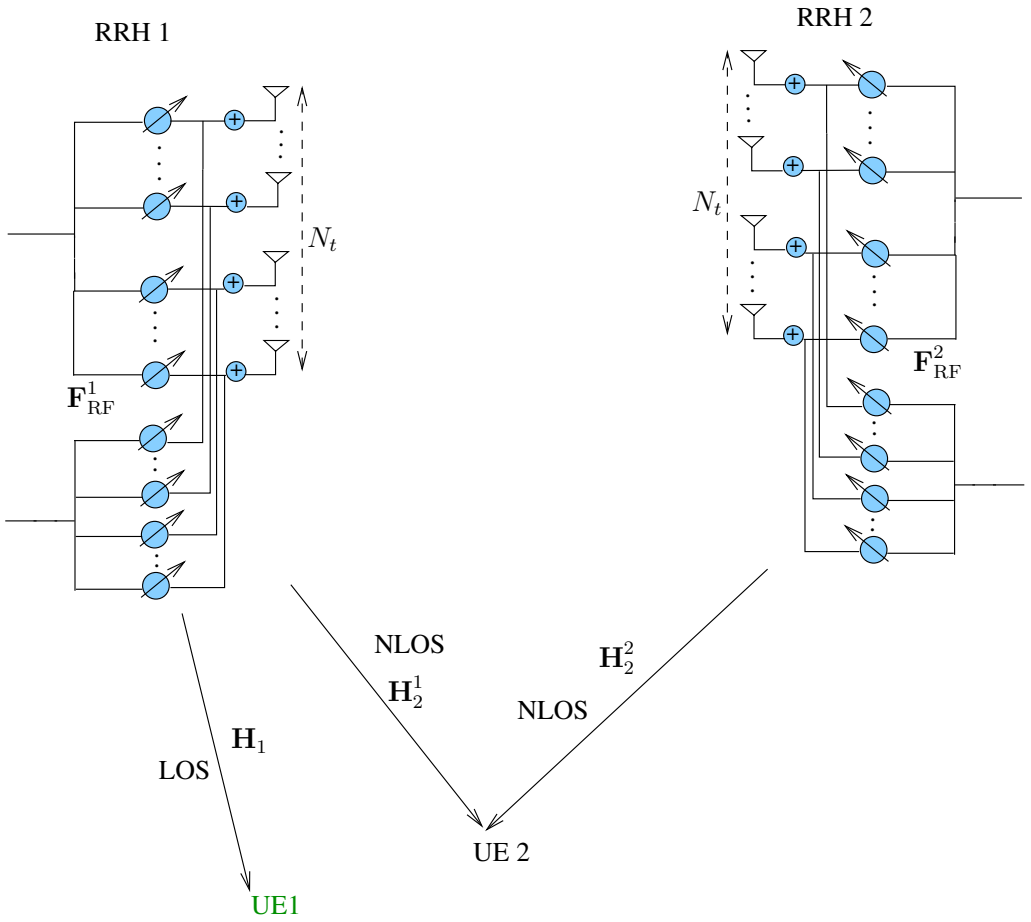


Figure 3.21: Example illustrating multiuser scenario, where a particular user is associated with 2 RRHs.

Algorithm 3.3 Virtual Cell Formation for Multi-User

For an N user system
for each user $n \in N$ **do**
    Repeat algorithm 1
end for
Let N_{RRH} be the number of users associated with a particular RRH
if $N_{\mathrm{RRH}} > 1$ **then**
    Form virtual cell with N_t AEs as fully-connected architecture or forming a 2-sub-array-connected architecture with another RRH
else
    Form virtual cell with the number of antennas required by the single user.
end if

we presented an adaptive array design, where the adaptation is carried out in two-stages, relying on the channel characteristics. By exploiting our adaptive array design, we proposed a beneficial virtual cell formation for a C-RAN.

More explicitly, in Sec. 3.2, a new HBF architecture is proposed by partitioning the AA into sub-arrays, where each emerging sub-array is separated by a sufficiently large

distance so that the correlation between the sub-arrays is minimized. Table 3.5 contrasts the proposed dual-function ASA architecture having 2-sub-arrays with both the conventional ASA and the fully-connected architectures. We summarized in Table 3.5 the beamforming, diversity and multiplexing gains of the different designs. While both the fully-connected and conventional ASA designs can provide strong beamforming gains at the cost of compromising the diversity/multiplexing gain, our proposed dual-function hybrid architecture provides both high beamforming and diversity gains. For this reason, our proposed design enhances the throughput of the system, as shown in Fig. 3.3. For example, observe from Table 3.5 that to achieve the rate of 5.5 bpcu, the fully-connected and the conventional ASA designs require SNRs of -9.2 dB and -6.1 dB, respectively, while our proposed design achieves this rate at a lower SNR of -10 dB.

However, we note that the diversity gain obtained reduces as the number of sub-arrays increases — this is due to the diminishing returns of the diversity gains. Furthermore, in our design, each sub-array may be fully-connected as shown in Fig. 3.2. Additionally, the total number of phase shifters in our design is $N_t N_t^{RF}/N_{sub}$, which is the same as that of the conventional ASA, while it is $N_t N_t^{RF}$ in the fully-connected design.

Table 3.5: Summary of the HBF transmission schemes relying on Fig. 3.3 for Figures 2.23, 2.24, and 3.2 of Chapter 2.

	Proposed ASA Dual-Function	Conventional ASA	Fully-Connected
Beamforming	Strong	Strong	Strong
Diversity	Strong	Weak	Weak
Multiplexing	Strong	Weak	Weak
SNR at rate 5.5 bpcu Fig. 3.3	-10 dB	-6.1 dB	-9.2 dB
Number of phase shifters	Fully-Connected ASA $N_t N_t^{RF}$ Sub-connected ASA $N_t N_{sub}^{RF}/N_{sub}$	$N_t N_t^{RF}/N_{sub}$	$N_t N_t^{RF}$

In Sec. 3.3, we investigated the performance of the codebook based digital precoder and analog RF beamformer. Table 3.6 shows a summary of Fig. 3.8 at a glance for a rate of 4 bpcu, where the DFT-MUB codebook performs similarly to the OMP and SVD precoder designs despite having a significantly reduced complexity. The rationale for choosing MUB as our codebook in the baseband is because the entries of the MUBs are the powers of i. More particularly, each codeword is constructed from a finite alphabet drawn from the set of points $\{-1, -i, i, 1\}$, which has the twin-fold benefits of reducing the power imbalance and eliminating the complex multiplications, as shown in Table 3.6. Furthermore, in Sec. 3.4, we have shown that when the channel is LOS nature, analog a pure beamformer, such as the DFT, based beamformer would capture the dominant

Table 3.6: Summary of the proposed DFT-MUB codebook for mmWave HBF systems shown in Figures 2.23, 2.24, and 3.2 of Chapter 2. The simulation parameters are listed in Table 3.1.

Schemes	DFT-MUB	OMP	SVD
SNR (at R=4 bpcu) for 32×16 MIMO	-10 dB	-11 dB	-12.25 dB
SNR (at R=4 bpcu) for 8×8 MIMO	-2.5 dB	-2.5 dB	-3.3 dB
Computations	$\mathcal{O}(N_s N_r^{\mathrm{RF}} N_t^{\mathrm{RF}})$	$\mathcal{O}(N_s N_r^{\mathrm{RF}} N_t^{\mathrm{RF}})+$ $\mathcal{O}(N_s N_r^{\mathrm{RF}}(N_t^{\mathrm{RF}}-1))$	$\mathcal{O}(N_s N_r^{\mathrm{RF}} N_t^{\mathrm{RF}})+$ $\mathcal{O}(N_s N_r^{\mathrm{RF}}(N_t^{\mathrm{RF}}-1))$

paths, followed by employing the digital TPC in the baseband, such as the MUB, albeit the latter becomes redundant. Additionally, in Fig. 3.17 we have shown that the fully-connected design performs better than the dual-function design having 2-sub-arrays, and otherwise having NLOS channel conditions. By exploiting this philosophy, an adaptive array design is proposed, which is shown in Fig. 3.14 and Table 3.4.

Next in Sec. 3.4.2, we extended our adaptive design in the context of C-RANs, where a virtual cell is formed with a fixed number of RRHs. Table 3.7 shows the summary of the possible combinations, in terms of the architecture, the beamforming technique, and the number of RRHs. To elaborate a little further, Table 3.7 classifies the RRH architecture into two categories: generic and 'best strategy'. By the generic architecture we mean that the RRH can be either fully-connected or sub-connected, regardless of the channel conditions. In other words, in the context of C-RAN, the user may be served by a single RRH having N_t AEs or by two RRHs having $N_t/2$ AEs, which acts as two sub-arrays, without considering the nature of the channel. However, we have shown that when the channel is of LOS nature, a single RRH with N_t AEs would provide better performance. Therefore, the best strategy may be classified into two types: if the channel is LOS nature, then 1 RRH having N_t AEs is associated with a user while employing analog only beamforming, and on the other hand, if the channel is of NLOS nature, 2 RRHs having $N_t/2$ AEs each are associated with a user while employing both digital and ABF.

Table 3.7: Summary of the dual-stage adaptation and virtual cell formation based on channel characteristics, which are shown in Figures 3.14 and 3.16.

	Architecture	Channel Conditions	Beamforming	Number of RRHs Association
Generic	Fully-Connected or Sub-Connected	LOS	Analog	1 RRH
		NLOS	Hybrid	2 RRH
Best Strategy	Fully-Connected	LOS	Analog	1 RRH
	Sub-Connected	NLOS	Hybrid	2 RRH

In the next chapter, we will exploit the beamforming for mitigating the self-interference observed in full duplex systems.

Chapter 4

Hybrid Beamforming for Full Duplex mmWave Systems

Communication is a two-way street. Make sure you are willing to listen as much as you are open to speak.

—Inspired by Warren Buffet

4.1 Introduction

It is generally assumed that bidirectional radio communication in the same frequency band is not possible because of the high SI that results from the transmitter's own transmission at the local receiver. However, the research conducted in [6, 7] proposed a solution to this challenge, where simultaneous transmission and reception in the same frequency were designed, which is termed as full-duplex (FD) wireless communication. Henceforth, the FD communication philosophy has attracted the attention of the wireless communication researchers, given its potential to double the spectral efficiency. However, the significant amount of SI at the local receiver tends to prevent FD communication from attaining satisfactory performance gains. Typically, the SI cancellation at the local receiver is carried out by a combination of passive and active methods [6, 8]. The passive methods aim for increasing the path loss of the interfering signal through antenna isolation so as to reduce the SI power [9]. By contrast, the active methods rely on the knowledge of the transmitted signal and they are generally carried out in three different approaches [10], namely RF cancellation, antenna cancellation, and digital cancellation. In the RF cancellation, the known transmitted signal at the local transmitter is used as a reference signal in the RF chain to reconstruct and then to subtract the SI. In the antenna cancellation, two replicas of the transmitted signal having opposite phases are generated using multiple transmit and receive antennas and the SI is removed by

adding the two opposite replicas [11]. In contrast to RF and antenna cancellation, digital cancellation is typically employed together with the RF or antenna cancellation to further suppress the SI in the baseband [11, 12]. However, the benefits obtained by cascading RF/digital cancellation to remove the SI may remain limited because of the distortions at various stages, such as the power amplifier non-linearity, I/Q imbalance, and phase noise [3]. This is especially more pronounced in mmWave communications because of the high non-linearities of the ADCs/DACs at mmWave frequencies [13]. Hence, considering these non-linearities caused by the ADCs/DACs, power amplifiers, I/Q imbalance and phase noise, the residual SI would be significantly high. In the state-of-the-art contributions on the FD, most of the work has been focused on SI cancellation for MIMO relay systems [14–18], while some other related work includes dynamic resource allocation for FD systems [19]. More recently, Wang *et al.* [34] has derived a closed-form expression for achievable rates for FD MIMO relay systems over Rician fading when linear receivers are employed. Considering the limitations of the input circuitry, Day *et al.* derived upper and lowers bounds on achievable sum rate for FD MIMO in [21]. Everett *et al.* [22] studied the performance analysis of passive SI suppression for FD, where the authors demonstrate that as high as 70 dB of SI suppression is possible in certain environments.

Although mmWave frequencies can provide large contiguous bandwidths, the efficient use of spectral resources is crucial to meet the escalating data rate demands. Hence, the efficient employment of FD communication at mmWave frequencies would further improve the attainable spectral efficiency. As far as FD communication is concerned, there is a paucity of literature on mmWave communications relying on FD techniques. Additionally, employing BF would further mitigate the SI. More recently, Xiao *et al.* [10] proposed a BF based FD for mmWave communication, where the authors designed transmit and receive BF weights to cancel the SI. In [111], Snow *et al.* demonstrated that SI cancellation of upto 40 dB using digital beamforming can be achieved relying on multiple coordinated transceivers. Despite the application of both the analog and digital cancellations, there is still a significant residual SI. Therefore, the SI cancellation relying on BF techniques is crucial at mmWave frequencies in addition to the conventional analog and digital SI suppression methods in the face of the strongly non-ideal nature of the ADCs/DACs and power amplifiers, which result in high residual SI [13]. Unfortunately, the designs in [10, 111] cannot be easily extended to HBF systems in mmWave communications, given that they are proposed for a single RF chain and an attempt to extend them to multiple RF chains becomes mathematical intractable because of the constraints imposed on the analog RF beamformer matrix. Hence, we propose a HBF design, where the RF beamformer and baseband precoder are obtained from the fully-digital precoder. Our contributions are summarized as follows.

1. *We consider FD communication at mmWave frequencies relying on HBF, where we aim for mitigating the SI by jointly designing the transmit and receive RF*

beamformer weights and the precoder as well as combiner matrices. To design the beamformer, precoder and combiner matrices, we first obtain the fully-digital solution, where we resort to an iterative algorithm relying on the idealized simplifying assumption of having perfect CSI. Then we derive the HBF solution from the obtained digital solution using least-squares approximation [42]. The proposed solution preserves the signal's dimensionality[1]*, while mitigating the SI. Based on our simulation results, we show that the proposed design is capable of reducing the SI by upto 30 dB and hence performing similarly to the interference-free FD system.*

2. *We present the mathematical proof for the convergence of the proposed iterative design, where the objective function is minimized in each iteration. Furthermore, we show that the value of the objective function reduces in every iteration and it is lower-bounded by zero.*

3. *We present quantitative comparisons of our proposed design through simulation results, where we demonstrate that our proposed design achieves better performance gains than EBF, especially when the SI power is high.*

4. *We then extend our design to the K-user frequency selective interference channels, where the precoder and combiner is designed to minimize both the SI and MI in mmWave systems using BF. In this design, we aim for preserving the signal dimension, while mitigating both the SI and MI.*

5. *Finally, we develop an iterative matrix decomposition for hybrid precoding aided OFDM systems, where the digital TPC weights are employed in the OFDM scheme's frequency-domain, while the analog RF beamformer weights are applied to the time-domain signal.*

The remainder of this chapter is organized as shown in the Fig. 4.1. In Sec. 4.2, we present our full-duplex HBF design for single user. More explicitly, in Sec. 4.2.1, we first propose digital precoder and combiner solution for the single user full-duplex system. Then subsequently in Sec. 4.2.3 we discuss the hybrid decomposition of the digital solution. In Sec. 4.3, we extend our proposed algorithm to multi-user full duplex system. Finally, our conclusions are presented in Sec. 4.4.

4.2 Single-User Full Duplex Transceiver

We first focus on a point-to-point communication of Fig. 4.2 in order to show the precoder and combiner design in details[2]. The transmitter and the receiver of node j are equipped

[1] In contrast to spatial suppression [43], where the signal is projected into the null space of the interference, which may result in a reduced signal dimension, our design preserves the dimensionality of the signal subspace where the rank of the received signal is equal to the number of signal streams transmitted.

[2] In the next section, we will be showing how the system would work with multiple nodes (users > 2).

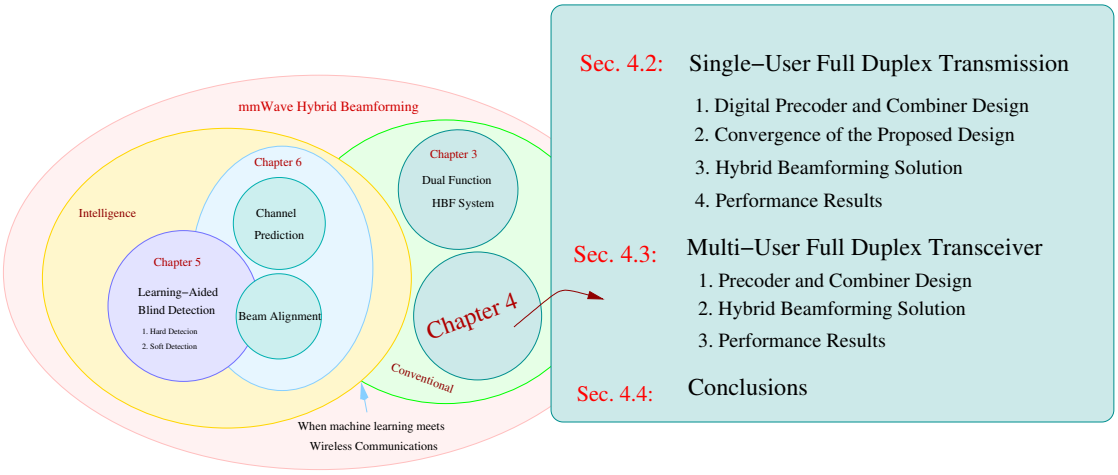


Figure 4.1: The organization of this chapter.

with $N_t^{(j)}$ and $N_r^{(j)}$ antennas as well as by $N_{r_{(j)}}^{\mathrm{RF}}$ and $N_{t_{(j)}}^{\mathrm{RF}}$ RF chains, respectively. The transmitter of node j maps the bits to QAM symbols and employs both RF BF and precoding using the matrices $\mathbf{F}_{\mathrm{RF}}^{(j)}$ and $\mathbf{F}_{\mathrm{BB}}^{(j)}$ of size $N_t^{(j)} \times N_{t_{(j)}}^{\mathrm{RF}}$ and $N_{t_{(j)}}^{\mathrm{RF}} \times N_s^{(j)}$, respectively. The beamformed symbols are transmitted over the mmWave channel. Then, the receiver at the node i combines the signal using the RF combiner $\mathbf{W}_{\mathrm{RF}}^{(i)}$ and baseband combiner $\mathbf{W}_{\mathrm{BB}}^{(i)}$ of sizes $N_r^{(i)} \times N_{r_{(i)}}^{\mathrm{RF}}$ and $N_{r_{(i)}}^{\mathrm{RF}} \times N_s^{(j)}$, respectively. Then, the received signal vector $\mathbf{y}^{(i)}$ at the receiver of node i after RF and baseband combining is given by[3]

$$\mathbf{y}^{(i)} = \underbrace{\mathbf{W}_{\mathrm{BB}}^{H(i)}\mathbf{W}_{\mathrm{RF}}^{H(i)}\mathbf{H}_{\mathrm{ji}}\mathbf{F}_{\mathrm{RF}}^{(j)}\mathbf{F}_{\mathrm{BB}}^{(j)}\mathbf{s}^{(j)}}_{\text{desired signal}} + \underbrace{\mathbf{W}_{\mathrm{BB}}^{H(i)}\mathbf{W}_{\mathrm{RF}}^{H(i)}\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}_{\mathrm{RF}}^{(i)}\mathbf{F}_{\mathrm{BB}}^{(i)}\mathbf{s}^{(i)}}_{\text{SI}} + \underbrace{\mathbf{W}_{\mathrm{BB}}^{H(i)}\mathbf{W}_{\mathrm{RF}}^{H(i)}\mathbf{n}^{(i)}}_{\text{noise}}, \tag{4.1}$$

where $\mathbf{n}$ is the noise vector of dimension $N_r^{(i)}$ whose elements are i.i.d. with distribution $\mathcal{CN}(0, \sigma^2)$, $\mathbf{s}^{(j)}$ is the signal vector of dimension $N_s^{(j)}$, $\mathbf{H}_{ji}$ is the mmWave channel spanning from the transmitter in the j^{th} node to the receiver in the i^{th} node, whenever $i \neq j$ and it is given by [190]

$$\mathbf{H}_{ji} = \sqrt{\frac{N_r^{(i)}N_t^{(j)}}{N_c N_{\mathrm{ray}}}} \sum_{n_c=1}^{N_c}\sum_{n_{\mathrm{p}}=1}^{N_{\mathrm{p}}} \alpha_{ji}^{(n_{\mathrm{p}},n_c)}\mathbf{a}_r(\phi_{ji}^{n_{\mathrm{p}},n_c})\mathbf{a}_t^T(\theta_{ji}^{n_{\mathrm{p}},n_c}). \tag{4.2}$$

To elaborate further, $\mathbf{H}_{ji}$ is the statistical channel matrix of size $N_r^{(i)} \times N_t^{(j)}$ so that $\mathbb{E}[\|\mathbf{H}_{\mathrm{ji}}\|_F^2] = N_t^{(j)}N_r^{(i)}$, while $\alpha_{n_c}^{n_{\mathrm{p}}} \sim \mathcal{CN}(0,1)$ is a complex-valued Gaussian random variable, whose amplitude and phase are Rayleigh and uniformly distributed, respectively. We note that the SI term is captured in the second term of (4.1). For a uniform linear

[3] The signal model considered is valid after both analog and digital SI cancellation.

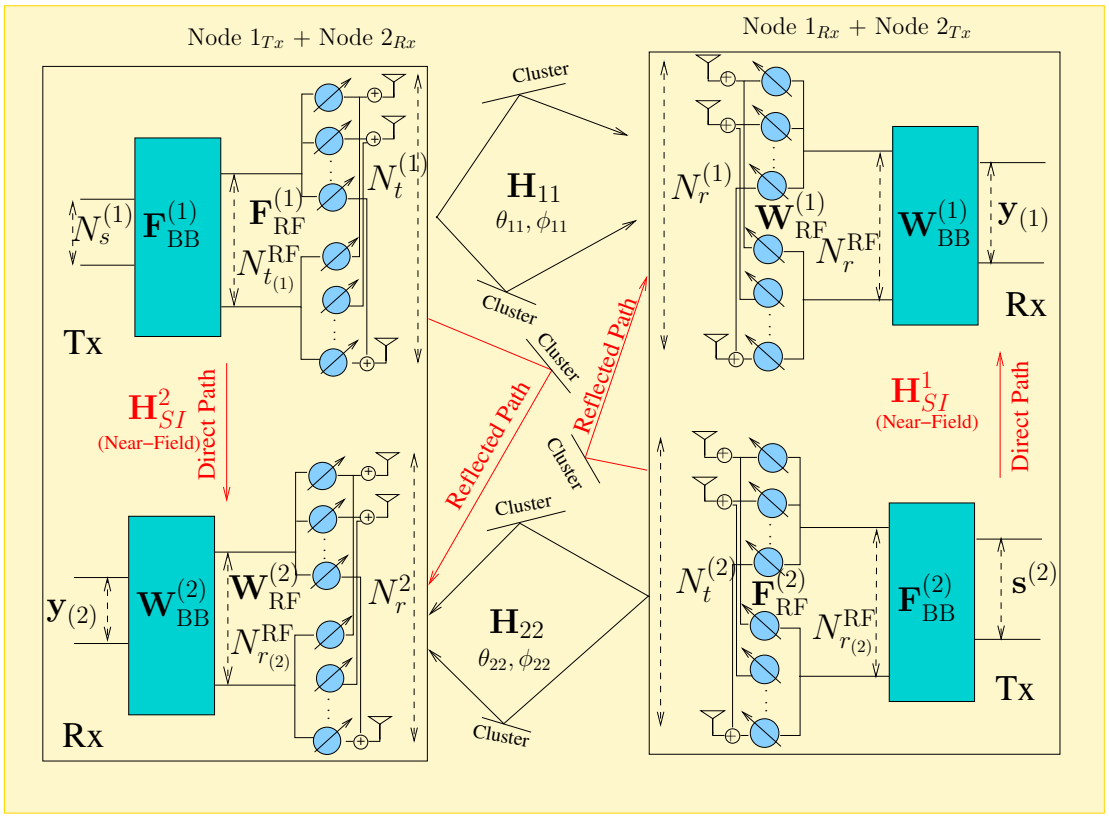


Figure 4.2: *Full-duplex with HBF in mmWave communications.*

array (ULA) having N_r and N_t AEs the response vectors $\mathbf{a}_r$ and $\mathbf{a}_t$ are expressed as:

$$\mathbf{a}_r(\phi_r) = [1 \; e^{j\frac{2\pi}{\lambda}d\cos(\phi_r)} \ldots e^{j\frac{2\pi}{\lambda}(N_r-1)d\cos(\phi_r)}]^T, \tag{4.3}$$

$$\mathbf{a}_t(\theta_t) = [1 \; e^{j\frac{2\pi}{\lambda}d\cos(\phi_t)} \ldots e^{j\frac{2\pi}{\lambda}(N_t-1)d\cos(\phi_t)}]^T. \tag{4.4}$$

Finally, θ_t and ϕ_r, N_c and N_{p} are the angles of departure (AOD) and arrival (AOA), and the number of clusters and rays, respectively.

Note that the node i is operating in FD mode and hence can transmit at the same time, while receiving the signal from node j. Hence, the second term in (4.1) is the interference due to its own transmission. The matrix $\mathbf{H}_{\text{SI}}^{(i)}$ is the near-field channel of node i and is given as

$$\mathbf{H}_{\text{SI}}^{(i)} = \sqrt{\frac{\kappa}{\kappa+1}}\mathbf{H}_{\text{ii}} + \sqrt{\frac{1}{\kappa+1}}\mathbf{H}_{rp}, \tag{4.5}$$

where κ is the Rician factor, $\mathbf{H}_{\text{ii}}$ is the near-field channel's LOS component (direct path as shown in Fig. 4.2) and $\mathbf{H}_{rp}$ is the reflected path of the SI. Fig. 4.3 shows the arrangement of the transmit and receive AAs, where the arrays are separated by a distance D at an angle Θ. To elaborate further, the distance D of Fig. 4.3 would be on the order of a few wavelengths when the transmitter and the local receiver AAs share the same AA, while it is zero if they share the same array [10]. Hence, the planar wave assumption does not

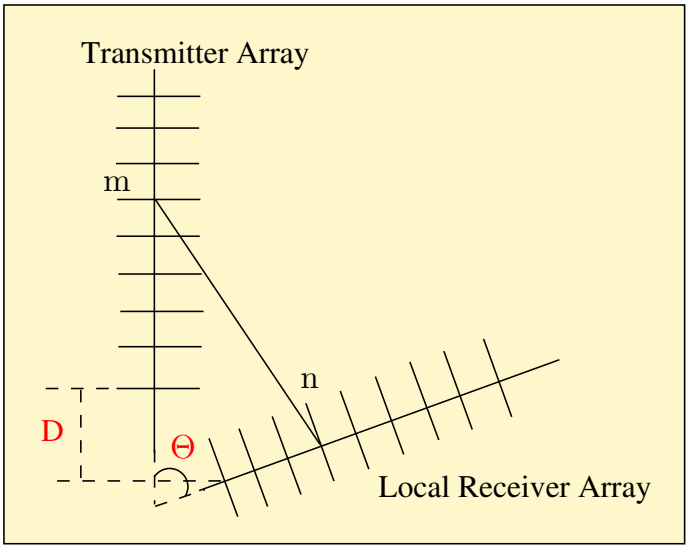


Figure 4.3: *The configuration of the transmitter and the local receiver AAs of a node.*

hold, as it violates the far-field condition, i.e. $R_o \geq 2\mathcal{D}^2/\lambda$ where $\mathcal{D}$ is the diameter of the antenna aperture as shown in Fig. 4.4. Typically, in the far-field as shown in Fig. 4.4

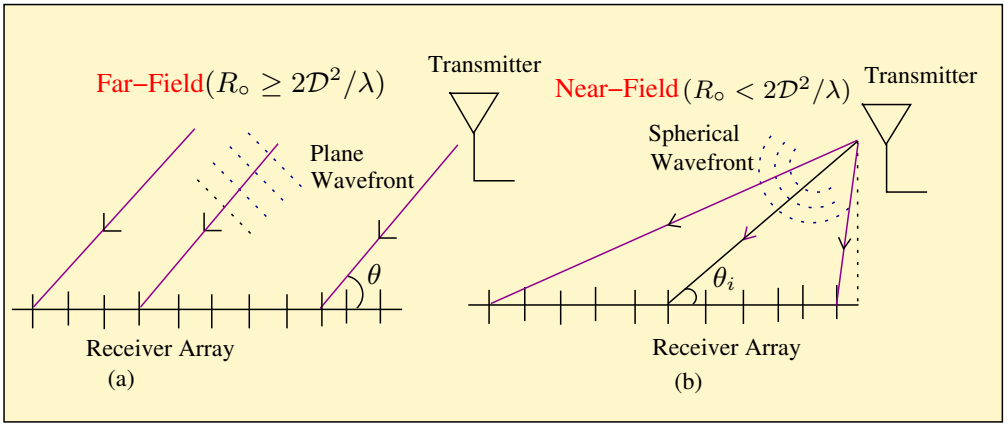


Figure 4.4: *Illustration of far-field and near-field models. (a) In the far-field scenario* $(R_o \geq 2\mathcal{D}^2/\lambda)$*, the signal is assumed to strike the array as a planar wave. (b) the signal is assumed to be a spherical wavefront as* $(R_o < 2\mathcal{D}^2/\lambda)$.

(a), the underpinning assumption is that the signal impinges on the AA as a planar wave, which means that all the AEs observe the same path loss in the received signal and the phase difference between the adjacent AEs depends only on the angle of arrival and AEs spacing [191]. By contrast, in the FD scenario, since the transmitter and the local receiver are closely placed, these assumptions do not hold for the SI channel, as it is a near-field channel, which is shown in Fig. 4.4 (b). Therefore, a more realistic channel model, which is the spherical wave propagation model is considered for the near-field LOS channel matrix $\mathbf{H}_{\text{ii}}^{(i)}$ [23] and the channel coefficient of the n^{th} row and m^{th} column

entry of $\mathbf{H}_{\text{ii}}^{(i)}$ is given by [10, 23]

$$[\mathbf{H}_{\text{ii}}]_{mn}^{(i)} = \frac{\rho}{r_{mn}} \exp(-j2\pi \frac{r_{mn}}{\lambda}), \tag{4.6}$$

where ρ is the power normalization constant invoked for ensuring $\mathbb{E}[\|\mathbf{H}_{\text{ii}}^{(i)}\|_F^2] = N_t^{(i)} N_r^{(i)}$, and r_{mn} is the distance between the m^{th} element of the transmitter and n^{th} element of the the receiver,

$$r_{mn} = \sqrt{a^2 + b^2 - 2ab\cos(\Theta)}, \tag{4.7}$$

where

$$a = \left(\frac{D}{\tan(\Theta)} + (n-1)\frac{\lambda}{2}\right),$$
$$b = \left(\frac{D}{\sin(\Theta)} + (m-1)\frac{\lambda}{2}\right),$$

where λ, D, and Θ are the wavelength, the distance between the first AEs of the transmitter and receiver arrays, and the angle between the arrays, respectively. The channel matrix $\mathbf{H}_{rp}$ of the reflected path is modeled as in (5.14) with the appropriate angle of departure and arrival. It is instructive to note that the near-field channel model considered in (4.6) is a simplified one as in [192]. It does not account for the mutual coupling effects between the AEs and the signal reflections.

For the system model in (4.1), the achievable sum rate of the system is given in (4.8), where the term $\mathbf{Z}_i^{-1}$ captures the SI power due to the transmitter of node i plus noise power at the receiver of node i. Furthermore, $P^{(j)}$ and $P_{\text{SI}}^{(i)}$ denotes the transmit power of node j and the SI power of node i, respectively.

$$C = \sum_j \log_2 \det\left(\mathbf{I}_{N_s} + \frac{P^{(j)}}{N_\circ}\mathbf{Z}_i^{-1}(\mathbf{F}_{\text{RF}}^{(j)}\mathbf{F}_{\text{BB}}^{(j)})^H \mathbf{H}_{ji}^H \mathbf{W}_{\text{BB}}^{(i)}\mathbf{W}_{\text{RF}}^{(i)}\mathbf{W}_{\text{BB}}^{H(i)}\mathbf{W}_{\text{RF}}^{H(i)}\mathbf{H}_{ji}\mathbf{F}_{\text{RF}}^{(j)}\mathbf{F}_{\text{BB}}^{(j)}\right), \tag{4.8}$$
$$\forall j = 1,2;\ i = 1,2;\ j \neq i,$$

where

$$\begin{aligned}\mathbf{Z}_i =& P_{\text{SI}}^{(i)}\mathbf{W}_{\text{BB}}^{H(i)}\mathbf{W}_{\text{RF}}^{H(i)}\mathbf{H}_{\text{SI}}^{i}\mathbf{F}_{\text{RF}}^{(i)}\mathbf{F}_{\text{BB}}^{(i)}\left(\mathbf{W}_{\text{BB}}^{H(i)}\mathbf{W}_{\text{RF}}^{H(i)}\mathbf{H}_{\text{SI}}^{i}\mathbf{F}_{\text{RF}}^{(i)}\mathbf{F}_{\text{BB}}^{(i)}\right)^H \\ &+ \sigma^2\mathbf{W}_{\text{BB}}^{H(i)}\mathbf{W}_{\text{RF}}^{H(i)}\mathbf{W}_{\text{BB}}^{(i)}\mathbf{W}_{\text{RF}}^{(i)}\end{aligned}$$

Remark 4.1: The employment of FD HBF is more plausible in a downlink scenario, where the BS is receiving a signal from one user at an angle ϕ_1 and simultaneously transmitting

to another user at an angle ϕ_2. In other words, the transmitter transmits its own signal in one beam and receives the signal from a user in a different beam, hence mitigating the SI.

In the next section, we design the beamformer, precoder and combiner matrices to suppress the interference, while preserving the signal dimensions.

4.2.1 Digital Precoder and Combiner Design

In this section, we present our proposed design, where we first conceive the local optimal digital precoder and combiner matrices for the system model of (4.1). Then, we split the digital precoder obtained into an analog RF beamformer matrix $\mathbf{F}_{\mathrm{RF}}^{(j)}$ and a digital precoder matrix $\mathbf{F}_{\mathrm{BB}}^{(j)}$. Consider the system model in (4.1), where only digital precoding and combining are employed, hence we set $\mathbf{F}^{(j)} = \mathbf{F}_{\mathrm{RF}}^{(j)}\mathbf{F}_{\mathrm{BB}}^{(j)}$ and $\mathbf{W}^{H^{(i)}} = \mathbf{W}_{\mathrm{BB}}^{H^{(i)}}\mathbf{W}_{\mathrm{RF}}^{H^{(i)}}$. Then, the system model in (4.1) reduces to

$$\mathbf{y}^{(i)} = \underbrace{\mathbf{W}^{H^{(i)}}\mathbf{H}_{ji}\mathbf{F}^{(j)}\mathbf{s}^{(j)}}_{\text{desired signal}} + \underbrace{\mathbf{W}^{H^{(i)}}\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(i)}\mathbf{s}^{(i)}}_{\text{SI}} + \underbrace{\mathbf{W}^{H^{(i)}}\mathbf{n}^{(i)}}_{\text{noise}}. \tag{4.9}$$

In this section, we aim for designing the precoder matrix $\mathbf{F}$ for a fixed combiner matrix $\mathbf{W}$ and then design the combiner matrix $\mathbf{W}$ for a fixed precoder matrix $\mathbf{F}$. We carry out this process of optimization in iterations so that the effect of SI is minimized. The SI plus noise at the receiver node i for fixed precoder matrix $\mathbf{F}^{(i)}$ is

$$\mathbf{SI}^{(i)} = \mathbf{W}^{H^{(i)}}\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(i)}\mathbf{s}^{(i)} + \mathbf{W}^{H^{(i)}}\mathbf{n}^{(i)}, \tag{4.10}$$

where the co-variance matrix $\mathbf{Q}^{(i)}$ of the interference plus noise at the receiver node i is given by

$$\mathbf{Q}^{(i)} = \mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\mathbf{R}^{(i)}\mathbf{W}^{(i)}\right), \tag{4.11}$$

with $\mathbf{R}^{(i)}$ given by

$$\mathbf{R}^{(i)} = P_{\mathrm{SI}}^{(i)}\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(i)}(\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(i)})^H + \mathbf{I}_n, \tag{4.12}$$

and $\mathbf{I}_n$ is the noise power. Our objective is to design $\mathbf{W}$ and $\mathbf{F}$ so that we minimize the SI power $\mathbf{Q}^{(i)}$ at the receiver and simultaneously preserve the dimension of the signal space, i.e. $\mathrm{rank}\left(\mathbf{W}^{H^{(i)}}\mathbf{H}_{ji}\mathbf{F}^{(j)}\right) = N_s^{(j)}$.

Proposition 4.1: For the equality constrained optimization problem of

$$\begin{aligned} &\min_{\mathbf{W}^{(i)}} \mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\mathbf{R}^{(i)}\mathbf{W}^{(i)}\right) \\ &\quad s.t.\ \mathbf{W}^{H^{(i)}}\mathbf{H}_{ji}\mathbf{F}^{(j)} = \alpha\mathbf{I}_{N_s} \end{aligned} \tag{4.13}$$

where $\mathbf{R}^{(i)}$ is a positive definite matrix ($\mathbf{R}^{(i)} \succ 0$), the local optimal solution is given by

$$\mathbf{W}_{\text{opt}}^{(i)} = \alpha \mathbf{R}^{(i)^{-1}} \mathbf{H}_{ji} \mathbf{F}^{(j)} \left((\mathbf{H}_{ji} \mathbf{F}^{(j)})^H \mathbf{R}^{(i)^{-1}} (\mathbf{H}_{ji} \mathbf{F}^{(j)}) \right)^{-1} \tag{4.14}$$

and α is the normalization constant expressed as

$$\alpha = \frac{1}{\sqrt{\text{Tr}\left(\mathbf{W}_{\text{opt}}^{H(i)} \mathbf{W}_{\text{opt}}^{(i)} \right)}}. \tag{4.15}$$

Proof: The objective is to minimize the equality constrained problem in (4.13), while maintaining the signal dimensions. To solve the above optimization problem, we begin by forming the Lagrangian function given as (4.16).

$$\mathcal{L}(\mathbf{W}^{(i)}, z) = \left(\mathbf{W}^{H(i)} \mathbf{R}^{(i)} \mathbf{W}^{(i)} \right) + z \left(\mathbf{W}^{H(i)} \mathbf{H}_{ji} \mathbf{F}^{(j)} - \mathbf{I}_{N_s} \right) \tag{4.16}$$

Then, the Lagrangian conditions for this problem are

$$\nabla_{\mathbf{W}_{\text{opt}}^{H(i)}} \mathcal{L} = 0 \tag{4.17}$$

$$z^* \left(\mathbf{W}_{\text{opt}}^{H(i)} \mathbf{H}_{ji} \mathbf{F}^{(j)} - \alpha \mathbf{I}_{N_s} \right) = 0 \tag{4.18}$$

Explicitly (4.17) can be written as,

$$\begin{aligned} &\nabla_{\mathbf{W}_{\text{opt}}^{H(i)}} \text{Tr}\left(\mathbf{W}_{\text{opt}}^{H(i)} \mathbf{R}^{(i)} \mathbf{W}_{\text{opt}}^{(i)} \right) \\ &+ z^* \nabla_{\mathbf{W}_{\text{opt}}^{H(i)}} \left(\mathbf{W}_{\text{opt}}^{H(i)} \mathbf{H}_{ji} \mathbf{F}^{(j)} - \mathbf{I}_{N_s} \right) = 0 \end{aligned} \tag{4.19}$$

where ∇ is the gradient operation and z^* is the Lagrangian multiplier.

By taking the derivative with the respect to $\mathbf{W}^{H(i)}$ in equation (4.19), we obtain

$$\mathbf{R}^{(i)} \mathbf{W}_{\text{opt}}^{(i)} + z \mathbf{H}_{ji} \mathbf{F}^{(j)} = 0 \tag{4.20}$$

$$\mathbf{W}_{\text{opt}}^{(i)} = -\mathbf{R}^{(i)^{-1}} \mathbf{H}_{ji} \mathbf{F}^{(j)} z. \tag{4.21}$$

Upon substituting $\mathbf{W}_{\text{opt}}^{(i)}$ in (4.18), we get

$$\left(-\mathbf{R}^{(i)^{-1}} \mathbf{H}_{ji} \mathbf{F}^{(j)} z \right)^H \mathbf{H}_{ji} \mathbf{F}^{(j)} = \alpha \mathbf{I}_{N_s} \tag{4.22}$$

$$z = -\alpha \left((\mathbf{H}_{ji} \mathbf{F}^{(j)})^H \mathbf{R}^{(i)^{-1}} \mathbf{H}_{ji} \mathbf{F}^{(j)} \right)^{-1}. \tag{4.23}$$

Hence,

$$\mathbf{W}_{\text{opt}}^{(i)} = \alpha \mathbf{R}^{(i)^{-1}} \mathbf{H}_{ji} \mathbf{F}^{(j)} \left((\mathbf{H}_{ji} \mathbf{F}^{(j)})^H \mathbf{R}^{(i)^{-1}} (\mathbf{H}_{ji} \mathbf{F}^{(j)}) \right)^{-1} \quad \square \tag{4.24}$$

As seen from (4.14), $\mathbf{W}_{\text{opt}}^{(i)}$ is a function of $\mathbf{F}^{(j)}$, while our objective is to design $\mathbf{W}$ and $\mathbf{F}$ jointly. Hence, we opt for an iterative design, where we initially set $\mathbf{F}^{(j)}$ to the right singular vector of the channel matrix $\mathbf{H}_{ji}$ to get $\mathbf{W}_{\text{opt}}^{(i)}$ and then having obtained the combiner matrix $\mathbf{W}^{(i)}$, we now proceed to design the precoder matrix $\mathbf{F}^{(j)}$ by considering the interference caused by the transmitter at the node j to the local receiver.

The SI power imposed by the transmitter of node j on the receiver of node j is given by

$$\mathbf{J}^{(j)} = \text{Tr}\left(\mathbf{F}^{H^{(j)}}\mathbf{S}^{(j)}\mathbf{F}^{(j)}\right), \tag{4.25}$$

where $\mathbf{S}^{(j)}$ is given by[4]

$$\mathbf{S}^{(j)} = \left(P_{\text{SI}}^{(j)}\mathbf{W}^{H^{(j)}}\mathbf{H}_{\text{SI}}^{(j)}\right)^{H}\left(\mathbf{W}^{H^{(j)}}\mathbf{H}_{\text{SI}}^{(j)}\right) + \mathbf{I}. \tag{4.26}$$

Similarly, our objective here is to minimize the SI power $\mathbf{S}^{(j)}$ caused by the transmitter.

Thus, the constrained optimization problem is given by

$$\begin{aligned} &\min_{\mathbf{F}^{(j)}} \ \text{Tr}\left(\mathbf{F}^{H^{(j)}}\mathbf{S}^{(j)}\mathbf{F}^{(j)}\right) \\ &s.t. \ \mathbf{W}^{H^{(i)}}\mathbf{H}_{ji}\mathbf{F}^{(j)} = \beta\mathbf{I}_{N_s}. \end{aligned} \tag{4.27}$$

By using Proposition 4.1, we obtain the local optimal solution as

$$\mathbf{F}_{\text{opt}}^{(j)} = \beta\mathbf{S}^{(j)^{-1}}\mathbf{H}_{ji}^{H}\mathbf{W}^{H^{(j)}}\left(\mathbf{W}^{H^{(j)}}\mathbf{H}_{ji}\mathbf{S}^{(j)^{-1}}\left(\mathbf{W}^{H^{(j)}}\mathbf{H}_{ji}\right)^{H}\right)^{-1} \tag{4.28}$$

and β is the normalization constant expressed as

$$\beta = \frac{1}{\sqrt{\text{Tr}\left(\mathbf{F}_{\text{opt}}^{H(j)}\mathbf{F}_{\text{opt}}^{(j)}\right)}}. \tag{4.29}$$

This is an iterative process and it is continued until the convergence of the objective functions (4.13) and (4.27) is reached. The pseudo code for the algorithm is presented in Algorithm 4.4. The proof of convergence is presented in the next section.

In this chapter, we have assumed that both the nodes 1 and 2 can simultaneously transmit and receive the data. However, in practice, given the power constraints, it is reasonable to assume that only the BS is equipped with a complex FD transceiver while the mobile station is BS (HD). In this scenario, the mobile station does not experience any SI,

[4]The term $\mathbf{I}$ is added in (4.26) to ensure that the matrix $\mathbf{S}^{(j)}$ is invertible. The physical meaning of this is that it minimizes the norm of the precoder matrix. However, its effect is counteracted by β, which ensures that the transmit power is constant, i.e. $||\mathbf{F}^{(j)}\mathbf{s}^{(j)}||_{\text{F}}^2 = N_s$. Therefore, the identity matrix role in (19) is to ensure that $\mathbf{S}^{(j)}$ is invertible, but it has no influence on the design regardless of the SNR/INR. It can be removed if $\mathbf{S}^{(j)}$ is invertible in practice.

Algorithm 4.4 Proposed Design for Full Duplex HBF system

1: Initialize $\mathbf{F}^{(j)}$ to right singular vectors of the channel matrix $\mathbf{H}_{ji}, j = 1, 2; i = 1, 2; j \neq i$;
2: Compute the matrices $\mathbf{R}^{(i)}, i = 1, 2$;
3: Obtain $\mathbf{W}^{(i)} =$
$$\mathbf{R}^{(i)^{-1}}\mathbf{H}_{ji}\mathbf{F}^{(j)}\left(\left(\mathbf{H}_{ji}\mathbf{F}^{(j)}\right)^{H}\mathbf{R}^{(i)^{-1}}\left(\mathbf{H}_{ji}\mathbf{F}^{(j)}\right)\right)^{-1};$$
4: Then normalize $\mathbf{W}^{(i)}$ by multiplication factor $\frac{1}{\sqrt{\mathrm{Tr}\left(\mathbf{W}_{\mathrm{opt}}^{H(i)}\mathbf{W}_{\mathrm{opt}}^{(i)}\right)}}$;
5: Compute the matrices $\mathbf{S}^{(j)}, j = 1, 2$;
6: Obtain $\mathbf{F}^{(j)} =$
$$\mathbf{H}_{ji}^{H}\mathbf{W}^{H(j)}\left(\mathbf{W}^{H(j)}\mathbf{H}_{ji}\mathbf{S}^{(j)^{-1}}\left(\mathbf{W}^{H(j)}\mathbf{H}_{ji}\right)^{H}\right)^{-1};$$
7: Then normalize $\mathbf{F}^{(j)}$ by multiplication factor $\frac{1}{\sqrt{\mathrm{Tr}\left(\mathbf{F}_{\mathrm{opt}}^{H(j)}\mathbf{F}_{\mathrm{opt}}^{(j)}\right)}}$;
8: Repeat steps from 2 until convergence of functions (4.13) and (4.27) is reached.

hence the precoder for the mobile station can be designed by setting $\mathbf{H}_{\mathrm{SI}} = 0$ in (4.26). However, the precoder and combiner weights for the BS can be obtained from Algorithm 1.

Remark 4.2: It is worth observing from (4.13) and (4.27) that the objective functions not only minimize the interference leakage into the desired receiver but also preserve the degrees of freedom for the desired signal, which is formulated in the constraint. The physical significance of the constraints in both the objective functions of (15) and (20) is ensuring that the signal does not suffer from distortions. The proposed objective functions can also be interpreted as maximizing the SINR, where the numerator (signal fidelity) is kept unaltered while reducing the denominator (interference) [193].

4.2.2 Convergence of the Proposed Design

In this subsection, we present the convergence of the objective function in (4.13) and (4.27). We show that the value of the objective function decreases in every iteration and converges to its local optimum in a few iterations, which makes it more computationally efficient in practical system designs. Here, we provide the proof for the convergence of the proposed algorithm. The total SI plus noise power at all receivers is given by

$$\mathbf{Q} = \sum_{i=1}^{2}\mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\mathbf{R}^{(i)}\mathbf{W}^{(i)}\right) \tag{4.30}$$

$$= \sum_{i=1}^{2}\mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\left(\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(j)}(\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(j)})^{H} + \mathbf{I}_{n}\right)\mathbf{W}^{(i)}\right) \tag{4.31}$$

$$= \sum_{i=1}^{2}\mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(j)}(\mathbf{H}_{\mathrm{SI}}^{(i)}\mathbf{F}^{(j)})^{H}\mathbf{W}^{(i)}\right) + 1. \tag{4.32}$$

Similarly, the SI plus noise power at node j is given by

$$\mathbf{J} = \sum_{j=1}^{2} \mathrm{Tr}\left(\mathbf{F}^{H^{(j)}}\mathbf{S}^{(j)}\mathbf{F}^{(j)}\right), \tag{4.33}$$

$$= \sum_{j=1}^{2} \mathrm{Tr}\left(\mathbf{F}^{H^{(j)}}\left((\mathbf{W}^{H^{(j)}}\mathbf{H}_{\mathrm{SI}}^{(j)})^{H}(\mathbf{W}^{H^{(j)}}\mathbf{H}_{\mathrm{SI}}^{(j)}) + \mathbf{I}_n\right)\mathbf{F}^{(j)}\right) \tag{4.34}$$

$$= \sum_{j=1}^{2} \mathrm{Tr}\left((\mathbf{W}^{H^{(j)}}\mathbf{H}_{\mathrm{SI}}^{(j)}\mathbf{F}^{(j)})^{H}(\mathbf{W}^{H^{(j)}}\mathbf{H}_{\mathrm{SI}}^{(j)}\mathbf{F}^{(j)})\right) + 1. \tag{4.35}$$

It is instructive to note that the objective functions in (4.30) and (4.33) are identical. Hence, $\mathbf{W}$ which minimizes (4.30) also minimizes (4.33) and similarly $\mathbf{F}$ that minimizes (4.33) also minimizes (4.30). Furthermore, the local optimal solutions of the objective functions (4.13) and (4.27) are given by Proposition 1, and they are sure to converge to the locally optimal solution as it is guaranteed by the Algorithm 1 of the proposed design and the value of the objective function decreases in every iteration and it is lower bounded by zero.

Fig. 4.5 shows the convergence of the objective function (4.13) for the proposed design. It can be seen from the figure that the objective function converges to a local optimum solution in just five iterations. Furthermore, since objective functions for designing precoder and combiner are the same, as proved in (4.30) and (4.33), Fig. 4.5 holds for (4.27).

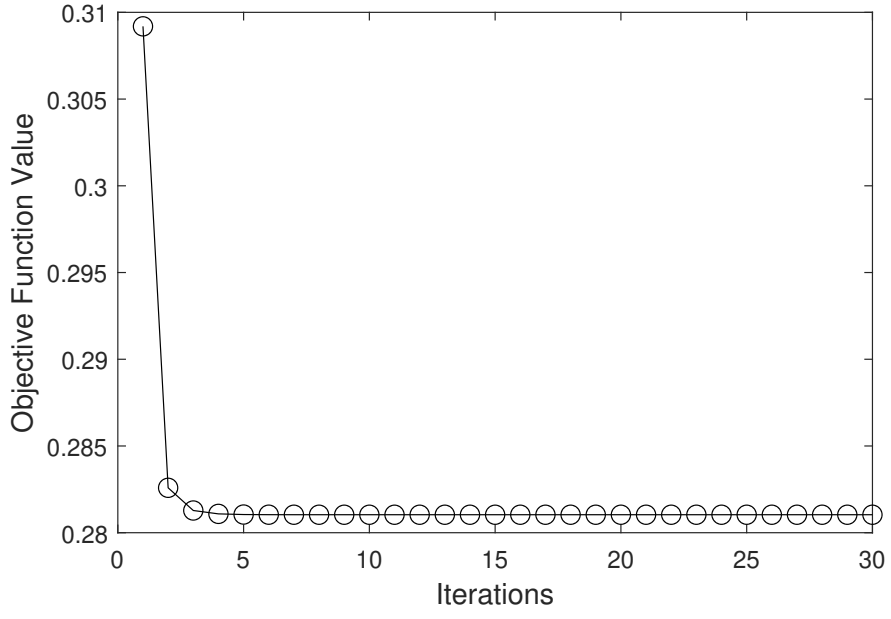


Figure 4.5: *Convergence of the objective function in (4.13) for the proposed design, SIR= 0 dB and SNR= -10 dB.*

4.2.3 Hybrid Beamforming Solution

In the previous sections, we presented the digital precoder and combiner design so as to minimize the SI. However, at the time of writing, the obtained resultant digital precoder and combiner design cannot be readily implemented in practice, since dedicating individual DACs/ADCs (RF chains) to each element of a large AA would impose a potentially excessive hardware complexity and a high power consumption. Therefore, to circumvent these challenges, typically a hybrid beamformer is employed, which has an analog RF beamformer and a baseband precoder. Thus, having obtained the fully-digital precoder solution, we now aim for decomposing the digital solution into a hybrid product, where we invoke the algorithm proposed in [194], which relies on the classic LS solution. Then similarly, the decomposition of the digital combiner matrix into an analog RF beamformer $\mathbf{W}_{\mathrm{RF}}$ and a baseband combiner $\mathbf{W}_{\mathrm{BB}}$ can be carried out.

It is important to note that decomposing the matrix $\mathbf{F}_{\mathrm{opt}}^{(j)}$ into the product of $\mathbf{F}_{\mathrm{RF}}^{(j)}$ and $\mathbf{F}_{\mathrm{BB}}^{(j)}$ is not straightforward because of the constraints imposed on the $\mathbf{F}_{\mathrm{RF}}^{(j)}$ matrix, where the entries in $\mathbf{F}_{\mathrm{RF}}^{(j)}$ must have constant modulus, otherwise we would have a power imbalance in the system. Therefore, we aim for finding the product of two matrices $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ which closely approximates the local optimal precoder $\mathbf{F}_{\mathrm{opt}}^{(j)}$. The optimization problem can be formulated as

$$\min_{\mathbf{F}_{\mathrm{RF}}^{(j)},\mathbf{F}_{\mathrm{BB}}^{(j)}} \|\mathbf{F}_{\mathrm{opt}}^{(j)} - \mathbf{F}_{\mathrm{RF}}^{(j)}\mathbf{F}_{\mathrm{BB}}^{(j)}\|_F^2. \tag{4.36}$$

$$s.t. |\mathbf{F}_{\mathrm{RF}}(m,n)|^2 = 1. \tag{4.37}$$

To solve the above optimization problem, we first fix $\mathbf{F}_{\mathrm{RF}}^{(j)}$ to the angle $\angle\mathbf{F}_{\mathrm{opt}}^{(j)}$ and then find the sub optimal $\mathbf{F}_{\mathrm{BB}}^{(j)}$. This solution is then used for obtaining the unconstrained $\mathbf{F}_{\mathrm{RF}}^{(j)}$ in the next step.

Thus, the problem in (4.36) reduces to a two-stage optimization problem, which can be formulated as

$$\mathbf{F}_{\mathrm{BB}_{k+1}}^{(j)} \triangleq \min_{\mathbf{F}_{\mathrm{BB}}^{(j)}} \|\mathbf{F}_{\mathrm{opt}}^{(j)} - \mathbf{F}_{\mathrm{RF}_k}^{(j)}\mathbf{F}_{\mathrm{BB}_k}^{(j)}\|_F^2, \tag{4.38}$$

$$\mathbf{F}_{\mathrm{RF}_{k+1}}^{(j)} \triangleq \min_{\mathbf{F}_{\mathrm{RF}}^{(j)}} \|\mathbf{F}_{\mathrm{opt}}^{(j)} - \mathbf{F}_{\mathrm{RF}_k}^{(j)}\mathbf{F}_{\mathrm{BB}_k}^{(j)}\|_F^2, \tag{4.39}$$

where k is the iteration index. Furthermore, Equations (4.38) and (4.39) represent a convex quadratically-constrained quadratic programming (QCQP) problem, whose solutions are given as [42, 194]

$$\mathbf{F}_{\mathrm{BB}_{k+1}}^{(j)} = \left(\mathbf{F}_{\mathrm{RF}_k}^{H(j)}\mathbf{F}_{\mathrm{RF}_k}^{(j)}\right)^{-1}\mathbf{F}_{\mathrm{RF}_k}^{H(j)}\mathbf{F}_{\mathrm{opt}}^{(j)}, \tag{4.40}$$

$$\mathbf{F}_{\mathrm{RF}_{k+1}}^{(j)} = \mathbf{F}_{\mathrm{opt}}^{(j)}\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(j)}\left(\mathbf{F}_{\mathrm{BB}_{k+1}}^{(j)}\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(j)}\right)^{-1}. \tag{4.41}$$

Since the $\mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}}$ obtained in (4.41) is unconstrained, we apply the Proposition 2 [194] to obtain the constrained $\mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}}$, where the magnitudes of the entries are constant.

Proposition 4.22 [194]: Suppose $\mathbf{A} \in \mathbb{C}^{N_t \times N_s}$ is expressed as $\mathbf{A}(m,n) = |\mathbf{A}(m,n)|e^{j\angle \mathbf{A}(m,n)}, \forall m,n$. Let $\mathcal{S} = \{\mathbf{B} \in \mathbb{C}^{N_t \times N_s} \mid \ |\mathbf{B}(m,n)| = 1/\sqrt{N_t}, \forall m,n\}$ and

$$\mathbf{U}' = f(\mathbf{A}) \triangleq \arg\min_{\mathbf{U}\in\mathcal{S}} \|\mathbf{A} - \mathbf{U}\|_{\mathrm{F}}^2 \tag{4.42}$$

is the Euclidean projection of $\mathbf{A}$ onto the set $\mathcal{S}$. Then, $\mathbf{U}' = \frac{1}{\sqrt{N_t}} e^{\angle \mathbf{A}(m,n)}$.

Proof: We refer readers to [194].

Thus, $\mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}} = f\left(\mathbf{F}^{(j)}_{\mathrm{opt}} \mathbf{F}^{H(j)}_{\mathrm{BB}_{k+1}} \left(\mathbf{F}^{(j)}_{\mathrm{BB}_{k+1}} \mathbf{F}^{H(j)}_{\mathrm{BB}_{k+1}}\right)^{-1}\right)$.

In other words, we set the magnitude of each entry in $\mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}}$ to $1/\sqrt{N_t^{(j)}}$ and the phase as $\angle \mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}}$. A similar procedure is followed to obtain $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$.

The pseudo-code for hybrid precoding is presented in Algorithm 4.5.

Algorithm 4.5 Iterative matrix decomposition for HBF system

1: Employ algorithm 1 to get $\mathbf{F}^{(j)}_{\mathrm{opt}}$;
2: Initialize $\mathbf{F}_{\mathrm{RF}_k} = \angle \mathbf{F}^{(j)}_{\mathrm{opt}}$;
3: Obtain $\mathbf{F}^{(j)}_{\mathrm{BB}_{k+1}} \leftarrow \left(\mathbf{F}^{H(j)}_{\mathrm{RF}_k} \mathbf{F}^{(j)}_{\mathrm{RF}_k}\right)^{-1} \mathbf{F}^{H(j)}_{\mathrm{RF}_k} \mathbf{F}^{(j)}_{\mathrm{opt}}$;
4: Obtain $\mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}} \leftarrow \mathbf{F}^{(j)}_{\mathrm{opt}} \mathbf{F}^{H(j)}_{\mathrm{BB}_{k+1}} \left(\mathbf{F}^{(j)}_{\mathrm{BB}_{k+1}} \mathbf{F}^{H(j)}_{\mathrm{BB}_{k+1}}\right)^{-1}$;
5: Then, set $\mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}} = 1/\sqrt{N_t^{(j)}} \angle \mathbf{F}^{(j)}_{\mathrm{RF}_{k+1}}$;
6: Repeat steps 3-5 until convergence. Do similarly for $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$;

4.2.4 Complexity

Having acquired the CSI, we now present the complexity in terms of the number of computations involved in both the digital precoder/combiner design and the channel prediction. Having obtained the co-variance matrix $\mathbf{R}$ in (4.48), the number of complex multiplications required to obtain the digital precoder and combiner matrix for $N_t = N_r = N_s = N$ would be $\mathcal{O}(N^3)$. More explicitly, for a different number of transmit and receiver antennas, the total number of computations required is shown in Table 4.1.

4.2.5 Performance Results

Having obtained the hybrid solution, we now characterize the performance of the proposed design and compare it to that of EBF, where the left and right singular matrices

Table 4.1: Number of Computations

Design	*Computations*
Proposed Digital Solution	$N_r^2 N_t + 3N_r N_t N_s +$ $N_s N_r^2 + 2N_r N_s^2$
SVD	$N_r^2 N_t + N_t^2 N_r + N_r^3$

of the corresponding channel are employed. We use the sum rate expression of (4.8) as the performance metric for our simulations. The parameters used in our simulations are summarized in Table 4.2. The Rician factor in the SI channel model in (4.5) is set as high as 50 dB in our simulations, since the SI channel is LOS dominant.

Table 4.2: Simulation parameters

Parameters	Values
Node 1 transmit antennas $N_t^{(1)}$	128
Node 2 transmit antennas $N_t^{(2)}$	128
Node 1 receive antennas $N_r^{(1)}$	32
Node 2 receive antennas $N_r^{(2)}$	32
Node 1 transmit RF chains $N_{t_1}^{\mathrm{RF}}$	2
Node 2 transmit RF chains $N_{t_2}^{\mathrm{RF}}$	2
Node 1 receive RF chains $N_{t_1}^{\mathrm{RF}}$	2
Node 2 receive RF chains $N_{t_2}^{\mathrm{RF}}$	2
Node 1 symbols N_s^1	2
Node 2 symbols N_s^2	2
Rician factor κ [dB]	50
AoA ϕ	$\mathcal{U}(-\pi/6, \pi/6)$
AoD θ	$\mathcal{U}(-\pi/6, \pi/6)$
Modulation	QPSK

Fig. 4.6 compares the sum rate of both the proposed design and of EBF, where the dominant eigen vectors are chosen as the precoder matrix. Furthermore, the curves shown in the figure are without HBF, i.e. with the fully-digital BF. We also show in Fig. 4.6 the achievable sum rate, when the system is interference free. It is seen in Fig. 4.6 that for the signal-to-interference ratios (SIRs) of - 10 dB as well as -30 dB and when $\Theta = 0$ and $D = 0$, the proposed design yields significant sum rate gains. Moreover, when the SIR is as low as -30 dB, the proposed design almost completely eliminates the

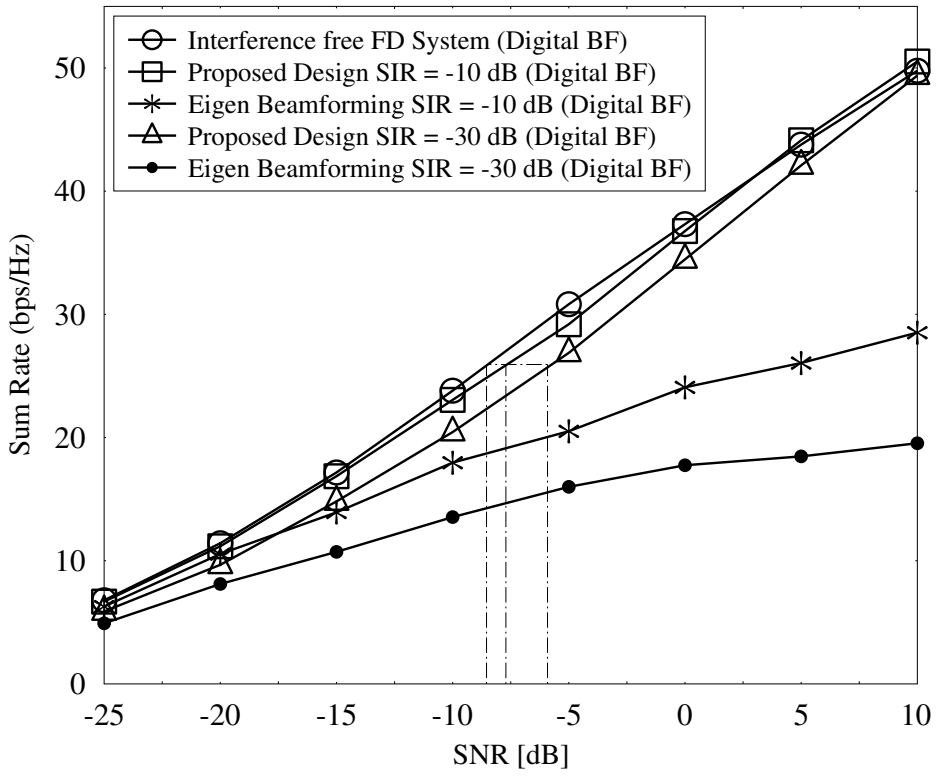


Figure 4.6: Digital precoding of the proposed design for SIR= -10 dB, -30 dB. The distance between the transmit and receive AAs is D=0 and the angle between the arrays is $\Theta = 0°$.

interference and performs close to the system subjected to no interference. On the other hand, the performance of EBF is inferior to the proposed design by a large margin.

Fig. 4.7 shows the sum rate performance of the unconstrained (fully-digital) and constrained (HBF) designs of the proposed algorithm. It can be seen from the figure that for the interference-to-noise ratio (INR) of 0 dB the performance of the fully-digital design and of the HBF designs is similar. However, when the INR is increased to 10 dB, the HBF of the proposed design starts deviating from the fully-digital design. This is because of the constraint in the RF BF to have constant gain entries. The error in splitting the fully-digital matrix into HBF matrices is amplified as the INR is increased.

The comparison of the sum rate of the proposed design and of the EBF using HBF is shown in Fig. 4.8. The curves portrayed in the figure are simulated for SIRs $= -20$ dB and -25 dB, when D=0 and $\Theta = 0°$. It is seen that the proposed BF solution outperforms EBF by about 5 dB for both the SIRs considered.

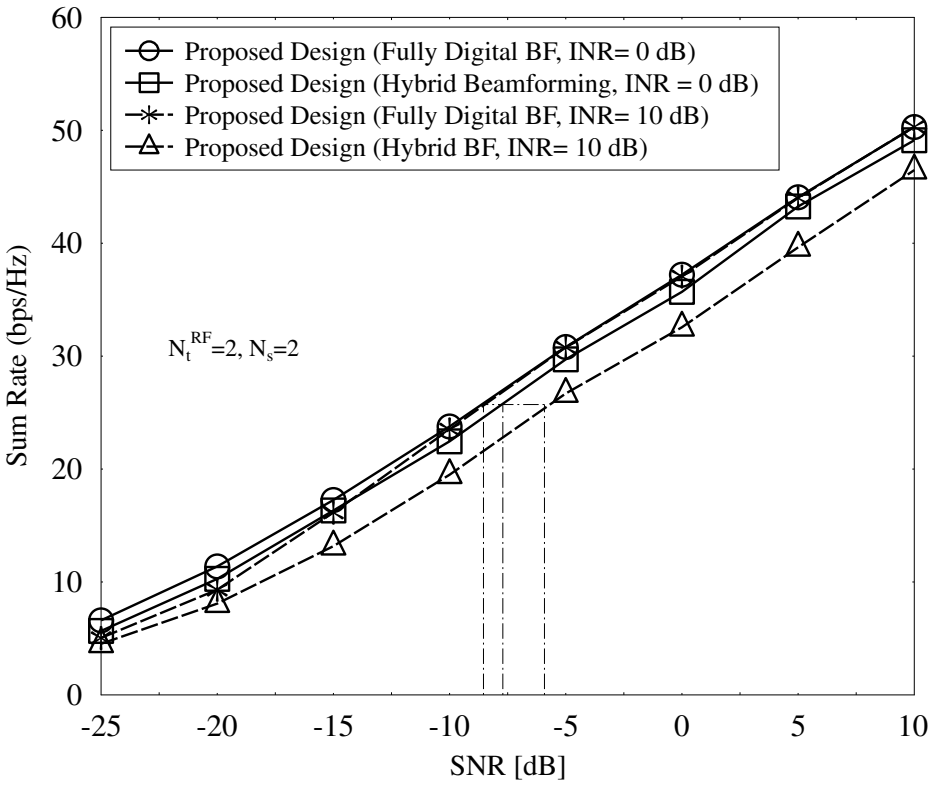


Figure 4.7: Sum rate of HBF relying on the proposed design for INR=0 dB as well as 10 dB. The distance between the transmit and receive AAs is D=0 and the angle between the arrays is $\Theta = 10°$.

To further examine the sum rate of both systems, Fig. 4.9 shows the sum rate of the proposed design for INR= 10 dB, 25 dB. Since the residual SI in FD is significant in practice, an INR as high as 25 dB is considered for Fig. 4.9. It is again evident from the figure that for both the INRs of 10 dB and 25 dB, the proposed design achieves better sum rate than the EBF based HBF. The proposed design is capable of better mitigating the SI and hence provides about 5 dB gain against the EBF system.

Fig. 4.10 (a) shows the achievable rate of both FD and HD systems for a 128×32 element MIMO. It is interesting to note that for higher levels of INR at low SNRs the HD mode provides better sum rate than the FD systems, which can be observed at the cross-over points in the Fig. 4.10 for INR levels of 10 dB and 15 dB. Therefore, in system design, it is important to consider the operating region of FD and HD modes depending on the INR levels, especially at mmWave frequencies because of the dominant residual SI. Furthermore, it is instructive to observe that when the MIMO size is reduced to 64×32, the cross-over shifts further to right as observed in Fig. 4.10 (b). In other words, the

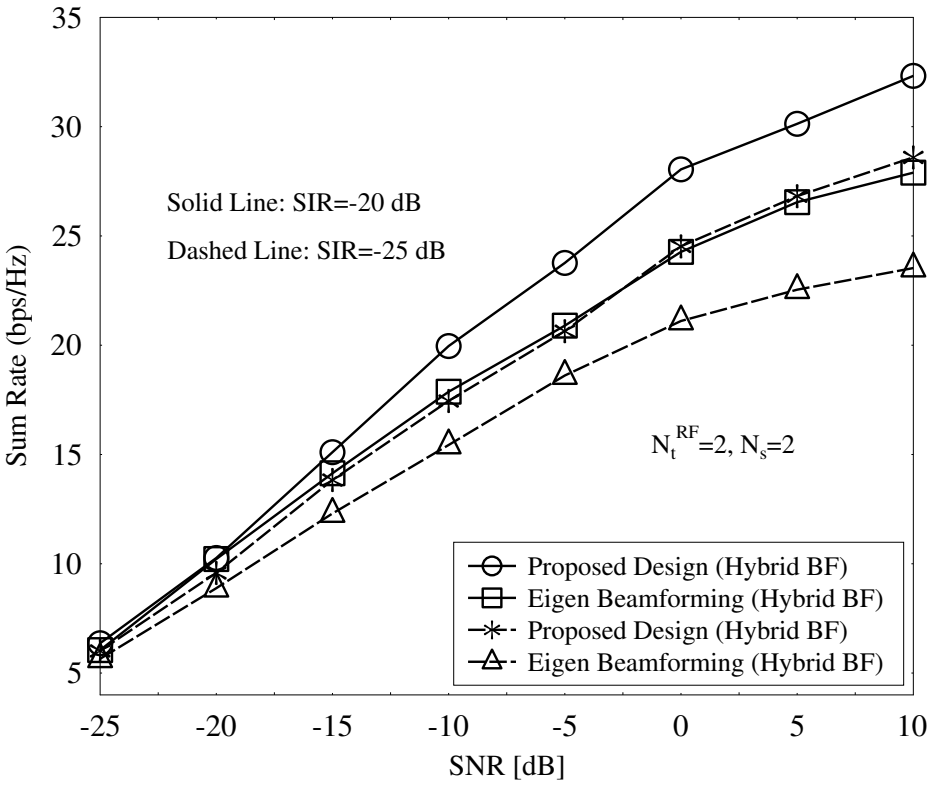


Figure 4.8: Sum rate comparison of the proposed design and of EBF using HBF for different SIRs. The distance between the transmit and receive AAs is D=0 and the angle between the arrays is $\Theta = 10°$.

128×32 element MIMO provides higher BF gain with narrow beams to circumvent the SI when compared to that of 64×32 element MIMO. Thus, large AAs may be employed for directional transmission to suppress the SI efficiently.

Additionally, the performance of the system also depends on the position of the transmit and receive AAs. To study the system performance for both designs with respect to the transmit and receive AAs placement, Fig. 4.11 shows the sum rate for different values of D. It is worth observing from the figure that as the distance D between the transmit and receive antenna is increased from 2λ to 20λ, the performance of the system improves significantly. This is because for the increased separation, the SI would experience higher path loss, which results in passive interference suppression, thus a gain of about 5 dB is observed when D is set to 20λ. Hence, superior performance is observed.

Another important parameter that affects the system's performance is the angle between the transmit and receive AAs. Fig. 4.12 shows the performance of both system designs, when the angle between the arrays is varied. It is seen that for both designs, the sum

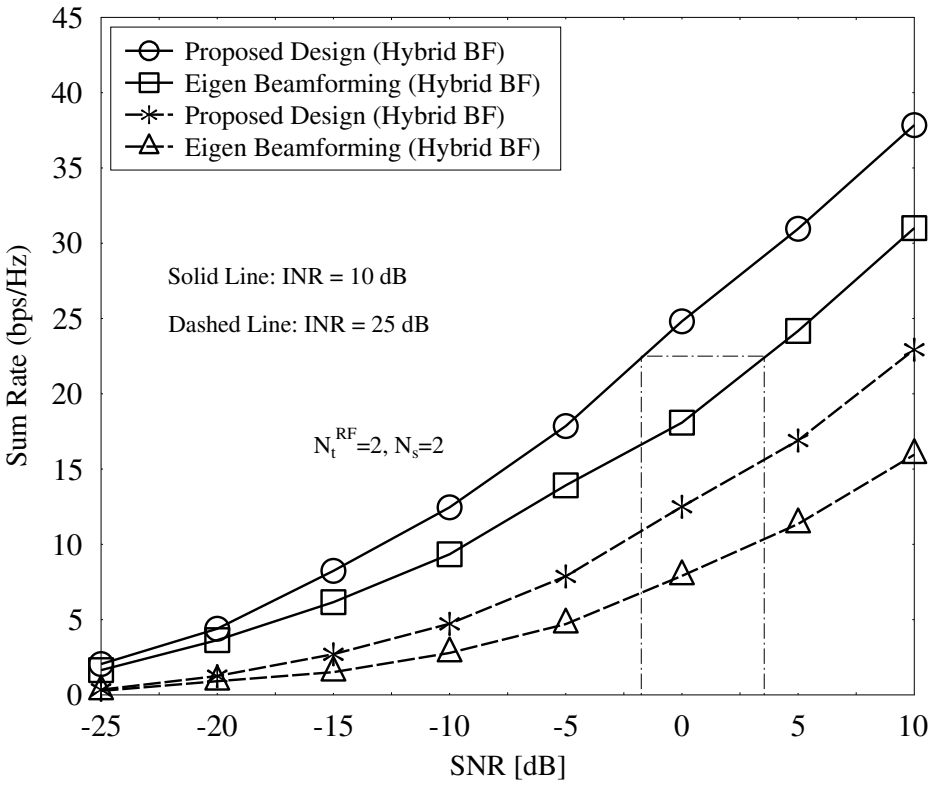
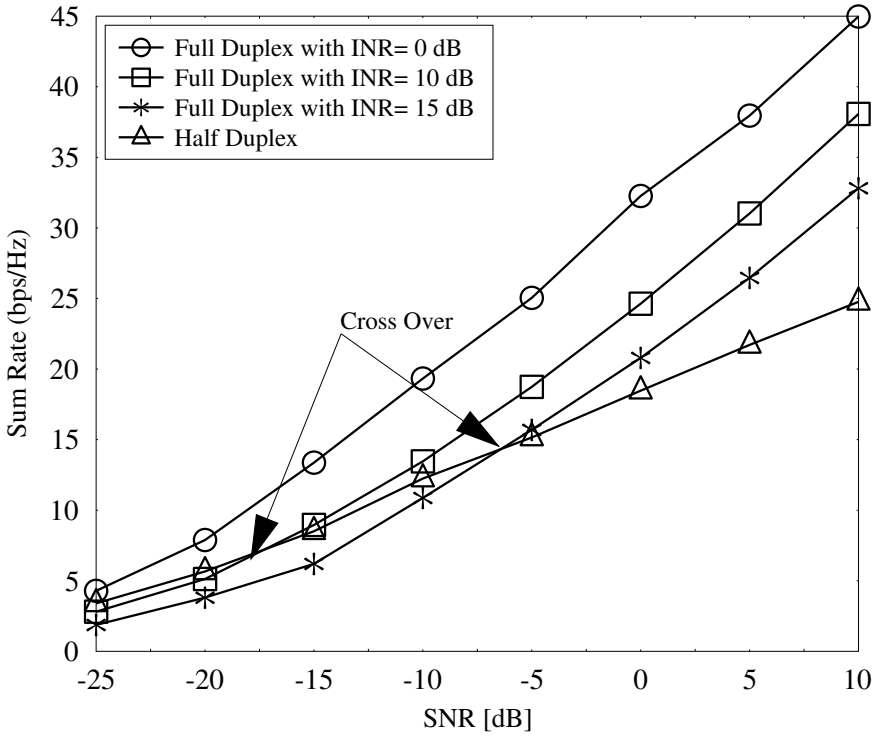


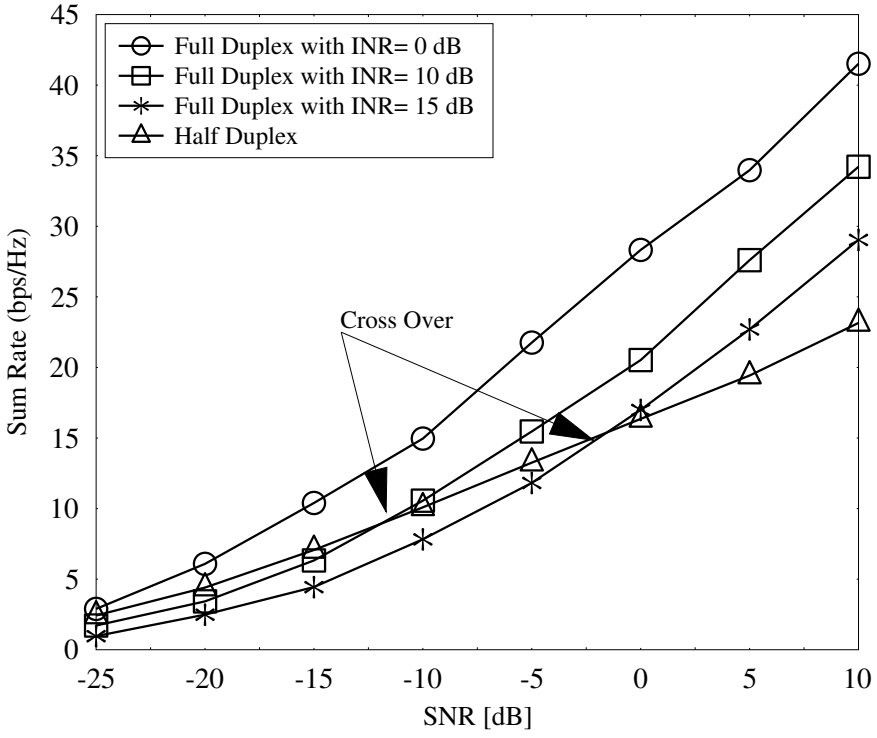
Figure 4.9: Comparison of the sum rate of the proposed design and of EBF using HBF for different INRs. The distance between the transmit and receive AAs is D=0 and the angle between the arrays is $\Theta = 0°$.

rate remains low when the angle between the arrays $\Theta = 0°, 180°$ and $360°$ and D=0. This is expected, because when they share same AA (as $\Theta = 0°$ and D=0), the SI would be high, which has detrimental effects on the system's performance. However, when the arrays are placed at different angles, the sum rate improves, because the angles of arrival and departure would be different, which makes the transmit and receive signal separation more efficiently.

Furthermore, in order to understand the effect of channel estimation error on the performance of the proposed design, we simulated the BER performance of the proposed system, when considering perfect channel knowledge as well as erroneous CSI at the transmitter and receiver. Fig. 4.13 shows the BER curve for both perfect and imperfect CSI, when the precoder and combiner of Section 4.2.3 are employed. In this simulation the error in the channel estimate is assumed to be distributed as complex Gaussian with mean 0 and variance σ_h^2 i.e, $\sim \mathcal{CN}(0, \sigma_h^2)$ [195]. We have investigated the BER performance for σ_h^2 as low as 0.1, and for σ_h^2 as high as 25 [195]. It can be seen from the



(a)



(b)

Figure 4.10: Comparison of the sum rate of FD and HD modes at different SI levels — (a) 128×32 (b) 64×32. The distance between the transmit and receive AAs is D=0 and the angle between the arrays is $\Theta = 0°$.

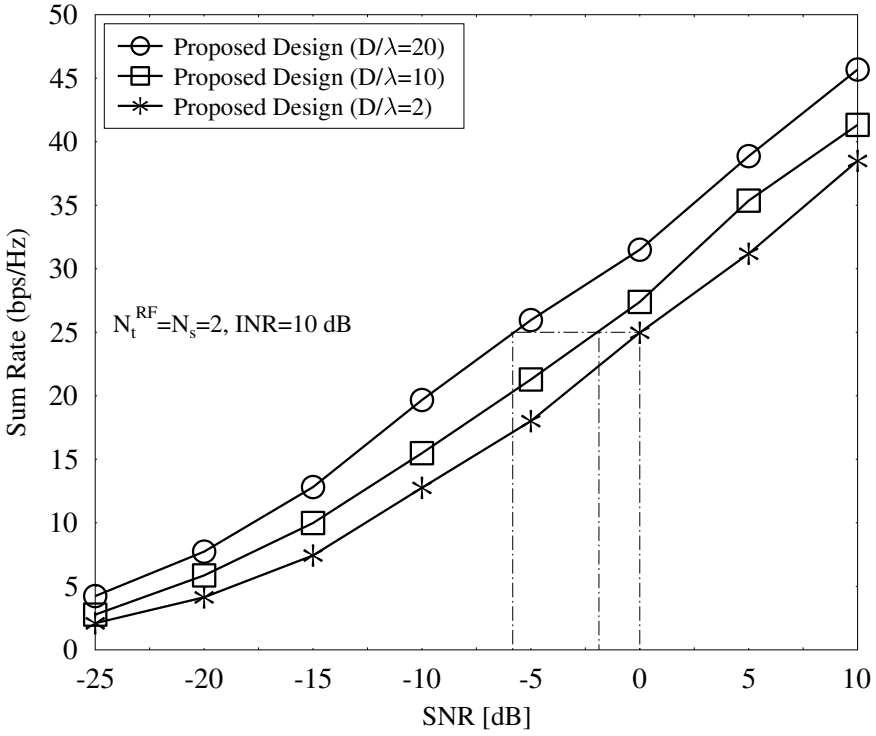


Figure 4.11: Sum rate of the proposed design for different values of D when INR= 10 dB and angle between the transmit and receive AAs $\Theta = 0°$.

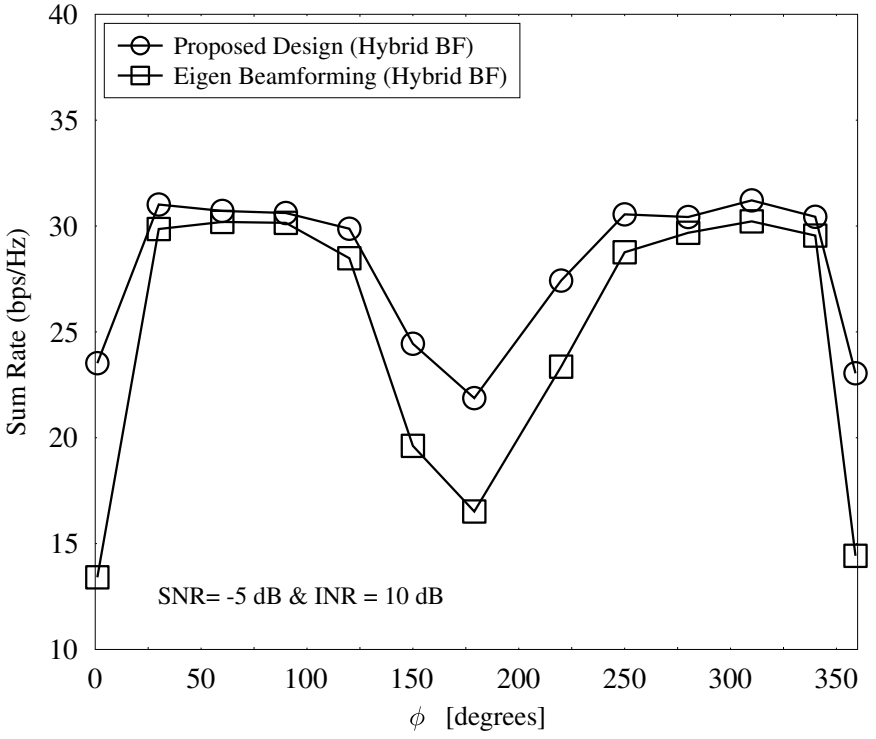


Figure 4.12: Sum rate performance vs angle between the arrays when D=0, SNR= -5 dB and INR =10 dB.

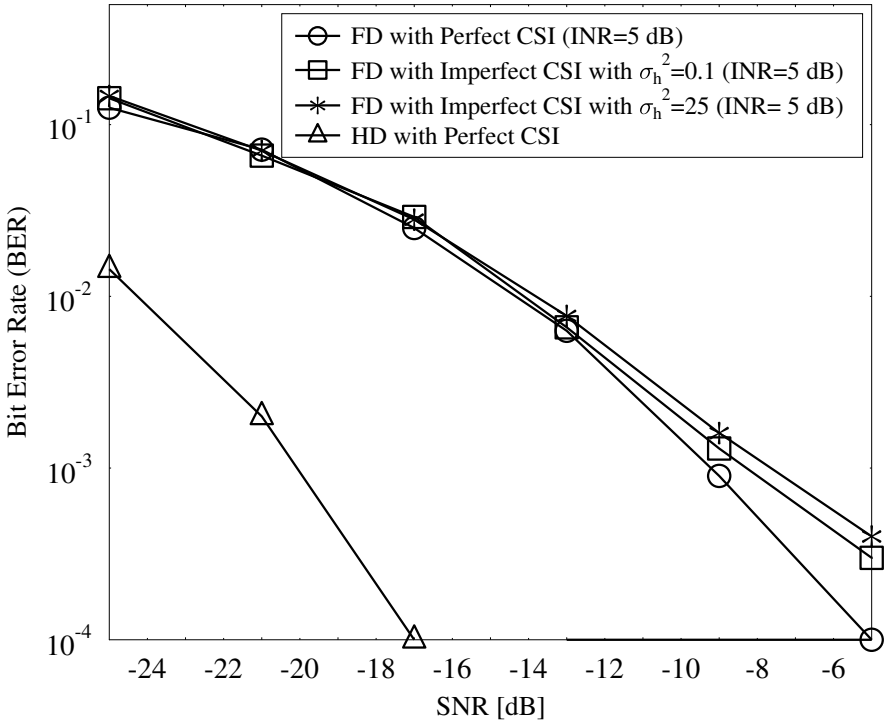


Figure 4.13: BER of the proposed design for FD under imperfect CSI. To obtain this curve, a single stream with QPSK modulation is transmitted using 128×32 element MIMO.

Fig. 4.13 that there is a degradation in the performance of the BER when the channel is imperfect to that of when there is perfect CSI. It is instructive to note that the proposed designed is for maximizing the sum rate, which need not necessarily minimize the BER. Furthermore, the FD is conceived for capacity enhancement, not for BER-reduction. Although the uncoded BER performance of the FD may remain inferior to that of HD as in Fig. 4.13, the coded BER performance of the FD scheme is capable of approaching the capacity [196].

In the next section, we extend our proposed design in the context of multiuser MIMO, where the FD node experiences interference not only from its own transmission but also from external users.

4.3 Multi-User Full Duplex Transceiver

We now focus our attention on full duplex HBF for K-user frequency selective interference channels. Let us consider a K-user interference channel, where the users are communicating over a mmWave channel. To counter act the channel-induced dispersion, each

user employs OFDM based transmission [197]. Fig. 4.14 depicts the K-user interference channel, where each user communicates with its intended receiver. As an example, in Fig. 4.14, Nodes 1 and 2 operate in full duplex mode, while all other nodes operate in half duplex mode[5]. To elaborate further, User 1 at Node 1 is transmitting to its receiver at Node 2 over the wireless channel $\mathbf{H}_{11}$, while User 2 at Node 2 is transmitting to its receiver at Node 1 over the channel $\mathbf{H}_{22}$. Likewise, the user at Node 3 is transmitting to its respective receiver over the channel characterized by the matrix $\mathbf{H}_{33}$ and so forth, while the channel matrix $\mathbf{H}_{ij}, i \neq j$ represents the interfering channel from transmitter i to the receiver j. For example, in Fig. 4.14, $\mathbf{H}_{31}$ is the interfering channel between transmitter 3 and receiver 1.

In the multi-user scenario, we aim for designing the hybrid precoders that minimize the MI for users operating in the half duplex mode and both the MI as well as SI for the users in the FD mode. A similar approach can be used when all the users are in full duplex mode by considering the respective interference channels. It is instructive to note that the users at Nodes 1 and 2 of Fig. 4.14 transmit and receive simultaneously, as shown in Fig. 4.2. On the other hand, the users at other nodes rely on HD communication using the fully-connected design. Following the similar approach discussed for the single-user case, the received signal vector $\mathbf{y}$ for the n^{th} sub-carrier after combining at the i^{th} receiver node is given by[6]

$$\mathbf{y}^{(i)}[n] = \underbrace{\mathbf{W}_{\mathrm{BB}}^{H^{(i)}}[n]\mathbf{W}_{\mathrm{RF}}^{H^{(i)}}\mathbf{H}_{\mathrm{ii}}[n]\mathbf{F}_{\mathrm{RF}}^{(i)}\mathbf{F}_{\mathrm{BB}}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{desired signal}} + \underbrace{\mathbf{W}_{\mathrm{BB}}^{H^{(i)}}[n]\mathbf{W}_{\mathrm{RF}}^{H^{(i)}}\mathbf{H}_{\mathrm{SI}}^{(i)}[n]\mathbf{F}_{\mathrm{RF_{SI}}}^{(i)}\mathbf{F}_{\mathrm{BB_{SI}}}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{self interference (SI)}}$$
$$+ \underbrace{\sum_{j=1}^{K}\mathbf{W}_{\mathrm{BB}}^{H^{(i)}}[n]\mathbf{W}_{\mathrm{RF}}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\mathbf{F}_{\mathrm{RF}}^{(j)}\mathbf{F}_{\mathrm{BB}}^{(j)}[n]\mathbf{s}^{(j)}[n]}_{\text{multi-user interference (MI)}} + \underbrace{\mathbf{W}_{\mathrm{BB}}[n]^{H^{(i)}}\mathbf{W}_{\mathrm{RF}}^{H^{(i)}}\mathbf{n}^{(i)}}_{\text{noise}}, \quad \forall j \neq i, \tag{4.43}$$

where $\mathbf{n}$ is the noise vector of dimension $N_r^{(i)}$, whose elements obey the distribution $\mathcal{CN}(0, \sigma^2)$, $\mathbf{s}^{(i)}$ is the signal vector of dimension $N_s^{(i)}$, $\mathbf{H}_{\mathrm{SI}}^{(i)}[n]$ is the SI channel at the i^{th} node given by (4.5), while $\mathbf{H}_{\mathrm{ji}}$ is the mmWave channel matrix of size $N_r \times N_t$ spanning from the j^{th} node transmitter to the i^{th} node receiver so that $\mathbb{E}[\|\mathbf{H}_{\mathrm{ji}}\|_F^2] = N_t N_r$, as discussed in Sec. 4.1, which is given by

$$\mathbf{H}_{ji}[n] = \sqrt{\frac{N_r^{(i)}N_t^{(j)}}{N_c N_{\mathrm{ray}}}} \times \sum_{n_c=1}^{N_c}\sum_{n_{\mathrm{p}}=1}^{N_{\mathrm{p}}} \alpha_{ji}^{(n_{\mathrm{p}},N_c)}\mathbf{a}_r(\phi_{ji}^{n_{\mathrm{p}},n_c})\mathbf{a}_t^T(\theta_{ji}^{n_{\mathrm{p}},n_c})e^{-\hat{j}2\pi(n_c-1)n/N}, \tag{4.44}$$

[5]The design applies to any configuration, where any node can be FD or HD. In this chapter, we used this example for simplifying explanation.

[6]With slight abuse of notation, we represent the matrices $\mathbf{F}_{\mathrm{RF_{SI}}}^{(i)}\mathbf{F}_{\mathrm{BB_{SI}}}^{(i)}[n]$ as the precoder matrices of the transmitter causing SI at the node i.

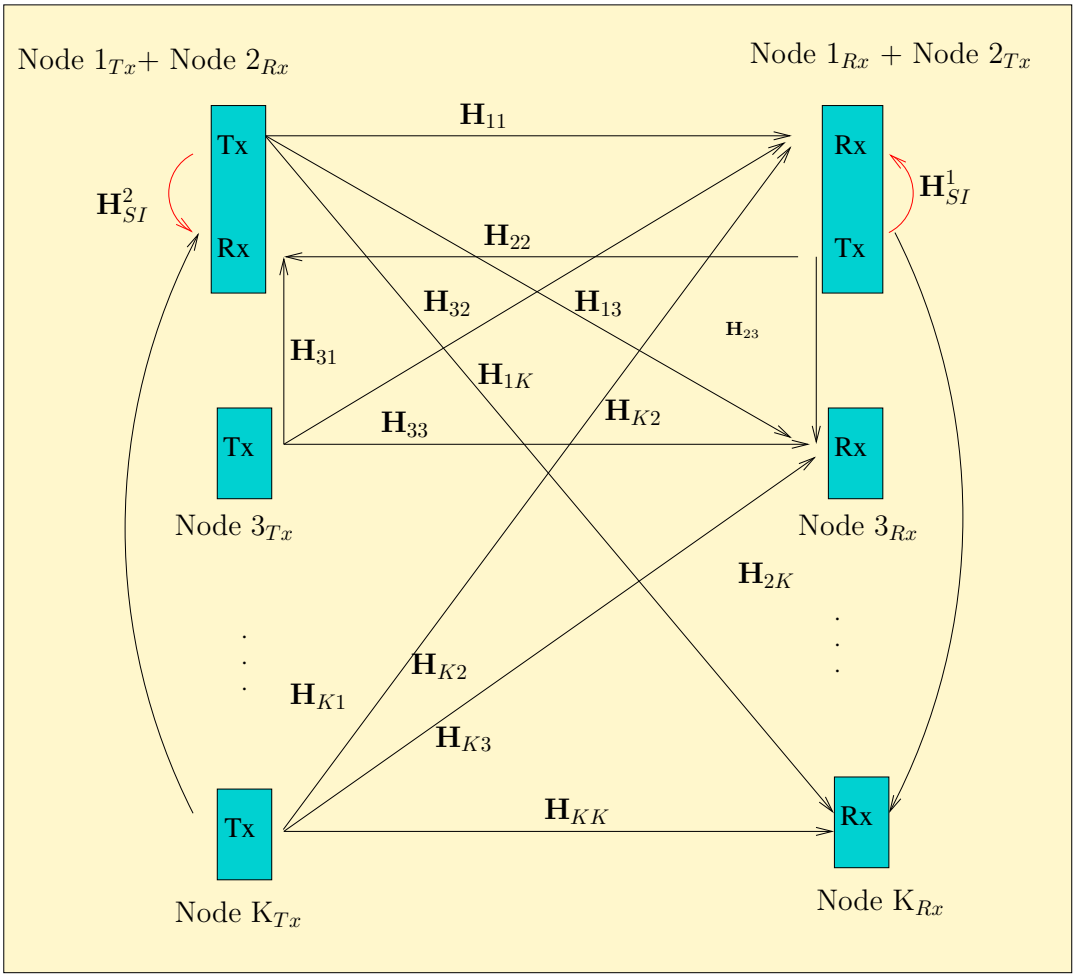


Figure 4.14: *K-user interference channel, where User 1 and 2 rely on FD communication, while the remaining users operate in HD mode.*

where $\hat{j} = \sqrt{-1}$, N is the number of sub-carriers, while the distribution of $\alpha_{n_c}^{n_\mathrm{p}}$ is $\sim \mathcal{CN}(0,1)$.

4.3.1 Precoder and Combiner Design

Following the procedure discussed in Sec. 4.2.1, we first derive the fully-digital solution by setting $\mathbf{F}^{(i)} = \mathbf{F}^{(i)}_{\mathrm{RF}}\mathbf{F}^{(i)}_{\mathrm{BB}}$ and $\mathbf{W}^{H^{(i)}} = \mathbf{W}^{H^{(i)}}_{\mathrm{BB}}\mathbf{W}^{H^{(i)}}_{\mathrm{RF}}$. Then, the system model in (4.43) reduces to

$$\mathbf{y}^{(i)}[n] = \underbrace{\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\mathbf{s}^{(i)}[n]}_{\text{desired signal}} + \underbrace{\mathbf{W}^{H^{(i)}}[n]\mathbf{H}^{(i)}_{\mathrm{SI}}[n]\mathbf{F}^{(i)}_{\mathrm{SI}}[n]\mathbf{s}^{(i)}[n]}_{\text{SI}} + \underbrace{\sum_{j=1}^{K}\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\mathbf{s}^{(j)}[n]}_{\text{MI}} + \underbrace{\mathbf{W}^{H^{(i)}}[n]\mathbf{n}^{(i)}}_{\text{noise}}, \forall j \neq i. \tag{4.45}$$

The second and third terms of (4.43) represent the SI and the MI, respectively. Then, the total interference plus noise (**TI**) at the receiver Node i is

$$\mathbf{TI}^{(i)}[n] = \mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{\mathrm{SI}}^{(i)}[n]\mathbf{F}^{(i)}[n]\mathbf{s}^{(i)}[n] + \sum_{j=1}^{K} \mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}\mathbf{s}^{(j)}[n] + \mathbf{W}^{H^{(i)}}[n]\mathbf{n}^{(i)}[n]. \tag{4.46}$$

Upon denoting the co-variance matrix of the interference plus noise power at the receiver Node i by $\mathbf{Q}^{(i)}[n]$, we have

$$\mathbf{Q}^{(i)}[n] = \mathrm{Tr}\left(\mathbf{W}^{H^{(i)}[n]}\mathbf{R}^{(i)}[n]\mathbf{W}^{(i)}[n]\right), \tag{4.47}$$

with $\mathbf{R}^{(i)}[n]$ represented by

$$\mathbf{R}^{(i)}[n] = \underbrace{P_{\mathrm{SI}}^{(i)}\mathbf{H}_{\mathrm{SI}}^{(i)}[n]\mathbf{F}_{\mathrm{SI}}^{(i)}[n](\mathbf{H}_{\mathrm{SI}}^{(i)}[n]\mathbf{F}_{\mathrm{SI}}^{(i)}[n])^{H}}_{\mathrm{SI}} + \underbrace{\sum_{j=1}^{K} P^{(j)}\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\left(\mathbf{H}_{ji}[n]\mathbf{F}^{(j)}[n]\right)^{H}}_{\mathrm{MI}} + \sigma^2\mathbf{I}_n, \tag{4.48}$$

where $P_{\mathrm{SI}}^{(i)}$ is the SI power, $P^{(j)}$ is the power of the j^{th} user, and σ^2 is the noise power. Our aim is to evaluate $\mathbf{W}$ and $\mathbf{F}$ so that the interference power $\mathbf{Q}^{(i)}$ at the receiver is minimized while also preserving the signal dimensionality, i.e. $\mathrm{rank}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\right) = N_s^{(i)}$.

Proposition 4.3: For the optimization problem of

$$\min_{\mathbf{W}^{(i)}[n]} \mathrm{Tr}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{R}^{(i)}[n]\mathbf{W}^{(i)[n]}\right) \tag{4.49}$$
$$s.t.\ \mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n] = \alpha\mathbf{I}_{N_s},$$

where $\mathbf{R}^{(i)}[n]$ is a positive definite matrix ($\mathbf{R}^{(i)}[n] \succ 0$), the optimal solution is given by Proposition 4.1, where we have

$$\mathbf{W}_{\mathrm{opt}}^{(i)}[n] = \alpha\mathbf{R}^{(i)^{-1}}[n]\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n] * \tag{4.50}$$
$$\left(\left(\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\right)^{H}\mathbf{R}^{(i)^{-1}}[n]\left(\mathbf{H}_{ii}[n]\mathbf{F}^{(i)}[n]\right)\right)^{-1} \tag{4.51}$$

and α is the normalization factor given by

$$\alpha = \frac{1}{\sqrt{\mathrm{Tr}\left(\mathbf{W}^{H(i)}[n]_{\mathrm{opt}}\mathbf{W}_{\mathrm{opt}}^{(i)}[n]\right)}}. \tag{4.52}$$

We then proceed to design the precoder matrix $\mathbf{F}^{(i)}$ by considering the interference

imposed by the transmitter of Node i upon the local receiver of the same node. More explicitly, we aim for specifically designing the precoder to minimize the interference inflicted upon the unintended receivers due to its own transmission.

As in Sec. 4.1, let us consider the total interference imposed by the transmitter j on the unintended receivers, which is given by

$$\mathbf{J}^{(j)}[n] = \mathrm{Tr}\left(\mathbf{F}^{H^{(j)}}[n]\mathbf{S}^{(j)}[n]\mathbf{F}^{(j)}[n]\right), \tag{4.53}$$

where $\mathbf{S}^{(j)}[n]$ is given by[7]

$$\begin{aligned}\mathbf{S}^{(j)}[n] &= (P_{\mathrm{SI}}^{(j)}\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{\mathrm{SI}}^{(j)}[n])^H(\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{\mathrm{SI}}^{(j)}[n])+ \\ &\sum_{i=1}^{K} P^{(i)}\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\right)^H\left(\mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{ji}[n]\right)+\mathbf{I}, \forall i \neq j.\end{aligned} \tag{4.54}$$

Similar to (4.48), our objective here is to minimize the interference power $\mathbf{S}^{(j)}$ caused by the transmitter.

Thus, the associated constrained optimization problem is given by

$$\begin{aligned}&\min_{\mathbf{F}^{(j)}[n]} \ \mathrm{Tr}\left(\mathbf{F}^{H^{(j)}}[n]\mathbf{S}^{(j)}[n]\mathbf{F}^{(j)}[n]\right) \\ &s.t.\ \mathbf{W}^{H^{(i)}}[n]\mathbf{H}_{jj}[n]\mathbf{F}^{(j)}[n] = \beta\mathbf{I}_{N_s}.\end{aligned} \tag{4.55}$$

By invoking Proposition 1, the optimal solution is expressed as

$$\mathbf{F}_{\mathrm{opt}}^{(j)}[n] = \beta{\mathbf{S}^{(j)}}^{-1}[n]\mathbf{H}_{jj}^H[n]\mathbf{W}^{H^{(j)}}[n]* \tag{4.56}$$

$$\left(\left(\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{jj}[n]{\mathbf{S}^{(j)}}^{-1}[n]\right)\left(\mathbf{W}^{H^{(j)}}[n]\mathbf{H}_{jj}[n]\right)^H\right)^{-1} \tag{4.57}$$

and β is a normalization factor given as

$$\beta = \frac{1}{\sqrt{\mathrm{Tr}\left(\mathbf{F}_{\mathrm{opt}}^{H(j)}[n]\mathbf{F}_{\mathrm{opt}}^{(j)}[n]\right)}}. \tag{4.58}$$

The pseudo code of the multi-user precoder and combiner design follows Algorithm 4.4. Hence, in this section, we omit the convergence analysis.

4.3.2 Hybrid Beamforming Solution

Similar to the classic single carrier systems, the analog RF beamformer of OFDM systems is also applied in the time-domain. By contrast the digital baseband TPC is employed in

[7] The term $\mathbf{I}$ in (4.54) ensures the matrix $\mathbf{S}^{(j)}[n]$ is invertible.

the frequency-domain, unlike the TPC discussed in Sec. 4.1. Therefore, to decompose the fully-digital solution obtained into the analog and digital precoder matrices, we pursue a similar approach to [198], where the objective is to minimize the Frobenius norm between the fully-digital and hybrid precoders. This can be formulated as [198]

$$\min_{\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}^{(i)}[n]} \sum_{n=1}^{N} \|\mathbf{F}_{\mathrm{opt}}^{(i)}[n] - \mathbf{F}_{\mathrm{RF}_k}^{(i)}\mathbf{F}_{\mathrm{BB}_k}^{(i)}[n]\|_F^2, \tag{4.59}$$
$$s.t. \begin{cases} |\mathbf{F}_{\mathrm{RF}}(a,b)| = 1 \\ \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[n]\|_{\mathrm{F}}^2 = N_s \end{cases}.$$

In order to solve the above optimization problem, we first fix the matrix $\mathbf{F}_{\mathrm{RF}}^{(i)}$ to $\angle\mathbf{F}_{\mathrm{opt}}^{(i)}$ and then obtain a locally optimal solution $\mathbf{F}_{\mathrm{BB}}^{(i)}$, which is then utilized for obtaining the unconstrained $\mathbf{F}_{\mathrm{RF}}^{(i)}$ in the next step.

Thus, the optimization problem of (4.59) can be interpreted as a two-stage problem. Furthermore, since the baseband TPC is obtained separately for each sub-carrier, we can get rid of the summation in (4.59), while optimizing the baseband TPC $\mathbf{F}_{\mathrm{BB}}[n]$. Therefore, for k^{th} iteration we have

$$\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n] \triangleq \min_{\mathbf{F}_{\mathrm{BB}}^{(i)}[n]} \|\mathbf{F}_{\mathrm{opt}}^{(i)}[n] - \mathbf{F}_{\mathrm{RF}_k}^{(i)}\mathbf{F}_{\mathrm{BB}_k}^{(i)}[n]\|_F^2 \tag{4.60}$$

$$\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)} \triangleq \sum_{n=1}^{N} \min_{\mathbf{F}_{\mathrm{RF}}^{(i)}} \|\mathbf{F}_{\mathrm{opt}}^{(i)}[n] - \mathbf{F}_{\mathrm{RF}_k}^{(i)}\mathbf{F}_{\mathrm{BB}_k}^{(i)}[n]\|_F^2. \tag{4.61}$$

By invoking the classic least-squares solution, we have

$$\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n] = \left(\mathbf{F}_{\mathrm{RF}_k}^{H(i)}\mathbf{F}_{\mathrm{RF}_k}^{(i)}\right)^{-1}\mathbf{F}_{\mathrm{RF}_k}^{H(i)}\mathbf{F}_{\mathrm{opt}}^{(i)}[n], \tag{4.62}$$

(4.63)

$$\mathbf{F}_{\mathrm{RF}_{k+1}}^{(i)} = \left(\sum_{n=1}^{N}\mathbf{F}_{\mathrm{opt}}^{(i)}[n]\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(i)}[n]\right)\left(\sum_{n=1}^{N}\mathbf{F}_{\mathrm{BB}_{k+1}}^{(i)}[n]\mathbf{F}_{\mathrm{BB}_{k+1}}^{H(i)}[n]\right)^{-1}. \tag{4.64}$$

The pseudo-code of the hybrid decomposition developed for FD OFDM systems is presented in Algorithm 4.6.

4.3.3 Performance Results

We now present our simulations characterizing the sum rate of the system relying on the proposed hybrid precoder design. We performed Monte Carlo simulations for studying the performance difference between the proposed and the EBF designs. Our simulation

Algorithm 4.6 Hybrid Decomposition for FD OFDM system

1: Invoke algorithm 1 to obtain $\mathbf{F}_{\text{opt}}^{(i)}$
2: Set $\mathbf{F}_{\text{RF}_k}^{(i)} = \angle \mathbf{F}_{\text{opt}}^{(i)}$
3: Compute $\mathbf{F}_{\text{BB}_{k+1}}^{(i)}[n] \leftarrow \left(\mathbf{F}_{\text{RF}_k}^{H(i)}\mathbf{F}_{\text{RF}_k}^{(i)}\right)^{-1}\mathbf{F}_{\text{RF}_k}^{H(i)}\mathbf{F}_{\text{opt}}^{(i)}[n]$
4: Compute $\mathbf{F}_{\text{RF}_{k+1}}^{(i)} \leftarrow$
$\left(\sum_{n=1}^{N}\mathbf{F}_{\text{opt}}^{(i)}[n]\mathbf{F}_{\text{BB}_{k+1}}^{H(i)}[n]\right)\left(\sum_{n=1}^{N}\mathbf{F}_{\text{BB}_{k+1}}^{(i)}[n]\mathbf{F}_{\text{BB}_{k+1}}^{H(i)}[n]\right)^{-1}$
5: Then, set $\mathbf{F}_{\text{RF}_{k+1}}^{(i)} = 1/\sqrt{N_t^{(i)}}\angle\mathbf{F}_{\text{RF}_{k+1}}^{(i)}$
6: Continue from Step 3 until convergence.

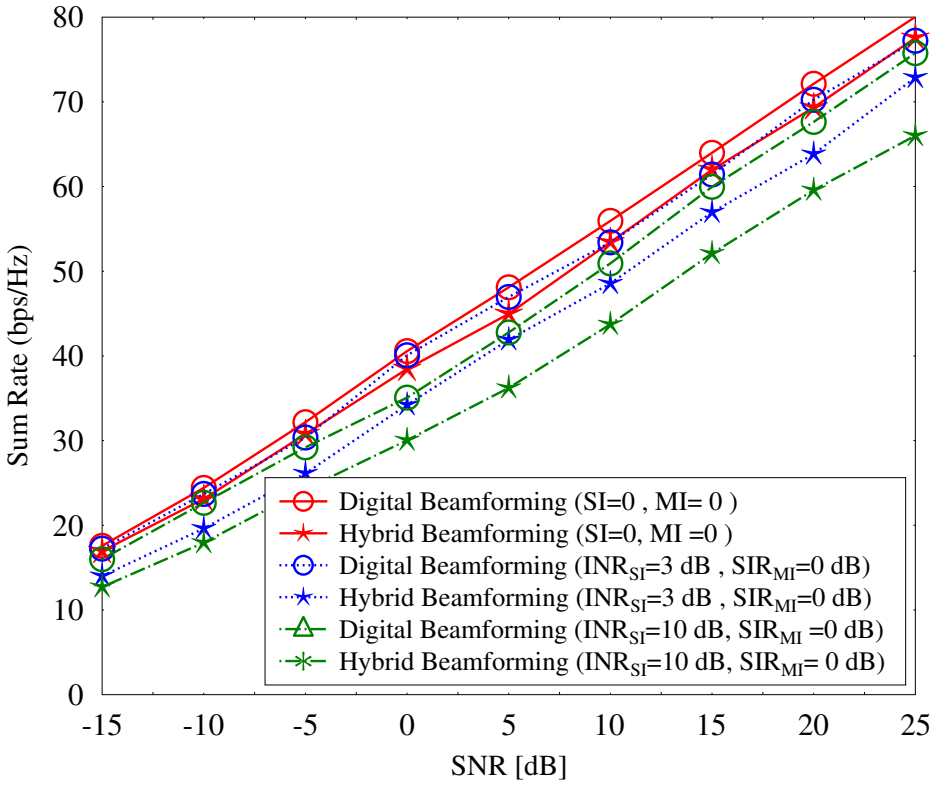


Figure 4.15: Characterizing the sum rate performance of the proposed design having digital solution and of the proposed design having hybrid solution.

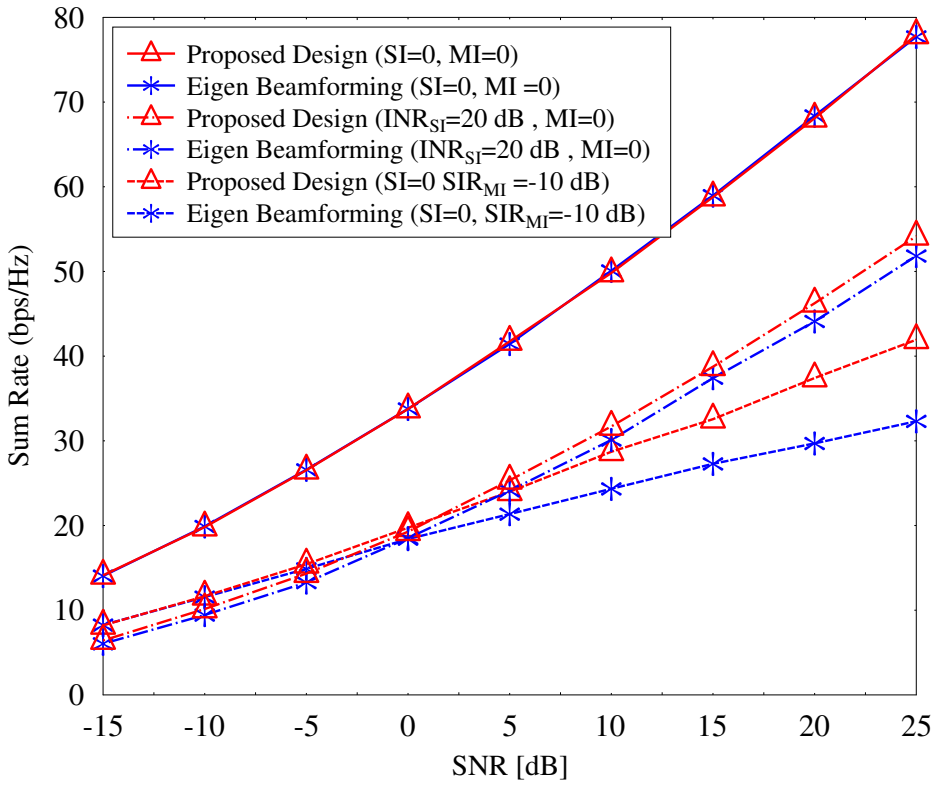


Figure 4.16: Characterizing the sum rate performance of our FD HBF design and of the EBF for different interference configurations. The parameters in Table 4.3 are used for simulations.

parameters are summarized in Table 4.3. The angles of arrivals and departures obey Laplacian distribution with an angular spread of 5°, with the mean angles uniformly distributed between $(0, 60°)$. In these simulations, we have used the INR INR_{SI} as a measure of SI level, where we aim for reducing the SI to the level of the noise floor, while the SIR_{MI} is used as a measure of MI levels. Note that in all our simulation results characterizing the sum-rate performance, the channel employed is the predicted channel for all users.

Fig. 4.15 shows the sum rate performance of the proposed design relying on both the digital and on the HBF solutions. It can be seen from Fig. 4.15 that the HBF solution performs closely to the digital solution, provided that the interference is low. However, when the INR_{SI} is set to 3 dB, there is around 2.5 dB loss at the sum rate of 50 bps/Hz, while it is around 4 dB when the INR_{SI} is set to as high as 10 dB.

Fig. 4.16 shows the sum rate performance of the proposed design and of the EBF design.

Table 4.3: System parameters.

Parameters	*Values*
Number of sub-carrier N (sub-carriers)	32
Node 1 & 2 transmit antennas N_t	128
Node 1 & 2 receive antennas N_r	32
Node 1 & 2 transmit RF chains N_t^{RF}	2
Node 1 & 2 transmit RF chains N_r^{RF}	2
Node 1 & 2 symbols N_s	2
AoA θ	Laplacian distributed
AoD ϕ	Laplacian distributed

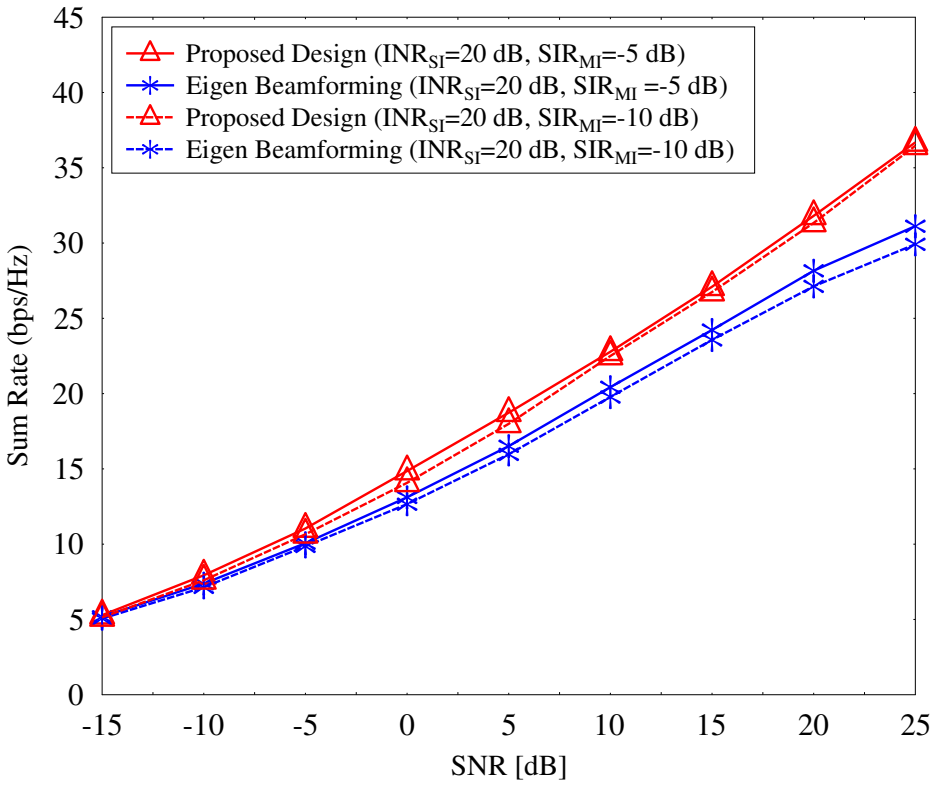


Figure 4.17: Characterizing the sum rate performance of our FD HBF design and of the EBF for a given SI power with different MI levels. The parameters in Table 4.3 are used for simulations. In this simulation, channel with prediction is used.

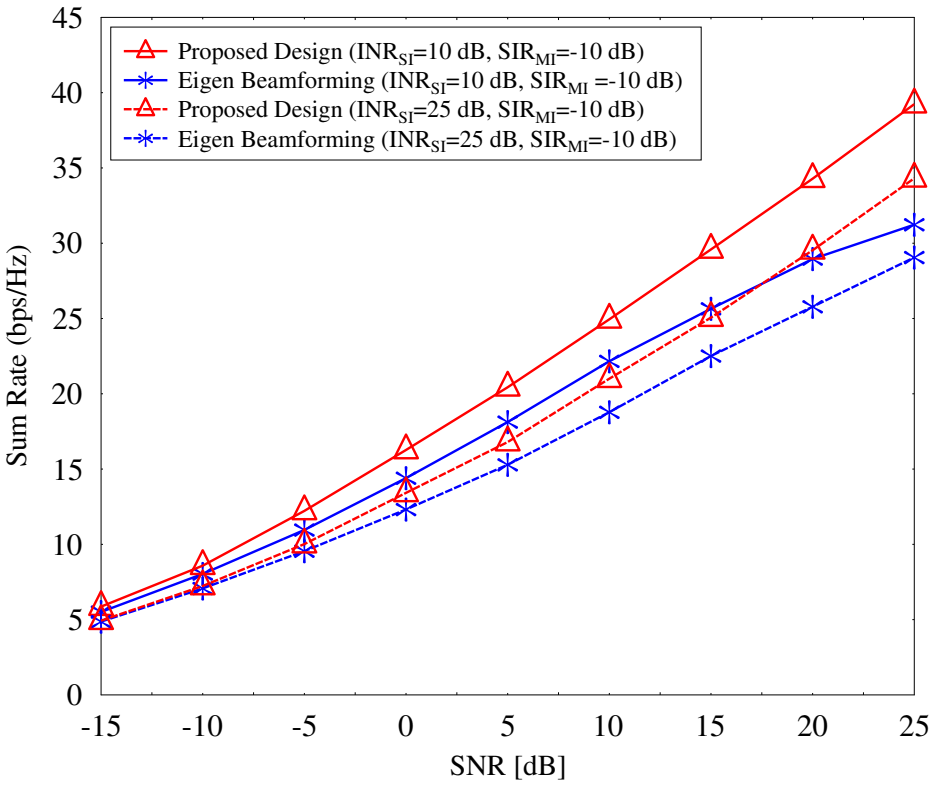


Figure 4.18: Characterizing the sum rate performance of our FD HBF design and of the EBF for a given MI power with different SI levels. The parameters in Table 4.3 are used for simulations. In this simulation, channel with prediction is used.

The simulations are carried out for three interference configurations, namely in the presence of i) no interference; ii) MI; iii) SI. It can be seen in Fig. 4.16 that when there is no MI or SI, the performances of the proposed and of the EBF designs are similar. This is because, in the absence of interference, our proposed hybrid precoder design reduces to the EBF solution, which is optimal in this setting. However, when the SI-to-noise ratio (INR_{SI}) is 20 dB, the performance gap between the two designs becomes clear. Furthermore, when the signal-to-MI ratio (SIR_{MI}) is -10 dB, the proposed design outperforms the EBF design significantly, and the performance gap between the two designs is as high as 10 dB.

Fig. 4.17 shows the sum rate performance of both the proposed design and of the EBF for a given INR_{SI} of 20 dB. It is evident from Fig. 4.17 that the proposed technique outperforms the design relying on EBF by about 4 dB for SIR_{MI} =-5 dB. Moreover, when the SIR_{MI} is set to -10 dB, the EBF degrades the performance. By contrast, the

performance of our proposed technique does not degrade as much as the EBF design and has a similar sum rate to that at SIR_{MI} =-5 dB.

To further understand the behavior of our design, Fig. 4.18 plots the sum rate of our design in comparison to that of the EBF for a fixed SIR_{MI} of -10 dB with varying INR_{SI}. It can be seen from Fig. 4.18 that our proposed design performs about 5 dB better than the EBF design at a sum rate of 30 bps/Hz. It is also worth observing the trajectory of the EBF, indicating that slope of the sum rate starts to reduce as it approaches high SNRs for both INR_{SI}=10 dB and 25 dB. It is instructive to note that despite having high INR_{SI} of 25 dB, our proposed design outperforms the EBF by a significant margin.

4.4 Conclusions

In this chapter, we have advocated the employment of BF in conjunction with both active and passive cancellations. Explicitly, we proposed a digital precoder and combiner solution for suppressing the residual SI relying on BF, and then we followed the classic least square solution to obtain the hybrid product in order to mitigate the hardware complexity by dispensing with power-hungry ADCs/DACs operating at mmWave frequencies, as discussed in Chapters 2 and 3.

Amongst the SI cancellation techniques categorized in Table 4.4, we have opted for BF-aided SI reduction. The rationale of choosing BF is that it can minimize amount of SI leakage into the received signal. More particularly, in Sec. 4.2, we conceived the single-user HBF for full duplex systems. In our system model, since the transmit and receive AAs are closely spaced, we have taken into account the near-field effects in conjunction with the far-field effects for the SI channel. We then proposed a solution for obtaining the digital precoder and combiner weights to suppress the SI.

Table 4.4: Summary of the full duplex SI cancellation methods.

SI Cancellation Techniques	
Active Cancellation	Analog Cancellation
	Digital Cancellation
Passive Cancellation	Antenna Cancellation
	Passive Cancellation
	BF Cancellation

More explicitly, we considered a single-user full duplex mmWave system in Sec. 4.2, where we designed the precoder and combiners for mitigating the SI while also preserving the signal dimensionality. Table 4.5 shows a summary of the proposed design by contrasting it to EBF. Fig. 4.8 captures the scenario, when the SIR is as low as -25 dB. As seen in

Table 4.5: Summary of the full duplex HBF for both single user and multi-user systems based on Figures 4.8, 4.9, and 4.17, 4.18. More particularly, it shows the SNR required to achieve the rate of 22.5 bpcu and 30 bpcu for single and multiuser scenarios, respectively.

	Interference Levels	Figure	EBF	Our Design
Single User Sec. 4.2	SNR (R= 22.5 bpcu & SIR =-25 dB)	Fig. 4.8	3 dB	-2.5 dB
	SNR (R=22.5 bpcu & INR =10 dB)	Fig. 4.9	3 dB	-2.2 dB
Multi-User Sec. 4.3	SNR (R= 30 bpc & SIR_{MI} =-5 dB, INR_{SI} ==20 dB)	Fig. 4.17	22.5 dB	19 dB
	SNR (R=30 bpcu & SIR_{MI} =-10 dB, INR_{SI} ==10 dB)	Fig. 4.18	20 dB	15 dB

Table 4.5, to achieve a rate of 22.5 bpcu, EBF would require an SNR of 3 dB, while our proposed design would only need -2.5 dB. On the other hand, when the INR is set to 10 dB, the SNR requirements are around 3 dB and -2.2 dB for an eigen-beamformer and for our proposed design, respectively, in order to achieve the same rate of 22.5 bpcu.

In Sec. 4.3, we analyzed our BF technique in the context of multi-user mmWave OFDM systems. In this scenario, our precoder and combiner weights aimed for minimizing both the SI and MI, while also preserving the signal dimensions. Figures 4.17 and 4.18 captured the performance of the system for different SIR and INR values. More particularly, we observed in Fig. 4.17, that when we have SIR_{MI} =-5 dB, INR_{SI} ==20 dB, eigen-BF has required an SNR of 22.5 dB, while a lower SNR of 19 dB was required by our design. Similarly, an SNR of 20 dB and 15 dB was necessitated for eigen-beamformer and for our design, respectively, when SIR_{MI} =-10 dB, INR_{SI} ==10 dB.

In chapter 3, we have analyzed the dual-function transceiver designs. More explicitly, we have studied a transceiver design that was capable of providing both diversity and BF gains, while in this Chapter we advocated a BF-assisted transceiver design that can simultaneously transmit and receive data. In the next chapter, we extend our design to multi-functional MIMO systems relying on HBF, where we invoke learning tools for detection.

Chapter 5

Learning-Aided Transceiver Design for Multi-functional mmWave Hybrid Systems

Tell me and I forget. Teach me and I remember. Involve me and I learn.

— Benjamin Franklin

5.1 Introduction

As discussed in chapter 2, the radio-link between the transmitter and receiver may be subjected to fading and high propagation loss, which results in a low SNR at the receiver. Furthermore, the imperfections in the channel and impairments in the ADCs/DACs may exacerbate the received SNR. Therefore, to counter the aforementioned challenges, leveraging machine learning techniques can provide promising solution. In other words, learning tools can be employed for prediction and functional mapping between two metrics to counter the impediments caused by both hardware and channel estimation process, which we will discuss later in this chapter. More particularly, in this chapter we present learning strategy for link-adaptation and symbol detection in mmWave hybrid systems.

Since the received SNR is contingent on the propagation and fading losses experienced by the radio link, matching the modulation and coding rate to the observed SNR becomes necessary for successful symbol detection. Thus to support the escalating data rate demands, typically link-adaptation is used to maximize the data rate while simultaneously meeting the bit error rate (BER) targets [199], where a threshold is used as a criterion to adapt the link based on the pre-defined modes of transmission. In other

words, a BER versus rate look-up table is constructed for each of the legitimate transmission modes. Then the SNR after post-processing is compared against the average SNR threshold values, and the specific mode having the highest throughput as well as satisfying the BER target is activated. Prior work on link-adaptation, such as in [184,200–202], employs adaptation relying on hard threshold values of the average SNR. By contrast the authors of [4, 196] employ adaptation based on the rapidly-fluctuating time-variant channels. However, the ever-changing wireless channel and the non-linearities in the amplifiers erode the performance of conventional adaptation [30]. This is because the decision activated on the transmission scheme is based on the distorted threshold values due to the time-varying nature of the channel and owing to the non-linearities in the amplifiers. Hence, to circumvent the limitations of conventional link-adaptation, machine learning algorithms may be invoked based on the training data used for observation, regardless of the nature of imperfections imposed at the various processing stages [203]. In other words, no threshold values are used for making a decision on the transmission scheme activated. Instead, a more confident decision is made based on the model learned during the training stage.

Furthermore, the employment of multiple-input multiple-output (MIMO) transmission has been proved beneficial for the enhancement of the data rates, as explained in chapter 3. In the literature, there is a vast body of MIMO-aided transmission schemes, where Bell-Labs Layered Space-Time (BLAST) [204] stands as the quintessential and pioneering design for achieving high multiplexing gains. For the benefit of obtaining diversity gains, a variant of MIMO assisted transmission is proposed, which is exemplified by orthogonal space-time block coding (OSTBC) [205] and its generalization to space time block coding (STBC) [206]. With the aforementioned designs, a new MIMO transmission scheme has been born by the amalgamation of diversity, multiplexing and beamforming, which is referred to as multi-functional MIMO (MF MIMO) [207, 208]. In this design, both the diversity and multiplexing as well as the beamforming gains can be obtained to enhance the capacity by manyfold. The MF MIMO is an amalgamation of two or more MIMO schemes. For example, Satyanarayana *et al.* [4] designed a MF MIMO that plays a dual role by providing both diversity and beamforming (BF) gains. Another example is the layered steered space-time coding (LSSTC) [209], where both diversity and multiplexing gains as well as the beamforming gain is achieved by the amalgamation of V-BLAST, STBC and BF. Amongst other MF MIMO techniques, space-time shift keying (STSK) [210, 211] is popular for its design trade-off between multiplexing and diversity gains. The STSK design is conceived as an extension to the concept of spatial modulation [212], where a single antenna is activated at any time. To elaborate a little further, in the STSK design, a single dispersion matrix (DM) is activated amongst a set of DMs at any time. In other words, information is conveyed implicitly by the index of the DM in addition to the complex-valued signal drawn from M-ary constellation. As an extension of the STSK, a multi-set (MS) STSK was proposed [213,214], which is formed by combining the concepts of the STSK and SM. This design is capable of increasing the

data rate, since the information is implicitly carried by both M-ary alphabet and DM index as well as the antenna index combination. In mmWave communications, where the channel supports a few cluster of rays, the data rate of the MS-STSK design can be further enhanced by coupling with the concept of beam index modulation (BIM) [215]. In the BIM aided transmission, information is implicitly conveyed by the index of the beams in addition to the M-ary constellation.

However, a growing concern in the index modulation transmission schemes, such as the MS-STSK is the search complexity at the receiver [213]. Furthermore, like in any communication system, it imposes the requirement of having accurate CSI at the receiver to achieve a low bit error rate (BER) [216]. In frequency division duplex (FDD) systems, the CSI estimate is typically carried out relying on pilots sent prior to the data transmission. Thus, this results in a reduced data rate in addition to the extra complexity involved for channel estimation before the final stage of detection is ensued.

To circumvent this problem, a machine learning approach may be employed, where the symbol detection is carried out without explicit CSI knowledge. *In other words, the receiver relying on the learning strategy employed during the training process for the detection, turns a 'blind eye' to the knowledge of CSI — thereby evading the pilot-overhead involved in channel estimation. This philosophy makes the design more spectral-efficient.*

Against this backdrop, in this chapter, we invoke a supervised learning based algorithm, where the decision/prediction is made based on the observation samples collected during the training phase. Both the BER and the instantaneous post processing SNR are taken as feature spaces to capture the channel conditions as well as the implementation losses imposed by the imperfections of the amplifiers. In our chapter, the dimension of the feature space is 2. It is also instructive to note that if the dimension of the feature space is high, machine learning algorithms would result in erroneous solutions, unless they are provided with an exceedingly large number of training samples for decision-making [31].

On the other hand, in the mmWave HBF literature, analog BF is always combined with digital BF, regardless of the nature of the channel. Contrary to popular belief, we show in this chapter that activating both a digital precoder and combiner when the channel has only a single dominant path is redundant. Therefore, switching off both the digital precoder and the digital combiner and activating analog only BF is energy-efficient.

Against this background, our contributions in this chapter are summarized as follows.

1. *We propose a learning assisted adaptive transceiver design for each user link based on the near-instantaneous post-processed SNR, where the adaptation switches between multiplexing versus diversity oriented transmission modes as well as by appropriately configuring the modulation employed so as to facilitate both high-reliability and high-rate operation. The receiver relies on the instantaneous post-processed*

SNR to decide on the transmitter's multiplexing versus diversity aided transmission mode and on the choice of the specific modulation scheme to be employed with the aid of supervised learning relying on the feed-back information forwarded to the BS. In this chapter, we invoke the KNN classification technique at the receiver for decision making, as a design example.

2. *We show through simulations that at a target BER of 10^{-3}, the learning-assisted adaptation achieves a significantly higher rate with a SNR gain of about 5 dB, while maintaining the required target BER compared to that of conventional link-adaptation carried out based on hard threshold values. Furthermore, we demonstrate by simulations that the performance of the system relying on HBF and on analog only BF is similar in an environment, where there is only a single dominant path for communications.*

3. *We also propose a deep learning assisted semi-blind detection for index modulation mmWave MIMO systems. More particularity, we propose a detection design for MS-STSK transmitter without relying on the explicit knowledge of the CSI at the receiver. We also extend our design to a transmission scheme, where the MS-STSK is coupled with the BIM but dispensing with the CSI knowledge for detection. This philosophy makes our design spectral-efficient, since it evades the need for the pilot-assisted channel estimation.*

4. *We demonstrate by simulations that our proposed design detects the MS-STSK information with integrity while circumventing the CSI estimation, unlike the conventional maximum-likelihood (ML) detector.*

5. *We show by simulations that our design outperforms the ML-aided detection in the face of channel impairments introduced during the CSI estimation. Furthermore, we demonstrate that the net DCMC capacity of the ML-aided detection is lower than that of our proposed learning-assisted detection. In other words, the DCMC capacity is significantly affected by the pilot overhead when having ML-aided detection, since it requires accurate CSI in every frame of the transmission.*

6. *To improve the fidelity of our learning-assisted detection, we propose recalibration of the NN weights after a certain number of frames, which is contingent on the Doppler spread. The recalibration of the NN weights is carried out using side-information or training information, which is known at both the transmitter and receiver. Hence, we call our design semi-blind, since it requires the transmission of the training data for recalibration of the NN weights. We show that the overhead needed for this training data is negligible compared to the pilot overhead for channel estimation.*

7. *We show by simulations that our proposed design outperforms ML-aided detection in the face of channel estimation errors whose variance is as low as 0.15.*

8. *A qualitative complexity discussion is presented both for the learning-assisted detection and for the ML-aided detection in terms of search space as well as the number of computations.*

9. *We then extend our NN design to provide the soft values of the MS-STSK symbol without relying on the explicit CSI. Furthermore, we demonstrate by simulations that despite both its incognizance about the CSI and the low complexity, the NN design performs closely to the conventional soft-demodulator having perfect CSI, while it outperforms the conventional design in the face of imperfections in the CSI at the receiver.*

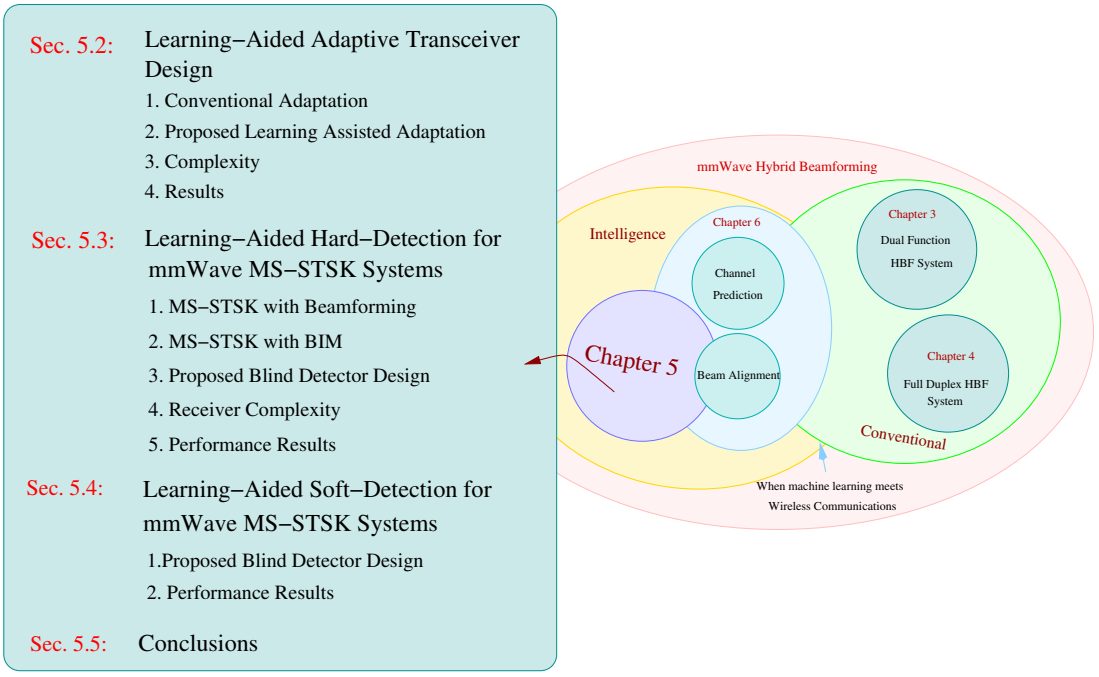


Figure 5.1: The organization of the chapter.

The remainder of this chapter is organized as shown in the Fig. 5.1. In Sec. 5.2, we propose learning-aided link-adaption while contrasting with conventional designs in mmWave hybrid systems. Then, we extend our philosophy to MS-STSK mmWave systems in Sec. 5.3 and Sec. 5.4 using learning for hard- and soft-detection, respectively. Finally, our conclusions are presented in Sec. 5.5.

Having discussed the allocation of phase shifters and RF chains, we now focus our attention to a single-link of (5.1), where the BS design its $\mathbf{F}_{\mathrm{RF}}^{k}$ and $\mathbf{F}_{\mathrm{BB}}^{k}$ as well as the modulation and transmission scheme depending on the nature of the channel.

5.2 Learning-Aided Adaptive Transceiver Design

Let us again consider Fig. 2.25, where K users communicating with the base station. After allocating the appropriate number of phase shifters and RF chains to the users as discussed in Sec. 2.5, each user link can be modeled relying on the fully-connected system model of Fig. 2.23, where user[1] k receives its signal transmitted from N_{t_k} active antennas, which is then digitally processed using N_k^{RF} RF chains. However, note that the analysis can readily be extended to sub-array-connected designs of Fig. 2.24 and Fig. 3.2. Then the vector of received signal for user k using fully-connected HBF design is given by

$$\mathbf{y}_k = \mathbf{W}_{\mathrm{BB}}^{kH}\mathbf{W}_{\mathrm{RF}}^{kH}\mathbf{H}_k\mathbf{F}_{\mathrm{RF}}^{k}\mathbf{F}_{\mathrm{BB}}^{k}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^{kH}\mathbf{W}_{\mathrm{RF}}^{kH}\mathbf{n}, \tag{5.1}$$

It is important to emphasize that the architecture of Fig. 2.26 may be deemed to be equivalent to that of Fig. 2.23, where the BS of Fig. 2.26 processes the signals digitally in the baseband using a digital TPC matrix $\mathbf{F}_{\mathrm{BB}}$ of size $N_t^{\mathrm{RF}} \times N_s$ and then the digitally precoded signal is phase shifted using the RF beamformer matrix $\mathbf{F}_{\mathrm{RF}}$ of size $N_t \times N_t^{\mathrm{RF}}$ before transmission. Then the collective downlink received signal vector $\mathbf{y}$ for all K users of Fig. 2.25 after both RF and baseband processing using the constituent matrices $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ of sizes $N_r \times N_r^{\mathrm{RF}}$ and $N_r^{\mathrm{RF}} \times N_s$, respectively, is given by

$$\mathbf{y} = \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{n}, \tag{5.2}$$

where $\mathbf{n}$ is the Gaussian noise, $\mathbf{W}_{\mathrm{RF}} = \mathrm{diag}\left[\mathbf{W}_{\mathrm{RF}}^{1}, \ldots, \mathbf{W}_{\mathrm{RF}}^{K}\right]$, $\mathbf{W}_{\mathrm{BB}} = \mathrm{diag}\left[\mathbf{W}_{\mathrm{BB}}^{1}, \ldots, \mathbf{W}_{\mathrm{BB}}^{K}\right]$. Furthermore, $\mathbf{H} = \mathrm{diag}\left[\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_K\right]$ and $\mathbf{s}$ is the transmit symbol vector of size N_s, while $\mathbf{y} = [\bar{y}_1^T \ldots \bar{y}_k^T \ldots \bar{y}_K^T]^T$. Similarly, $\mathbf{F}_{\mathrm{RF}} = \mathrm{diag}\left[\mathbf{F}_{\mathrm{RF}}^{1}, \mathbf{F}_{\mathrm{RF}}^{2}, \ldots, \mathbf{F}_{\mathrm{RF}}^{K}\right]$, while $\mathbf{F}_{\mathrm{BB}} = \mathrm{diag}\left[\mathbf{F}_{\mathrm{BB}}^{1}, \mathbf{F}_{\mathrm{BB}}^{2}, \ldots, \mathbf{F}_{\mathrm{BB}}^{K}\right]$.

Traditionally, the link-adaptation is carried out based on the average threshold values. Upon receiving the signal, the receiver calculates the instantaneous post-processing SNR, based on which the receiver makes the decision concerning the most appropriate transmission mode by comparing it against the pre-defined average SNR threshold values [184].

The total post-processing SNR, which is the SNR calculated after combining using the matrix $\mathbf{W}_k$ at the receiver of user k, for a given channel realization $\mathbf{H}_k$ and noise variance

[1] As a design example of Fig. 2.25, we assumed K groups with one user in each. However, there may be more than one user in any group at any point of time. For brevity we considered one user in each group — and each user in a group may be served by a plurality of RF chains provided that the user's channel supports multiple beams. Having plurality of RF chains implies plurality of multiple beams for transmission. However, the number of beams is governed by the rank of the channel matrix.

σ_n^2 is given by

$$\text{SNR} = \frac{\text{tr}\left(\left(\mathbf{W}^{k^H}\mathbf{H}_k\mathbf{F}^k\right)^H \mathbf{W}^{k^H}\mathbf{H}_k\mathbf{F}^k\right)}{\text{tr}\left(\mathbf{W}^k\mathbf{W}^{k^H}\sigma_n^2\right)}, \tag{5.3}$$

where $\mathbf{W}^k = \mathbf{W}^k_{\text{RF}}\mathbf{W}^k_{\text{BB}}$ and $\mathbf{F}^k = \mathbf{F}^k_{\text{RF}}\mathbf{F}^k_{\text{BB}}$.

The matrices $\mathbf{F}^k_{\text{RF}}$ and $\mathbf{F}^k_{\text{BB}}$ are designed by maximizing the capacity of the hybrid precoder. More explicitly, the objective function formulation results in minimizing the Frobenius norm between the optimal matrix and the matrix product $\mathbf{F}^k_{\text{RF}}\mathbf{F}^k_{\text{BB}}$ [194], which is formulated as:

$$\min_{\mathbf{F}^k_{\text{RF}},\mathbf{F}^k_{\text{BB}}} \|\mathbf{F}^k_{\text{opt}} - \mathbf{F}^k_{\text{RF}}\mathbf{F}^k_{\text{BB}}\|_{\text{F}}, \tag{5.4}$$

$$s.t. \quad \|\mathbf{F}^k_{\text{RF}}\mathbf{F}^k_{\text{BB}}\|^2_{\text{F}} = N_s, \tag{5.5}$$

$$|\mathbf{F}^k_{\text{RF}}(m,n)| = 1. \tag{5.6}$$

The optimal precoder matrix $\mathbf{F}^k_{\text{opt}}$ of the system model in (5.2) is obtained by the SVD of the channel matrix $\mathbf{H}_k = \mathbf{U}_k\Sigma_k\mathbf{V}_k^H$, where the first N_s columns of the right singular matrix $\mathbf{V}_k$ are chosen to construct the matrix $\mathbf{F}^k_{\text{opt}}$. We note that in order to decompose the digital precoder into hybrid product, the solutions presented in Table 2.3 may be invoked. In the next section, we detail both the conventional and the proposed link-adaptation.

Remark 5.1. It is instructive to note that the system considered is not very different from the scenario of having unknown interferences. The proposed design can be readily extended to the situation when there is interference from the unintended transmitters by using the SINR instead of the SNR. In this setting the denominator in (5.3) would contain one more additional term, which is the interference. Moreover, according to the central limit theorem, when there is a large number of interferes, the interference caused by them can be modeled as additional Gaussian noise. In this case (5.3) would remain the same, except for an increased noise variance.

5.2.1 Conventional Adaptation

In conventional adaptation, the receiver makes a decision concerning the mode of transmission using the post-processed SNR based on the pre-defined threshold SNR values, which are set to meet a required BER. Fig. 5.2(a) shows the plot of average BER against the average SNR for different transmission schemes using 64×32 element MIMO scheme relying on two RF chains at both the BS as well as at the receiver and communicating over the mmWave channel model. In this plot, two spatial streams are used for

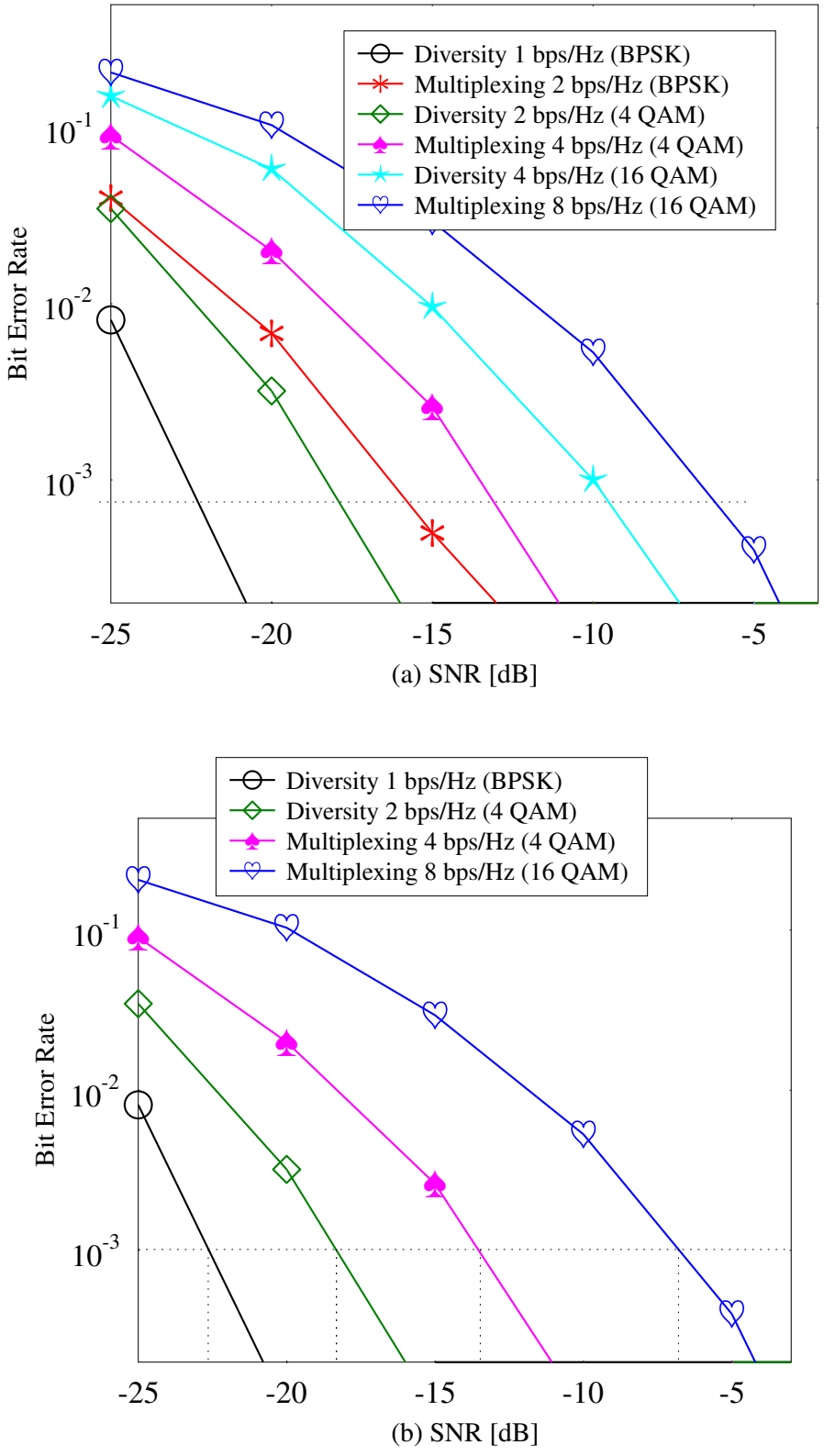


Figure 5.2: Average BER versus average SNR of spatial diversity and spatial multiplexing for different transmission rates. In this configuration, two spatial streams are transmitted using a 64×32 element MIMO with N_t^{RF} and $N_r^{\mathrm{RF}} = 2$, while channel is NLOS in nature.

spatial multiplexing, while only a single spatial stream is used when aiming for diversity. Observe from the plot that the diversity-oriented QPSK scheme performs better than spatial multiplexing using two BPSK streams. Furthermore, spatial multiplexing with two QPSK streams achieves better BER than diversity-aided 16 QAM. Therefore, Fig. 5.2(b) excludes both the inferior spatial multiplexing aided BPSK streams and the diversity aided 16 QAM, hence only considering the schemes that provide a better BER for a given rate.

Following the conventional link-adaptation, the specific threshold values designed for each scheme characterized in Fig. 5.2(b) that attain the target BER of 10^{-3} are shown by vertical lines. After the calculation of the instantaneous post-processed SNR the receiver decides on both the type of transmission scheme as well as on the modulation mode by comparing it against the pre-defined threshold values. Explicitly, the receiver compares the post-processed SNR against the vertical lines in Fig. 5.2(b) and relays the requested mode information to the BS. We note that the post-processed SNR values are calculated offline and stored in a memory so that the receiver does not have to do any calculations.

On the other hand, in the case of only one dominant path, the digital BF is switched off and adaptation is performed amongst the modulation schemes only, since the channel does not support multiplexing/diversity.

5.2.2 Proposed Learning Assisted Adaptation

In contrast to conventional adaptation, in the proposed design, the receiver relies on learning, hence dispensing with any pre-defined threshold values. Here we conceive a classification algorithm. However, most classification algorithms require an explicit functional mapping between the feature set and the classifiers. The feature set in our work includes the SNR and the BER, while the classifiers are spatial multiplexing associated with different modulations modes and spatial diversity also using different modulations. Unfortunately, there is a paucity of information about the functional mapping between the two. Hence, we conceive a KNN classification algorithm, which is non-parametric, since it does not require any information about the functional mapping [203]. Furthermore, the theoretical assumptions made related to its mathematical tractability may become invalid in practical environments, where the data may not obey a specific distribution. In such scenarios, the KNN algorithm would be a promising choice, because again, it does not depend on any assumptions or knowledge about the data distribution [203]. The operating principle of the KNN algorithm is illustrated in Fig. 5.3, where the training data of two classifiers are distinguished with squares and circles. When a testing data point is given, a circle with the testing data point as its center is drawn so that it encircles K points from the classifiers, as shown in Fig. 5.3. Then the class associated with more points in the circle will be chosen by invoking majority voting. If

the number of points in each class is the same, then the class having a better throughput is selected. The feature set used for link-adaptation is $\mathcal{F} = \{\text{SNR}, \text{BER}\}$, and the class

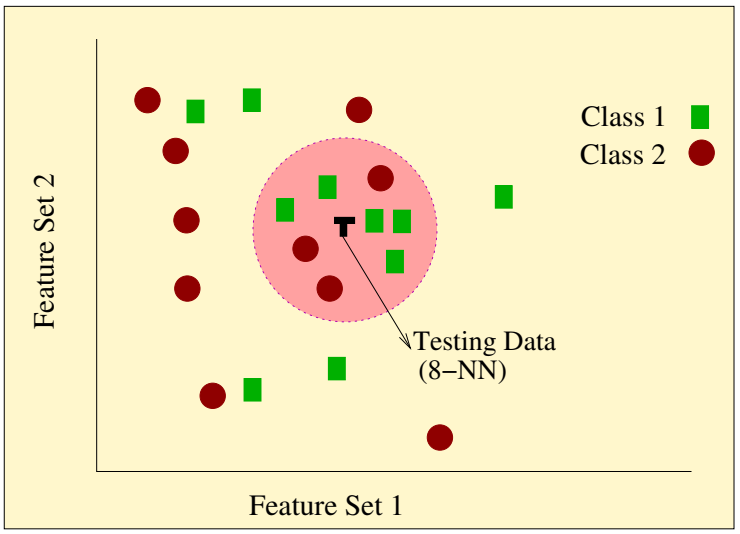


Figure 5.3: Illustration of K-Nearest Neighborhood Algorithm.

set is defined as $C = \{\text{Div}_{\text{BPSK}}, \text{Div}_{\text{QPSK}}, \text{Mux}_{\text{QPSK}}, \text{Mux}_{16\text{QAM}}\}$[2] for NLOS channel, while $C = \{\text{QPSK}, 16\text{QAM}\}$ when the channel has only a single dominant path since it has no multiplexing/diversity gain. In other words, when the channel has only a single dominant path, adaptation during the second stage only reconfigures the modulation scheme.

The proposed learning adaptation is comprised of two phases: the training phase and the testing phase. In the training phase, both the BER and the instantaneous post-processing SNR are calculated for each channel realization and stored in memory. Having accumulated the training data, the testing phase ensues. When a new data point is received, the post-processed SNR is calculated. Then, equipped with the post-processed SNR and the required BER as the parameters, the K nearest neighbors are chosen from the set C. Finally, the specific class which has more points in the neighborhood is selected. To elaborate further, Fig. 5.4 shows the training data for 100 channel realizations at 5 different noise levels, where 500 symbols are transmitted for every channel realization. Then the BER is computed for 500 symbols. This process is carried out for different tranmission techniques. For example, in Fig. 5.4 the diversity associated with 2 bps/Hz is obtained by transmitting a 4-QAM symbol using two RF chains, while multiplexing associated with 4 bps/Hz is obtained by transmitting two 4-QAM symbols using two RF chains. Similarly, multiplexing associated with 8 bps/Hz is obtained by transmitting two 16-QAM symbols. Thus, Fig. 5.4 was obtained for different transmission modes by plotting the *instantaneous BER versus instantaneous SNR using (5.3), as opposed to the average BER versus average SNR portrayed in* Fig. 5.2.

[2]Div denotes diversity, while Mux denotes multiplexing.

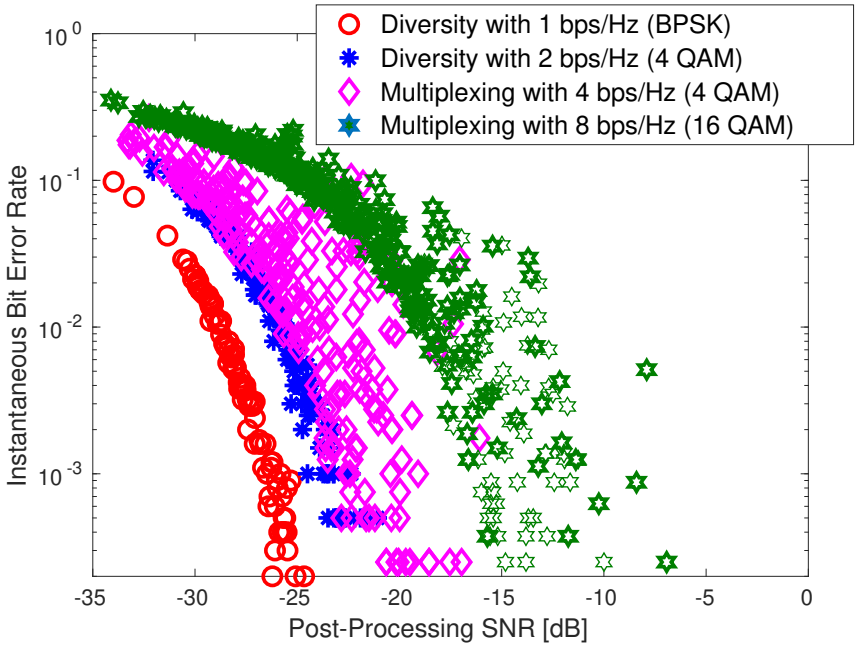


Figure 5.4: Instantaneous BER of different transmission schemes.

It is worth observing from Fig. 5.4 that the selection of the class from set C based on the average threshold values would result in low rates because of the wide-ranging scattering of the instantaneous post-processing SNR values. It can clearly be seen in Fig. 5.4 that the boundaries are not hard, which makes it difficult to decide the choice of the class based on the threshold values of the conventional link-adaptation. During the testing phase, upon estimating the channel state information, the receiver calculates the post-processing SNR assuming that spatial multiplexing is used relying on (5.3) and then finds the K-nearest neighbors with the aid of the post-processed SNR calculated and the required BER, followed by selecting the class through majority voting.

Remark 5.2. When two different color points of Fig. 5.4 overlap, which means that the instantaneous BERs of the two schemes are similar, the specific transmission scheme having the higher rate is selected.

Remark 5.3. Unlike the parametric algorithms discussed in Sec. 2.7, where the parameters such as weights and biases are designed, the KNN learning algorithm invoked in this chapter is a non-parametric classification method [203], which does not depend neither on weight/bias learning parameters, nor on communication parameters such bandwidth, or sampling rate. The advantage of choosing a non-parametric learning algorithm is that it need not have any functional mapping between the input and output samples. To expound a little further, it only considers the post-processing SNR, as well as the target BER as the feature sets and then decides upon the specific class having the higher rate based on majority vote, while satisfying the target BER for the post-processing SNR observed.

5.2.3 Complexity

In this section, we present a qualitative discussion on the KNN algorithm's complexity in practical implementations. There are two significant components of the KNN, which dominate the complexity, namely the search complexity and the memory required for saving data points. Given the advances in the storage capabilities of the devices, the latter may be of less concern than the former. Hence, we focus our discussions on the search complexity. Let us assume that we have n training samples in d dimensions. Then a brute-force KNN search would have the search complexity of $O(kdn)$ for K nearest-neighbors. Although in our chapter we only deal with the BER, rate and post-processing SNR, the brute-force search across the design-space of these three parameters results in a high complexity for higher dimensions. Therefore, this method is not effective as it does not exploit the structure of the training data. However, there is a vast body of literature on fast-KNN techniques search [217–219], which restructure the training data into clusters. By restructuring the data into clusters each having $O(\sqrt{n})$ objects in each cluster, the search complexity is reduced to $O(kd\sqrt{n}\log nd + kn)$, which is lower than the brute-force KNN search complexity [218, 219].

On the other hand, in conventional link-adaptation, a BER and rate look-up table is constructed for different average SNR values. Then the SNR after post-processing is compared against the average SNR values and the scheme whose SNR value is higher than the threshold is selected. Although the computational complexity in conventional link-adaptation is lower, the quality-of-the-service of the system is significantly affected.

5.2.4 Performance Results

In this section, we present our simulation results for characterizing the rate achieved by the proposed learning assisted link adaptation for a user k. We performed Monte Carlo simulations for studying the performance gap between the proposed design and the conventional design, where the average is calculated using 100 channel realizations and a total of 1000 symbols are used for each channel realization. Furthermore, in these results, the desired/required BER is set to 10^{-3}. The simulation parameters are listed in Table 5.1.

Fig. 5.5(a) shows the histogram-based probability density function (PDF) versus the average SNR for all the classes in the set C using learning. It can be seen from Fig. 5.5(a) that at -20 dB of post-processed SNR there is a similar probability of transmission for the classes of $\text{Div}_{\text{BPSK}}, \text{Div}_{\text{QPSK}}$. Furthermore, the probability of the class Div_{BPSK} falls gradually as the SNR increases, while the Mux_{QPSK} mode starts to share the probability of being activated with the Div_{QPSK} mode, where the probabilities are obtained by evaluating the relative frequency of each class using the proposed learning assisted link-adaptation based on both the post-processing SNR and on the BER target. Similarly,

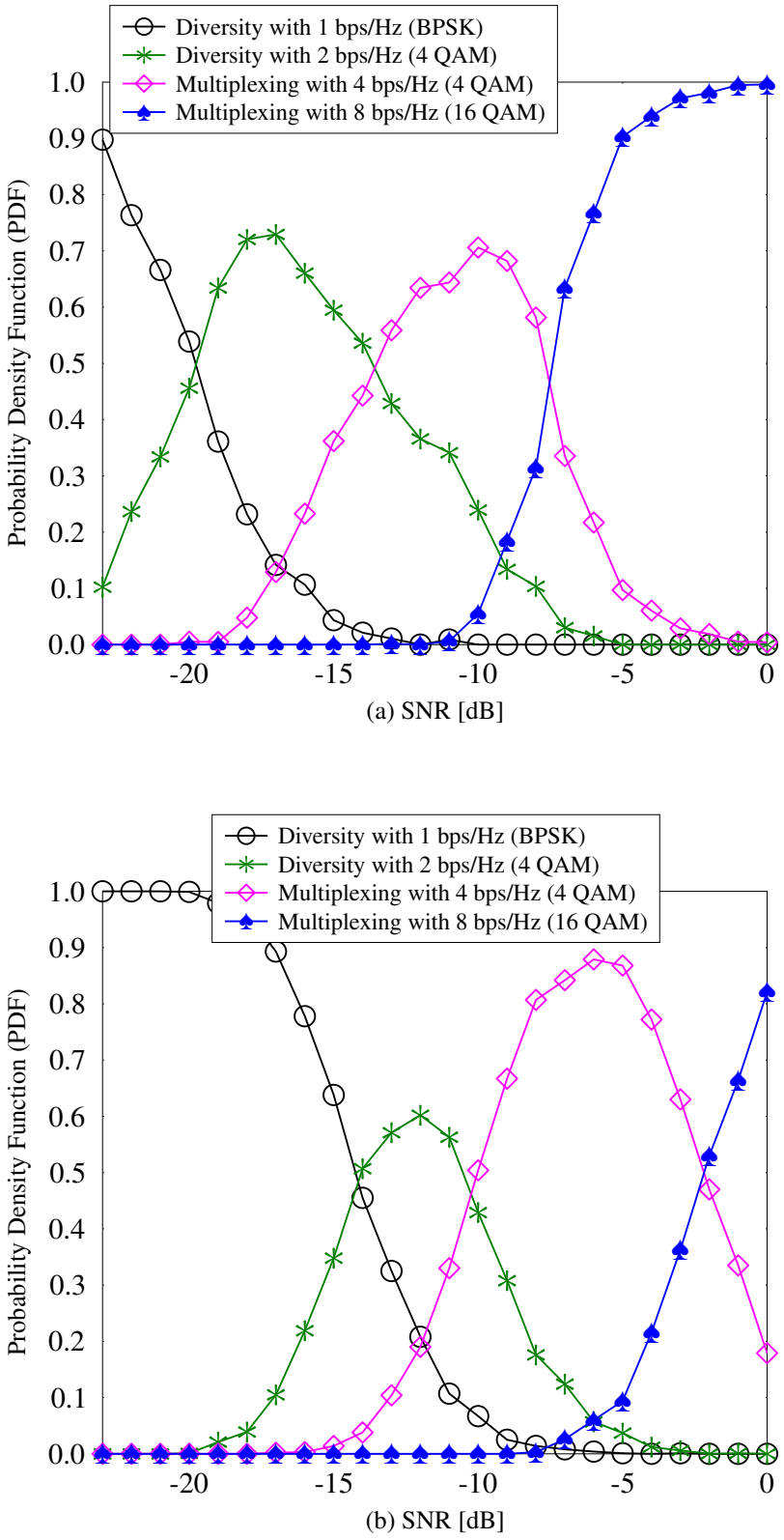


Figure 5.5: Probability density functions for different classes as a function of the average SNR. Simulation parameters used are listed in Table I. (a) Learning assisted link-adaptation. (b) Conventional link-adaptation. In this setting, the channel is NLOS in nature.

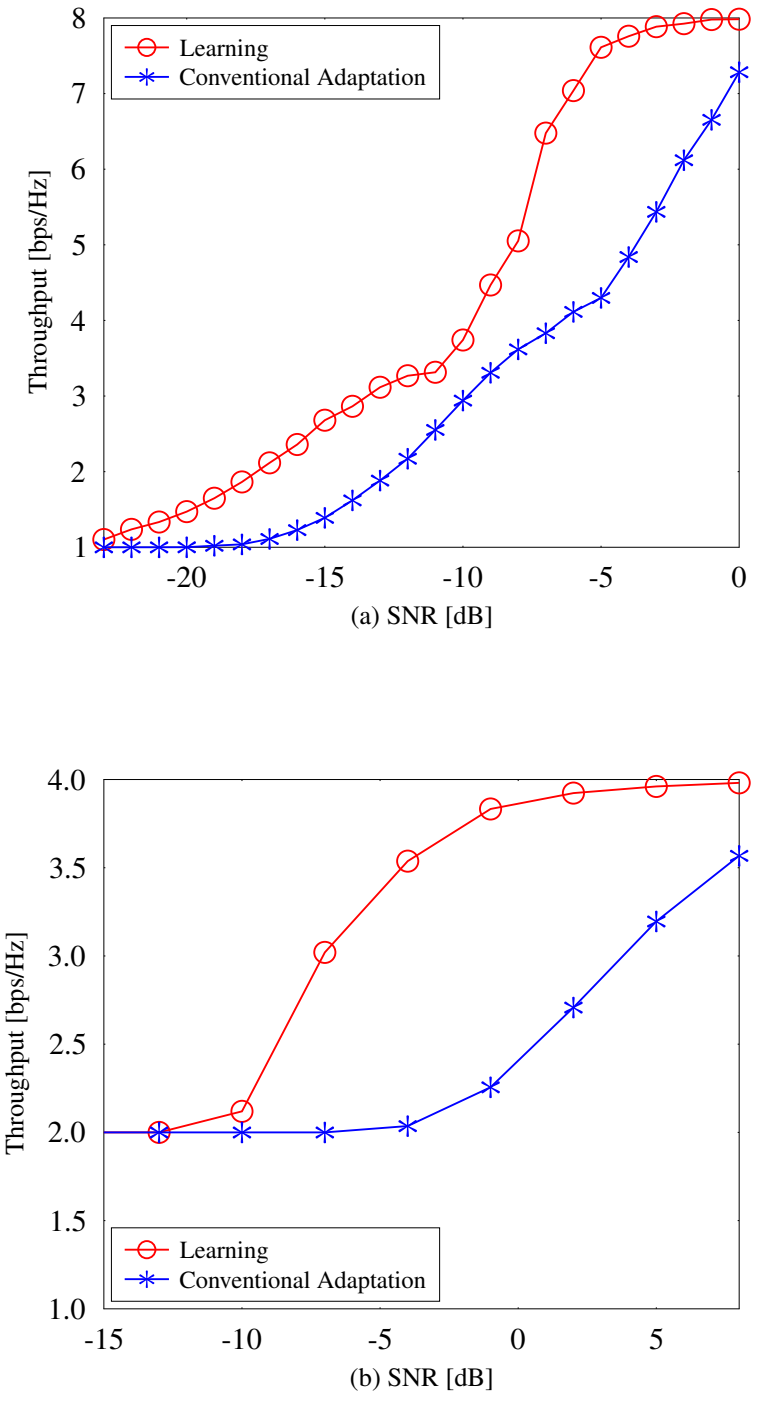


Figure 5.6: Capacity of the proposed design and of the conventional adaptation as a function of the average SNR — (a) NLOS (b) Only one dominant path. Simulation parameters used are listed in Table 5.1.

Table 5.1: Simulation parameters.

Parameters	*Values*
N_t	64
N_r	32
N_t^{RF}	2
N_r^{RF}	2
N_s	1, 2
K	20
$\alpha_{n_c}^{n_{\mathrm{ray}}}$	$\mathcal{CN}(0,1)$
$\phi_{n_c}^{n_{\mathrm{ray}}}$	Laplacian distributed
$\theta_{n_c}^{n_{\mathrm{ray}}}$	Laplacian distributed
BER	10^{-3}

the probability of choosing the $\mathrm{Mux}_{\mathrm{QPSK}}$ mode increases at high SNRs. By contrast, the conventional link-adaptation shown in Fig. 5.5(b) has a very low probability of selecting the class $\mathrm{Div}_{\mathrm{QPSK}}$ in the SNR region $-20 < \mathrm{SNR}\ [\mathrm{dB}] < -15$. Similarly, $\mathrm{Mux}_{\mathrm{QPSK}}$ has a low probability, when we have $-10 < \mathrm{SNR}\ [\mathrm{dB}] < -3$ as opposed to Fig. 5.5(a), where the decision is made based on the pre-defined threshold values.

Fig. 5.6(a) shows the throughput of the system for both the proposed as well as for the conventional link-adaptations. The curves in Fig. 5.6(a) are obtained by calculating the average number of bits transmitted per channel realization using learning and conventional link-adaptations, respectively. It is readily seen that the proposed learning assisted link-adaptation achieves a superior throughput and the throughput difference is substantial, especially at the border lines, during the transition between adjacent classes. In Fig. 5.6(a), for example, it can be clearly seen that in the high-SNR region the proposed design outperforms the conventional adaptation[3] by about 5 dB.

On the other hand, as an example we have also shown in Fig. 5.6(b) the throughput of the system when the channel has only a single dominant path, where the adaptation takes place between the QPSK and 16QAM schemes. It can be seen from Fig. 5.6(b) that the learning-aided design achieves superior performance over the conventional design.

Fig. 5.7(a) shows the BER performance of both the proposed design and of the conventional design as a function of the average SNR. It can be observed from Fig. 5.7(a) that the proposed design meets the target BER of 10^{-3} whilst providing a higher data rate, as shown in Fig. 5.6 (a). Although the conventional adaptation typically provides a lower BER than the target, it fails to reach the highest data rate possible. Similarly, Fig. 5.7(b)

[3] In conventional adaptation, the threshold is selected based on the target BER. In our simulations, we have set a target BER of 10^{-3}; then the SNR versus rate look-up table is constructed by using Fig. 5.2 (b).

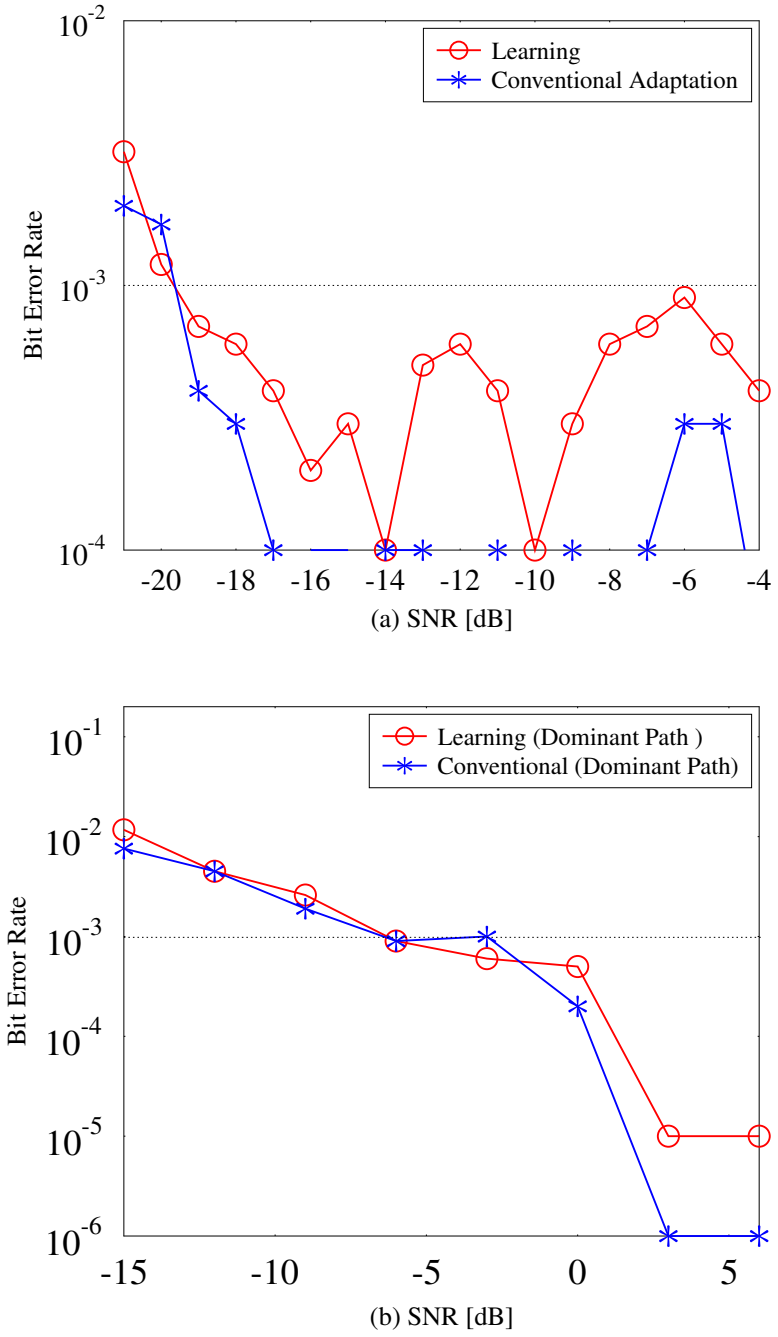


Figure 5.7: BER performance of the proposed design and of the conventional design as a function of the average SNR —(a) NLOS (b) Only one dominant path. Simulation parameters used are listed in Table 5.1.

shows the BER performance when the channel has only a single dominant path, where the learning assisted design meets the target BER of 10^{-3}, whilst also providing a higher data rate, as shown in Fig. 5.6(b).

Remark 5.4. The KNN algorithm's superior performance can be attributed to the learning strategy, which is invoked offline. Furthermore, the KNN algorithm also records the post-processing SNR, the instantaneous BER and the rate as a data point during its operation transmission and updates its data points. These data points are further used for decision-making during the next transmission time slot. These attributes make the KNN algorithm aided system superior to the conventional technique.

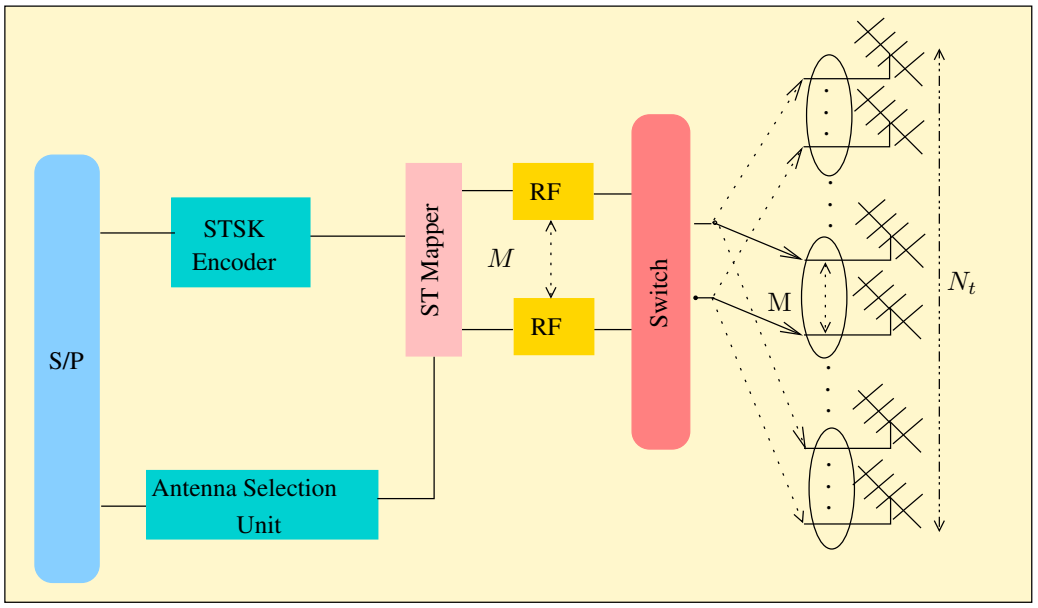


Figure 5.8: Block diagram of the MS-STSK encoder.

Having discussed the learning-aided link-adaptation in mmWave hybrid systems using a non-parametric machine learning tool, we now focus our attention on the learning-aided detection in multifunctional mmWave systems whose receiver complexity is typically high.

5.3 Learning-Aided Hard-Detection for mmWave MS-STSK Systems

In previous chapters we have analyzed the dual-function transceiver designs. More explicitly, in Chapter 3 we have studied the transceiver design that can provide both diversity and beamforming gains, while in Chapter we advocated a beamforming-assisted transceiver design that can simultaneously transmit and receive data. In this Section, we extend our design to multifunctional MIMO system with HBF, where we invoke NN network for detection. Let us consider the system model shown in Fig. 5.8, where the

transmitter is equipped with N_t AA [4] of K AE each. In Fig. 5.8, the transmitter employs an MS-STSK scheme, where the information is conveyed by both the STSK symbols and antenna combination (AC) information. In our system model of Fig. 5.8, the MS-STSK scheme relies on using M AAs (RF chains), where the AC selection is performed by selecting M AA out of N_t AA. More explicitly, the MS-STSK codeword is comprised of two parts, where the first part conveys $\log_2(M_c M_Q)$ bits, with M_c being the constellation size and M_Q is the number of dispersion matrices. The remaining $\log_2(N_t/M)$ bits are mapped to a specific AC in the antenna selection unit of Fig. 5.8. It is important to emphasize that during the MS-STSK transmission only M AA are activated at any symbol interval, while the other antennas remain silent.

The output of a typical STSK encoder is given by

$$\mathbf{X} = \mathbf{A}_q x_l, \tag{5.7}$$

where x_l is the M-QAM/PSK symbol, and $\mathbf{A}_q$ is the q^{th} dispersion matrix of size $M \times T$ from the set $\mathcal{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_q, \ldots, \mathbf{A}_{M_Q}\}$, where $\|\mathbf{A}\|_q^2 = T$. The physical significance of the matrix $\mathbf{A}_q$ is that it disperses the symbol x_l over M AA during T time slots. For example, for a 4-bit-sequence '0110', where the first two bits, '10', are mapped to one of the classic 4-QAM symbols, while the remaining two bits, '01', are mapped to one of the four dispersion matrices from the set $\mathcal{A}$ having the cardinality of 4, i.e. $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}$. It is also possible that the first three bits are mapped to 8-QAM while the last bit is used for the selection of one of two dispersion matrices, depending on the specific design requirements.

Having expounded on the MS-STSK design, in the next subsections, we focus our attention on the system model of the MS-STSK design combined with beamforming followed by the description of amalgamating MS-STSK with the BIM concept.

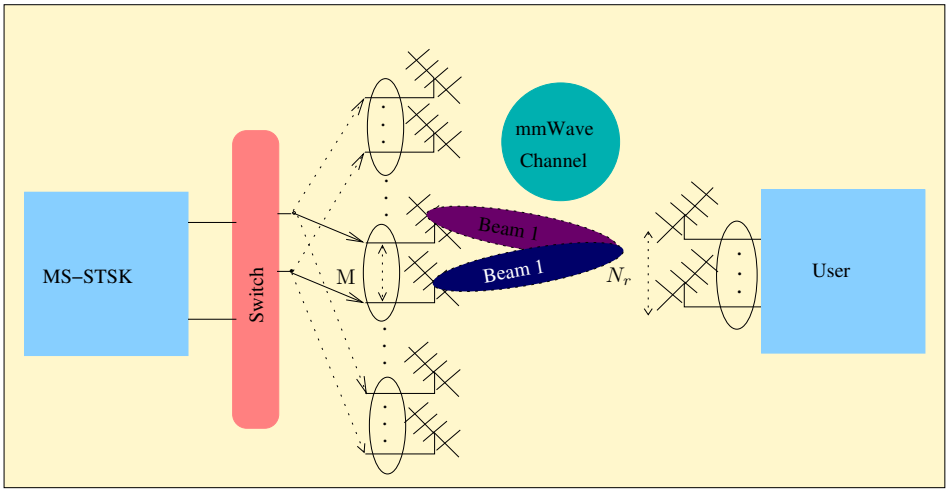
5.3.1 MS-STSK Combined with Beamforming

It is important to emphasize that (5.7) represents the STSK symbol associated with a specific combination of active AA. However, since the transmitter is equipped with N_t AA, the total number of ACs (N_c) for the STSK symbol transmission is

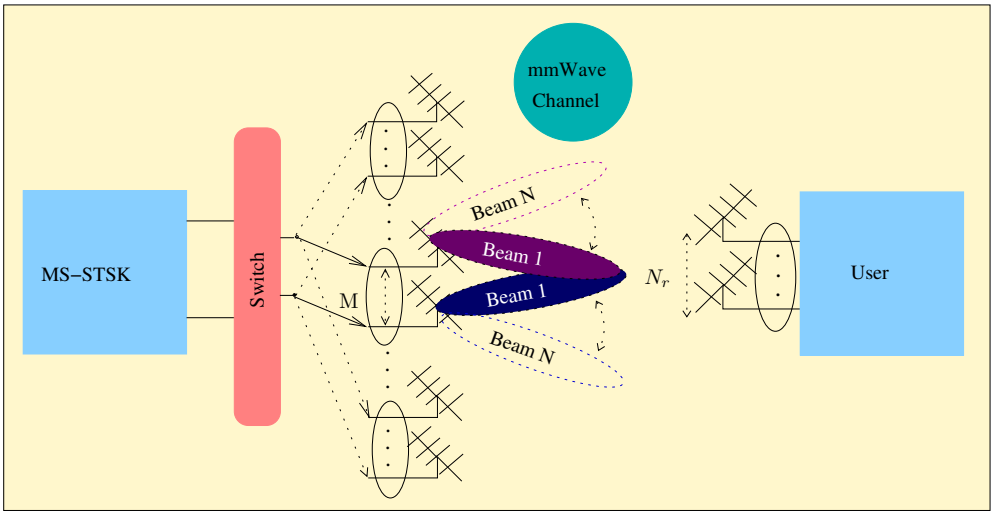
$$N_c = 2^k, k = \lfloor \log_2 N_t/M \rfloor . \tag{5.8}$$

The N_c value is rounded down using the floor function to allow only an integer number of bits. Note that the number of ACs (N_c) is calculated by combining the set of M consecutive antennas together.

[4] The transmitter is equipped with a beamforming AA to derive for compensating the propagation loss at mmWave frequencies.



(a) Beamformed MS-STSK symbol in the direction of the desired user. In this design, a total of $\log_2(N_c M_c M_Q)$ bits are mapped to one of the classic M_c-QAM symbols, one of the M_Q dispersion matrices and one of the N_c ACs.



(b) Beamformed MS-STSK symbol coupled with the beam index in the direction of the desired user. In this design, a total of $\log_2(N_b N_c M_c M_Q)$ bits are mapped to one of the classic M_c-QAM symbols, one of the M_Q dispersion matrices, one of the N_c ACs and one of the N_b beams.

Figure 5.9: MS-STSK symbol transmission - (a) when the channel supports only one beam (b) amalgamated with beam index when the channel supports a plurality of beams.

During transmission, an MS-STSK symbol is formed when an STSK symbol is fed to the ST mapper of Fig. 5.8, where a specific AC is selected depending upon the input bit-sequence. In other words, a part of the input bit-sequence determines the specific AC to be selected for transmission. To expound a little further, let us again consider the bit-sequence where two additional bits are appended to the left of the aforementioned bits, i.e., '110110'. In this scenario, the bits '11' convey the index of one of four ACs[5]. Thus, in this design we have a total of $\log_2(N_c M_c M_Q)$ buts/channel use (bpcu), where the additional $\log_2(N_c)$ bits pass the information of the AC. Therefore, an MS-STSK symbol formed at the output of the ST mapper of Fig. 5.9 from the c^{th} AC can be

[5]In this setting, the total number of AC assumed is 4.

expressed as

$$\tilde{\mathbf{X}} = \tilde{\mathbf{A}}_{q,c} s_l, \tag{5.9}$$

where $\tilde{\mathbf{A}}_{q,c}$ is the MS-STSK dispersion matrix whose entries are constitute by the selected AC dispersion matrix and it is given by

$$\tilde{\mathbf{A}}_{q,c} = [\mathbf{0} \dots \mathbf{A}_q \dots \mathbf{0}]^T. \tag{5.10}$$

Then the MS-STSK symbol is steered in the desired direction for transmission over the mmWave channel using the RF analog RF BF matrix $\mathbf{F}_{\text{RF}}$ to the desired user, as shown in Fig. 5.9(a), where the block-based received signal Y of size $N_r \times T$ after RF analog combining using $\mathbf{W}_{\text{RF}}$ matrix is given by

$$\mathbf{Y} = \mathbf{W}_{\text{RF}} \mathbf{H} \mathbf{F}_{\text{RF}} \tilde{\mathbf{X}} + \mathbf{V}, \tag{5.11}$$

where $\mathbf{V}$ is the Gaussian noise having the distribution of $\mathcal{CN}(0, \sigma^2)$, while $\mathbf{F}_{\text{RF}}$ is expressed as

$$\mathbf{F}_{\text{RF}} = [\mathbf{O} \dots \mathbf{F}_{\text{RF}_q} \dots \mathbf{O}] \in \mathcal{C}^{KN_t \times N_t}, \tag{5.12}$$

$$\mathbf{F}_{\text{RF}_q} = \text{diag}(\mathbf{F}^1_{\text{RF}} \mathbf{F}^2_{\text{RF}} \dots \mathbf{F}^M_{\text{RF}}), \tag{5.13}$$

where $\mathbf{F}^i_{\text{RF}}$ is the BF vector of the i^{th} AA of size $K \times 1$. Similarly, $\mathbf{W}_{\text{RF}}$ is the analog RF combining matrix of size $M \times N_r$. Furthermore, $\mathbf{H}$ represents the statistical mmWave channel model expressed as

$$\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_{N_t}], \tag{5.14}$$

where $\mathbf{H}_i$ is a statistical channel matrix of size $N_r \times K$, which is expressed as

$$\mathbf{H}_i = \sqrt{\frac{N_r N_t}{N_c N_{\text{ray}}}} \sum_{n_c=1}^{N_c} \sum_{n_{\text{ray}}=1}^{N_{\text{ray}}} \alpha_{n_c}^{n_{\text{ray}}} \mathbf{a}_r(\phi_{n_c}^{n_{\text{ray}}}) \mathbf{a}_t^T(\phi_{n_c}^{n_{\text{ray}}}). \tag{5.15}$$

The variables N_c, N_{ray} in (5.15) are the number of clusters and number of rays, respectively, while α obeys the distribution of $\mathcal{NC}(0,1)$, and $\mathbf{a}_i$ as well as $\mathbf{a}_r$ represent the array response vectors at the i^{th} AA of the transmitter and the array at the receiver, respectively.

Note that our system model of (5.11) corresponds to Fig. 5.9(a), where the MS-STSK symbol is transmitted over the channel matrix $\mathbf{H}$ to its intended receiver with the aid of beamforming, where all the beams (one or many) supported by the channel are utilized.

Additionally, by letting $\mathbf{W}_{\text{RF}} \mathbf{H} \mathbf{F}_{\text{RF}} = \mathbf{H}_{\text{eff}}$ and invoking the vectorial stacking operation, Eq. (5.11) becomes equivalent to an SM system, which is detailed in [211]. In other

words, Eq. (5.11) can be re-written in the vector form of [211,213]:

$$\mathbf{y} = \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l} + \tilde{\mathbf{V}}, \tag{5.16}$$

where the vectorized constituent matrices are expressed as

$$\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{N_r T \times 1}, \tag{5.17}$$
$$\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}_{\text{eff}} \in \mathbb{C}^{N_r T \times N_t T}, \tag{5.18}$$
$$\tilde{\mathbf{V}} = \text{vec}(\mathbf{V}) \in \mathbb{C}^{N_r T \times 1}, \tag{5.19}$$
$$\mathcal{X} = [\text{vec}(\tilde{\mathbf{A}}_{1,1}) \dots \text{vec}(\tilde{\mathbf{A}}_{q,c}) \dots \text{vec}(\tilde{\mathbf{A}}_{q,N_c})] \tag{5.20}$$
$$\in \mathbb{C}^{KN_tT \times N_c M_Q}, \tag{5.21}$$
$$\mathbf{K} = [\underbrace{0 \dots 0}_{q-1} s_l \underbrace{0 \dots 0}_{M_Q - q}]^T. \tag{5.22}$$

At the receiver, the vectorized received signal $\mathbf{y}$ is used during the detection process. Conventionally, the detection of the MS-STSK symbol, where the estimates $\hat{q}.\hat{l}, \hat{c}$ of (5.11) are obtained, is carried out by employing ML detection relying on the CSI estimated at the receiver and it is expressed as

$$< \hat{q}, \hat{l}, \hat{c} >= \arg\min_{q,l,c} \|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2. \tag{5.23}$$

Additionally, the DCMC capacity of the MS-STSK scheme is given by [220]

$$C_{\text{DCMC}}^{\text{MS-STSK}} = \log_2(N_c M_c M_Q) - \frac{1}{(N_c M_c M_Q)} \sum\nolimits_{q,l,c} \mathbb{E}\left[\log_2 \sum\nolimits_{q',l',c'} \exp(\psi_{\text{MS-STSK}} \middle| \mathbf{K}\right], \tag{5.24}$$

where

$$\psi_{\text{MS-STSK}} = -\frac{\left\|\mathbf{H}\mathcal{X}\left(\mathcal{I}_{c'}\mathbf{K}_{q,l} - \mathcal{I}_{c'}\mathbf{K}_{q',l'}\right) + \tilde{\mathbf{V}}\right\|^2 - \|\tilde{\mathbf{V}}\|^2}{\sigma^2}. \tag{5.25}$$

By taking into account the pilot overhead f_p, which is the ratio of the number of pilots (N_p) to the total number of symbols in the block, the effective DCMC capacity becomes:

$$C_{\text{eDCMC}}^{\text{MS-STSK}} = (1 - f_p) C_{\text{DCMC}}^{\text{MS-STSK}},$$

For N_d number of symbols and N_p number of pilots in a frame, f_p is expressed as

$$f_p = \frac{N_p}{N_d + N_p}. \tag{5.26}$$

Furthermore, when the minimum mean squared error (MMSE) channel estimate is considered, the channel estimate error variance (σ_h^2) for a total transmission signal power ρ_t is given by [221]

$$\sigma_h^2 = \frac{1}{1 + 100\rho_t f_p}. \tag{5.27}$$

5.3.2 MS-STSK Amalgamated with BIM

In contrast to Fig. 5.9(a), Fig. 5.9(b) shows the MS-STSK symbol coupled with the beam index before the final transmission. In this model of Fig. 5.9(b) information is also conveyed by the index of the beam in addition to the information conveyed by MS-STSK symbol. More explicitly, when the channel of (5.14) supports a plurality of beams, say N_b beams, instead of transmitting in all beams at once, the transmitter selects a specific beam for transmission depending upon the input bit-sequence. Thus, this design is capable of achieving an additional bit rate of $\log_2(N_b)$ bits per channel use than that of its counterpart dispensing with BIM.

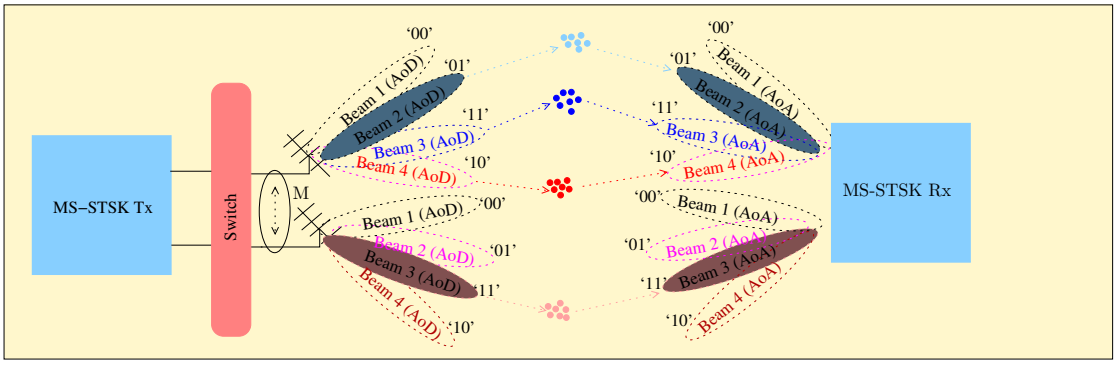


Figure 5.10: Illustration of beam index modulation coupled with MS-STSK.

To elaborate a little further, let us consider the 'toy' example shown in Fig. 5.10, where the channel seen from each AA supports a total of 4 beams for transmission. In order to increase the spectral efficiency, the BIM is invoked by relying on the index of the beam used for transmission. In other words, MS-STSK transmission can be carried out by one of the four beams from each AA by allowing additional bits to be conveyed by the index of the beam. Naturally, this philosophy only can be exploited when there are more than one beams. At this stage the question arises — how do we best configure our N_tK-element antenna for a specific diversity-, spatial multiplexing- and beamforming order? Naturally, this depends on the particular application in mind, as well as on its specifications. In the example of Fig. 5.10, we opted for invoking BIM for implicitly conveying two extra bits instead of the classic space-division multiple access principle, because BF relies on $\lambda/2$-spaced elements; but such a tight element-spacing would result in a modest STSK multiplexing gains, because the adjacent AEs receive correlated signals, which are hard

to separate at the receiver. In a nutshell, the specific assignment of AEs to the baseband signal processing functions has to be carefully considered. If there is only a single beam, then MS-STSK will be combined with conventional beamforming. As an example, in Fig. 5.10, beam 2 is selected for MS-STSK transmission according to the bit sequence '01' representing the beam index, while beam 3 is selected from the other AA for the bits '11'. Thus, the total number of bits per channel use when BIM is coupled with the MS-STSK example discussed in Section 5.3.1 is 10, i.e., '$\underbrace{1101}_{\text{BI}}$ $\underbrace{110110}_{MS-STSK}$.'

Now let us again consider Fig. 5.9(b), where the channel seen from each AA at the transmitter supports N_b beams and only one of the N_b beams is selected for transmission depending on the bit-sequence. Thus there are N_b possible combinations of beamforming for MS-STSK symbol transmission conveying a total of $\log_2(N_b N_c M_c M_Q)$ information bits in contrast to the $\log_2(N_c M_c M_Q)$ bits of the MS-STSK scheme of [211], transmission dispensing with the beam index mode. The block-based received signal Y_{BI} of this scenario can be expressed as

$$\mathbf{Y}_{\mathrm{BI}} = \mathbf{W}_{\mathrm{RF}}^{n} \mathbf{H}_{\mathrm{BI}}^{n} \mathbf{F}_{\mathrm{RF}}^{n} \tilde{\mathbf{X}} + \mathbf{N}, \tag{5.28}$$

where $\mathbf{H}_{\mathrm{BI}}^{n}$ is the statistical channel model of (5.14) in the n^{th} beam, while the sizes of the matrices $\mathbf{H}_{\mathrm{BI}}^{n}, \mathbf{W}_{\mathrm{RF}}^{n}, \mathbf{F}_{\mathrm{RF}}^{n}$ and $\tilde{\mathbf{X}}$ are still the same, as described in Sec. 5.3.1. We observe that the only difference between (5.11) and (5.28) lies in the manner of exploiting the beams.

Similar to (5.16), Eq. (5.28) can be vectorized for the n^{th} beam. At the receiver, the vectorized received signal $\mathbf{y}_{\mathrm{BI}}$ is used during the detection process. In this setting, the detection of the estimates $\hat{q}.\hat{l}, \hat{c}, \hat{n}$ of (5.11) is obtained by employing ML detection on the CSI of the n^{th} beam at the receiver and it is expressed as

$$< \hat{q}, \hat{l}, \hat{c}, \hat{n} >= \arg \min_{q,l,c,n} \|\mathbf{y}_{\mathrm{BI}} - \mathbf{H}_{\mathrm{BI}}^{n} \mathcal{X} \mathcal{I} \mathbf{K}_{q,l}\|^2; \tag{5.29}$$

It is important to emphasize that both (5.23) and (5.29) are heavily reliant on the availability of accurate CSI for the successful detection of the symbols, thereby imposing both the usual pilot overhead required for channel estimation and an additional channel estimation complexity. Furthermore, we will show in the subsequent section that, both (5.23) and (5.29) produce an error floor when the CSI estimate error variance is set to 0.15.

Additionally, the DCMC capacity of the MS-STSK scheme combined with BIM is expressed as

$$C_{\text{DCMC}}^{\text{MS-STSK-BIM}} = \log_2(N_b N_c M_c M_Q) - \frac{1}{(N_b N_c M_c M_Q)} \times \sum_{n,q,l,c} \mathbb{E}\left[\log_2 \sum_{n',q',l',c'} \exp(\psi_{\text{MS-STSK-BIM}}) \middle| \mathbf{K}\right], \tag{5.30}$$

where we have:

$$\psi_{\text{MS-STSK-BIM}} = -\frac{\left\|\mathbf{H}_n \mathcal{X} \mathcal{I}_c \mathbf{K}_{q,l} - \mathbf{H}_{n'} \mathcal{X} \mathcal{I}_{c'} \mathbf{K}_{q',l'} + \tilde{\mathbf{V}}\right\|^2 - \|\tilde{\mathbf{V}}\|^2}{\sigma^2}. \tag{5.31}$$

By taking into account the pilot percentage f_p the effective DCMC capacity becomes:

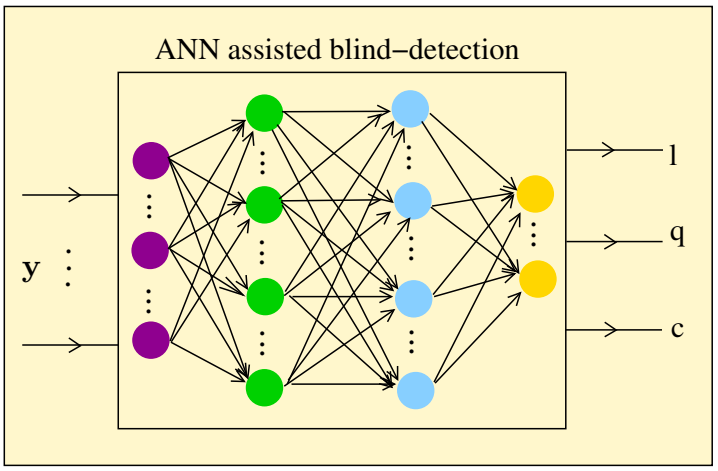
$$C_{\text{eDCMC}}^{\text{MS-STSK-BIM}} = (1 - f_p) C_{\text{DCMC}}^{\text{MS-STSK-BIM}}.$$

In the next section, we propose a design, where we employ blind detection relying on a neural network while dispensing with the requirement of having CSI knowledge at the receiver. The advantage of our design is that it avoids the reliance on CSI and consequently circumventing the pilot overhead and complexity of channel estimation. This philosophy makes our design spectral-efficient.

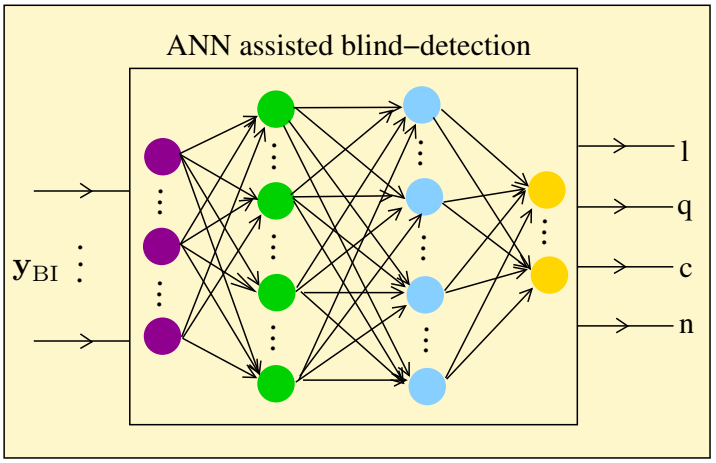
5.3.3 Proposed Semi-Blind Detector Design

Let us now focus our attention on the learning-assisted blind-detection for our system model. As discussed previously in machine learning section in Chapter 2, we first aim for designing the training weights and biases for our neural network. In our design, the number of hidden layers is set to 2, while the number of neurons is adjusted in such a way that it faithfully reproduces the output during the training stage. In our MS-STSK symbol transmission of Eq. (5.16), the vectorized received signal $\mathbf{y}$ serves as the input to the neural network; while the detected dispersion matrix index, the antenna index and the complex-valued classic symbol drawn from the M-QAM constellation constitute the output vector, as shown in Fig. 5.11(a). Similarly, when the MS-STSK encoder is amalgamated with a beam-index, as in Eq. (5.28), the vectorized matrix $\mathbf{y}_{\text{BI}}$ serves as the input of the ANN. In this scenario, we have an additional element, at the output of the ANN which is the beam index,as shown in Fig. 5.11(b).

Having defined the input and the output vectors of the neural network, training of the network is carried out using a set of known input and output samples. We note that before the training process, the weight matrices and the biases vectors of all layers are set to random values from the distribution $\mathcal{N}(0, 1)$ [112]. Furthermore, in our design, a member of the hyperbolic tangent function is used as the activating function as a benefit



(a) Learning assisted detection for MS-STSK transmission.



(b) Learning assisted detection for MS-STSK transmission.

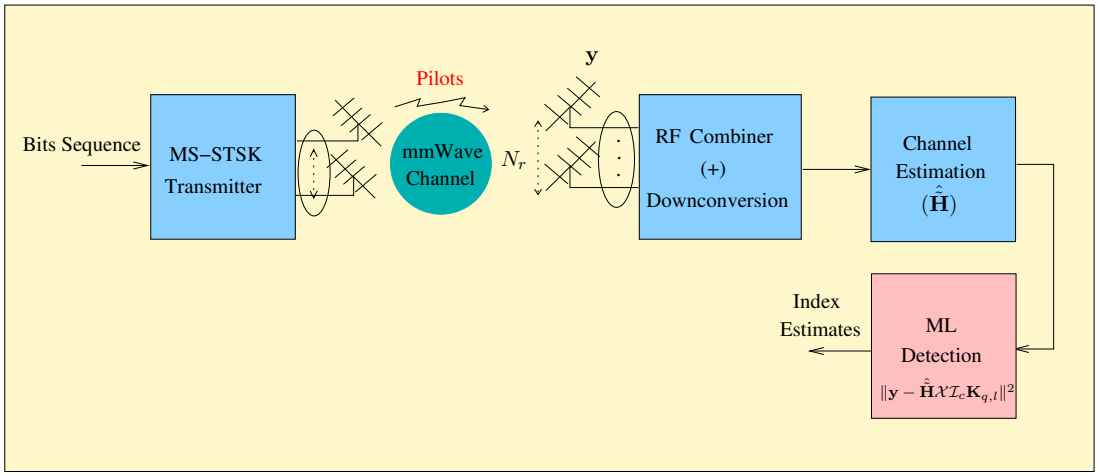
Figure 5.11: ANN assisted blind-detection at the receiver. Since this design is capable of estimating the output while dispensing with the CSI, the pilot overhead is markedly reduced.

of its smoothness and asymptotic properties, which is given as [128]

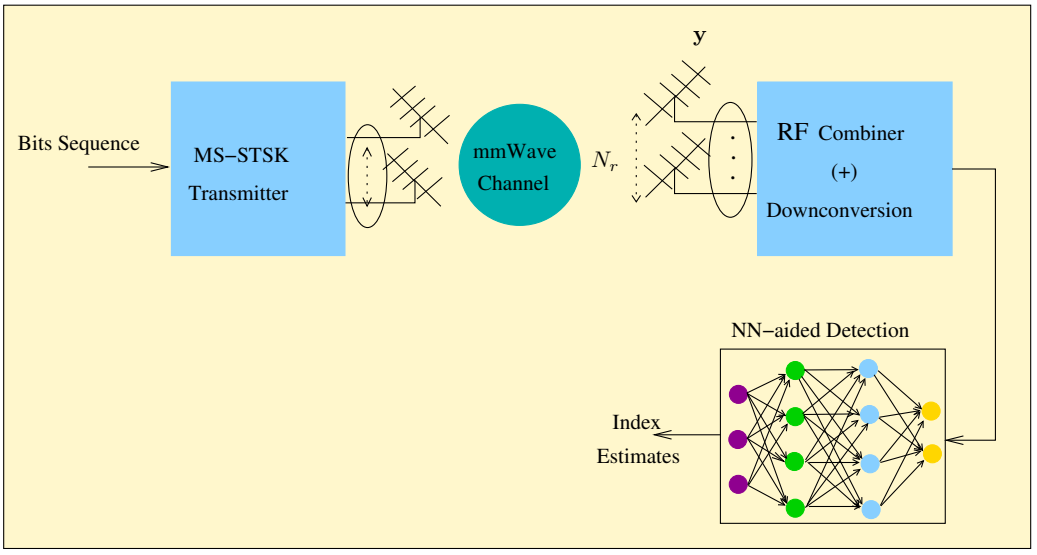
$$f(x,a,v) = \frac{2}{1+e^{-2a(x-v)}} - 1, \tag{5.32}$$

where $f(x,a,v)$ is mapping on x, a is the slope parameter, and v affects the function position.

The activating function of (5.32) is applied at every neuron of the network whose output is fed as the input to the next layer of neurons. In other words, the activating function of (5.32) maps its input vector $\mathbf{y}_i$ of the i^{th} training sample using the weight matrix $\mathbf{W}$ and bias vector $\mathbf{b}$ of that layer. Then this mapping serves as the input vector to its succeeding layer and so forth. After the final mapping, which is at the output layer, the



(a) ML assisted detection for MS-STSK transmission. This design entails having CSI estimation for symbol detection.



(b) Learning assisted detection for MS-STSK transmission. This design dispenses with the CSI estimation for symbol detection.

Figure 5.12: ANN assisted blind-detection at the receiver. Since this design is capable of estimating the output while dispensing with the CSI, the pilot overhead is markedly reduced.

error between the known output and the predicted output is computed. The error is formulated for each layer as a loss function given by [131]

$$L = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\| + \rho_1 \|\mathbf{W}_1\|_2^2 + \rho_2 \|\mathbf{W}_2\|_2^2 + \rho_3 \|\mathbf{W}_3\|_2^2, \tag{5.33}$$

where $\mathcal{S}$ is the cardinality of the training set, $\hat{\mathbf{y}}_i$ and $\mathbf{y}_i$ are the predicted and the known output vector of the i^{th} training sample, respectively, while ρ_1, ρ_2, ρ_3 are the regularization factors used for avoiding over-fitting [131].

Then, the weight matrices and bias vectors are designed for minimizing the loss function of (5.33). This is typically carried out using the technique of back-propagation. In back-propagation, the gradient of (5.33) is evaluated with respect to the weight matrices $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$ and bias vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$. A more detailed explanation of back-propagation is presented by Chauvin *et al.* [132].

Note that in our design we assumed that the channel evolves in time according to Jakes' model, where the channel's correlation coefficient in time is defined by the zero-order Bessel-function of the first kind as [222]

$$\zeta = J_{\circ}(2\pi f_d \tau), \tag{5.34}$$

where f_d is the maximum Doppler frequency and τ is the sample time. Therefore, the number of neurons and the time required for designing the weights and biases during the training phase depend on the Doppler spread[6] $f_d\tau$. Furthermore, the Doppler spread also plays a key role in deciding how often the training of the weights and biases is required for estimating the indices with a high integrity.

After designing the neural network parameters, the testing phase ensues, where the vector $\mathbf{y}$(or $\mathbf{y}_{\mathrm{BI}}$) from the receive AA is fed to the input of the neural network. Here the weights and biases computed during the training phase are applied to the input vector for estimating the indices at the output of the network.

Remark 5.5. The input of the neural network takes only real values; therefore, we have split the received vector $\mathbf{y}$ into real part $\mathcal{R}(\mathbf{y})$ and imaginary part $\mathcal{I}(\mathbf{y})$ before feeding it to the neural network.

5.3.4 Receiver Complexity

In this section, we focus our attention on the receiver design of the proposed model. Fig. 5.12 illustrates the block diagram of a typical receiver employing ML detection as well as the learning-assisted blind-detection. To expound further, Fig. 5.12(a) shows the schematic of ML detection. In this design, the receiver first combines the signal in the RF stage and then performs down-conversion for further digital processing in the baseband. Given the necessity of having the CSI, channel estimation is carried out with the aid of pilots prior to the detection. The ML detection is invoked for estimating the MS-STSK symbol after the CSI estimate is obtained.

By contrast, Fig. 5.12(b) shows the receiver design relying on a neural network, where the training weights are designed offline. It can be seen in Fig. 5.12(b) that, in contrast to Fig. 5.12(a), this design dispenses with the channel estimation stage. More explicitly,

[6] At high Doppler spread, more number of neurons may be required for training.

the signal received after down-conversion is fed to an ANN, where the NN parameters learned during the training phase are applied for estimating the MS-STSK indices.

Having briefly discussed the receiver structure of both designs, let us now focus our attention on the complexity quantified in terms of search space volume and the number of computations. Let us consider again the MS-STSK symbol of (5.11) as a 'toy' example, where there are M_Q dispersion matrices, M_c complex-valued symbols and N_c ACs. In this scenario, the ML detection of (5.23) has to estimate both the index of the dispersion matrix and of the M-QAM symbol, as well as of the AC. Thus, the run-time complexity relying on ML detection would be on the order of $\mathcal{O}(M_c M_Q N_c)$. Furthermore, the ML detection requires CSI knowledge which relies on pilots and imposes additional complexity during the channel estimation stage of Fig. 5.12 (a), while also significantly reducing the data rate because of the pilot overhead.

In contrast to the ML detection, semi-blind learning assisted detection improves the data rate by eliminating the pilot overhead. In other words, once trained the neural network turns a 'blind eye' to the CSI. This philosophy makes our design spectral-efficient. Furthermore, the pre-determined parameters of the network learned during the training process allows us to estimate the indices of (5.11) with a high integrity as we will show later in Section 5.3.5.

On the other hand, the complexity of the proposed design depends on the number of neurons in each hidden layer. More explicitly, the complexity of a typical NN is jointly determined by the forward propagation, and backward propagation. To elaborate a little further, let us assume that there are n neurons in a hidden layer. Let us also assume that the input and output vectors are of sizes n_i and n_o, respectively. We know that for each layer's activating function of (5.32) is computed using the network parameters of the respective layer. In other words, the pre-determined weight matrix and bias vector values are substituted in the activating function of (5.32) relying on the input vector n_i for computing the intermediate output n_o, which serves as the input of the next layer. In (5.32), x is the input, a is the weight and v is the bias. Contrasting it to the ML receiver's search complexity by considering each search operation as a node of Fig. 5.11, the complexity of the proposed design would be on the order of $\mathcal{O}(n_i n_{h_1} n_{h_2} n_o)$. It is important to emphasize that in contrast to the ML of Fig. 5.12(a), this design does not require additional computations for channel estimation and also avoids the pilot overhead.

Let us now delve into the complexity in terms of the number of computations for both designs. The total number of complex multiplications required by ML detection for the transmission parameters of (5.11) is $\mathcal{O}(N_t N_r N_c M_Q T^2) + \mathcal{O}(N_c M_Q^3)$, while it is $\mathcal{O}(N_b N_t N_r N_c M_Q T^2) + \mathcal{O}(N_b N_c M_Q^3)$ for (5.28). By contrast, for the NN associated with the aforementioned parameters, the number of multiplications is $\mathcal{O}(n_i n_{h_1}) + \mathcal{O}(n_{h_1} n_{h_2}) + \mathcal{O}(n_{h_2} n_o)$.

Table 5.2 illustrates the number of computations required for (5.23) with the simulation parameters listed in Table. 5.3.

Table 5.2: Quantifying the complexity in terms of number of complex multiplications.

Design	Computations
Learning assisted Blind Detection for (5.11)	330
ML-Aided Detection for (5.11)	384
Learning assisted Blind Detection for (5.28)	1146
ML-Aided Detection for (5.28)	768

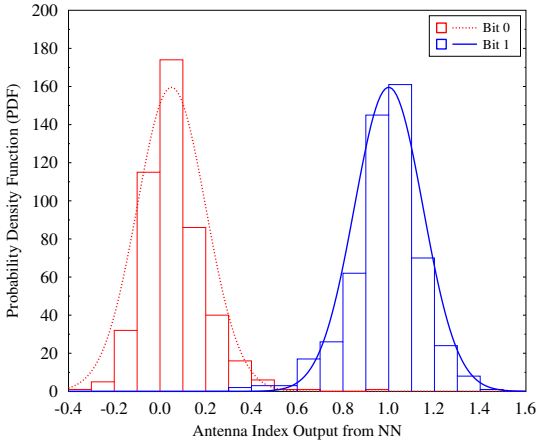
Table 5.3: Simulation Parameters.

Parameters	Values
Number of AEs in each AA at Tx (K)	64
Number of AEs in each AA at Rx (L)	16
Number of AAs at Tx (N_t)	4
Number of AAs at Rx (N_r)	2
Size of QAM symbol (M_c)	4
Number of dispersion matrices M_Q	4
Size of consecutive AA selected (M)	2
Time slots (T)	2
Number of clusters (N_c)	1 & 2
Number of rays (N_{ray})	1
Number of beams (N_b)	2
AoA (ϕ_r)	variable
ϕ_t	variable

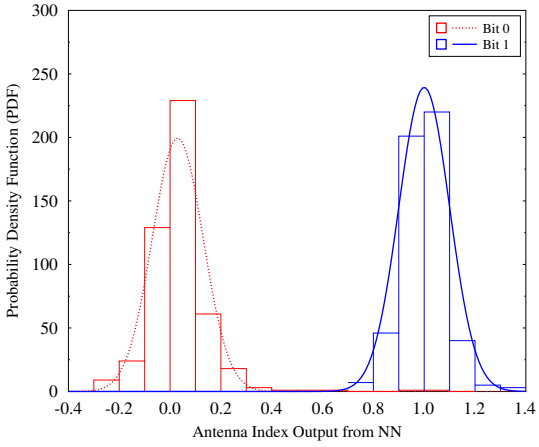
5.3.5 Performance Results

In this section we present our simulations characterizing the performance of the proposed design and of the ML detection. More particularly, we performed Monte Carlo simulations for comparing the performance of the learning-assisted blind detection and of the ML-aided detection. Our simulation parameters are listed in Table 5.5.

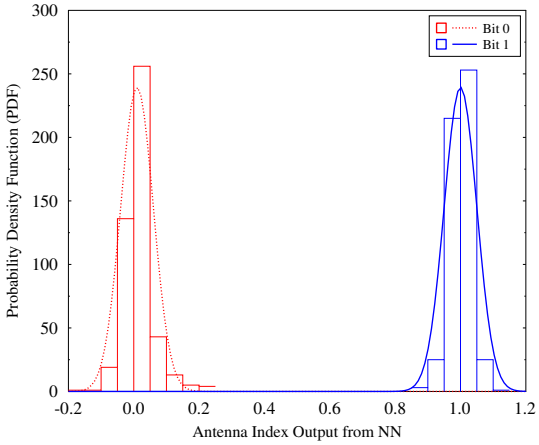
Fig. 5.13 shows the probability density function (PDF) of the antenna index estimate generated from the output of the NN for the system model of Fig. 5.9(a). The plots shown in Fig. 5.23 are generated during the testing phase of the NN for SNRs of $0, 5$ and 15 dB. To obtain this plot, we have set the number of hidden layers to 2, where the NN



(a) The PDF of antenna index estimate at SNR 0 dB.



(b) The PDF of antenna index estimate at SNR 5 dB.



(c) The PDF of antenna index estimate at SNR 15 dB.

Figure 5.13: Probability distribution of the antenna index estimated at the output of the NN for (a) SNR 0 dB (b) SNR 5 dB (c) SNR 15 dB.

underwent training using 2000 samples. Additionally, we note that the input vector of the NN of Fig. 5.11 takes only real values as discussed in Section 5.3.3, hence the received vector y is split into its real and imaginary parts. In this setting, we empirically observed that the NN makes an accurate inference between the output and the input vectors, when the number of neurons is set to 12 and 13 in hidden layers 1 and 2, respectively for both the real and imaginary parts of the network.

In this setting, Fig. 5.13(a) shows the PDF of the antenna index output from the NN for an SNR of 0 dB. Ideally, the estimate of the output representing the antenna index is expected to be either '0' or '1', but it can be seen from the figure that the output of the NN is not exactly binary but a set of continuous values spanning from -0.4-to-1.5. More explicitly, for the bit '0' the output ranges from -0.4-to-0.5; while for the bit '1', the output of the NN is a real value between 0.3-to-1.6. It is observed empirically that the output for the bit '0' follows a near-Gaussian distribution with a mean of 0 and variance of 0.15. Similar trends are valid for the bit '1'. However, as the SNR increases from 0-to-5-to-15, as shown in Figures 5.13(b) and 5.13(c), it may be observed that the variances of the distributions reduce gradually. In other words, the range of values seen for an SNR of 0 dB becomes narrower as the SNR increases, which results in a confident estimate of the antenna index. To understand the behavior of the NN in terms of index estimation, we have analyzed the output of the antenna index, as an example. The PDFs for the other indices can also be readily obtained.

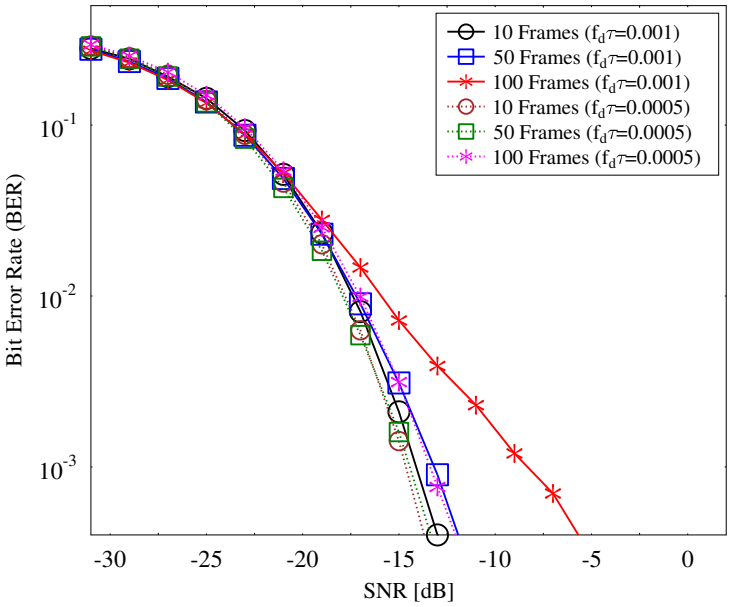


Figure 5.14: BER of the proposed design for different number of frames as channel evolves in time according to Jakes' correlation coefficient (5.34). The simulation parameters are listed in Table 5.5.

Fig. 5.14 analyzes the BER of the proposed design for different number of frames as channel evolves in time according to Jakes' correlation coefficient of (5.34). Note that

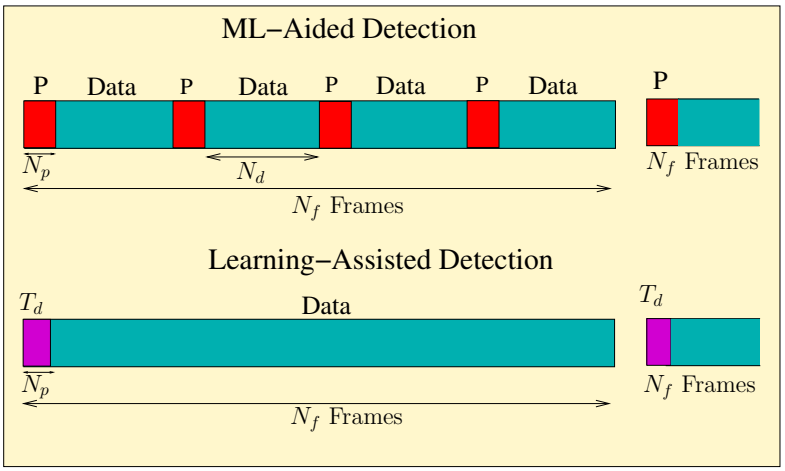


Figure 5.15: Pilots transmission for the proposed learning-assisted design and conventional design. In ML-aided pilots P are transmitted in every frame which significantly affects the spectral-efficiency, while training data T_d is requested by the user only after N_f frames which is contingent on the Doppler spread.

here we define frame as a block comprising of both data and pilot symbols, as shown in the Fig. 5.15. If the number of pilots in a frame is zero, then the the frame consist of only data symbols. The number of frames that can be transmitted before retraining of the NN depends on the Doppler frequency. More particularly,it is evident from Fig. 5.14 that as the number of frames increases from 10-to-100, the BER of the proposed design degrades. This phenomenon is observed because the training weights designed during the first few frames become outdated after a certain number of frames; hence retraining the NN parameters becomes necessary. We note that as expected, the number of frames transmitted before the NN weights become outdated directly depends on the Doppler spread. For example, in Fig. 5.14, the number of frames transmitted before the BER starts to degrade for the normalized Doppler spread of 0.0005 is higher than for 0.001. Therefore, in this scenario, the receiver requests the BS to transmit pilots to recalibrate its weights depending on the BER observed. Fig. 5.15 shows the schematic of the pilot transmission for both the learning-assisted design and ML detection. It is important to emphasize that the pilots are transmitted for every single frame in the case of ML detection. By contrast, our learning-assisted detection requires the training data for recalibrating the NN weights only after every N_f frames, as shown in Fig. 5.15, while performing blind-detection in the rest of the frames. This can also be interpreted as online learning. Thus our system is semi-blind, since it requires recalibration of its weights relying on the training data, while employing blind-detection thereafter. It is of salient importance to note that the amount of training data required for recalibration decreases with the reduction of the Doppler spread. The overhead involved in retraining the NN network of our design is proportional to $N_p/\left[N_f(N_d+N_p)\right]$, where N_f is the number of frames, N_d is the number of data streams, and N_p is the number of pilots; while the pilot overhead involved in the channel estimation for ML detection is $N_pN_f/\left[N_f(N_d+N_p)\right]$.

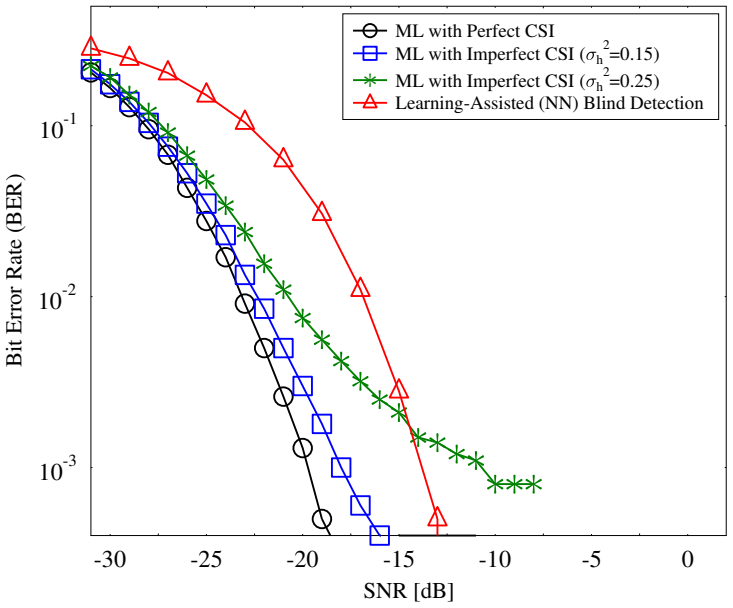


Figure 5.16: BER of the MS-STSK transmission. In this design the NN estimates indices of antenna, dispersion matrix and the symbol. The simulation parameters are listed in Table 5.5.

Fig. 5.16 shows the BER of both the learning-assisted blind detection and of the ML-aided detection with perfect CSI, as well as of the ML detection with imperfect CSI for the MS-STSK transmission dispensing with the BIM. Since no BF index is considered, it can be assumed that the channel supports only a single beam, or all potential beams are utilized for the transmission. It can be seen in Fig. 5.16 that for the aforementioned NN parameters, there is around 6 dB gap at the BER of 10^{-3} between the learning-assisted detection and the ML-aided detection relying on perfect CSI. Although the ML-aided detection relying on perfect CSI outperforms the learning-assisted detection by 6 dB SNR gain, this is achieved under the idealized simplifying assumption of having perfect CSI. On the other hand, for the CSI estimate having an error variance of 0.15, the SNR gap reduces to around 3 dB. Furthermore, the ML-aided detection produces an error floor for the CSI error variance of 0.25. By contrast, the learning-assisted blind detection remains capable of accurately estimating the indices of the MS-STSK transmission regardless of the nature of the CSI, while also circumventing pilot-assisted channel estimation all together.

While Fig. 5.16 shows an SNR gap between the proposed scheme and ML detection, it is also pertinent to study the effective throughput of both designs for the sake of fairness, since the SNR gain observed for ML detection in Fig. 5.16 is critically contingent on the CSI estimation accuracy, which increases proportionally to the pilot density. However,

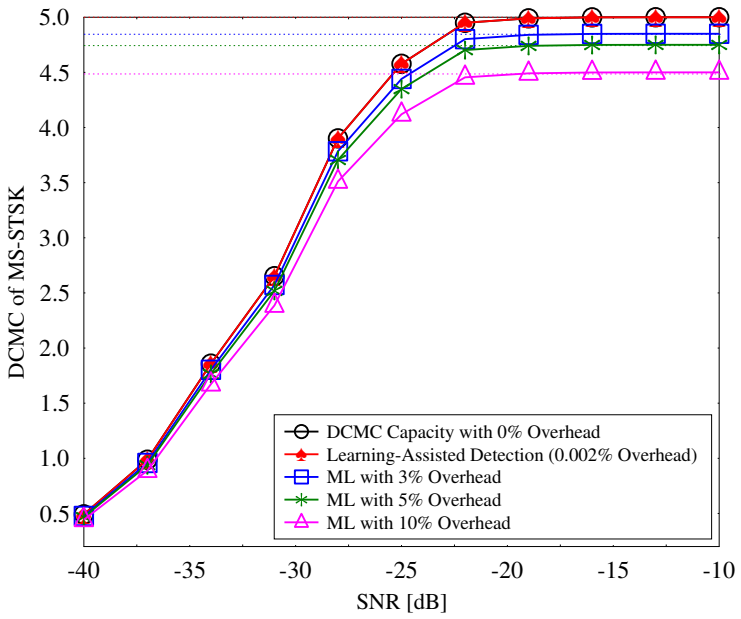


Figure 5.17: Discrete-input Continuous-output Memoryless Channel capacity of the learning-assisted blind detection and of the ML-aided detection with different pilot overheads for channel estimation when MS-STSK transmission with beamforming is considered. The simulation parameters are listed in Table 5.5.

increasing the pilot density would reduce the effective throughput of the design. Therefore, Fig. 5.17 characterizes the DCMC capacity of the learning-assisted blind detection and of the ML detection for different pilot overheads. It can be seen in the figure that the capacity of the design is strictly governed by the pilot overhead. More explicitly, for the simulation parameters summarized in Table 5.5, the capacity of the ML detection for 3% pilot-overhead is limited to a maximum value of 4.85 [bpcu], while it is 4.75 [bpcu] for 5% pilot overhead. Furthermore, when the pilot overhead is increased to 10%, the DCMC capacity of the ML detection is reduced to a maximum value of 4.5 [bpcu]. By contrast, the DCMC capacity of the learning-assisted detection reaches the maximum value of 4.99 [bpcu], which is close to the DCMC capacity of 5 [bpcu], since the overhead involved in recalibrating the weights is marginally lower and detects the MS-STSK symbols blindly. Additionally, we note that there is a SNR gain of 0.5 dB and 1 dB for our learning-assisted detection over the ML-detection with 5% overhead when aiming for achieving a rate of 3 [bpcu] and 4 [bpcu], respectively. This gain is around 1 dB and 3 dB for the learning-assisted detection over the ML-detection with 10% overhead at the rate of 3 [bpcu] and 4 [bpcu], respectively.

Fig. 5.18 shows the BER of both the learning-assisted blind detection and of the ML-aided detection with perfect CSI as well as of the ML-aided detection with imperfect

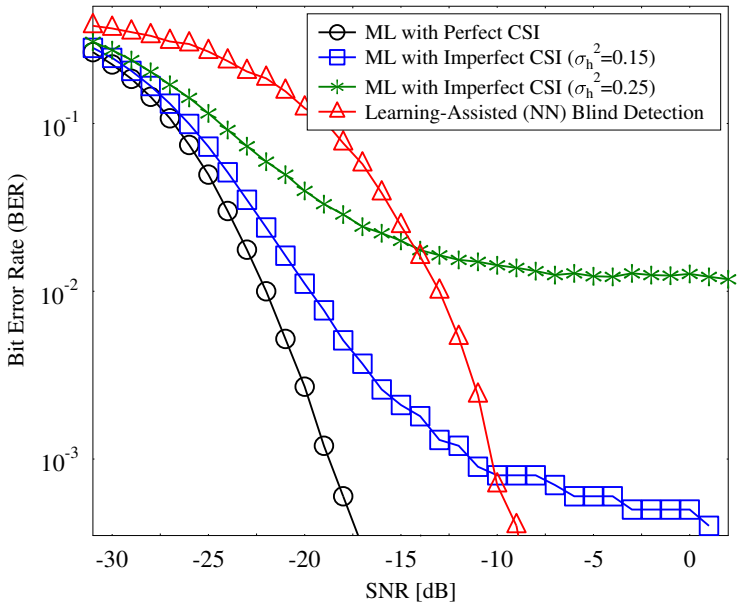


Figure 5.18: BER of the MS-STSK transmission amalgamated with the BIM. The simulation parameters are listed in Table 5.5.

CSI for MS-STSK transmission in conjunction with BIM. In this setting, we assumed that the channel supports two beams for each AA of Fig. 5.9(b). In other words, in Fig. 5.9(b) there is an extra index to be estimated, which is the beam index. In this scenario, it is empirically observed during the training phase that the number of neurons has to be set to 30 for both the real and imaginary constituents of the NN. It can be seen from Fig. 5.18 that adding an additional index for estimation increased the SNR gap between the learning-assisted blind detection and the ML detection using perfect CSI to 8 dB. Again, the superior performance of the ML detection is because of the unrealistic assumption of having perfect CSI. However, when a CSI estimate associated with the error variance of 0.15 is introduced, the ML-aided detection starts to produce an error floor from around -10 dB, while the BER remains flat for the CSI error variance of 0.25. On the other hand, despite the absence of CSI, the learning-assisted detection estimates both the MS-STSK indices and the beam index with high integrity.

Fig. 5.19 characterizes the DCMC capacity of the proposed design and of the ML design when MS-STSK transmission is amalgamated with the BIM. In this simulation, it is assumed that the channel supports two beams and BIM is employed for exploiting these two beams, where only one beam is activated depending on the input bit-sequence. It is evident from the figure that the DCMC capacity of the ML detection is inferior to that of e learning-assisted blind detection. This is because of the overhead imposed by the pilots used for channel estimation to aid the ML detection process. This becomes especially

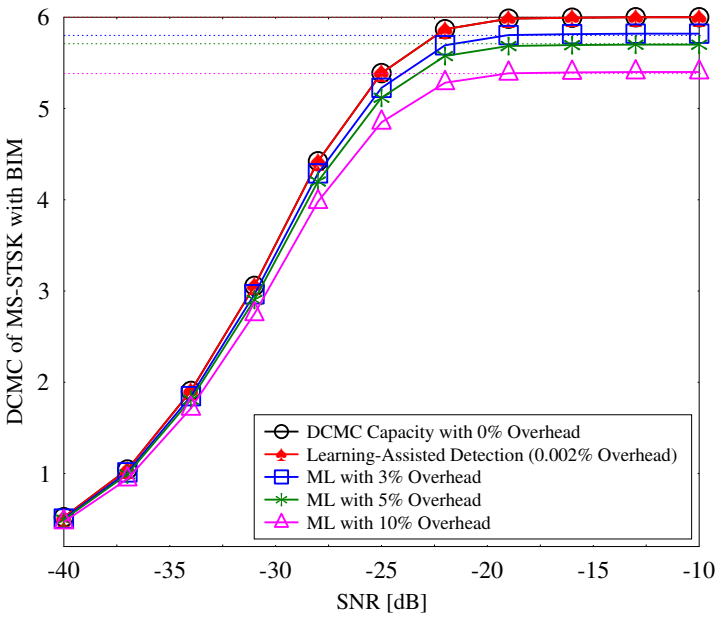


Figure 5.19: Discrete-input Continuous-output Memoryless Channel capacity of the learning-assisted blind detection and of the ML-aided detection with different pilot overheads for channel estimation when MS-STSK transmission with BIM is considered. The simulation parameters are listed in Table 5.5.

pronounced, when the pilot overhead is set to 10% as seen in the figure, where the DCMC capacity is less than 5.5 [bpcu], while that of learning-aided detection is 6 [bpcu]. In other words, the necessity of having pilots for estimating the CSI partly consumes the physical resources, thereby reducing the effective capacity of the system. However, this behavior is avoided by the learning-assisted system, since it achieves accurate symbol detection at a retraining overhead as low as 0.002%. Furthermore, we observe an SNR gain of 3 dB at the rate of 5 [bpcu] for our learning-assisted detection over the ML detection at 10% overhead, while it is around 2 dB at the rate of 4 [bpcu].

In the foregoing section, we discussed learning-aided hard detection for mmWave MS-STSK systems. In the next section, we present soft-detection technique with aid of learning for the system model of (5.11).

5.4 Learning-Aided Soft-Detection for mmWave MS-STSK Systems

In the previous section, we focused our attention on the hard detection of MS-STSK symbols. However, hard detection may lead to erroneous results, since it makes decision

with high confidence whether a bit is '0' or '1'. By contrast, soft detection makes decision by investigating how *likely* is that particular bit '0' or '1'. To elaborate further, it deals with the probabilities of bit being either '0' or '1', where a hypothesis testing is formulated using likelihood ratio test with threshold Θ. It is instructive to note that while dealing with soft detection, bits are encoded using a channel encoder, where the bits are decoded using an appropriate channel decoder at the receiver. The purpose of channel coding is to correct or detect the error in the received bits that may have experienced hostile channel conditions. Therefore, in this section, we commence by outlining conventional soft-detection for mmWave MS-STSK systems with beamforming, followed by exposition of our proposed learning-aided soft-detection.

In the conventional MS-STSK aided transceiver design, soft-decision detection is carried out by computing the log-likelihood ratio (LLR) of the MS-STSK demodulator. The LLR of a bit is defined as the ratio of the probabilities associated with the logical bit '1' and logical bit '0', which is formulated as:

$$L(b) = \log \frac{p(b=1)}{p(b=0)}, \tag{5.35}$$

where $p(b=1)$ and $p(b=0)$ are the probabilities associated with the logical bit '1' and logical bit '0', respectively. The sign of $L(b)$ indicates the logical bit '1' or '0', while the magnitude indicates the confidence in that specific bit.

Then, the probability of receiving the signal $\mathbf{y}$ of (5.16) given that the symbol $\mathbf{K}_{q,l}$ is transmitted from the c^{th} AC is given by

$$p(\mathbf{y}|\mathbf{K}_{q,l,c}) = \frac{1}{(\pi\sigma^2)^{KN_rT}} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2}{\sigma^2}\right). \tag{5.36}$$

On the other hand, the received signal $\mathbf{y}$ conveys the bit sequence $B = [b_1, \ldots, b_{N_b}]$, where $N_b = \log(N_c M_Q M_c)$. Then the LLR of the bit b_i is given by

$$L(b_i) = \log \frac{p(\mathbf{y}|b_i=1)}{p(\mathbf{y}|b_i=0)}, \tag{5.37}$$

where

$$p(\mathbf{y}|b_i=1) = \frac{1}{(\pi\sigma^2)^{KN_rT}} \sum_{k,q,c\in b_i=1} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2}{\sigma^2}\right) \tag{5.38}$$

and

$$p(\mathbf{y}|b_i=0) = \frac{1}{(\pi\sigma^2)^{KN_rT}} \sum_{k,q,c\in b_i=0} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{H}}\mathcal{X}\mathcal{I}_c\mathbf{K}_{q,l}\|^2}{\sigma^2}\right). \tag{5.39}$$

It is important to emphasize that (5.36) relies on the knowledge of CSI at the receiver. This requires pilots for channel estimation, hence reducing the effective data rate [223].

We now discuss the proposed learning-aided soft-decoding. Note that the rationale of using learning over conventional soft-decoding is that the former pilot-assisted channel estimation can be eliminated. In contrast to (5.37), where an exhaustive search is carried out over all the legitimate combinations of the bit being either logical '0' or '1', the learning-aided design provides the soft-LLRs by employing the ANN weights designed during the training phase. In this design, the input training samples of the ANN are the received signal vectors $\mathbf{y}$, while the output labels are the LLRs, as shown in Fig. 5.20. Then, the ANN is trained to infer the functional mapping between the input and output samples. However, since the received signal vectors are affected by noise, the ANN may fail to accurately infer the function. Therefore, the choice of the SNR during the training is crucial, which can only be obtained empirically by varying the SNR[7].

For a given SNR, the ANN predicts the LLR value by employing the AF $f(v)$ in each layer of the network as shown in Fig. 5.21, where the input of each AF is the output of the preceding layer. In this paper, we opted for the sigmoid function as the AF, as a benefit of its smoothness which is formulated as

$$f(v) = \frac{1}{1+e^{-v}}. \tag{5.40}$$

For example, in Fig. 5.21, the input of the first neuron in the second layer is $v_1 = w_{11}y_1 + w_{21}y_2 + w_{31}y_3 + b_1$. Then the AF of (5.40) is applied to v_1 to obtain its mapping, which serves as one of the inputs to the next layer. In this way, the output of the AF of each layer is then fed to its subsequent layer as shown in Fig. 5.20. This process is carried on until the output layer is reached, where the final predicted values are obtained. Note that initially the weights of each layer are assigned to random values obeying the distribution $\mathcal{N}(0,1)$. These weights are then updated for ensuring that they minimize the error between the predicted LLR and the actual LLR. Mathematically, it is formulated by a loss function (LF) given by

$$\text{LF} = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \left\| \hat{\mathbf{L}}_i - \mathbf{L}_i \right\| + \rho_1 \|\mathbf{W}_1\|_2^2 + \rho_2 \|\mathbf{W}_2\|_2^2 + \rho_3 \|\mathbf{W}_3\|_2^2, \tag{5.41}$$

where $\mathcal{S}$ is the cardinality of the training set, $\hat{\mathbf{L}}_i$ and $\mathbf{L}_i$ are the predicted and the known LLR value, respectively, of the i^{th} training sample, while ρ_1, ρ_2, ρ_3 are the regularization factors used for avoiding over-fitting [131].

To minimize the loss function of (5.41), the gradient of the loss with respect to the weights is computed and used for updating the weight values in a gradient descent procedure known as back-propagation. While only a local —rather than global— minimum is

[7] There is a temptation to train the network at high SNRs, where the input samples are noise-free. However, using noiseless training data may result in performance degradation under realistic conditions.

ensured, the procedure has been shown to be practical. A more detailed discourse on back-propagation is presented by Chauvin *et al.* in [132]. These weights, which are learned during the training phase are then stored in memory and are invoked during the testing phase. In other words, the ANN predicts the LLR value from the received signal vector $\mathbf{y}$ by employing the pre-determined weights. Thus, this design does not depend on the channel knowledge to obtain the LLR values. It is instructive to note that during the training of ANN weights, we assume the AoAs and AoDs of the channel matrix to be time-invarying, while the small-scale fading coefficient is assumed to evolve in time according to Jakes' model.

The soft-LLR values predicted from the MS-STSK's ANN demodulator are then passed to the turbo channel decoder. In the next section, we examine the complexity of both the traditional MS-STSK receiver and the learning-aided MS-STSK receiver.

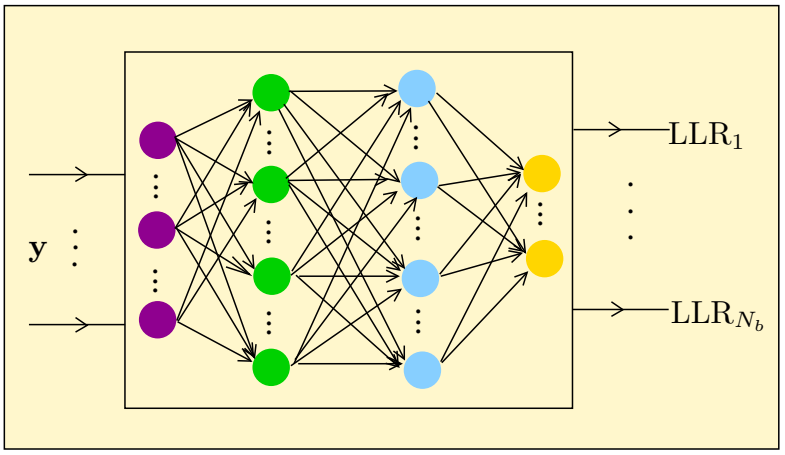


Figure 5.20: The input and output of the ANN proposed.

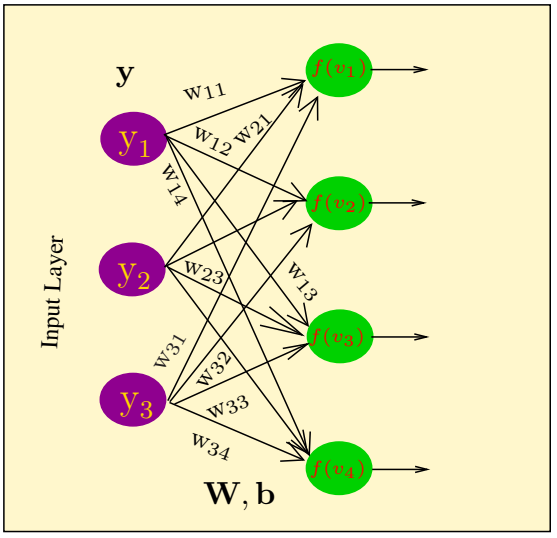
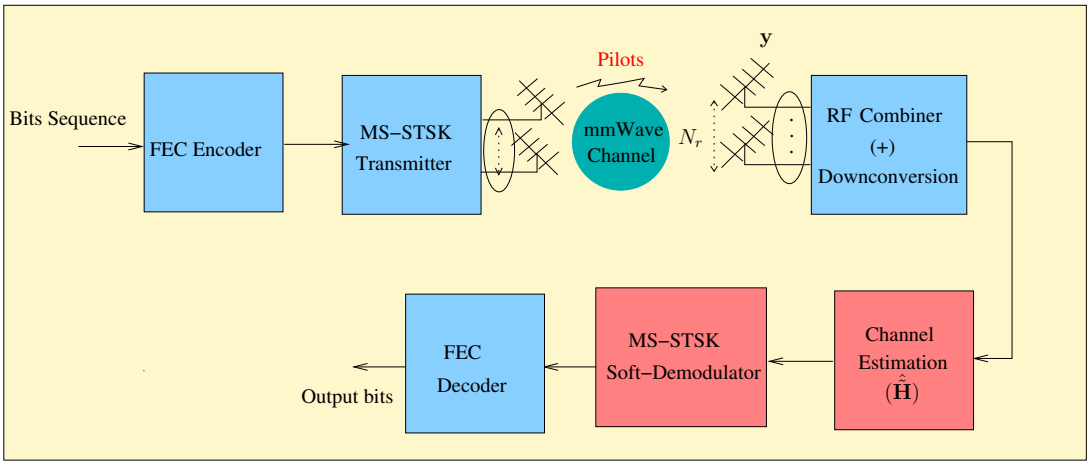
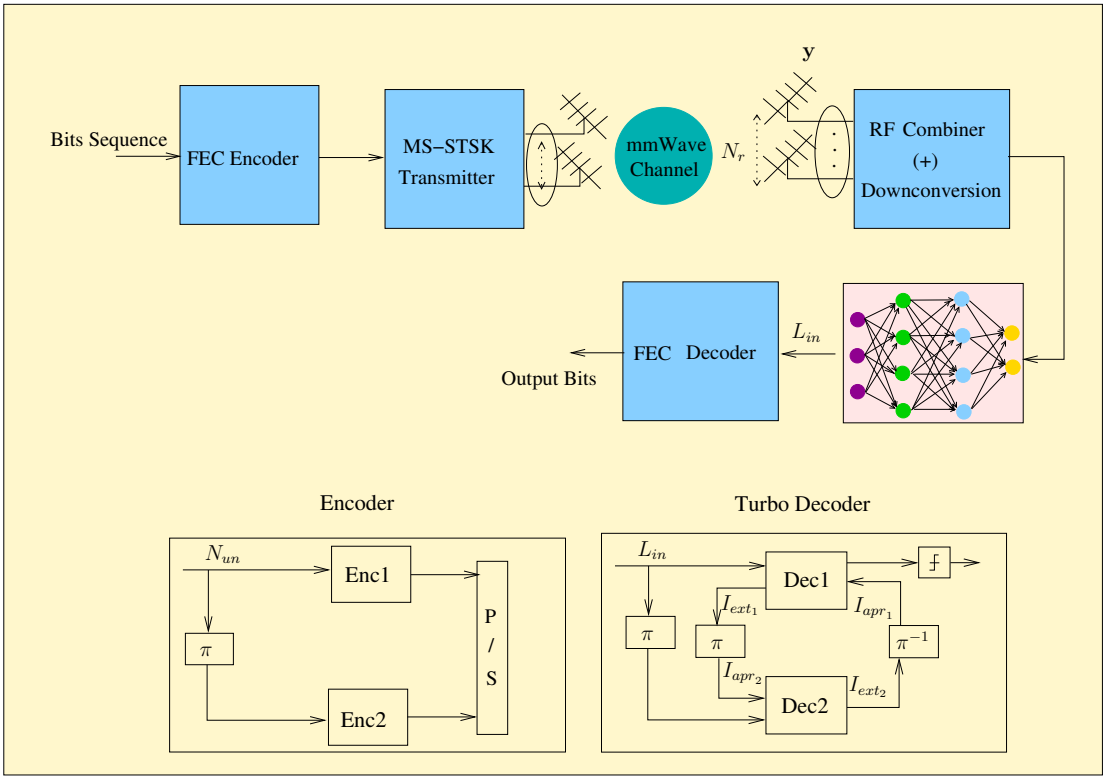


Figure 5.21: Illustration of the input to the AF of each neuron.



(a) Conventional MS-STSK soft-demodulator.



(b) Learning MS-STSK soft-demodulator.

Figure 5.22: Block diagram of MS-STSK soft-demodulators: Conventional design and proposed ANN-assisted design.

5.4.1 Receiver Complexity

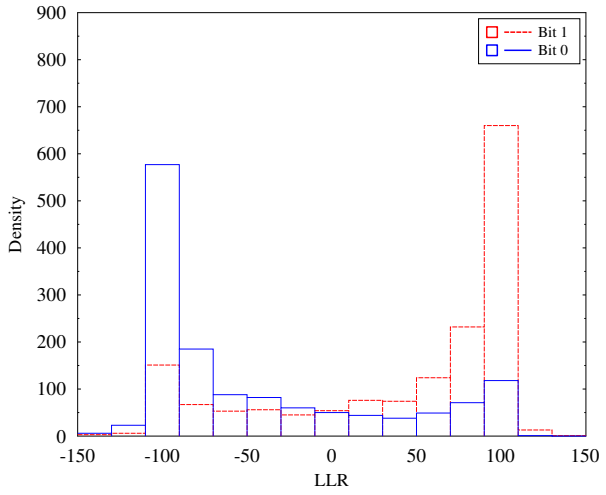
In this section, we commence by discussing the overhead associated with the MS-STSK soft-demodulator, followed by quantifying the receiver's complexity in terms of the number of computations for both the conventional and for the learning-aided soft-demodulator.

Fig. 5.22 shows the block diagram of both of the conventional MS-STSK and of the ANN assisted MS-STSK soft-demodulators. To expound further, Fig. 5.22(a) employs the conventional MS-STSK soft-demodulator, where first the signal is processed by analog RF combining and down-converted to the baseband. Then, the receiver estimates the channel matrix $\mathbf{H}$ with the aid of pilots. After the estimation of the channel the receiver employs the soft-demodulator of (5.37) to obtain the LLR values, which are then passed to the channel decoder. It is important to emphasize that in this design, the receiver has to estimate the channel using a pilot overhead of f_p, which has to be high enough for sampling the channel's complex-valued envelop at multiples of the Nyquist frequency for mitigating the effects of channel noise.

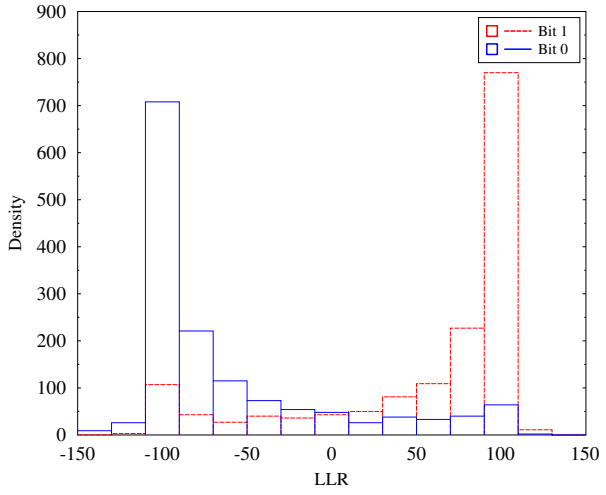
By contrast, Fig. 5.22(b) shows our proposed learning-assisted soft-demodulator. Like in the conventional design, the receiver first employs analog RF combining followed by the down-conversion of the received signal. However, in contrast to the conventional design, our proposed learning-aided soft-demodulator does not require the knowledge of the CSI. The down-converted signal vector is fed to the ANN, which then employs the learned weights to obtain the LLRs without explicit CSI. In other words, our design by passes the channel estimation stage, which is indispensable in the conventional design. The LLRs gleamed from the ANN are then passed to the turbo decoder, where the decoded bits are retrieved.

Having discussed the pilot overhead, we now focus our attention on the complexity in terms of the number of computations. To quantify the complexity, let us assume that the input and output vectors of the ANN shown in Fig. 5.20 are of sizes n_i and n_o, respectively. Let us also assume that the number of neurons in each hidden layer is n_h. The AF (5.40) is computed for each neuron of each layer by employing the corresponding layer's weights and biases, as discussed in Sections 5.3.3 and 5.4. Then the total number of computations required for the ANN of Fig. 5.20 is on the order of $\mathcal{O}(n_i n_h) + \mathcal{O}(n_h^2) + \mathcal{O}(n_h n_o)$.

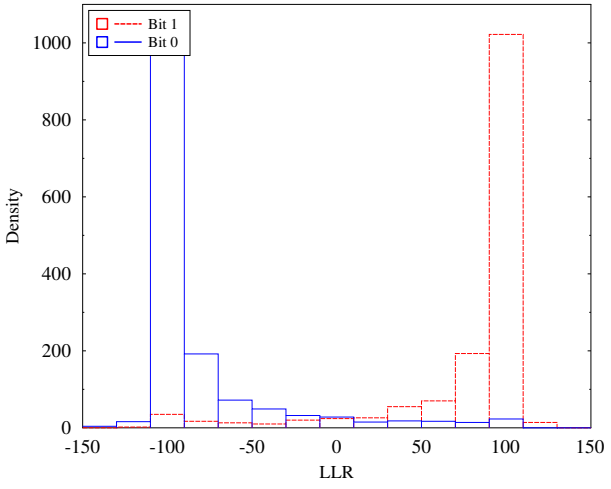
On the other hand, the number of computations required for the conventional soft-demodulator is $2^{N_b}\left(\mathcal{O}(N_t N_r N_c M_Q T^2) + \mathcal{O}(N_c M_Q^3)\right)$. Table 5.4 shows the number of computations required for the simulation parameters summarized in Table 5.5.



(a)



(b)



(c)

Figure 5.23: Histogram depicting the densities of the LLRs at the output of the ANN assisted MS-STSK demodulator for (a) SNR 0 dB (b) SNR 2 dB (c) SNR 4 dB.

Table 5.4: *Quantifying the complexity in terms of the number of complex-valued multiplications.*

Design	Computations
Learning assisted Blind Detection	330
ML-Aided Detection for	12288

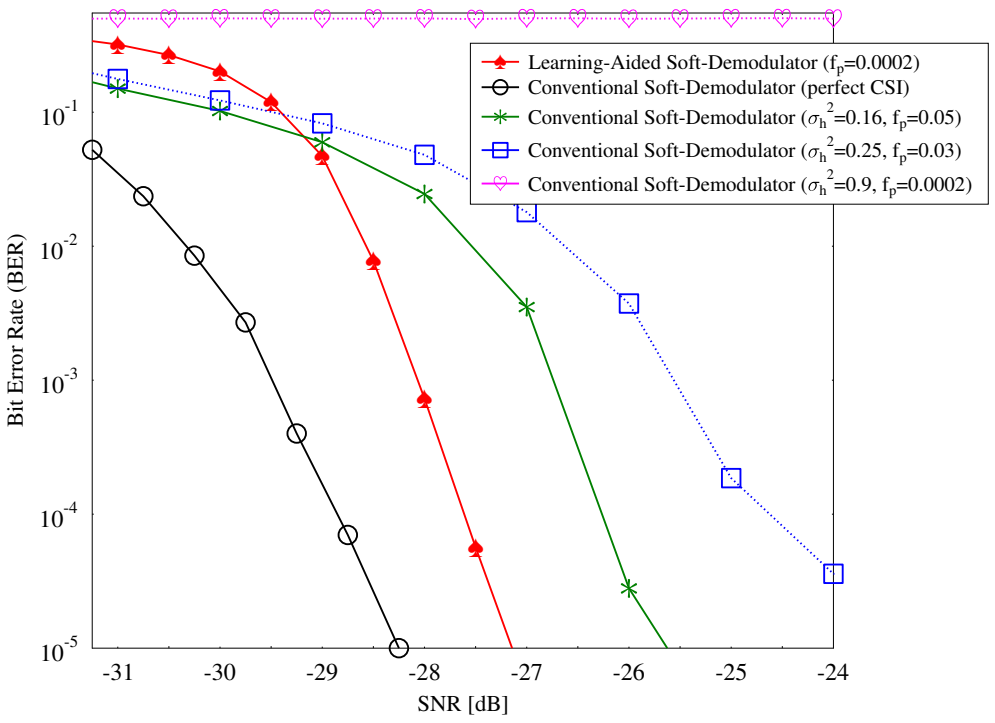


Figure 5.24: Bit error ratio of the proposed learning-aided soft-demodulator and of the conventional soft-demodulator. The simulation parameters are listed in Table 5.5.

5.4.2 Performance Results

In this section, we present our simulation results characterizing the performance of our proposed learning aided soft-demodulator and of the conventional soft-demodulator. We performed Monte Carlo simulations for analyzing the performance of both designs. Furthermore, in our simulations, we employed a half-rate turbo encoder using the LTE generator polynomials as that of the LTE in Fig. 5.22. The simulation parameters are summarized in Table 5.5.

Fig. 5.23 shows the histogram of the LLRs at the output of the ANN for different SNR values. In our simulation, we have limited the maximum and minimum values of the

Table 5.5: Simulation Parameters.

Parameters	Values
Number of AEs in each AA at Tx (K)	64
Number of AEs in each AA at Rx (L)	16
Number of AAs at Tx (N_t)	4
Number of AAs at Rx (N_r)	2
Size of QAM symbol (M_c)	4
Number of dispersion matrices (M_Q)	4
Size of consecutive AA selected (M)	2
Time slots (T)	2
Number of clusters (N_c)	1 & 2
Number of rays (N_{ray})	1
AoA (ϕ_r)	variable
ϕ_t	variable

LLR to ± 100 for the conventional MS-STSK soft-demodulator. Therefore, we have trained the ANN using discrete output samples, i.e $+100$ or -100, according to the bits transmitted. Furthermore, the ANN adopted in our simulations is of continuous regression, since the values from the soft-demodulator are continuous-valued. Therefore, the maximum and minimum values of the ANN's LLRs hover around ± 100. It can be seen from Fig. 5.23 that for the SNR value of 0 dB, the densities of the ANN's LLRs for bit '1' and bit '0' overlap with a larger area. The physical significance of this is that the LLRs observed for bit '1' fall with a higher probability on the wrong side of the bit, i.e. bit '0', hence forwarding less reliable information to the turbo channel decoder. However, as the SNR is increased from 0-to-2-to-4 dB, observe in Fig. 5.23(a)-to-Fig. 5.23(c), that the overlapping area between the two histograms is gradually reduced. In other words, the ANN provides more reliable LLRs.

Fig. 5.24 characterizes the bit error rate of both the proposed learning-aided soft-demodulator and of the conventional MS-STSK soft-demodulator relying on the idealized simplifying assumption of having perfect CSI knowledge. It is evident from Fig. 5.24 that despite the lack of CSI knowledge, the learning-aided soft-demodulator performs closely to the conventional soft-demodulation. More particularly, at BER $< 10^{-4}$, the SNR gap between the two is 1.5 dB. Furthermore, when the realistic imperfect CSI is considered, the conventional soft-demodulation exhibits inferior performance. This becomes evident in Fig. 5.24, where the channel error variance is increased from as low as 0.16 to as high as 0.9. In other words, the conventional design is incapable of achieving a low BER for low pilot overheads. We note that the normalized pilot overhead required for training

the learning-aided design and used in Fig. 5.24 is around 0.002. For the same pilot overhead, the conventional design produces an error floor. It is important to note that our learning-aided soft-demodulator relies on an extremely low pilot overhead as well as much lower search complexity than the conventional MS-STSK demodulator, as seen Table 5.4 and discussed in Sec. 3.3.3.

Our learning-aided soft demodulator requires a low pilot overhead for training the NN. We have found empirically that the pilot overhead required for normalized Doppler frequency of 10^{-3} is 0.002, while it is around 5% for conventional design, when MMSE based channel estimation is employed. We also note that by using (5.27) the channel estimation variance is around 0.16 for a pilot overhead of 5% [221]. Given these parameters, we observed in Fig. 5.24 that our proposed learning-aided soft-demodulation outperforms the conventional detection relying on practical MMSE channel estimation. However, in our design, retraining of the weights may be required depending on the environment.

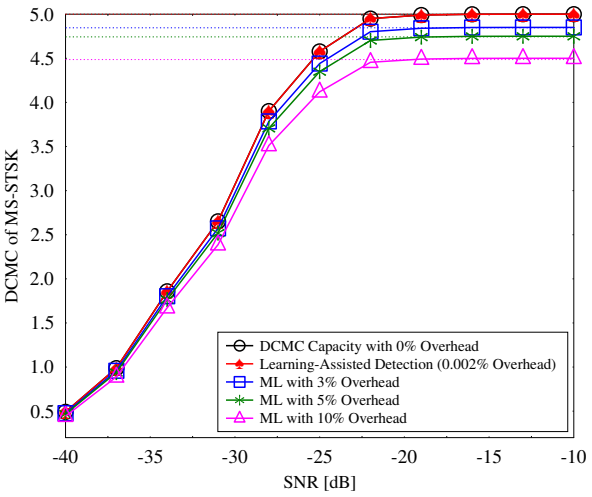
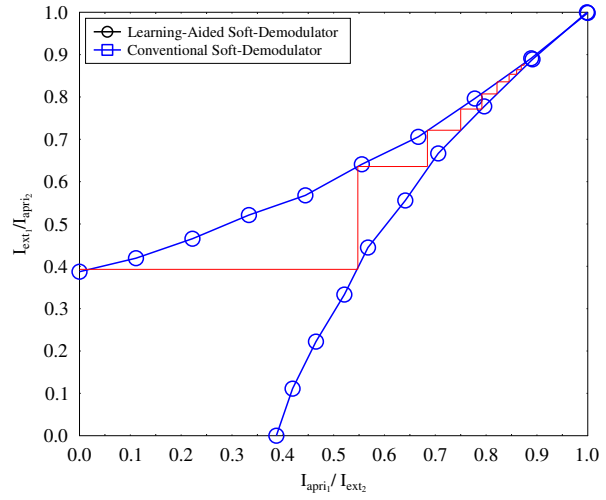
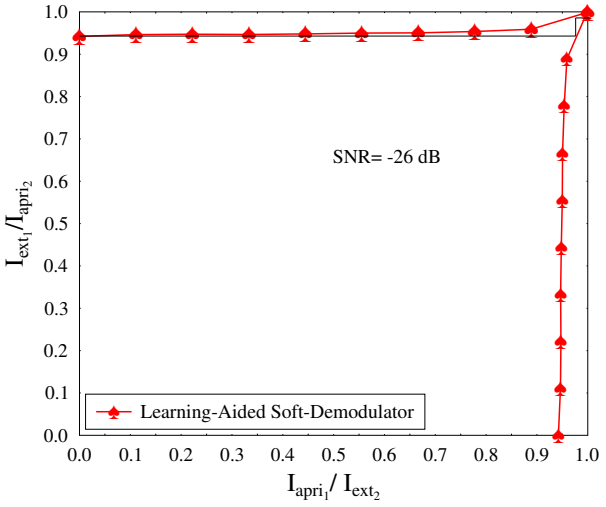


Figure 5.25: The DCMC capacity of the proposed design, and of the conventional design at 3%, 5% and 10% pilot overheads. The simulation parameters are listed in Table 5.5.

Fig. 5.25 shows the Discrete-input Continuous-output Memoryless Channel (DCMC) capacity of both the proposed design, and of the conventional design. Whilst our proposed learning-assisted design exhibits a 1.5 dB SNR loss at a BER of 10^{-4} as shown in Fig. 5.24, observe in Fig. 5.25 that it provides a pilot-overhead-dependent DCMC capacity reduction compared to the conventional proposed design. More explicitly, this is because of the pilot overhead required by the conventional design for channel estimation. For example, the pilot overhead of the conventional design may span from 3% to 10% of the data rate, depending on the Doppler spread. By contrast, the learning-assisted soft-demodulator does not rely on channel estimation, hence almost totally eliminating the pilot overhead, while providing a higher data rate. It is evident from Fig. 5.25 that



(a)



(b)

Figure 5.26: Exit chart (a) of the conventional design; (b) of the learning-assisted design.

the learning assisted design provides an SNR gain of 3 dB over the conventional design having a 10% pilot overhead at a throughput rate of 4 bps/Hz, while it is around 1.5 and 0.8 dB, when having 5.7% and 3% pilot overhead, respectively.

Fig. 5.26 shows the EXtrinsic Information transfer (EXIT) chart of our proposed design and of the conventional design. More particularly, Fig. 5.26 shows the EXIT chart of the conventional design at the SNR of -28 dB, where extrinsic soft-information is exchanged between the upper and lower recursive systematic convolutional turbo decoders, as shown in Fig. 5.22 , while the SNR is -26 dB for our learning-assisted design. More

explicitly, the stair-case shaped curve of Fig. 5.26 portrays the extrinsic information exchange between I_{apri_1}/I_{ext_2} and I_{apri_2}/I_{ext_1}, when the input LLRs are provided by: a) conventional MS-STSK soft-demodulator and b) the learning-assisted soft-demodulator; where I_{apri_1}, I_{ext_1} denote the input and output mutual information of the Decoder 1[8], while I_{apri_2}, I_{ext_2} denote the input and output mutual information of the Decoder 2, as shown in Fig. 5.22. We observe from Fig. 5.26 that for the conventional design relying on perfect CSI, the tunnel begins to open at SNR of -28 dB allowing the stair-case-shaped decoding trajectory to reach the $(1, 1)$ point of perfect convergence to a vanishingly low BER. By contrast, the machine learning aided design achieves this at SNR of -26 dB with the aid of as few as two iterations.

5.5 Conclusions

In this chapter, we proposed a learning-aided transceiver design for mmWave systems. Table 5.6 provides the summary of the learning-aided designs presented in this Chapter. More explicitly, we proposed a learning assisted adaptive transceiver design for each user link of system model presented in Sec. 2.5, based on the near-instantaneous post-processed SNR, where the adaptation switches between multiplexing versus diversity oriented transmission modes as well as by appropriately configuring the modulation employed so as to facilitate both high-reliability and high-rate operation.

Table 5.6: Summary of learning-aided transceiver design.

Section	**System Model**	**Learning Method**
Sec. 5.2	1. User-specific allocation of transmit AEs 2. Grouping of RF chains based on beam-availability	Aim: Link-Adaptation Learning model: KNN
Sec. 5.3	1. MS-STSK transmission with beamforming 2. MS-STSK in conjunction with beam index modulation	Aim: Blind detection of MS-STSK symbols Learning Model: Feed-forward neural network
Sec. 5.4	MS-STSK transmission with beamforming	Aim: Soft decoding Learning Model: Feed-forward

Next in Section 5.3 and 5.4, we proposed a learning assisted blind detection for MS-STSK symbols in conjunction with beam index modulation using soft and hard detections, respectively. More explicitly, Figures 5.16 and 5.24 compares the BER of the MS-STSK

[8]The decoder 1 and decoder 2 correspond to the constituent decoders in the turbo decoder, as shown in Fig. 5.22(b).

system with learning-assisted detection against the optimal receiver. This philosophy makes our design spectral-efficient, since it evades the need for the pilot-assisted channel estimation. In other words, the DCMC capacity is significantly affected by the pilot overhead when ML-aided detection is employed, since it requires accurate CSI in every frame of the transmission. Table 5.7 summarizes the performance observed in Figures 5.16 and 5.24. We noted in Fig. 5.16, although the ML-assisted detection relying on perfect CSI has around 6 dB gain when compared to that of the learning, it starts to degrade when channel imperfections are considered. It is also observed in Fig. 5.4 that the BER of the learning-assisted detection is close to that of ML-assisted detection. On the other hand, for the soft-detection scenario, the SNR gap at BER of 10^{-5} between ML-assisted detection relying on perfect CSI and the learning-aided detection is as low as 1.5 dB.

Table 5.7: Summary of the BER and rate loss observed in Figures 5.16, 5.17, and 5.24.

	ML Assisted Detection	**Learning-Assisted Detection**
Fig. 5.16, hard-detection **SNR** (at BER 10^{-3})	-20.1 dB ($\sigma_h^2 = 0$)	-14.5 dB
	Error Floor ($\sigma_h = 0.25$)	
Fig. 5.18, soft-detection **SNR** (at BER 10^{-4})	-28.9 dB ($\sigma_h^2 = 0$)	-28.4 dB
	to-do (σ_h=0)	
Fig. 5.17, **Rate loss**	0.25 bpcu (overhead=5%)	0.001 bpcu
	0.5 bpcu (overhead =10%)	

Chapter 6

Deep Learning-Aided CSI Acquisition for mmWave Systems

While human's desire to learn for its own sake is diminishing, machine's thirst to learn is thriving.

—The Author, K. Satyanarayana

6.1 Introduction

In Chapters 3 and 4 we have designed the HBF design for both dual-function MIMO transceiver and for full duplex mmWave systems. It is important to emphasize that the HBF designs proposed in the previous chapters rely on the knowledge of the CSI at the transmitter, which is estimated at the receiver using pilots and then fed back to the transmitter. We have seen in Chapter 2 that the statistical mmWave channel model given in (2.20) is primarily comprised of two constituents: their respective angles of arrival and departure, as well as their small scale fading coefficient. However, because of the narrow beams and high susceptibility of mmWave frequencies to blockages, aligning the beams of departure and arrival is challenging. The probability of beam misalignment is even higher when the environment is continuously changing as in vehicular environments [224, 225], which maybe attributed to the ever-changing traffic because of the time-varying movement of vehicles. The mobility of vehicles is different at different times of the day [226]. Furthermore, the traffic density is dependent on the topology of roads [226]. Thus, by considering all the parameters such as spatial randomness and vehicular speeds, it was found that the number of vehicles at any time obeys Poisson distribution with a mean of λ [226,227]. Therefore, in this treatise, we model the traffic density by a Poisson process.

On the other hand, the CIR[1] rapidly becomes outdated and this phenomenon is also referred to as channel aging. To elaborate, the channel aging is a phenomenon that mainly arises because of the mobility of the users relative to the BS. As a result, the channel information arriving at the transmitter would become outdated [228]. In other words, the channel estimate acquired from the receiver at the transmitter would no longer be accurate for the current channel conditions. Furthermore, the multiple local oscillators used at the transmitter would further contribute to the Doppler shift because of the phase noise [229], hence increasing the inaccuracies in the CSI at the transmitter. Therefore, regardless of how sophisticated channel estimation is used at the receiver, the CSI would become outdated at the transmitter. We also note that the loss in the accuracy of the CSI is also dependent on the mobility (speed) of the users. Therefore, it is of crucial importance to predict the channel at the transmitter from the CSI estimate acquired at the receiver.

Against this background, our contributions in this chapter are summarized as follows.

1. *We propose a multi-fingerprint based database, where the fingerprints are collected for different traffic densities in a given location. Then, the BS intelligently chooses the fingerprint based on the traffic density and location information, where we invoke deep learning for the selection of the fingerprint. More explicitly, we invoke a convolutional neural network, namely the softmax linear classifier, where the training weights are designed offline for the selection of the fingerprint. Upon the selection of the fingerprint, the BS then relays the information of the fingerprint selected to the mobile station using the legacy frequency. Thereafter, the BS starts the training process to select the beam-pair from the fingerprint which meets the target received signal power. The mobile station feeds back the index of the beam-pair from the selected fingerprint if it meets the threshold. This significantly reduces the search complexity involved.*

2. *We demonstrate by our simulations that having multiple fingerprint-based beam-alignment provides a superior performance over the single fingerprint based beam-alignment. Furthermore, we show that our learning-aided multiple fingerprint design provides better fidelity than that of the scheme employing multiple fingerprints but dispensing with learning. Additionally, our proposed learning-aided reduced-complexity beam-alignment design performs similarly to beam-sweeping based beam-alignment relying on high-complexity exhaustive beam-search. More explicitly, our design is capable of maintaining the target RSS in dense vehicular environments, while both single fingerprint and LOS based beam-alignment suffer from blockages.*

3. *We propose a multi-functional beam transmission scheme as an application example of our proposed design, where multiple beams that satisfy the target received*

[1]Once the beam-alignment is carried out, the index of the AoA-AoD pair changes very slowly, which can be assumed to be constant.

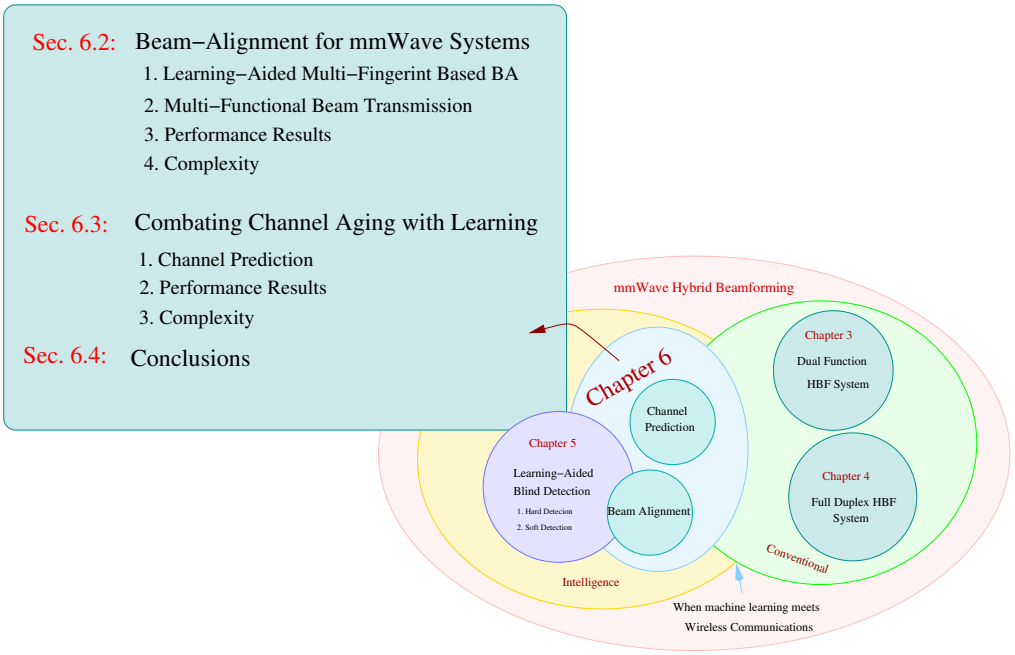


Figure 6.1: The organization of the chapter.

power are selected. Then, depending on the mobile station's requirement, the BS can leverage the additional beams for increasing the multiplexing/diversity gains. Furthermore, if the number of RF chains is lower than the number of beam pairs available that meets the target receiver power, the BS then employs beam index modulation to increase the spectral efficiency.

4. *To counteract the effects of channel aging, we propose a learning-assisted channel prediction relying on a radial basis function neural network, where we show by simulation that upon involving sufficient training neural network-aided channel prediction can faithfully reproduce the current channel.*

5. *We empirically show that depending on the Doppler spread, the neural network used for channel predictions has to be periodically retrained. In other words, we demonstrate by simulations that the overhead involved in the CSI feedback for faithful reproduction of the actual channel at the transmitter is low for lower Doppler spreads.*

The remainder of this chapter is organized as shown in the Fig. 6.1. In Sec. 6.2, we present beam-alignment for mmWave hybrid systems, where we first discuss multi-fingerprint based design in Sec. 6.2.1. Then, in Sec. 6.2.2, we introduce learning to the proposed design whose complexity and performance are discussed in Sec. 6.2.3 and Sec. 6.2.5, respectively. Next, in Sec. 6.3, we propose channel prediction to combat the channel aging phenomenon in mmWave hybrid systems. While the complexity and performance of this design are discussed in Sec. 6.3.3 and Sec. 6.3.2, respectively, our conclusions are presented in Sec. 6.4

6.2 Beam-Alignment for mmWave Systems

Let us consider a vehicular scenario, where a BS serves the vehicles (users) of its cell. The number of vehicles N_v at any point of time obeys the Poisson distribution with a mean of λ and variance of λ [230]. Let us also assume that the serving cell is partitioned into N locations, where the BS is equipped with the knowledge of the fingerprints (FP) for each location in its database, as shown in Fig. 6.2. In this chapter, we define fingerprints as beam-pairs, whose AoA-AoD values are pre-determined offline. We refer the readers to [231–233] for a detailed account on fingerprint based communication. Fingerprints, i.e. AoA-AoD pairs in our design, are typically obtained by employing beam-sweeping, where a high-resolution scanning of beam-pairs is carried out. However, given the time-varying traffic density, corresponding to the varying number of vehicles at any point of time, a single pre-determined fingerprint would not be able to provide improved performance at a reduced search complexity. This is because the AoA-AoD pair which provides high received signal power may be blocked/suppressed by the neighboring vehicles. Hence a multi-fingerprint based regime is conceived for different traffic densities at a given location.

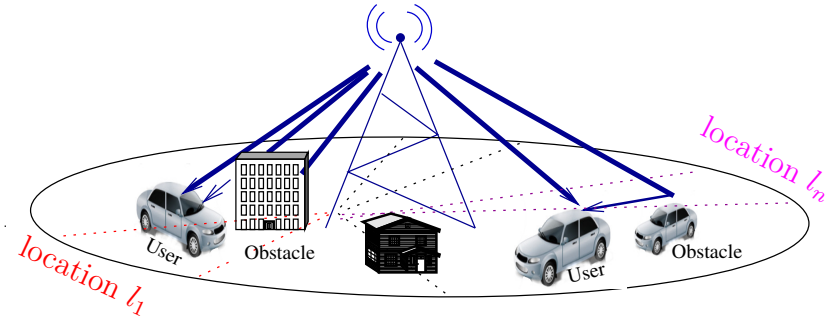


Figure 6.2: Fingerprint Based beam-alignment.

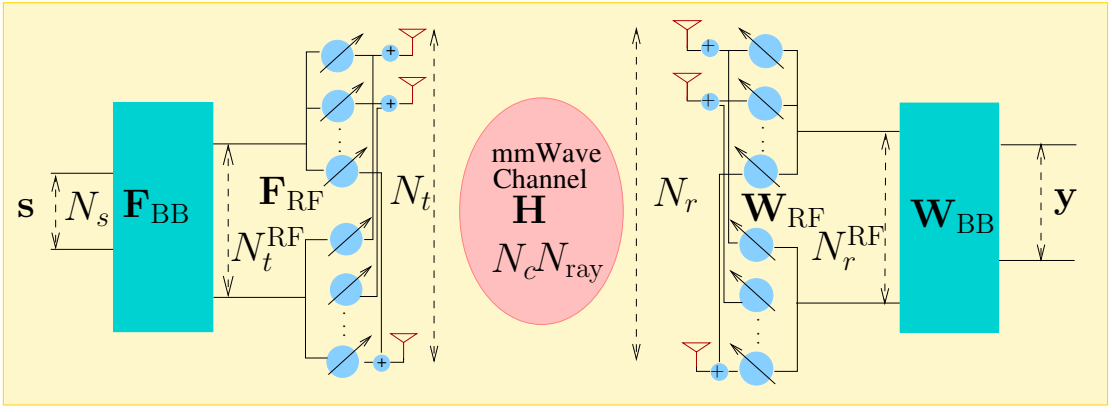


Figure 6.3: HBF architecture.

Fig. 6.2 shows our system model, where the BS is serving a user in location l_1 for a traffic density λ using the fingerprint $\text{FP}_\lambda(l_1)$[2]. The architecture of the BS and the user, which is detailed in Sec. 2.4.2.1, is shown in Fig. 6.3, where the BS (transmitter) is equipped with N_t transmit antennas and N_t^{RF} chains, while the user (receiver) is equipped with N_r receive antennas and N_r^{RF} chains. Furthermore, let us assume that $\{(\mathbf{f}_{\text{RF}}^1, \mathbf{w}_{\text{RF}}^1), (\mathbf{f}_{\text{RF}}^2, \mathbf{w}_{\text{RF}}^2), \ldots, (\mathbf{f}_{\text{RF}}^N, \mathbf{w}_{\text{RF}}^N)\}$ are chosen as the BF vectors at the BS and at the user end, respectively, for a traffic density of λ. Then the received signal vector at the user is given by

$$\mathbf{y} = \sqrt{P_t}\mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{n}, \tag{6.1}$$

where $\mathbf{F}_{\text{RF}}$ is the transmit BF matrix of size $N_t \times N_t^{\text{RF}}$ at the BS, where N_t^{RF} columns are constructed from the potential AoD set $\{\mathbf{f}_{\text{RF}}^1 \ldots \mathbf{f}_{\text{RF}}^N\}$. Similarly, $\mathbf{W}_{\text{RF}}$ is the receive BF matrix of size $N_r \times N_r^{\text{RF}}$ at the user end, where N_r^{RF} columns are constructed from the potential AoA set $\{\mathbf{w}_{\text{RF}}^1 \ldots \mathbf{w}_{\text{RF}}^N\}$, while the other parameters are same as discussed in previous chapters.

In this chapter, we consider the RSS[3] [234] as the performance metric to determine the beam-pair for a successful transmission. Furthermore, for the construction of the fingerprints database, we accounted for the attenuation and blockage caused by neighboring vehicles in addition to the path loss experienced by the mmWave carrier. To achieve this, we have invoked the multiple knife-edge model recommended by ITU-R [235]. For a given location, each fingerprint is constructed for each traffic density, by accounting for the total attenuation caused by the vehicles. The attenuation caused by each vehicles using single knife-edge is given by [224,235]

$$A = \begin{cases} 6.9 + 20\log_{10}\left[\sqrt{(v-1)^2+1} + v - 0.1\right], \forall v > -0.7 \\ 0 \qquad \text{Otherwise}, \end{cases} \tag{6.2}$$

where $v = \sqrt{2}h/r_f$, h is the height of the obstacle from the line joining the BS and the user, and r_f is the Fresnel ellipsoid radius expressed as

$$r_f = \sqrt{\frac{\Lambda d_{\text{obstacle}} d_{\text{user}}(d_{\text{user}} - d_{\text{obstacle}})}{d_{\text{user}}}},$$

where Λ is the wavelength, d_{user} is the distance between the transmitter and user, while d_{obstacle} is the distance between the transmitter and an obstacle as shown in Fig. 6.4.

[2]The location information is obtained from the user using lower frequency, such as LTE, where the transmission is omni-directional.

[3]Note that other performance indicators may also be used, such signal-to-noise ratio (SNR), rate, received signal quality.

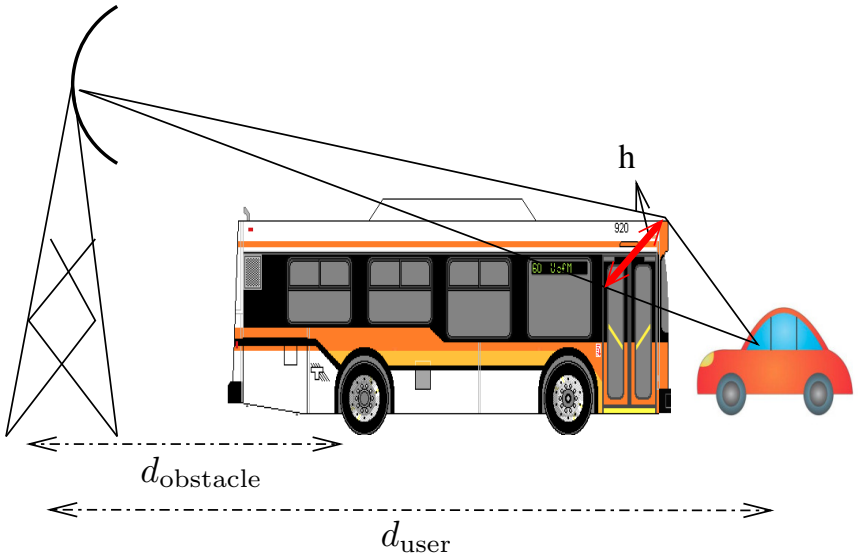


Figure 6.4: Knife-edge diffraction caused by the vehicle obstacle.

On the other hand, the path loss experienced by the signal at a distance of d for a transmit and receive antenna gain of G_t and G_r, respectively, is given by [1]

$$PL_{[dB]} = PL_0 + 10n_p \log\left(\frac{d}{d_0}\right) + S_{\sigma_s} + G_t + G_r, \tag{6.3}$$

where d_0 is the close-in reference distance, S_{σ_s} is the shadowing factor, while PL_0 is the free-space path loss. Furthermore, for a given wavelength (Λ) the free-space path loss is expressed as

$$PL_0 = 20\log_{10}\left(\frac{4\pi d_0}{\Lambda}\right).$$

Therefore, the total received power after accounting for both the path loss and the attenuation caused by a vehicle is

$$P_r = P_t - PL_{[dB]} - A. \tag{6.4}$$

It is instructive to note that (6.4) is due to large scale fading. However, considering the small scale fading as well as the BF and combining effects, the net received power $P_{r_{\text{net}}}$ is expressed as

$$P_{r_{\text{net}}} = P_r + 10\log_{10}\left\|\left(\mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right)\right\|_{\text{F}}^2, \tag{6.5}$$

while the capacity is given by

$$C = \log_2\left(\mathbf{I} + P_r\mathbf{Z}^{-1}\left(\mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right)\left(\mathbf{W}_{\text{BB}}^H\mathbf{W}_{\text{RF}}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right)^H\right), \tag{6.6}$$

where $\mathbf{Z} = \sigma_n^2 \left(\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H\right) \left(\mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}}\right)$, σ_n^2 is the noise variance. The net received power $P_{r_{\mathrm{net}}}$ is observed for the fingerprint construction. A beam sweeping action is carried out initially for N locations at discrete traffic densities and then the specific beam-pairs which achieve the target RSS are stored in the database. In other words, the AoA-AoD fingerprint is obtained by conducting a high-resolution beam-search campaign by accounting for all blockages caused by obstacles. The fingerprint construction is typically carried out offline either by computer-generated environment simulations, or in real time during the BS installation.

In the next section, we discuss multi-fingerprint construction as well as the learning assisted fingerprint adaptation, followed by multi-functional beam transmission.

6.2.1 Multi-Fingerprint Based Beam-Alignment

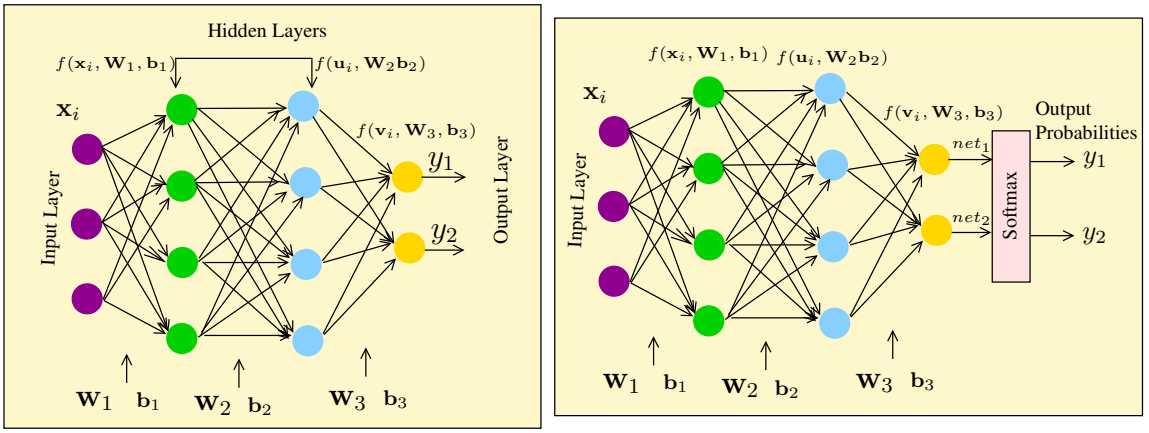
In this section, we commence by outlining our multi-fingerprint based beam-alignment, where we propose multiple fingerprints for a given location for different traffic densities. Then, we aim for improving the design by adopting a learning based approach for fingerprint adaptation followed by a discussion on the concept of multi-functional beam transmission, where a plurality of beam-pairs satisfying the RSS threshold are selected for attaining both multiplexing and diversity gains. Furthermore, we also propose that if the number of beams are higher than the number of RF chains, beam index modulation is employed to attain a high spectral efficiency.

Achieving accurate beam-alignment in directional transmission systems is challenging, especially at mmWave frequencies because of its high susceptibility to blockages. To circumvent this problem a fingerprint based beam-alignment technique may be employed. In broader terms, having accurate fingerprint may be viewed as side-information, which is used for enhancing the system performance. In our design, the fingerprint is comprised of a set of possible beam-pairs for a given location, over which a communication link can be established. Typically, a fingerprint is constructed for each location during the BS installation stage by taking the surrounding environment into account, such as buildings, lamp posts, vehicles, etc. However, the ever-changing vehicular traffic may strictly limit the performance of the fingerprint. This is because the direction of the beam or the number of beam-pairs available is highly dependent on the density and position of vehicles on the road. In other words, it depends on the traffic density. Furthermore, the traffic density is time-varying. For example, the traffic density in the morning is different from that in the afternoon, or during special events. This necessitates using multiple fingerprints for a given location, depending on the traffic density. A typical multi-fingerprint based database is shown in Table 6.1. It is constructed by calculating the net received power using (6.4) for each location at different traffic densities. It is interesting to note that this multi-fingerprint adaptation is akin to the classical link-adaptation involving multiple modulation coding scheme (MCS), BER and SNR, where the transmitter adapts

among different schemes to provide high data rates under given channel conditions while also meeting the target BER.

Table 6.1: An example illustrating multi-fingerprint database

Location	**Traffic Density**	**Beam-Pairs (BP)**
L_1	λ_1	$BP_2, BP_{10}, BP_{23}, \ldots$
L_1	λ_2	$BP_{320}, BP_{210}, BP_3, \ldots$
$\vdots$	$\vdots$	$\vdots$
L_k	λ_1	$BP_1, BP_9, BP_{300}, \ldots$
$\vdots$	$\vdots$	$\vdots$
L_n	λ_3	$BP_{21}, BP_6, BP_{250}, \ldots$



(a) A typical neural network

(b) A softmax convolutional neural network

Figure 6.5: Architecture of a neural network.

This relationship can be represented by a look-up table for link-adaptation, where depending on the traffic density and on the location, its corresponding fingerprint is selected. For example, if the user is at the location L_1 and the BS estimates the traffic density as λ_2, then the BS selects the fingerprint consisting of beams $\{BP_{320}, BP_{210}, BP_3, \ldots\}$. The BS then shares this information with the user, whereupon the BS and user invoke training for identifying the best beam-pair amongst the available beam-pairs of the fingerprint selected. This significantly reduces the search space involved in the beam-alignment. To further reduce the search complexity, a RSS threshold is set, where the user selects the specific beam-pair whose observed RSS value is higher than the threshold. Having selected the beam pair, the user relays this information to the BS, eliminating the search over the successive beam-pairs. While this procedure is different for transmission over a plurality of beam-pairs invoked in order to achieve diversity or spatial multiplexing gains, where the search is continued until the required number of beam-pairs are found.

However, given the ever-changing nature of the channel and the non-linearities imposed by the ADC/DAC which is even more pronounced at mmWave frequencies, a look-up table assisted fingerprint adaptation would be significantly affected, which we will discuss in the simulation results.

6.2.2 Learning-Aided Multi-Fingerprint Based Beam-Alignment

We have discussed in the previous section that a multi-fingerprint based beam-alignment would account for beam-pairs that are capable of handling different traffic densities, where the BS selects the fingerprint that is pertinent to the traffic density and location observed. This is carried out relying on a look-up table. However, while this design enhances the performance compared to that of a single fingerprint based beam-alignment, the performance gain may become limited if the threshold values, such as RSS observed in a fingerprint at a given location L for a traffic density λ of the look-up table becomes outdated. In other words, the performance of a look-up table based fingerprint adaptation may limit the performance because of the uncertainties mentioned in Sec. 6.2.1, and hence it becomes extremely challenging to attain perfect beam-alignment. Therefore, we resort to a learning-aided multi-fingerprint based beam-alignment. The rationale for employing learning is to eliminate the dependence on traffic density threshold values, as enumerated in Table 6.1 during the fingerprint selection. This is because, the values may become outdated owing to the imperfections in the channel and impairments in the ADC/DACs.

Therefore, in this chapter, we conceive a neural network based approach for intelligent adaptation between multiple fingerprints. The rationale for choosing a neural network is its reduced complexity while providing a superior performance. The learning approach pursued in this chapter relies on two stages: the training phase and the testing phase. In the training phase, the weight vectors of the network are computed using training samples, where the input and output are known. This is classified as a supervised learning technique, where the training weights are designed using supervision. It is instructive to note that the training weights are calculated offline, hence this calculation does not impose any real time overhead on the system. A typical neural network is shown in Fig. 6.5(a).

This network is referred to as a deep neural network if the number of hidden layers is higher than one, as shown in Fig. 6.5, where the number of hidden layers is two. As seen in the figure, the training samples are passed to the input layer and the weights are designed for minimizing the error, which is the difference between the true and predicted output values. In Fig. 6.5(a), $\mathbf{W}_1, \mathbf{W}_2$, and $\mathbf{W}_3$ are the weights for the input layer, hidden layer, and output layer, respectively, where $\mathbf{W}(p, q)$ denotes the weight attached between the nodes p and q, while $\mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i$ and y_1, y_2 are the inputs and outputs of the network, respectively, where i denotes the class. In our design, each fingerprint corresponds to a class. Furthermore, the output of each hidden layer is defined by an

activation function $f(.)$, also called a score function, which determines the performance of the system. The choice of the activation function is dependent on the analytical tractability, computational complexity and the type of output signal. On the other hand, to yield better results, a loss function is introduced which captures the error (loss) between the predicted outcome and the real outcome in the training samples.

In our proposed design, the choice of the fingerprint (or class) is dictated by the output probabilities associated with each fingerprint. In other words, the fingerprint having a high probability is selected. Since we deal with the probabilities at the output, it can be interpreted as a logistic regression having multiple classes, where each fingerprint constitutes a class. Hence, an activation (score) function which deals with the probability is considered, where in our case we use a so-called softmax function, which we will discuss in detail later in this section. The resultant neural networks are also sometimes referred to as softmax convolutional networks [236].

Each layer of Fig. 6.5(b) is assigned a score by the activating function. However, since we deal with the probabilities of specific outcomes, the linear weights are translated to probabilities by a softmax function at the output, as shown in Fig. 6.5(b) and expressed as [236]

$$f_i(z) = \frac{e^{z_i}}{\sum_k e^{z_k}}, \tag{6.7}$$

where z is the score vector at the layer before the output probabilities are calculated, which is given by [236]

$$z = f(\mathbf{x}_i, \mathbf{W}, b) = \mathbf{W}\mathbf{x}_i + \mathbf{b}. \tag{6.8}$$

In (6.8) $\mathbf{W}$ is the weight matrix of the layer, $\mathbf{b}$ is the bias vector, while $\mathbf{x}_i$ is the input of the layer, as shown in Fig. 6.5(b). As an example, let us consider Fig. 6.5(b), where the softmax function is applied at the output layer to the scores net_1 and net_2, which are calculated using the weight matrix $\mathbf{W}_3$ and bias vector $\mathbf{b}_3$. Note that the softmax function is employed only at the output layer, while for all other layers, only the scores attained using (6.8) are calculated as shown in Fig. 6.5(b).

In our chapter, the input vector $\mathbf{x}_i$ is a three-dimensional vector holding the location, traffic density and RSS values, while the outputs represent the probabilities associated with each fingerprint. In other words, the output takes the form of $[0\ldots,0,1,0,\ldots,0]^T$, where 1 is the probability associated with that particular fingerprint.

It is instructive to note that the weight matrices are initially chosen as random from the distribution $\mathcal{N}(0,1)$, hence the prediction at the output would be erroneous. Therefore, to improve the prediction of the fingerprint for a given traffic density and location, a loss function is introduced which is a measure of difference between the predicted probability and the true probability associated with the given class. In other words, by considering

the loss function, the weight matrices are optimized for ensuring the loss is minimized. More explicitly, in our design we aim for minimizing the divergence between the real and predicted probability distributions. This loss function can also be interpreted as the Kullback-Leibler divergence between two distributions. Thus, for distributions p and q it is expressed as [237]

$$D_{KL}(p\|q) = -\sum_{i} p(i)\log q(i), \tag{6.9}$$

where in our case, $p(i)$ is the probability of the correct class i, which is $p(i) = [0, \ldots, 1, \ldots, 0]$, while q is the function in (6.7). Upon substituting (6.7) into (6.9), we get

$$D_{KL}(p\|q) = -\sum_{i=1}^{S} \log \frac{e^{z_i}}{\sum_k e^{z_k}}, \tag{6.10}$$

where S is the number of training samples.

Additionally, we have the cross entropy of $\mathcal{H}(p,q) = \mathcal{H}(p) + D_{KL}(p\|q)$, where $\mathcal{H}(p) = 0$ holds, since there is no uncertainty in the correct class. Therefore, this loss function may also be referred to as cross entropy loss.

Having defined the cross entropy loss, the total loss function over all classes associated with a regularization penalty of $R(\mathbf{W})$ is given as [236]

$$L = -\frac{1}{S}\sum_{i=1}^{S} \log \frac{e^{z_i}}{\sum_k e^{z_k}} + R(\mathbf{W}), \tag{6.11}$$

where we have $R(\mathbf{W}) = \frac{\lambda}{2}\|\mathbf{W}\|_2^2$. The rationale behind adding a regularization term in (6.11) is to ensure that it does not result in over-fitting [236].

We now aim for minimizing (6.11) by computing the gradient with respect to the weight matrix $\mathbf{W}_3$ and the bias $\mathbf{b}_3$ of Fig. 6.5(b). To achieve this, we compute the gradient for each class whose corresponding weight vectors are now $\mathbf{w}_3^i$ and bias b_3^i, where i represents the class.[4] Note that z_i in (6.11) is a function of the weight vector $\mathbf{w}_3^i$ and bias $\mathbf{b}_i$. After a series of steps, the gradient with respect to $\mathbf{w}_i$ and b_i is given as [236]

$$\frac{\partial L}{\partial \mathbf{w}_3^i} = (f_i - \delta_{ik})\mathbf{x} + \lambda \mathbf{w}_i \quad \delta_{ik} = \begin{cases} 1, i = k \\ 0, \text{otherwise} \end{cases} \tag{6.12}$$

$$\frac{\partial L}{\partial b_3^i} = f_i - \delta_{ik}. \tag{6.13}$$

[4]Each row of the weight matrix represents the weights associated with the respective class. For example, the i^{th} row vector $\mathbf{w}_3^i$ of $\mathbf{W}_3$ represents the weights associated with the i^{th} class.

Thus, by employing gradient-descent we arrive at

$$\mathbf{w}_3^i = \mathbf{w}_3^i - \alpha \frac{\partial L}{\partial \mathbf{w}_3^i}, \tag{6.14}$$

$$b_3^i = b_3^i - \alpha \frac{\partial L}{\partial b_3^i}, \tag{6.15}$$

where α is the step-size. Similarly, weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$, and bias vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ are obtained by employing the gradient of the loss function in (6.12) with the respective matrix $\mathbf{W}$ and vector $\mathbf{b}$. This process is called as error backpropagation [236]. The pseudo-code of the proposed learning-aided fingerprint based algorithm is presented in Algorithm 6.7.

Algorithm 6.7 Proposed Learning-Aided beam-alignment

Offline:

1: Input: Training samples for all training locations
2: Output: Fingerprint
3: Initialize $\mathbf{W}_1$, $\mathbf{W}_1$ and $\mathbf{W}_3$ as random
4: Compute the output of $f(.)$ using (6.8) for each layer with the assigned weights
5: Apply softmax function using (6.7) to obtain the probabilities for each class
6: Obtain the weight matrices and bias vectors by (6.14) and (6.15)
7: Do error backpropagation
8: Repeat 4-7 until convergence of (6.12).

Online:

1: Input the location, number of vehicles, Target RSS
2: Apply training weights found during offline phase
3: Output: desired fingerprint

It is important to emphasize that the weight matrix $\mathbf{W}$ and bias vector $\mathbf{b}$ of the network are computed offline and stored in memory. This entire process is carried out during the training phase.

Having discussed the softmax training phase, we now focus our attention on the real time, the testing phase. During the initial access the BS communicates with the user by employing omni-directional transmission using lower frequencies such as that of LTE, where the BS estimates the location of the user. The BS also has the information of the number of users (vehicles) in its cell, which is exploited for estimating the vehicular density (λ). Furthermore, the user relays the information of its RSS threshold requirements to the BS. Given the number of vehicles, the RSS threshold, and the location as the input parameters, the BS employs the softmax algorithm with the weights and bias calculated during the training stage. The algorithm then predicts the probability of each class, i.e. each fingerprint, and then the BS selects the fingerprint having the highest probability. After the selection of the fingerprint, it informs the user about the possible beam-pairs, and conducts beam-search over these selected beam-pairs. The user then

feeds back the specific beam-pair index which meets its post-processing RSS threshold. A schematic diagram illustrating the algorithm is shown in Fig. 6.6.

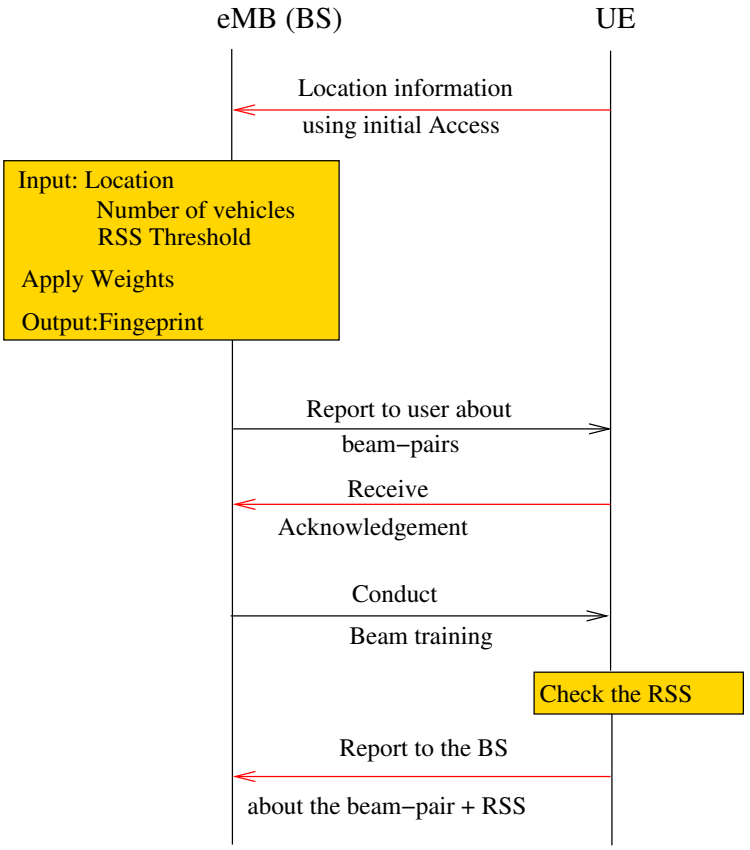


Figure 6.6: The beam-alignment procedure.

In the next section, we present our multi-functional BF as an application of the proposed design.

6.2.3 Multi-functional Beam Transmission: An Application

In this section, we propose a multi-functional beam transmission by allowing some tolerance in the beam search complexity. The rationale for employing multi-functional beam transmission is to increase the spectral efficiency as well as to enhance the performance. This philosophy works under the assumption that there exist a plurality of beam-pairs which satisfy the RSS threshold. As discussed in Sec. 6.2.1 and Sec. 6.2.2, the BS conducts beam-search in the selected fingerprint, where the user chooses the specific beam-pair which satisfies the target RSS and feeds back the index of the beam-pair to the BS using LTE and avoids the beam search in the successive beam-pairs.

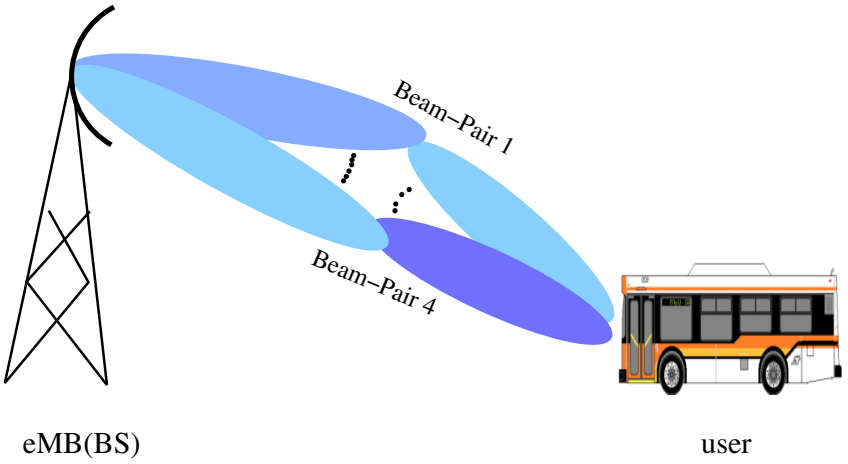


Figure 6.7: A plurality of beam-pairs selected to attain diversity/multiplexing gains.

By contrast, in the multi-functional beam transmission, the user chooses several beam-pairs which satisfies the RSS threshold at the expense of increased search complexity[5] as shown in Fig. 6.7.

These additional beam-pairs can be leveraged to achieve diversity and/or multiplexing gains. It is important to emphasize that the number of beams that can be exploited is limited by the number of RF chains. Given the eligible beam-pairs observed, the BS-user pair may also employ link-adaptation depending on the nature of the channel in each beam. In other words, depending on the post-processing SNR observed at the user, the BS may employ diversity or multiplexing. Explicitly, if the channel is in deep fade, the user may opt for employing diversity, whilst for multiplexing otherwise. Furthermore, the BS may also optimize the power allocation for each beam in conjunction with multiplexing- and diversity-aided transmission.

On the other hand, if the number of beam-pairs observed by the user that satisfy the target RSS is higher than the number of RF chains[6], the BS may decide to employ beam-index modulation[7] using the beam-pairs reported by the user as shown in Fig. 6.8.

Fig. 6.8 illustrates the typical beam index modulation employed at the BS to increase the data rate. Note that here the number of beam pairs N_b that meets the target RSS is higher than the number of RF chains N_t^{RF}. This is akin to the spatial modulation, where the antenna index carries the information. In contrast to the antenna index in spatial modulation, in this design, beam index is used for the information conveyance.

[5]It is important to emphasize that there may also be a scenario, where there is no an additional beam-pair despite the increased beam search in the successive beam-pairs.

[6]Note that the maximum number of beam-pairs the BS can leverage is equal to the number of RF chains.

[7]Beam-index modulation allows us to convey extra implicit information by inferring at the receiver which specific beam was activated.

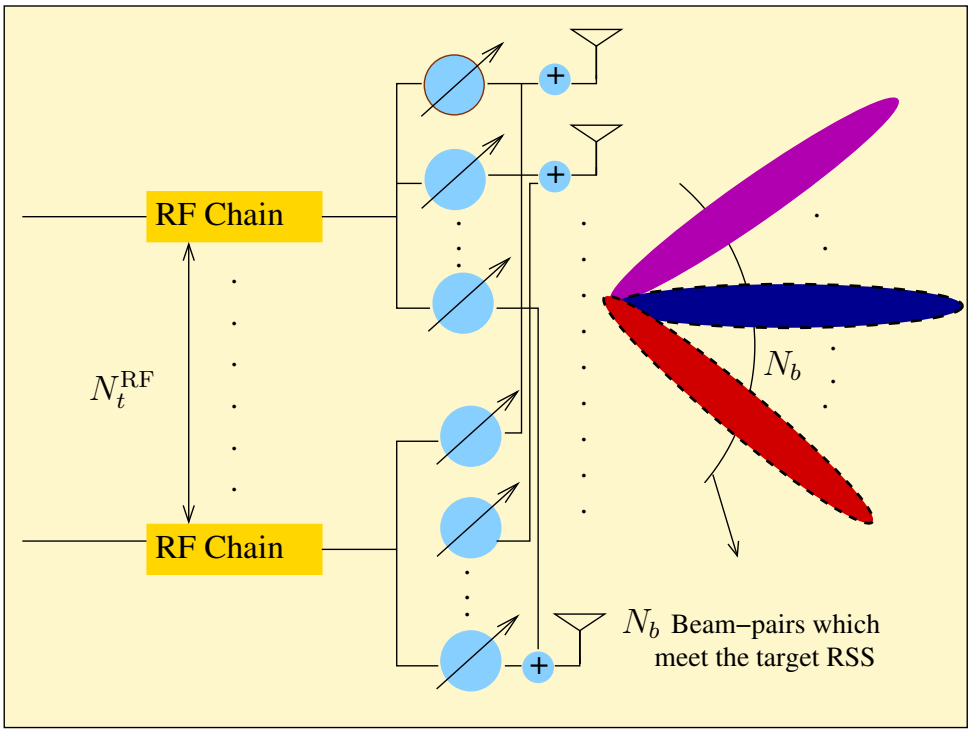


Figure 6.8: Beam index modulation at the BS.

Thus the total number of bits that can be transmitted per second per channel per user is

$$\text{Number of bits} = \underbrace{\log_2 M}_{\text{M-ary bits}} + \underbrace{\log_2 N_b}_{\text{number of beam pairs}} .$$

We note that in addition to the beam-pairs selected for the transmission based on the input signal stream during beam index modulation, diversity or multiplexing gains may also be attained by the selected beam-pairs depending on the channel's nature in these beam-pairs.

In the next section, we present our simulation results of the proposed design.

6.2.4 Performance Results

In this section, we present our simulation results for characterizing the performance of the proposed design. More explicitly, we characterize the performance of the multi-fingerprint aided beam-alignment scheme relying on learning and of the benchmark design dispensing with learning. In our simulations, the number of vehicles at any point of time follows Poisson distribution $\sim$ Poisson(λ) having both a mean and variance of λ. Furthermore, the blockages caused by vehicles are random obeying the distribution $\mathcal{U}(0, N_v)$, where N_v is the number of vehicles having a Poisson distribution $\sim$ Poisson(λ). At the time

of writing, there is no mathematical model or distribution for blockage, only empirical models [224]. Hence, we assumed in our simulations that the maximum number of blockages is equal to the number of vehicles. Furthermore, from the database on the vehicle dimension [238], it was found that the height of the vehicles follows normal distribution with a mean of μ_h and standard deviation of σ_h [224]. Our simulation parameters are summarized in Table 6.2. In this chapter, we have considered three fingerprints constructed for three different traffic densities.

Table 6.2: Simulation parameters

Parameters	Values
P_t	20 dBm
N_t	32
N_r	8
G_t	10 dBi
G_r	5 dBi
Λ(28 GHz)	0.0107
Number of Vehicles (N_v)	Poisson (λ)
Vehicles height [meters]	$\mathcal{N}(150, 8.6)$
Blockages	rand$(0, N_v)$
$d_{\text{user}}, d_{\text{obstacle}}$	rand(.)

Fig. 6.9 shows an example, where a BS is located at the center of the cell. Furthermore, the cell is partitioned into four locations, and each location has its own fingerprints for the different traffic densities. It is important to emphasize that the number of vehicles in each location is Poisson distributed, while the distance of the vehicles from the BS in each cell is uniformly distributed. Observe in Fig. 6.9 that the BS is serving two users in Locations 2 and 3, while the vehicles near the users are treated as obstacles.

To analyze the loss function of the softmax neural network presented in (6.11), Fig. 6.10 shows the cross-entropy of the network during training versus the number of epochs[8]. It can be seen from the figure that the network's best performance is reached after approximately 60 epochs. Furthermore, it can be noticed from the figure that the validation error is only slightly higher than the training error — which implies that the neural network weights designed provide a good fit in terms of the mapping between the input and the output samples. It is instructive to note that Fig. 6.10 can be used to study how well the neural network parameters are designed. In other words, if the validation

[8]The network is said to have reached '1 Epoch' when an entire dataset, i.e., when each and every sample of the data is passed through the network for designing the network parameters for one time. It can be loosely interpreted as the number of times (or iterations) the dataset is used for designing the parameters.

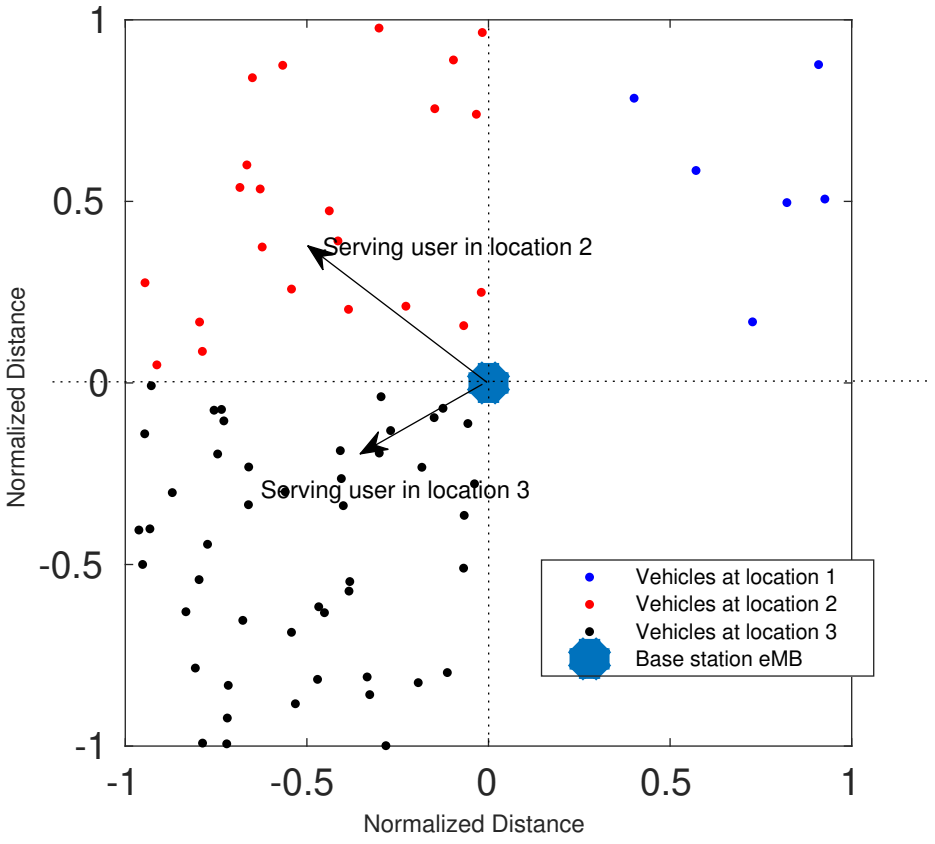


Figure 6.9: Distribution of users at four locations.

error in the above plot is high while the training error is low, it implies over-fitting of the network and hence regularization parameters may be adjusted; on the other hand, if both the validation and training errors are high, it implies under-fitting and hence the number of neurons (nodes) may be adjusted.

Fig. 6.11 shows the instantaneous RSS values for three fingerprints, when the number of vehicles is Poisson distributed, as shown in Fig. 6.9 having means of 5, 20 and 45. More explicitly, Fig. 6.11(a) depicts the RSS values for the scenarios, where the difference in the height of the obstacle to the line joining the transmitter and receiver is such that v of (6.2) is less than -0.7, i.e. the vehicular attenuation is zero. In other words, the attenuation of the signal is purely due to path loss. However, the fluctuations observed in Fig. 6.11a are owing to the fading introduced by the channel, which is captured in (6.5).

On the other hand, Fig. 6.11(b) depicts the RSS values for the scenario, where the difference in the height of obstacle to the line joining the transmitter and receiver is such that v of (6.2) is greater than -0.7. In other words, the signal experiences attenuation due to both the vehicles and path loss, which becomes evident from the plot, where there are points scattered around having RSS values as low as -125 dB. The physical meaning of this is that some of the beam-pairs available in the fingerprints are subjected

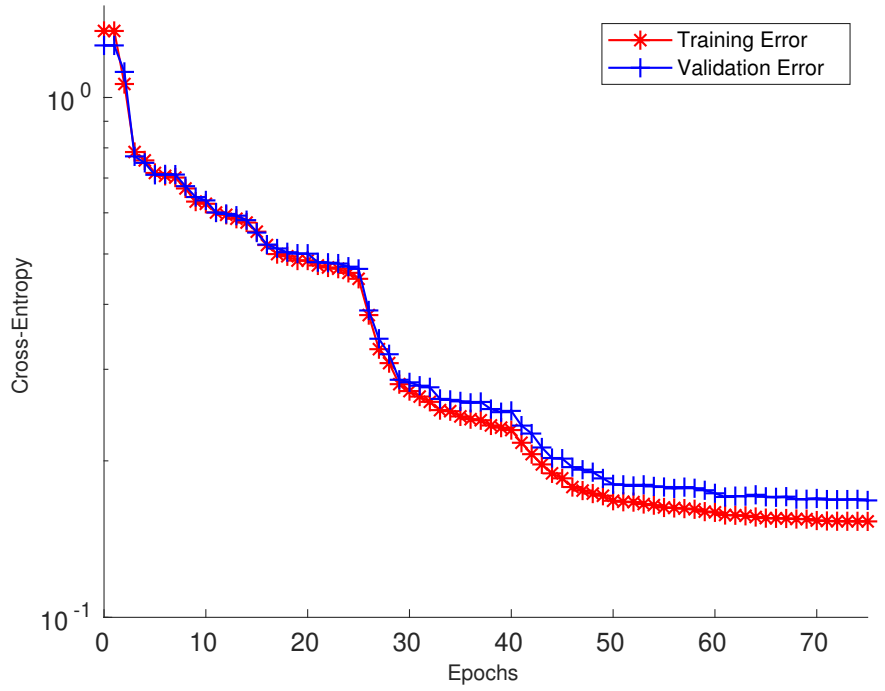


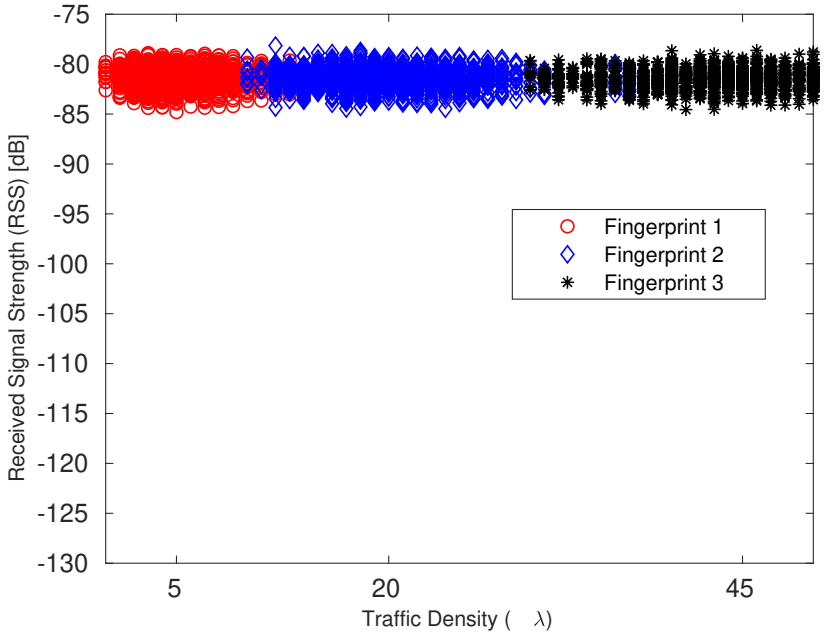
Figure 6.10: The training and validation curves of the neural network employed.

to blockages because of the vehicular obstruction. We note that in the absence of these blockages, Fig. 6.11(b) would show similar behavior to Fig. 6.11(a), where both both Fig. 6.11(a) and Fig. 6.11(b) are plotted for a non-adaptive system.

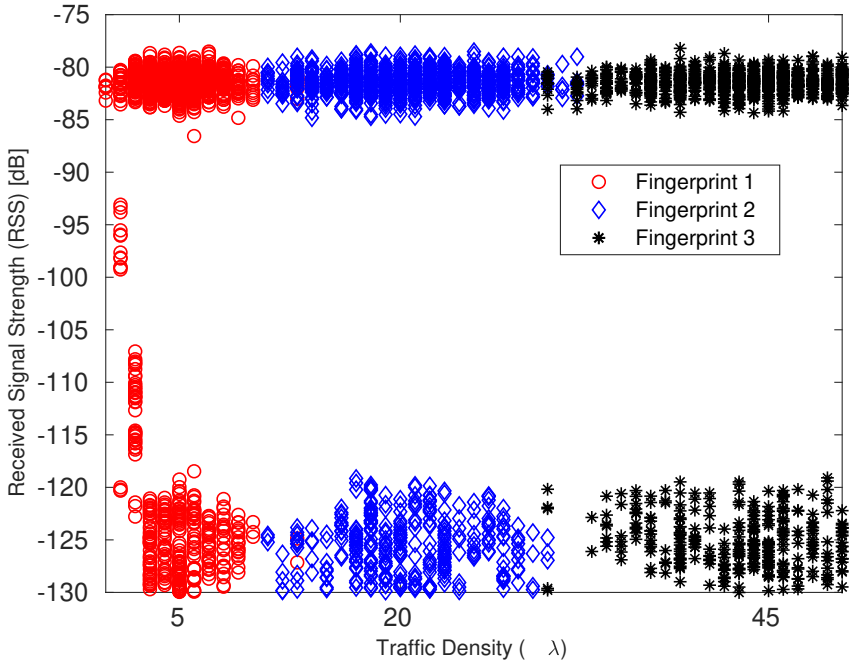
Fig. 6.12 shows the RSS observed for a system employing multiple fingerprints both with and without learning, a single fingerprint as well as for LOS[9] propagation. In this simulation, we have set the target RSS to -82 dBm. It is evident from Fig. 6.12 that multiple fingerprint based beam-alignment provides superior performance, while the performance of single fingerprint based beam-alignment and of LOS falls down precipitously with the increase of traffic density. This is because with the increase of traffic density, the probability of LOS blockage becomes high, hence resulting in low RSS. On the other hand, the single fingerprint based beam-alignment, which is designed for a given traffic density[10] and uses the same fingerprint for other traffic densities suffers from blockages, since the beam-pair suitable to one setting is unsuitable for another. Furthermore, the desired AoA-AoD pair which is suitable for that specific setting may be absent from the fingerprint constructed for another setting. Therefore, the performance is significantly affected. By contrast, the multiple fingerprint based design provides better performance; however, it can be seen that multiple fingerprint based adaptation without learning is unable to always maintain the target RSS. Instead it is hovering around it owing to the

[9]In LOS based beam-alignment, the beam is steered in the direction of the user, which is obtained from location, without account for blockages.

[10]In our simulation, we have designed the single fingerprint for a traffic density λ of 5.



(a) Instantaneous RSS values observed for $v < -0.7$ for three fingerprints at three traffic densities.



(b) Instantaneous RSS values observed for $v < -0.7$ for three fingerprints at three traffic densities

Figure 6.11: (a) Instantaneous RSS values observed for $v < -0.7$ for three fingerprints at three traffic densities. (b) Instantaneous RSS values observed for $v < -0.7$ for three fingerprints at three traffic densities.

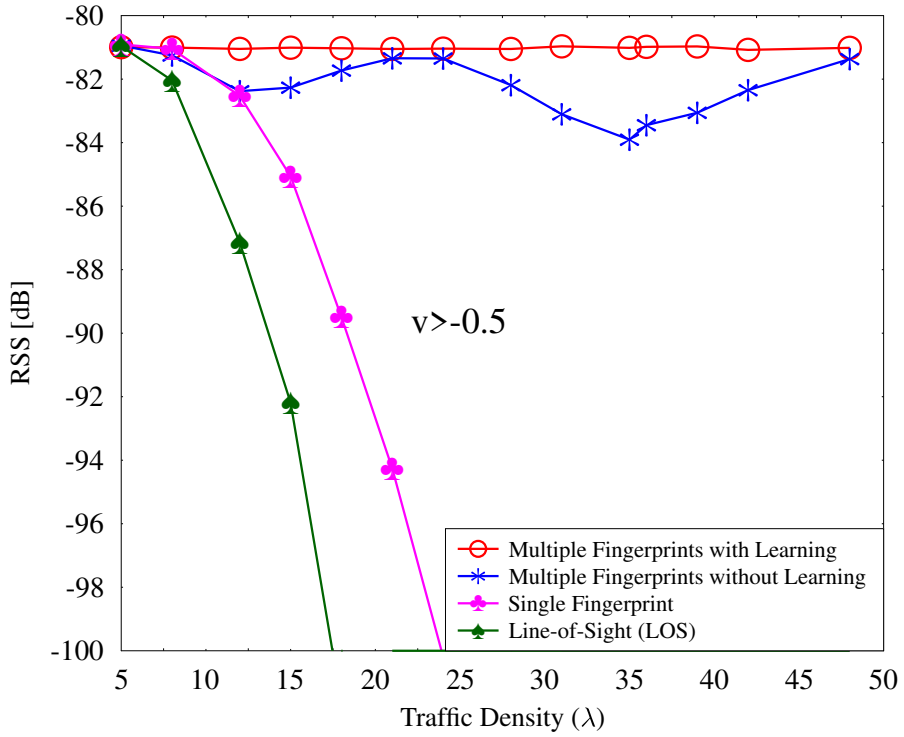


Figure 6.12: RSS of the different schemes against traffic density.

ever-changing channel statistics imposed by the changes in the environment. Hence, employing learning in our multiple fingerprint based design would intelligently adapt to the time-variant environment so that it always meets the target RSS, as shown in Fig. 6.12.

Fig. 6.13 shows the average RSS values over all the beam-pairs in the selected fingerprint after adaptation versus the traffic density. More explicitly, it is a measure of the total average received power observed within a given fingerprint, which is the power observed after post-processing, if the transmit power is equally shared across the beam-pairs in the fingerprint. The rationale of choosing this metric is to study the impact of the average RSS power over all the beam-pairs when the transmit power in each beam is constant. It is evident from the figure that the fingerprint selected using learning performs similarly to perfect beam-alignment, where an exhaustive beam sweeping is carried out. The physical significance is that the receiver is able to capture the signal from all the directions predicted by our learning algorithm. In other words, the learning accurately predicted the fingerprint that is comprised of maximum possible number of beam-pairs for successful link-connection, which is also observed in the average RSS. On the other hand, the fingerprint selection dispensing with the learning is significantly affected, as seen in Fig. 6.13, due to poor selection of the fingerprint, i.e. when the fingerprint associated with the wrong beam-pairs is selected. To elaborate further, the fingerprint that is selected does not contain the beam-pairs for successful transmission, which means

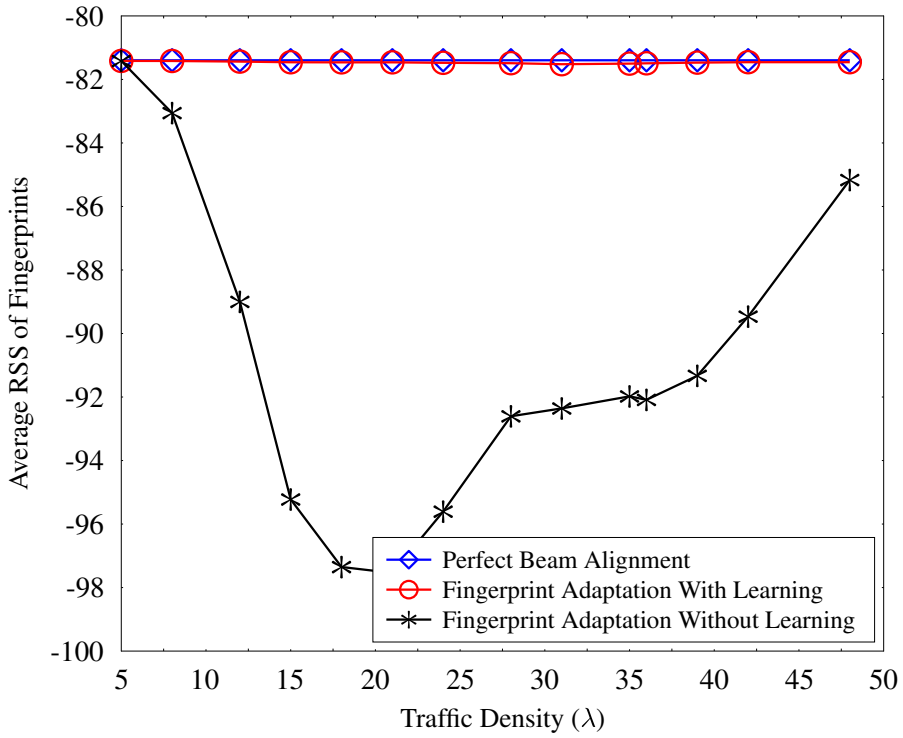
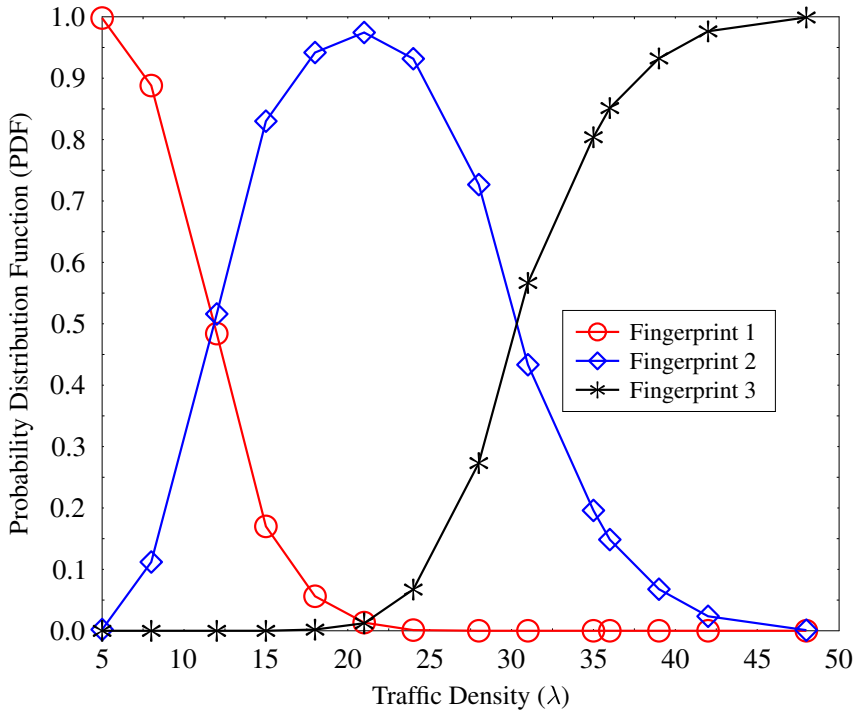


Figure 6.13: Average RSS values over all the beam-pairs in the fingerprint observed at the user for $v < -0.7$.

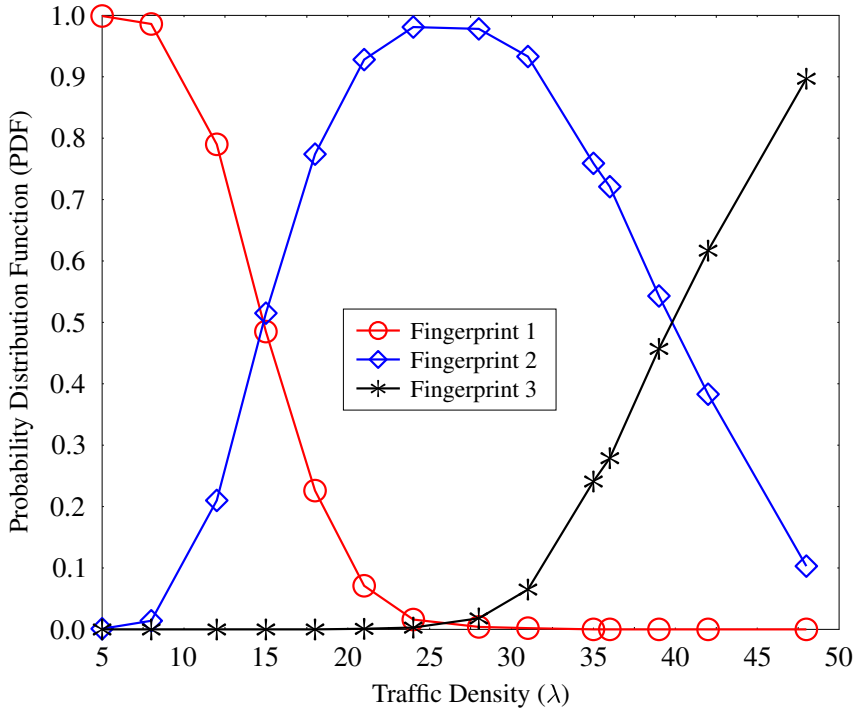
that the transmit power allotted to these beam-pairs suffer blockages or experience deep fading[11], hence resulting in low average RSS values at the user. It is important to emphasize that our proposed design achieves the performance of perfect beam-alignment at a significantly reduced search complexity.

Fig. 6.14 shows the PDF of the fingerprints against the traffic density. It can be seen in Fig. 6.14(a) that as the number of vehicles or as the traffic density increases, the PDF of fingerprint 1 starts falling gradually, while fingerprint 2 increases monotonically. Similarly, as the the traffic density increases further, the PDF of fingerprint 2 falls and the PDF of fingerprint 3 starts to increase. This implies that the set of beam-pairs in the fingerprint that provides a successful alignment starts to fall because of the increase in blockages caused by the increase in traffic density. Hence, the fingerprint which has the beam-pairs suitable for that environment is selected for a successful transmission. For example, when considering the traffic density range between 5-to-20, the PDF of fingerprint 1 falls because, whenever the set of beams in it is blocked due to the increased density of vehicles, it selects fingerprint 2 with the aid of the learning model developed during the training phase, since it provides alternate beam-pairs for link-connection.

[11] In this plot, since $v < -0.7$, the reduced RSS values should be attributed to only channel fading, as there is no attenuation due to vehicles. Hence the look-up table based fingerprint selection degrades the performance because of the channel imperfections as discussed in Sec. 6.2.1.



(a)



(b)

Figure 6.14: (a) PDF of fingerprints when learning is employed. (b) PDF of fingerprints dispensing with learning.

Fig. 6.14(b) shows the PDF of the fingerprints, when dispensing with learning. It can be noticed here that switching from one fingerprint to other fingerprint is different from that observed in Fig. 6.14(a), as the switching is decided based solely on the look-up table, which may be outdated because of the time-varying nature of the channel.

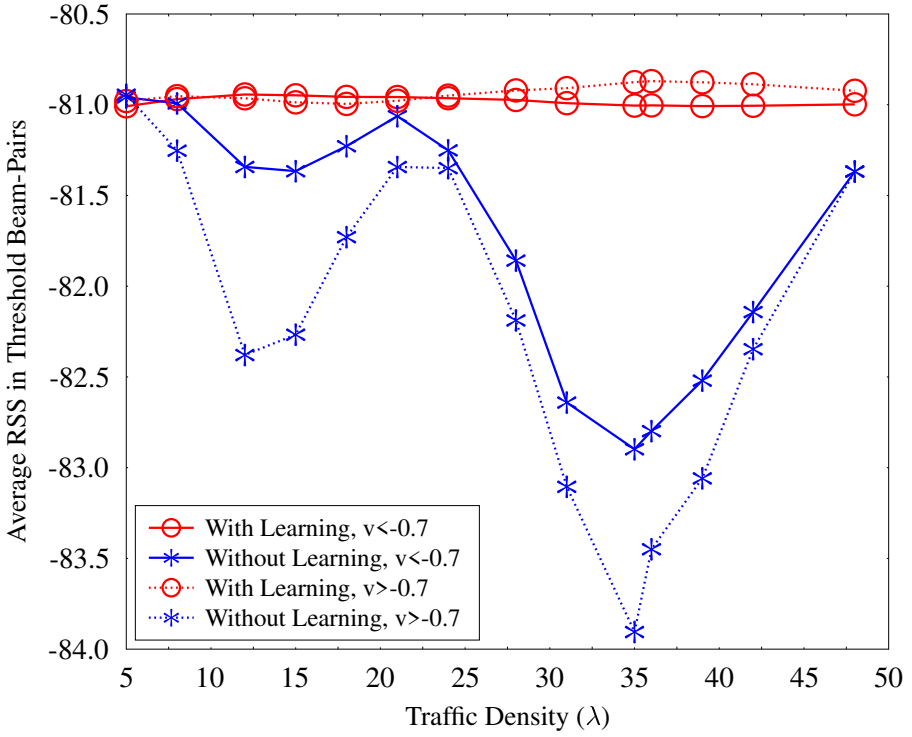


Figure 6.15: Average RSS values observed at the receiver in the specifically selected beam-pairs which meet the target RSS.

Fig. 6.15 shows the average RSS observed in the beam-pairs chosen from the selected fingerprint at the receiver. It is important to note the difference between Fig. 6.15 and Fig. 6.13, where the latter is the average RSS observed in all the beam-pairs of the selected fingerprint. Fig. 6.15 depicts the RSS for two values of v, i.e. for different vehicular heights. In this investigation, we have set the target RSS to -82 dB. It can be seen that for both cases of v the learning aided design performs about 1 dB higher than the target RSS. By contrast, the design relying on look-up table suffers significantly, especially in the traffic density regions between 10-15 and 30-40. This performance is akin to that of the conventional link-adaptation designs.

Fig. 6.16 shows the rate of our proposed design with learning and dispensing with learning. It can be seen from Fig. 6.16 that the rate of the design dispensing with learning is inferior to that of the learning-aided beam-alignment, while the design with learning provides the maximum rate of 2 bpcu regardless of the vehicular density. This is because

the design without learning relies on the threshold values of the look-up table which suffers from blockages in the traffic density regions between 10-15 and 30-40.

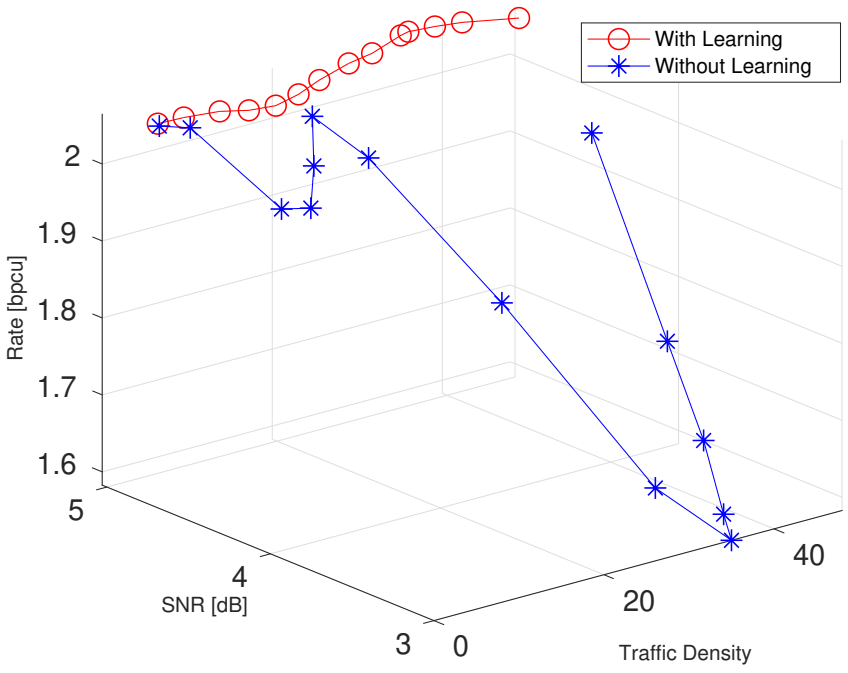


Figure 6.16: Rate observed at the receiver in the selected beam-pairs.

6.2.5 Complexity

In this section, we discuss the complexity of the designs presented. The conceptually simple beam-sweeping technique, imposes substantial complexity during the beam-search. Let us consider a signal departing at an angle $\theta_d \in (0, 360^\circ)$ from the transmitter, which is being received at an angle $\theta_r \in (0, 360^\circ)$ by the receiver. Let us also assume that the half-power beamwidth (HPBW) of the signal ray is β. Then a beam-sweeping based beam-alignment has to perform an exhaustive search over $\left(\frac{360}{\beta}\right)^2$ beam-pair combinations. On the other hand, the fingerprint based beam-alignment imposes a much lower complexity, since the number of beam-pairs is significantly reduced. We note that the complexity of the proposed design predominantly arises in the learning stage, which is carried out in an offline fashion. In other words, the weights of Fig. 6.5(b) are designed prior to any communication between the BS and the user, which then are stored in memory. However, this does not mean that our proposed design has no complexity constraints. The complexity in our design depends on the number of computations it has to perform while applying the weights in real time. To elaborate further, let us consider Fig. 6.5(b), and for the sake of simplicity let us assume that there are n neurons in each hidden layer, while the input vector $\mathbf{x}_i$ also has a dimension of n. Then the total number

of computations, i.e. additions and multiplications, needed for h hidden layers is on the order of $\mathcal{O}(hn^2)$. Furthermore, it is important to emphasize that the complexity of any learning algorithm depends on the input dimensionality [203,236] (the input vector $\mathbf{x}_i$ is referred as *feature set* in machine learning), which is however, only three-dimensional in our design.

6.3 Combating Channel Aging with Learning

Typically, in the FDD systems, the channel is estimated at the receiver in the downlink and fedback to the transmitter in the uplink, where the transmitter leverages the CSI to enhance the system performance. However, the CSI acquired at the transmitter is imperfect owing to the estimation errors introduced by the pilot-assisted CSI estimate as well as by the non-linear components in the hardware. The errors in the CSI at the transmitter would be exacerbated when the ubiquitous phenomenon of channel aging is considered [228]. To elaborate, the channel aging is a phenomenon that arises due to the time-varying nature of the channel. As a result the channel information arriving at the transmitter would become outdated. In this scenario, any signal processing performed at the transmitter relying on the outdated CSI would result in performance losses. Therefore, to circumvent the problem of channel aging, it is pertinent to predict the channel from the current estimates — where in this chapter we resort to a learning approach. In other words, we use the pilot-based channel estimate at the receiver, which would become outdated by the time it reaches the transmitting terminal, to predict the channel estimate at the transmitter. Then the transmitter employs the proposed HBF using the predicted channel estimate. More explicitly, in this section, we employ a learning strategy to predict the small scale fading coefficient of the channel matrix (Equation 2.19), where we invoke the radial basis neural network.

6.3.1 Channel Prediction

In learning-aided channel prediction design, the training is performed to design the weights using the outdated channel estimate and the actual channel. A pictorial illustration of outdated, of current and of predicted channel estimates is shown in Fig. 6.18.

More explicitly, the outdated channel samples and the corresponding actual (current) channel samples are used as the input and output of the network, respectively. In other words, the outdated channel is fed to the input layer, whose initial weights are randomly assigned from the distribution $\mathcal{N}(0,1)$ to obtain scores for the hidden layer, where the radial basis function $f_r(.)$ is employed. The score of neuron, v in the hidden layer of the

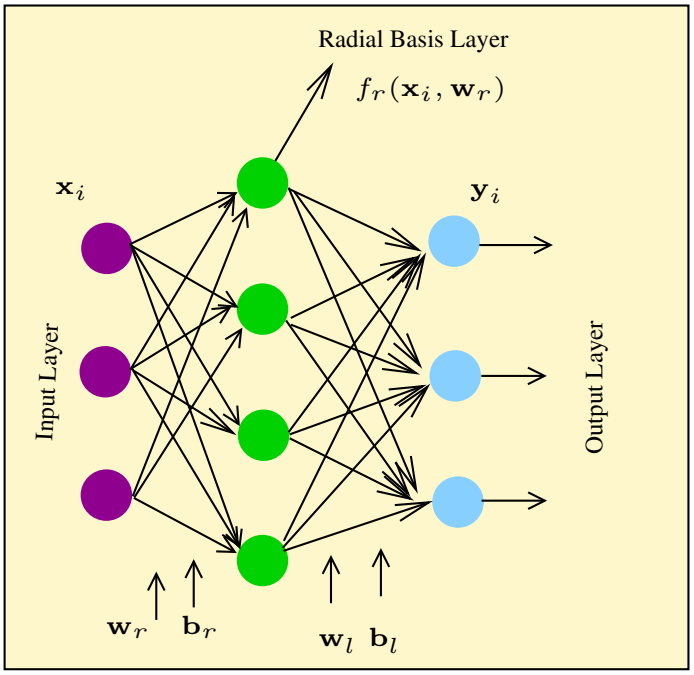


Figure 6.17: A typical radial basis neural network.

radial basis function is given by [131]

$$o_v = f_r(\|\mathbf{w}_r - \mathbf{x}_i\| \, \mathrm{b}_v), \tag{6.16}$$

where $\mathbf{w}_r, \mathbf{x}_i$ and b_v are the weight vector, the input and the bias, respectively, while $f(.)$ is the radial basis function expressed as [131]

$$f_r(n) = e^{-n^2}. \tag{6.17}$$

The score computed using (6.16) serves as the input to its succeeding layer, where a linear function is applied on the score obtained, which is expressed as

$$\mathbf{y}_i = \mathbf{W}_l \mathbf{o}_r + \mathbf{b}_l. \tag{6.18}$$

The score $\mathbf{y}_i$ obtained using (6.18) in the output layer is the predicted channel, which is then compared to the known actual channel to compute the error. The error between the predicted and the actual channel is evaluated using a loss function. The loss function used in this design is the mean square error given by [131]

$$L = \frac{1}{\mathcal{S}} \sum_{i=1}^{\mathcal{S}} \|\mathbf{y}_i - \mathbf{y}_i^t\| + \rho_1 \|\mathbf{W}_r\|_2^2 + \rho_2 \|\mathbf{W}_l\|_2^2, \tag{6.19}$$

where $\mathcal{S}$ is the total number of training samples, $\mathbf{y}^t$ is the current channel, while ρ_1, ρ_2 are the regularization factors used to avoid over-fitting [131]. Having obtained the loss function, the weights are then designed to minimize the loss function using back-propagation.

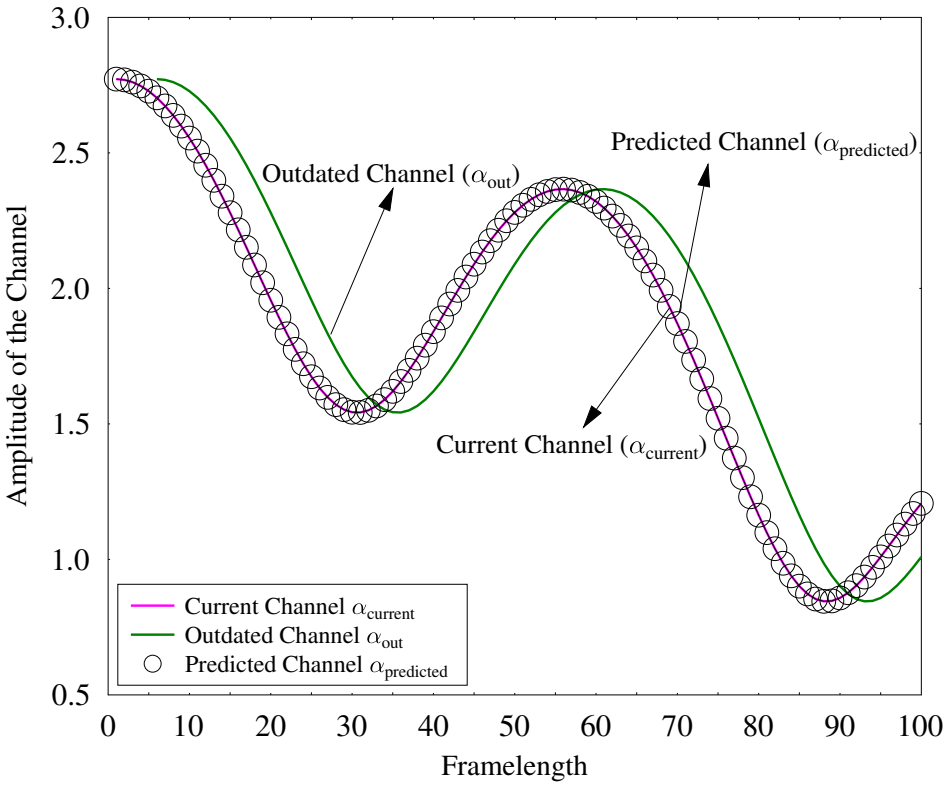
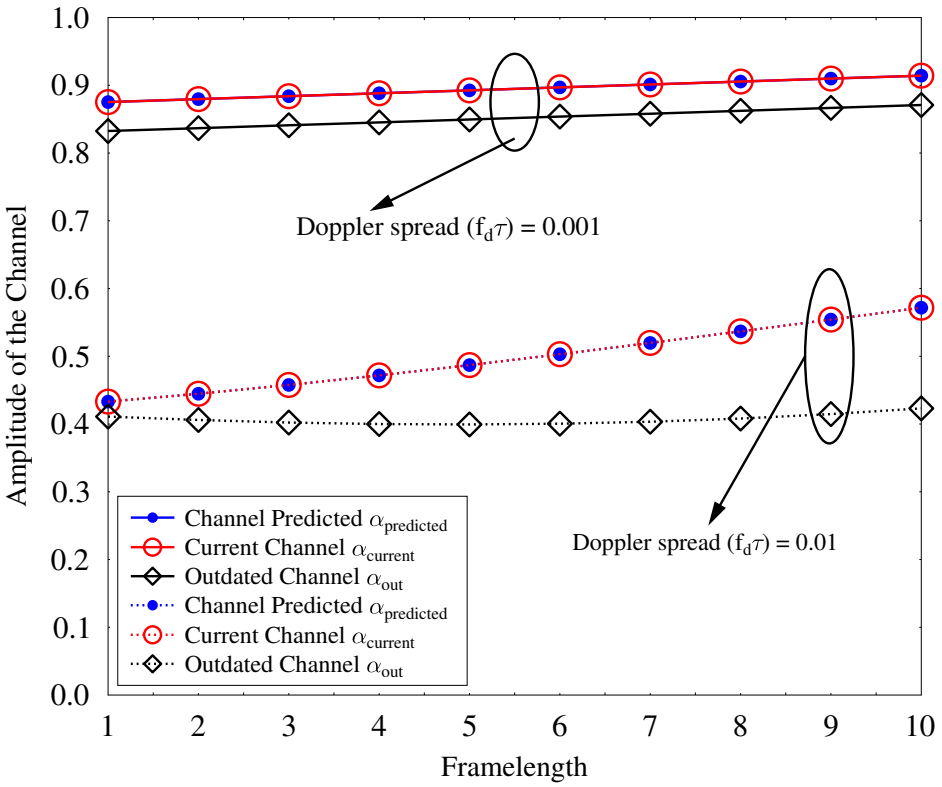


Figure 6.18: Pictorial representation of outdated, current and predicted channel estimates.

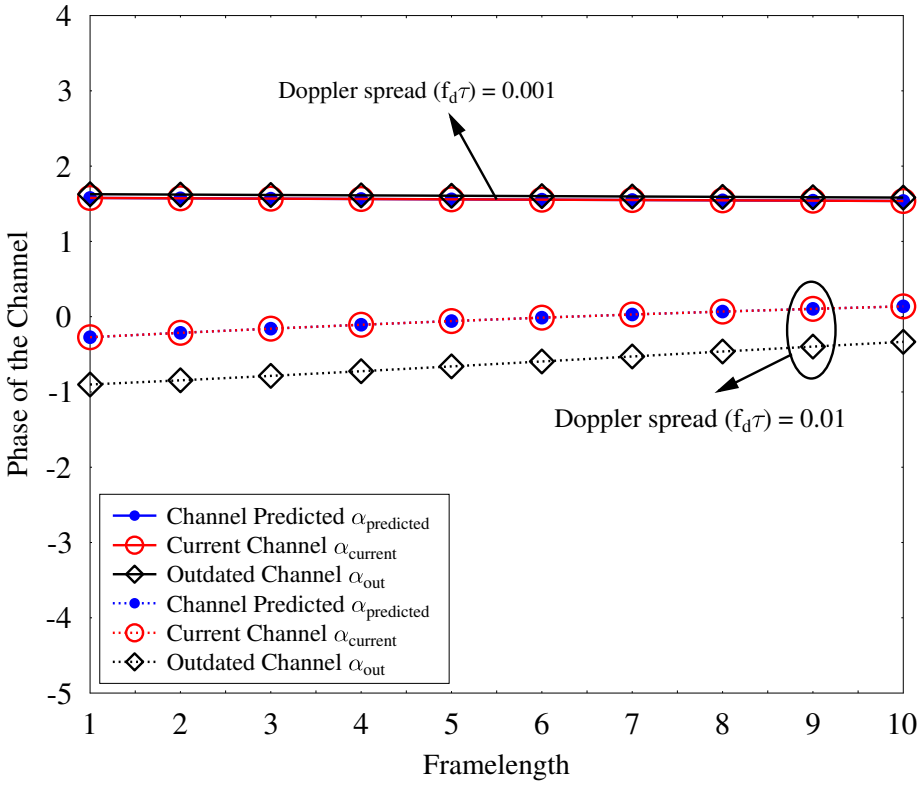
In back-propagation, the gradient of the loss function with respect to weights and biases is computed for each layer. A more detailed discussion on back-propagation is presented by Chauvin *et al.* in [132]. This process is continued until convergence is reached. In our design, $\mathbf{x}_i$ is the outdated channel vector α_{out} and $\mathbf{y}_i$ is the predicted channel vector α_{predict}. More explicitly, α_{out} is the channel learned at the receiver during pilot transmission which would become outdated at the transmitter because of the time-varying nature of the channel.

Let us now again consider the channel presented in (5.14) in the face of channel aging, where the AoA and AoD change slowly when compared to the small-scale fading coefficient α, which is the instantaneous CSI. More explicitly, α evolves with time according to the Jakes' model, whose autocorrelation is given by the zero order Bessel function of the first kind [222]. The correlation coefficient is given by

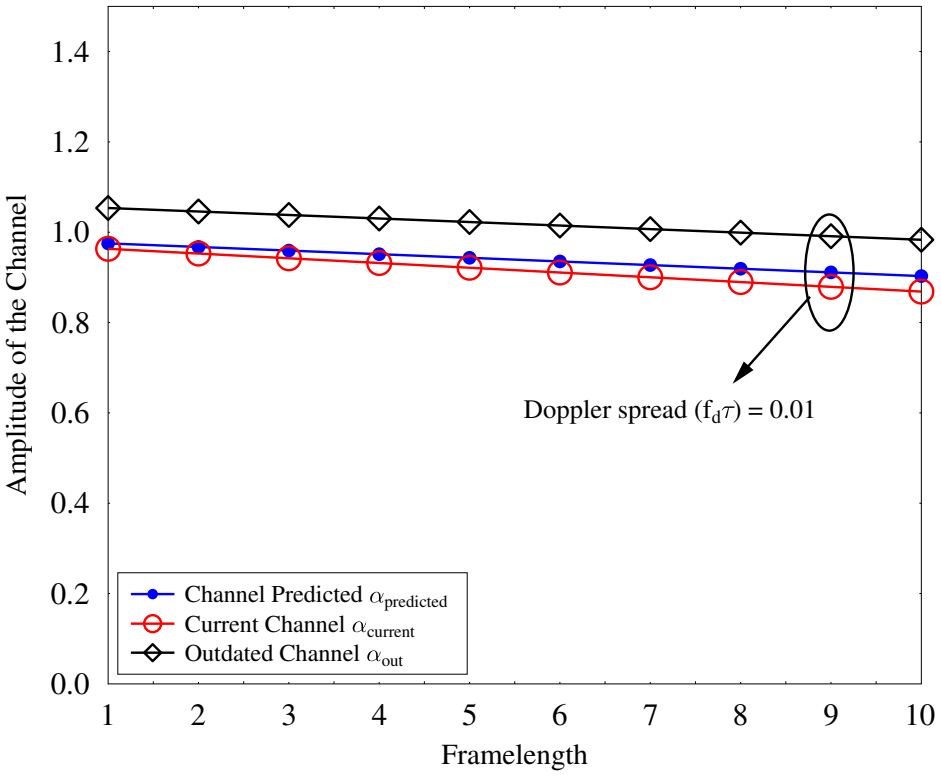
$$\zeta = J_{\circ}(2\pi f_d \tau),$$



(a) Amplitude of the CSI with 200 training samples.



(b) Phase of the CSI with 200 training samples.



(c) Amplitude of the CSI with 100 training samples.

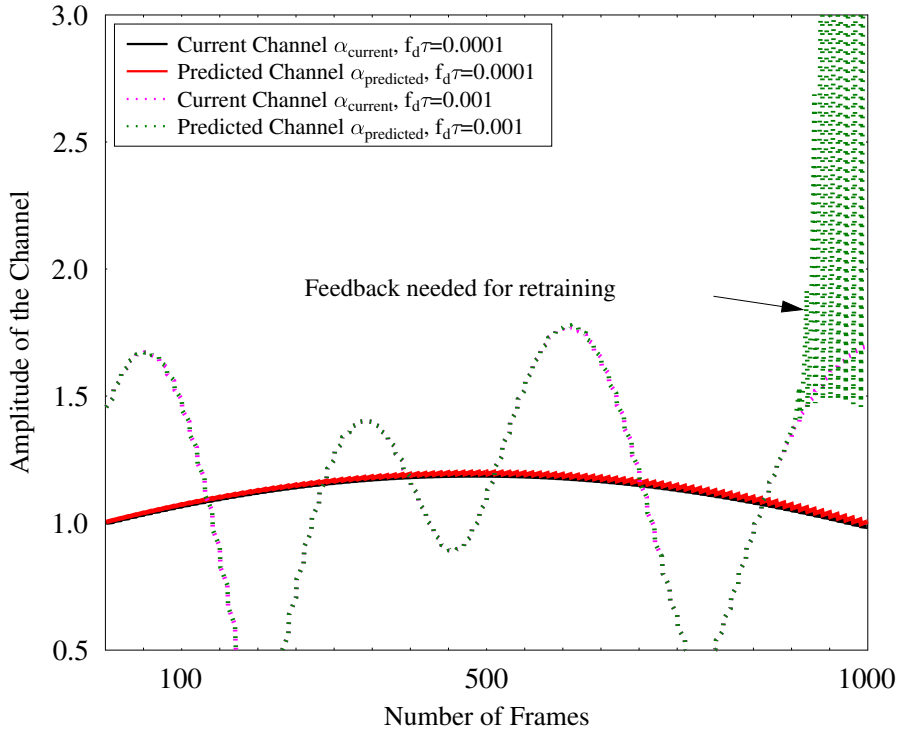
Figure 6.19: Amplitude and phase of the outdated channel, of the current channel and of the predicted channel.

where f_d is the maximum Doppler frequency and τ is the sample time.

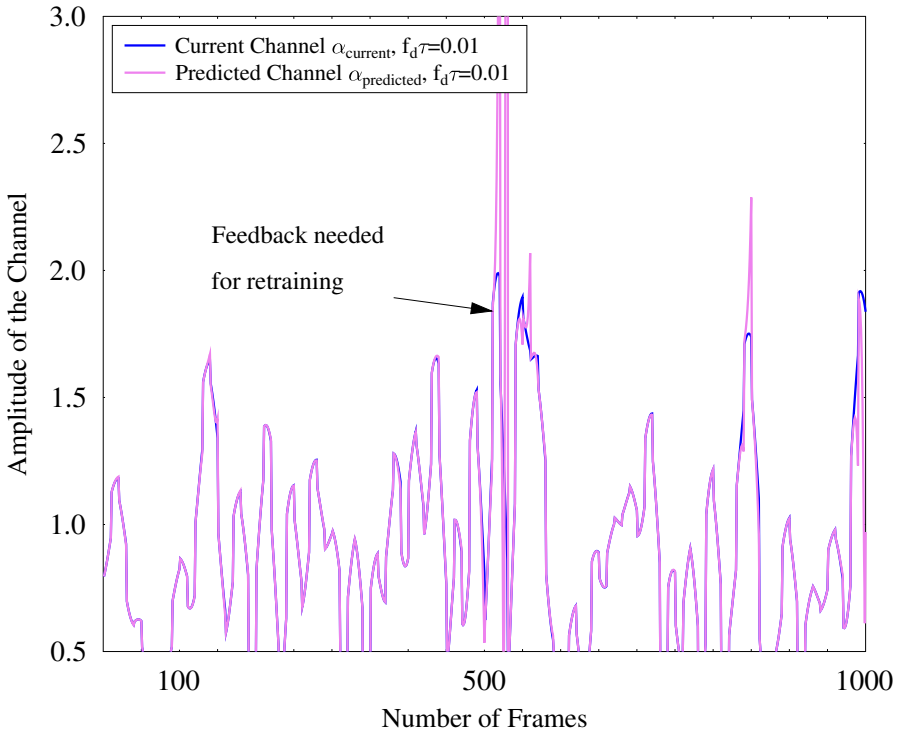
It is important to emphasize that the pilot-assisted channel estimate at the receiver which is outdated because of the mobility of the users is $\alpha_{\text{out}} = \alpha(\tau - 1)$, while the current channel estimate is $\alpha_{\text{current}} = \alpha(\tau)$. Note that during the training phase of the learning α_{out} and α_{current} are used as the training samples to design the weights of the neural network as discussed in the previous section. Having designed the training weights, the transmitter predicts the current channel estimate α_{predict} from the outdated channel estimate α_{out} in real time.

6.3.2 Performance Results

Fig. 6.19 shows the channel response in for Doppler spreads of 0.001 and 0.01. More explicitly, Fig. 6.19(a) and Fig. 6.19(b) characterizes the amplitude and phase of the



(a) Amplitude of the predicted channel and the current channel for $f_d\tau = 0.0001$ and 0.001.



(b) Amplitude of the predicted channel and the current channel for $f_d\tau = 0.01$.

Figure 6.20: Amplitude of the predicted channel and the current channel for different Doppler spread values.

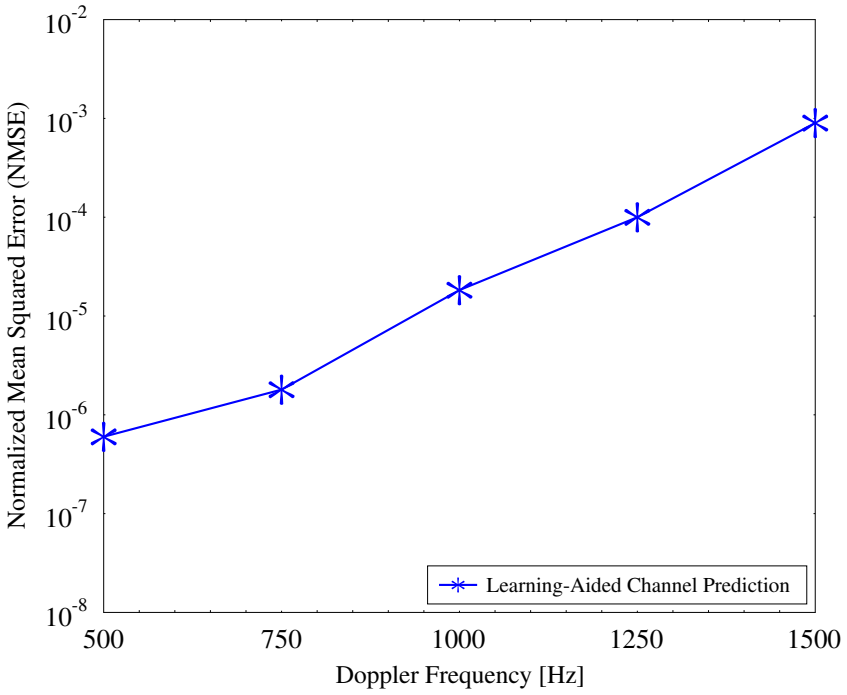
channel response, where 200 training samples are used to design the training weights of the neural network. It can be seen from Figures 6.19(a) and 6.19(b) that the learning-aided channel prediction has faithfully reproduced the current channel from the outdated channel both for a Doppler spread as high as 0.01 and for Doppler spread as low as 0.001. It is important to emphasize that the fidelity of the learning assisted channel prediction relies on the number of training samples used to design the training weights. This is evident from Fig. 6.19(c), where the predicted channel is a bit off the mark from the current channel, when only 100 training samples are used[12].

Fig. 6.20 shows the amplitude of the predicted channel and the current channel for different Doppler spread values. It is evident from the Fig. 6.20(a) that for a normalized Doppler spread as low as 0.0001, the neural network weights evaluated during training at the start of the frame are capable of predicting the channel even after 1000 frames. In other words, the CSI feedback from the receiver, which is estimated using pilots, is not needed at the transmitter. However, it can be also seen from Fig. 6.20(a) that as the Doppler spread is increased to 0.001, the weights of the neural network evaluated during training at the transmitter becomes unable to faithfully reproduce the channel after 800 frames. Therefore, feedback from the receiver becomes necessary for the neural network retraining after around 800 frames of transmission. Observe from Fig. 6.20(b) that as the Doppler spread is increased to 0.01, the number of frames transmitted using the same weights decreases. In other words, more frequent feedback from the receiver becomes necessary. Therefore, this implies that the overhead required for retraining the neural network depends on the Doppler spread.

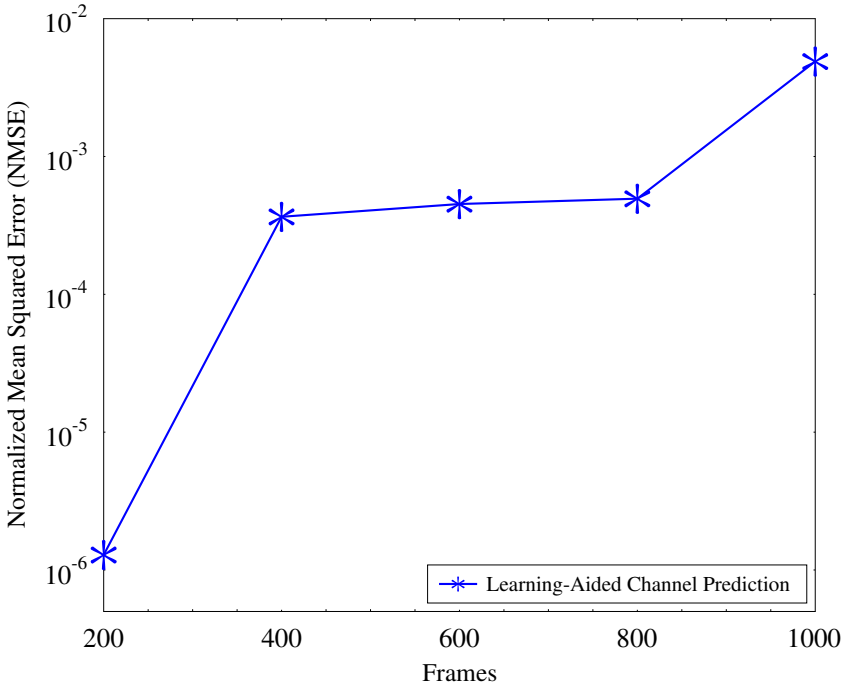
Fig. 6.21 shows the normalized mean squared error (NMSE) between the predicted channel and the current channel. More explicitly, Fig. 6.21(a) shows the NMSE at different Doppler frequencies. It can be noted from Fig. 6.21(a) that the NMSE is substantially lower between the predicted and current channel. In other words, the ANN is capable of reproducing the current channel with high accuracy. Although the NMSE increases with the Doppler frequency, it still is on the order of $\mathcal{O}(10^{-3})$. On the other hand, Fig. 6.21(b) shows the NMSE between the predicted and current channels when transmitting at different number of frames between a pair of training events. It can be seen from Fig. 6.21(b) that NMSE becomes insignificant for a low number of frames, while it is on the order of $\mathcal{O}(10^{-3})$ for say 800 frames. More explicitly, the physical significance of this is that the training weights designed before the transmission of the first frame can be applied without retraining for upto 800-1000 frame durations depending on the required NMSE.

Having procured both the AoA-AoD information and the current channel response, we now analyze the performance of the full duplex system presented in Chapter 4 with the predicted channel.

[12]The number of training samples required for faithful channel prediction would also depend on the Doppler spread.



(a) Normalized mean squared error between the predicted channel and the accurate channel at different Doppler frequencies.



(b) Normalized mean squared error between the predicted channel and the accurate channel at different number of frames for a Doppler spread of 0.001.

Figure 6.21: Normalized mean squared error between the predicted channel and the accurate channel.

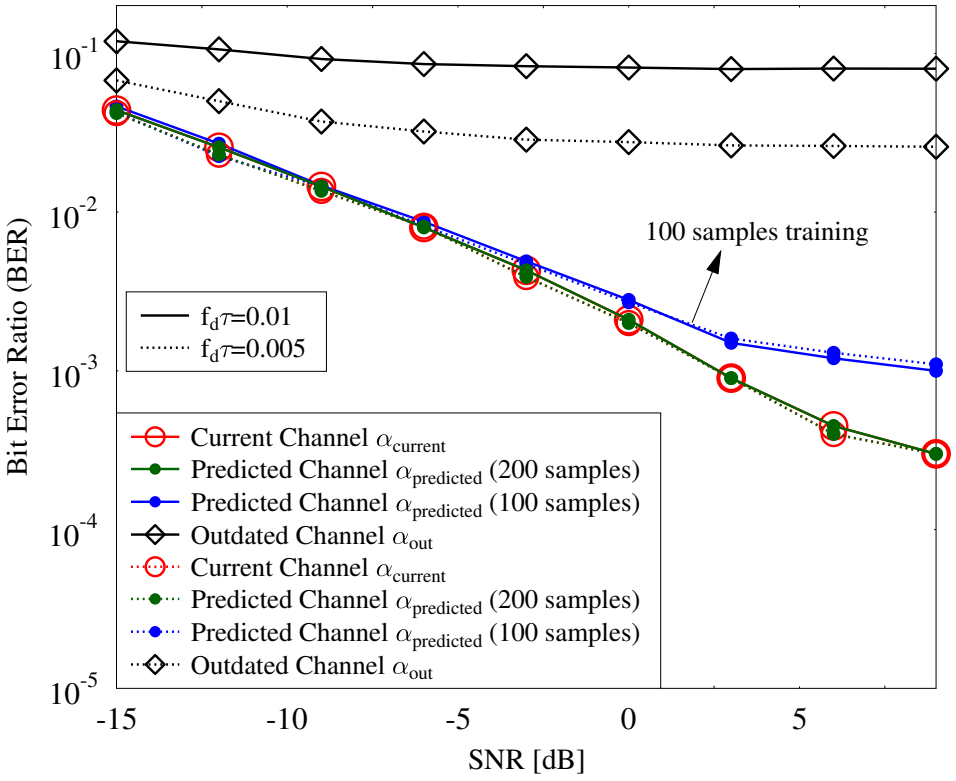


Figure 6.22: Characterizing the BER performance for Doppler spreads of 0.01 and 0.005 with predicted channel, outdated channel and current channel. In this simulation, BER performance is studied using predicted channel with both 100 and 200 training samples for designing the weights of the neural network. Furthermore, in this setting SI and MI is set to 3 dB.

Fig. 6.22 studies the BER performance of the system for Doppler spreads of 0.01 and 0.005. More particularly, it characterizes the BER performance of the channel with prediction, of the current channel and of the outdated channel. Furthermore, in Fig. 6.22 we study the BER performance using channel predicted with both 100 and 200 training samples for designing the training weights of the learning network. In this setting, we have set the SI and MI to 3 dB. It can be seen from Fig. 6.22 that for a Doppler spread as high as 0.01 and 0.005, the BER of the system relying on channel prediction is close to that of the system using 100 training samples, while it performs similarly to the current channel, when the number of training samples is increased to 200. On the other hand, the outdated channel produces an error floor for both Doppler spreads.

6.3.3 Complexity

The complexity of the radial basis neural network used for channel prediction involves two phases: offline and online. During the offline phase, which is the training phase, the weights of the neural network are designed and stored in memory. If the neural network has n neurons, then the number of computations required to train the weights is $\mathcal{O}(n^4)+\mathcal{O}(n^3)$. We note that the design of weights is carried out only once and stored in the memory during the training phase. Furthermore, the weights are designed at the BS, which may be equipped with high computational resources for channel prediction. However, during the online stage, the weights are applied to generate the output, where the number of computations is $\mathcal{O}(n^3)$.

For example, in our simulations, where we have set $N_t = 128, N_r = 32, N_s = 2$, the number of complex multiplications required for our propose design is $157,952$, while for SVD it is $688,128$. Furthermore, in our neural network setting, we have set the number of hidden layers to 1 and the number of neurons in the hidden layer to 250.

6.4 Conclusions

In this chapter, we have aimed for obtaining the CSI for employment by both the precoder and combiner designs discussed in Chapters 3 and 4. In mmWave systems, the channel information is comprised of two constituents: the AoA/AoD and the small-scale fading factor α, as discussed in Chapter 2.

Table 6.3 summarizes the basic methods for obtaining the AoA/AoD information.

Table 6.3: Summary of different methods to acquire AoA/AoD information.

AoA/AoD Acquisition Methods	**Features**
Conventional Estimation	Fails to exploit BF, hence resulting in low SNR
Beam sweeping	High search complexity
Hierarchical Codebook	Moderate search complexity, which is lower than that of beam sweeping
Fingerprint based alignment	Low search complexity

In this chapter, we primarily focused our discussions on fingerprint based beam alignment, as a benefit of its lower search complexity. More particularly, in Sec. 6.2.1 we

proposed a multi-fingerprint based database, where the fingerprints are collected for different traffic densities in a given location. Then, we improved our design by intelligently choosing the fingerprint based on the traffic density and location information relying on a neural network. Fig. 6.6 shows the signaling between the user and the BS at various stages. Figures 6.12 and 6.13 show the superior performance of the multifingerprint based beam-alignment assisted by learning. Table 6.4 provides a summary of all schemes considered in Figures 6.12 and 6.13 at a glance.

Table 6.4: Summary of maximum RSS and average RSS observed for different schemes of Figures 6.12 and 6.13.

		Schemes	**Max. RSS Strength**
Fig. 6.12	Traffic Density =15	LOS	-91 dBm
		Single Fingerprint	-84 dBm
		Multiple Fingerprints without Learning	-82.5 dBm
		Multiple Fingerprints with Learning	-81 dBm
	Traffic Density =30	LOS	< -100 dBm
		Single Fingerprint	< -100 dBm
		Multiple Fingerprints without Learning	-82 dBm
		Multiple Fingerprints with Learning	-81 dBm
		Schemes	**Avg. RSS Strength**
Fig. 6.13	Traffic Density =15	Fingerprint Adaptation without Learning	-95 dBm
		Fingerprint Adaptation with Learning	-81 dBm
		Perfect Beam Alignment	-81 dBm
	Traffic Density =30	Fingerprint Adaptation without Learning	-95 dBm
		Fingerprint Adaptation with Learning	-81 dBm
		Perfect Beam Alignment	-81 dBm

It is noted that for a target RSS threshold of -81 dBm, having multiple fingerprints

can provide a maximum RSS strength of -82.5 dBm and -82 dBm at traffic densities of 15 and 30 units, respectively. On the other hand, multiple fingerprints associated with learning capabilities achieve the target RSS for both traffic densities, while the LOS-based and single-fingerprint based schemes' RSS degrades substantially upon increasing traffic density. Furthermore, we have analyzed in Fig. 6.13 the average RSS observed in conjunction with fingerprint adaptation. We noted that fingerprint adaptation combined with learning is capable of providing a stable RSS of -81 dBm, which is similar to that of perfect beam alignment. requirements.

Based on these findings, we have also proposed a multi-functional beam transmission scheme, where multiple beams that satisfy the target received power are selected. Then, depending on the mobile station's requirement, the BS can opt for exploiting the additional beams for increasing the multiplexing/diversity gains. Furthermore, if the number of RF chains is lower than the number of beam pairs that satisfy the target receiver power, the BS then employs beam index modulation for increasing the spectral efficiency.

Then in Sec. 6.3, we have counteracted the effects of channel aging by proposing a learning-assisted channel prediction relying on a radial basis function neural network, where we showed by simulation that upon employing sufficient training our neural network aided channel prediction became capable of faithfully reproducing the channel. Fig. 6.18 shows the evolution of the current channel, outdated and predicted channel. We have observed in Fig. 6.19 that with a sufficient number of training samples the channel can be faithfully reproduced. However, we have noted in Fig. 6.20 that upon increasing the Doppler spread, the frequency of having to retrain the neural network weights increases. In other words, neural network needs to be retrained after a certain number of data frames, which is dependent on the Doppler spread, as seen in Fig. 6.20.

Chapter 7

Conclusions and Future Work

... when the characters in this thesis have told their story, you start again with new characters and equations for an exhilarating story for the posterity!

—The Author, K. Satyanarayana

In this chapter, we provide the summary and conclusions of the thesis, followed by future research directions for mmWave hybrid transceiver designs and their compelling applications.

7.1 Summary Conclusions

In this thesis, we have conceived the mmWave hybrid transceiver designs relying on machine learning tools. We commenced our study by analyzing the mmWave radio wave propagation effects, the array gain and link-budget analysis in Chapter 2. Then, in order to achieve both diversity and beamforming gains, in Chapter 3 we conceived a dual-function hybrid architecture and compared the sub-array and full-array solution. To enhance the throughput of the system, we have proposed a beamforming assisted self-interference cancellation technique for full duplex systems in Chapter 4. Both Chapters 3 and 4 deal with the dual-function transceivers. Subsequently, we extended hybrid beamforming to the MF MIMO system concept in Chapter 5, where we amalgamated MF MIMO transmitter with a learning assisted receiver. We note that the transceiver designs presented in Chapters 2- 5 rely on the perfect knowledge of CSI and beam alignment angle, while in Chapter 6, we dispense with these idealized simplifying assumptions. To elaborate further, in Chapter 6 we have proposed learning-assisted channel beam alignment as well as channel prediction to combat the channel aging phenomenon. The summary of the chapters and our major findings are provided below in a little more detail.

- **Chapter 2:** We studied the radio wave propagation effects of mmWave carriers in Sec. 2.1.1. Then in Sec. 2.2, we analyzed the array gain requirements for compensation of propagation losses, followed by our discussions on beamsteering, beamwidth and antenna types, namely on patch and horn antennas. Subsequently in Sec. 2.3, we further elaborated on the link-budget analysis discussing the number of AEs required for achieving a given spectral efficiency for a given transmit power. Furthermore, in Sec. 2.4 we presented the state-of-the-art both in mmWave hybrid architectures and in beamforming solutions, while in Sec. 2.5 we conceived an energy-efficient hybrid beamforming design by exploiting the specific nature of the mmWave channel. Additionally, in Sec. 2.6 we have also studied the state-of-the-art designs in the context of full duplex transceivers, where we employed beamforming techniques for self-interference cancellation. Finally, a rudimentary exposition on machine learning in the context of wireless communications is presented in Sec. 2.7.5.

- **Chapter 3:** In Sec. 3.2, a new hybrid BF architecture is conceived, where an AA is *partitioned* into two or more sub-arrays. In this architecture, the sub-arrays emerging from a full array are separated by a *sufficiently large* distance that minimizes the correlation between sub-arrays by exploiting their independent fading. The rationale behind the sub-array separation is to attain diversity gains in addition to the BF gains derived from the sub-arrays. In Sec. 3.2.1, we have analyzed the capacity of the proposed architecture and concluded that using *two sub-arrays* strikes an attractive compromise, since the diversity gain obtained from the two sub-arrays out-weighs the loss in the BF gain imposed by splitting the array. By contrast, dividing the array into more than two sub-arrays would result in a reduction of BF gain, which the additional diversity gain cannot compensate.

  Observe from the summary in Table 3.5 that to achieve a rate of 5.5 bpcu, the fully-connected and the conventional ASA designs require SNRs of -9.2 dB and -6.1 dB, respectively, while our proposed design achieves this rate at a lower SNR of -10 dB. However, we note that the diversity gain obtained reduces as the number of sub-arrays increases — this is due to the diminishing returns in terms of the diversity gains. Furthermore, in our design, each sub-array may be fully-connected, as shown in Fig. 3.2. We observed that the total number of phase shifters in our design is $N_t N_t^{RF}/N_{sub}$ as that of the conventional ASA, while it is $N_t N_t^{RF}$ in the fully-connected design.

  Then, in Sec. 3.3 we analyzed a codebook based on MUBs amalgamated with DFT-based analog BF in the RF stage for the aforementioned mmWave MIMO systems. Additionally, we investigated the performance of hybrid beamforming systems in the face of mutual coupling relying on a practical codebook using both DFT and MUB precoders of [98].

Subsequently in Sec. 3.4, we proposed an adaptive array design for hybrid beamforming in mmWave communication that adapts the architecture as well as the digital precoding depending on whether the channel is LOS or NLOS nature. Explicitly, we designed an architecture relying on an adaptive array, where the array is full-array-connected for communications over LOS channel with an antenna spacing of $\lambda/2$, and its sub-array-connected for NLOS channels. Explicitly, for the latter where the AA is partitioned into sub-arrays with sufficiently large separation so that each sub-array experiences independent fading.

Finally in Sec. 3.4.2, we utilized our adaptive array design of Section Sec. 3.4 to propose an adaptive virtual cell formation in the PHY layer of C-RAN for mmWave hybrid beamforming aided systems. More explicitly, the dynamic selection of RRHs is similar to a mmWave transmitter that adaptively switches between the fully-connected and ASA designs. In virtual cell formation, a user may maintain its connection with one or two RRH(s) depending on our proposed design criterion. We first present an algorithm for virtual cell formation designed for a single mobile user, where the RRHs connected to the user experience different channel conditions (LOS/NLOS). Then we extended our algorithm to a multi-user scenario, where the mobile users are distributed randomly and experience different channel conditions (LOS/NLOS) with respect to the RRHs. Moreover, given the typical fronthaul constraints of C-RANs, for the multi-user setting, we investigated considered the maximum number of users that can be supported by the fronthaul link.

- **Chapter 4:** In Sec. 4.2, we considered FD communication at mmWave frequencies relying on hybrid beamforming, where we aimed for mitigating the SI by jointly designing the transmit and receive RF beamformer weights and the precoder as well as combiner matrices. To design the beamformer, precoder and combiner matrices, we first obtained the fully-digital solution, where we resorted to an iterative algorithm relying on the idealized simplifying assumption of having perfect CSI knowledge. Then we derived the HBF solution from the digital solution using least-squares approximation. The proposed solution preserves the signal's dimensionality, while mitigating the SI. We show that the proposed design is capable of reducing the SI by upto 30 dB, hence performing similarly to the hypothetical interference-free FD system.

  In Sec. 4.2.2, we presented the mathematical proof for the convergence of the proposed iterative design, where the objective function is minimized in each iteration. Furthermore, we showed that the value of the objective function reduces in every iteration and it is lower-bounded by zero. We presented quantitative comparisons of our proposed design by simulation results in Sec 4.2.5, where we demonstrated that our proposed design achieves better performance gains than eigen beamforming, especially when the SI power is high. Based on our simulation results, we observed in Fig. 4.8 that when the SIR is as low as -25 dB, eigen beamforming

would require an SNR of 3 dB to achieve a rate of 22.5 bpcu, while our proposed design would need only -2.5 dB. On the other hand, when the INR is set to 10 dB, the SNR requirements are around 3 dB and -2.2 dB for the eigen beamformer and for our proposed design, respectively, to achieve the same rate of 22.5 bpcu.

Subsequently, we extended our design to the K-user frequency selective interference channels in Sec. 4.3, where the precoder and combiner is designed for minimizing both the SI and MI in mmWave systems using beamforming. In this design, we aimed for preserving the signal dimensionality, while mitigating both the SI and MI. Then, we developed an iterative matrix decomposition for hybrid precoding aided OFDM systems, where the digital TPC weights are employed in the OFDM scheme's frequency-domain, while the analog RF beamformer weights are applied to the time-domain signal. Figures 4.17 and 4.18 captured the performance of the system for different SIR and INR values. More particularly, we observed in Fig. 4.17 that when SIR_{MI} =-5 dB, INR_{SI}=-20 dB, the eigen beamforming weights required an SNR of 22.5 dB, which was as low as 19 dB for our design. Similarly, it is 20 dB and 15 dB for the eigen beamformer and for our design, respectively, when we have SIR_{MI} =-10 dB, INR_{SI} =-10 dB.

- **Chapter 5:** In Sec. 5.2, we proposed a learning assisted adaptive transceiver design for each user link based on the near-instantaneous post-processed SNR, where the adaptation regime switches between multiplexing versus diversity oriented transmission modes as well as by appropriately configuring the modulation mode employed so as to facilitate both high-reliability and high-rate operation. In this design, the receiver relies on the instantaneous post-processed SNR to decide on the transmitter's multiplexing versus diversity aided transmission mode and on the choice of the specific modulation scheme to be employed with the aid of supervised learning relying on the feed-back information forwarded to the BS. We invoked the KNN classification technique at the receiver for decision making, as a design example.

  We demonstrated by our simulations, presented in Sec. 5.2.4, that at a target BER of 10^{-3}, the learning-assisted adaptation achieves a significant SNR gain of about 5 dB, while maintaining the required target BER compared to that of conventional link-adaptation carried out based on hard threshold values. A qualitative discussion on the complexity of the KNN algorithm is presented in Sec. 5.2.3. In Sec. 5.3, we proposed deep learning assisted semi-blind detection for index modulation aided mmWave MIMO systems. More particularity, we proposed a detector design for MS-STSK operating without relying on the explicit knowledge of the CSI at the receiver. Then, we also extended our design to a BIM-aided MS-STSK scheme dispensing explicit CSI knowledge at the detector. This philosophy makes our design spectral-efficient, since it does not the need the pilots for channel estimation.

In Sec. 5.3.5, we demonstrated by simulations that our proposed design detects the MS-STSK information with high integrity while circumventing CSI estimation. Furthermore, in Fig. 5.16 we showed by simulations that our design outperforms the ML-aided detection in the face of channel impairments introduced during the CSI estimation. This has become possible by amalgamating the originally separate channel estimation and data detection, which lends an improved grade of flexibility to the receiver. Additionally, we demonstrated in Fig. 5.17 that the net DCMC capacity of the ML-aided detection is lower than that of our proposed learning-assisted detection. In other words, the DCMC capacity is significantly affected by the pilot overhead when having ML-aided detection, since it requires accurate CSI for every transmission frame.

To improve the fidelity of our learning-assisted detection, we proposed recalibration of the NN weights after a certain number of frames, which is contingent on the Doppler spread. The recalibration of the NN weights is carried out using side-information or training information, which is known at both the transmitter and receiver. Hence, we referred to our design semi-blind, since it requires the transmission of the training data for recalibration of the NN weights. We showed that the overhead needed for this training data is negligible compared to the pilot overhead for channel estimation. We show by simulations that our proposed design outperforms ML-aided detection in the face of channel estimation errors whose variance is as low as 0.15. A qualitative complexity discussion is presented both for the learning-assisted detection and for the ML-aided detection in terms of search space as well as the number of computations in Sec. 5.3.4 and Sec. 5.4.1.

- **Chapter 6:** In Sec. 6.2 we proposed a multi-fingerprint based database, where the fingerprints are collected for different traffic densities in a given location. Then, the BS intelligently chooses the fingerprint based on the traffic density and location information, where we invoked deep learning for the selection of the fingerprint. Then, in Sec. 6.2.2 we invoked a convolutional neural network, where the training weights are designed offline for the selection of the fingerprint. Upon the selection of the fingerprint, the BS then relays the information of the fingerprint selected to the mobile station using the control channel. Thereafter, the BS starts the training process to select the beam-pair from the fingerprint, which meets the target received signal power. The mobile station feeds back the index of the beam-pair from the selected fingerprint if it meets the threshold condition. This significantly reduces the search complexity involved.

  Additionally, in Section 6.2.3 we proposed a multi-functional beam transmission scheme as an application example of our proposed design, where multiple beams that satisfy the target received power are selected. Then, depending on the mobile station's requirement, the BS can leverage the additional beams for increasing the multiplexing/diversity gains. Furthermore, if the number of RF chains is lower

than the number of beam pairs available that meets the target received power, the BS then employs beam index modulation to increase the throughput.

Subsequently, in Sec. 6.2.4 we demonstrated by our simulations that having multiple fingerprint-based beam-alignment provides a superior performance over the single fingerprint based beam-alignment. Furthermore, we showed that our learning-aided multiple fingerprint design provides better fidelity than that of the scheme employing multiple fingerprints but dispensing with learning. Our discussion on the complexity of our proposed learning-aided beam-alignment are presented in Sec. 6.2.5, where we demonstrated that our design performs similarly to beam-sweeping based beam-alignment relying on high-complexity exhaustive beam-search.

On the other hand, to counteract the effects of channel aging, in Sec. 6.3 we proposed learning-assisted channel prediction relying on a radial basis function neural network, where we showed in Fig. 6.19 that upon involving sufficient training, the radial basis neural network-aided channel prediction can faithfully reproduce the current channel. We empirically showed in Sec. 6.3.2 that depending on the Doppler spread, the radial basis neural network used for channel predictions has to be periodically retrained. To this end, we demonstrated in Fig. 6.20 that the overhead involved in the CSI feedback for faithful reproduction of the actual channel at the transmitter is low for lower Doppler spreads.

7.2 Future Work

In this thesis, we have presented research ideas combining both conventional and machine learning conceived for mmWave hybrid beamforming systems. In this section, we briefly present some of the future work inspired by the ideas developed in this thesis.

7.2.1 Exploitation of Polarization

The conventional wisdom about mmWave communication channel is that it always exhibits LOS propagation, which may be exploited with the aid of directional antennas. However, it has been shown in [1,46] that in mobile communications, mmWave frequencies suffer from frequent LOS blockages and exhibit multipath propagation, which motivates us to study and analysis of the polarimetric properties of the mmWave channel. While the exploitation of polarization is promising, the cross-polarization associated with it can be detrimental to the overall performance which requires in-depth investigations.

Let us consider the single-user dual-polarized mmWave MIMO system shown in Fig. 7.1, where the transmitter and the receiver are equipped with $N_t/2$ and $N_r/2$ antennas, respectively. Furthermore, the AAs of both the transmitter and the receiver of Fig. 7.1 are dual-polarized, having both horizontal (H) and vertical (V) polarizations. Then the

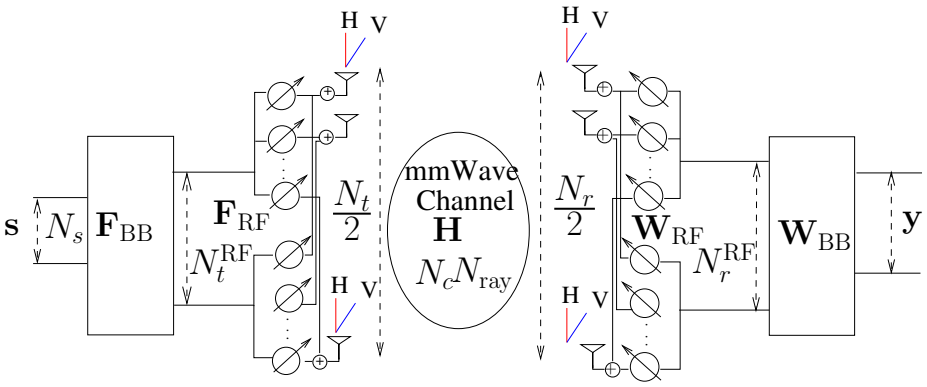


Figure 7.1: *Dual-polarized hybrid beamforming architecture.*

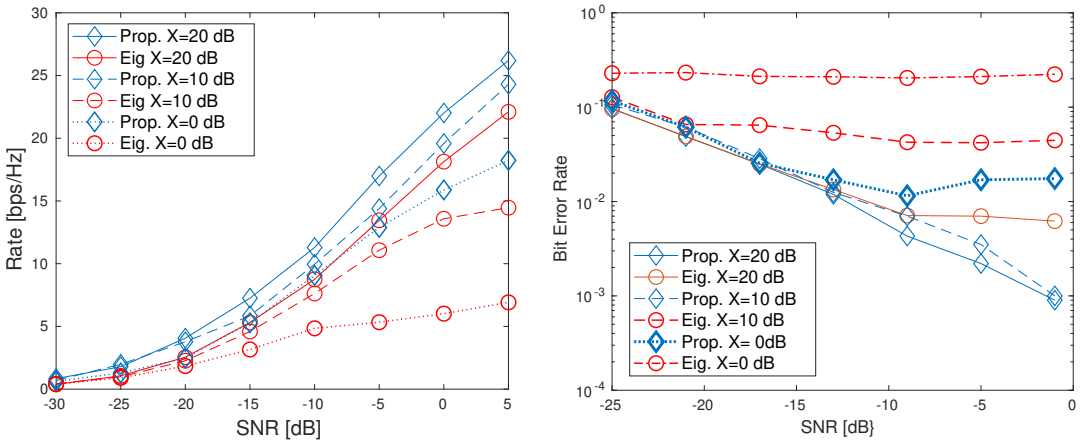
received signal vector $\mathbf{y}$ at the output of the RF stage and baseband combining is given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{W}_{\mathrm{BB}}^{\dagger}\mathbf{W}_{\mathrm{RF}}^{\dagger}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^{\dagger}\mathbf{W}_{\mathrm{RF}}^{\dagger}\mathbf{n}. \tag{7.1}$$

Furthermore, the channel matrix $\mathbf{H}$ of the dual-polarized AA of Fig. 7.1, is given by [40]

$$\mathbf{H} = \left[\mathbf{X} \odot \left\{\begin{bmatrix} e^{\mathrm{j}\angle\alpha_{\mathrm{hh},n_c}^{n_{\mathrm{ray}}}} & e^{\mathrm{j}\angle\alpha_{\mathrm{hv},n_c}^{n_{\mathrm{ray}}}} \\ e^{\mathrm{j}\angle\alpha_{\mathrm{vh},n_c}^{n_{\mathrm{ray}}}} & e^{\mathrm{j}\angle\alpha_{\mathrm{vv},n_c}^{n_{\mathrm{ray}}}} \end{bmatrix}\right\} \otimes \mathbf{H}'\right] \mathbf{G} = \begin{bmatrix} \mathbf{H}_{\mathrm{HH}} & \mathbf{H}_{\mathrm{HV}} \\ \mathbf{H}_{\mathrm{VH}} & \mathbf{H}_{\mathrm{VV}} \end{bmatrix},$$

where $\angle\alpha_{\mathrm{xy},n_c}^{n_{\mathrm{ray}}}$ is the initial random phase of the cluster n_c and ray n_{ray} that departs from the polarization 'x' and arrives in the polarization direction 'y', while $\mathbf{X}$ is the power-imbalance due to polarization and $\mathbf{H}'$ is the mmWave statistical channel model. Furthermore, $\mathbf{G}$ represents Givens' rotation that captures the difference in orientation between the transmitter and the receiver.



(a) Achievable rate. (b) Bit error rate.

Figure 7.2: Performance of 128×32 dual-polarized antenna (a) achievable rate (b) bit error rate.

Fig. 7.2(a) shows preliminary results of the rate of the algorithm presented in Chapter (4) as well as of the eigen beamforming design using 128×32-element dual-polarized AA. Explicitly, the parameters in (2.20), namely θ, ϕ are uniformly distributed from 0-to-2π and $N_t^{\mathrm{RF}}, N_r^{\mathrm{RF}}, N_s$ are set to 2 whilst $\mathcal{X}$ is varied from 0-to-20 dB, where $\mathcal{X}$ is defined as the ratio of co-polarization power (ρ_{xx}) to cross polarization power (ρ_{xy}). In this configuration, two streams are transmitted using two RF chains. It can be seen from Fig. 7.2(a) that for lower values of $\mathcal{X}$, the eigen beamforming rate-curve saturates at lower rates, while the proposed design outperforms eigen beamforming with SNR gain of more than 10 dB, especially when $\mathcal{X}$ is as low as 0 and 10 dB. Furthermore, when $\mathcal{X}$ is increased to 20 dB, the proposed design represented by the diamonds outperforms the eigen beamforming by about 5 dB.

Additionally, to understand the reliability of the system using cross-polarization, Fig. 7.2(b) shows some early results on the BER of both the proposed as well as of the eigen beamforming designs. It is interesting to note that at SNR of -1 dB, the proposed design achieves a BER as low as 10^{-3} while the eigen beamforming produces an error floor for $\mathcal{X} = 10$ dB. By contrast, when the cross-polarization power leakage is high, i.e. $\mathcal{X} = 0$ dB, both produce an error floor. However, the proposed design produces an error floor at a low BER. On the other hand, when the cross-polarization is low, i.e. $\mathcal{X} = 20$ dB, the proposed design outperforms the eigen beamforming by a significant margin.

7.2.2 Beam Squint Compensation

In this thesis, we assumed that the AA response vector experiences the same phase shift for all frequencies. However, we know that the phase-shifting of $\phi = \omega t$ is a linear function of the time delay. However, for signals having a high bandwidth the group-delay tends to be different at different frequencies. Hence the phase of the high complex-valued array responses also becomes different at different frequencies. This change or variation of array response with frequencies results in frequency-dependent beam-direction. More explicitly, the beam direction for a frequency of $(f_\circ + \delta f)$ points to an angle of $(\theta_\circ + \delta\theta)$, instead of at $\theta_\circ$, as shown in Fig 7.3. The new array response vector is expressed as

$$\mathbf{a}(\theta) = [1 \; e^{j\frac{2\pi}{\lambda}d\cos(\theta_\circ)+j\frac{2\pi}{\lambda}d\cos(\delta\theta)} \ldots e^{j\frac{2\pi}{(}N_t-1)\lambda d\cos(\theta_\circ)+j\frac{2\pi}{\lambda}d\cos(\delta\theta)}]^T. \tag{7.2}$$

$$\tag{7.3}$$

We can see from Fig. 7.3 that the signal at frequency f_2 is squinted by an angle of $\delta\theta$, which significantly reduces the beamforming gain and hence the throughput of the system, since it suffers from reduced SNR. Therefore, in order to have a flawless signal transmission, compensation of the beam squinting becomes necessary [239, 240].

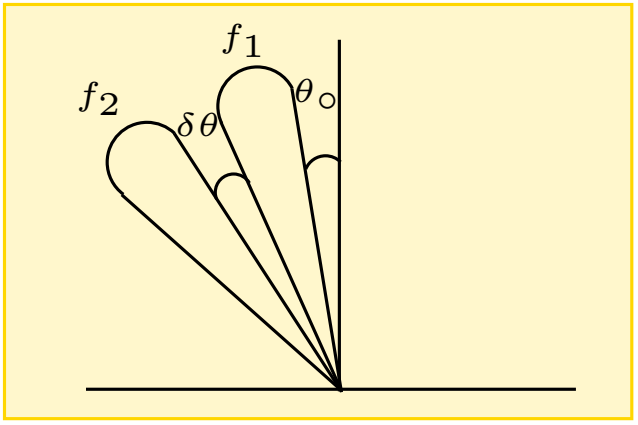


Figure 7.3: Beam squint for two frequencies f_1 and f_2.

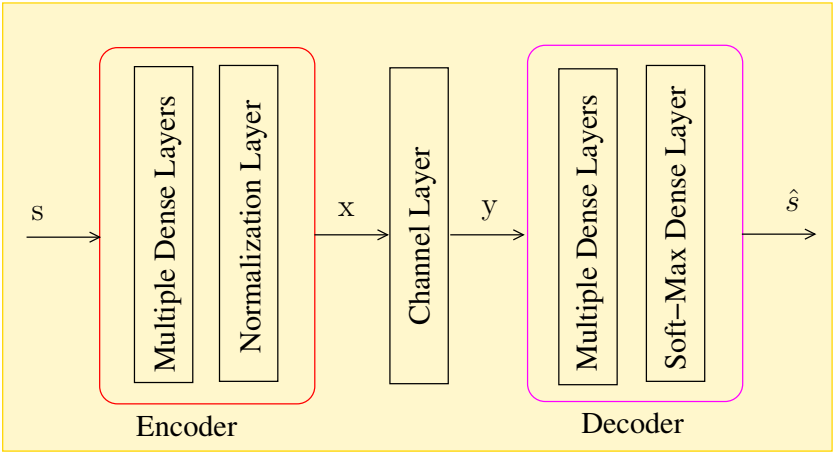


Figure 7.4: A typical SISO autoencoder.

7.2.3 Autoencoder-Assisted mmWave Hybrid Systems

In Chapter 6 we have used learning-aided soft-detection for reducing the detection complexity. However, the performance of the system may be further improved by efficient precoding and decoding using machine learning [35, 146, 148]. More particularly, autoencoder based learning design may be employed, where the neural network intelligently choses the precoding and combing weights. The philosophy of autoencoders is that they can learn an approximate function between the input and output, where the output is similar to the input. In contrast to the supervised learning discussed in Chapters 5 and 6, autoencoders rely on unsupervised learning. A typical SISO autoencoder is shown in Fig. 7.4. The autoencoder of Fig. 7.4 maps the input bit sequence s to a stream of complex-valued signals x, which is then transmitted. Then the received signal vector y is demapped to reconstruct the bit sequence $\hat{s}$. Mathematically, this can be expressed as

$$\mathbf{x} = \mathbf{W}\mathbf{s} + \mathbf{b} \tag{7.4}$$

$$\hat{\mathbf{s}} = \mathbf{W}'\mathbf{y} + \mathbf{b}', \tag{7.5}$$

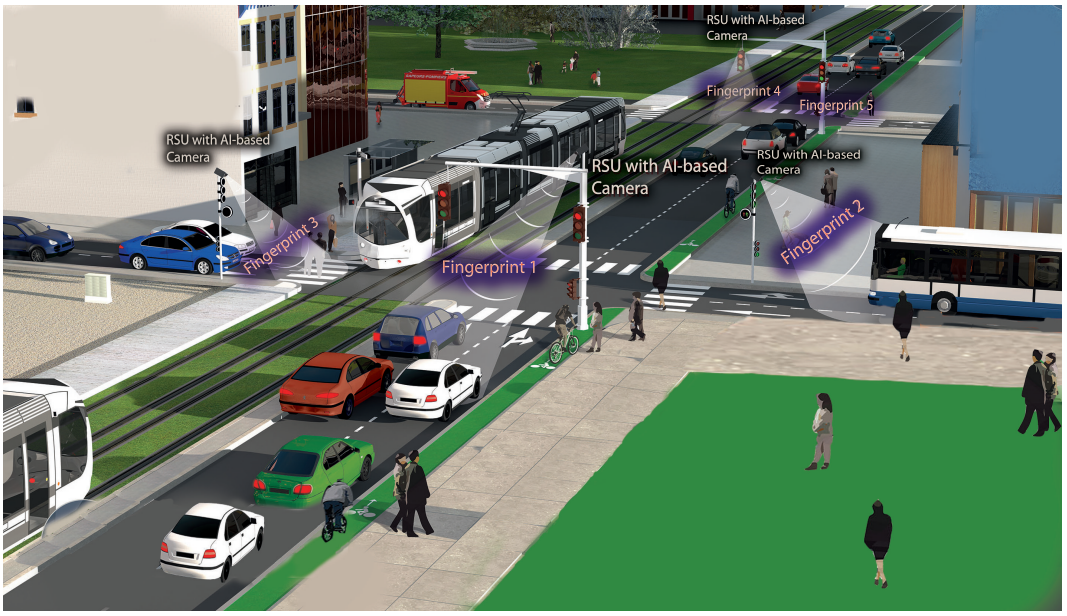


Figure 7.5: Fingerprint based beam-alignment for fast initial access.

where $\mathbf{W}, \mathbf{W}', \mathbf{b}$, and $\mathbf{b}'$ are the weights and biases of the neural network.

At this point, a question arises: can autoencoder shown in Fig. 7.4 be utilized to replace the conventional precoder and combiner designs for mmWave hybrid systems, given the AoA and AoD information as the input to the encoder?

7.2.4 Beneficial Applications

An immediate application of the design proposed in Chapter 6 can readily be seen for fast initial access in mmWave systems relying on hybrid beamforming [241, 242].

For example, Fig. 7.5 shows an application of fingerprint based beam-alignment for prompt initial access at mmWave frequencies. In Fig. 7.5 the road side unit (RSU) equipped with all AI-based camera is communicating with the users on the pavement. In this scenario, the RSU typically would aim for aligning the beam-pairs by performing an exhaustive beam-sweeping. This would pose high search-complexity. Therefore, to circumvent the exhaustive search, our design proposed in Chapter 5 may be invoked. More particularly, our deep learning-aided design based on the location of the user and on the traffic density, rapidly aligns the beam-pairs. Furthermore, given the mobility of the users, especially among the vehicular users of Fig. 7.5, the CSI necessary for detection becomes outdated. In this scenario, our channel prediction algorithm proposed in 6.3 can be invoked for faithful reproduction of the CSI.

References

[1] I. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions and link-budget," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2018.

[2] G. Maccartney, M. Samimi, and T. Rappaport, "Exploiting directionality for millimeter-wave wireless system improvement," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 2416–2422.

[3] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, July 2016.

[4] K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Dual-function hybrid beamforming and transmit diversity aided millimeter wave architecture," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–1, 2017.

[5] K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A. Mourad, and L. Hanzo, "Millimeter wave hybrid beamforming with DFT-MUB aided precoder codebook design," *Proc. VTC*, Sept. 2017.

[6] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," *ACM SIGCOMM*, vol. 43, no. 4, pp. 375–386, 2013.

[7] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," in *Proc. ASILOMAR*, 2011, pp. 2002–2006.

[8] A. Sahai, G. Patel, C. Dick, and A. Sabharwal, "On the impact of phase noise on active cancelation in wireless full-duplex," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4494–4510, Nov 2013.

[9] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. ASILOMAR*, 2010, pp. 1558–1562.

[10] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," *Sensors*, vol. 16, no. 7, 2016. [Online]. Available: http://www.mdpi.com/1424-8220/16/7/1130

[11] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sept 2014.

[12] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, December 2012.

[13] Z. Xiao, P. Xia, and X. G. Xia, "Full-duplex millimeter-wave communication," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 136–143, Dec 2017.

[14] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec 2011.

[15] Y. Liu, X. G. Xia, and H. Zhang, "Distributed linear convolutional space-time coding for two-relay full-duplex asynchronous cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6406–6417, December 2013.

[16] I. Krikidis, H. A. Suraweera, P. J. Smith, and C. Yuen, "Full-duplex relay selection for amplify-and-forward cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4381–4393, December 2012.

[17] L. Li, C. Dong, L. Wang, and L. Hanzo, "Spectral-efficient bidirectional decode-and-forward relaying for full-duplex communication," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7010–7020, Sept 2016.

[18] C. T. Lin, F. S. Tseng, and W. R. Wu, "MMSE transceiver design for full-duplex MIMO relay systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6849–6861, Aug 2017.

[19] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in mimo-ofdma systems with full-duplex and hybrid relaying," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1291–1304, May 2012.

[20] S. Wang, Y. Liu, W. Zhang, and H. Zhang, "Achievable rates of full-duplex massive mimo relay systems over rician fading channels," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 9825–9837, Nov 2017.

[21] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex mimo relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541–1553, September 2012.

[22] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, February 2014.

[23] J.-S. Jiang and M. A. Ingram, "Spherical-wave model for short-range MIMO," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1534–1541, Sept 2005.

[24] Z. Wei, Y. Zhao, X. Liu, and Z. Feng, "DoA-LF: A location fingerprint positioning algorithm with millimeter-wave," *IEEE Access*, vol. 5, pp. 22 678–22 688, 2017.

[25] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things J.*, vol. 3, no. 6, pp. 1113–1123, Dec 2016.

[26] Z. E. Khatab, A. Hajihoseini, and S. A. Ghorashi, "A fingerprint method for indoor localization using autoencoder based deep extreme learning machine," *IEEE Sensors Letters*, vol. 2, no. 1, pp. 1–4, March 2018.

[27] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan 2017.

[28] A. Capone, I. Filippini, V. Sciancalepore, and D. Tremolada, "Obstacle avoidance cell discovery using mm-waves directive antennas in 5G networks," in *Proc. PIMRC*, Aug 2015, pp. 2349–2353.

[29] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "Confi: Convolutional neural networks based indoor wi-fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[30] R. C. Daniels, C. M. Caramanis, and R. Heath, "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 114–126, Jan 2010.

[31] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2012.

[32] R. Daniels and R. Heath, "Online adaptive modulation and coding with support vector machines," in *in Proc. European Wireless Conf.*, April 2010, pp. 718–724.

[33] Z. Puljiz, M. Park, and R. Heath, "A machine learning approach to link adaptation for SC-FDE system," in *Proc. Globecom*, Dec 2011, pp. 1–5.

[34] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan 2017.

[35] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *CoRR*, vol. abs/1805.07631, 2018. [Online]. Available: http://arxiv.org/abs/1805.07631

[36] K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1394–1404, Feb 2019.

[37] ——, "Adaptive transceiver design for C-RAN in mmwave communications," *IEEE Access*, vol. 6, pp. 16 770–16 782, 2018.

[38] K. Satyanarayana, M. El-Hajjar, A. Mourad, and L. Hanzo, "Multi-user full duplex transceiver design for mmwave systems using learning-aided channel prediction," *IEEE Access*, vol. 7, pp. 66 068–66 083, 2019.

[39] ——, "Deep learning aided fingerprint based beam alignment for mmwave vehicular communication," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2019.

[40] K. Satyanarayana, T. Ivanescu, M. El-Hajjar, P. . Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for dual-polarised millimetre wave mimo systems," *Electronics Letters*, vol. 54, no. 22, pp. 1257–1258, 2018.

[41] D. Prisiceanu, K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Effects of mutual coupling on lattice reduction-aided millimeter wave hybrid beamforming," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2018, pp. 1–5.

[42] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[43] T. Riihonen, S. Werner, and R. Wichman, "Spatial loop interference suppression in full-duplex mimo relays," in *Proc. ASILOMAR*, Nov 2009, pp. 1508–1512.

[44] T. Rappaport, R. Heath, R. Daniels, and J. Murdock, *Millimeter Wave Wireless Communications*, ser. Prentice Hall Communications Engineering and Emerging Technologies Series. Prentice Hall, 2015. [Online]. Available: https://books.google.co.uk/books?id=_Tt_BAAAQBAJ

[45] H. T. Friis, "A Note on a Simple Transmission Formula," in *Proceedings of the IRE*, May 1946, pp. 254–256.

[46] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5g) wireless networksâĂŤwith a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, Dec 2017.

[47] T. S. Rappaport and S. Deng, "73 ghz wideband millimeter-wave foliage and ground reflection measurements and models," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1238–1243.

[48] J. Rosen and L. Gothard, *Encyclopedia of Physical Science*, ser. Facts on File Science Library. Facts On File, 2010. [Online]. Available: https://books.google.co.uk/books?id=avyQ64LIJa0C

[49] R. H. Clarke, "A statistical theory of mobile-radio reception," *The Bell System Technical Journal*, vol. 47, no. 6, pp. 957–1000, July 1968.

[50] M. J. Gans, "A power-spectral theory of propagation in the mobile-radio environment," *IEEE Transactions on Vehicular Technology*, vol. 21, no. 1, pp. 27–38, Feb 1972.

[51] *Wireless Communications: Principles and Practice*, ser. Prentice Hall communications engineering and emerging technologies series. Dorling Kindersley, 2009. [Online]. Available: https://books.google.co.uk/books?id=11qEWkNFFwQC

[52] A. Goldsmith. WIRELESS COMMUNICATIONS. [Online]. Available: http://web.cs.ucdavis.edu/~liu/289I/Material/book-goldsmith.pdf

[53] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Wiley-Interscience, 2005.

[54] S. J. Orfanidis. (2016, June) Electromagnetic Waves and Antennas. [Online]. Available: http://www.ece.rutgers.edu/~orfanidi/ewa/#pod

[55] K.-C. Huang and D. J. Edwards, *Millimeter Wave Antennas*, 1st ed. Wiley, 2008.

[56] D. M. Pozar and D. H. Schaubert, *Mircostrip Antennas*, 1st ed. Wiley-IEEE Press, 1995.

[57] S. Rajagopal, Shadi Abu Surra, Zhouyue Pi, and Farooq Khan, "Antenna Array Design for Multi-Gbps mmWave Mobile Broadband Communication," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011, pp. 1–6.

[58] Samsung. (2015, Aug.) 5G Vision. [Online]. Available: http://www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-2.pdf

[59] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[60] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges," in *Proceedings of the IEEE*, vol. 103, no. 3, 2014, pp. 366–385.

[61] Agilent Technologies. Fundamentals of RF and Microwave Noise Figure Measurements. [Online]. Available: http://cp.literature.agilent.com/litweb/pdf/5952-8255E.pdf

[62] B. Le *et al.*, "Analog-to-digital converters," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 69–77, 2005.

[63] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave (mmwave) communications for 5g: Opportunities and challenges," *CoRR*, vol. abs/1502.07228, 2015. [Online]. Available: http://arxiv.org/abs/1502.07228

[64] D. Tse and P. Vishwanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge University Press, 2005.

[65] P. M. C. Nathan, *Wireless Communications*, 1st ed. Prentice Hall, 2010.

[66] Z. Li, S. Han, and A. F. Molisch, "Hybrid Beamforming Design for Millimeter-Wave Multi-User Massive MIMO Downlink," in *IEEE International Conference on Communications (ICC),*, May 2016.

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

[68] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *2013 Information Theory and Applications Workshop (ITA)*, Feb 2013, pp. 1–5.

[69] G. Kwon, Y. Shim, H. Park, and H. M. Kwon, "Design of millimeter wave hybrid beamforming systems," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sep. 2014, pp. 1–5.

[70] D. Zhu, B. Li, and P. Liang, "A novel hybrid beamforming algorithm with unified analog beamforming by subspace construction based on partial csi for massive mimo-ofdm systems," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 594–607, Feb 2017.

[71] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct 2014.

[72] E. ZÃűchmann, S. Schwarz, and M. Rupp, "Comparing antenna selection and hybrid precoding for millimeter wave wireless communications," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016, pp. 1–5.

[73] H. Chiang, T. Kadur, W. Rave, and G. Fettweis, "Low-complexity spatial channel estimation and hybrid beamforming for millimeter wave links," in *2016 IEEE 27th*

Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Sep. 2016, pp. 1–7.

[74] K. Venugopal, A. Alkhateeb, N. GonzÃąlez Prelcic, and R. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

[75] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.

[76] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive mimo with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, June 2016.

[77] F. Sohrabi and W. Yu, "Hybrid Digital and Analog Beamforming Design for Large-Scale MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501 – 513, Jan 2016.

[78] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.

[79] R. Mai, D. H. N. Nguyen, and T. Le-Ngoc, "Mmse hybrid precoder design for millimeter-wave massive mimo systems," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.

[80] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical methods for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 927–949, Oct. 1993. [Online]. Available: http://dx.doi.org/10.1137/0614062

[81] J. Zhang, A. Wiesel, and M. Haardt, "Low rank approximation based hybrid precoding schemes for multi-carrier single-user massive mimo systems," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3281–3285.

[82] M. T. Vu and A. Paulraj, "Mimo wireless linear precoding," *IEEE Signal Processing Magazine*, vol. 24, pp. 86–105, 2007.

[83] A. Alkhateeb and R. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *CoRR*, vol. abs/1510.00609, 2015. [Online]. Available: http://arxiv.org/abs/1510.00609

[84] L. Dai, X. Gao, J. Quan, S. Han, and C. I, "Near-optimal hybrid analog and digital precoding for downlink mmwave massive mimo systems," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1334–1339.

[85] C. Rusu, R. MÃÍndez-Rial, N. GonzÃąlez-Prelcic, and R. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8380–8393, Dec 2016.

[86] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar 1966. [Online]. Available: https://doi.org/10.1007/BF02289451

[87] J. Cai, B. Rong, and S. Sun, "A low complexity hybrid precoding scheme for massive mimo system," in *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*, Sep. 2016, pp. 638–641.

[88] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *IEEE International Conference on Communications, (ICC)*, June 2015.

[89] A. Alkhateeb, G. Leus, and R. Heath, "Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 3337 – 3349, 2015.

[90] M. E. Eltayeb, A. Alkhateeb, R. Heath, and T. Y. Al-Naffouri, "Opportunistic beam training with hybrid analog/digital codebooks for mmwave systems," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015.

[91] A. Liu and V. K. N. Lau, "Impact of csi knowledge on the codebook-based hybrid beamforming in massive mimo," *IEEE Transactions on Signal Processing*, vol. 64, no. 24, pp. 6545–6556, Dec 2016.

[92] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analogâĂŞdigital precoding revisited under realistic rf modeling," *IEEE Wireless Communications Letters*, vol. 5, no. 5, pp. 528–531, Oct 2016.

[93] V. Venkateswaran, F. Pivit, and L. Guan, "Hybrid RF and digital beamformer for cellular networks: Algorithms, microwave architectures and measurements," *CoRR*, vol. abs/1510.02822, 2015. [Online]. Available: http://arxiv.org/abs/1510.02822

[94] R. Rajashekar and L. Hanzo, "Hybrid Beamforming in mm-Wave MIMO Systems Having a Finite Input Alphabet," *IEEE Transactions on Communications*, vol. 64, no. 8, pp. 3337 – 3349, 2016.

[95] S. Gimenez, S. Roger, D. MartÃŋn-SacristÃąn, J. F. Monserrat, P. Baracca, V. Braun, and H. Halbauer, "Performance of hybrid beamforming for mmw multi-antenna systems in dense urban scenarios," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2016, pp. 1–6.

[96] A. Alkhateeb and R. Heath, "Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1801 – 1818, April 2016.

[97] B. Clerckx, Y. Zhou, and S. Kim, "Practical codebook design for limited feedback spatial multiplexing," in *IEEE International Conference on Communications*, 2008, pp. 3982–3987.

[98] K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A. Mourad, and L. Hanzo, "Millimeter wave hybrid beamforming with DFT-MUB aided precoder codebook design," in *Proc. VTC (Fall)*, Sept. 2017, pp. 1–5.

[99] J. S. Lemos, F. RosÃąrio, F. A. Monteiro, J. Xavier, and A. Rodrigues, "Massive mimo full-duplex relaying with optimal power allocation for independent multi-pairs," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 306–310.

[100] H. Ju, E. Oh, and D. Hong, "Improving efficiency of resource usage in two-hop full duplex relay systems based on resource sharing and interference cancellation," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 3933–3938, August 2009.

[101] D. Nguyen, L. Tran, P. Pirinen, and M. Latva-aho, "Precoding for full duplex multiuser mimo systems: Spectral and energy efficiency maximization," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4038–4050, Aug 2013.

[102] ——, "On the spectral efficiency of full-duplex small cell wireless systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4896–4910, Sep. 2014.

[103] M. Mohammadi, B. K. Chalise, H. A. Suraweera, C. Zhong, G. Zheng, and I. Krikidis, "Throughput analysis and optimization of wireless-powered multiple antenna full-duplex relay systems," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1769–1785, April 2016.

[104] A. C. Cirik, M. J. Rahman, and L. Lampe, "Robust fairness transceiver design for a full-duplex mimo multi-cell system," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1027–1041, March 2018.

[105] H. A. Suraweera, I. Krikidis, and C. Yuen, "Antenna selection in the full-duplex multi-antenna relay channel," in *2013 IEEE International Conference on Communications (ICC)*, June 2013, pp. 4823–4828.

[106] T. Riihonen, S. Werner, and R. Wichman, "Hybrid full-duplex/half-duplex relaying with transmit power adaptation," *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 3074–3085, Sep. 2011.

[107] ——, "Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov 2011, pp. 1408–1412.

[108] Q. Meng, W. Feng, G. Zheng, S. Chatzinotas, and B. O. Ten, "Fixed full duplex relaying for wireless broadband communication," in *2012 International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2012, pp. 1–5.

[109] A. Sahai, S. Diggavi, and A. Sabharwal, "On degrees-of-freedom of full-duplex uplink/downlink channel," in *2013 IEEE Information Theory Workshop (ITW)*, Sep. 2013, pp. 1–5.

[110] ——, "On uplink/downlink full-duplex networks," in *2013 Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 14–18.

[111] T. Snow, C. Fulton, and W. J. Chappell, "Transmit—receive duplexing using digital beamforming system to cancel self-interference," *IEEE Trans. Micro. Theory Techn.*, vol. 59, no. 12, pp. 3494–3503, Dec 2011.

[112] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[113] D. Freedman, *Statistical Models: Theory and Practice*. Cambridge University Press, 2009. [Online]. Available: https://books.google.co.uk/books?id=k0hJ9HpAhnkC

[114] W. S. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943. [Online]. Available: http://www.cse.chalmers.se/~coquand/AUTOMATA/mcp.pdf

[115] A. Turing, "Computing machinery and intelligence," *Mind*, vol. 5, 1950. [Online]. Available: https://www.csee.umbc.edu/courses/471/papers/turing.pdf

[116] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal*, vol. 3, no. 3, 1959. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf

[117] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain1," *Psychological Review*, vol. 65, no. 6, 1958. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf

[118] B. Widrow and M. E. Hoff, "Neurocomputing: Foundations of research," J. A. Anderson and E. Rosenfeld, Eds. Cambridge, MA, USA: MIT Press, 1988, ch. Adaptive Switching Circuits, pp. 123–134. [Online]. Available: http://dl.acm.org/citation.cfm?id=65669.104390

[119] J. V. Neumann, "The first draft report on the edvac," *IEEE Annals of the History of Computing*, vol. 15, no. 4, 1993.

[120] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, 1982. [Online]. Available: https://www.pnas.org/content/79/8/2554

[121] G. E. H. David E. Rumelhart and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 79, no. 8, pp. 533–536, 1986.

[122] IBM, "Deep blue„" 1997. [Online]. Available: https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/

[123] Google, "Googlebrain," 2012. [Online]. Available: https://googleblog.blogspot.com/2012/06/using-large-scale-brain-simulations-for.html

[124] I. S. Alex Krizhevsky and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 2Nd International Conference on Neural Information Processing Systems*, ser. NIPS'12, 2012.

[125] DeepMind. [Online]. Available: https://deepmind.com/research/alphago/

[126] D. A. Heger, "An introduction to artificial neural networks methods, abstraction, and usage," 2015.

[127] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *CoRR*, vol. abs/1710.02913, 2017. [Online]. Available: http://arxiv.org/abs/1710.02913

[128] S. Haykin, *Neural Networks: A Comprehensive Foundation*, ser. International edition. Prentice Hall, 1999. [Online]. Available: https://books.google.co.uk/books?id=M5abQgAACAAJ

[129] R. Livni, S. Shalev-Shwartz, and O. Shamir, "On the computational efficiency of training neural networks," *CoRR*, vol. abs/1410.1141, 2014. [Online]. Available: http://arxiv.org/abs/1410.1141

[130] A. L. Maas, "Rectifier nonlinearities improve neural network acoustic models," 2013.

[131] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[132] Y. Chauvin and D. E. Rumelhart, Eds., *Backpropagation: Theory, Architectures, and Applications*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1995.

[133] S. Ruder, "An overview of gradient descent optimization algorithms," *ArXiv*, vol. abs/1609.04747, 2016.

[134] R. Tibshirani, "Regression shrinkage and selection via the lasso," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 58, pp. 267–288, 1994.

[135] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *CoRR*, vol. abs/1710.02913, 2017. [Online]. Available: http://arxiv.org/abs/1710.02913

[136] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *Trans. Evol. Comp*, vol. 1, no. 1, pp. 67–82, Apr. 1997. [Online]. Available: https://doi.org/10.1109/4235.585893

[137] J. SjÃűberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, no. 12, pp. 1691 – 1724, 1995, trends in System Identification. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005109895001208

[138] R. J. Williams and D. Zipser, "Experimental analysis of the real-time recurrent learning algorithm," *Connection Science*, vol. 1, no. 1, pp. 87–111, 1989. [Online]. Available: https://doi.org/10.1080/09540098908915631

[139] G. E. Hinton, S. Osindero, and Y. whye Teh, "A fast learning algorithm for deep belief nets," 2006.

[140] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[141] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[142] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, Dec 1989. [Online]. Available: https://doi.org/10.1007/BF02551274

[143] R. Daniels and R. Heath, "Online adaptive modulation and coding with support vector machines," in *2010 European Wireless Conference (EW)*, April 2010, pp. 718–724.

[144] S. Dorner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Topics Sig. Proc.*, vol. 12, no. 1, pp. 132–143, Feb 2018.

[145] X. Jin and H.-N. Kim, "Deep learning detection networks in MIMO decode-forward relay channels," *arXiv e-prints*, p. arXiv:1807.09571, Jul. 2018.

[146] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. BeâĂŹery, "Deep learning methods for improved decoding of linear codes," vol. 12, no. 1, pp. 119–131, Feb 2018.

[147] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," vol. 12, no. 1, pp. 144–159, Feb 2018.

[148] X. Yan, F. Long, J. Wang, N. Fu, W. Ou, and B. Liu, "Signal detection of MIMO-OFDM system based on auto encoder and extreme learning machine," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 1602–1606.

[149] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[150] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive mimo system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[151] A. Balatsoukas-Stimming, "Non-Linear Digital Self-Interference Cancellation for In-Band Full-Duplex Radios Using Neural Networks," *arXiv e-prints*, p. arXiv:1711.00379, Nov 2017.

[152] H. Guo, N. Zhang, S. AlQarni, and S. Wu, "DSIC: deep learning based self-interference cancellation for in-band full duplex wireless," *CoRR*, vol. abs/1811.01498, 2018. [Online]. Available: http://arxiv.org/abs/1811.01498

[153] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, Feb 2018.

[154] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," *Trans. Sig. Proc.*, vol. 66, no. 21, pp. 5663–5678, Nov. 2018. [Online]. Available: https://doi.org/10.1109/TSP.2018.2868322

[155] L. Hanzo, H. Haas, S. Imre, D. O'brien, M. Rupp, and L. Gyongyosi, "Wireless Myths, Realities, and Futures: From 3G/4G to Optical and Quantum Wireless," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1853–1888, May 2012.

[156] Cisco. (2016, February) Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020. Cisco. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[157] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based rf-baseband codesign for mimo antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, 2005.

[158] A. Alkhateeb, J. Mo, N. GonzÃąlez-Prelcic, and R. Heath, "Mimo precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122 – 131, 2014.

[159] S. Han *et al.*, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5g," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 186–194, 2015.

[160] Y. S. Cho *et al.*, *MIMO-OFDM Wireless Communications with Matlab*, 1st ed. Wiley, 2010.

[161] Jaspreet *et al.*, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, Jan 2015.

[162] X. Gao *et al.*, "Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 1499 – 1513, April 2016.

[163] B. Mondal, T. A. Thomas, and M. Harrison, "Rank-independent codebook design from a quaternary alphabet," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, 2007, pp. 297 – 301.

[164] T. Inoue and R. Heath, "Kerdock codes for limited feedback precoded mimo systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3711 – 3716, 2009.

[165] D. J. Love, R. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735 – 2747, 2003.

[166] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *Published in European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November/December 1999.

[167] M. El-Hajjar and L. Hanzo, "Multifunctional MIMO systems: A combined diversity and multiplexing design perspective," *IEEE Wireless Commun*, vol. 17, no. 2, pp. 73–79, April 2010.

[168] B. Friedlander and S. Scherzer, "Beamforming vs transmit diversity in the downlink of a cellular communications system," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1023–1034, July 2004.

[169] D. Yang, L.-L. Yang, and L. Hanzo, "DFT-based beamforming weight-vector codebook design for spatially correlated channels in the unitary precoding aided multiuser downlink," in *IEEE International Conference on Communications (ICC)*, May 2010.

[170] T. Instruments, "âĂIJCodebook design for E-UTRA MIMO Precoding," *R1-062016*, vol. 3GPP TSG RAN WG1, 2006.

[171] K. Pan, D. Grieco, and R. Olesen, "Constant modulus mimo precoding for constraining transmit antenna power for differential feedback," Dec. 24 2008, wO Patent App. PCT/US2008/067,372. [Online]. Available: http://www.google.co.uk/patents/WO2008157620A2?cl=en

[172] J. Schwinger, "Unitary operator bases," in *Proc Natl Acad Sci U S A.*, April 1960, pp. 570–579.

[173] Bandyopadhyay, Boykin, Roychowdhury, and Vatan, "A new proof for the existence of unbiased bases," *Algorithmica*, vol. 34, no. 4, pp. 512–528, 2002.

[174] G. A. Kabatiansky and V. I. Levenshtein, "On bounds for packings on a sphere and in space," *Problems of Information Transmission*, vol. 14, no. 1, pp. 3–25, 2002.

[175] P. Delsarte, J. Goethals, and J. Seidel, "Bounds for systems of lines, and jacobi polynomials," *Philips Research Reports*, vol. 30, no. 1, pp. 91–105, 1975.

[176] A. Klappenecker and M. Rotteler, "Constructions of Mutually Unbiased Bases," *CoRR*, vol. quant-ph/0309120v1, 2003. [Online]. Available: http://arxiv.org/abs/quant-ph/0309120v1

[177] G. Zauner. (1999, June) Quantum Designs: Foundations of a non-commutative Design Theory. [Online]. Available: http://gerhardzauner.at/documents/gz-quantumdesigns.pdf

[178] C. K. Vala, M. French, A. Acharyya, and B. M. Al-Hashimi, "Low-complexity architecture for cyber-physical systems model identification," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 8, pp. 1416–1420, Aug 2019.

[179] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Wiley-Interscience, 2005.

[180] B. Clerckx, D. Vanhoenacker-Janvier, C. Oestges, and L. Vandendorpe, "Mutual coupling effects on the channel capacity and the space-time processing of MIMO communication systems," in *Proc. ICC*, vol. 4, May 2003, pp. 2638–2642.

[181] I. Guvenc, C.-C. Chong, and F. Watanabe, "NLOS Identification and Mitigation for UWB Localization Systems," in *IEEE Wireless Communications and Networking Conference*, March 2007.

[182] A. Checko *et al.*, "Cloud RAN for mobile networks: A technology overview," *IEEE Commun. Surveys Tutorials.*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[183] R. G. Stephen and R. Zhang, "Joint millimeter-wave fronthaul and OFDMA resource allocation in ultra-dense CRAN," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1411–1423, March 2017.

[184] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, April 2015.

[185] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks — A technology overview," *IEEE Commun. Surveys Tutorials.*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[186] J. Li, M. Peng, A. Cheng, Y. Yu, and C. Wang, "Resource allocation optimization for delay-sensitive traffic in fronthaul constrained cloud radio access networks," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2017.

[187] V. N. Ha, L. B. Le, and N. D. Oao, "Cooperative transmission in cloud RAN considering fronthaul capacity and cloud processing constraints," in *Proc. WCNC*, April 2014, pp. 1862–1867.

[188] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint decompression and decoding for cloud radio access networks," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 503–506, May 2013.

[189] J. D. Bakker and R. Prasad, "Handover in a virtual cellular network," in *Proc. VTC*, vol. 1, 1999, pp. 544–548 vol.1.

[190] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Micro. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, July 2016.

[191] *Near-field self-interference cancellation and quality of service multicast beamforming in full-duplex*, vol. 9902, 2016.

[192] Y.-P. Liao, H. Han, and Q. Guo, "Design of robust near-field multi-beam forming based on improved LCMV algorithm," *J. Inf. Hiding Multimed. Signal Process*, vol. 6, pp. 783–791, 2015.

[193] E. Antonio-RodrÃŋguez, R. LÃşpez-Valcarce, T. Riihonen, S. Werner, and R. Wichman, "Sinr optimization in wideband full-duplex mimo relays under limited dynamic range," in *n Proc. SAM Signal Process. Workshop*, June 2014, pp. 177–180.

[194] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Sig. Proc.*, vol. 10, no. 3, pp. 528–542, April 2016.

[195] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, April 2015.

[196] M. El-Hajjar, S. Zummo, and L. Hanzo, "Near-instantaneously adaptive cooperative uplink schemes based on space-time block codes and V-Blast," in *Proc. VTC*, April 2007, pp. 2200–2204.

[197] L. Hanzo, M. Münster, B. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-user Communications, WLANs and Broadcasting*. John Wiley & Sons, May 2003. [Online]. Available: http://eprints.soton.ac.uk/258228/

[198] C.-S. Lee and W.-H. Chung, "Max-Min Hybrid Precoding in Millimeter Wave Cooperative MISO Systems," in *IEEE International Conference on Communications (ICC),*, May 2016.

[199] L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalised and Space-Time Coded TDMA, CDMA, MC-CDMA and OFDM Systems*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[200] L. Chen, Y. Yang, X. Chen, and G. Wei, "Energy-efficient link adaptation on rayleigh fading channel for OSTBC MIMO system with imperfect CSIT," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1577–1585, May 2013.

[201] E. Eraslan and B. Daneshrad, "Low-complexity link adaptation for energy efficiency maximization in MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5102–5114, Aug 2017.

[202] T. L. Jensen, S. Kant, J. Wehinger, and B. H. Fleury, "Fast link adaptation for MIMO OFDM," *IEEE Trans. Veh. Technol.*, vol. 59, no. 8, pp. 3766–3778, Oct 2010.

[203] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. New York, NY, USA: Wiley-Interscience, 2000.

[204] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *in Proc. URSI Int. Symp. Signals, Syst., Electron.*, Oct 1998, pp. 295–300.

[205] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct 1998.

[206] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[207] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*, ser. Wiley - IEEE. Wiley, 2009. [Online]. Available: http://books.google.co.uk/books?id=590-JItOJREC

[208] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-Capacity Wireless Transceivers and Cooperative Communications in the MIMO Era: Evolution of Standards, Waveform Design, and Future Perspectives," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1343–1385, Aug 2011.

[209] M. El-Hajjar, O. Alamri, J. Wang, S. Zummo, and L. Hanzo, "Layered steered space-time codes using multi-dimensional sphere packing modulation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3335–3340, July 2009.

[210] S. Sugiura, S. Chen, and L. Hanzo, "Space-time shift keying: A unified mimo architecture," in *Proc. Globecom*, Dec 2010, pp. 1–5.

[211] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Layered multi-group steered space-time shift-keying for millimeter-wave communications," *IEEE Access*, vol. 4, pp. 3708–3718, 2016.

[212] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, July 2008.

[213] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Multi-set space-time shift-keying with reduced detection complexity," *IEEE Access*, vol. 4, pp. 4234–4246, 2016.

[214] I. A. Hemadeh, M. El-Hajjar, S. Won, and L. Hanzo, "Multiuser steered multiset space-time shift keying for millimeter-wave communications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5491–5495, June 2017.

[215] Y. Ding, V. Fusco, A. Shitvov, Y. Xiao, and H. Li, "Beam index modulation wireless communication with analog beamforming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6340–6354, July 2018.

[216] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.

[217] S. Ougiaroglou and G. Evangelidis, "RHC: a non-parametric cluster-based data reduction for efficient k-NN classification," *Pattern Analysis and Applications*, vol. 19, no. 1, pp. 93–109, 2016.

[218] X. Wang, "A fast exact k-nearest neighbors algorithm for high dimensional search using k-means clustering and triangle inequality," in *Int. Jr. Conf. Neural Netw.*, April 2012, pp. 2351—2358.

[219] Rokach and Lior, *A Survey of Clustering Algorithms: Data Mining and Knowledge Discovery Handbook (2nd Edition)*. Springer US, 2010.

[220] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multidimensional signal sets," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 528–536, March 2006.

[221] A. Vakili, M. Sharif, and B. Hassibi, "The effect of channel estimation error on the throughput of broadcast channels," in *Proc. ICASSP*, vol. 4, May 2006, pp. IV–IV.

[222] W. C. Jakes and D. C. Cox, Eds., *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.

[223] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. Wiley Publishing, 2010.

[224] M. Boban, T. T. V. Vinhoza, M. Ferreira, J. Barros, and O. K. Tonguz, "Impact of vehicles as obstacles in vehicular ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 15–28, January 2011.

[225] V. Va, T. Shimizu, G. Bansal, and R. Heath, "Beam design for beam switching based millimeter wave vehicle-to-infrastructure communications," in *Proc. ICC*, May 2016, pp. 1–6.

[226] G. H. Mohimani, F. Ashtiani, A. Javanmard, and M. Hamdi, "Mobility modeling, spatial traffic distribution, and probability of connectivity for sparse and dense vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1998–2007, May 2009.

[227] M. J. Farooq, H. ElSawy, and M. Alouini, "A stochastic geometry model for multi-hop highway vehicular communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2276–2291, March 2016.

[228] K. T. Truong and R. Heath, "Effects of channel aging in massive MIMO systems," *J. of Commun. and Networks*, vol. 15, no. 4, pp. 338–351, Aug 2013.

[229] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *CoRR*, vol. abs/1605.07661, 2016. [Online]. Available: http://arxiv.org/abs/1605.07661

[230] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. Boston: McGraw Hill, 2002. [Online]. Available: http://www.worldcat.org/search?qt=worldcat_org_all&q=0071226613

[231] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 12, pp. 1983–1993, Dec 2012.

[232] S. h. Fang, T. n. Lin, and K. c. Lee, "A novel algorithm for multipath fingerprinting in indoor WLAN environments," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3579–3588, September 2008.

[233] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based indoor localization," vol. 24, no. 7, pp. 1300–1309, July 2013.

[234] Lte; physical layer — measurements. [Online]. Available: https://www.etsi.org/deliver/etsi

[235] ITU-R. Propagation by diffraction. [Online]. Available: https://www.itu.int/rec/R-REC-P.526-11-200910-S/en

[236] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[237] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. New York, NY, USA: Wiley-Interscience, 2006.

[238] Automotive technical data and specifications. [Online]. Available: http://www.carfolio.com/

[239] X. Gao, L. Dai, S. Zhou, A. M. Sayeed, and L. Hanzo, "Wideband beamspace channel estimation for millimeter-wave mimo systems relying on lens antenna arrays," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4809–4824, Sep. 2019.

[240] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave mimo systems," *IEEE Transactions on Communications*, pp. 1–1, 2019.

[241] H. Soleimani, R. Parada, S. Tomasin, and M. Zorzi, "Fast initial access for mmwave 5g systems with hybrid beamforming using online statistics learning," *IEEE Communications Magazine*, vol. 57, no. 9, pp. 132–137, Sep. 2019.

[242] C. N. Barati, S. A. Hosseini, M. Mezzavilla, P. Amiri-Eliasi, S. Rangan, T. Korakis, S. S. Panwar, and M. Zorzi, "Directional initial access for millimeter wave cellular systems," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, Nov 2015, pp. 307–311.

Author Index

Capone, A. 4, 74

Caramanis, C. M. 4, 73, 75, 157

Chalise, B. K. 55, 57

Challita, Ursula 60, 70, 72, 73

Chappell, W. J. 56, 57, 123

Chatzinotas, S. 56, 57

Checko, A. 112, 114

Chen, D. 207

Chen, H. 4, 74, 75

Chen, L. 157

Chen, Mingzhe 60, 70, 72, 73

Chen, S. 157

Chen, X. 157

Cheng, A. 114

Chiang, H. 43, 45

Cho, Yong Soo 77, 193

Choi, B.J. 141

Choi, Jinho 4, 53, 56, 57, 122, 123, 127, 128

Choi, Junil 44, 45, 77, 83

Chong, Chia-Chin 107, 111

Christiansen, H. L. 114

Chung, Wei-Ho 148

Cirik, A. C. 55, 57

Cisco 76

Clarke, R. H. 19

Clerckx, B. 100, 103

Clerckx, Bruno 45, 77, 90

Courville, Aaron 57, 58, 66, 68, 69, 71, 72, 179

Cover, Thomas M. 32, 214

Cybenko, G. 73

Dai, L. 43–45

Daneshrad, B. 157

Daniels, R. 4, 73–75

Daniels, R. C. 4, 73, 75, 157

Daniels, R.C. 15, 17, 19

David E. Rumelhart, Geoffrey E. Hinton 60

Day, B. P. 4, 56, 123

Debbah, Mérouane 60, 70, 72, 73

DeepMind 60

Delsarte, Ph. 91

Delyon, Bernard 72

Deng, S. 18

Dick, C. 3, 4, 53, 122, 123

Dick, Chris 3, 52, 122

Diggavi, S. 56, 57

Ding, Y. 158

Diskin, Tzvi 5, 74, 248

Dittmann, L. 114

Dong, C. 4, 55, 123

Dorner, S. 74, 75

Duarte, M. 4, 53, 123

Duarte, Melissa 3, 52, 53, 122

Duda, Richard O. 157, 164, 166, 228

Edwards, David J. 30–33

El Ayach, O. 43, 45

El-Hajjar, M. 157, 246

El-Hajjar, Mohammed 3, 40, 49, 78, 100, 101, 157

ElSawy, H. 204

Eltayeb, Mohammed E. 45, 77

Eraslan, E. 157

Erkip, E. 32

Evangelidis, Georgios 167

Everett, E. 4, 56, 123

Everett, Evan 3, 52, 122

Farooq Khan 32

Farooq, M. J. 204

Farsad, Nariman 75

Feng, C. 207

Feng, W. 56, 57

Feng, Z. 4, 74, 75

Ferreira, M. 204, 208, 219

Fettweis, G. 43, 45

Filippini, I. 4, 74

Fleury, B. H. 157

Foschini, G. J. 157

Freedman, D. 58

French, M. 95

Friedlander, B. 83

Friis, H. T. 17

Fu, N. 74, 75, 248

Fulton, C. 56, 57, 123

Fusco, V. 158

Gans, M. J. 19

Gao, L. 4, 5, 74, 75, 123

Gao, X. 44, 45

Gao, Xinyu 77

Gao, Z. 43, 45

Garcia-Rodriguez, A. 45

Ghauch, H. 134, 135, 162

Ghorashi, S. A. 4, 74, 75

Gimenez, S. 45

Glorennec, Pierre-Yves 72

Goethals, J.M. 91

Golden, G. D. 157

Goldsmith, Andrea 19, 75

GonzÃąlez-Prelcic, N. 44, 45

GonzÃąlez-Prelcic, Nuria 77

Goodfellow, Ian 57, 58, 66, 68, 69, 71, 72, 179

Google 60

Gothard, L.Q. 18

Grieco, D.M. 90

Gross, W. J. 74, 75, 248

Guan, Lei 45

Gui, G. 74, 75

Guo, D. 4, 53, 123

Guo, Hanqing 75

Guo, Qiang 128

Guvenc, Ismail 107, 111

Gyongyosi, L. 76

h. Fang, S. 207

Ha, V. N. 114

Haardt, M. 44, 45

Haas, H. 76, 157

Hajihoseini, A. 4, 74, 75

Halbauer, H. 45

Hamdi, M. 204

Han, Hua 128

Han, S. 44, 45

Han, Shengqian 40

Han, Shuangfeng 77

Hanzo, L. 157, 246

Harrison, Mark 78, 91–93

Hart, Peter E. 157, 164, 166, 228

Haykin, S.S. 63, 64, 180

Heath, R.W. 4, 5, 15, 17, 19, 43–45, 73–75, 77, 78, 83, 91, 93–97, 157, 204, 205, 228

Heger, Dominique A. 61

Hemadeh, I. 2, 14, 15, 17, 47, 49, 50, 209, 245

Hemadeh, I. A. 157

Hinton, Geoffrey E. 60, 72

Hjalmarsson, HÃěkan 72

Hochreiter, Sepp 72

Hoff, Marcian E. 60

Hong, D. 55

Hopfield, J J 60

Hosseini, S. A. 249

Hoydis, J. 74, 75

Hu, C. 43, 45

Huang, H. 74, 75

Huang, Kao-Cheng 30–33

I, C. 44, 45

IBM 60

Imre, S. 76

Ingram, M. A. 4, 128

Inoue, Takao 78, 91, 93, 95

Instruments, Texas 90

ITU-R 208

Ivanescu, T. 246

Jafarkhani, H. 157

Jaspreet 77

Javanmard, A. 204

Jensen, T. L. 157

Jiang, Jeng-Shiann 4, 128

Jin, Depeng 39, 76

Jin, Xianglan 74, 75

Ju, H. 55

Juang, B. 75

Juditsky, Anatoli 72

Kabatiansky, G. A. 91

Kadur, T. 43, 45

Kang, Chung G. 193

Kant, S. 157

Kardaras, G. 114

Katti, Sachin 3, 52, 53, 122

Keller, T. 141

Khatab, Z. E. 4, 74, 75

Kim, Hyoung-Nam 74, 75

Kim, Jaekwon 193

Kim, Sungjin 45, 77, 90

Kim, T. 134, 135, 162

Klappenecker, Andreas 91, 93

Korakis, T. 249

Krikidis, I. 55, 57

Kung, Sun-Yuan 76

Kuo, P. . 246

Kuo, Ping-Heng 3, 40, 49, 78, 100, 101

Kwon, G. 43, 45

Kwon, H. M. 43, 45

Lampe, L. 55, 57

Latva-aho, M. 55, 57

Lau, V. 114, 157, 161

Lau, V. K. N. 45

Le, B. 38, 76

Le, L. B. 114

Le-Ngoc, T. 44, 45

Lee, Chang-Shen 148

Lemos, J. S. 55, 57

Letaief, Khaled B. 44, 45

Leus, G. 43, 45

Leus, Geert 44, 45, 94

Levenshtein, V. I. 91

Li, B. 43, 45

Li, G. Y. 75

Li, H. 158

Li, J. 114

Li, L. 4, 55, 123

Li, W. 4, 74, 75

Li, Yong 39, 76

Li, Zheda 40

Liang, F. 74, 75

Liang, P. 43, 45

Liao, Yan-Ping 128

Lin, C. T. 4, 55, 57, 123

Lior 167

Liu, A. 45

Liu, B. 74, 75, 248

Liu, X. 4, 74, 75

Liu, Xiao 4, 53, 56, 57, 122, 123, 127, 128

Liu, Y. 4, 56

Livni, Roi 64, 73

Ljung, Lennart 72

Lo, E. S. 4, 55, 123

Long, F. 74, 75, 248

Long, K. 3, 4, 53, 54, 123

Love, D. J. 78

Love, David J. 44, 45, 77, 83

Lugosch, L. 74, 75, 248

Luo, X. 207

LÃşpez-Valcarce, R. 132

Maas, Andrew L. 64, 73

MacCartney, G. R. 17, 245

Maccartney, G.R. 2, 15, 76

Macready, W. G. 70

Mai, R. 44, 45

Mao, S. 4, 5, 74, 75, 123

Marciano, E. 74, 75, 248

Margetts, A. R. 4, 56, 123

MartÃŋn-SacristÃąn, D. 45

Masouros, C. 45

Mcculloch, Warren S. 60

McMilin, Emily 3, 52, 53, 122

Mehrmann, Volker 44

Mellios, E. 17, 245

Meng, Q. 56, 57

Mesleh, R. Y. 157

Mezzavilla, M. 249

Mo, Jianhua 77

Mohammadi, M. 55, 57

Mohimani, G. H. 204

Mohri, Mehryar 4, 73, 158

Molisch, A. F. 76

Molisch, Andreas F. 40

Mondal, Bishwarup 78, 91–93

Monserrat, J. F. 45

Monteiro, F. A. 55, 57

Mourad, A. 246

Mourad, Alain 3, 40, 49, 78, 100, 101, 106, 109, 114, 157

Münster, M. 141

Murdock, J.N. 15, 17, 19

MÃĺndez-Rial, R. 44, 45

n. Lin, T. 207

Nachmani, E. 74, 75, 248

Nathan, P. Muthu Chindambara 40, 90

Neumann, John Von 60

Ng, D. W. K. 4, 55, 123

Ng, Soon Xin 176

Nguyen, D. 55, 57

Nguyen, D. H. N. 44, 45

Ni, L. M. 207

Niu, Yong 39, 76

Oao, N. D. 114

O'brien, D. 76

Oestges, C. 100, 103

Oh, E. 55

Olesen, R.L. 90

Orfanidis, Sophocles J. 20, 21, 24, 27, 30

Osindero, Simon 72

Ou, W. 74, 75, 248

Ougiaroglou, Stefanos 167

Pan, K.J.L. 90

Pandey, S. 4, 5, 74, 123

Panwar, S. S. 249

Papazafeiropoulos, Anastasios K. 205

Parada, R. 249

Park, H. 43, 45

Park, M. 5, 74, 75

Park, S. H. 114

Patel, G. 3, 53, 122

Paulraj, A. 44

Peng, M. 114, 157, 161

Pi, Zhouyue 43, 45, 83, 95–97

Pirinen, P. 55, 57

Pitts, Walter 60

Pivit, Florian 45

Poor, H. V. 114, 157, 161

Pozar, David M. 30

Prasad, R. 115

Puljiz, Z. 5, 74, 75

Quan, J. 44, 45

Qureshi, S. U. H. 158

Rahman, M. J. 55, 57

Rajagopal, S. 43, 45, 83, 95–97

Rajagopal, Sridhar 32

Rajashekar, Rakshith 45, 77

Rangan, S. 249

Rangarajan, S. 4, 53, 123

Rappaport, T. S. 32, 125

Rappaport, T.S. 2, 15, 17, 19, 76

Ratnarajah, T. 140

Rave, W. 43, 45

Riihonen, T. 4, 8, 55, 57, 123, 124, 132

Rodrigues, A. 55, 57

Roger, S. 45

Rokach 167

Rong, B. 44, 45

Rosen, J. 18

Rosenblatt, Frank 60

Rostamizadeh, Afshin 4, 73, 158

RosÃąrio, F. 55, 57

Rotteler, Martin 91, 93

Roychowdhury 91

Ruder, Sebastian 67

Rulikowski, P. 45

Rupp, M. 76

Rusu, C. 44, 45

Saad, Walid 60, 70, 72, 73

Sabharwal, A. 56, 57

Sabharwal, Ashutosh 3, 52, 53, 122

Sahai, A. 56, 57

Sahin, O. 114

Samimi, M. K. 125

Samimi, M.K. 2, 15, 76

Samsung 32

Samuel, Arthur L. 60

Samuel, Neev 5, 74, 248

Sari, H. 74, 75

Satyanarayana, K. 246

Schaubert, Daniel H. 30

Scherzer, S. 83

Schmidhuber, JÃijrgen 72

Schniter, P. 4, 53, 56, 123

Schober, R. 4, 55, 123

Schönemann, Peter H. 44

Schwarz, S. 43, 45

Schwinger, Julian 91

Sciancalepore, V. 4, 74

Scolari, L. 114

Seidel, J.J. 91

Shadi Abu Surra 32

Shalev-Shwartz, Shai 64, 73

Shamai, S. 114

Shamir, Ohad 64, 73

Shao, Shihai 128

Shen, C. 74, 75

Shen, Juei-Chin 44, 45

Shim, Y. 43, 45

Shimizu, T. 204

Shitvov, A. 158

Simeone, O. 114

Sinanovic, S. 157

SjÃűberg, Jonas 72

Skoglund, M. 134, 135, 162

Smith, P. J. 4, 55, 123

Snow, T. 56, 57, 123

Sohrabi, Foad 44, 45, 84

Soleimani, H. 249

Song, Jiho 44, 45, 77, 83

Song, Y. 74, 75

Stephen, R. G. 112

Stork, David G. 157, 164, 166, 228

Strohmer, T. 78

Su, Li 39, 76

Sugiura, S. 157

Sun, S. 44, 45

Suraweera, H. A. 55, 57

Sutton, Richard S. 72

t. Brink, S. 74, 75

Talwalkar, Ameet 4, 73, 158

Tan, Z. 207

Tang, Youxi 128

Tao, X. 4, 74, 75

Tarokh, V. 157

Telatar, Emre 82

Ten, B. O. 56, 57

Thomas, Joy A. 32, 214

Thomas, Timothy A. 78, 91–93

Tibshirani, Robert 68, 69

Tomasin, S. 249

Tonguz, O. K. 204, 208, 219

Tran, L. 55, 57

Tremolada, D. 4, 74

Truong, K. T. 205, 228

Tse, David 39, 42, 84, 85

Tseng, F. S. 4, 55, 57, 123

Turing, Alan 60

Va, V. 204

Vala, C. K. 95

Valaee, S. 207

Valenzuela, R. A. 157

Vandenberghe, Lieven 8, 124, 135

Vandendorpe, L. 100, 103

Vanhoenacker-Janvier, D. 100, 103

Vasilakos, A. V. 3, 4, 53, 54, 123

Vasilakos, Athanasios V. 39, 76

Vatan 91

Venkateswaran, V. 45

Venkateswaran, Vijay 45

Venugopal, K. 43, 45

Vinhoza, T. T. V. 204, 208, 219

Vishwanath, Pramod 39, 42, 84, 85

Vu, Mai Trang 44

Wang, C. 114, 157, 161

Wang, J. 74, 75, 157, 248

Wang, L. 4, 55, 123

Wang, S. 4, 56

Wang, X. 4, 5, 74, 75, 123

Wang, Xueyi 167

Wang, Z. 43, 45

Watanabe, Fujio 107, 111

Wehinger, J. 157

Wei, G. 157

Wei, Z. 4, 74, 75

Werner, S. 4, 8, 55, 57, 123, 124, 132

whye Teh, Yee 72

Wichman, R. 4, 8, 53, 55, 57, 123, 124, 132

Widrow, Bernard 60

Wiesel, A. 44, 45

Wiesel, Ami 5, 74, 248

Williams, Ronald J. 60, 72

Wolniansky, P. W. 157

Wolpert, D. H. 70

Won, S. 157

Wong, C. H. 156

Wu, F. 74, 75

Wu, Fei 128

Wu, K. 207

Wu, Nan 157

Wu, Shaoen 75

Wu, W. R. 4, 55, 57, 123

Xavier, J. 55, 57

Xia, P. 4, 54, 56, 123

Xia, Pengfei 4, 53, 56, 57, 122, 123, 127, 128

Xia, X. G. 4, 54–57, 114, 123

Xia, Xiang-Gen 4, 53, 56, 57, 122, 123, 127, 128

Xiao, J. 207

Xiao, Y. 158

Xiao, Z. 4, 54, 56, 123

Xiao, Zhenyu 4, 53, 56, 57, 122, 123, 127, 128

Xing, Y. 17, 245

Yan, X. 74, 75, 248

Yan, Y. 114

Yang, D. 88

Yang, J. 74, 75

Yang, L.-L. 88

Yang, Won Young 193

Yang, Y. 157

Ye, H. 75

Yee, M. S. 156

Yi, Y. 207

Yin, Changchuan 60, 70, 72, 73

Yu, Wei 44, 45, 84

Yu, Xianghao 44, 45

Yu, Y. 114

Yuen, C. 55, 57

Yun, S. 157

Zauner, Gerhard 91

Zhang, H. 4, 56

Zhang, J. 17, 44, 45, 245

Zhang, Jun 44, 45

Zhang, Nan 75

Zhang, P. 4, 74, 75

Zhang, Qinghua 72

Zhang, R. 112

Zhang, T. 43

Zhang, W. 4, 56

Zhang, Xinying 76

Zhang, Y. 4, 74, 75

Zhang, Z. 3, 4, 53, 54, 123

Zhao, Y. 4, 74, 75

Zheng, G. 55–57

Zhong, C. 55, 57

Zhou, YongXing 45, 77, 90

Zhouyue Pi 32

Zhu, D. 43, 45

Zipser, David 72

Zorzi, M. 249

Zummo, S. 140, 157

ZÃűchmann, E. 43, 45