

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**Compression of Channel State Information in
Multiple Input Multiple Output Mobile Systems**

Brenda Vilas Boas

DM: 15/2019

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2019

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Brenda Vilas Boas

**Compression of Channel State Information in
Multiple Input Multiple Output Mobile Systems**

DM: 15/2019

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2019

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Brenda Vilas Boas

**Compression of Channel State Information in
Multiple Input Multiple Output Mobile Systems**

Dissertation submitted to the examination committee of the graduate department of Electrical Engineering at Federal University of Pará in fulfillment of the requirements for the degree of Master of Science in Electrical Engineering with emphasis in Telecommunications.

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2019

Dados Internacionais de Catalogação - na – Publicação (CIP) Sistema de
Bibliotecas da UFPA

V697c Vilas Boas, Brenda, 1992-

Compression of Channel state information in multiple input multiple output
mobile systems / Brenda Vilas Boas.-2019.

Orientador: Aldebaro Barreto da Rocha Klautau Júnior
Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de
Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2019.

1. Processamento de sinais – Técnicas digitais. 2. Telefonia celular –
Inovações tecnológicas. 3. Modulação digital. I. Título.

CDD 23. ed. 621.3822

Elaborada por Lucicléa S. de Oliveira – CRB -2/648

**Compression of Channel State Information in
Multiple Input Multiple Output Mobile Systems**

Dissertação de mestrado submetida à avaliação da banca examinadora aprovada pelo colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Pará e julgada adequada para obtenção do Grau de Mestre em Engenharia Elétrica na área de Telecomunicações.

Aprovada em ____ / ____ / ____

Prof. Dr. Aldebaro Barreto da Rocha Klautau Junior

ORIENTADOR

Prof. Dr. Fabrício José Brito Barros

MEMBRO DA BANCA EXAMINADORA

Prof. Dr. Marcus Vinicius Lamar

MEMBRO DA BANCA EXAMINADORA

Prof^ª. Dr^ª. Maria Emília de Lima Tostes

DIRETOR DA PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

To my family for all the support.

Acknowledgments

I would like to thank my parents for the emotional, financial, logistics and all kinds of support they could provide me.

My gratitude to Aldebaro Klautau for the trust, mentorship, patience and friendship. It was quite challenging to pursue masters with the master advisor abroad, but we made it!

More than a laboratory, LASSE is a family! So, I would like to thank everybody for our time together, either studying, attending conferences, developing projects, or simple having lunch together. Specially, I would like to thank Joary, Yuichi, and Leonardo for the insightful conversations about Multiple input multiple output (MIMO) and compression methods.

Last, but not least, I would like to thank God for the strength he has given me to overcome the obstacles within my way.

Brenda Vilas Boas

May 2019

*"I was taught that the way of progress was neither swift
nor easy."*

Marie Curie

List of Acronyms

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
3GPP	Third generation partnership project
5G	Fifth generation of wireless networks
AWGN	Additive white Gaussian noise
BS	Base station
CS	Compressive sensing
CSI	Channel state information
CSIR	Channel state information at receiver
CSIT	Channel state information at transmitter
DCT	Discrete cosine transform
DoF	Degrees of freedom
DFT	Discrete Fourier transform
eMBB	Enhanced mobile broadband
FDD	Frequency division duplexing
FW	Fixed wireless
IoT	Internet of things

ITU-R	International Telecommunications Union Radiocommunication sector
KLT	Karhunen-Loève transform
LBG	Linde-Buzo-Gray
MIMO	Multiple input multiple output
MSE	Mean square error
mMTC	Massive machine-type communication
mmWaves	Millimeter waves
OMP	Orthogonal matching pursuit
RT	Ray tracing
SNR	Signal to noise ratio
SQNR	Signal to quantization noise ratio
TDD	Time division duplexing
UE	User equipment
ULA	Uniform linear array
UPA	Uniform planar array
VQ	Vector quantization
V2E	Vehicle to everything
URLLC	Ultra-reliable and low latency communication

List of Symbols

$(\cdot)^{-1}$ - Inverse matrix

$(\cdot)^H$ - Hermitian operation

α_l - Complex power of l^{th} multi-path component

Δ_ϕ - Angular spread

Δ - Step size of a scalar uniform quantizer

η - Compression ratio

λ - Wavelength

θ - Root mean square value of input signal

θ_d^2 - Distortion of discarded coefficients

θ_i^2 - Variance of the i^{th} transform coefficient

θ_q^2 - Quantization distortion

θ_l^A - Elevation angle of arrival of l^{th} multipath component

θ_l^D - Elevation angle of departure of l^{th} multipath component

μ_ϕ - Mean azimuth angle of arrival

τ - Delay

ϕ_l^A - Azimuth angle of arrival of l^{th} multipath component

ϕ_l^D - Azimuth angle of departure of l^{th} multipath component

Φ - Compressive sensing measurement matrix

Ψ_{DCT} - Orthogonal basis function of DCT

Ψ_{KLT} - Orthogonal basis function of KLT

Ψ_x - Generic orthogonal basis function

\mathbf{a} - Array steer vector

\mathbf{A} - Generic transform basis

b - Number of bits

b_{in} - Number of bits at system input

b_{out} - Number of bits after quantization

B - Range of a quantizer

BW - Bandwidth

d - Distance among antenna elements

d_λ - Normalized distance between antenna elements

$d(x_n, \hat{x}_n)$ - Distortion of n^{th} sample

\mathbf{D} - Matrix of eigenvalues

D_f - Degree of freedom

$D_{f,\text{MAX}}$ - Maximum degree of freedom

G_{TC} - Transform coding gain

h_s - \mathbf{H} Vectorized by columns and sorted coefficients

h_z - \mathbf{H} Vectorized by zigzag

\mathbf{H} - MIMO complex channel gain matrix

\mathbf{H}_{idd} - Complex Gaussian matrix

H_v - Virtual channel

H_x - H in transform domain

\mathbf{I} - Identity matrix

\mathcal{I} - Set of indices with selected coefficients

K - Amount of relevant signal components

L - Total number of multipath components

N_{rx} - Number of antennas at receiver

N_{tx} - Number of antennas at transmitter

N_{UE} - Number of UEs

Q - Quantizer

r - Radius of the scattering ring

r_{sq} - Code rate of a scalar quantizer

r_{vq} - Code rate of a vector quantizer

\mathbf{R}_{rx} - Covariance matrix at the receiver

\mathbf{R}_{tx} - Covariance matrix at the transmitter

\mathbf{R}_{UPA} - Covariance matrix of an UPA

\mathbf{R}_{xx} - Correlation matrix of any signal

\mathcal{R} - Output levels of a quantizer

s - Distance between transmitter and receiver

\mathbf{s} - Input signal at sparse domain

\mathbf{U} - Matrix of eigenvectors

t - Time

T_w - Transmit time of the waveform across the antenna array

\mathbf{x} - Input block signal

\hat{x} - Reconstructed input signal

X - Input signal at transform domain

\hat{X} - Reconstructed signal at transform domain

List of Figures

1.1	The importance of key capabilities for generic 5G services.	3
1.2	Massive MIMO in Fixed wireless (FW) use case exploiting beamforming.	3
2.1	Multipath propagation of a signal.	10
2.2	Array steering modeling for Uniform linear array (ULA) assuming planar waves.	12
2.3	one-ring scatter model used to compute covariance matrix.	13
3.1	General transform coding block diagram.	17
3.2	Staircase structure of a scalar quantizer.	24
3.3	Example of VQ for $M = 4$, $N = 5$, and $b = 2$ bits.	26
3.4	Generic block diagram of the transform coding methods evaluated and proposed in this dissertation.	31
3.5	Algorithm for implementing Discrete cosine transform (DCT) or Discrete Fourier transform (DFT) zigzag.	33
3.6	Algorithm for implementing DCT or DFT Signal to noise ratio (SNR) target.	34
4.1	Parameters used to model the covariance matrices at transmitter and receiver of Kronecker channel model.	38
4.2	Covariance matrices for the Kronecker channel model with antenna element spacing of $d = 0.1\lambda$, distance in between transmitter and receiver of 100 m, scattering ring at receiver of 30 m, and 60 m transmitter height.	39
4.3	Covariance matrices for the Kronecker channel model with antenna element spacing of $d = 0.5\lambda$, distance in between transmitter and receiver of 100 m, scattering ring at receiver of 30 m, and 60 m transmitter height.	39

4.4	Rossllyn urban canyon scenario at Ray tracing (RT) software for Vehicle to everything (V2E) use case, with receiver antennas on top of vehicles and transmitter antenna placed in a roadside unit.	40
4.5	Delay spread over scenes of RT simulations at 2.8 GHz FW use case for 10 receivers.	41
4.6	Delay spread over scenes of RT simulations at 60 GHz FW use case for 10 receivers.	42
4.7	Delay spread for a receiver over different episodes for V2E use case simulated on RT.	42
4.8	Plot of a sample channel matrix, its covariance, and its transforms for Kronecker channel model with antenna element spacing of $d = 0.1\lambda$	44
4.9	Plot of a sample channel matrix, its covariance, and its transforms for Kronecker channel model with antenna element spacing of $d = 0.5\lambda$	44
4.10	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Kronecker channel model with antenna element spacing of $d = 0.1\lambda$	45
4.11	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Kronecker channel model with antenna element spacing of $d = 0.5\lambda$	45
4.12	Plot of a sample channel matrix, its covariance, and its transforms for 2.8 GHz FW Raymobtime dataset with antenna element spacing of $d = 0.1\lambda$	46
4.13	Plot of a sample channel matrix, its covariance, and its transforms for 2.8 GHz FW Raymobtime dataset with antenna element spacing of $d = 0.5\lambda$	47
4.14	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Raymobtime 2.8 GHz fixed environment with antenna element spacing of $d = 0.1\lambda$	48
4.15	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Raymobtime 2.8 GHz fixed environment with antenna element spacing of $d = 0.5\lambda$	48
4.16	Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz FW Raymobtime dataset with antenna element spacing of $d = 0.1\lambda$	50
4.17	Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz FW Raymobtime dataset with antenna element spacing of $d = 0.5\lambda$	50
4.18	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Raymobtime 60 GHz fixed environment with antenna element spacing of $d = 0.1\lambda$	51

4.19	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Raymobtime 60 GHz fixed environment with antenna element spacing of $d = 0.5\lambda$.	51
4.20	Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz V2E Raymobtime dataset with antenna element spacing of $d = 0.1\lambda$.	53
4.21	Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz V2E Raymobtime dataset with antenna element spacing of $d = 0.5\lambda$.	53
4.22	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Raymobtime 60 GHz mobile environment with antenna element spacing of $d = 0.1\lambda$.	54
4.23	Channel coefficients evaluation for DCT and DFT with <i>zigzag</i> and <i>sort</i> for Raymobtime 60 GHz mobile environment with antenna element spacing of $d = 0.5\lambda$.	54
4.24	SNR results for Kronecker channel model with target SNR of 6 dB.	57
4.25	Compression factor for Kronecker channel model with target SNR of 6 dB.	58
4.26	SNR results for Raymobtime 2.8 GHz FW dataset with target SNR of 6 dB.	58
4.27	Compression factor considering quantization for Raymobtime 2.8 GHz FW dataset with target SNR of 6 dB.	59
4.28	SNR results for Raymobtime 60 GHz FW dataset with target SNR of 6 dB.	60
4.29	Compression factor for Raymobtime 60 GHz FW dataset with target SNR of 6 dB.	60
4.30	SNR results for Raymobtime 60 GHz V2E dataset with target SNR of 6 dB.	61
4.31	Compression factor for Raymobtime 60 GHz V2E dataset with target SNR of 6 dB.	61

List of Tables

1.1	Use cases and Requirements for 5G.	4
3.1	Techniques for transform coding of Channel state information (CSI) employed in this work.	31
4.1	Description of the Raymobtime dataset used for compression.	41
4.2	Achieved compression ratio.	56

Contents

Acknowledgment	v
List of Figures	xiii
List of Tables	xvi
Contents	xvii
Abstract	xix
Resumo	1
1 Introduction	2
1.1 Dissertation outline	6
2 Fundamentals of MIMO processing	7
2.1 MIMO and CSI	8
2.2 MIMO propagation characteristics	9
2.3 MIMO channel modeling	11
2.3.1 Narrowband multipath channel model	11
2.3.2 Narrowband Kronecker channel model	12
2.4 MIMO channel sparsity	14
3 MIMO CSI Compression	15
3.1 Transform coding	15
3.2 Signal processing for sparse signals	17
3.2.1 Karhunen-Loève transform	18
3.2.2 Discrete cosine transform	18

3.2.3	Discrete Fourier transform	19
3.2.4	Compressive sensing	20
3.3	Side information	21
3.4	Quantization	22
3.4.1	Scalar quantization	23
3.4.2	Vector quantization	25
3.5	Compression metrics	27
3.6	CSI compression for limited rate feedback	28
3.7	Proposed low-complexity CSI transform coding	30
3.7.1	Sparsifying transform	31
3.7.2	Determining the relevant coefficients based on target distortion	32
3.7.3	Encoding few relevant coefficients	35
3.7.4	Quantization	35
4	Simulations and results	37
4.1	Acquisition and simulation of channel matrices	37
4.1.1	Simulation of Kronecker channel model	37
4.1.2	Ray tracing (RT) simulation and multipath channel model	38
4.2	Analysis of channel model and orthogonal transform	43
4.2.1	Transform analysis of Kronecker channel model	43
4.2.2	Transform analysis of RT dataset at sub-6 GHz	46
4.2.3	Transform analysis of RT dataset at Millimeter waves (mmWaves)	49
4.3	Evaluation of compression metrics	55
4.3.1	Compression ratio	55
4.3.2	Signal to noise ratio and compression factor performance	55
5	Conclusion	62
5.1	Future works	63
5.2	Published Articles	63
	Bibliography	64

Abstract

The first trials of the Fifth generation of wireless networks (5G) are taking place worldwide. A variety of use cases are envisaged, requiring flexible and scalable technologies to meet their key performance indicators. Massive MIMO is a 5G-enabling technology that improves spectral efficiency. To exploit the advantages of MIMO, the transmitter needs to have information about the channel condition (CSI) of each User equipment (UE). 5G is being standardized in Frequency division duplexing (FDD) and Time division duplexing (TDD) operational modes; hence, MIMO has to be feasible in both duplexing modes. As TDD operates downlink and uplink on the same frequency, it can rely on channel reciprocity to acquire the CSI needed to further design precoding and user scheduling, for instance. However, FDD cannot exploit channel reciprocity; therefore, massive MIMO in FDD mode is more challenging because the increasing number of antennas may turn the feedback of CSI impractical. Hence, compressing CSI in MIMO FDD systems is of interest. Furthermore, the use of vast spectrum ranges, sub-6 GHz and mmWaves bands, leads to different channel characteristics. Moreover, the close packaging of antenna elements increases the spatial correlation among a MIMO array. Consequently, this correlation can be exploited to leverage compression of CSI. This dissertation presents an overview of CSI compression methods, proposes a heuristic transform coding method with low computational cost, and does a systematic evaluation of the transform coding methods based on realistic MIMO channel simulations. Besides, the impact of different 5G use cases and design of transmitter and receiver antennas are also included on the evaluation to choose the transform coding method for compressing CSI in MIMO FDD systems.

Index Terms— MIMO, FDD, CSI compression, transform coding.

Resumo

Os primeiros testes de redes de quinta geração (5G) de telefonia celular estão sendo feitos em diversas partes do mundo. Vários cenários de uso estão sendo planejados, o que torna necessário a adoção de tecnologias flexíveis e escaláveis para atender seus indicadores de performance. MIMO massivo é uma tecnologia chave para o 5G, pois aumenta a eficiência espectral do sistema. Para explorar as vantagens do MIMO o transmissor precisa ter informações sobre a condição do canal (CSI) de cada equipamento de usuário (UE). O 5G está sendo padronizado para operar em duplexação por divisão de tempo (TDD) e duplexação por divisão de frequência (FDD). Desta forma, os custos de operar MIMO em TDD e FDD devem ser factíveis. Como TDD opera uplink e downlink na mesma frequência, a característica de reciprocidade do canal pode ser usada para extrair o CSI necessário para projetar precodificador ou alocação de usuário, por exemplo. No entanto, FDD não pode explorar reciprocidade do canal. Assim, operar MIMO em FDD é desafiador, pois com o aumento no número de antenas a quantidade de informação a ser repassada ao transmissor se torna impraticável. Desta forma, comprimir CSI de sistemas MIMO FDD é de interesse. Além disso, o uso de vastas faixas de espectro, bandas abaixo de 6 GHz e ondas milimétricas, leva a diferentes características de canal. Também, a proximidade de encapsulamento dos elementos de antena aumentam a correlação espacial do arranjo MIMO. Consequentemente, essa correlação pode ser explorada para prover compressão do CSI. Essa dissertação apresenta uma visão geral sobre compressão de CSI e propõe uma meta heurística que usa codificação por transformada de baixo custo computacional, e faz uma avaliação sistemática dos métodos de compressão baseado em simulações realistas de canais MIMO. Além, disso, o impacto de diferentes aplicações do 5G e o projeto das antenas dos transmissores e receptores também são incluídos na avaliação para escolha do método de codificação por transformada para comprimir CSI em sistemas MIMO FDD.

Palavras-chave — MIMO, FDD, compressão de CSI, codificação por transformada.

Chapter 1

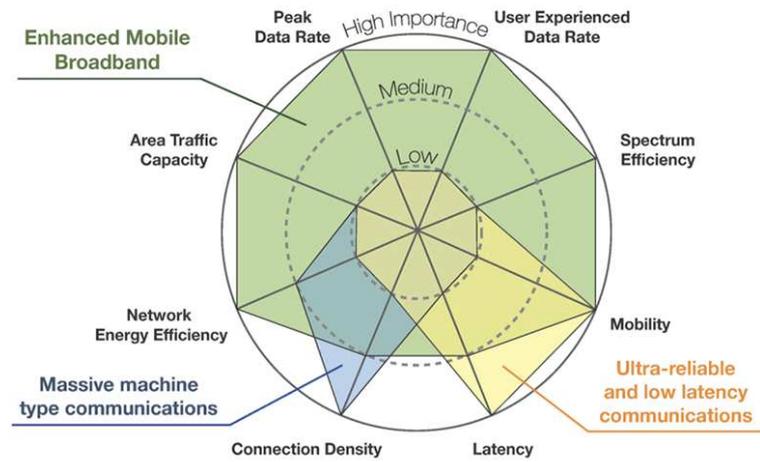
Introduction

The Fifth generation of wireless networks (5G) is already a reality. The Third generation partnership project (3GPP) has released its first standard [1] and many trials are claimed by operators and telecommunication industry worldwide [2]. 5G is revolutionary because it is a convergent network which allows a variety of diverse use cases [3,4]. Table 1.1 shows some of the expected use cases and their network requirements. It can be noticed that diverse industry verticals are leveraged by 5G, such as autonomous driving, remote surgery, Internet of things (IoT), media on demand and gaming. Therefore, 5G should be flexible and scalable to provide massive system capacity, very high data rates, ultrahigh reliability and availability, very low latency, low device cost and energy consumption, as well as be an energy efficient network [3,5].

In order to group use cases with similar requirements, International Telecommunications Union Radiocommunication sector (ITU-R) has defined three generic 5G services: Massive machine-type communication (mMTC), Enhanced mobile broadband (eMBB), and Ultra-reliable and low latency communication (URLLC). Figure 1.1 presents radar chart with the importance of each key performance indicator to each of the three generic 5G services [6]. Among the use cases expected for 5G are Fixed wireless (FW) access and Vehicle to everything (V2E) which are eMBB and URLLC services, respectively [4]. FW access is the provision of 5G broadband services to homes or small business, Figure 1.2. This is specially interesting for places where the deployment of fiber is difficult; for instance, historical heritage sites and places which require high deployment investments [7]. V2E is the use case for vehicles that connect to other vehicles, road side units, road infrastructure and others. V2E can enhance driving experience, reducing accidents and road congestion; besides, it is a premises for autonomous driving [8].

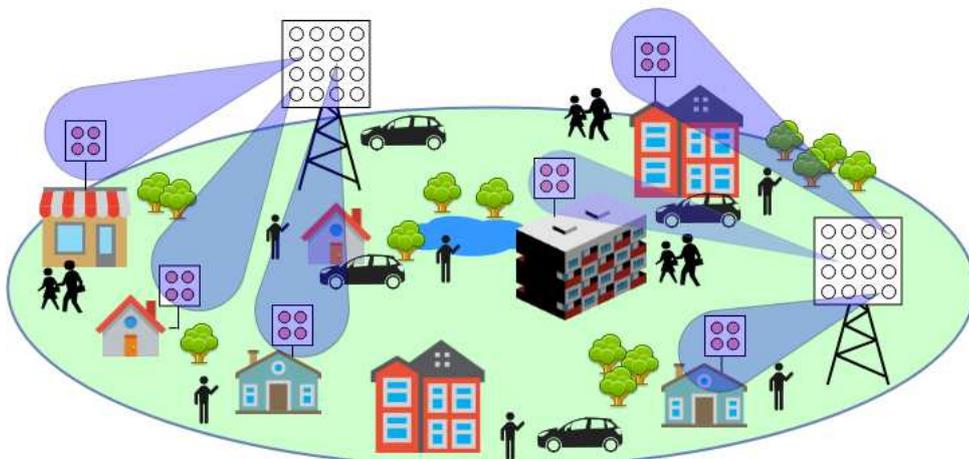
In ~~this~~ context, massive MIMO is an enabling technology for 5G because it can improve

Figure 1.1: The importance of key capabilities for generic 5G services.



Source: [6].

Figure 1.2: Massive MIMO in FW use case exploiting beamforming.



Source: Author.

Table 1.1: Use cases and Requirements for 5G.

Use cases	Requirements	Desired value
Autonomous vehicle control	latency availability reliability	5 ms 99.999% 99.999%
Emergency communication	availability energy efficiency	99.9% victim discovery rate 1 week battery life
Factory cell automation	latency reliability	down to below 1 ms down to packet loss of less than 10^{-9}
High-speed train	traffic volume density experienced user throughput mobility latency	100 Gbps/km ² in DL, and 50 Gbps/km ² in UL 50 Mbps in DL, and 25 Mbps in UL 500 km/h 10 ms
Large outdoor event	experienced user throughput traffic volume density connection density reliability	30 Mbps 900 Gbps/km ² 4 subscribers per m ² outage probability < 1%
Massive amount of geographically spread devices	connection density availability energy efficiency	1, 000, 000 devices per km ² 99.9% coverage 10 years battery life
Media on demand	experienced user throughput latency connection density traffic volume density availability	15 Mbps 5 s (start application), 200 ms (after possible link interruptions) 4000 devices per km ² 60 Gbps/km ² 95% coverage
Remote surgery and examination	latency reliability	down to below 1 ms 99.999%
Shopping mall	experienced user throughput availability reliability	300 Mbps in DL, and 60 Mbps in UL at least 95% for all applications, and 99% for safety-related applications at least 95% for all applications, and 99% for safety-related applications
Smart city	experienced user throughput traffic volume density connection density	300 Mbps in DL, and 60 Mbps in UL 700 Gbps/km ² 200 000 users/km ²
Stadium	experienced user throughput traffic volume density	0.320 Mbps 0.110 Mbps/m ²
Teleprotection in smart grid network	latency reliability	8 ms 99.999%
Traffic jam entertainment	traffic volume density experienced user throughput availability	480 Gbps/km ² 100 Mbps in DL, and 20 Mbps in UL 95%
Virtual and augmented reality	experienced user throughput latency	428 Gbps 10 ms RTT

Source: [9].

spectrum efficiency which leverages high mobile data traffic [10, 11]. MIMO relies on the use of a set of antennas at both transmitter and receiver to increase system reliability due to spatial diversity, and system capacity by allowing parallel transmissions to an UE [12]. In order to achieve optimal MIMO capacity, the Base station (BS) needs to know the Channel state information (CSI) of each UE at uplink and downlink channels. Downlink is the physical link from a BS to an UE, uplink is the link from an UE to a BS, and CSI can refer to the channel coefficients, the modulus of the channel, the statistics of the channel or the noise variance at the receiver [13]. In this dissertation, CSI corresponds to the “rich” information representing a complex channel impulse response (eventually time varying) or its corresponding frequency response.

Pilot signaling is the main method for CSI acquisition, which consists in sending a predefined orthogonal pilot signal from each transmitter antenna that all receive antennas can “listen” to estimate their channel [11]. In TDD mode the CSI can be obtained at BS by estimation from the received uplink pilot signals. Relying on channel reciprocity, the BS can directly use the uplink estimations as legitimate downlink channel information. Hence, in TDD the number of pilots depends on the amount of antennas at each UE and the number of UEs served by the same BS [10, 11, 14]. In FDD mode the UE may learn the downlink channel from pilots sent by the BS and send the estimated CSI back to the BS over a control channel. The FDD feedback is increasingly costly because the amount of feedback needed is proportional to the number of antennas at both sides, BS and UE [10, 15]. Nevertheless, there are vast amount of spectrum reserved for FDD operation; besides, FDD MIMO is key for 5G backward compatibility [11]. Therefore, compression of CSI on the feedback link is of interest to allow efficient massive MIMO operation on FDD mode.

As presented in this chapter, operating massive MIMO in FDD mode is challenging because CSI overhead is detrimental. However, there is market and spectrum for this technology. Therefore, compression of the CSI to be sent from an UE to a BS over the feedback control channel is considered in this dissertation. A primary goal is to investigate transform coding solutions which can efficiently compress the CSI. Besides, the compression solution should be flexible and require low computational cost. In this context, an heuristic method is proposed for compression of CSI. In addition, an analysis based on simulations is provided and the results validates the efficiency of the *SNR sort* transform coding method, the proposed method. The following section presents the structure of this dissertation.

1.1 Dissertation outline

- Chapter 2 presents fundamental concepts on MIMO, and its channel modeling. MIMO channel characteristics are presented and further explored to allow compression of CSI in FDD mode.
- Chapter 3 introduces the concept of transform coding and gives an overview of the state of the art in compressing CSI in MIMO FDD systems. In addition, the chapter presents the proposed heuristic compression method which is based on transform coding.
- Chapter 4 describes the simulations developed to evaluate the studied compression methods. Based on the results, the chapter discusses the performance that can be achieved in more practical scenarios.
- Chapter 5 presents the conclusions and describes possible future extensions of the work.

Chapter 2

Fundamentals of MIMO processing

Availability of spectrum, electromagnetic propagation laws, and information theory principles are among the factors which limit the performance of the physical layer in wireless communications [10]. In order to improve efficiency of wireless communications, there are basically three options: increase the density of access points, increase the spectrum usage, improve spectral efficiency which is the amount of bits that can be conveyed per second per unit of bandwidth [10, 11].

Multiple input multiple output (MIMO) technology refers to wireless communications systems that leverage multipath scattering between transmitter and receiver to substantially increase the spectral efficiency [10, 16]. The use of MIMO can increase data rates through multiplexing or diversity. For diversity, communication reliability is improved by transmitting replicas of the same signal through the fading channel [16, 17]. Multiplexing can be used to send independent data through independent signaling paths, which are obtained by exploiting the structure of the MIMO channel matrix [17]. Because of that, MIMO is considered a fundamental technology to achieve the huge growth of data demand in 5G systems.

To leverage MIMO potential to increase spectral efficiency, the BS has to have knowledge of the CSI which is composed, in this dissertation, by the channel matrix coefficients. Therefore, this chapter presents MIMO channel models and discuss their propagation characteristics. The chapter ends presenting the MIMO channel sparsity characteristic which leverages compression of CSI in massive MIMO FDD systems.

2.1 MIMO and CSI

Knowledge of CSI is fundamental for mobile communication functioning. The information regarding the channel gain matrix \mathbf{H} that is available at transmitter and receiver is called, respectively, Channel state information at transmitter (CSIT) and Channel state information at receiver (CSIR). Different assumptions can be made about their knowledge. For instance, CSIR for static channels can be fairly estimated from pilot sequences sent by the transmitter. If a feedback path is available, CSIR can be sent to the transmitter to provide CSIT. Nevertheless, CSIT can be acquired without feedback path if reciprocal propagation properties can be exploited [17]. CSIT enable to increase the transmission rate, enhance coverage, and reduce receiver complexity in MIMO wireless systems [18].

A communication system that aims to achieve the Shannon capacity of MIMO channels depends on what is known about the channel gain matrix \mathbf{H} . Exact knowledge of CSIT and CSIR allows the transmitter to adapt both power and rate to the channel gain at a given time. This adaptation leads to high bit rate because the signals transmitted or received over the multiple antennas can be coherently combined to maximize the channel Signal to noise ratio (SNR) [17]. Precoding is a technique which exploits CSIT for enhancing MIMO system performance [18]. Basically, precoding consists in choosing different phase and/or amplitude for each data signal sent from all antennas aiming to direct the signal spatially. Angular beams (or digital beamforming) are a special case of precoding that is useful in line of sight propagation. For non-line of sight channels the transmitted signal might not have a distinct angular directivity, but it can still be precoded such that the multipath components are received coherently at the UE [11].

Practical channel acquisition based on reciprocity is referred to as the open-loop method and may be applicable in TDD if the time lag between the forward and reverse transmissions is smaller than the channel coherence time. In FDD mode, however, the frequency offset between downlink and uplink is larger than the channel coherence bandwidth; hence, reciprocity cannot be exploited [18]. Channel acquisition using feedback is referred to as the closed-loop method. Feedback can be used in both TDD and FDD modes, but it is more common in FDD mode [18].

Channel sounding is the technique used for open-loop channel acquisition, the UE is scheduled to send a sounding transmission to the BS. The sounding signals are orthogonal among simultaneously scheduled users. Hence, the overhead of open-loop channel acquisition is the product of the number of training pilots on the uplink, which is proportional to the number

of antennas at each UE, and the number of users participating on the uplink channel sounding $\text{CSI}_{\text{TDD}} \sim N_{\text{UE}}N_{\text{rx}}$ [18]. Therefore, TDD operation mode is independent of the number of antennas at the BS [10].

In closed-loop MIMO, downlink training overhead is independent of the number of users, but is proportional to the number of transmitter antennas N_{tx} [10]. The overhead for uplink is proportional to the number of scheduled users multiplied by the size of the channel matrix, which is a product of the number of transmit antennas at the BS N_{tx} and the number of receive antennas at each UE N_{rx} [18]. Therefore, FDD feedback is proportional to $\text{CSI}_{\text{FDD}} \sim N_{\text{tx}} + N_{\text{UE}}N_{\text{rx}}N_{\text{tx}}$ which requires precious transmission resources, leading to the interest in compressing the feedback data.

2.2 MIMO propagation characteristics

There are three main factors that influence signal propagation in MIMO wireless communication systems: multipath propagation, Doppler effect, and spatial characteristics. Obstacles within the propagation environment cause the received signal to be a superposition of signals that propagate over different paths between transmitter and receivers, as depicted in Figure 2.1. Because of multipath, signals arrive at a receiver with different delays, attenuation, and phase shifts. Considering mobility, the Doppler effect can be considerable, resulting in frequency shift of the received signal [19]. For MIMO transceivers, the angles of multipath components characterize the spatial features. These factors influence drastically the channel impulse response (CIR). Channel characteristics are quantified by channel statistical properties such as delay spread, angle spread, Doppler spread, power angle spread (PAS), power delay profile (PDP), and spatial-temporal correlation function (STCF) [19].

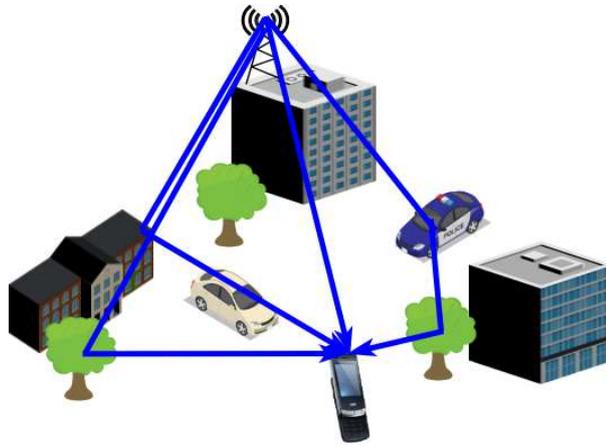
The angular spread measures the power dispersed over incident angles. For instance, the angular spread Δ at azimuth direction is

$$\Delta_{\phi} = \sqrt{\frac{\sum_{l=0}^{L-1} (\phi_l - \mu_{\phi})^2 \alpha_l^2}{\sum_{l=0}^{L-1} \alpha_l^2}}, \quad \text{with} \quad (2.1)$$

$$\mu_{\phi} = \frac{\sum_{l=0}^{L-1} \phi_l \alpha_l^2}{\sum_{l=0}^{L-1} \alpha_l^2},$$

where Δ_{ϕ} is the angular spread for azimuth angle of arrival ϕ , α_l^2 is the absolute gain of the l^{th} multipath component, and μ_{ϕ} is the mean azimuth angle of arrival. The angular spread of

Figure 2.1: Multipath propagation of a signal.



Source: Author.

azimuth angle of departure as well as elevation angular spread of arrival and departure angles can be calculated using the same Equation 2.1, but changing the correspondent parameters. Small angular spread means that the multipath components arrive/departs with angles close the mean angle of arrival/departure. Hence, better spatial multiplexing is achieved due to small intersector interference [20].

MIMO was originally conceived for conventional sub-6 GHz frequencies. With the electronic industry advancement, MIMO can reach massive number of antennas within frequency bands in the range 30-300 GHz, known as mmWaves. Propagation in those bands, sub-6 GHz and mmWaves, can suffer different physical effects, such as diffraction, shadowing, attenuation, blockage, etc [21]. On frequencies below 6 GHz, propagation depends more on path loss, shadowing, and multipath. Measurement campaigns of sub-6 GHz massive MIMO have shown that increasing the number of antenna elements leads to UE channels close to orthogonal. On mmWaves the wavelengths are very small, which increases attenuation and susceptibility to blockage given the Fresnel zone reduces allowing obstruction by small objects, such as human body and foliage [21,22]. However, mmWaves have larger spectral channels and distance based path-loss can be mitigated by directional transmissions [22].

2.3 MIMO channel modeling

Considering a MIMO channel with N_{tx} and N_{rx} antenna elements at transmitter and receiver respectively, the MIMO channel matrix $\mathbf{H}(\tau, t)$ of the time-varying channel impulse response $h_{i,j}(\tau, t)$ between the antenna elements is

$$\mathbf{H}(\tau, t) = \begin{bmatrix} h_{1,1}(\tau, t) & h_{1,2}(\tau, t) & \dots & h_{1,N_{\text{tx}}}(\tau, t) \\ h_{2,1}(\tau, t) & h_{2,2}(\tau, t) & \dots & h_{2,N_{\text{tx}}}(\tau, t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_{\text{rx}},1}(\tau, t) & h_{N_{\text{rx}},2}(\tau, t) & \dots & h_{N_{\text{rx}},N_{\text{tx}}}(\tau, t) \end{bmatrix}, \quad (2.2)$$

where $i = \{1, 2, \dots, N_{\text{rx}}\}$ indicates an antenna element of the receiver and $j = \{1, 2, \dots, N_{\text{tx}}\}$ an antenna element at the transmitter. Each column of $\mathbf{H}(\tau, t)$ is referred as the spatial-temporal signature of the j^{th} transmitter element across the receiver antenna array [23]. Given a signal x_j transmitted from the transmitter array, the signal received at the i^{th} antenna is

$$y_i(t) = \sum_{j=1}^{N_{\text{tx}}} h_{i,j}(\tau, t) * x_j(t) + n_i(t), \quad i = 1, 2, \dots, N_{\text{rx}}, \quad (2.3)$$

where $n_i(t)$ is the receiver additive noise.

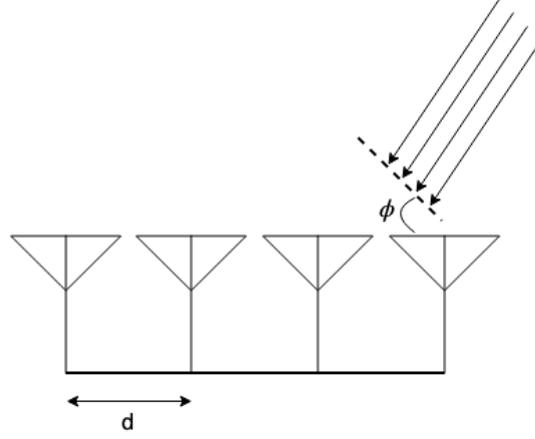
2.3.1 Narrowband multipath channel model

For convenience, it is assumed that the bandwidth BW of the waveform is much smaller than the reciprocal of the transit time T_w of the wavefront across the antenna array, $BW \ll 1/T_w$ [23]. Therefore, the signal received at the antenna element $i+1$ is a phase shifted version of the signal received at antenna element i . The corresponding phase shift depends on the array geometry and the angle of arrival of the wavefront. In this work, Uniform linear array (ULA) and Uniform planar array (UPA) are considered, Figure 2.2 depicts an ULA with four antenna elements and rays arriving with angle ϕ from the array plane. The array steering vector at the azimuth direction ϕ_ℓ is

$$\mathbf{a}(\phi_\ell) = [1, e^{-i2\pi \sin(\phi_\ell)d_\lambda}, \dots, e^{-i2\pi \sin(\phi_\ell)d_\lambda(N_{\text{tx}}-1)}], \quad (2.4)$$

where $d_\lambda = d/\lambda$ is the distance among the antenna elements normalized by the wavelength [24]. The array steering vector in the elevation direction $\mathbf{a}(\theta_\ell)$ is computed similarly, and the array steering vector at elevation and azimuth direction can be calculated as $\mathbf{a}(\phi_\ell, \theta_\ell) = \mathbf{a}(\phi_\ell) \otimes \mathbf{a}(\theta_\ell)$ [22, 24].

Figure 2.2: Array steering modeling for ULA assuming planar waves.



Source: Author.

Knowing the array steering vector, the multipath narrowband model for the array channel matrix is

$$\mathbf{H} = \sqrt{N_{\text{tx}}N_{\text{rx}}} \sum_{\ell=1}^L \alpha_{\ell} \mathbf{a}_{\text{rx}}(\phi_{\ell}^A, \theta_{\ell}^A) \mathbf{a}_{\text{tx}}^H(\phi_{\ell}^D, \theta_{\ell}^D), \quad (2.5)$$

where L is the total number of multi-path components, α_{ℓ} is the complex channel gain of the ℓ^{th} path, $\mathbf{a}_{\text{rx}}(\phi_{\ell}^A, \theta_{\ell}^A)$ is the steering vector for the angle of arrival of the ℓ^{th} path at the receiver, and $\mathbf{a}_{\text{tx}}^H(\phi_{\ell}^D, \theta_{\ell}^D)$ is for the angle of departure of the ℓ^{th} path at transmitter.

2.3.2 Narrowband Kronecker channel model

The Kronecker channel model is appropriate in FW where the BS is usually elevated and unobstructed by local scatterers, and the UE is often surrounded by local scatterers [25]. The Kronecker channel model is among the analytical models of a random MIMO channel matrix that explicitly incorporates correlation between pairs of antennas [16]. Mathematically, it is defined as

$$\mathbf{H} = \frac{1}{\sqrt{\text{tr}(\mathbf{R}_{\text{rx}})}} \mathbf{R}_{\text{rx}}^{\frac{1}{2}} \mathbf{H}_{\text{iid}} \mathbf{R}_{\text{tx}}^{\frac{1}{2}}, \quad (2.6)$$

where \mathbf{R}_{tx} and \mathbf{R}_{rx} are the covariance matrices for the transmitter and receiver, respectively, \mathbf{H}_{iid} is a complex Gaussian matrix which undergoes Rayleigh fading, and tr is the trace operation which is the sum of the elements on the matrix main diagonal [16, 26].

The elements of \mathbf{R}_{tx} are the correlations between all pairs of transmit antennas at a given

receiver, and the elements of \mathbf{R}_{rx} are the correlations between all pairs of receive antennas when the signal is emitted by a transmitter. From the structure of the covariance matrix it follows that the top and bottom halves of \mathbf{R}_{rx} and \mathbf{R}_{tx} matrices are simply the complex conjugates of each other [16]. When the scattered energy at the receiver is assumed to arrive over a uniformly distributed range of angles, the covariance matrix of an ULA can be computed as

$$[\mathbf{R}_{\text{rx}}]_{p,q} = \frac{1}{2\Delta_\phi} \int_{-\Delta_\phi + \phi_k}^{\Delta_\phi - \phi_k} e^{-j2\pi(p-q)\frac{d}{\lambda} \sin(\alpha)} d\alpha, \quad (2.7)$$

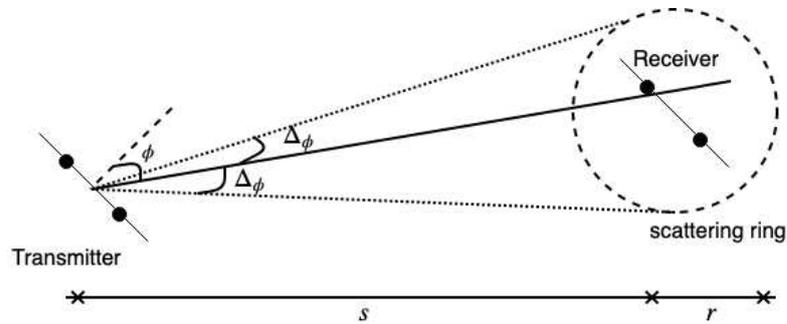
where p and q are the covariance matrix indexes, Δ_ϕ is the angular spread, d is the antenna elements separation, ϕ_k is the angle of arrival for the k^{th} receiver [25,26]. The covariance matrix of an UPA can be expressed by $\mathbf{R}_{\text{UPA}} = \mathbf{R}_{\text{V}} \otimes \mathbf{R}_{\text{H}}$ which is the Kronecker product combination of an ULA modeled at horizontal direction \mathbf{R}_{H} and other ULA modeled on vertical direction \mathbf{R}_{V} [26]. Assuming the distance from transmitter to receiver is larger than the antenna spacing, the angle of arrival and angular spread for an ULA can be computed as [25, 26]

$$\Delta_{\text{H}} = \arctan\left(\frac{r}{s}\right), \quad (2.8)$$

$$\phi_{\text{H}} \in (-\pi, \pi],$$

where r is the radius of the scattering ring for the receiver, and s is the distance between transmitter and receiver as depicted in Figure 2.3 [25,26].

Figure 2.3: one-ring scatter model used to compute covariance matrix.



Source: Author.

Generally, an increase in antenna elements spacing causes \mathbf{R}_{rx} and \mathbf{R}_{tx} to become more diagonal, therefore the antenna elements are less correlated [16, 25]. When antenna elements

spacing increases, the electromagnetic coupling among the antenna elements reduces leading to less correlation between antenna elements [16]. A decrease in angular spread increases the correlation within \mathbf{R}_{rx} as it becomes increasingly difficult for the receiver to distinguish among the transmissions of each transmit antenna element [25]. Besides, when the angle over which radio waves arrive at the receiver increases, multipath causes the signal to appear increasingly dissimilar at different antennas [16]. A counter example is when the plane wave signal arrives at the receive array moving perpendicular to the array. In this scenario, the signal at each receive antenna is the same and the antenna elements are perfectly correlated as long as electromagnetic coupling is negligible.

2.4 MIMO channel sparsity

Multipath signal propagation is a key concept for the functioning of wireless communications. Over the years, statistical modeling of wireless channels have assumed rich scattering environment. However, scientific research has shown that real wireless channels exhibits sparse multipath structure which is more noticeable when the signal spatial dimensions increase; for instance, when there is an increase in bandwidth or in number of antennas [27].

In order to explain the concept of channel sparsity, the channel is analyzed in a virtual channel representation which leverages inference of the effects of scattering and array characteristics on channel capacity and diversity [24]. Besides, virtual channel representation, also called angular domain, keeps the essence of physical modeling without its complexity and provides a tractable linear channel characterization as it employs fixed spatial basis functions defined by the spatial resolution of the antenna array [24]. Mathematically, the virtual channel \mathbf{H}_v is related to the physical channel \mathbf{H} as $\mathbf{H}_v = \mathbf{A}_{\text{rx}}^H \mathbf{H} \mathbf{A}_{\text{tx}}$, where \mathbf{A}_{rx} and \mathbf{A}_{tx} are unitary DFT matrices; hence, \mathbf{H}_v is a 2D discrete Fourier transform of \mathbf{H} [24].

The number of dominant non-vanishing coefficients of the virtual channel represents the statistically independent Degrees of freedom (DoF) in the channel which governs channel capacity and diversity [27]. The maximal number of DoF in a channel is $D_{f,\text{MAX}} = N_{\text{rx}} N_{\text{tx}}$ and achieving it requires a rich scattering environment [24]. A sparse multipath channel has very few DoF $D_f \ll D_{f,\text{MAX}}$ which represents a sparse distribution of resolvable paths in the angular domain, causing \mathbf{H}_v to have many entries close to zero [27].

Chapter 3

MIMO CSI Compression

As discussed, accurate knowledge of CSI at the BS is essential to achieve the high data rates enabled by MIMO and massive MIMO. However, the amount of CSI increases prohibitively in FDD mode as the number of antennas increases. Hence, compression of the CSI is needed to reduce the feedback overhead [26].

Transform codes are used for data compression because it is easy to apply at any rate and even with very large sequence length [28]. Modularization is the key aspect of transform coding. First, a linear transform is applied to the input sequence x , then the transform coefficients are quantized and, optionally, entropy code can be applied on the quantized indexes. Those operations are reversed at the decoder where an approximation \hat{x} of the input sequence is reconstructed [28].

This chapter presents an overview of the basic elements of transform coding, and discusses the state of art in CSI compression for FDD massive MIMO. The chapter finishes presenting the transform coding methods proposed in this dissertation.

3.1 Transform coding

Representing a waveform as a sequence of bits is referred as signal coding and is useful for reducing cost of storage or transmission through a digital channel. The design of digital coding aims to minimize the number of bits used to represent the signal while maintaining a fidelity criterion [29]. Sampling and quantization, representing analog signals in digital format, are basic operations in signal coding. In quantization, the key design goal is to balance the amount of distortion imposed by the limited number of discrete levels in the quantizer and

the amount of bits required to code these output levels. The original aim of transform coding is to extract the redundancy existing between elements of random vectors, as in multichannel environment with cross coupling among channels or in random vectors formed by blocks of consecutive samples of a scalar signal, by applying a transform and then quantize the transformed vector [29]. Quantizing the transformed coefficients allows employing less bits. Due to the energy compaction provided by the chosen transform, fewer coefficients can be encoded and transmitted along with their location in the new sequence. This results in an efficient data compression method in many applications [30].

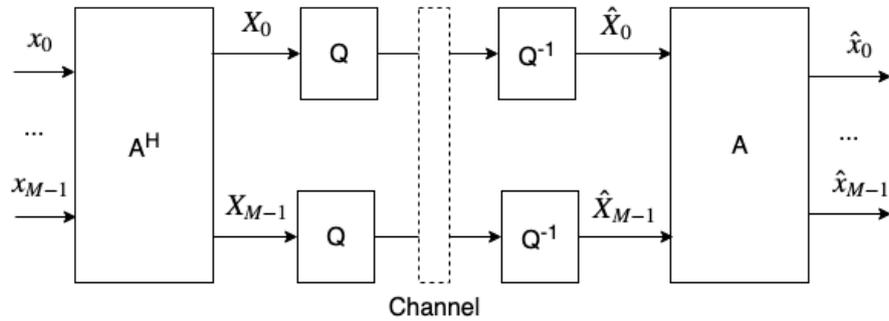
Figure 3.1 presents a general block diagram for transform coding. At the encoder, the direct linear transform of the sampled input signal \mathbf{x} of size M is $\mathbf{X} = \mathbf{A}^H \mathbf{x}$, where \mathbf{X} is \mathbf{x} in the transform domain. \mathbf{A} is assumed orthonormal and invertible, $\mathbf{A}^H \mathbf{A} = \mathbf{A} \mathbf{A}^H = \mathbf{I}$ where \mathbf{I} is the identity matrix. Hence, \mathbf{A}^H is the direct transform matrix $\Psi_{\mathbf{x}}$ and \mathbf{A} is the inverse transform matrix $\Psi_{\mathbf{x}}^H$ in order for the columns of \mathbf{A} to contain the basis vectors of the transform [29]. \mathbf{X} is then quantized to $\tilde{\mathbf{X}} = Q(\mathbf{X})$, and $\tilde{\mathbf{X}}$ is transmitted over the physical channel. Assuming the channel is noiseless, $\tilde{\mathbf{X}}$ is decoded as $\hat{\mathbf{X}} = Q^{-1}(\tilde{\mathbf{X}})$. Applying the inverse transform matrix \mathbf{A} , the reconstructed input signal $\hat{\mathbf{x}}$ is $\hat{\mathbf{x}} = (\mathbf{A}^H)^{-1} \hat{\mathbf{X}} = \mathbf{A} \hat{\mathbf{X}}$ [29]. This example used data and transforms of one dimension (1D). For two dimensional (2D) datasets like images, the 2D transformation of the matrix \mathbf{x} can be obtained from applying a 1D transform at all columns of \mathbf{x} as $\mathbf{Y} = \mathbf{A}^H \mathbf{x}$ and then transforming all rows of \mathbf{x} as $\mathbf{X} = (\mathbf{A}^H \mathbf{Y}^H)^H$. Substituting equations, the separable 2D transform is calculated as

$$\mathbf{X} = (\mathbf{A}^H (\mathbf{A}^H \mathbf{x})^H)^H = \mathbf{A}^H \mathbf{x} \mathbf{A} = \Psi_{\mathbf{x}} \mathbf{x} \Psi_{\mathbf{x}}^H. \quad (3.1)$$

Some of the most employed transforms for transform coding are presented in Section 3.2. Orthonormal transforms are usually applied because of its energy conservation property $\|\mathbf{X}\| = \|\mathbf{x}\| = \sum_{i=1}^M |x_i|^2$, and needfulness in computing inverse matrix to get the direct or inverse transform. Nevertheless, the efficiency of a transform depends on how much energy compaction it provides. The transform coding gain G_{TC} measures the energy compaction of a given transform and is computed as

$$G_{TC} = \frac{\frac{1}{M} \sum_{i=0}^{M-1} \sigma_i^2}{\left(\prod_{i=0}^{M-1} \sigma_i^2 \right)^{\frac{1}{M}}}, \quad (3.2)$$

where σ_i^2 is the variance of the i^{th} transform coefficient [29, 30].

Figure 3.1: General transform coding block diagram.

Source: Author.

3.2 Signal processing for sparse signals

Leverage channel sparsity for compression of MIMO channel matrices is essential to allow feasibility of massive MIMO in FDD mode. However, not all channel matrices are sparse at the physical domain; therefore, sparsifying basis are applied to find the few dominant coefficients which represent the channel. This section presents some orthogonal basis, or transforms, used for sparsifying the channel matrices and also introduces the concept of Compressive sensing (CS).

Primarily, transforms are used to reduce the complexity of mathematical problems. For data compression, the desirable transform should present the following characteristics: data decorrelation, data-independent basis functions, and fast implementation [31]. Data decorrelation is when the transform compacts most of the signal energy in a few number of coefficients, allowing to further discard coefficients. The optimum transform is usually data-dependent due to the large statistical variations among data [31]. Then, finding those data-dependent basis functions may be computationally difficult and the functions need to be informed to the decoder in the case of signal compression. Besides, the basis functions should be updated more frequently when the data blocks are highly non-stationary. Hence, it is desirable to trade optimum performance for a transform whose basis functions are data-independent [31]. Nonetheless, fast transform implementations are desired to reduce computational complexity.

3.2.1 Karhunen-Loève transform

Karhunen-Loève transform (KLT) is the transform which exactly diagonalizes the correlation matrix of any signal $\mathbf{R}_{xx} = \mathbf{E}[\mathbf{xx}^H]$ [32]. In essence, the KLT can be obtained from the eigen-decomposition of a given random signal autocorrelation matrix $\mathbf{R}_{xx} = \mathbf{U}\mathbf{D}\mathbf{U}^H$ where \mathbf{U} is the matrix of column eigenvectors which are the orthogonal basis function of the KLT $\Psi_{\text{KLT}} = \mathbf{U}$, and \mathbf{D} is a diagonal matrix with the corresponding eigenvalues of \mathbf{R}_{xx} .

The KLT is said an optimal transform because it completely decorrelates the signal in the transform domain [28, 32]. Besides, it contains the most variance in the fewest number of transform coefficients, which allows to minimize the Mean square error (MSE) in data compression. Even though KLT has the ideal property of data decorrelation, its signal dependence, lack of a fast transform algorithm, and the need to transmit its basis make it difficult to adopt it in compression. However, KLT serves as benchmark against which other discrete transforms may be evaluated [32].

3.2.2 Discrete cosine transform

DCT is a sinusoidal unitary transform which is invertible and its kernel describes a set of complete orthogonal discrete cosine basis functions [31]. The 2-D DCT for an $N \times N$ input data matrix $x_{m,n}$ with $m, n = [0, 1, \dots, N - 1]$ is defined by the following relation

$$\Psi_{\text{DCT}(k,l)} = \frac{4\epsilon_k\epsilon_l}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x_{m,n} \cos\left(\frac{\pi(2m+1)k}{2N}\right) \cos\left(\frac{\pi(2n+1)l}{2N}\right), \quad (3.3)$$

$$k, l = 0, 1, \dots, N - 1,$$

$$\epsilon_p = \begin{cases} \frac{1}{\sqrt{2}} & p = 0 \\ 1 & \text{otherwise.} \end{cases}$$

For first-order stationary Markov (Markov-1) processes, the current stage of the process just depends on the last stage $P(s_t | s_{t-1}, s_{t-2}, s_{t-3}, \dots, s_0) = P(s_t | s_{t-1} \forall t)$, where s_t is the current stage and $P(\cdot)$ is the probability, and the dynamics of the process does not change over time $P(s'_t | s'_{t-1}) = P(s_t | s_{t-1}) \forall t$. According to [31, 32], the DCT approaches asymptotically the KLT performance for (Markov-1) processes with high correlation, independent of the size N of the signal sequence. This is true because DCT provides good eigenvectors approximation for Toeplitz matrices [33]. Besides, [34] has compared the asymptotic equivalence of DCT and DFT with KLT, and proved that DCT performs closer to optimal than DFT for Markov-1

signals. Due to its powerful performance in the bit-rate reduction, DCT have been employed in international image and video coding standards, such as JPEG, MPEG, H.261, H.263 [31].

3.2.3 Discrete Fourier transform

The 2D-DFT for an $M \times N$ data matrix $x_{m,n}$ with $m = [0, 1, \dots, M - 1]$ and $n = [0, 1, \dots, N - 1]$ is defined as

$$\Psi_{\text{DFT}(k,l)} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{m,n} e^{-j2\pi\left(\frac{k}{M}m + \frac{l}{N}n\right)}, \quad (3.4)$$

$$k = 0, 1, \dots, M - 1 \quad \text{and} \quad l = 0, 1, \dots, N - 1,$$

For a stationary Markov-1 signal, the correlation matrix is a symmetric Toeplitz matrix which may be treated as asymptotically equivalent to a circulant matrix. As the eigenvectors of the circulant matrix are the basis vectors of the discrete Fourier transform, KLT and DFT are asymptotically equivalent [32]. In addition, DFT can be used to compute DCT [33].

For One dimensional (1D) transforms, an M point DCT can be computed from $2M$ point DFT as

$$\Psi_{\text{DCT}}(0) = \frac{\sqrt{2}}{M} \sum_{m=0}^{M-1} x_m$$

$$\Psi_{\text{DCT}}(k) = \frac{2}{M} \Re\left\{ e^{-\frac{ik\pi}{2M}} \sum_{m=0}^{2M-1} x_m e^{-\frac{i2\pi km}{2M}} \right\} \quad (3.5)$$

$$k = 1, 2, \dots, M - 1 \quad \text{and} \quad x_m = 0, m = M, (M + 1), \dots, (2M - 1).$$

As it is a separable transform, the Two dimensional (2D) transform can be obtained from the 1D transform version. Even though DCT can be computed from DFT, DCT performs better in terms of compression for Markov sources with high correlation coefficient. To compute DFT of length M , it is assumed that the sequence is periodic with period M . This repetition of the sequence at each M points causes sharp discontinuities at the edges of the sequence which leads to high frequency components. The representation of high frequency components affects other sequence points besides its endpoints. Therefore, discarding high frequency coefficients in DFT leads to additional distortion [30].

3.2.4 Compressive sensing

Compressive sensing (CS) exploits the compressibility of a signal to reduce the amount of measurements required to completely describe a signal [35]. This technique can be applied to sparse signals where only K components of the $N \times 1$ signal vector are non-zero, $K \ll N$. The sparse representation \mathbf{s} of the time signal \mathbf{x} is $\mathbf{x} = \Psi_{\mathbf{x}}\mathbf{s}$. Following, the CS operation is defined as

$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi_{\mathbf{x}}\mathbf{s} = \Theta\mathbf{s}, \quad (3.6)$$

where Φ is the $M \times N$ random measurements matrix with $M < N$, $\Psi_{\mathbf{x}}$ is a sparsifying basis (DCT, DFT, KLT, etc), $\mathbf{s} = \Psi_{\mathbf{x}}^H\mathbf{x}$, and $\Theta = \Phi\Psi_{\mathbf{x}}$.

The aim of CS is to take M measurements from which the N -length \mathbf{x} can be stably reconstructed, or equivalently \mathbf{s} in the $\Psi_{\mathbf{x}}$ domain. In general, this linear problem has less equations than unknowns as $M < N$. However, the K -sparsity of \mathbf{s} can be exploited where \mathbf{y} simplifies to a combination of the K -columns of Θ . As $K \leq M$, the $M \times K$ system is well conditioned if Θ preserves the length of the particular K -sparse vectors [35]. From that, the elements of the measurement matrix Φ can be generated from random variables with Gaussian or Bernoulli distributions, and $\Theta = \Phi\Psi_{\mathbf{x}}$ has to present restricted isometry property (RIP) [35, 36]. The compressive capability is bounded by $M \geq cK \log(N/K)$ for c of small value [26, 35]. For reconstruction of \mathbf{x} from \mathbf{y} , the decoder should solve

$$\hat{\mathbf{s}} = \arg \min \|\mathbf{s}\|_{l_1} \quad \text{such that} \quad \Theta\mathbf{s} = \mathbf{y}. \quad (3.7)$$

Equation (3.7) is a convex optimization problem that can be solved by linear programming (LP), basic pursuit (BP), and Orthogonal matching pursuit (OMP) [36]. OMP is an iterative greedy algorithm that is presented at Algorithm 1. At each step, OMP algorithm selects a column which is most correlated with the current residuals r_s , as in line 6 of Algorithm 1. This column is added to the set of selected columns W , as in line 7 of Algorithm 1. The algorithm updates the residuals, line 8 of Algorithm 1, by projecting the observations into the linear subspace spanned by the columns which have been selected, as in line 7 of Algorithm 1, and the algorithm iterates. OMP has simple and fast implementation if compared with other alternative methods [26, 37].

Algorithm 1: OMP

Input: Measurements \mathbf{y} Measurement matrix Φ Sparsifying basis Ψ_x Sparsity K **Output:** $\hat{\mathbf{h}} = \Psi_x \hat{\mathbf{s}}$ **begin**

```

1    $k \leftarrow 0$ 
2    $\mathbf{r}_s^0 \leftarrow \mathbf{y}$ 
3    $W^0 \leftarrow \emptyset$ 
4   while  $k < K$  do
5      $k = k + 1$ 
6      $w^k = \arg \max_j |\langle \mathbf{r}_s^{k-1}, \Phi_j \rangle|$ 
7      $W^k = W^{k-1} \cup w^k$ 
8      $\hat{\mathbf{s}}_{W^k} = \arg \min_{\mathbf{s}} \|\mathbf{y} - \Phi_{W^k} \mathbf{s}\|_2$ 
9      $\mathbf{r}_s^k = \mathbf{y} - \Phi_{W^k} \hat{\mathbf{s}}_{W^k}$ 
   Reconstruction:  $\hat{\mathbf{s}} = \arg \max_{\mathbf{s}: \text{supp}(\mathbf{s})=W^k} \|\Phi \mathbf{s}\|_2$ 

```

3.3 Side information

In coding literature, such as in image processing, sparsifying basis are commonly applied and coefficients are discarded. The encoder uses some bits to inform which coefficients are relevant and, therefore, not discarded. Eventually, assuming scalar quantization, the encoder needs to specify the amount of bits allocated to each coefficient, unless this amount is fixed over time. All the bits used to encode information other than the coefficient values themselves is called *side information*.

One key aspect of conventional transform coding is to encode the relevancy of the coefficients. For example, the JPEG image encoder assumes a fixed relevancy order of coefficients called *zigzag* scan (top-bottom, left-right), which eliminates the need for some side information on the relevant coefficients. In [26], the zigzag scan was adopted to encode the MIMO CSI. It should be noted that compressive sensing consists of a technique that does not require side information to encode the coefficients significance. The relevancy of each coefficient in compressive sensing is found via optimization algorithms.

In general, the selection of relevant coefficients may not follow a fixed scheme as in the zigzag scan. In these cases, the encoder has to send side information for the decoder to be able to restore the compressed data correctly. For instance, a binary map can be used to indicate the set of significant transformed coefficients. The binary map representation of the chosen transform coefficients is referred as *significance map* and it has been used in signal compression strategies based on transforms such as wavelets, and in video coding standards [38–40]. Significance maps of channel matrices use one bit per coefficient, to indicate that coefficient was discarded or not [40]. Therefore, a significance map always require the same number of bits, which coincides with the total number of coefficients.

Other option for encoding position of significant coefficients is the run length encoding (RLE) which is adopted in the PCX image format. When a RLE encoder detects long sequence of the same value, it indicates this specific value and its number of occurrences. This can be very efficient for images, especially for encoding the background pixels. However, for CSI compression, with complex-values entries in the \mathbf{H} matrix, RLE could not be applied to the \mathbf{H} values. One could still execute RLE on the binary matrix that is used in the *bitmap* representation. However, RLE applied to binary matrices has some complications. Encoding a run of bits 0 or 1 would require bits to represent a given range of integer numbers. For example, a run of 150 bits zero, would require 8 bits to represent the number of occurrences. Unless some entropy coding such as Huffmans code is adopted, the number of bits required to represent the maximum run would be needed each time a bit is flipped from 0 to 1 or vice-versa. RLE and entropy coding would make the encoding more complex which is not desired in this dissertation.

3.4 Quantization

Quantization is one of the simplest and most general ideas of lossy compression where large amounts of information should be represented in a small number of pre-defined symbols (codewords) [30]. For lossy compression schemes, the design of a quantizer has significant impact on the amount of compression obtained and loss incurred. The input and output values of a quantizer can be scalar or vectors which leads to scalar quantizers and vector quantizers, respectively [30]. The following sub-sections present the basics of each of those quantizers.

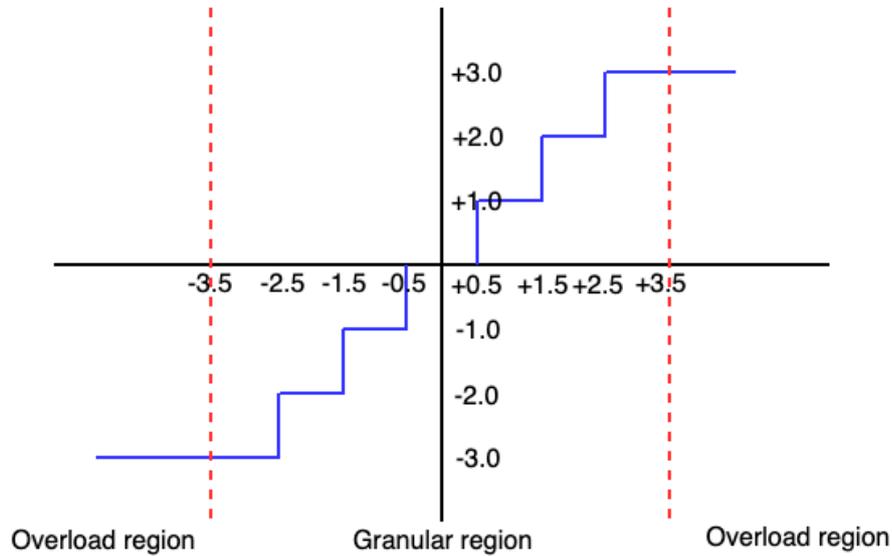
3.4.1 Scalar quantization

Encoder and decoder mapping are the basic parts of a quantization. At the encoder, the range of source values generated are divided into a number of intervals where each interval is represented by a unique codeword. Following, the encoder represents all the source values that fall on a particular interval by its respective codeword. As knowing the codeword just tells the interval the source value belongs, quantization is said irreversible. However, the decoder happens to reconstruct the source values with some error. The decoder outputs a value that best represents all the values in the interval represented by a codeword [30].

A scalar quantizer encodes each of its input to a real scalar output, generating codewords for each fixed-length quantization interval. More precisely, an M -point scalar (one dimensional) quantizer $Q : \mathcal{R} \rightarrow \mathcal{C}$ where \mathcal{R} is the real line, $\mathcal{C} = \{y_1, y_2, \dots, y_M\}$ is the codebook with size $|\mathcal{C}| = M$, and $y_i \in \mathcal{R}$ are the output levels, or codewords [41]. In practical applications, M is finite to allow output values to be specified by a finite number of binary digits $M = 2^b$. For instance, consider a digital thermometer which displays temperatures in the range $[0, 1, 2, \dots, 99]$ with 1° resolution. \mathcal{R} is the set of all possible temperatures to which the thermometer can be exposed, \mathcal{C} is the range of temperatures it can display. The mapping Q is determined by the sensor resolution which is 1° Celsius. Therefore, for any actual temperature in the range of 0 to 99 degrees, the output reading gives the true value rounded to the nearest integer [41].

Associated with every M -point quantizer is a partition of the real line \mathcal{R} into M cells or atoms R_i for $i = 1, 2, \dots, M$. The i^{th} cell is given by $R_i = \{x \in \mathcal{R} : Q(x) = y_i\} \equiv Q^{-1}(y_i)$, the inverse image of y_i under Q . From this definition, there is no overlap between partitions cells. A quantizer Q is completely described by its output levels $\{y_i : i = 1, 2, \dots, M\}$ and the corresponding partition cells $\{R_i : i = 1, 2, \dots, M\}$ [41]. The endpoints of the intervals are known as decision boundaries $y_i \in (x_{i-1}, x_i)$. Each bounded cell is called a granular cell, an unbounded cell is referred as overload cell. Usually, $x_0 = -\infty$ and $x_M = \infty$ and the cells R_0 and R_M are overload cells. The total length of the granular cells is called range B where for an unbounded regular quantizer is $B = x_{M-1} - x_1$. Figure 3.2 presents the structure of a uniform scalar quantizer and highlights its granular and overload regions.

Code rate r_{sq} of a scalar quantizer is defined as $r_{sq} = \log_2(M)$ and measures the number of bits needed to uniquely specify a quantized value. The code rate impacts on the accuracy with which the source values are described. For an integer value of code rate, each y_i could be assigned to a unique r -tuple. This is, a fixed rate code is being used to represent the quan-

Figure 3.2: Staircase structure of a scalar quantizer.

Source: Author.

tized value [41]. The quantization performance is directly dependent on the partition boundary values, the output points and the input source statistics. Nevertheless, the load factor measures the size of the highest decision level x_{M-1} relative to the root mean square (RMS) value σ of the input signal. Hence, the loading factor mediates a trade-off between granular and overload distortion. Mathematically, the loading factor γ_f is given by

$$\gamma_f = \frac{V}{\sigma}, \quad (3.8)$$

where V is the peak signal magnitude that can be quantized without incurring overload error. The value of V , is usually defined as x_{M-1} or y_i . For the case of symmetric quantizer $V = B/2$ where $B = x_{M-1} - x_1$ is the total length of the granular cells.

The uniform quantizer is the simplest type of quantizer where all intervals are of the same size, except possibly for the two outer intervals which may form the overload region. For the uniform quantizer, the decision boundaries are spaced evenly as well as the reconstruction values which have the same spacing as the decision boundaries [30]. This constant spacing is usually referred to as the step size Δ . For instance, considering a source with uniformly distributed entries on the bounded interval $[-X_{\text{MAX}}, X_{\text{MAX}}]$, we aim to design a M -level uniform quantizer for this source. Therefore, the interval $[-X_{\text{MAX}}, X_{\text{MAX}}]$ should be divided into M

equally sized intervals where the step size is given by

$$\Delta_u = \frac{X_{\text{MAX}} - (-X_{\text{MAX}})}{M} = \frac{2X_{\text{MAX}}}{M}. \quad (3.9)$$

The distortion measures the overall quality degradation incurred by quantization. For a uniform quantizer with levels distributed equally for positive and negative levels, and source probability distribution function (*pdf*) $f_x(x)$, the quantization distortion σ_q^2 is

$$\sigma_q^2 = \mathbb{E}[(x - \hat{x})^2] = 2 \underbrace{\sum_{i=1}^{M/2} \int_{(i-1)\Delta}^{i\Delta} \left(x - \frac{2i-1}{2}\Delta\right)^2 f_x(x) dx}_{\text{granular distortion}} + 2 \underbrace{\int_{(M/2)\Delta}^{\infty} \left(x - \frac{M-1}{2}\Delta\right)^2 f_x(x) dx}_{\text{overload distortion}}. \quad (3.10)$$

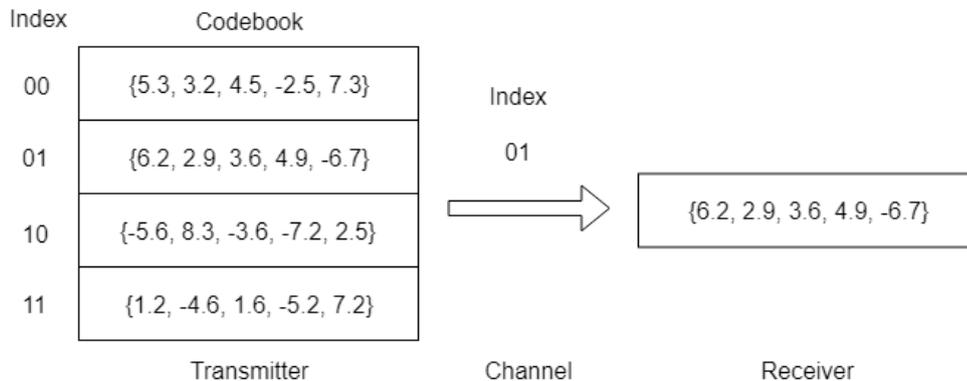
Assuming all quantizer intervals are bounded, the number of levels provide high resolution, and the distortion is uniformly distributed over the granular region, the quantization error can be simplified as

$$\sigma_{q,u}^2 = 2 \sum_{i=1}^{M/2} \int_{(i-1)\Delta}^{i\Delta} \left(x - \frac{2i-1}{2}\Delta\right)^2 \frac{1}{2X_{\text{MAX}}} dx = \frac{\Delta^2}{12} [30]. \quad (3.11)$$

3.4.2 Vector quantization

Vector quantization (VQ) is a generalization of scalar quantization to quantize vectors. A block of N consecutive source samples are grouped into blocks, and those blocks are the input to a vector quantizer. The N dimensional codebook is known at encoder and decoder. The vectors in this codebook are refereed as code-vectors, they are selected to be representative of the vectors generated from the source. Each code-vector is assigned to a binary index. The encoder compares the input vector to each code-vector and decides on the code-vector which leads to the smallest distortion; therefore, the code-vector is the quantized value of the source block. The binary index of the selected code-vector is informed to the decoder to allow recovery [30]. Figure 3.3 presents an example of VQ. The source block is composed of 5 samples, and the transmitter just sends the index of the code-vector to the receiver that recovers the original source block.

There are several techniques to reduce its computational cost, but VQ may require considerable amount of computation at the encoder in order to find the closest reproduction vector to the block of source outputs. However, the decoder is a simple look up table. Hence, vector quantization is rather attractive for applications where the resources available for decoding are

Figure 3.3: Example of VQ for $M = 4$, $N = 5$, and $b = 2$ bits.

Source: Author.

scarce if compared to available resources for encoding [30]. Comparing with scalar quantization, VQ gives flexibility to represent the source outputs; besides, if the source outputs are correlated, this structure can be explored to generate efficient representations. Thus, codebook design is one of the most important aspects of VQ design.

Rigorously, a N dimensional vector quantizer Q with size M is a mapping from a vector in N -dimensional Euclidean space \mathcal{R}^N , into a finite set \mathcal{C} containing M output or reproduction points, $Q : \mathcal{R}^N \rightarrow \mathcal{C}$ where $\mathcal{C} = \{y_1, y_2, \dots, y_M\}$ and $y_i \in \mathcal{R}^N$ for each $\mathcal{J} \equiv \{1, 2, \dots, M\}$ [41]. The set \mathcal{C} is called codebook and has size M , with each entry being a vector in \mathcal{R}^N . The code rate of a vector quantizer is $r_{vq} = (\log_2 M)/N$, it measures the number of bits per vector component which is used to encode the input vector [41].

Associated with every M point vector quantizer is a partition of \mathcal{R}^N into M regions or cells, R_i for $i \in \mathcal{J}$. The i^{th} cell is defined as

$$R_i = \{x \in \mathcal{R}^N : Q(x) = y_i\}, \quad (3.12)$$

denoted more concisely as $R_i = Q^{-1}(y_i)$. From this definition it follows that $\bigcup_i R_i = \mathcal{R}^N$, and $R_i \cap R_j = \emptyset$ for $i \neq j$, hence the cells form a non-overlapping partition of \mathcal{R}^N [41].

Linde-Buzo-Gray (LBG) algorithm is a popular approach to design the quantizer codebook. LBG is very similar to the k -means algorithm from pattern recognition applications, differing on its initialization [30]. If the goal is to have M codewords, k -means starts with $k = M$ number of means or random codewords while LBG designs codebooks with $k = 1, 2, 4, 8, \dots, M$ means/centroids recursively. Both algorithms depend on a training set which

is a large set of output vectors from the source. Each of the training elements is assigned to the closest codeword. After all elements are assigned, the codewords are updated by computing the centroids of the training set vectors assigned to them. When the assignment process is complete, there will be M new groups of vectors clustered around each of the output points, the code-vectors [30]. Both LBG and k -means algorithms guarantee that the distortion from one iteration to the next will not increase. However, there is no guarantee that the procedure will converge to the optimal solution. The initial conditions influence the convergence of the algorithms.

3.5 Compression metrics

The performance goal of lossy compression schemes is to achieve low distortion using fewer bits to represent the information-bearing sequence. Fidelity of the reconstructed source sequence to the original is accounted by some measurement (for example, a function of the difference between the two sequences) of the distortion introduced by the compression process [30]. The square error measure $d(x_n, \hat{x}_n) = (x_n - \hat{x}_n)^2$ and the absolute difference measure $d(x_n, \hat{x}_n) = |x_n - \hat{x}_n|$ are popular measures of distortion where x_n is the source sample and \hat{x}_n is the reconstructed sample. Average measures are used instead of term-by-term distortion; hence, the average of the squared error measures, also called MSE, is most used. Modeling a sequence of length N as realizations of random variables X_n and their reconstructions \hat{X}_n , the MSE is

$$D_{MSE} = \mathbb{E} \left\{ \frac{1}{N} \sum_{n=1}^N (X_n - \hat{X}_n)^2 \right\} = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[(X_n - \hat{X}_n)^2], \quad (3.13)$$

where the expectation is with relation to the joint distribution of X_n and \hat{X}_n . In practice, the information and reconstruction are assumed ergodic processes and Equation (3.13) can be simplified to

$$\sigma_d^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{x}_n)^2, \quad (3.14)$$

where the variance σ_d^2 of the distortion sequence $d(x_n, \hat{x}_n)$ is the MSE [30] in case the difference has zero-mean.

Other measure related to distortion is the Signal to noise ratio (SNR), which is the power of the incurred error relative to the input signal power. It is computed as

$$\text{SNR} = \frac{\sigma_x^2}{\sigma_d^2}, \quad (3.15)$$

where σ_x^2 is the average squared value of the source signal (zero-mean signals are assumed hereafter, unless otherwise noted). SNR can be computed in *decibels* as $\text{SNR}(dB) = 10 \log_{10}(\text{SNR})$.

There are several definitions related to compression ratios to measure the efficiency of data compression techniques. This dissertation adopts two metrics, the first one is referred as compression ratio η that concerns only the number of discarded coefficients $N_{\text{discarded}}$, not taking in account quantization, and is given by

$$\eta = \frac{N_{\text{discarded}}}{N_{\text{total}}}, \quad (3.16)$$

where N_{total} is the total number of channel coefficients. The second metric is the compression factor defined by

$$\text{CF} = \frac{b_{\text{in}}}{b_{\text{out}}}, \quad (3.17)$$

where b_{in} is the number of input bits required to represent a given channel matrix \mathbf{H} , and b_{out} is the number of bits required to represent \mathbf{H} after quantization. CF and η can be used for a single realization of a channel matrix, or in the average sense. One can also divide the numerator and denominator of CF and interpret CF as the ratio of the average number of bits per channel matrix coefficient. These metrics will be useful for comparing different compression methods in this dissertation.

3.6 CSI compression for limited rate feedback

Feedback in communications systems can enable the transmitter to exploit channel conditions and avoid interference. For instance, feedback can be used to specify a precoding matrix at the MIMO transmitter, which leverages the strongest channel modes [42]. Designing limited feedback MIMO systems is a nontrivial problem with the potential for substantial performance gains. The two main approaches for feedback design is quantizing the channel or quantizing the properties of the transmitted signal. Allowing the receiver to send a quantized version of the forward channel to the transmitter gives more flexibility for the transmitter to choose the most suitable space-time signaling technique [42]. For a fixed transmission technique, performance gains can be achieved by improving the quantization of the needed information to adapt the transmitted signal to current channel conditions.

The article [43] provides an overview of many methods for limited feedback designed to quantize the signal properties, such as covariance matrix and beamforming designed at receiver,

either in single user or multiuser wireless MIMO systems. However, those techniques rely on some assumption that may actually limit system performance. The work in [44] employed DCT to sparsify channel vectors, and data compression based on VQ. The method relies on a power factor which defines a zero-one filtering matrix that will further be applied to discard coefficients of the DCT transformed channel vector; the filtering matrix is sent as side information to the transmitter. In addition, VQ encoding is performed at UE side employing Grassmanian codebook. The authors provide a mathematical analysis of the distortion induced by this method and conclude that there is a trade off between the power factor and amount of quantization bits. However, the method proposed in [44] is computationally difficult because it operates VQ encoding at UE side which has relatively high computational cost. Besides, the results were evaluated for high amount of bits and the distortion was just presented in graphical format.

An adaptive CS approach and a rank-1 beamforming KLT method are proposed in [36] for compressing CSI of spatially correlated antenna arrays. The adaptive CS scheme is designed to have different levels of compression depending on the channel gain available and its tolerance to error. As example, it defines a two level adaptation where M_1 coefficients are random measured if channel gain is greater than a threshold γ , or M_2 coefficients are random measured if channel gain is below γ , and $M_1 < M_2 \ll N_{tx}N_{rx}$. This adaptive method employs DCT as sparsifying basis and the method overhead is to inform the transmitter the values of M . The rank-1 beamforming KLT method relies on slow varying channels to exploit the first singular value of the channel matrix. The side information consists of informing the KLT basis used; however, it is updated according to a MSE threshold. If the KLT basis become obsolete, the system is reset and CS is applied to update the channel estimate that is further used to the rank-1 beamforming approach [36]. Despite the reduction in feedback, the method assumes the UE emulates the recovering process in order to cope with KLT basis update, which increases complexity at receiver side. Besides, the method is limited to slow varying circumstances which may not happen in many 5G scenarios.

The article [45] proposed a low-complexity adaptive feedback to exploit spatial and temporal correlation of slow varying MIMO channels. It is assumed that the transmitter and receiver know the channel statistics, and the system has memory of past feedback. It proposes the use of KLT for compressing the first channel of K block channels to M dominant eigenvectors, and the following blocks are feedback according to its difference from a previews block. The block differences are also projected to a M dimensional vector. The projection basis are de-

rived from the spatial correlation matrix \mathbf{R} considering temporal correlation. The performance of this method at average SNR gain reaches the one corresponding to having perfect CSI within 10 channel blocks at 95 % compression ratio. However, the method proposed in [45] does not consider quantization error.

CS and transform coding with DCT and KLT are compared in [26]. CS is proposed for a scenario where the UE does not know which basis is employed, while DCT or KLT with dimensionality reduction is proposed for a scenario where the UE knows the adopted basis. While using DCT, the channel coefficients are read in *zigzag* order and K low frequency dominant components are kept for quantization. For KLT, the eigenvalues and eigenvectors are rearranged in descending order and K dominant elements are selected for quantization. The innovation on the use of CS is the introduction of modified OMP which reduces the reconstruction complexity. VQ with codebook generated by Linde-Buzo-Gray (LBG) algorithm is used for quantization. The results in [26] shows that better compression performance is achieved when antennas are closely packed, which corresponds to highly correlated channels. Besides, dimensionality reduction methods outperform CS method [26].

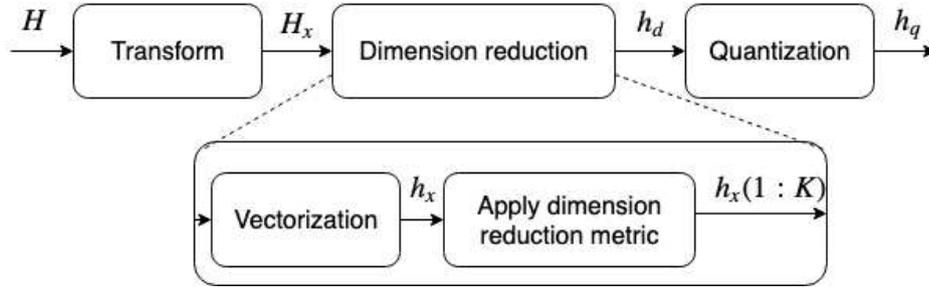
Inspired by those works, the next section presents the combination of techniques that compose the methods proposed in this dissertation. Focus is given to transform coding methods which applies DCT or DFT with dimensionality reduction and scalar quantization to reduce complexity.

3.7 Proposed low-complexity CSI transform coding

In this dissertation, the adopted transform coding method for compression of CSI is composed by three phases: sparsifying basis, dimensionality reduction, and scalar quantization. The dimensionality reduction phase is divided in matrix vectorization and application of the dimensionality reduction metric. Figure 3.4 presents the general block diagram of the compression methods and Table 3.1 presents the combinations of techniques which are explored therein.

The following sections will explain the nomenclature in Table 3.1. For example, the compression ratio is equivalent to arbitrarily choosing the number K of relevant coefficients, while “SNR target” adopts a value K derived from the target SNR.

Figure 3.4: Generic block diagram of the transform coding methods evaluated and proposed in this dissertation.



Source: Author.

Table 3.1: Techniques for transform coding of CSI employed in this work.

Sparsifying basis	Matrix vectorization	Dimension reduction metric	Reference name
DCT	<i>zigzag</i>	compression ratio	DCT <i>zigzag</i>
	<i>sort</i>	SNR target	DCT <i>SNR target</i>
DFT	<i>zigzag</i>	compression ratio	DFT <i>zigzag</i>
	<i>sort</i>	SNR target	DFT <i>SNR target</i>

Source: Author.

3.7.1 Sparsifying transform

As presented in Section 2.4, MIMO channel matrices may be sparse. However, sparsity is usually not clear on the channel physical domain; hence, sparsifying basis are needed to reveal the dominant channel coefficients. The orthogonal transform basis is Ψ_X and the 2D transformed channel matrix is defined as

$$\mathbf{H}_X = \Psi_X \mathbf{H} \Psi_X^H, \quad (3.18)$$

where $(\cdot)_X$ refers to the chosen transform, and \mathbf{H} is the channel matrix in the physical domain, coming from multipath channel model or Kronecker channel model, for instance. Here, DCT and DFT are evaluated as candidates for the proposed low complexity transform coding method.

3.7.2 Determining the relevant coefficients based on target distortion

Following the change of basis, the coefficients matrix \mathbf{H}_X should be vectorized and further undergoes dimensionality reduction. This section presents the two main methods explored in this dissertation for vectorization and dimensionality reduction: *zigzag* method and *SNR target* method.

3.7.2.1 Zigzag method

After channel matrix transform and inspired by JPEG, the vectorization is made by reading \mathbf{H}_X in *zigzag* manner $h_z = \text{zigzag}(\mathbf{H}_X)$, as proposed in [26]. For dimensionality reduction, [26] uses the compression ratio η as the design metric. Knowing the size of the antenna arrays at transmitter N_{tx} and receiver N_{rx} , the total number of channel coefficients is $N_{total} = N_{tx}N_{rx}$ which allows to compute the number of coefficients that should be discarded $N_{discarded}$ from a specified value for η . The number of coefficients that should be coded is $K = N_{total} - N_{discarded}$. Therefore, the first K coefficients of h_z are quantized and the value of K is sent as side information to the decoder. Figure 3.5 provides the flowchart for *DCT zigzag* and *DFT zigzag* compression methods.

3.7.2.2 SNR target method

Following the change of basis, our proposed option [46] is to vectorize $h_v = \text{vec}(\mathbf{H}_X)$ and sort in ascending order its coefficients according to their absolute values $h_s = \text{sort}(|\mathbf{H}_X(:)|)$, where $(:)$ is a column-wise matrix vectorization (one could also adopt a row-wise vectorization). The next phase is dimensionality reduction where a number of h_s coefficients are zeroed according to the SNR target (SNR_t) design metric as explained in the sequel.

The desired SNR_t is set and the corresponding total distortion σ_t^2 is computed as

$$\sigma_t^2 = \sigma_q^2 + \sigma_d^2 \quad (3.19)$$

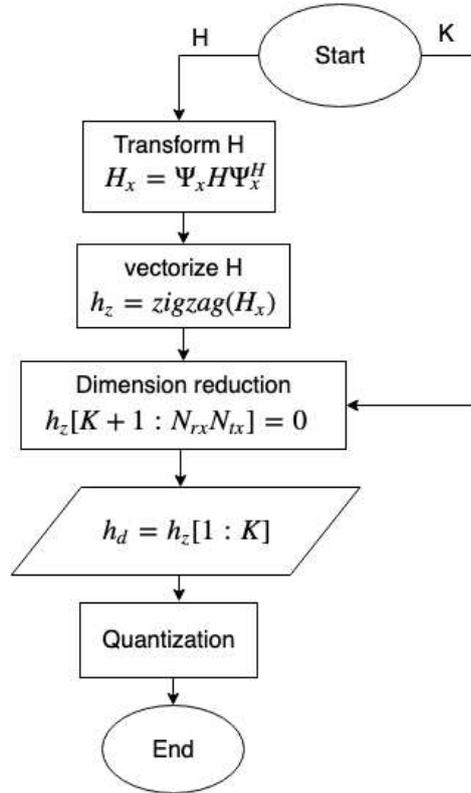
where σ_q^2 is the quantization distortion and σ_d^2 is the dimensionality reduction distortion.

The cumulative sum of the sorted coefficients are computed as

$$C_{sum}(i) = \sum_{i=1}^{N_{tx}N_{rx}} |h_{s,i}|^2 \quad (3.20)$$

where C_{sum} is a vector with the same dimension as h_s , and each index i stores the energy of the coefficients which are located from the first index of h_s up to the current index $i =$

Figure 3.5: Algorithm for implementing DCT or DFT zigzag.



Source: Author.

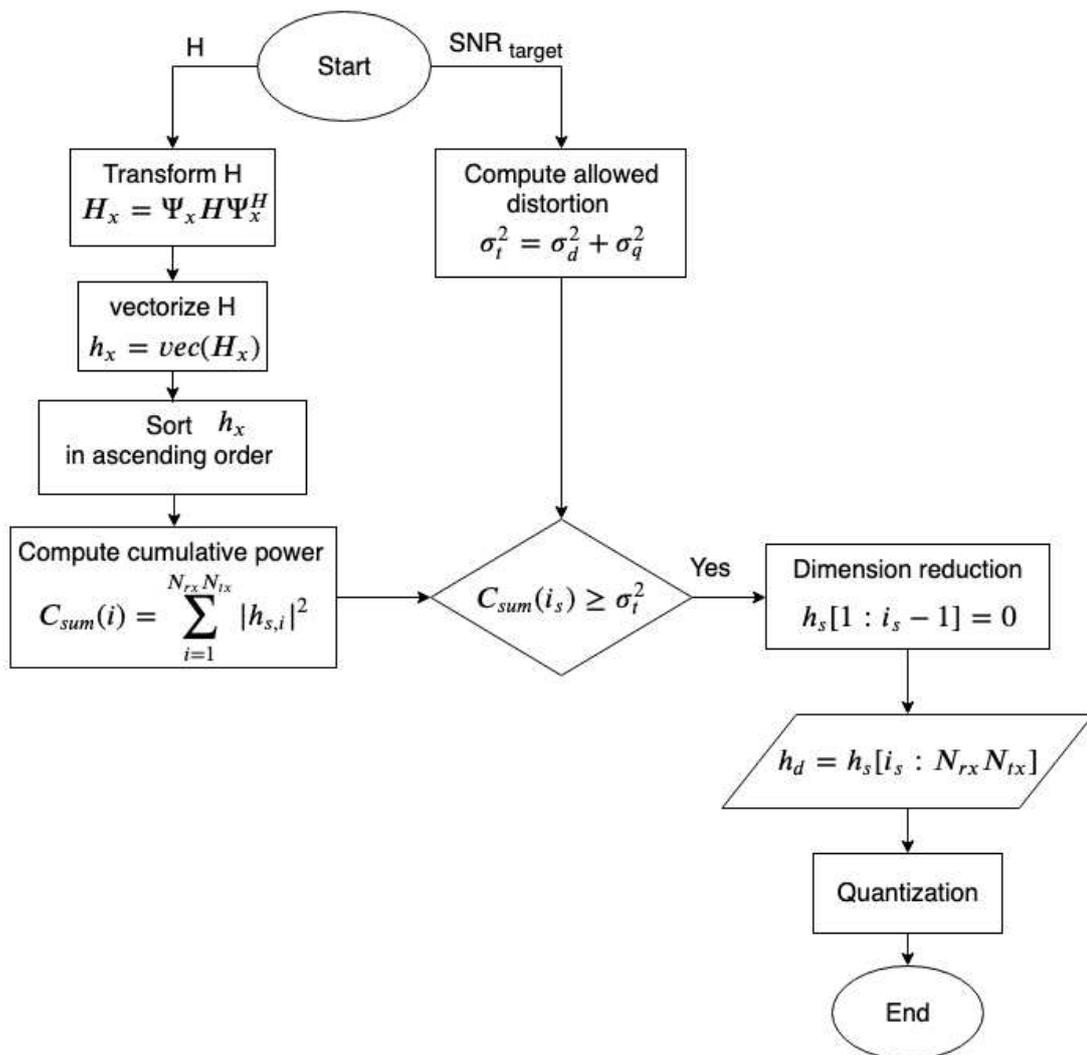
$[1, 2, \dots, N_{tx}N_{rx}]$. Equation (3.19) and Equation (3.20) are combined to set a point for dimensionality reduction which comply with the SNR_t design metric that follows Equation (3.15). To ease computation, quantization is assumed perfect and the total distortion reduces to $\sigma_t^2 = \sigma_d^2$, and $\text{SNR}_t = \frac{\sigma_{h_s}^2}{\sigma_d^2}$. The quantization distortion is included later during the analysis of the quantizer. The discard of coefficients can be made as $h_s(1 : i_s - 1) = 0$ where i_s is the first index that satisfies $C_{sum}(i_s) \geq \sigma_d^2$. From that, the number of complex coefficients which is sent for quantization is $K = N_{rx}N_{tx} - (i_s - 1)$, and an *index set* containing the index location of the K selected coefficients in h_v is sent as side information. Figure 3.6 provides a flowchart for implementing *DCT SNR target* and *DFT SNR target* methods.

For instance, \mathbf{H}_x is a 2×2 complex channel matrix on a transform domain as

$$\mathbf{H}_x = \begin{pmatrix} 0 & 0.1 + 0.2j \\ 4 + 5j & 0.5 - 1j \end{pmatrix} \quad (3.21)$$

which should be compressed with SNR target of 4 dB, assuming perfect quantization. The

Figure 3.6: Algorithm for implementing DCT or DFT SNR target.



Source: Author.

matrix \mathbf{H}_x is vectorized column-wise and normalized by l_2 norm which gives $h_x = [0, 0.6150 + 0.7688j, 0.0154 + 0.0308j, 0.0769 - 0.1538j]$. The coefficients sorted according to their absolute square value in ascending order as $h_s = [0, 0.0012, 0.0296, 0.9693]$. The cumulative sum of each sorted element is computed leading to $C_{sum} = [0, 0.0012, 0.0307, 1.0000]$. From the SNR formula, the power of the channel matrix is 0.2500 mW, and the allowed distortion is 0.0995 mW. Therefore, $i_s = 4$ and the firsts 3 coefficients can be discarded $h_s(1 : 3) = 0$ and just $h_s(4) = [0.6150 + 0.7688j]$ coefficients are to be quantized, $K = 1$. The norm is quantized and sent to the decoder for either SNR target or zigzag methods. As the norm overhead is the same for the evaluated methods, its consideration is skipped in this dissertation.

3.7.3 Encoding few relevant coefficients

As discussed, the relevant coefficients can be encoded via a relevancy map that will require a fixed number of bits. For example, a significance map can identify which \mathbf{h}_s coefficients were not discarded [38]. However, when there are very few relevant coefficients, instead of considering the whole map as in image processing applications, we propose in this dissertation to transmit only the positions and values of these few coefficients as side information for the BS [38]. The proposed method simply indicates the selected coefficients by informing the *index set* to the decoder. In this *index* format, the position of the K selected coefficient components within \mathbf{h}_s is specified using a set of integers \mathcal{I} of cardinality K . Each index requires $\log_2(N_{rx}N_{tx})$ bits. As entropy coding is not considered, sending the index set requires a total amount of $K \log_2(N_{rx}N_{tx})$ bits of side information. Adopting index set allows the decoder to know the original positions of the coefficients. From the previous example, the indexes of h_s are $\mathcal{I} = \{1, 3, 4, 2\}$, and the side information to recover the position of the quantized coefficient is $\mathcal{I}_s = \{4, 2\}$.

3.7.4 Quantization

A scalar quantizer is designed for each real and imaginary parts of a transform channel matrix to avoid operation on overloading region. The number b of bits used to represent the quantized values is varied in order to account for σ_q^2 and find an operational point close to the projected value of total distortion σ_t^2 . The number of quantization levels is $M = 2^b$ and the quantization step is $\Delta = |H_{max} - H_{min}|/M$. The quantizer of the real part of the transform

matrix and the quantizer of the imaginary part are designed with the same amount of bits. Those computed quantizer parameters are stored in a look-up table for further compressions.

The total number of bits used to encode DFT and DCT zigzag methods is

$$b_{z,t} = 2bK + \log_2(N_{\text{rx}}N_{\text{tx}}) \quad (3.22)$$

where K is the side information to allow the decoder to know how many channel coefficients were encoded. The total number of bits used to encode DFT and DCT SNR target methods is

$$b_{s,t} = 2bK + K \log_2(N_{\text{rx}}N_{\text{tx}}) \quad (3.23)$$

where $K \log_2(N_{\text{rx}}N_{\text{tx}})$ is the side information due to the index set \mathcal{I} .

While [26] proposed DCT *zigzag* to compress MIMO CSI using the LBG quantizer, here the DCT *zigzag* method is evaluated using scalar quantizer to reduce its computational complexity. VQ with code-vector search has a computational cost of $\mathcal{O}(N_{\text{rx}}N_{\text{tx}}K)$, while scalar quantization only requires $\mathcal{O}(1)$. The *SNR target* method is proposed in [46] and is evaluated with detail in this dissertation.

Chapter 4

Simulations and results

This chapter presents the results obtained by simulating the methods described in Chapter 3. The MIMO channel modeling was performed in Wireless Insite ray tracing software according to Raymobtime approach [47]. The simulations also used the Kronecker channel model. Further processing and simulations were done in MATLAB to analyze the channel compression methods in MIMO FDD systems.

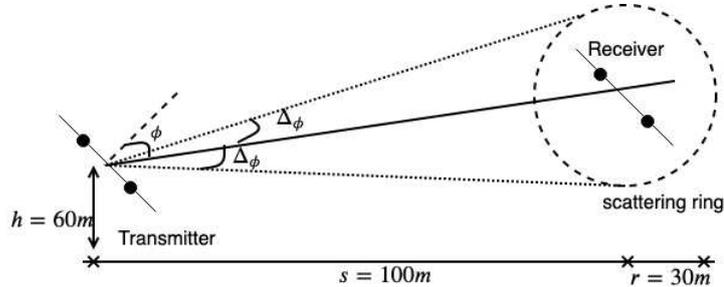
4.1 Acquisition and simulation of channel matrices

In order to analyze the compression methods, two channel models are used to evaluate the compression techniques: the Kronecker channel model and the multipath channel model with Raymobtime dataset. Both channel models are used to simulate an ULA 8×1 at receiver and transmitter. However, the antenna elements spacing is changed as well as the use cases and operational frequencies to analyze the impact of those factors on the compression methods. This section explains how each model was applied and provides some characteristics of the simulated channels.

4.1.1 Simulation of Kronecker channel model

The Kronecker channel model was implemented as described in Section 2.3.2, assuming the parameter values specified in [26]. The distance between transmitter and receiver was set to 100 m, the scattering ring was set at 30 m from the receiver, and the transmitter was at a height of 60 m, as shown in Figure 4.1. The covariance matrices of transmitter \mathbf{R}_{tx} and receiver \mathbf{R}_{rx} are computed as in Equation (2.7).

Figure 4.1: Parameters used to model the covariance matrices at transmitter and receiver of Kronecker channel model.



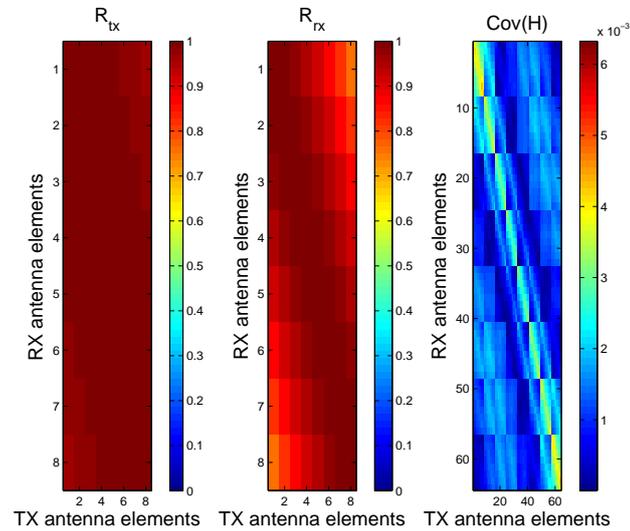
Source: Author.

One thousand realizations of the channel matrix \mathbf{H} are generated from Equation (2.6), and the average covariance matrix of the final channel is computed. In this context, final channel covariance matrices show the tendency in linear relation between antenna elements of transmitter and receiver. For the analyzed cases, an ULA of 8 elements was used at transmitter and receiver side; hence, the channel matrix \mathbf{H} has size 8×8 . The channel covariance matrix is computed from the column-wise vectorization of \mathbf{H} which gives a covariance matrix of size 64×64 . \mathbf{R}_{tx} , \mathbf{R}_{rx} , and the average final channel covariance matrix are presented in Figure 4.2 for 0.1λ antenna spacing, and in Figure 4.3 for 0.5λ antenna spacing. It can be visually noticed that increasing the antenna spacing reduces the amount of pixels with high correlation. This reduction of correlated pixels on final covariance matrix will further influence performance of the compression methods presented in Chapter 3.

4.1.2 Ray tracing (RT) simulation and multipath channel model

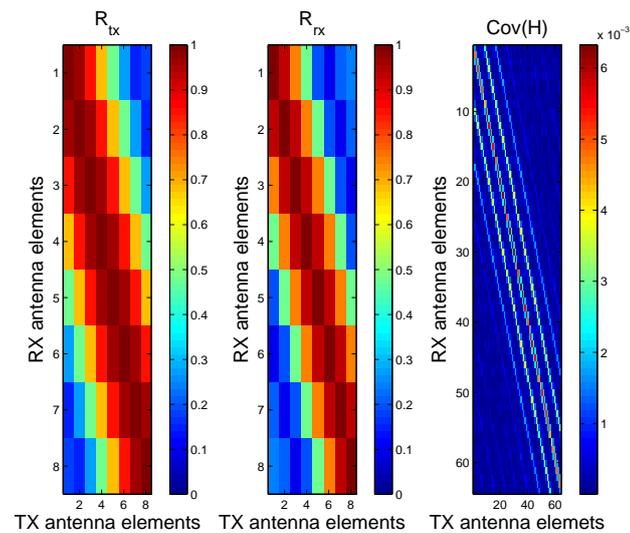
The multipath channel model was adopted as described in Section 2.3.1, employing data from the Raymobtime dataset [47]. The Raymobtime dataset is built by combining ray tracing simulations from REMCOM Wireless Insite software and traffic simulations of vehicles and pedestrians from Simulator for Urban Mobility (SUMO). This combination allows Raymobtime dataset to consider mobility of receivers and/or mobility of the propagation scenario, which are relevant for more realistic simulations of V2E and FW use cases. Figure 4.4 presents the urban canyon environment simulated.

Figure 4.2: Covariance matrices for the Kronecker channel model with antenna element spacing of $d = 0.1\lambda$, distance in between transmitter and receiver of 100 m, scattering ring at receiver of 30 m, and 60 m transmitter height.



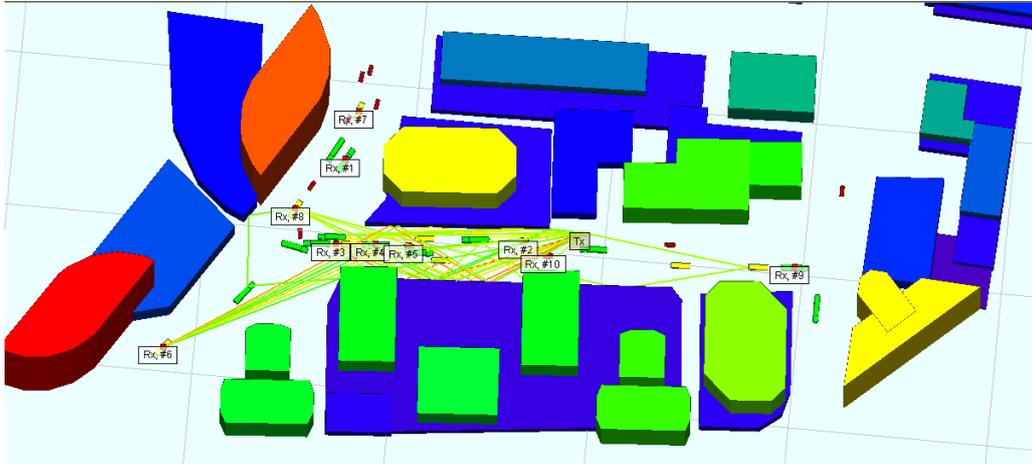
Source: Author.

Figure 4.3: Covariance matrices for the Kronecker channel model with antenna element spacing of $d = 0.5\lambda$, distance in between transmitter and receiver of 100 m, scattering ring at receiver of 30 m, and 60 m transmitter height.



Source: Author.

Figure 4.4: Rosslyn urban canyon scenario at RT software for V2E use case, with receiver antennas on top of vehicles and transmitter antenna placed in a roadside unit.



Source: [48].

The Raymobtime dataset was developed as described in [47]. It is spatially consistent and keeps a history of the time evolution by defining *scenes* and *episodes* in the simulation. A scene is a simulation snapshot in time, each scene is sampled periodically at time $t = nT_{\text{sam}}$ where T_{sam} is the sampling period. Besides, each scene can contain various transmitters and receivers, and depending on the scenario simulated, their location may also change. An episode constitutes a set of N_{sce} scenes in an observation window T_{epi} , where $N_{\text{sce}} = T_{\text{epi}}/T_{\text{sam}}$. After each T_{epi} , there is a random waiting time before a new episode is recorded to improve scene diversity.

Following Raymobtime nomenclature [48], datasets s003, s004, and s006 are used for analyzing the compression methods in FW and V2E. The s004 dataset is a mobile configuration with receivers on top of vehicles and 60 GHz operational frequency. The s003 and s006 are FW configurations with receivers on buildings. In these FW scenarios, the vehicles move around the urban canyon scenario operating on frequencies of 2.8, and 60 GHz respectively, and influence the ray scattering. Table 4.1 specifies the characteristics of the dataset used in this dissertation.

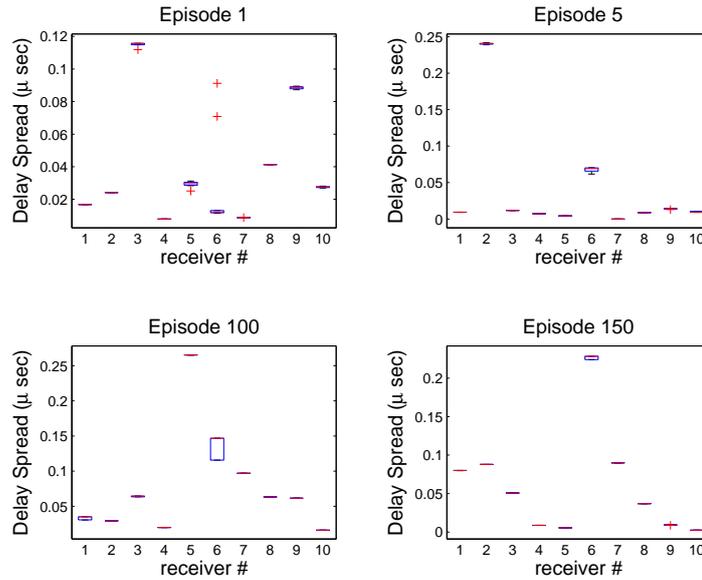
Figure 4.5 presents the delay spread for 4 episodes of the 2.8 GHz FW use case. It can be noticed that the box plot is mostly concentrated around the mean delay value for most of the receivers on the analyzed episodes. This indicates that the channel impulse response changes very little over the scenes within an episode. Figure 4.6 presents a box plot for delay spread of

Table 4.1: Description of the Raymobtime dataset used for compression.

Dataset name	3D scenario	Frequency	Number of receivers and type	T_{sam}	N_{epi}	N_{sce} per episode
s004	Urban canyon	60 GHz	10 Mobile	1 sec	3608	1
s003	Urban canyon	2.8 GHz	10 Fixed	1 ms	200	10
s006	Urban canyon	60 GHz	10 Fixed	1 ms	200	10

Source: [48]

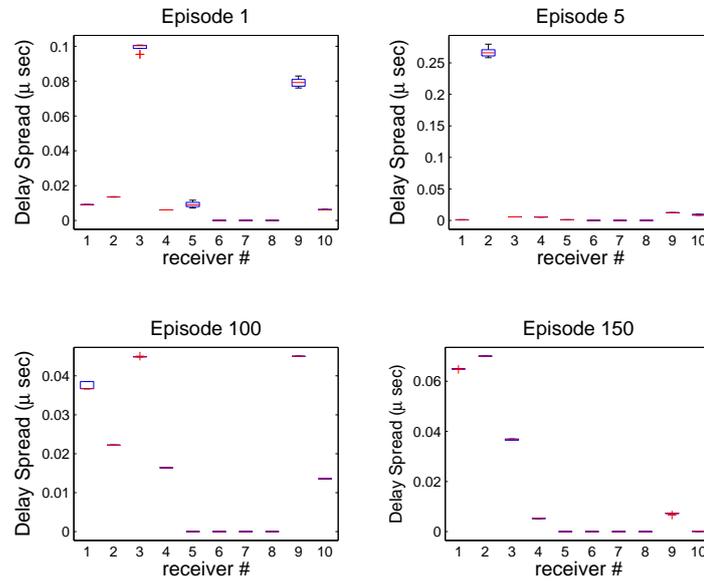
4 episodes of the 60 GHz FW use case. The variability of delay spread from the mean value is also small. Hence, the channel can be said constant over a simulated episode $T_{\text{epi}} = 10$ ms for FW use cases.

Figure 4.5: Delay spread over scenes of RT simulations at 2.8 GHz FW use case for 10 receivers.

Source: Author.

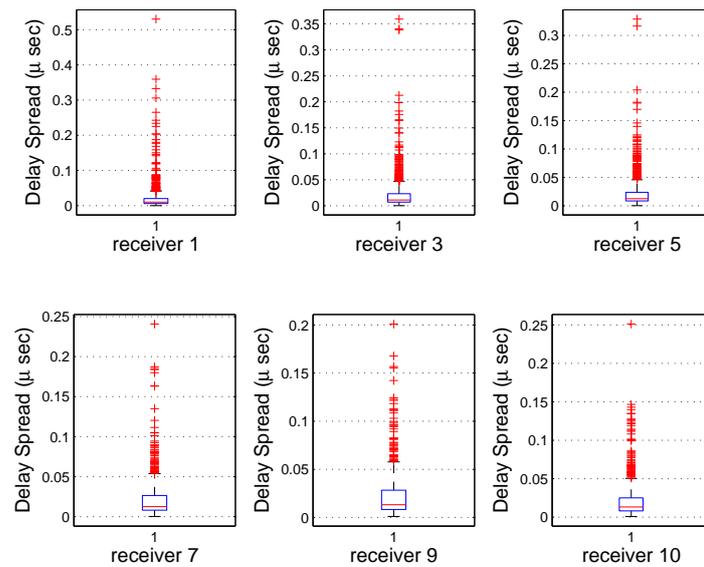
As the dataset for V2E is recorded with only one scene per episode, the delay spread analysis is conducted differently. Figure 4.7 presents the delay spread for some of the receives where the delay spread was computed over all valid episodes. Valid episodes, or channels, are the ones in which the receiver is within transmitter coverage. The increase in delay spread was expected because of receiver mobility as well as a longer episode duration of $T_{\text{epi}} = 1$ second.

Figure 4.6: Delay spread over scenes of RT simulations at 60 GHz FW use case for 10 receivers.



Source: Author.

Figure 4.7: Delay spread for a receiver over different episodes for V2E use case simulated on RT.



Source: Author.

Indoor scenarios were not simulated. However, sparsity can still be leveraged mainly if operating in mmWaves bands [22]. Note that the covariance matrices \mathbf{R} from the Kronecker channel model are not computed here because RT is a deterministic simulation approach.

4.2 Analysis of channel model and orthogonal transform

This section provides insights about the energy compaction of each orthogonal transform presented in Section 3.7 of this dissertation. Besides, the effectiveness of the matrix vectorization methods employed, *zigzag* and *sort*, is also analyzed.

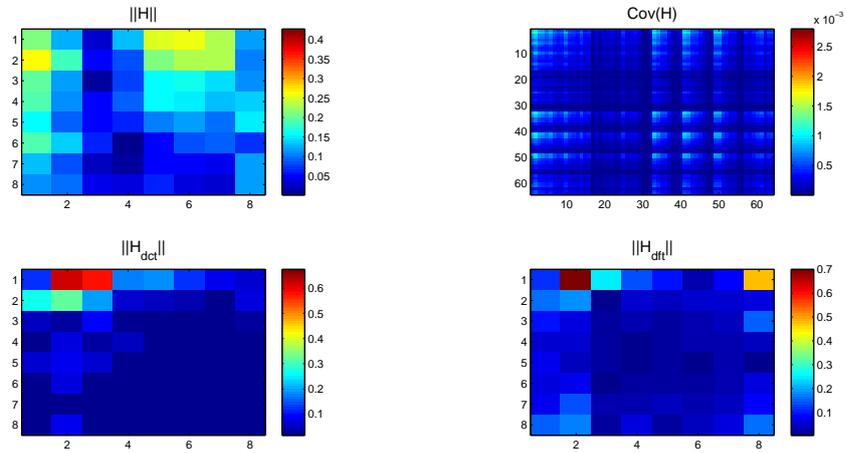
4.2.1 Transform analysis of Kronecker channel model

Figure 4.8 presents a sample of a channel matrix \mathbf{H} generated from the Kronecker channel model with 0.1λ antenna spacing as well as the covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . It can be noticed that the pixels in \mathbf{H} are more correlated on the top part of the channel matrix. This intuition is proved at the covariance matrix plot. Both DCT and DFT happen to compress \mathbf{H} . However, the coefficients with most energy are located differently in each transform matrix.

Figure 4.9 shows a sample \mathbf{H} matrix from the Kronecker channel dataset with 0.5λ antenna spacing, its covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} are also plotted. \mathbf{H} seems more random, less correlated. This lack of correlation is confirmed by its covariance matrix, which shows very few points in red (more correlated). DCT and DFT provide energy compaction. However, visually, more pixels are relevant if compared to the 0.1λ antenna spacing sample of Figure 4.8. Therefore, less compression is expected in this case.

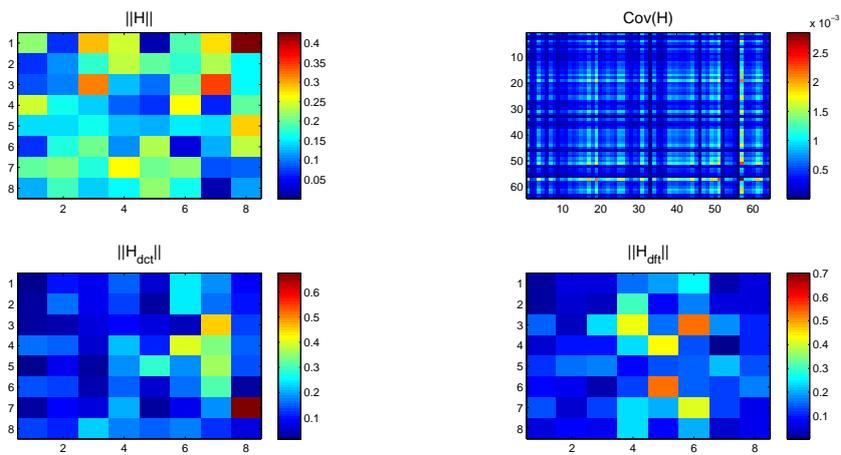
To better observe the statistical behavior of channel matrices \mathbf{H} and their DCT and DFT transforms, their average values and standard deviation are plotted using column-wise vectorization, except for the *zigzag* case. Figures 4.10 and 4.11 show the energy compaction of applying DCT and DFT to the Kronecker dataset with 0.1λ and 0.5λ antenna spacing, respectively. It can be noticed that DCT with *zigzag* provides good compaction of data when antenna spacing is 0.1λ ; however, this energy compaction does not hold when the antenna spacing increases. As expected, sorting the transform coefficients leads to a well-behaved dynamic range and, consequently, a better compaction trend for both antenna spacing scenarios and transforms applied.

Figure 4.8: Plot of a sample channel matrix, its covariance, and its transforms for Kronecker channel model with antenna element spacing of $d = 0.1\lambda$.



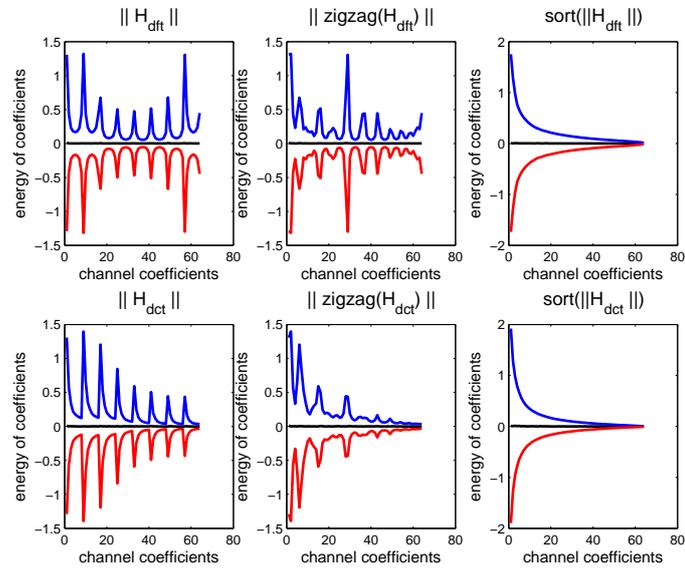
Source: Author.

Figure 4.9: Plot of a sample channel matrix, its covariance, and its transforms for Kronecker channel model with antenna element spacing of $d = 0.5\lambda$.



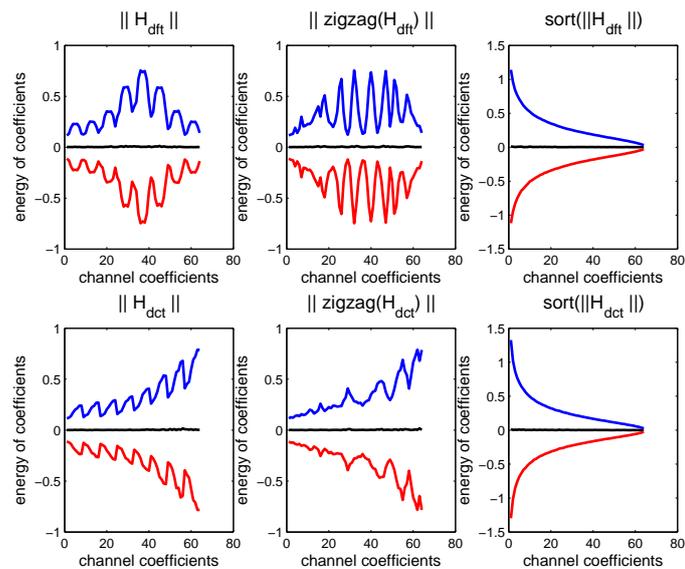
Source: Author.

Figure 4.10: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Kronecker channel model with antenna element spacing of $d = 0.1\lambda$.



Source: Author.

Figure 4.11: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Kronecker channel model with antenna element spacing of $d = 0.5\lambda$.

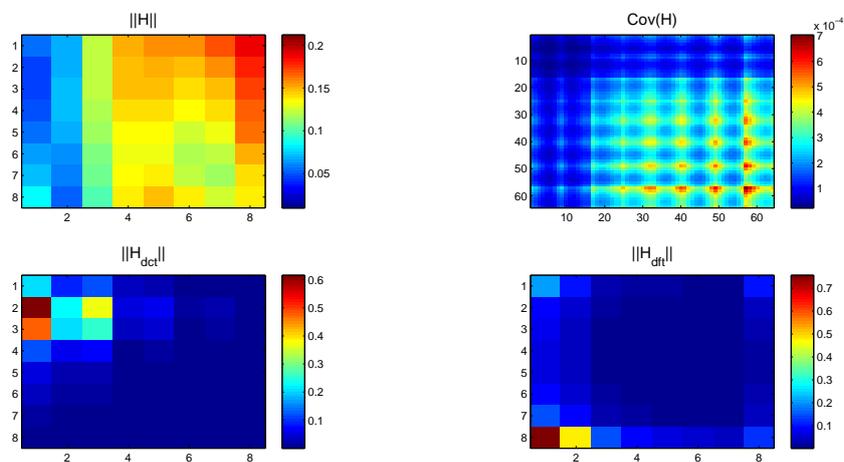


Source: Author.

4.2.2 Transform analysis of RT dataset at sub-6 GHz

Figure 4.12 presents a sample of a channel matrix \mathbf{H} generated from the RT 2.8 GHz FW dataset with 0.1λ antenna spacing as well as the covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . It can be noticed that there is few energy on the first two columns of \mathbf{H} which means that the two first antenna elements of the ULA 8×1 transmitter contributes very little to the received signal at any antenna element. The covariance matrix is computed from the column-wise vectorization of the channel matrix \mathbf{H} . Visually, one can notice that the first 16 entries of rows and columns of this sample covariance matrix have very few correlation values which is consistent with the observation of \mathbf{H} . Both DCT and DFT happen to compress \mathbf{H} . However, the coefficients with most energy are located differently in each transform matrix. In this sample example, the DCT transform matrix groups the significant coefficients on top-left portion of its matrix and the DFT transform matrix groups the relevant coefficients on its bottom-left corner.

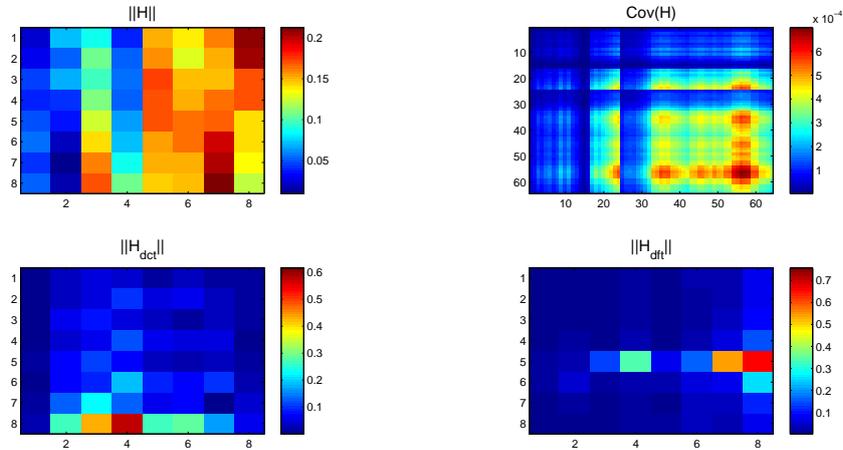
Figure 4.12: Plot of a sample channel matrix, its covariance, and its transforms for 2.8 GHz FW Ray-mobtime dataset with antenna element spacing of $d = 0.1\lambda$.



Source: Author.

Figure 4.13 resents a sample of a channel matrix \mathbf{H} generated from the RT 2.8 GHz FW dataset with 0.5λ antenna spacing as well as the covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . It can be noticed that there is more energy on the right part of \mathbf{H} . The last 2 columns of \mathbf{H} have the most energy (tend to red color), and there are 3 columns with almost

Figure 4.13: Plot of a sample channel matrix, its covariance, and its transforms for 2.8 GHz FW Ray-mobtime dataset with antenna element spacing of $d = 0.5\lambda$.

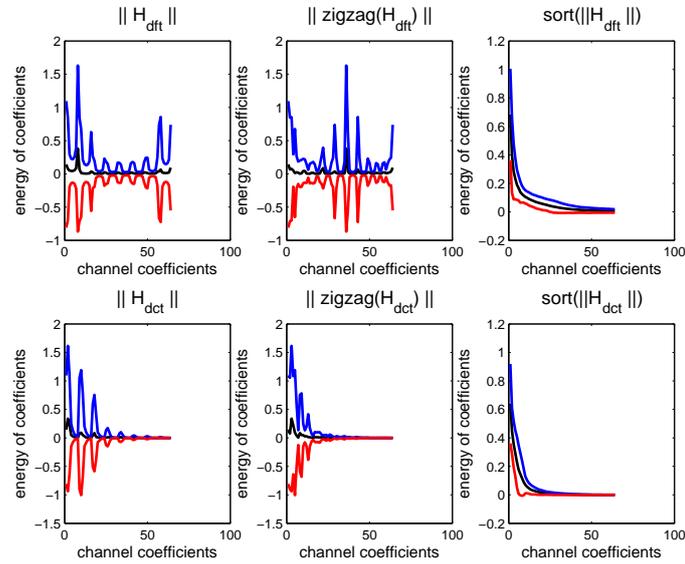


Source: Author.

no energy (tend to blue color). The covariance matrix is consistent with the observation of \mathbf{H} , the first 16 entries have low correlation and the correlation is more accentuated on the bottom-right of the covariance matrix. DCT and DFT compresses \mathbf{H} and concentrates the coefficients with most energy on different locations of the respective transform matrix. Figure 4.12 and Figure 4.13 are for the same sample channel matrix changing only the antenna spacing. It can be noticed that the location of the relevant transform coefficients varies considerably when the correlation of antenna elements decreases, due to the increase on distance between antenna elements.

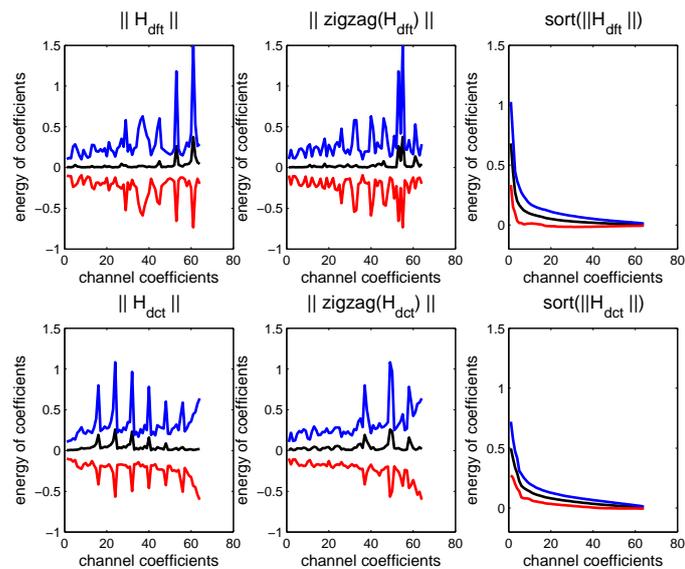
Following, the statistical behavior of channel matrices \mathbf{H} , and their DCT and DFT transforms are observed by plotting their average values and standard deviation using column-wise vectorization, except for the zigzag case. Figures 4.14 and 4.15 present the energy compaction for applying DCT and DFT to the 2.8 GHz FW Raymobtime dataset with 0.1λ and 0.5λ antenna spacing, respectively, obtained by post-processing according to the multipath channel model in Equation (2.5). Despite the different channel model, DCT with *zigzag* in 0.1λ antenna spacing scenario still presents a good energy compaction characteristic. Nevertheless, for 0.5λ antenna spacing, sorting the channel transform coefficients is the best way among the adopted alternatives to exploit energy compaction effectively for both DCT and DFT.

Figure 4.14: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Raymobtime 2.8 GHz fixed environment with antenna element spacing of $d = 0.1\lambda$.



Source: Author.

Figure 4.15: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Raymobtime 2.8 GHz fixed environment with antenna element spacing of $d = 0.5\lambda$.



Source: Author.

4.2.3 Transform analysis of RT dataset at mmWaves

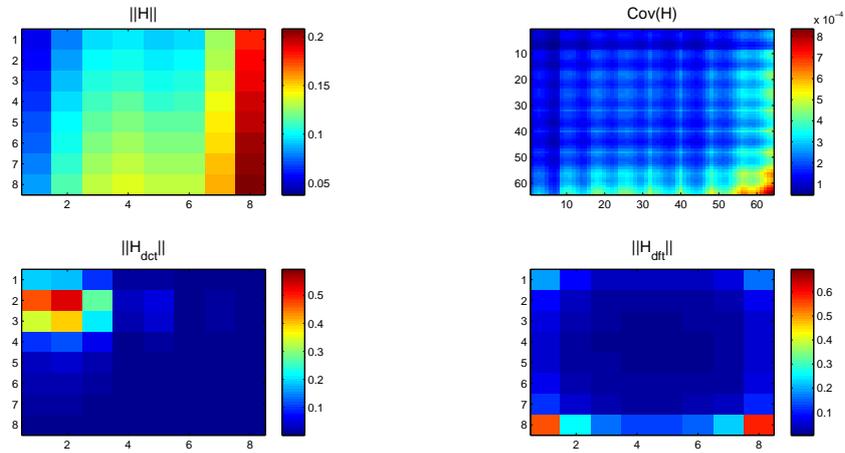
Figure 4.16 presents a sample of a channel matrix \mathbf{H} generated from the RT 60 GHz FW dataset with 0.1λ antenna spacing as well as its covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . From observation of the channel matrix, the last column of \mathbf{H} is the one with most energy. This can be confirmed at the covariance matrix which shows more correlation on its bottom-right corner. Application of DCT transform leads to concentration of most energy relevant coefficients on the top-left portion of \mathbf{H}_{DCT} . Applying DFT transform leads to spread of most relevant coefficients on the corners of \mathbf{H}_{DFT} .

Figure 4.17 shows a sample of a channel matrix \mathbf{H} generated from the RT 60 GHz FW dataset with 0.5λ antenna spacing as well as its covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . From the \mathbf{H} matrix, it can be noticed that the transmitter antenna elements 5 and 7 are the ones which send most energy to all antenna elements of the receiver. The covariance matrix confirms the high correlation among those antenna elements by plotting large pixels in red. DCT and DFT compresses \mathbf{H} and concentrates the coefficients with most energy on different locations of their respective transform matrix. Figure 4.16 and Figure 4.17 plot the same sample channel matrix changing only the antenna spacing used in post-processing. It can be noticed that the location of the relevant transform coefficients varies considerably when the correlation of antenna elements decreases. Despite the difference in frequency, the location of relevant coefficients for antenna spacing of 0.1λ and 0.5λ have a localization trend.

To better observe the statistical behavior of \mathbf{H} , \mathbf{H}_{DCT} and \mathbf{H}_{DFT} , their average values and standard deviation are plotted using column-wise vectorization, except for the zigzag case. Figures 4.18 and 4.19 present the energy compaction for applying DCT and DFT to the 60 GHz FW Raymobtime dataset with post processing of multipath channel model with 0.1λ and 0.5λ antenna spacing respectively. The standard deviation has very different behavior if compared to Kronecker channel model; however, DCT with *zigzag* for 0.1λ antenna spacing still presents a good energy compaction characteristic. Nevertheless, for 0.5λ antenna spacing, just *sort* achieves good energy compaction for both DCT and DFT.

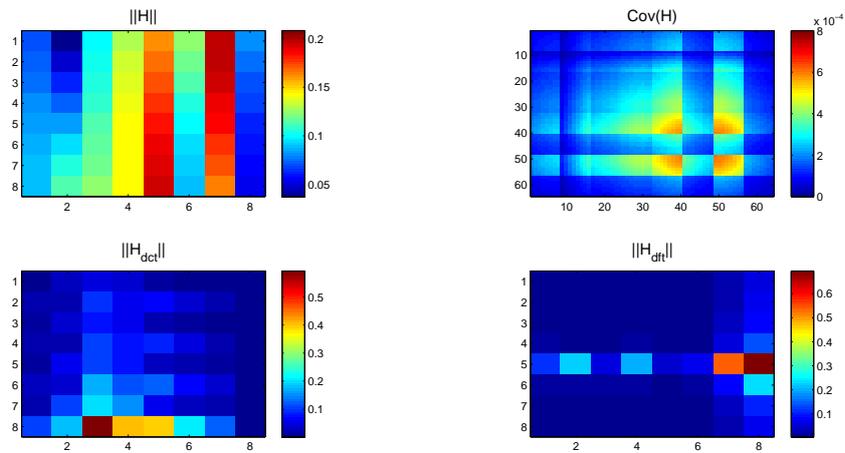
Figure 4.20 presents a sample of a channel matrix \mathbf{H} generated from the RT 60 GHz V2E dataset with 0.1λ antenna spacing as well as its covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . The \mathbf{H} components with most energy are on the top-right to bottom-left diagonal. At the covariance matrix, this diagonal generates high correlation points at every 8 elements spacing, being them more pronounced at top-left part of the covariance matrix because

Figure 4.16: Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz FW Ray-mobtime dataset with antenna element spacing of $d = 0.1\lambda$.



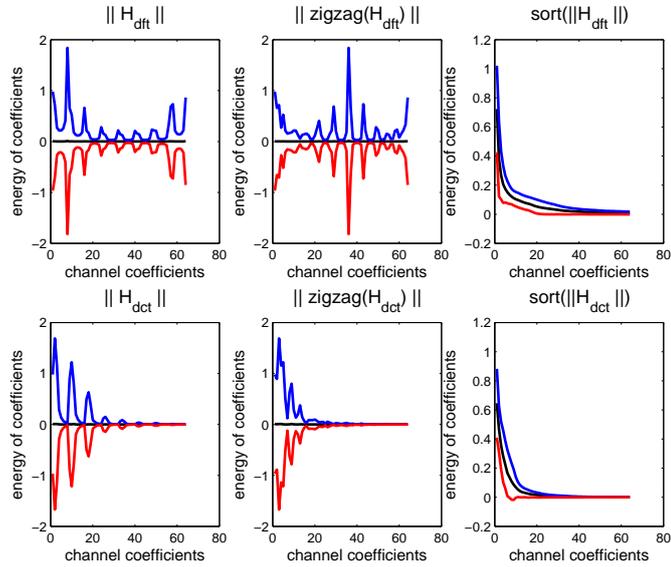
Source: Author.

Figure 4.17: Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz FW Ray-mobtime dataset with antenna element spacing of $d = 0.5\lambda$.



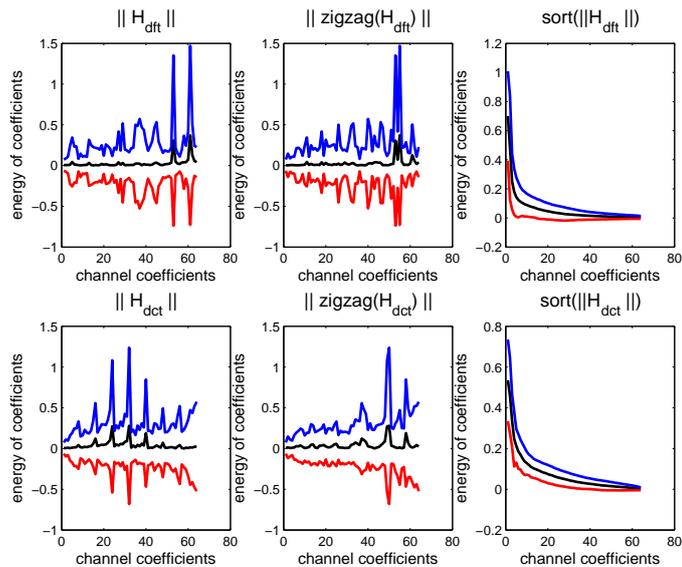
Source: Author.

Figure 4.18: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Raymobtime 60 GHz fixed environment with antenna element spacing of $d = 0.1\lambda$.



Source: Author.

Figure 4.19: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Raymobtime 60 GHz fixed environment with antenna element spacing of $d = 0.5\lambda$.



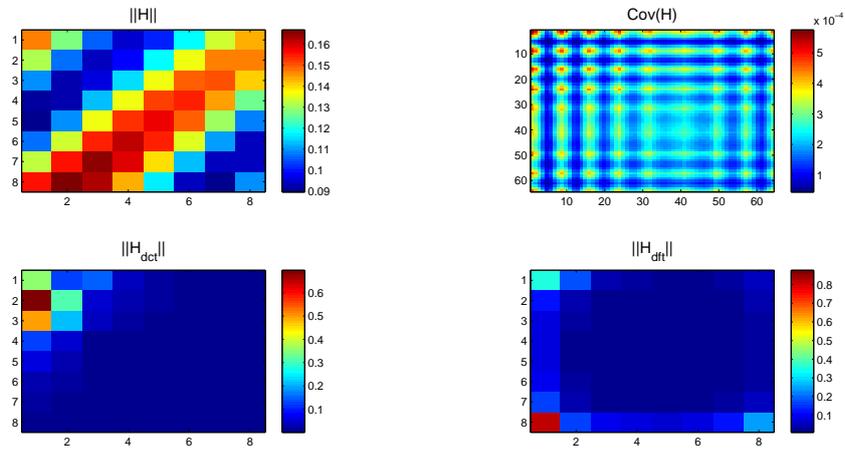
Source: Author.

\mathbf{H} has more energy at its bottom-left part. DCT and DFT compresses \mathbf{H} and concentrates the more energy relevant coefficients on different locations of their respective transform matrix.

Figure 4.21 presents a sample of a channel matrix \mathbf{H} generated from the RT 60 GHz V2E dataset with 0.5λ antenna spacing as well as its covariance matrix, and its transformed matrices \mathbf{H}_{DCT} and \mathbf{H}_{DFT} . For this case, \mathbf{H} has no defined region of high energy; however, all receiver antenna elements receive signals from at least one transmitter element with high energy except for the receiver antenna element 5. The covariance matrix has very few rows and columns with low correlation (blue color), some points with high correlation (red color), and many middle correlation areas (green color). These observations of the covariance matrix are compatible with the observation of \mathbf{H} . DCT and DFT compress the \mathbf{H} matrix to very few relevant coefficients which are located differently within each transform matrix. This example is the one with most channel matrix energy variability. This phenomenon can be physically explained by mobility of the receiver and the larger distance among antenna elements.

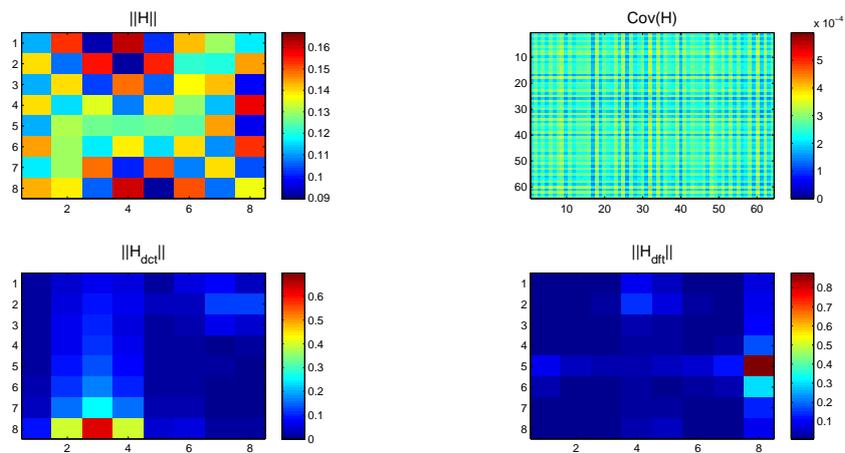
To better observe the statistical behavior of \mathbf{H} , \mathbf{H}_{DCT} and \mathbf{H}_{DFT} , their average values and standard deviation are plotted using column-wise vectorization, except for the *zigzag* case. Figures 4.22 and 4.23 present the energy compaction for applying DCT and DFT to the 60 GHz V2E Raymobtime dataset with post processing of multipath channel model with 0.1λ and 0.5λ antenna spacing respectively. Despite the receiver mobility DCT with *zigzag* for 0.1λ antenna spacing arrives to put together the coefficients with most standard deviation. However, the energy compaction of DCT with *zigzag* is not perceived when antenna spacing is set to 0.5λ . For 0.5λ antenna spacing, just *sort* of coefficients matrix in decreasing order achieves good energy compaction for both DCT and DFT.

Figure 4.20: Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz V2E Ray-mobtime dataset with antenna element spacing of $d = 0.1\lambda$.



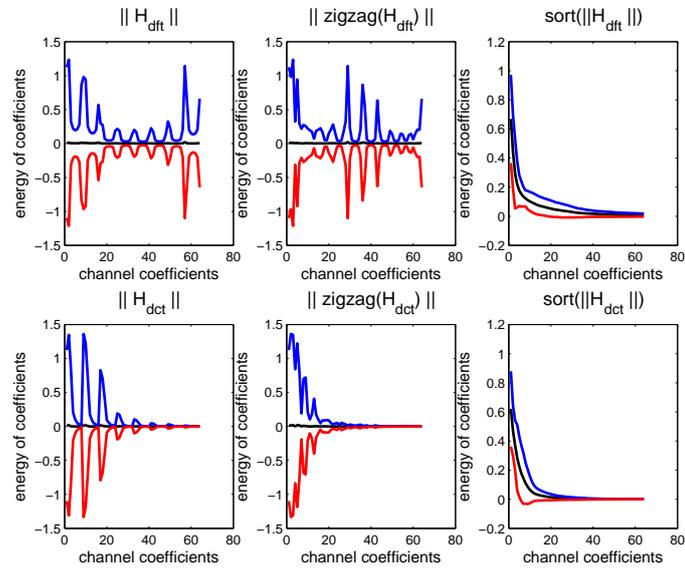
Source: Author.

Figure 4.21: Plot of a sample channel matrix, its covariance, and its transforms for 60 GHz V2E Ray-mobtime dataset with antenna element spacing of $d = 0.5\lambda$.



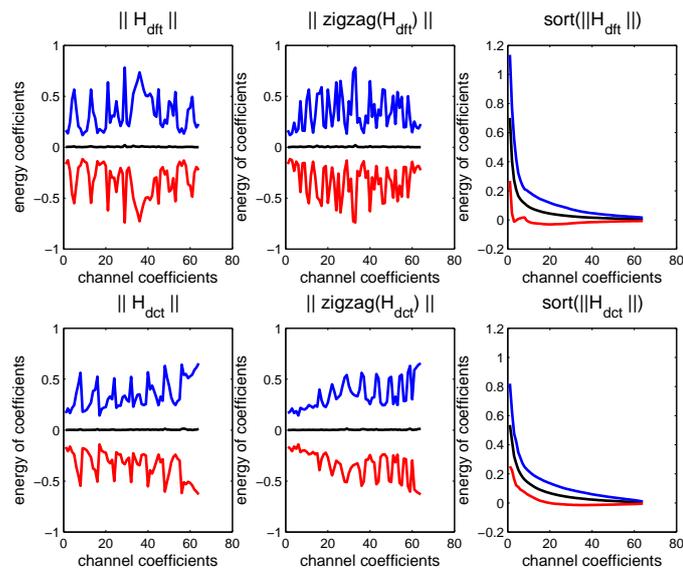
Source: Author.

Figure 4.22: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Raymobtime 60 GHz mobile environment with antenna element spacing of $d = 0.1\lambda$.



Source: Author.

Figure 4.23: Channel coefficients evaluation for DCT and DFT with *zigzag* and *sort* for Raymobtime 60 GHz mobile environment with antenna element spacing of $d = 0.5\lambda$.



Source: Author.

4.3 Evaluation of compression metrics

This section presents the simulation results for the various channel models described previously, Section 4.1. The results are analyzed in regard to SNR, and compression ratio where the design metric is set to 6 dB of SNR with the least amount of bits per complex channel coefficient, considering side information. In order to compare the performance of *zigzag* methods with *SNR target* methods, the simulation starts by computing SNR target method to define the number of coefficients K which must be quantized to achieve the desired SNR. After, the computed K is used as input to the *zigzag* methods as the number of coefficients which should be coded. Therefore, the methods employing *zigzag* and *SNR target* are compressing the same number of transform coefficients; however, these coefficients do not contain the same amount of energy as it is seen henceforth.

4.3.1 Compression ratio

Table 4.2 lists the mean compression ratio η , which does not take quantization in account, achieved with the different channel datasets, varying transform and antenna spacing applied. For the Kronecker channel model, the compression ratio varies considerably when the antenna spacing increases from 0.1λ to 0.5λ ; such effect can be explained by the reduction on channel correlation shown in Figures 4.2 and 4.3. Besides, DCT becomes the best choice for both antenna spacing, in this case, because of the covariance structure of Kronecker channel model. For the Raymobtime dataset, DCT also allows better compression when the antenna spacing is 0.1λ ; however, when the channel spacing increases to 0.5λ , DFT has better compression ratio for either FW or V2E use cases. Moreover, the compression ratio achieved by DFT is quite stable to changes on antenna spacing, while DCT suffers around 6% reduction on its compression performance when antenna spacing increases from 0.1λ to 0.5λ . DCT reduces performance at 0.5λ because more coefficients are needed to meet the SNR desired. The stability on DFT compression performance can also be an effect of the steering vectors used to model linear arrays in multipath channel model because the steering structure is similar to a DFT matrix.

4.3.2 Signal to noise ratio and compression factor performance

In the following paragraphs, the SNR is analyzed in regard to the amount of bits employed to code each complex channel coefficient and its side information. Besides, the compression

Table 4.2: Achieved compression ratio.

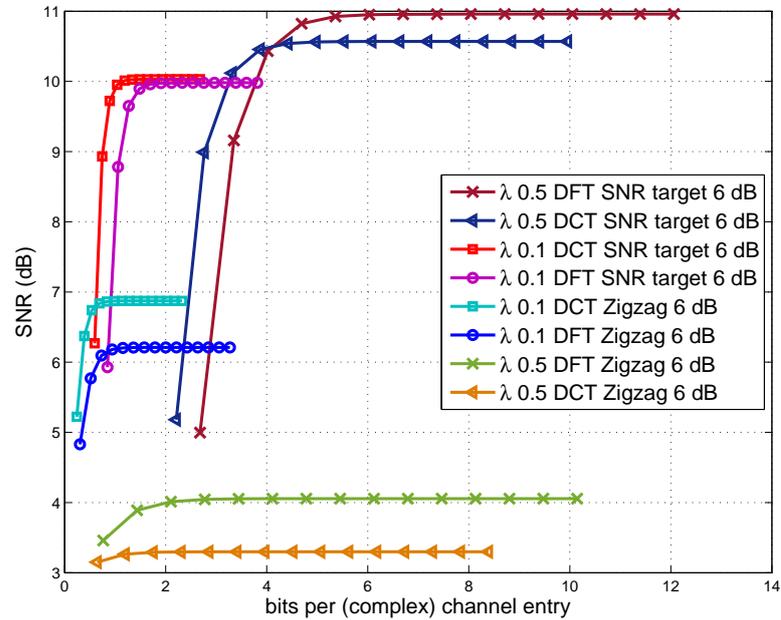
Channel model	Antenna spacing	Transform	Compression ratio
Kronecker	0.1	DCT	92.56%
		DFT	89.41%
	0.5	DCT	72.35%
		DFT	66.50%
2.8 GHz FW	0.1	DCT	94.34%
		DFT	93.56%
	0.5	DCT	86.02%
		DFT	93.67%
60 GHz FW	0.1	DCT	94.73%
		DFT	94.18%
	0.5	DCT	89.38%
		DFT	95.20%
60 GHz Mobile	0.1	DCT	94.40%
		DFT	94.02%
	0.5	DCT	88.71%
		DFT	94.35%

Source: Author.

factor CF is analyzed in this section for each amount of bits at quantizers. The SNR analysis considers quantization error, being it computed after all blocks of Figure 3.4.

Figure 4.24 presents the achieved SNR when considering the Kronecker channel model which curves are consistent with the values in Table 4.2. Although the desirable SNR is achieved for DCT zigzag when dataset has 0.1λ antenna spacing, an increase in antenna spacing heavily reduces the efficiency of the compression method. Figure 4.25 shows the compression factor for each tested amount of bits at the quantizers when using Kronecker channel model. As CF just takes in account the number of bits, zigzag compression methods were expected to perform better. It is noticed that DCT zigzag provides the higher compression factor when $d_\lambda = 0.1$; however, its performance decreases by more than half when $d_\lambda = 0.5$. This decrease in compression factor when increasing antenna spacing is noticed in all compression metrics of Figure 4.25. Despite DCT zigzag seems the best method in terms of compression factor, Fig-

Figure 4.24: SNR results for Kronecker channel model with target SNR of 6 dB.



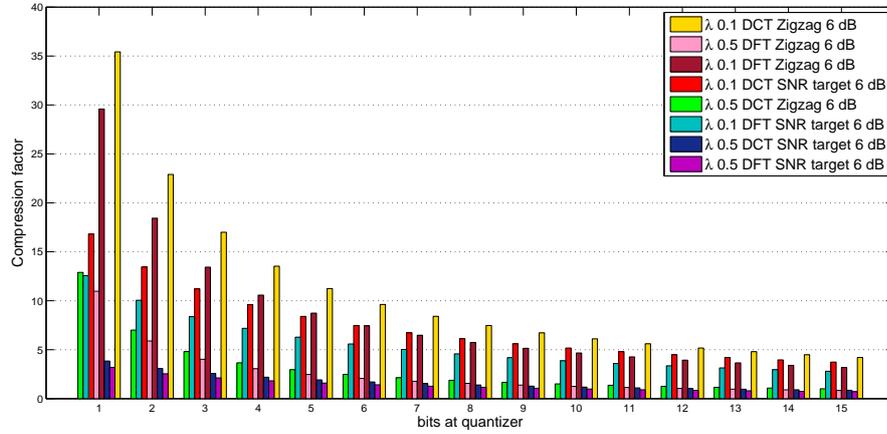
Source: Author.

Figure 4.24 shows this method does not have good reconstruction performance. Considering the design metric of achieving 6 dB of SNR with the least amount of bits, the SNR target method employing DCT would be chosen because its performance requires the least amount of bits and does not go below the SNR target if the antenna spacing increases.

Figures 4.26, 4.28, and 4.30 present the SNR results for the Raymobtime dataset. Figure 4.26 shows the results for the FW use case operating at 2.8 GHz. It can be noticed that the antenna spacing has less influence on this dataset. Figure 4.27 shows again that the 3 methods with higher CF do not achieve the SNR desired. Hence, the chosen transform coding for this case is the SNR target employing DFT because it uses the least amount of bits and provides good performance in both antenna spacing scenarios. Even though SNR target with DCT achieves higher performance in 0.1λ antenna spacing, it requires more bits to have the same performance as SNR target employing DFT at 0.5λ antenna spacing. From Figure 4.27, it can be noticed that DFT SNR target method has higher compression factor when $d_\lambda = 0.5$ than for $d_\lambda = 0.1$, this shift does not happen in Kronecker channel model dataset.

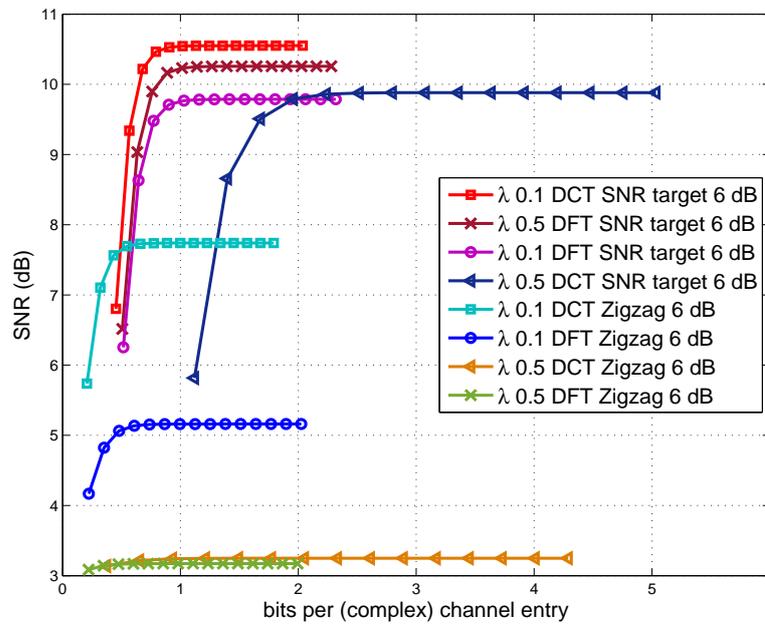
SNR results for FW use case operating at 60 GHz are shown in Figure 4.28, and CF

Figure 4.25: Compression factor for Kronecker channel model with target SNR of 6 dB.



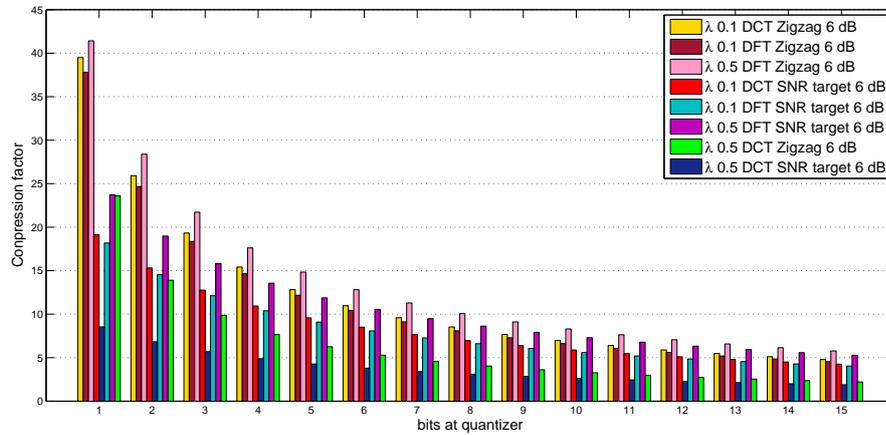
Source: Author.

Figure 4.26: SNR results for Raymobtime 2.8 GHz FW dataset with target SNR of 6 dB.



Source: Author.

Figure 4.27: Compression factor considering quantization for Raymobtime 2.8 GHz FW dataset with target SNR of 6 dB.

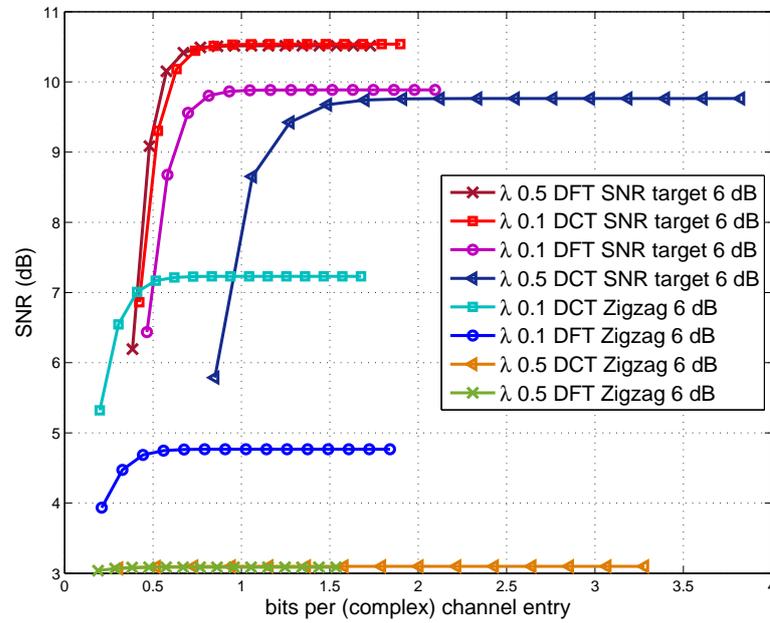


Source: Author.

results are presented in Figure 4.29. Jointly analyzing both graphics, SNR target with DFT also provides a good trade off between amount of bits needed and achieved SNR for both antenna spacing scenarios. SNR target with DCT is not chosen because its SNR performance at 0.5λ degrades and requires more bits. Figure 4.30 presents the SNR results and Figure 4.31 presents the CF results for V2E use case operating at 60 GHz. From Figure 4.30, SNR target employing DFT at 0.5λ antenna spacing has superior SNR performance. Figure 4.31 shows there is little loss in CF performance when the antenna spacing increases from 0.1λ to 0.5λ . In this scenario, SNR target with DFT is also the chosen method for transform coding given its higher SNR performance and reasonable amount of bits needed in both antenna spacing scenario.

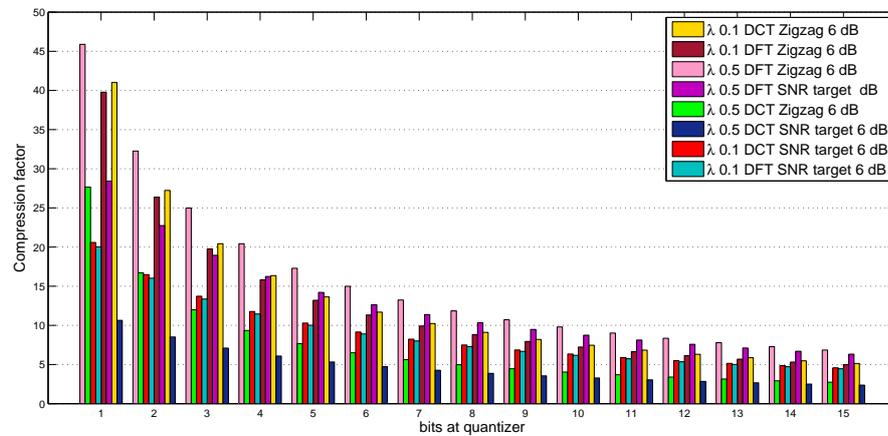
The saturation seen in Figures 4.24, 4.26, 4.28, and 4.30 is an effect of the dimensionality reduction. Due to the limitation on the amount of transform coefficients quantized, there is a point where an increase in the number of bits at the quantizer does not improve the compression performance. Besides, the shift to the right in all curves are due to the amount of side information that each method incurs on the number of bits needed.

Figure 4.28: SNR results for Raymobtime 60 GHz FW dataset with target SNR of 6 dB.



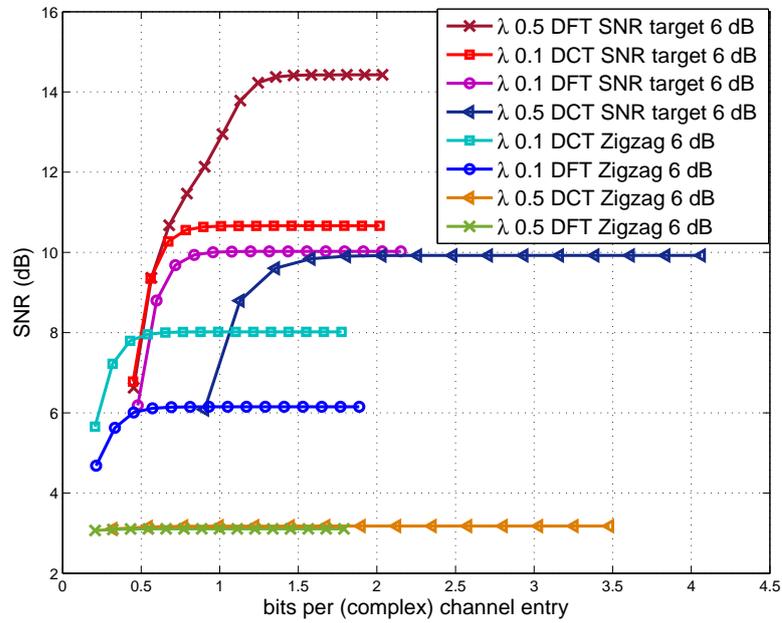
Source: Author.

Figure 4.29: Compression factor for Raymobtime 60 GHz FW dataset with target SNR of 6 dB.



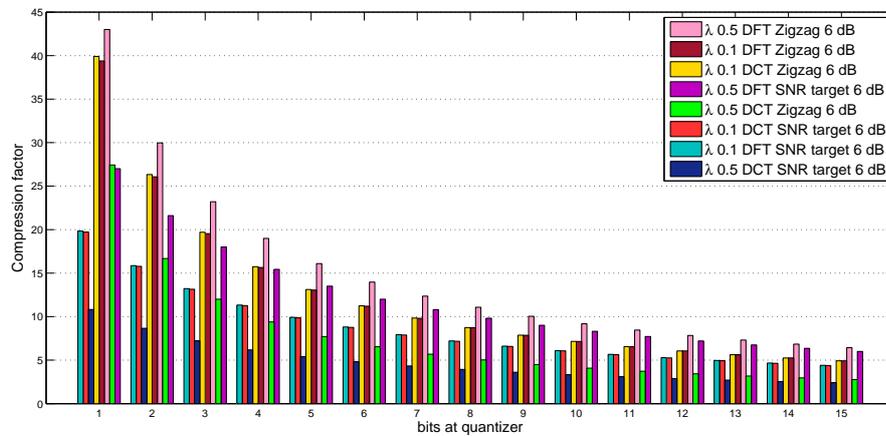
Source: Author.

Figure 4.30: SNR results for Raymobtime 60 GHz V2E dataset with target SNR of 6 dB.



Source: Author.

Figure 4.31: Compression factor for Raymobtime 60 GHz V2E dataset with target SNR of 6 dB.



Source: Author.

Chapter 5

Conclusion

Massive MIMO is mostly related with operation in TDD mode due to the needless use of CSI feedback when relying on channel reciprocity. Nonetheless, massive MIMO should also operate on FDD mode due to available spectrum and backward compatibility of 5G with previous mobile generations. FDD, however, cannot rely on channel reciprocity and the increasing number of antennas at BS and/or UE makes the feedback of CSI impractical as most radio resources would be used for signaling. Due to its practical interest, this dissertation has investigated compressing CSI in MIMO FDD systems.

The first contribution of this dissertation is an heuristic transform coding method that leads to good results and has a relatively low computational cost. This dissertation provides a comparison of some transform coding methods for MIMO FDD channel matrix compression on feedback channel. The DFT SNR target method has shown superior performance in almost all scenarios analyzed in terms of achieved SNR, number of bits, and amount of side information. Therefore, the DFT SNR target method is flexible, operating well for use cases as diverse as FW and V2E.

Another contribution of this dissertation is its evaluation methodology, which allows a more comprehensive comparison of state-of-art methods. It uses compression ratio or desirable SNR as lossy compression design metric. In the literature, one can find papers such as [26] that used relatively simple evaluations and datasets which may lead to biased conclusions. Here, realistic RT simulation datasets were used to evaluate and compare compression performance. For example, this work highlighted that the antenna spacing significantly impacts the compression results. Besides, changing the use case in which compression is employed also impacts results.

5.1 Future works

Future extension of the work developed in this dissertation shall include taking time correlation into account. From the delay spread box plot graphics, the FW use case has very small time variability over a scene, which can be further explored to reduce the frequency within CSI is sent over the feedback channel. Another option is to get inspiration on video coding and perform intraframe and interframe compression, where intraframe compress an entire channel matrix, as seen in this dissertation, and an interframe compression just informs the relevant changes in the channel matrix within a coherence interval which is related to “tracking” the channel.

The inclusion of entropy coding to further compress the side information of the transform code method can be investigated. Besides, CS can also be compared to the SNR target transform coding method proposed in this dissertation. In addition, the transform coding method presented in this dissertation could be compared with deep-learning techniques such as the autoencoder in [49]. One can evaluate the computational costs of choosing deep-learning techniques together with the compression ratio it can provide.

5.2 Published Articles

1. B. V. Boas, N. Fonseca, A. Klautau and N. González-Prelcic, “**Low Complexity Transform Coding for Millimeter Wave MIMO CSI Compression**”, 2018 52nd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2018, pp. 1582-1586. doi: 10.1109/ACSSC.2018.8645142.
2. N. Fonseca, B. V. Boas, N. González-Prelcic, and A. Klautau. “**Channel Matrix Transform Coding for Reduced Feedback in mmWave MIMO Systems**”, MOMAG 2018 - 18th Brazilian Symposium of Microwaves and Optoelectronic (SBMO) e 13th Brazilian Congress of Electromagnetism (CBMAG), Santa Rita do Sapucaí, Minas Gerais, Brazil, 2018.

Bibliography

- [1] “Release 15: TR 21.915.” [Online]. Available: <https://www.3gpp.org/release-15>
- [2] “The European 5G Observatory, 5G Trials,” April 2019. [Online]. Available: <https://5gobservatory.eu/5g-trial/major-european-5g-trials-and-pilots/>
- [3] Nokia, “5G use cases and requirements,” Nokia, Finland, Tech. Rep., 2016.
- [4] A. A. Zaidi, R. Baldemair, M. Andersson, S. Faxér, V. Molés-Cases, and Z. Wang, “Designing for the future: the 5G NR physical layer,” *Ericsson Technology Review*, p. 14, 2017.
- [5] Ericsson, “5G radio access,” Ericsson, Tech. Rep., April 2016, uen 284 23-3204 Rev C.
- [6] ITU-R, “IMT vision framework and overall objectives of the future development of IMT for 2020 and beyond,” Electronic Publication, September 2015, recommendation “ITU-R” M.2083-0.
- [7] H. Olofsson, A. Ericsson, F. Kronestedt, and S. Hellsten, “Leveraging LTE and 5G NR networks for fixed wireless access,” *Ericsson Technology Review*, p. 12, 2018.
- [8] T. Shimizu, V. Va, G. Bansal, and R. W. Heath, “Millimeter wave V2X communications: Use cases and design considerations of beam management,” in *2018 Asia-Pacific Microwave Conference (APMC)*, Nov 2018, pp. 183–185.
- [9] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G Mobile and Wireless Communications Technology*, 1st ed. New York: Cambridge University Press, June 2016.
- [10] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

- [11] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [12] A. Katalinic, R. Nagy, and R. Zentner, “Benefits of MIMO systems in practice: increased capacity, reliability and spectrum efficiency,” in *Proceedings ELMAR 2006*. IEEE, 2006, pp. 263–266.
- [13] J. K. Ojala and T. E. Lunttila, “Reporting Channel State Information,” U.S. Patent 9 001 747B2, April 7, 2015. [Online]. Available: <https://patentimages.storage.googleapis.com/6a/66/2a/8f27a92088af90/US9001747.pdf>
- [14] X. Li, E. Björnson, S. Zhou, and J. Wang, “Massive MIMO with multi-antenna users: When are additional user antennas beneficial?” in *2016 23rd International Conference on Telecommunications (ICT)*. IEEE, 2016, pp. 1–6.
- [15] E. Björnson, E. G. Larsson, and T. L. Marzetta, “Massive MIMO: Ten myths and one critical question,” *arXiv preprint arXiv:1503.06854*, 2015.
- [16] J. R. Hampton, *Introduction to MIMO communications*. Cambridge university press, 2013.
- [17] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.
- [18] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO wireless communications*. Cambridge university press, 2007.
- [19] P. Zhang, J. Chen, X. Yang, N. Ma, and Z. Zhang, “Recent research on massive MIMO propagation channels: A survey,” *IEEE Communications Magazine*, vol. 56, no. 12, pp. 22–29, 2018.
- [20] R. Zhang, X. Lu, J. Zhao, L. Cai, and J. Wang, “Measurement and modeling of angular spreads of three-dimensional urban street radio channels,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3555–3570, 2017.
- [21] E. Björnson, L. Van der Perre, S. Buzzi, and E. G. Larsson, “Massive MIMO in sub-6 ghz and mmwave: Physical, practical, and use-case differences,” *IEEE Wireless Communications*, 2019.

- [22] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.
- [23] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, “An overview of MIMO communications - a key to gigabit wireless,” *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb 2004.
- [24] A. M. Sayeed, “Deconstructing multiantenna fading channels,” *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct 2002.
- [25] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, “Fading correlation and its effect on the capacity of multielement antenna systems,” *IEEE Transactions on communications*, vol. 48, no. 3, pp. 502–513, 2000.
- [26] M. S. Sim, J. Park, C. Chae, and R. W. Heath, “Compressed channel feedback for correlated massive MIMO systems,” *Journal of Communications and Networks*, vol. 18, no. 1, pp. 95–104, 2016.
- [27] W. U. Bajwa, A. Sayeed, and R. Nowak, “Sparse multipath channels: Modeling and estimation,” in *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*. IEEE, 2009, pp. 320–325.
- [28] V. K. Goyal, “Theoretical foundations of transform coding,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 9–21, Sep. 2001.
- [29] H. S. Malvar, *Signal processing with lapped transforms*. Artech House, Inc., 1992.
- [30] K. Sayood, *Introduction to data compression*. Newnes, 2012.
- [31] K. R. Rao and P. C. Yip, *The transform and data compression handbook*. CRC Press, 2000.
- [32] K. R. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages, applications*. Academic Press, 2014.
- [33] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

- [34] M. Hamidi and J. Pearl, "Comparison of the cosine and fourier transforms of markov-1 signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 5, pp. 428–429, 1976.
- [35] R. G. Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.
- [36] P.-H. Kuo, H. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2012, pp. 492–497.
- [37] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," in *IEEE Transactions on Information Theory*, vol. 57, no. 7, July 2011, pp. 4680–4688.
- [38] W. Yang and J. D. Gibson, "Coefficient rate and significance maps in transform coding," in *Asilomar*, vol. 2, Nov 1997, pp. 1373–1377 vol.2.
- [39] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas, "Transform coefficient coding in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1765–1777, Dec 2012.
- [40] H. Kirchhoffer, H. Schwarz, T. Nguyen, D. Marpe, and T. Wiegand, "Coding of significance maps and transform coefficient blocks," U.S. Patent 0051 459A1, February 28, 2013. [Online]. Available: <https://patentimages.storage.googleapis.com/8a/70/af/3f975fe490d230/US20130051459A1.pdf>
- [41] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 1992, vol. 159.
- [42] D. J. Love, R. W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE communications magazine*, vol. 42, no. 10, pp. 54–59, 2004.
- [43] D. J. Love, R. W. Heath, V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on selected areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.

- [44] P. Cao and E. Jorswieck, “DCT and VQ based limited feedback in spatially-correlated massive MIMO systems,” in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2014, pp. 285–288.
- [45] Y. Han, W. Shin, and J. Lee, “Projection based feedback compression for FDD massive MIMO systems,” in *2014 IEEE Globecom Workshops*, Dec. 2014, pp. 364–369.
- [46] B. V. Boas, N. Fonseca, A. Klautau, and N. González-Prelcic, “Low complexity transform coding for millimeter wave MIMO CSI compression,” in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Oct 2018, pp. 1582–1586.
- [47] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, “5G MIMO data for machine learning: Application to beam-selection using deep learning,” in *2018 Information Theory and Applications Workshop (ITA)*, Feb 2018, pp. 1–9. [Online]. Available: <https://www.lasse.ufpa.br/raymobtime/>
- [48] “Raymobtime dataset,” Feb 2019. [Online]. Available: <https://github.com/lasseufpa/Raymobtime/wiki/Available-Raymobtime-Datasets>
- [49] C.-K. Wen, W.-T. Shih, and S. Jin, “Deep learning for massive MIMO CSI feedback,” *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.

