TECHNISCHE UNIVERSITÄT DRESDEN

Unified Framework for Multicarrier and Multiple Access based on Generalized Frequency Division Multiplexing

Ahmad Nimr

der Fakultät Elektrotechnik und Informationstechnik
der Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs
(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender:	Prof. Dr.-Ing. Peter Birkholz
Gutachter:	Prof. Dr.-Ing. Dr. h.c. Gerhard Fettweis
	Prof. Dr. Erik G. Larsson

Tag der Einreichung:	01. Dezember 2020
Tag der Verteidigung:	28. April 2021

Ahmad Nimr
Unified Framework for Multicarrier and Multiple Access based on Generalized Frequency Division Multiplexing
Dissertation, 28. April 2021

Vodafone Chair Mobile Communications Systems
Institut für Nachrichtentechnik
Fakultät Elektrotechnik und Informationstechnik
Technische Universität Dresden
01062 Dresden

Abstract

The advancements in wireless communications are the key-enablers of new applications with stringent requirements in low-latency, ultra-reliability, high data rate, high mobility, and massive connectivity. Diverse types of devices, ranging from tiny sensors to vehicles, with different capabilities need to be connected under various channel conditions. Thus, modern connectivity and network techniques at all layers are essential to overcome these challenges. In particular, the physical layer (PHY) transmission is required to achieve certain link reliability, data rate, and latency. In modern digital communications systems, the transmission is performed by means of a digital signal processing module that derives analog hardware. The performance of the analog part is influenced by the quality of the hardware and the baseband signal denoted as waveform. In most of the modern systems such as fifth generation (5G) and WiFi, orthogonal frequency division multiplexing (OFDM) is adopted as a favorite waveform due to its low-complexity advantages in terms of signal processing. However, OFDM requires strict requirements on hardware quality.

Many devices are equipped with simplified analog hardware to reduce the cost. In this case, OFDM does not work properly as a result of its high peak-to-average power ratio (PAPR) and sensitivity to synchronization errors. To tackle these problems, many waveforms design have been recently proposed in the literature. Some of these designs are modified versions of OFDM or based on conventional single subcarrier. Moreover, multicarrier frameworks, such as generalized frequency division multiplexing (GFDM), have been proposed to realize varieties of conventional waveforms. Furthermore, recent studies show the potential of using non-conventional waveforms for increasing the link reliability with affordable complexity. Based on that, flexible waveforms and transmission techniques are necessary to adapt the system for different hardware and channel constraints in order to fulfill the applications requirements while optimizing the resources.

The objective of this thesis is to provide a holistic view of waveforms and the related multiple access (MA) techniques to enable efficient study and evaluation of different approaches. First, the wireless communications system is reviewed with specific focus on the impact of hardware impairments and the wireless channel on the waveform design. Then, generalized model of waveforms and MA are presented highlighting various special cases. Finally, this work introduces low-complexity architectures for hardware implementation of flexible waveforms. Integrating such designs with software-defined radio (SDR) contributes to the development of practical real-time flexible PHY.

Zusammenfassung

Die Weiterentwicklung der drahtlosen Kommunikation ist der Schlüssel zu neuen Anwendungen mit hohen Anforderungen an niedrige Latenzzeiten, maximale Zuverlässigkeit, hohe Datenrate, hohe Mobilität und massive Verbindungen. Verschiedene Gerätetypen, von kleinen Sensoren bis hin zu Fahrzeugen, mit unterschiedlichen Fähigkeiten müssen unter verschiedenen Kanalbedingungen verbunden werden. Daher sind moderne Konnektivitäts- und Netzwerktechniken auf allen Schichten notwendig, um diese Herausforderungen zu erfüllen. Insbesondere die übertragung auf der physikalischen Schicht (PHY) ist erforderlich, um eine bestimmte Zuverlässigkeit der Verbindung, Datenrate und Latenzzeit zu erreichen. In modernen digitalen Kommunikationssystemen wird die übertragung mit Hilfe eines digitalen Signalverarbeitungsmoduls durchgeführt, das analoge Hardware ableitet. Die Leistung des analogen Teils wird durch die Qualität der Hardware und des Basisbandsignals, das als Signalform bezeichnet wird, beeinflusst. In den meisten modernen Systemen, wie z. B. der fünften Generation (5G) und WiFi, wird das orthogonale Frequenzmultiplexverfahren (OFDM) aufgrund seiner Vorteile bei der Signalverarbeitung mit geringer Komplexität als bevorzugte Signalform eingesetzt. OFDM erfordert jedoch strenge Anforderungen an die Hardwarequalität.

Viele Geräte sind mit einfacher analoger Hardware ausgestattet, um die Kosten zu senken. In diesem Fall funktioniert OFDM aufgrund des hohen Peak-to-Average Power Ratio (PAPR) und der Sensitivität gegenüber Synchronisationsfehlern nicht richtig. Um diese Probleme in den Griff zu bekommen, wurden kürzlich in der Literatur viele Wellenformdesigns vorgeschlagen. Einige sind modifizierte Versionen von OFDM oder basieren auf klassischen einzelnen Unterträgern. Darüber hinaus wurden Mehrträger-Frameworks, wie z. B. Generalized Frequency Division Multiplexing (GFDM), vorgeschlagen, um verschiedene Versionen der bisherigen Signalformen zu realisieren. Außerdem zeigen neuere Studien das Potenzial der Verwendung neuer Signale zur Erhöhung der Verbindungssicherheit bei akzeptabler Komplexität. Darauf aufbauend sind flexible Wellenformen und übertragungstechniken notwendig, damit das System an verschiedene Hardware- und Kanalbeschränkungen angepasst werden kann, um die Anforderungen der Anwendungen zu erfüllen und gleichzeitig die Ressourcen zu optimieren.

Das Ziel dieser Arbeit ist es, eine holistische Sicht auf Wellenformen und die zugehörigen Mehrfachzugriffstechniken (MA) zu bieten, um eine effiziente

Untersuchung und Bewertung verschiedener Verfahren zu ermöglichen. Zunächst wird das drahtlose Kommunikationssystem mit speziellem Fokus auf den Einfluss von Hardware-Einschränkungen und dem drahtlosen Kanal auf das Wellenformdesign untersucht. Dann werden allgemeinere Modelle von Wellenformen und MA vorgestellt, wobei verschiedene besondere Fälle hervorgehoben werden. Schließlich werden in dieser Arbeit Architekturen mit geringer Komplexität für die Hardware-Implementierung von flexiblen Wellenformen vorgestellt. Die Integration solcher Designs mit Software-Defined Radio (SDR) trägt zur Entwicklung praktischer flexibler Echtzeit-PHY bei.

Acknowledgements

I would like first to thank Prof. Gerhard Fettweis, for giving me the opportunity to work and conduct my Ph.D. at the Vodafone Chair Mobile Communications Systems at the Technische Universität Dresden. I am very thankful for being a member of vodafone chair, which provides many facilities in hardware, software, references, training seminars, academic and industrial events, in addition to unconstrained financial support for all research activities. With that, I was able to freely practice and explore my interests in theoretical research and practical implementation. During five years of the work on different projects, I enhanced my abilities in science and engineering, and shaped my own research personality. For that, I am also grateful to the head staff, Kathrin, Sylvia, Raffael, Rüdiger, Eva, and Yaning for keeping the chair running and supporting in all matters.

In relation to the thesis, I would like to express my gratitude to Prof. Gerhard Fettweis for his supervision, comments and careful review and to Prof. Erik G. Larsson for taking time to read and write the second review. I appreciate their supportive and motivating feedback. A special thanks to Dr. Marwa Chafii, for her important role in increasing the number of publications by valuating the results and pushing for publishing. And many thanks to Martin for exchanging practical experience, to Roberto for the insightful discussions on waveforms, and to Zhongju for his work on the implementation. In addition, I want to thank the former colleagues Dr. Dan Zhang, Dr. Ivan Gaspar and Dr. Maximilian Matthé for their technical support at the early stage of my work. And many thanks to the contemporary colleagues, Shahab, Ana, Zhitao for the nice time at the chair. And of course, I wont forget to thank the recent colleagues Ivo and Daniel for their support.

Besides the work, all the praise and appreciation is devoted to my dear wife Ramia for her unlimited support and taking a lot of responsibilities, especially in raising our two super active lovely boys, Joud and Karim. Together, we were able to harmonize our life and establish a happy family, which compensates us for being far away since years from our parents. I am very grateful to all the blessings that I have.

Finally, I want to thank every one had a positive impact on my life, starting from my parents, homeland friends, school and university teachers, and all the friends who I met abroad, in particular, the friends of Ilmenau and Dresden. A special thank to Usama who encouraged me to come to Dresden and hosted me several times, and to Lukas Landau for helping out with the move and driving the truck between Ilmenau and Dresden.

Contents

List of figures **ix**

List of tables **xi**

Abbreviations **xiii**

Mathematical Notations **xix**

1 Introduction **1**

1.1 Baseband transmission model 2

1.2 History of multicarrier systems 4

1.3 The state-of-the-art waveforms 6

1.4 Prior works related to GFDM 7

1.5 Objective and contributions 8

2 Fundamentals of Wireless Communications **11**

2.1 Wireless communications system 12

2.2 RF transceiver 13

2.2.1 Digital-analog conversion 13

2.2.2 QAM modulation 15

2.2.3 Effective channel 15

2.2.4 Hardware impairments 16

2.3 Linear waveform concept and techniques 19

2.3.1 Single-carrier waveform 21

2.3.2 Multicarrier waveform 23

2.3.3 MIMO-Waveforms 24

2.3.4 Waveform performance metrics 25

2.4 Wireless channel 29
2.4.1 Line-of-sight propagation 30
2.4.2 Multi path and fading process 31
2.4.3 General baseband statistical channel model 33
2.4.4 MIMO channel 35
2.5 Summary 35

3 Generic Block-based Waveforms 37
3.1 Block-based waveform formulation 38
3.1.1 Variable-rate multicarrier 39
3.1.2 General block-based multicarrier model 40
3.2 Waveform processing techniques 41
3.2.1 Linear and circular filtering 41
3.3 Generic multicarrier modem structure 45
3.3.1 Modulator 45
3.3.2 Demodulator 47
3.3.3 MIMO waveform processing 48
3.4 Detection 49
3.4.1 Maximum-likelihood detection 50
3.4.2 Linear detection 52
3.4.3 Iterative detection 56
3.4.4 Numerical examples and insights 58
3.5 Summary 73

4 Generic Multiple Access Schemes 77
4.1 Basic multiple access and multiplexing schemes 78
4.1.1 Infrastructure network system model 78
4.1.2 Duplex schemes 79
4.1.3 Common multiplexing and multiple access schemes 80
4.2 General multicarrier-based multiple access 83
4.2.1 Design with fixed set of pulses 84
4.2.2 Computational model 85
4.2.3 Asynchronous multiple access 87
4.3 Summary 90

5 Time-Frequency Analysis of Multicarrier **93**

5.1 General time-frequency representation 93

5.1.1 Block representation 94

5.1.2 Relation to Zak transform 96

5.2 Time-frequency spreading 97

5.3 Time-frequency block in LTV channel 99

5.3.1 Subcarrier and subsymbol numerology 100

5.3.2 Processing based on the time-domain signal 104

5.3.3 Processing based on the frequency-domain signal 105

5.3.4 Unified signal model 106

5.4 Summary 107

6 Generalized waveforms based on time-frequency shifts **109**

6.1 General time-frequency shift 110

6.1.1 Time-frequency shift design 110

6.1.2 Relation between the shifted pulses 111

6.2 Time-frequency shift in Gabor frame 112

6.2.1 Conventional GFDM 113

6.3 GFDM modulation 115

6.3.1 Filter bank representation 115

6.3.2 Block representation 116

6.3.3 GFDM matrix structure 117

6.3.4 GFDM demodulator 118

6.3.5 Alternative interpretation of GFDM 119

6.3.6 Orthogonal modulation and GFDM spreading 121

6.4 Summary 121

7 Modulation Framework: Architectures and Applications **123**

7.1 Modem architectures 124

7.1.1 General modulation matrix structure 124

7.1.2 Run-time flexibility 125

7.1.3 Generic GFDM-based architecture 126

7.1.4 Flexible parallel multiplications architecture 127

7.1.5 MIMO waveform architecture 128
7.2 Extended GFDM framework 130
7.2.1 Architectures complexity and flexibility analysis 132
7.2.2 Number of multiplications 132
7.2.3 Hardware analysis 134
7.3 Applications of the extended GFDM framework 137
7.3.1 Generalized FDMA 138
7.3.2 Enhancement of OFDM system 140
7.4 Summary 140

8 Conclusions and Future Works **143**

Publications of the author **151**

Bibliography **173**

List of Figures

1.1 Digital communications system. 2

1.2 Multicarrier systems. 4

2.1 Wireless communications system. 12

2.2 radio frequency (RF) Transceiver. 13

2.3 digital-to-analog (DAC) operations. 14

2.4 analog-to-digital (ADC) operations. 15

2.5 PAPR for discrete sequence $|x[n]| = 1$ and $\beta_{\text{tx}}(t) = \text{sinc}\,(F_s t)$. 18

2.6 Baseband processing functions. 20

2.7 Multicarrier waveform. 23

2.8 MIMO baseband signal. 24

2.9 out-of-band (OOB) for different finite pulse design. 27

2.10 PAPR evaluation for different waveforms using QPSK, SC (Dirichlet pulse), DFT-S-OFDM ($M = 32$, $K = 16$ subbands), and OFDM ($N = 1024$, 512 active subcarriers). 28

2.11 line of sight (LoS) channel model. 30

2.12 Jakes's model for a single path with $f_c = 3$ GHz, $v_x \in \{6, 60, 120\}$ Kmph. 32

2.13 Channel representation. 33

3.1 Overlapping between blocks. 38

3.2 Reformulation of multicarrier with variable rates. 40

3.3 Linear and circular filtering. 42

3.4 Block with zero-padding (ZP)/cyclic prefix (CP). 43

3.5 Block with ZP/CP. 44

3.6 Multicarrier waveforms generator stages. 45

3.7 Multicarrier waveforms demodulator. 47

3.8 2×2 MIMO waveform example. 49

3.9 General iterative parallel interference cancellation (PIC) receiver. 57

3.10 Average mutual information for QAM mapping in single-input, single-output (SISO) system with frequency selective channel under the assumption of perfect interference cancellation (Genie) and uniform power delay profile with L taps. 61

3.11 symbol error rate (SER) in different frequency selective channel. Low ($\gamma = 1$), and high ($\gamma = 3$) 64

3.12 SER in different spreading with perfect inter-symbol interference (ISI) cancellation ($\gamma = 3$). 64

3.13 SER for different spreading with linear minimum mean square error (LMMSE) equalization ($\gamma = 3$). 65

3.14 bit error rate (BER) comparison of SC and OFDM with interleaving [on] and non-interleaving [off] for different selectivity; high selectivity $\gamma = 3$ and low-selectivity $\gamma = 1$. 66

3.15 FER comparison of SC and OFDM with interleaving [on] and non-interleaving [off] for different selectivity; high selectivity $\gamma = 3$ and low-selectivity $\gamma = 1$. 66

3.16 frame error rate (FER) comparison of Walsh-Hadamard transform (WHT) and single-carrier (SC) in high selective channel ($\gamma = 3$). 67

3.17 Number of iterations of WHT and SC in high selective channel ($\gamma = 3$). . . 67

3.18 FER with different spreading matrices. Here, $M = 8$, $N = 64$ and $\gamma = 3$. . 68

3.19 Mutual information for different MIMO settings and QAM order. The spread case is calculated for the exponential channel with $\gamma = 3$ and assuming perfect interference cancellation (genie). 71

3.20 SER for multiple-input, multiple-output (MIMO) with uncorrelated channels ($\gamma = 3$) 71

3.21 FER for MIMO with uncorrelated channels and exponential power delay profile (PDP)($\gamma = 3$), $N = 64$ and, payload size $= K_p = L_c R_c N_T N$. 72

4.1 Wireless network infrastructure. 78

4.2 Generic SISO infrastructure multiuser system model. 79

4.3 Basic orthogonal SISO MA schemes. 81

4.4 Generalized FDMA-TDMA. 87

4.5 inter-user interference (IUI) using full allocation with orthogonal matrix. The right side employs discrete Fourier transform (DFT) or WHT matrix with null symbol. 88

4.6 Interference of user 2 to user 1 with different waveforms settings ΔN_g is the guard band in number of frequency bins. For OFDM it is explicitly specified by $[\Delta N_g = x]$, and it is fixed for DFT and WHT. 89

5.1 Time-frequency modulation. 94

5.2 Zak transform. 96

5.3 Discrete channel delay profile for extended ITU Veh. A. 103

5.4 Static channel approximation error within the subsymbol vs number of subcarriers using extended ITU Veh. A at $f_c = 5$ GHz, $F_s = 10$ MHz. . . . 104

6.1 Core block with multiple prototype pulses. 114

6.2 GFDM modem with circular convolution representation. 116

6.3 GFDM modulator 119

6.4 GFDM demodulator modulator 120

6.5 Orthogonal GFDM. 121

7.1 Generic GFDM-based modem structure. 127

7.2 Parallel convolution architecture. 128

7.3 Flexible MIMO architecture. 129

7.4 Unified architecture for time-domain (TD) and frequency-domain (FD) GFDM processing. 131

7.5 Complexity evaluation model. 132

7.6 Number of complex multiplications (CMs), where $N = 1024$ and $M = 2^m, m = 1 \cdots, 9$. 134

7.7 Efficient modem realizations. 134

7.8 Latency evaluation, where M is fixed, and $N = 2^n, n = 6, \cdots, 11$. 137

7.9 Precoded OFDM system model. 140

List of Tables

3.1 Summary of block-based waveform parameters. 41

3.2 Multicarrier waveform generation steps 46

3.3 Simulation parameters. 63

5.1 Extended ITU Veh. A PDP 102

5.2 Coherence time in samples for different velocity and $f_c = 5$ GHz. 102

5.3 General received block in time variant channel. 106

6.1 GFDM based waveform parameters 115

6.2 The four steps of conventional GFDM modulator. 122

7.1 GFDM unified representation parameters 130

7.2 Configuration parameters for conventional GFDM modem. 131

7.3 Total number of CMs considering FD equalization. 133

7.4 Number of required resources. 135

7.5 Xilinx FFT internet protocole (IP) processing latency in cycles. 136

7.6 Latency evaluation. 136

Abbreviations

2G	second generation
3G	third generation
4G	fourth generation
5G	fifth generation
A/D	analog-to-digital
ADC	analog-to-digital
AR	augmented reality
ASIC	application-specific integrated circuit
AWGN	additive white Gaussian noise
BER	bit error rate
BS	base stations
CCDF	complementary cumulative density function
CDMA	code-division multiple access
CFO	carrier frequency offset
CIR	channel impulse response
CM	complex multiplication
COFDM	coded-OFDM
COQAM	cyclic OQAM
CP	cyclic prefix
CRC	cyclic redundancy check
CS	cyclic suffix
CSI	channel state information
CSMA	carrier-sense multiple access
CWCU	component-wise conditionally unbiased
D/A	digital-to-analog
D2D	device-to-device
DAC	digital-to-analog
DC	direct current
DFT	discrete Fourier transform

DL	downlink
DMT	discrete multitone
DNN	deep neural network
DSL	digital subscriber line
DSP	digital signal processor
DTFT	discrete-time Fourier transform
DVB	digital video broadcasting
E2E	end-to-end
FBMC	filter bank multicarrier
FD	frequency-domain
FDD	frequency-division duplexing
FDE	frequency domain equalization
FDM	frequency division multiplex
FDMA	frequency-division multiple access
FEC	forward error correction
FER	frame error rate
FFT	fast Fourier transform
FIR	finite impulse response
FMT	filtered multi tone
FO	frequency offset
F-OFDM	filtered-OFDM
FPGA	field programmable gate array
FT	Fourier transform
FTN	faster-than-Nyquist signaling
GFDM	generalized frequency division multiplexing
GFDMA	generalized frequency division multiple access
GMC-CDM	generalized multicarrier code-division multiplexing
GSM	Groupe Spécial Mobile
H2H	human-to-human
H2M	human-to-machine
I	in-phase
i.i.d.	independent and identically distributed
IBI	inter-block interference
ICI	inter-carrier interference
ICT	information and communication technologies
IDFT	inverse discrete Fourier transform
IoT	Internet of Things

IP	internet protocole
ISDN	integrated services digital network
ISI	inter-symbol interference
IUI	inter-user interference
LAN	local area netwrok
LLR	log-likelihood ratio
LMMSE	linear minimum mean square error
LNA	low noise amplifier
LO	local oscillator
LoS	line of sight
LP	low-pass
LPF	low-pass filter
LS	least squares
LTE	Long Term Evolution
LTV	linear time variant
M2M	machine-to-machine
MA	multiple access
MAC	multiple access control
MAP	maximum a posteriori
MC	multicarrier
MCM	multicarrier modulation
MCS	modulation coding scheme
MF	matched filter
MIMO	multiple-input, multiple-output
MISO	multiple-input single-output
ML	machien learning
MLD	maximum likelihood detection
MRC	maximum ratio combining
MSE	mean squared error
MTC	machine type communication
MU	multi user
NOMA	non-orthogonal multiple access
NRZ	non-return-to-zero
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OOB	out-of-band
OQAM	offset quadrature amplitude modulation

OTFS	orthogonal time frequency space
PA	power amplifier
PAPR	peak-to-average power ratio
PDF	probability distribution function
PDP	power delay profile
PHY	physical layer
PIC	parallel interference cancellation
PLC	power line communication
PSD	power spectral density
Q	quadrature-phase
QAM	quadrature amplitude modulation
QoS	quality of service
R/W	read-or-write
RAM	random-access memmory
RAN	radio access network
RAT	radio access technologies
RF	radio frequency
RW	read-and-write
SC	single-carrier
SD	sphere decoding
SDD	space-division duplexing
SDMA	space division multiple access
SDR	software-defined radio
SER	symbol error rate
SIC	successive interference cancellation
SINR	signal-to-interference-plus-noise ratio
SIR	signal-to-interference ratio
SISO	single-input, single-output
SNR	signal-to-noise ratio
STC	space-time coding
SVD	singular value decomposition
TD	time-domain
TDD	time-division duplexing
TDMA	time-division multiple access
TO	time offset
UE	user equipment
UFMC	universally filtered multicarrier

UL	uplink
US	uncorrelated scattering
VR	virtual reality
WCP	windowing and CP
WHT	Walsh-Hadamard transform
WLAN	wireless local area network
W-OFDM	windowed-OFDM
WOLA	windowing and overlapping
WSS	wide-sense stationary
ZF	zero-forcing
ZMCSCG	zero-mean circularly-symmetric complex Gaussian
ZP	zero-padding
ZT	zero-tail

Mathematical Notations

Notations

$\mathbb{C}$	Complex number field
$\mathbb{R}$	Real number field
$\mathbb{Z}$	Set of integer set
$\boldsymbol{A}$, $\boldsymbol{a}$	Uppercase (Lowercase) and bold face for matrix (vector)
a, A	Scalar value
$\mathcal{A}$	Calligraphic letter for set
$\lvert\mathcal{A}\rvert$	Number of elements in the set $\mathcal{A}$
$\boldsymbol{A}[n,m]$, $\boldsymbol{a}[n]$	The (n,m)-th element (n-th element) of a matrix (vector). The result is zero if the index is out of the range.
$\boldsymbol{A}[n,:], \boldsymbol{A}[:,m]$	The n-th column vector and the m-th row vector, receptively
$\boldsymbol{A}[\mathcal{N},\mathcal{M}]$	A sub-matrix of size $\lvert\mathcal{N}\rvert \times \lvert\mathcal{M}\rvert$ by selecting the elments at the indexes $(n,m) \in \mathcal{N} \times \mathcal{M}$
$x[n]$	The discrete-time signal value at the sample n
$\tilde{x}(\nu)$	The discrete-time Fourier transform (DTFT), $X(\nu) = \sum_n x[n]e^{-j2\pi\nu n}$
$x(t)$	Continuous-time signal value at the instance t
$\tilde{x}(f)$	Fourier transform $X(f) = \int_t x(t)e^{-j2\pi ft}dt$
$\boldsymbol{F}_N$	The normalized N-DFT matrix, $\boldsymbol{F}_N[k,m] = \frac{1}{\sqrt{N}}e^{-2\pi\frac{nk}{N}}$
$\boldsymbol{I}_N, \mathbf{1}_N, \mathbf{0}_N$	Identity, all ones, and all zeros matrices of size $N \times N$
$\mathbf{1}_{N\times M}, \mathbf{0}_{N\times M}$	All ones, and all zeros matrices of size $N \times M$
$\tilde{\boldsymbol{a}} = \boldsymbol{F}_N\boldsymbol{a}$	The DFT of $\boldsymbol{a}$, $\boldsymbol{a} = \boldsymbol{F}_N^H\tilde{\boldsymbol{a}}$
$\langle\cdot\rangle_N$	Modulo-N operator
$\delta[n]$	Discrete Dirac pulse, $\delta[0] = 1$ and $\delta[n] = 0, n \neq 0$
δ_{ij}	Kronecker delta, $\delta_{ij} = 1, i = j, \delta_{ij} = 0, i \neq j$

Functions and operators

$\mathrm{vec}\left\{\boldsymbol{A}\right\}$	Vectorization operation
$\mathrm{unvec}_{N\times M}\left\{\boldsymbol{a}\right\}$	Inverse of vectorization
$\mathrm{diag}\left\{\boldsymbol{A}\right\}$	Diagonal elements of $\boldsymbol{A}$
$\boldsymbol{\Lambda}^{(\boldsymbol{a})}$	Diagonal matrix, where $\boldsymbol{\Lambda}^{(\boldsymbol{a})}[n,n]=\boldsymbol{a}[n]$
$\{\cdot\}^T$	Matrix or vector transpose
$\{\cdot\}^*$	Matrix or vector complex conjugate
$\{\cdot\}^H$	Matrix or vector Hermitian transpose $\boldsymbol{A}^H=\{\boldsymbol{A}^*\}^T$
$\{\cdot\}^{-1}$	Inverse of a square matrix
$\{\cdot\}^+$	Moore–Penrose inverse, $\boldsymbol{A}^+=\left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H$ or $\boldsymbol{A}^+=\boldsymbol{A}^H\left(\boldsymbol{A}\boldsymbol{A}^H\right)^{-1}$ depending on which inverse is valid
$\mathrm{trace}\left\{\cdot\right\}$	Trace of a square matrix $\boldsymbol{A}\in\mathbb{C}^{N\times N}$, $\mathrm{trace}\left\{\boldsymbol{A}\right\}=\sum\limits_{n=0}^{N-1}\boldsymbol{A}[n,n]$
$\lVert\cdot\rVert_F$	Frobenius norm of, $\lVert\cdot\rVert_F^2=\mathrm{trace}\left\{\boldsymbol{A}\boldsymbol{A}^H\right\}=\mathrm{trace}\left\{\boldsymbol{A}^H\boldsymbol{A}\right\}$
$\lVert\cdot\rVert$	Second order norm of a vector, $\lVert\boldsymbol{a}\rVert=\boldsymbol{a}^H\boldsymbol{a}$
$\mathrm{E}\left[\cdot\right]$	Expectation function
$\mathrm{Re}\left\{\cdot\right\}$	Real part of a complex-valued input
$\mathrm{Im}\left\{\cdot\right\}$	Imaginary part a complex-valued input
$\odot$	Element-wise product of matrices or vectors
$\oslash$	Element-wise division Element-wise product of matrices or vectors
$\otimes$	Kronecker product
$*$	Linear convolution, $y[n]=h[n]*x[n]=\sum\limits_{l}h[l]x[n-l]$
$\circledast$	Circular convolution, $\boldsymbol{y}[n]=\boldsymbol{h}[n]\circledast\boldsymbol{x}[n]=\sum\limits_{l=0}^{N-1}\boldsymbol{h}[l]\boldsymbol{x}[\langle n-l\rangle_N]$, $\tilde{\boldsymbol{y}}=\tilde{\boldsymbol{h}}\odot\tilde{\boldsymbol{x}}$, $\boldsymbol{y}=\boldsymbol{C}^{(\boldsymbol{h})}\boldsymbol{x}$

Chapter 1

Introduction

Communication is the process of exchanging meaningful information between entities for certain purposes. This implies several questions about the types and situation of involved entities, the objective of communication, the representation of information, and the means of exchanging. Inspired by human-to-human (H2H) [BK79] communication, machine-to-machine (M2M) [BEH12] and human-to-machine (H2M) [LS+13] have been developed leading to advancement and revolution in information and communication technologies (ICT). The purposes of H2H communication include entertainment, education, collaborative work, warnings, and social contact for sharing ideas and feelings. The messages can be represented in form of verbal and body language, facial expression, or eye contact, which are exchanged by means of speech, vision, or haptic contact using the human physiological system [Mor10]. The feasibility of such communication over long distance is limited by the necessity of physical presence, the abilities of human senses, and other barriers [Men96]. To overcome this limitation, new means of telecommunication and information presentation have emerged to realize different purposes [ITU06].

The early scope of telecommunication technologies was to enable voice calls and low-data rate leading to the invention of fixed wired telephony, e.g. integrated services digital network (ISDN) [Sta95] and wireless mobile networks, e.g. Groupe Spécial Mobile (GSM) (second generation (2G)) network [MPFBH92]. With the increase of the communication speeds provided by broadband technologies such as digital subscriber line (DSL) [OMH+09] and Long Term Evolution (LTE) (fourth generation (4G)) [GRM+10], high quality video calls become possible as well. The demand of industry has changed from the basic needs of device-to-device (D2D) connection [AWM14] to an industrial networks for automation, monitoring and remote operation [KL14]. The initial goal of inventing the local area netwroks (LANs) was to enable exchanging information between local computers [Spu00]. The idea has expanded towards the internet connecting computers over the globe, until reaching the Internet of Things (IoT) [AIM10], where everything is connected. The entertainment industry has changed the scope from broadcasting of contents to on-demand streaming. The need of communications at anytime and everywhere shifted the interest

from fixed to mobile networks [SSA16] to support mobility and maintain connectivity while traveling. To allow flexibility and save wiring costs, wireless links have replaced cables using technologies such as Bluetooth [BS01] and WiFi [CWKS97].

Recently, the fifth generation (5G) and beyond technologies will be the key-enablers of new applications with stringent requirements in low-latency, ultra-reliability, and high data rate, high mobility, and massive connectivity [All15, DB18]. Among the future use cases that have important social and economical impacts; 1) telepresence [DKU98, LHJ17] will become an alternative of personal meetings to reduce the number of flights and to fulfill social distancing during pandemics, 2) the internet of skills [DML+17] will enable high level of remote education and skill transfer over the globe, and 3) the 4th industrial revolution (Industry 4.0) [VT14, Lu17] will transform the manufacturing from mass production to customization of products relying on a network of intelligent robots that can be reprogrammed to do multiple instead of repetitive tasks. These use cases require heterogeneous services with different requirements including virtual reality (VR) and augmented reality (AR) with high-resolution and low-latency video, haptic feedback via the tactile internet [Fet14, SAD+16], a network of sensors with massive connectivity [BPW+18], in addition to the support of vehicular communication [SAIZ18]. Achieving that needs to consolidate efforts in the development of different system components including devices, network, control, and management.

The starting point of any network is the physical layer (PHY) transmission over dedicated or shared medium channel. The end-to-end (E2E) requirements of applications in terms of throughput and latency, as well as the channel condition influence the design of transmission systems to achieve certain link reliability, data rate, and latency. The focus of this thesis is to develop a unified PHY and multiple access framework that is able to realize legacy transmission techniques and propose alternatives.

1.1 Baseband transmission model

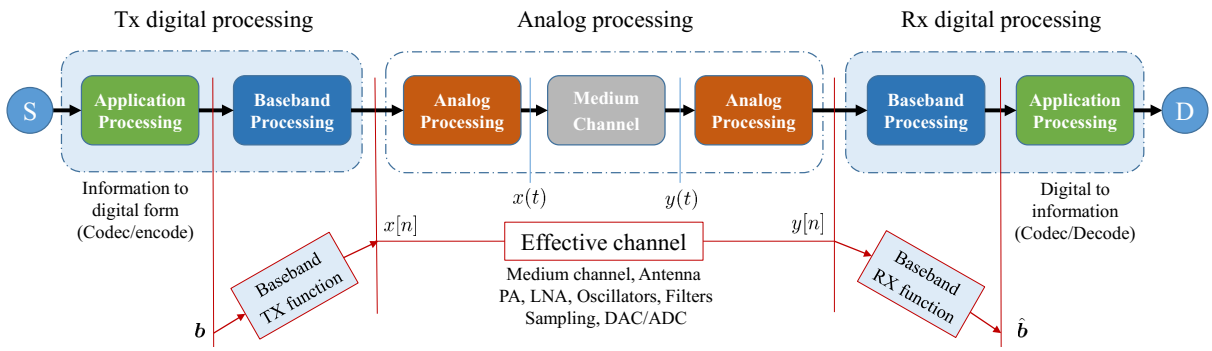

Figure 1.1: Digital communications system.

Modern telecommunication systems, as depicted in Figure 1.1, employ digital

representation of the information [Hay94]. The application processing converts the source (S) signal to a digital bit stream and vice versa at the destination (D). This include analog-to-digital (A/D) conversion, compression, etc. Electrical and electromagnetic signals are used to curry the information over wire, radio, optical, ultrasonic or other electromagnetic medium [ITU16]. For some mediums, discrete signals can be directly used such as in wired serial communication [Wic97]. The signal propagated through other mediums is naturally continuous, e.g. the electromagnetic waves over radio communication [SAZ07] or transmission lines [JE97]. The signal generation can be split into baseband processing, where a discreet sequence $x[n]$ is generated based on the input data $\boldsymbol{b}$. The analog processing is responsible of converting the discrete sequence to a continuous signal $x(t)$ suitable for transmission over the medium. The analog transmission may involve multiple-input, multiple-output (MIMO) techniques, where multiple analog signals are transmitted and/or received [LLL+10]. The inverse operations are performed at the receiver to obtain the discrete sequence $y[n]$, which is processed to reproduce the data.

The design of analog processing depends on the physical characteristics of the propagation medium and the properties of signals. The transceiver used DSL [JYL+99] over twisted pair is different from the one used for power line communication (PLC) [FC12]. Optical communications use light signals, which can be visible light or infra red. The design is different for free-space [ZK02] and fiber-optic communication [Pal88]. In radio communications, the design of radio frequency (RF) frontend depends on the range of carrier frequencies and signal bandwidth [AAB+06].

The baseband transmission model $y[n] = h(x[n]) + z[n] + v[n]$ is practically used as a mathematical model for communication following the Shannon model [Sha48]. Here, $v[n]$ is additive thermal noise, $z[n]$ interference caused by multiple access of the channel or from external sources, and $h(\cdot)$ represents the effective discrete channel function, which combines the overall effects of the medium and analog hardware impairments. This thesis focuses particularly on RF communications, however, the results concerning the baseband processing can be extended to other systems.

The design of baseband processing is essential to alleviate the analog hardware impairment such as power amplifier (PA) non-linearity [TPZ05], local oscillator (LO) synchronization error [Tom98]. Moreover, it influences the spectral properties of the analog signal and the out-of-band (OOB) [KB13] emission. It has also an impact on the detection complexity and reliability performance depending on the effective channel model [Sch04, FBM07]. Furthermore, multiple access techniques, such as time-division multiple access (TDMA), frequency-division multiple access (FDMA), code-division multiple access (CDMA) can be realized by the design of $x[n]$ [JBS93a].

The design of the baseband signal $x[n]$ is commonly known as waveform design [WJK+14]. Multicarrier waveform is a common design approach used for multiplexing the data from a single source, or for multiple access by assigning subsets of the subcarriers to multiple sources. In multicarrier systems, $x[n]$ is generated from the superposition of signals $\{x_k[n]\}$

denoted as subcarriers, such that $x[n] = \sum_{k \in \mathcal{K}_{on}} x_k[n]$. Each subcarrier can experience different effective channel. The design of multicarrier systems has evolved through the history and become dominant in many commercial systems.

1.2 History of multicarrier systems

The original motivation behind multicarrier modulation (MCM) is to tackle the frequency selectivity of the wireless channel for broadband communications [Bin90]. The idea is to transmit high data rate using parallel streams, each stream has a lower data rate and occupies narrower bandwidth. Thereby, the effective channel corresponds to each stream is approximately flat. This enables simple equalization of the MCM signal at the receiver and allows adaptive loading depending on the channel condition at each stream. The name carrier is inherited from the concept of frequency division multiplex (FDM), where several signals are transmitted simultaneously over the same medium using different carrier frequencies [SB08].

The carrier frequencies are spaced adequately to eliminate the cross talk (inter-carrier interference (ICI)). The earliest MCM schemes apply the FDM principle. As illustrated in

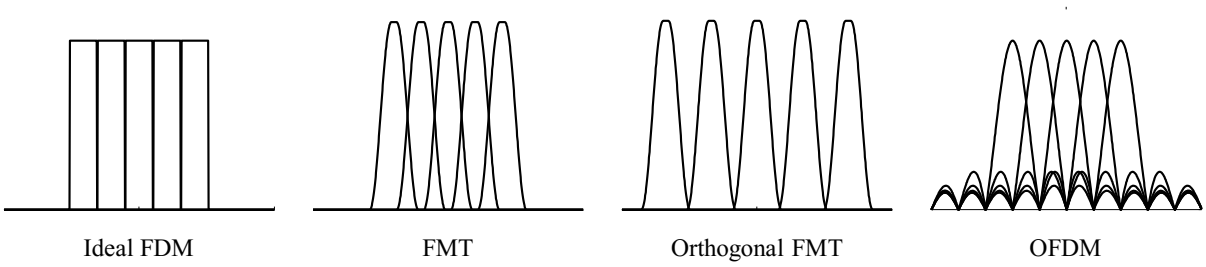

Figure 1.2: Multicarrier systems.

Figure 1.2, the available bandwidth is divided into several narrow bands centered around subcarriers. Filters are used to separate the bands. In the ideal case, the frequency response of the filter needs to be sharp to confine the signal within the narrow band. This requires infinite impulse response, which is practically infeasible. The practical subband filters can be generated from a prototype filter with frequency shift equals to the subcarrier spacing as in filtered multi tone (FMT) [CEO99]. The ICI depends on the overlap between adjacent filters. To eliminate the ICI and retain the orthogonality, the subcarrier spacing can be set larger than the filter bandwidth. With that, the modulation is orthogonal FDM. However, the spectral efficiency in this case is reduced. In order to preserve both orthogonality and spectral efficiency, two approaches emerged during the past years; the first leads to offset quadrature amplitude modulation (OQAM)-orthogonal frequency division multiplexing (OFDM) and the other leads to discrete Fourier transform (DFT)-OFDM [Wei09]. In OQAM-OFDM, the prototype filter is designed to be overlapped with the adjacent

subcarriers. The orthogonality is maintained at the in-phase and quadrature components of the complex data symbols. This is done by delaying the in-phase component by half the symbol period in one channel and vice versa on the adjacent channel [Sal67]. This approach is revised and reintroduced by the name filter bank multicarrier (FBMC) [Far11]. In the beginning of DFT-OFDM, the complex data symbols are transformed with DFT, the real part is passed through low-pass filter [WE71]. Later, similar approach is used with the consideration of the complex signal and using inverse discrete Fourier transform (IDFT) under the name discrete multitone (DMT) [CTC91], which is nowadays known as OFDM. The prototype filter used in OFDM has a sinc response, which overlaps with more than the adjacent subcarriers. However, plain OFDM does not serve the purpose of combating the fading channel, because inter-symbol interference (ISI) is introduced between successive OFDM symbols as a result of the channel delay spread. Nevertheless, with cyclic prefix (CP), which is invented prior to having the well-known form of OFDM [PR80], the frequency-domain processing of CP-OFDM allows the separation of the symbols carried on the multicarriers. Thereby, each received symbol is weighted with a channel gain. After removing the CP, the channel matrix is circular, and thus, it can be diagonalized with DFT without any knowledge of the channel state information (CSI) at the transmitter. An alternative approach called [KAC90] gits rid of the CP overhead by precoding the data with respect to the channel. The channel need to be known at the transmitter, and the data are precoded with respect to the singular value decomposition (SVD) of the channel.

Based on OFDM, orthogonal frequency division multiple access (OFDMA) is adopted as the multiple access technique to serve multiple users at the same time. Each user is assigned a set of subcarriers [SLK97] and the power assigned to each user depends on the CSI by means of power allocation [KH00]. Reference symbols known as pilot subcarriers [HL91] can be multiplexed with the data for channel estimation and synchronization.

Other existing multiple access schemes such as CDMA [JBS93b] are combined with OFDM [HP99], which is a particular case of more general multicarrier (MC)-CDMA [FBA94], to improve the detection in sever channels. In this scheme, the data symbols are precoded with Walsh-Hadamard matrix prior to mapping to the subcarriers. At the receiver, inverse precoding is applied after channel equalization. The precoded-OFDM approch can be extended to cover other cases. For instance, considering the frequency domain equalization (FDE) of single carrier system which is first proposed in [WS73], it can be seen as precoded OFDM, where the precoding matrix is the DFT matrix as noted in [SKJ95]. More details about the relation between single-carrier (SC)-FDE and OFDM is given in [NLF07]. The idea of precoded-OFDM is extended to span several OFDM symbols. Another famous precoding is DFT-spread-OFDM, which is exploited for the uplink multiple access under the name SC-FDMA [MLG06b, DFCS04].

Unlike other forms of FDM modulations such as FBMC, which are designed with expensive filter bank, the implementation of OFDM is simple and can be realized on digital signal processor (DSP) using fast Fourier transform (FFT) algorithm [CT65]. Therefore,

OFDM gains main attention among the other multicarrier and has been adopted by many standards for wired and wireless systems. The first sounding application was in DSL [SCC93] under the original name DMT. Moreover, OFDM is adopted by the WiFi (IEEE 802.11x) and WiMAX (IEEE 802.16) standards. It is also used in digital video broadcasting (DVB) standard [Rei98], where coded-OFDM (COFDM) [LAB95] is mainly used. This scheme is essentially OFDM, with additional processing steps applied to the data including channel coding, and frequency and time interleaving. After using CDMA in the third generation (3G) [OP98], OFDM is the technique used in the 4G and continues to be used in the 5G mobile communications [LSH+17].

1.3 The state-of-the-art waveforms

The well known problems associated with OFDM including high peak-to-average power ratio (PAPR) and strict synchronization, prompt the research for new alternatives. For the new machine type communication (MTC) use cases, low PAPR is necessary to reduce the complexity and cost of the radio frequency front end. OFDM requires strict synchronization in the uplink (UL) to maintain orthogonality among subcarriers that are allocated to different users. Because of the sinc response, a small offset leads to high inter-user interference (IUI). Thus, OFDM is not appropriate for asynchronous multiple access. The sinc response is also responsible for high OOB, which reduces its efficiency for dynamic spectrum access. To tackle these challenges, additional signal processing approaches on top of OFDM have been suggested, such as subband filtering [VWS+13] and windowing [Pra04]. Filtering is also used as per group of subcarriers mainly for reducing the OOB among users in the UL. This technique is named universally filtered multicarrier (UFMC) [VWS+13]. Other filtering is applied to the whole block and given the name filtered-OFDM (F-OFDM) [AJM15]. Windowing is an old technique used for wireless local area network (WLAN) standard [vAM+99]. However, some new publications associate the windowing with new waveforms proposals, calling it windowed-OFDM (W-OFDM) [AKR16], windowing and overlapping (WOLA)-OFDM [ZMSR16] when applied on OFDM, or windowing and CP (WCP)-cyclic OQAM (COQAM) when applied to COQAM [LS14]. Other processing on OFDM to solve the problems is done via precoding. One of the widely used precoding to reduce the PAPR is DFT-spread-OFDM, which is originally proposed for the UL [MLG06b]. By inserting zero subcarriers and spreading with DFT matrix, the resulting modulation is called zero-tail (ZT)-DFT-spread-OFDM [BTS+13]. Other precoding techniques target reducing the OOB by smoothing the transition between blocks such as N-continuous-OFDM [vB09]. To reduce the PAPR, index modulation is proposed [BAPP13]. In this approach, the number of allocated subcarrier is reduced, and their indexes are exploited to encode data to compensate for the data rate loss.

For the application of opportunistic spectrum access of the white space in the TV UHF bands, generalized frequency division multiplexing (GFDM) is first published in 2009

[FKB09]. The idea is to address the challenges faced in OFDM, such as high PAPR and high OOB with affordable complexity. Combining the block-based processing used in GFDM, where the successive blocks are isolated with a CP, and using the filter bank multicarrier techniques, the resulting GFDM modulation can be simply described as filter bank system with circular filtering. Different proposals with different names leads to exactly the same system. For instance CP-FMT is proposed four years later for the application of power line communications [TG13]. The circular filtering as a technique to allow block-based and adding CP is also proposed in 2011 for OQAM-OFDM under the name CP-OQAM-OFDM [GWX+11]. Later on, the same technique reinvented and denoted as COQAM[LS14]. Prior to that, the same authors propose in 2008 the use of CP with OQAM under the name CP-OFDM/OQAM [LS08], without explicitly mentioning the circular filtering. In fact, CP-OQAM-OFDM is similar to frequency-shift OQAM-GFDM[GMM+15]. The original GFDM description considers the filtering of the subcarrier streams after adding a CP. If linear filtering is used, then the equivalent channel delay is larger according to the filter tail. As a consequence, the CP length needs to be larger. Nevertheless, if tail biting is used, the CP length can be kept shorter. This process is equivalent to circular filtering followed by CP insertion, which becomes the well-known GFDM [MMG+14]. A system called generalized multicarrier code-division multiplexing (GMC-CDM) is introduced in [LYY10] has similar structure of GFDM.

1.4 Prior works related to GFDM

The flexibility of GFDM has attracted contributions in different aspects. In [MMF14a], the GFDM structure is represented by means of Gabor analysis. This work emphasis that GFDM is singular if both even number of subcarriers and subsymbols are used with real-valued symmetric prototype pulse. Several works focus on the decomposition of the modulation matrix using different terminologies leading to SVD-like structure with different permutations [CSH17, WXXL16, LW16]. Using this decomposition, the demodulation matrix is determined. Filter design is considered for reducing the OOB in [HSL17] and [CS17], for low PAPR in [TSC18] and [LDL19], and to mitigate the non-linearity of PA in [LHS18]. The influence of different pulse design on the performance is evaluated in [MMGF14], and for better-than-Nyquist pulse design in [KMB17]. Several works evaluated the performance of GFDM. The bit error rate (BER) performance in single-input, single-output (SISO) system is presented in [MKLF12] and [CKFM18]. Link-level simulation including comparison with other waveforms is presented in [ZFF17]. The sensitivity to synchronization error is studied in [GMP17] by the evaluation of the BER, and via signal-to-interference ratio (SIR) analysis in [LK17]. Low-complexity modem design in the time-domain (TD) based on small circular convolution is introduced in [D+15] and [FMD16]. The low-complexity frequency-domain (FD) modem design in [GMN+13] exploits the sparsity of the prototype pulse. Several works consider

GFDM as a massive MIMO expressed by the the joint channel and modulation matrix. Low-complexity linear minimum mean square error (LMMSE) is studied in [TD18] and [MGZF15] exploiting the joint matrix structure. Other works focus on receiver design by performing LMMSE channel equalization followed by demodulation as in [DMR19]. Additionally, matched filter (MF) and interference cancellation approach is introduced in [DMLF12]. Precoding is applied on the input data symbols for PAPR reduction [SOSF16], and for diversity using precoding matrices on the subsymbols such as DFT [HP17], Walsh-Hadamard transform (WHT) [MMM+15], and fractional DFT [WMSL18]. Other precoding approaches are presented in [MMG+16] and [TSB15]. In the context of generalized frequency division multiple access (GFDMA), asynchronous multiple access (MA) is evaluated in [MMF15], and further studied with interference cancellation approaches in [LK19] and [WS18]. A frame structure for compatibility with LTE standard is discussed in [GMM+14]. Several modifications of GFDM are proposed to eliminate self-interference without noise enhancement [LZZ+19, MF16, TA18], which lead to similar form of OQAM-GFDM[GMM+15]. The work in [AES+16] introduces GFDM with different subcarrier bandwidth. Well-known techniques for OFDM are restudied with GFDM such as cognitive radio [DMK+12], index modulation [OBC16], unique word [Dd17], IQ imbalance compensation [THX+17], relaying [NLZ+19], and power allocation [MBT19]. MIMO-GFDM works include transmission techniques such as space-time coding (STC) [MMF14b, MMG+15], implementation [DMG+15], detection [MZF18, MZF16], and link-level simulation [ZMMF17].

1.5 Objective and contributions

As discussed in the previous sections, many variants of waveforms have been proposed in the recent years to solve the problems associated with conventional OFDM. Some focus on features related to hardware impairments, for instance, to improve the PAPR such as DFT-spread-OFDM, others targets low OOB such as FBMC and GFDM. Recent waveforms such as orthogonal time frequency space (OTFS) [H+17] consider the reliability performance in linear time variant (LTV) channel. The emerging applications have different requirements on hardware and operate in different channel conditions. Based on that, flexible waveform and transmission techniques are required to solve all that problems and fulfill the application demands. The main objective of this thesis, is to provide a holistic view of waveform and related MA and fuse them in a unified framework. This enables efficient study and evaluation without a need to go through individual approaches separately. For instance, a general linear model can be considered for detection in many cases. However, considering the implementation complexity, some matrix structures can be exploited. For that purpose, this work aims at narrowing the generalization scope towards special cases that fulfill most of the waveform design needs while offering low-complexity architectures. Accordingly, the thesis is organized as follows:

Chapter 2 introduces a generic model for digital wireless communications systems with three connected modules. Namely, the information source produces by the communication protocols, the baseband digital signal processing, and the analog transceiver. The relation between the transmitted and received digital baseband signals is formulated in a generic channel function. This function is determined from the hardware impairments and wireless communication channel, which are briefly reviewed in this chapter. The baseband processing module is split into bit processing and waveform units. The waveform techniques and their influence on the communication systems design and performance are also investigated in details. Starting from the traditional SC waveforms, multicarrier are represented and the concept is further extended to MIMO waveforms.

Chapter 3 provides a general block-based framework for the representation and processing of all types of linear waveforms. In this model, the waveform is generated by multiplexing finite blocks. These blocks are generated using a generic modulation matrix. On top of that, further waveform techniques such as filtering and windowing are applied. The conventional linear filtering used in SC is reformulated by means of add and overlap. For detection, the received block are expressed in a form of general linear model. Moreover, generic linear and iterative detection are reviewed and evaluated for different modulation matrices and transmission schemes. The evaluation focuses on the theoretical achievable data rate and simulated frame error rate (FER).

Chapter 4 focuses on MA schemes for infrastructure network. Based on the techniques of Chapter 3, a generic flexible MA framework is developed with the consideration of waveforms as a tuning parameter to achieve certain performance with affordable complexity. To show the requirement of different waveforms, asynchronous MA is studied and evaluated. Moreover, this chapter shows that FDMA schemes used in LTE and 5G are very special cases that fit in the presented framework. A more generalized FDMA concept is introduced and extended to TDMA paving the way to study the concept of time-frequency modulation in Chapter 5.

Chapter 5 Reformulates the generic modulation in form of time-frequency structure where a 2-D modulation block is presented and expressed by means of subsymbols and subcarriers. This reformulation is studied in relation to LTV channels. The concrete numerology design of the length of subcarriers and subsymbols is studied in details based on the channel selectivity. Afterwards, the corresponding signal model for detection are derived with concrete representation of the involved channel.

Chapter 6 introduces a special case of time-frequency modulation, where the modulation matrix is generated from a prototype pulse by means of time and frequency shifts. These shifts can be arbitrary in general. A particular study case, where the shifts are equally-spaced, corresponds to Gabor frame. The relation of Gabor frame to conventional GFDM is studied in details showing the role of GFDM modulation structure in developing flexible multicarrier framework. Accordingly, this chapter provides through details on GFDM representations and modem structure. An advanced interpretation of GFDM

inspires the development of the GFDM framework in Chapter 7.

Chapter 7 develops the concept of modulation classes based on the SVD decomposition of modulation matrix. A class of modulation is defined by the structure of the left and right singular vectors, whereas the singular values are programmable parameters. The flexibility of modulation class is studied in terms of the number of configurable parameters other than the singular values. Conventional GFDM is one of the modulation classes that employs DFT transforms. This chapter focuses on the implementation of GFDM modems, with the scope of providing programmable hardware architectures that allow wide range of waveform processing. Based on that, extended GFDM framework is presented, which has more degrees of freedom added by means of simple bypass functions and memory indexing or by replacing the DFT transforms, e.g., with WHT. The developed hardware architecture is evaluated in terms of resource consumption, latency and throughput. Finally, the application of the developed modem is discussed for practical systems.

Chapter 8 concludes the thesis and provides outlooks for future works.

Chapter 2

Fundamentals of Wireless Communications

In the era of digitization, the information is represented in a form of bit streams generated by applications, such as images, data files, audio, etc. The goal of communication systems is to reliably convey the information messages from a source to destination over a medium using specific signals. The main function of the transmitter is to generate signals depending on the incoming data, whereas the receiver aims at detecting the signals and reproducing the data. The typical signals used for wireless communication are physically continuous, e.g. light, sound, and radio signals. However, performing analog signal processing is not an easy task. To overcome this challenge, the design of modern transceivers is split into digital and analog parts. The digital transceiver performs signal processing on discrete signals, and the analog transceiver is responsible of converting the discrete signals into the corresponding medium signals and vice versa. Different transmission techniques require using specific analog module, but they can reuse the same digital processing.

The objective of this chapter is to derive a generic digital transceiver model considering the effect of analog hardware and transmission medium on the discrete channel. A concrete model is developed for radio frequency (RF) wireless communication system. Section 2.1 provides a brief description of wireless communications system considering varieties of transmission schemes, e.g., multiple-input, multiple-output (MIMO) and multiuser. The point-to-point RF transceiver components are discussed in Section 2.2 to derive the effective single-input, single-output (SISO) discrete channel and the impact of hardware impairments. The baseband transceiver is presented in Section 2.3 with a particular focus on the discrete waveform design aspects and its impact on the RF hardware. In addition, MIMO-waveform is introduced as a generic model for different transmission schemes. Section 2.4 is dedicated to an overview of wireless channel models.

2.1 Wireless communications system

A typical digital communication system consists of a digital and an analog modules. The digital module includes protocol and digital baseband processing units, which can be reused with different analog modules corresponding to different types of channels. The analog module in RF wireless communication involves RF transceiver and antennas, as illustrated in Figure 2.1. The digital-to-analog (D/A) and analog-to-digital (A/D) are achieved using the corresponding converters [Gol05]. In this generic model, U denotes

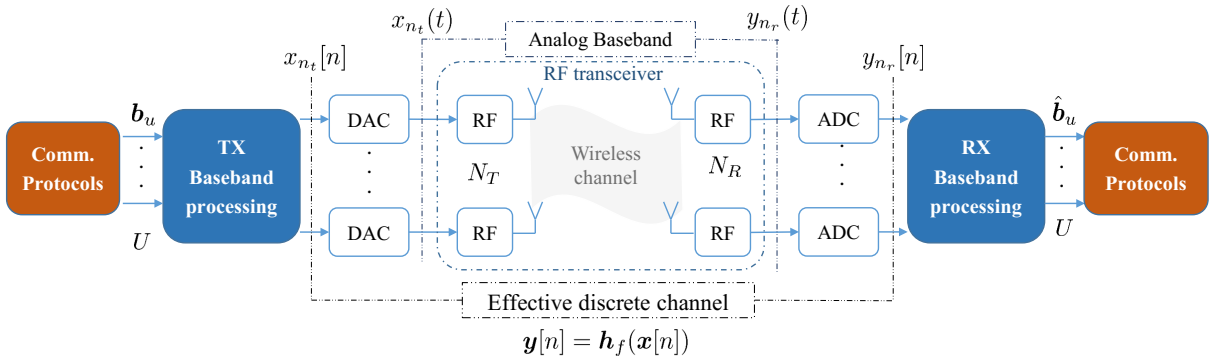

Figure 2.1: Wireless communications system.

the number of users, N_T and N_R the number of transmit and receive RF chains. This model is used to represent different transmission scenarios, including single user ($U = 1$) either in SISO ($N_T = N_R = 1$) or MIMO ($N_T > 1, N_R > 1$) [PGNB04]. Moreover, multiple access (MA) ($U > 1$) in the uplink (UL) and downlink (DL) can be represented in combination with conventional MIMO [SU09, SPSH04], and massive MIMO [NLM13]. Furthermore, it can be used for the schemes that employ different RF configurations, e.g. multi-connectivity, multi-radio access technologies (RAT) or carrier aggregation.

At the transmitter, the protocol unit provides the payload packets $\{\boldsymbol{b}_u\}$ and additional control data used to configure the transmission parameters. The baseband unit generates discrete signals $\{x_{n_t}[n]\}$ to carry the payload. The digital-to-analog (DAC) converts the discrete signal $x_{n_t}[n]$ to analog baseband signal $x_{n_t}(t)$, which is modulated by means of the RF frontend to a carrier frequency and transmitted over the air. The RF chains at the receiver converts the bandpass received signals to baseband signals $\{y_{n_r}(t)\}$. The analog-to-digital (ADC) generates the discrete signal $y_{n_r}[n]$, which is forwarded to the baseband unit for detection. The protocol unit validates the received packets and forwards them to the upper layers and it triggers a retransmission request for erroneous packets.

The effective discrete channel function $\boldsymbol{h}_f(\cdot)$ represents the relation between the discrete transmitted signals represented in the vector $\boldsymbol{x}[n] = [x_1[n], \cdots, x_{N_T}[n]]^T$ and the received signals expressed in the vector $\boldsymbol{y}[n] = [y_1[n], \cdots, y_{N_R}[n]]^T$, such that $\boldsymbol{y}[n] = \boldsymbol{h}_f(\boldsymbol{x}[n])$. In the case of ideal isolation between the RF chains, $y_{n_r}[n] = \sum_{n_t=1}^{N_T} h_{n_r n_t}(x_{n_t}[n])$, where $h_{n_r n_t}(\cdot)$ is the transfer function between the n_r-th receive and n_t-th transmit RF. This is

inherited from the superposition of the electromagnetic signals at the receive antennas [ABB+07]. In realistic system, the coupling effect [WSW04] can be expressed as an additional interference term $z_{n_r}[n]$. Therefore, considering the additive thermal noise $v[n]$

$$y_{n_r}[n] = \sum_{n_t=1}^{N_T} h_{n_r n_t}(x_{n_t}[n]) + z_{n_r}[n] + v_{n_r}[n]. \tag{2.1}$$

To define the individual channels $h_{n_r n_t}(\cdot)$, the RF transceiver model is used to elaborate both the effects of RF impairments and corresponding wireless channel.

2.2 RF transceiver

A simplified block diagram of the RF transceiver between a transmit and receive chain is illustrated in Figure 2.2 [FSCH04]. The transmitter converts a baseband complex signal $x(t) = x_I(t) + jx_Q(t)$ to a bandpass signal $s(t)$ centered at the carrier frequency f_c using quadrature amplitude modulation (QAM) mixer [MMF97]. The signal $x(t)$ results from the D/A conversion of the in-phase (I) and quadrature-phase (Q) discrete signals $x_I[n]$ and $x_Q[n]$, respectively.

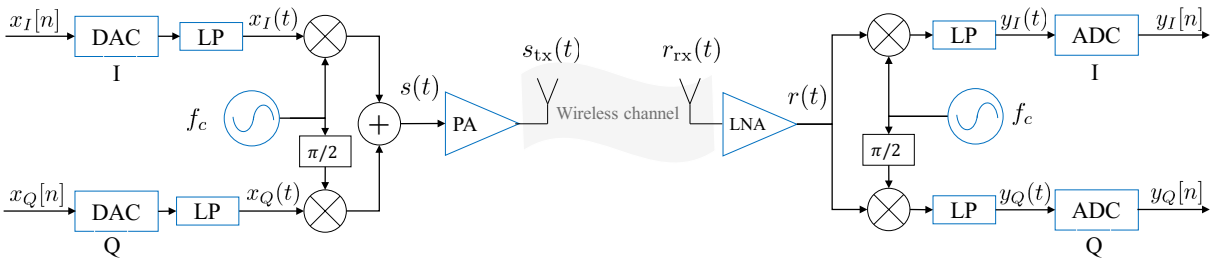

Figure 2.2: RF Transceiver.

2.2.1 Digital-analog conversion

The I and Q DACs convert the discrete sequence to analog baseband signal. Let $x[n]$ be a complex-valued discrete sequence, the corresponding discrete-time signal with sampling interval $T_s = \frac{1}{F_s}$ can be written as

$$x_s(t) = \sum_n x[n]\delta(t - nT_s) \Leftrightarrow \tilde{x}_s(f) = \sum_n x[n]e^{-j2\pi\frac{nf}{F_s}}. \tag{2.2}$$

The conversion is practically realized in several steps as illustrated in Figure 2.3. The DAC holds the input level for a duration equal to the sampling period. The corresponding frequency response of the hold operation is sinc function. A filter with inverse-sinc frequency responded is used in advance to obtain a frequency flat filter. To relax the filter design, it is sufficient to have a flat response within a specific bandwidth relative

to the DAC sampling rate F_{DAC}. Accordingly, the discrete signal is first upsampled to generate samples with rate F_{DAC}. To reject the upsampling images, a digital interpolation filter $\beta_{\text{int}}[n]$ with cutoff at the normalized frequency $\frac{F_s}{F_{\text{DAC}}}$ is used, such that $x_{\text{up}}[q] = \sum_n x[n]\beta_{\text{int}}[q - n\frac{F_s}{F_{\text{DAC}}}]$. Finally, the low-pass (LP) reconstruction filter $\beta_{\text{rec}}(t)$ is used

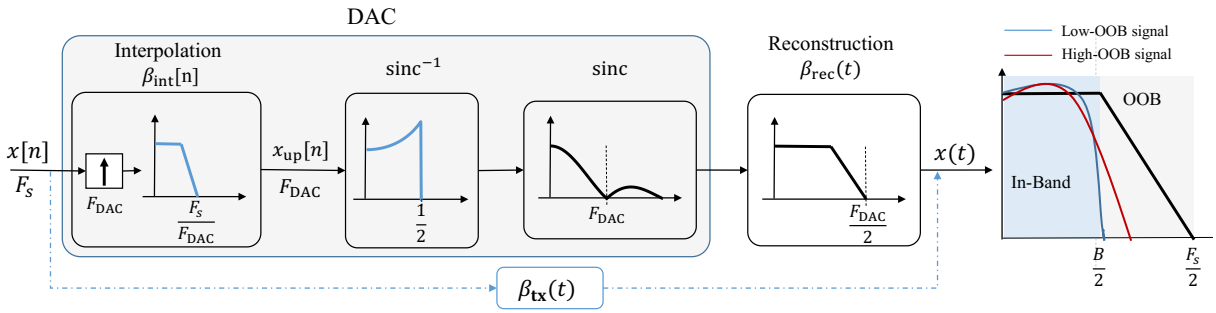

Figure 2.3: DAC operations.

to filter the images of $x_{s,\text{up}}(t) = \sum_q x_{\text{up}}[q]\delta(t - \frac{q}{F_{\text{DAC}}})$, and generate the baseband signal $x(t) = x_{s,\text{up}}(t) * \beta_{\text{rec}}(t)$. The reconstruction filter needs to have a cut-off frequency at $F_{\text{DAC}}/2$ and flat between $[0,\ F_s/2]$. As a result, the baseband signal $x(t)$ can be expressed as

$$x(t) = x_s(t) * \beta_{\text{tx}}(t) = \sum_n x[n]\beta_{\text{tx}}(t - nT_s),\ \tilde{x}(f) = \tilde{x}_s(f)\tilde{\beta}_{\text{tx}}(f), \tag{2.3}$$

where $\beta_{\text{tx}}(\cdot)$ summarizes the transfer function of the DAC. The inter-symbol interference (ISI) free criteria, refers to the case when $x(nT_s + \tau_0) = x[n]$, where τ_0 is a sampling time shift. When $\tilde{x}_s(f)$ is band limited, i.e $\tilde{x}_s(f) = 0, |\nu| < \frac{B}{2} < F_s/2$, there exists a band limited signal $x_0(t)$ that fulfills

$$\tilde{x}_s(f) = \frac{1}{T_s}\tilde{x}_0(f) * \sum_n \delta(f - nF_s) \text{ and } x[n] = x_0(nT_s). \tag{2.4}$$

When $\tilde{\beta}_{\text{tx}}(f)$ is flat for $|f| < \frac{B}{2}$ with linear phase, i.e. $\tilde{\beta}_{\text{tx}}(f) = e^{j2\pi f\tau_0}$, the ISI-free conversion can be achieved because $\tilde{x}(f) = \tilde{x}_0(f)e^{j2\pi f\tau_0}$, $x(t) = x_0(t - \tau_0)$, and therefore, $x(nT_s + \tau_0) = x[n]$. As a result, a proper design of the sequence $x[n]$ with low out-of-band (OOB) plays an important role in achieving ISI-free conversion with relaxed filter design.

The ADC, as shown in Figure 2.4, performs the inverse operations on the received baseband signal $y(t)$ at the output of the QAM demodulator. First the antialiasing filter guarantees the signal is confined within a bandwidth equal to the ADC sampling F_{ADC}. The sampled sequence is down-sampled to the sampling rate F_s, and a decimation LP filter is used to attenuate the signal level outside the expected bandwidth. As a result,

$$y[n] = y(t) * \beta_{\text{rx}}(t)|_{t=nT_s}. \tag{2.5}$$

In ideal channel and proper design of the discrete signal as well as the filters $\beta_{\text{tx}}(t)$ and $\beta_{\text{rx}}(t)$, the output is $y[n] = x[n - n_0]$, where n_0 corresponds to the filtering delay.

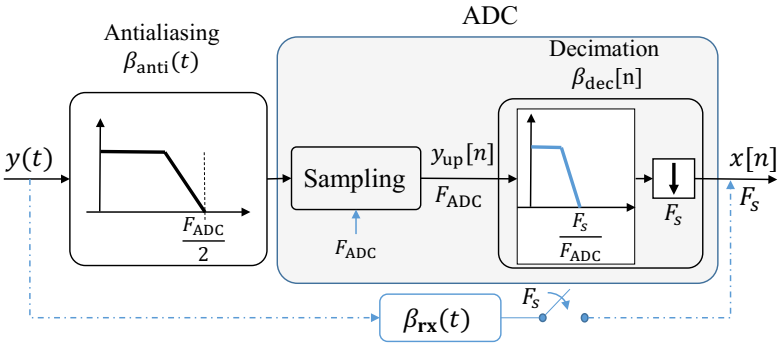

Figure 2.4: ADC operations.

2.2.2 QAM modulation

The baseband signal $x(t)$ of bandwidth B is modulated to a carrier frequency $f_c >> B$ using analog QAM. The ideal bandpass signal $s(t)$ can be expressed as

$$s(t) = x_I(t)\cos(2\pi f_c t) - x_Q(t)\sin(2\pi f_c t) = \text{Re}\left\{x(t)e^{j2\pi f_c t}\right\}. \tag{2.6}$$

This implies an ideal oscillator and exactly $\pi/2$ phase shift between the I and Q components. Then, $s(t)$ is amplified by means of a power amplifier (PA) and fed to the transmit antenna, which can be one element or antenna array [RASPK11]. In ideal case, the PA generates a signal $s_{\text{tx}}(t) = G_{\text{PA}}s(t)$, where G_{PA} is the PA gain.

The receive antenna generates the signal $r_{\text{rx}}(t)$, which is amplified by a low noise amplifier (LNA) resulting in the signal $r(t)$. The QAM demodulator followed by the LP antialiasing filter produces the I and Q components, which are forwarded to the ADC for sampling. Therefore, the input of the LP can be expressed by the virtual signal $y(t) = y_I(t) + jy_Q(t)$,

$$y_I(t) = [r(t)\cos(2\pi f_c t)] * \beta_{\text{LP}}(t), \quad y_Q(t) = -[r(t)\sin(2\pi f_c t)] * \beta_{\text{LP}}(t).$$

Here, $\beta_{\text{LP}}(t)$ represents a virtual ideal LP filter. In ideal channel $y(t) = x(t)$.

2.2.3 Effective channel

Assuming linear multipath channel and ideal analog processing hardware, then

$$r(t) = 2\sum_p \rho_p(t)s(t-\tau_p(t)) = 2\sum_p \text{Re}\left\{\rho_p(t)x(t-\tau_p(t))e^{j2\pi f_c(t-\tau_p(t))}\right\}, \tag{2.7}$$

were $\tau_p(t)$ and $\rho_p(t)$ are the delay and gain of the p-th path, respectively. Note that $\rho_p(t)$ includes the path loss and the gains of PA and LNA. Therefore, the baseband relation can be expressed as linear time variant (LTV) system

$$y(t) = \int_\tau h(\tau,t)x(t-\tau)d\tau, \text{ where } h(\tau,t) = \sum_p \rho_p(t)e^{-j2\pi\tau_p(t)f_c}\delta(\tau-\tau_p(t)). \tag{2.8}$$

Here, $h(\tau, t)$ denotes the channel gain corresponding to the delay τ at the time t. This gain is modeled as a random process as discussed in Section 2.4.

The effective channel is obtained by replacing (2.3) and (2.8) in (2.5)

$$\begin{aligned} y[n] &= \int_{\alpha} \beta_{\text{rx}}(\alpha) \int_{\tau} h(\tau, nT_s - \alpha) \sum_{q} x[q] \beta_{\text{tx}}(nT_s - \alpha - \tau - qT_s) d\tau d\alpha \\ &= \sum_{l} x[n-l] \int_{\alpha} \int_{\tau} \beta_{\text{rx}}(\alpha) h(\tau, nT_s - \alpha) \beta_{\text{tx}}(lT_s - \alpha - \tau) d\tau d\alpha \\ &= \sum_{l} h_e[l, n] x[n-l]. \end{aligned} \tag{2.9}$$

Here, $h_e[l, n]$ is the effective discrete channel assuming ideal hardware,

$$h_e[l, n] = |h_e[l, n]| e^{j(\phi(l,n))} = \int_{\alpha} \int_{\tau} \beta_{\text{rx}}(\alpha) h(\tau, nT_s - \alpha) \beta_{\text{tx}}(lT_s - \alpha - \tau) d\tau d\alpha. \tag{2.10}$$

For example, in multipath channel

$$\begin{aligned} h_e[l, n] &= \sum_{p=1}^{P} \int_{\alpha} \beta_{\text{rx}}(\alpha) h_p(nT_s - \alpha) \beta_{\text{tx}}(lT_s - \alpha - \tau_p) d\alpha \\ &\approx \sum_{p=1}^{P} h_p(nT_s) \int_{\alpha} \beta_{\text{rx}}(\alpha) \beta_{\text{tx}}(lT_s - \alpha - \tau_p) d\alpha \\ &= \sum_{p=1}^{P} h_p(nT_s) \beta(lT_s - \tau_p). \end{aligned} \tag{2.11}$$

Here $\beta(t) = \beta_{\text{rx}}(t) * \beta_{\text{tx}}(t)$, and the approximation $h_p(nT_s - \alpha) = h_p(nT_s)$ results from the assumption that the gain variation within a period of the filter delay is neglected.

2.2.4 Hardware impairments

In practice, the oscillators are subject to phase noise [DMR00]. Moreover, there is a gain difference between the I and Q components and the phase shift between them is different from $\pi/2$ resulting in IQ imbalance [TCVdP+05]. The IQ imbalance causes interference between the I and Q components. Furthermore, the practical PA is non-linear [CMP99], which induces non-linear distortion.

Phase noise. Assuming balanced IQ with phase noise defined by $\phi_{\text{tx}}(t)$ and $\phi_{\text{rx}}(t)$ at the transmitter and receiver, respectively. The actual transmitted baseband signal is $x_o(t) = e^{j\phi_{\text{tx}}(t)} x(t)$. Therefore, $y(t) = e^{-j\phi_{\text{rx}}(t)} \int_{\tau} h(\tau, t) e^{j\phi_{\text{tx}}(t-\tau)} x(t-\tau) d\tau$. In fact, the phase noise can be considered as part of the LTV channel $\bar{h}(\tau, t) = e^{-j\phi_{\text{rx}}(t)} h(\tau, t) e^{j\phi_{\text{tx}}(t-\tau)}$.

IQ imbalance. Let θ_t and A_t be the phase and gain differences at the transmitter such that $s(t) = x_I(t)\cos(2\pi f_c t + \phi_{\text{tx}}(t)) - A_t x_Q(t)\sin(2\pi f_c t + \phi_{\text{tx}}(t) + \theta_t)$. This can be expressed by means of a complex signal $x_{\text{im}}(t)$ as $s(t) = \text{Re}\left\{x_{\text{im}}(t)e^{j2\pi f_c t + j\phi_{\text{tx}}(t)}\right\}$, where

$$x_{\text{im}}(t) = [x_I(t) - A_t \sin(\theta_t) x_Q(t)] + jA_t \cos(\theta_t) x_Q(t).$$

Thus, the relation between the input and output I and Q signals is given by

$$\begin{bmatrix} x_{\text{im},I}(t) \\ x_{\text{im},Q}(t) \end{bmatrix} = \begin{bmatrix} 1 & -A_t \sin(\theta_t) \\ 0 & A_t \cos(\theta_t) \end{bmatrix} \begin{bmatrix} x_I(t) \\ x_Q(t) \end{bmatrix}. \tag{2.12}$$

The transmitter IQ imbalance correction is achieved by multiplying the I and Q by the inverse of the IQ imbalance matrix. Considering this correction at the transmitter and let θ_r and A_r be the phase and gain differences at the receiver

$$\begin{aligned} y_I(t) &= 2\left[r(t)\cos(2\pi f_c t + \phi_{\text{rx}}(t))\right] * \beta_{\text{LP}}(t) \\ &= \sum_p \rho_p(t)\left[x_I(t-\tau_p(t))\cos(\phi_p(t)) - x_Q(t-\tau_p(t))\sin(\phi_p(t))\right] \\ y_Q(t) &= -2\left[A_r r(t)\sin(2\pi f_c t + \phi_{\text{tx}}(t)\theta_r)\right] * \beta_{\text{LP}}(t) \\ &= A_r \sum_p \rho_p(t)\left[x_I(t-\tau_p(t))\sin(\phi_p(t)-\theta_r) + x_Q(t-\tau_p(t))\cos(\phi_p(t)-\phi_r)\right], \end{aligned}$$

where $\phi_p(t) = -2\pi f_c \tau_p(t) - \phi_{\text{rx}}(t) + \phi_{\text{tx}}(t-\tau_p(t))$. As a result

$$\begin{bmatrix} y_I(t) \\ y_Q(t) \end{bmatrix} = \sum_p \rho_p(t) \begin{bmatrix} \cos(\phi_p(t)) & -\sin(\phi_p(t)) \\ A_r \sin(\phi_p(t)-\theta_r) & A_r \cos(\phi_p(t)-\theta_r) \end{bmatrix} \begin{bmatrix} x_I(t-\tau_p(t)) \\ x_Q(t-\tau_p(t)) \end{bmatrix}. \tag{2.13}$$

For ideal mixer where $\theta_r = 0$ and $A_r = 1$, the complex baseband relation (2.8) holds. In practical hardware, the IQ imbalance introduces additional interference term to (2.8). However, the real-valued 2×2-MIMO channel for the I and Q can be considered instead of the complex model. Following the derivation (2.9), the effective discrete channel becomes

$$\begin{bmatrix} y_I[n] \\ y_Q[n] \end{bmatrix} = \sum_l |h_e[l,n]| \begin{bmatrix} \cos(\phi(l,n)) & -\sin(\phi(l,n)) \\ \sin(\phi(l,n)-\theta_r) & \cos(\phi(l,n)-\phi_r) \end{bmatrix} \begin{bmatrix} x_I[n-l] \\ x_Q[n-l] \end{bmatrix}. \tag{2.14}$$

PA non-linearity. In practice, the PA is non-linear [CMP99]. For example in the limiter model, the PA is linear within a signal level range $[-s_0, s_0]$ and saturates outside this range,

$$s_{\text{tx}}(t) = \left\{ \begin{matrix} G_{\text{PA}} s(t), & |s(t)| \le s_0 \\ G_{\text{PA}} s_0, & |s(t)| > s_0 \end{matrix} \right\}.$$

This leads to signal distortion and OOB emission. To alleviate this effect, the signal level $s(t)$ needs to be within the linear range of the PA. The effect of the baseband signal on the PA is studied by means of the peak-to-average power ratio (PAPR). From (2.6), it can be shown that $|s(t)| \le |x(t)|$. To minimize the distortion caused by the PA non-linearity,

it is sufficient to minimize the occurrence of $|x(t)| > s_0$. Assuming that $x(t)$ is a random variable with Gaussian distribution $C\mathcal{N}(0, \sigma_x^2)$, where $\sigma_x^2 = \mathrm{E}\left[|x(t)|^2\right]$, the distribution of $|x(t)|$ is Rayleigh i.e. $f_{x(t)}(x) = \frac{2x}{\sigma_x^2} e^{-\frac{x^2}{\sigma_x^2}}$. Let $x_{\max} = \max_t |x(t)|$, which corresponds to the

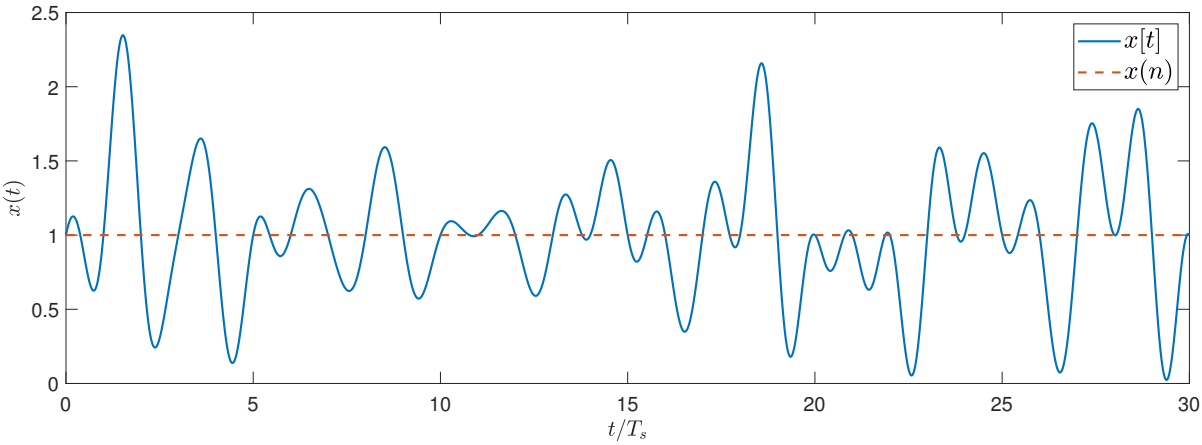

Figure 2.5: PAPR for discrete sequence $|x[n]| = 1$ and $\beta_{\mathrm{tx}}(t) = \mathrm{sinc}\,(F_s t)$.

peak value of $x(t)$, then for $x_{\max} > s_0$,

$$\Pr(|x(t)| > s_0) = \int_{x=s_0}^{x_{\max}} \frac{2x}{\sigma_x^2} e^{-\frac{x^2}{\sigma_x^2}} dx = e^{-\frac{s_0^2}{\sigma_x^2}} - e^{-\frac{x_{\max}^2}{\sigma_x^2}}. \tag{2.15}$$

Based on that, $\Pr(|x(t)| > s_0)$ increases with the increase of the ratio $\frac{x_{\max}^2}{\sigma_x^2}$. This ratio is denoted as the PAPR.

$$\mathrm{PAPR}_x = \frac{\max |x(t)|^2}{\mathrm{E}\left[|x(t)|^2\right]}. \tag{2.16}$$

The discrete signal $x[n]$ and the baseband filter $\beta_{\mathrm{tx}}(\cdot)$ influence the PAPR, as illustrated in Figure 2.5. Noting that $|x(t)| \leq \sum_n |x[n]||\beta_{\mathrm{tx}}(t - nT_s)| \leq \max |x[n]| \sum_n |\beta_{\mathrm{tx}}(t - nT_s)|$, thus, $\max |x[n]| \leq x_{\max} \leq \max |x[n]|\alpha(t)$, where $\alpha(t) = \sum_n |\beta_{\mathrm{tx}}(t - \frac{n}{F_s})|$. This means that the maximum value of the analog signal can be larger than that of the discrete signal. The PAPR is a simple indication, and in practice, the non-linearity can be evaluated by means of numerical simulation of $\Pr(|x(t)| > s_0)$.

Quantization noise. A practical DAC/ADC is limited with the number of bits used to represent the digital signal, and the quantization regions are uniform. Let $\bar{x}$ be the quantized value and ϵ_q the quantization error. The real-valued level x can be expressed as

$$x = \bar{x} + \epsilon_q. \tag{2.17}$$

For uniform DAC/ADC with dynamic range $2A$ and Q_d resolution bits, the quantization step is $\Delta_q = \frac{2A}{2^{Q_d}}$. Therefore, for $x \in [-A, A]$, the quantization error ϵ_q is uniformly distributed within the interval $[-\Delta_q/2, \Delta_q/2]$,

$$\mathrm{E}\left[\epsilon_q^2\right] = \frac{\Delta_q^2}{12} = \frac{A^2}{3} 2^{-2Q_d}. \tag{2.18}$$

For $|x| > A$, the quantization error depends on x, i.e. $\epsilon_q[x]$, with $\mathrm{E}\left[(\epsilon_q[x])^2\right] \geq \mathrm{E}\left[\epsilon_q^2\right]$. The average quantization error is given by

$$\sigma_q^2 = \frac{2A^2}{3} 2^{-2Q_d} \int\limits_{|x|\leq A} f_x(x)dx + \int\limits_{|x|>A} \mathrm{E}\left[(\epsilon_q[x])^2\right] f_x(x)dx \geq \frac{A^2}{3} 2^{-2Q_d}, \tag{2.19}$$

where $f_x(x)$ is the probability distribution function (PDF) of x. If the distribution of x is uniform within $[-A, A]$, where $\mathrm{E}\left[x^2\right] = \frac{A^2}{3}$, the average signal-to-noise ratio (SNR) after the quantization is

$$\mathrm{SNR}_q^{(u)} = 10 \log_{10} \left(\frac{\mathrm{E}\left[x^2\right]}{\mathrm{E}\left[\epsilon_q^2\right]} \right) \approx 6.02 Q_d \text{ dB}. \tag{2.20}$$

To examine other distributions within the range $[-A, A]$, consider $x = A\cos(2\pi t),\ t \in [-1/2, 1/2]$, so that $\mathrm{E}\left[x^2\right] = \int\limits_{-1/2}^{1/2} A^2 \cos^2(2\pi t)dt = \frac{A^2}{2}$. The quantization of this signal produces SNR better by 3/2 , i.e. 1.76 dB, than the uniform case. To improve the SNR, the signal needs to be scaled, offset to the dynamic range. The direct current (DC) offset correction insures that the signal is concentrated around zero. The scaling needs to be adjusted to balance clipping and provide enough resolution for small values.

Based on that, when the hardware impairments is not corrected, the discrete channel can be modeled in general as a LTV system in addition to self-interference term and additive noise, which consists of the quantization noise and the thermal noise, such that

$$y[n] = \sum_l h_e[l, n] x[n - l] + z[n] + v[n]. \tag{2.21}$$

A proper design of the baseband signal can reduce the effects of the hardware impairments.

As a result, the MIMO generic channel model (2.1) can be rewritten as

$$y_{n_r}[n] = \sum_{n_t=1}^{N_T} \sum_l h_{n_r n_t}[l, n] x_{n_t}[n - l] + z_{n_r}[n] + v_{n_r}[n]. \tag{2.22}$$

where $z_{n_r}[n]$ represents all types of hardware impairment interference, and v_{n_r} denotes the overall additive noise. In ideal hardware, $z_{n_r}[n] = 0$ and $v_{n_r}[n]$ is additive white Gaussian noise (AWGN).

2.3 Linear waveform concept and techniques

In conventional digital communication systems, the baseband processing at the transmitter can be split into bit processing, digital mapping, and waveform as depicted in Figure 2.6. The bit processing involves functions such as source coding, encryption, cyclic redundancy check (CRC) check, scrambling, channel coding, and interleaving. This is summarized by the transform $\boldsymbol{c}_u = \mu_u(\boldsymbol{b}_u)$, such that the information bit vector

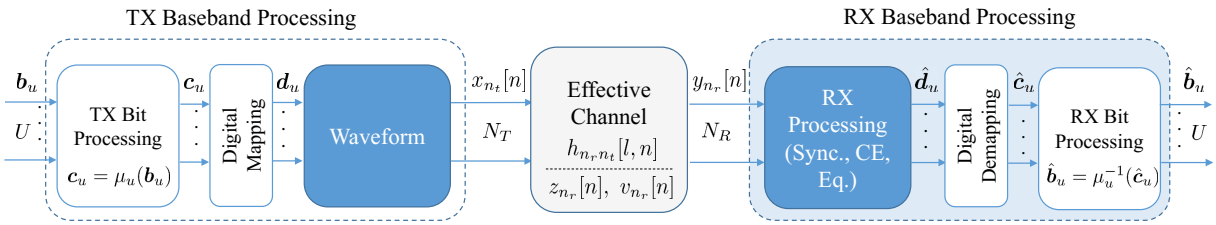

Figure 2.6: Baseband processing functions.

on the u-th stream $\boldsymbol{b}_u \in \mathbb{F}_2^{K_{b,u}\times 1}$ is mapped to a vector $\boldsymbol{c}_u \in \mathbb{F}_2^{K_{c,u}\times 1}$, where $\mathbb{F}_2$ is the Galois field that contains the values $\{0,1\}$. For $K_{c,u} > K_{b,u}$, the code rate is defined by $R_u = \frac{K_{b,u}}{K_{c,u}} \leq 1$. In the digital mapping step, $\boldsymbol{c}_u$ is mapped to a vector $\boldsymbol{d}_u \in \mathcal{M}_{c,u}^{K_{d,u}\times 1}$, where $\mathcal{M}_{c,u}$ is a finite set of $M_{c,u} = 2^{L_{c,u}}$ complex numbers denoted as the constellation set. Here, $L_{c,u}$ is the number of bits per symbol, and therefore, $K_{d,u} = K_{c,u}/L_{c,u}$. The data symbol vectors $\{\boldsymbol{d}_u\}$ are then forwarded to the waveform module to generate the digital baseband signals $\{x_{n_t}[n]\}$ provided to the N_T RF chains.

The baseband processing at the receiver aims at high reliability detection of the transmitted data on each stream, i.e. the estimated bit vector $\hat{\boldsymbol{b}}_u$ needs to be the same as $\boldsymbol{b}_u$ with very high probability. The receiver design follows different approaches to estimate the data from the received signals $\{y_{n_r}[n]\}$. The common receiver processing consists in *synchronization* to determine the start of the signal and to compensate for frequency misalignment, *channel estimation* to calculate the channel gains $h_{n_r n_t}[l,n]$, and *equalization* to inverse the effect of the channel. Afterwards, the equalized symbol vector $\hat{\boldsymbol{d}}_u$ is used thereafter in the decoding. The simplest form of decoding, known as hard decoding, involves binary digital demapping to obtain $\hat{\boldsymbol{c}}_u \in \mathbb{F}_2^{K_{c,u}}$ followed by the inverse bit processing such that $\hat{\boldsymbol{b}}_u = \mu_u^{-1}(\hat{\boldsymbol{c}}_u)$. Another type of decoding, soft decoding provides a reliability measure where $\hat{\boldsymbol{c}}_u \in \mathbb{R}^{K_{c,u}}$, and thus, $\mu_u^{-1}(\cdot)$ is composed of real-value and bit processing. Other receiver design include joint operations, such as joint synchronization and channel estimation [NLLNK09], and joint iterative equalization and decoding [KST04].

The complexity and reliability of the receiver are influenced by the waveform design, i.e. the relation between the generated baseband signal and the data symbols, and how that is reflected to the signal model. At the one hand, the waveform design has an important impact on the hardware impairment defined by the channel interference term $z_{n_t}[n]$ in (2.22), as discussed in Section 2.2.4. On the other hand, the relation with the channel $\sum_l h_{n_r n_t}[l,n] x_{n_t}[n-l]$ can be simplified for equalization using specific techniques. In multiuser scenarios, the waveform has an impact on the inter-user interference (IUI). The following subsections present an overview of conventional waveforms.

2.3.1 Single-carrier waveform

The single-carrier (SC) is the basic of digital modulation waveforms where the data symbols modulates a deterministic pulse analog $g(t)$ [MH81]. The analog form of SC is given by

$$x(t) = \sum_{i\in\mathbb{Z}} d_i g(t - iT) = g(t) * \sum_{i\in\mathbb{Z}} d_i \delta(t - iT). \tag{2.23}$$

Here, d_i denotes the random process of data, T is the duration of the symbol, i.e. $1/T$ is the symbol data rate. The pulse shape $g(t)$ can also be seen as a LP filter applied to the discrete time process $d(t) = \sum_{i\in\mathbb{Z}} d_i \delta(t - iT)$. Assuming the data process is stationary,

$$\forall i, \mathrm{E}\left[d_i\right] = \mu_d,\ \mathrm{E}\left[d_i d_{i-l}^*\right] = R_d[l].$$

This assumption can be fulfilled in practice by proper design of the data source by means of source/channel coding and scrambling. As a result, $x(t)$ is cyclostationary process [GNP06] of period T, because

$$\mu_x(t) = \mathrm{E}\left[x(t)\right] = \mu_d \sum_i g(t - iT), \tag{2.24}$$

$$R_x(t,\tau) = \mathrm{E}\left[x(t)x^*(t-\tau)\right] = \sum_l R_d[l] \sum_q g(t - (l+q)T) g^*(t - \tau - qT). \tag{2.25}$$

Therefore, $\mu_x(t) = \mu_x(t+T)$ and $R_x(t+T,\tau) = R_x(t,\tau)$, are periodic with period T. Moreover, the cyclic autocorrelation at the cyclic frequency k/T is given by

$$R_x^{k/T}(\tau) = \frac{1}{T}\int_{t=0}^{T} R_x(t,\tau) e^{j2\pi\frac{kt}{T}} = \frac{1}{T}\sum_l R_d[l] A_g^{k/T}(\tau - lT). \tag{2.26}$$

$$\text{where } A_g^{k/T}(\tau) = \int_{-\infty}^{\infty} g(t) g^*(t-\tau) e^{-j2\pi\frac{kt}{T}} dt. \tag{2.27}$$

Thus, $R_x^0(\tau) = \frac{1}{T}\sum_l R_d[l] A_g^0(\tau - lT)$, and the power spectral density (PSD) is given by

$$S_x(f) = \frac{1}{T}|g(f)|^2 \sum_l R_d[l] e^{-j2\pi l f T}. \tag{2.28}$$

For uncorrelated data symbols i.e. $R_d[l] = E_s\delta[l]$, where E_s denotes the energy per symbol, $S_x(f) = \frac{E_s}{T}|g(f)|^2$. As a result, the PSD of SC is influenced by the frequency response of the pulse shape, the symbols rate and the energy per symbol. The SC implementation can be achieved using analog filter with simple digital mapping such as direct non-return-to-zero (NRZ) coding of binary signal.

Discrete single-carrier representation. In discrete modulation, a discrete pulse shape $g[n]$ is used to generate the waveform samples $x[n]$, such that

$$x[n] = \sum_i d_i g[n - iN] = g[n] * \sum_i d_i \delta[n - iN], \tag{2.29}$$

where $1/N$ is the normalized symbol rate. Thus, $x[n]$ is cyclostationary with period N,

$$R_x[n,m] = \sum_l R_d[l] \sum_q g[n-[l+q]N]g^*[n-m-qN]. \tag{2.30}$$

Therefore, $R_x^0[m] = \frac{1}{N}\sum_l R_d[l]A_g[m-lN]$, where $A_g[m] = \sum\limits_{n=-\infty}^{\infty} g[n]g^*[n-m]$. The PSD can be expressed as

$$S_x(\nu) = \frac{1}{N}|\tilde{g}(\nu)|^2 \sum_l R_d[l]e^{-j2\pi Nl\nu}. \tag{2.31}$$

For uncorrelated data, $S_x(\nu) = \frac{E_s}{N}|\tilde{g}(\nu)|^2$ which is directly influenced by $g[n]$.

In special case of single-carrier, where $g[n] = \delta[n]$ and for uncorrelated data, $S_x(\nu) = \frac{E_s}{N}$. Thus, the symbol energy is spread equally over the frequency range $[-F_s/2, F_s/2]$, with power E_s/N. The discrete signal is $x[n] = \sum\limits_i d_i\delta[n-iN]$. After DAC conversion and filtering (2.3), the analog signal can be expressed in the form

$$x(t) = \sum_n x[n]\beta_{\text{tx}}(t-\frac{n}{F_s}) = \sum_i d_i\beta_{\text{tx}}(t-\frac{iN}{F_s}). \tag{2.32}$$

This is equivalent to the analog SC with using the LP filter $\beta_{\text{tx}}(t)$, where the symbol rate $T = \frac{N}{F_s}$. In the case of Nyquist filtering, the data symbols can be recovered with sampling rate F_s/N. For $N = 1$, the data symbols, $x[n] = d[n]$, are fed directly to the ADC with a rate F_s, which is the maximum possible symbol rate w.r.t. F_s. The signal bandwidth can be controlled with a low-pass filter (LPF).

Faster-than-Nyquist signaling. When the allowed bandwidth is $B < F_s$, using Nyquist filter is not feasible. However, a general LP filter $g_{\text{tx}}(t)$ can be used, such that

$$x(t) = \sum_i d[i]g_{\text{tx}}(t-\frac{i}{F_s}).$$

The sampled signal in ideal channel,

$$y[n] = \sum_i d_i g_{\text{tx}}(\frac{i-n}{F_s}) = \sum_i g_{\text{tx}}[i]d[n-i].$$

This results in ISI similar to the situation of multipath channel. However, if the interference can be resolved by the receiver, this approach allows the increase of the spectral efficiency. In other words, for the Nyquist case, the maximum spectral efficiency is L_b bits/s/Hz, where L_b is the number of bits per symbol. Whereas the general case may achieve $\frac{F_s}{B}L_b > L_b$. This approach refers to faster-than-Nyquist signaling (FTN) [Maz75], [FGZ+17] while the symbol rate is higher than the signal bandwidth. This concept can be directly applied in the digital processing by means of filtering. For a general discrete signal $x[n]$, a digital finite impulse response (FIR) LP filter $g_{\text{tx}}[n]$ with $\tilde{g}_{\text{tx}}(\nu) = 0$, $\nu > |\frac{B}{2Fs}|$ can be used to confine the signal in the given band. Thus,

$$x_f[n] = x[n] * g_{\text{tx}}[n] = \sum_q g_{\text{tx}}[q]x[n-q].$$

Recalling (2.29), the single-carrier results from the filtering of the random data process, $x[n] = g[n] * \sum_i d_i \delta[n - iN]$. Here the data rate is $\frac{F_s}{N}$, and the bandwidth is B. The FTN principle is achieved when $\frac{F_s}{BN} > 1$. Based on that, the generic form of single-carrier implicitly includes such FTN filtering in the modulation pulse.

2.3.2 Multicarrier waveform

Multicarrier system is a superposition of independent SC waveforms as shown in Figure 2.7. Several pulse shapes are used in the modulation, and the generic form of multicarrier is

$$x[n] = \sum_{k=0}^{K-1} \sum_i d_{k,i} g_k[n - iN_k], \tag{2.33}$$

where $g_k[n]$ is the pulse shape of the k-th subcarrier, $1/N_k$ is the normalized symbol rate.

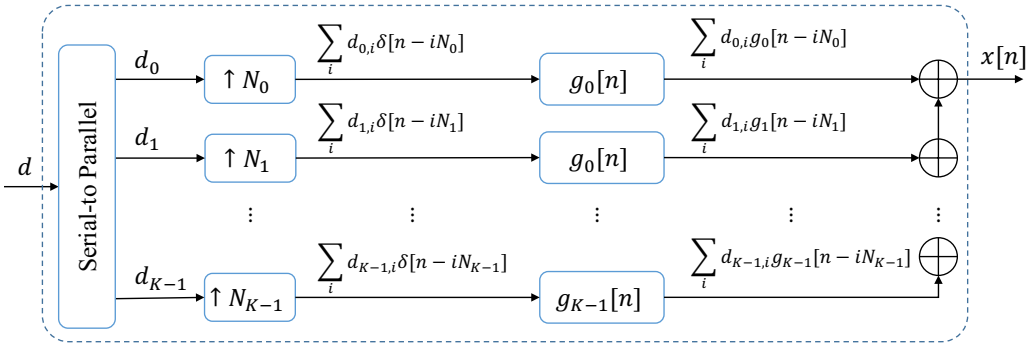

Figure 2.7: Multicarrier waveform.

Assuming the data in each stream are uncorrelated, $\mathrm{E}\left[d_{k,i} d_{q,i}^*\right] = 0, k \neq q$, then using the results from single-carrier, $R_x^0[m] = \sum_{k=0}^{K-1} R_{x,k}^0[m]$, where

$$R_{x,k}^0[m] = \frac{1}{N_k} \sum_l R_{d_k}[l] A_{g_k}[m - lN_k].$$

If the data are uncorrelated in each stream i.e. $R_{d_k}[l] = E_{s_k}\delta[l]$,

$$R_x^0[m] = \sum_{k=0}^{K-1} \frac{E_{s_k}}{N_k} A_{g_k}[m] = \sum_{k=0}^{K-1} \frac{E_{s_k}}{N_k} \sum_n g_k[n] g_k^*[n - m].$$

And therefore,

$$S_x(\nu) = \sum_{k=0}^{K-1} \frac{E_{s_k}}{N_k} |\tilde{g}_k(\nu)|^2. \tag{2.34}$$

Thereby, the PSD of the multicarrier system is a superposition of the PSD of individual single-carrier. Note that the PSD of a subcarrier is scaled with the data rate.

Finite-length pulse shape

Assuming that each of the pulses is of length N_t samples, and they are stacked in a matrix $\boldsymbol{G} \in \mathbb{C}^{N_t \times K}$, where $g_k[n] = \boldsymbol{G}[n,k]$, then (2.34) can be expanded as

$$\begin{aligned} S_x(\nu) &= \sum_{n=0}^{N_t-1} \sum_{q=0}^{N_t-1} \sum_{k=0}^{K-1} \frac{E_{s,k}}{N_k} \boldsymbol{G}[n,k] \boldsymbol{G}^H[k,q] e^{-j2\pi[n-q]\nu} \\ &= \frac{1}{N_t} \sum_{n=0}^{N_t-1} \sum_{q=0}^{N_t-1} \left(\boldsymbol{G} \boldsymbol{\Lambda}^{(R)2} \boldsymbol{G}^H \right) [n,q] e^{-j2\pi[n-q]\nu}. \end{aligned} \tag{2.35}$$

Here, $\boldsymbol{\Lambda}^{(R)} \in \mathbb{R}^{+K \times K}$ with $\boldsymbol{\Lambda}^{(R)2}[k,k] = \sqrt{\frac{N_t}{N_k} E_{s_k}}$ corresponds to the rate and power allocation. Therefore, the PSD depends on $\boldsymbol{R}_g = \boldsymbol{G} \boldsymbol{\Lambda}^{(R)} \boldsymbol{G}^H$. This matrix can be generated as the correlation matrix of the block $\boldsymbol{x} = \boldsymbol{G} \boldsymbol{\Lambda}^{(R)} \boldsymbol{d}$ given that $\mathrm{E}\left[\boldsymbol{d}\boldsymbol{d}^H\right] = \boldsymbol{I}_K$. Consequently, any waveform generated from set of pulse shapes $\boldsymbol{G}$ and allocation matrix $\boldsymbol{\Lambda}^{(R)}$ that produce the correlation $\mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^H\right] = \boldsymbol{R}_g$ has the same PSD.

In the case of correlated subcarrier, $\bar{\boldsymbol{d}}$ represents the data vector with the correlation matrix $\boldsymbol{R}_{\bar{d}} = \mathrm{E}\left[\bar{\boldsymbol{d}}\bar{\boldsymbol{d}}^H\right]$ that implicitly includes the allocation, $\bar{\boldsymbol{G}}$ the pulse matrix, and $\boldsymbol{x} = \bar{\boldsymbol{G}}\bar{\boldsymbol{d}}$. By defining the transformed data vector $\boldsymbol{d} = \boldsymbol{R}_{\bar{d}}^{-1/2}\bar{\boldsymbol{d}}$, which is uncorrelated, and thus $\bar{\boldsymbol{d}} = \boldsymbol{R}_{\bar{d}}^{1/2}\boldsymbol{d}$, the modulated block can be reformulated as

$$\boldsymbol{x} = \bar{\boldsymbol{G}}\bar{\boldsymbol{d}} = \bar{\boldsymbol{G}}\boldsymbol{R}_{\bar{d}}^{1/2}\boldsymbol{d} = \left[\bar{\boldsymbol{G}}\boldsymbol{R}_{\bar{d}}^{1/2}\boldsymbol{\Lambda}^{(R)-1}\right]\boldsymbol{\Lambda}^{(R)}\boldsymbol{d} = \boldsymbol{G}\boldsymbol{\Lambda}^{(R)}\boldsymbol{d}. \tag{2.36}$$

This means that the PSD of a waveform generated from correlated data with the correlation matrix $\boldsymbol{R}_{\bar{d}}$ using pulse matrix $\bar{\boldsymbol{G}}$ is the same as when using a modulation matrix $\boldsymbol{G} = \bar{\boldsymbol{G}}\boldsymbol{R}_{\bar{d}}^{1/2}\boldsymbol{\Lambda}^{(R)-1}$ and uncorrelated data with allocation $\boldsymbol{\Lambda}^{(R)}$.

2.3.3 MIMO-Waveforms

In MIMO system, the baseband signal $x_{n_t}[n]$ transmitted using the n_t-th RF chain can be generated using multicarrier waveform and precoding, such as space-time coding [LSL03], as illustrated in Figure 2.8.

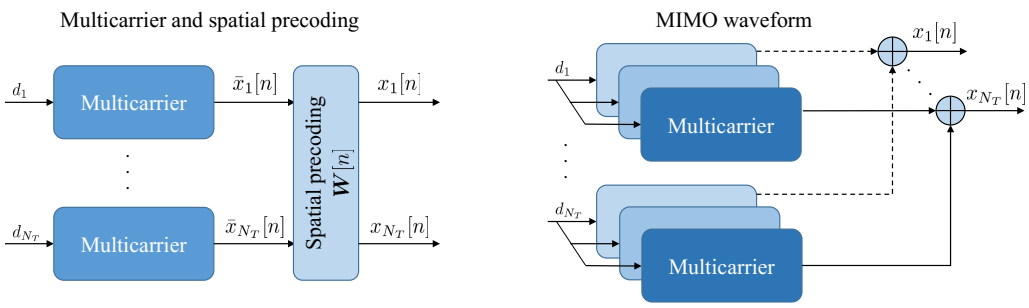

Figure 2.8: MIMO baseband signal.

First, a set of pulses $\{\bar{g}_{n_t,k}[n]\}$ are used to generate the m_t-th multicarrier signal

$$\bar{x}_{m_t}[n] = \sum_{k=0}^{K-1} \sum_{i} d_{m_t,k,i} \bar{g}_{m_t,k}[n - iN_{m_t,k}], \tag{2.37}$$

where $d_{m_t,k,i}$ is the data symbol corresponding to the m_t-th transmitter and k-th subcarrier with normalized rate $1/N_{m_t,k}$. Linear spatial precoding per sample is then applied using precoding weights $\{w_{n_t,m_t}[n]\}$ such that

$$x_{n_t}[n] = \sum_{m_t=1}^{N_T} w_{n_t,m_t}[n] \bar{x}_{m_t}[n] = \sum_{m_t=1}^{N_T} \sum_{k=0}^{K-1} \sum_{i} d_{m_t,k,i} w_{n_t,m_t}[n] \bar{g}_{m_t,k}[n - iN_{m_t,k}].$$

In more general case, the multicarrier and precoding can be realized by means of the pulses $g_{n_t,m_t,k}[n]$, which leads this thesis to define the term MIMO waveform as shown on the right side of Figure 2.8, such that

$$x_{n_t}[n] = \sum_{m_t=1}^{N_T} \sum_{k=0}^{K-1} \sum_{i} d_{m_t,k,i} g_{n_t,m_t,k}[n - iN_{m_t,k}]. \tag{2.38}$$

The pulse $g_{m_t,n_t,k}[n]$ introduces correlation between the data across the subcarriers and the transmitter chains. The power measured at certain location depends on the superposition of the signal from each transmitter. The PSD of the n_t-th transmitter is given by

$$S_{x_{n_t}}(\nu) = \sum_{m_t=1}^{N_T} \sum_{k=0}^{K-1} \frac{E_{s_{m_t,k}}}{N_{m_t,k}} |g_{n_t,m_t,k}(\nu)|^2. \tag{2.39}$$

2.3.4 Waveform performance metrics

Waveform processing can be seen as precoding that controls the correlation of digital samples. This is achieved by the filtering and superposition of uncorrelated data symbol streams. The waveform design aims at designing the pulses such that the generated baseband signal fulfills certain requirements. A suitable waveform design should jointly consider constraints of different performance metrics.

Out-of-band emission

In wireless transmission, the frequency is a shared resource within a given area. Different systems are assigned a finite band and must not introduce interference to neighboring bands. Therefore, the waveform design needs to guarantee low OOB emission. Let $B < F_s$ be the allowed bandwidth, the OOB and in-band power are given by

$$P_{\text{oob}} = P - P_{\text{inb}}, \quad \text{and } P_{\text{inb}} = \int_{f=-\frac{B}{2F_s}}^{\frac{B}{2F_s}} S_x(\nu) d\nu. \tag{2.40}$$

where $P = \int_{\nu=-1/2}^{1/2} S_x(\nu)d\nu$ is the total power. For multicarrier, using (2.34)

$$\begin{aligned} P_{\text{in}} &= \sum_{k=0}^{K-1} \frac{E_{s,k}}{N_k} \int_{\nu=-\frac{B}{2F_s}}^{\frac{B}{2F_s}} |\tilde{g}_k(\nu)|^2 d\nu, \\ P &= \sum_{k=0}^{K-1} \frac{E_{s,k}}{N_k} \int_{\nu=-1/2}^{1/2} |\tilde{g}_k(\nu)|^2 d\nu = \sum_{k=0}^{K-1} \frac{E_{s,k}}{N_k} \sum_n |g_k[n]|^2. \end{aligned} \tag{2.41}$$

The ratio $\gamma_{\text{oob}} = P_{\text{oob}}/P$ can be considered as a metric in the design of the waveform pulses and needs to be minimized. This depends on $\{\tilde{g}_k(\nu)\}$ and $\{\frac{E_{s,k}}{N_k}\}$ which allows more degrees of freedom in the waveform design using multicarrier in comparison with SC. Another important implication of low OOB is to allow asynchronous multiple access considering frequency-division multiple access (FDMA) multicarrier system. The subcarriers need to be well-confined in the frequency domain, where the low OOB emission within one subcarrier allows relaxed synchronizations of the local oscillator (LO) of different users transmitting at the same time on different subcarriers [NEM+19]. This is because such a relatively small frequency offset does not inject high interference. In comparison with orthogonal frequency division multiple access (OFDMA), the subcarriers are orthogonal when they are perfectly synchronized. In this case, the frequency samples of one subcarrier overlap with zero crossing of the side lobes of adjacent subcarriers. Under frequency offset (FO), the side lobe samples interfere with the subcarrier samples, and this becomes worse with the increase of the FO.

Latency

A long pulse shape can be used to fulfill the low OOB requirements. For example, in the ideal single-carrier case, where $|\tilde{g}(\nu)|$ is rectangular, then

$$g[n] = \int_{\nu=-\frac{B}{2F_s}}^{\frac{B}{2F_s}} e^{j2\pi n\nu} d\nu = \text{sinc}\left(\frac{Bn}{F_s}\right).$$

However, this pulse is infinitely long and cannot be made causal. To resolve this, the pulse shape is shortened using a window $w_{N_t}[n] = 1, -N_t/2 \leq n < N_t$ and shifted such that $g[n] = w_{N_t}[n - N_t/2]\text{sinc}\left(\frac{B[n-N_t/2]}{F_s}\right)$. This leads to the frequency response

$$\tilde{g}(\nu) = e^{-j\pi N_t \nu} \int_{u=-\frac{B}{2F_s}}^{\frac{B}{2F_s}} \tilde{w}_{N_t}(\nu - u)du.$$

In the case of rectangular window, $\tilde{w}_{\text{rec},N_t}(\nu) = e^{-j\pi\nu}\frac{\sin(\pi N_t \nu)}{\sin(\pi\nu)}$, and the OOB emission decreases with the increase of N_t. However, this increase comes at the cost of increased

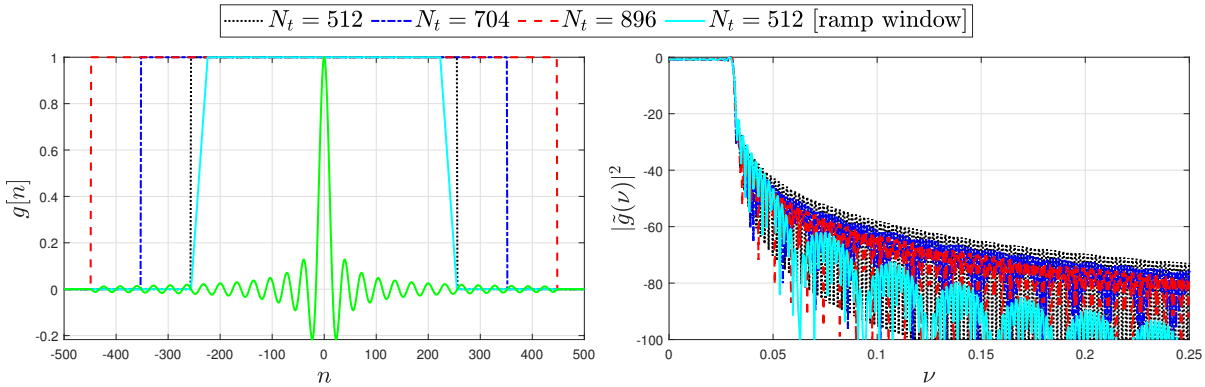

Figure 2.9: OOB for different finite pulse design.

latency for short packet transmission. And it also leads to non-efficient use of the time resources because of long filter tail. Another approach to reduce the OOB is achieved by using a window with lower side lobes, such as ramp-up ramp-down window, as illustrated in Figure 2.9.

PAPR

As discussed in Section 2.2.4, a low PAPR baseband signal $x(t) = \sum_n x[n]\beta_{\text{tx}}(t - nT_s)$ is required to overcome the non-linearity of PA. The goal is to reduce $\Pr(|x(t)| > |s_0|)$, where $|s_0|$ is the saturation level of the PA. For a given power $P = \mathrm{E}\left[|x(t)|^2\right]$, the PAPR can be defined within a signal duration T as [MLG06a]

$$\text{PAPR}(x(t)) = \max_{0 \le t < T} \frac{|x(t)|^2}{P}. \tag{2.42}$$

This form of PAPR is evaluated by means of the complementary cumulative density function (CCDF) defined by

$$\text{CCDF}\left(\frac{|s_0|^2}{P}\right) = \Pr\left(\text{PAPR}(x(t)) > \frac{|s_0|^2}{P}\right) = \Pr\left(\max_{0 \le t \le T} |x(t)]| > |s_0|\right). \tag{2.43}$$

Note that the PAPR depends also on the discrete signal $x[n]$ and the transmitter filter.

The PAPR of multicarrier signal is influenced by the data mapping and the pulses. For instance, this can be illustrated by evaluating the level of the discrete sample

$$|x[n]| = \left|\sum_{k=0}^{K-1} \sum_i d_{k,i} g_k[n - iN_k]\right| \le \sum_{k=0}^{K-1} \sum_i |d_{k,i}||g_k[n - iN_k]|.$$

The more the multicarrier pulses overlap in time, the higher is the PAPR. The worst case arises when the input symbols and the pulses produces maximum super position. As an example, in orthogonal frequency division multiplexing (OFDM), where $N_k = N$ and $g_k[0] = 1/\sqrt{N}$, then $x[iN] = \sum_{k=0}^{K-1} d_{k,i}$, which leads to maximum overlap when all the

data symbols are the same. For SC, $x[n] = \sum_i d_i g[n - iN]$, which results in PAPR that depends on $g[n]$. The highest peaks results because of the overlapping between successive symbols. Figure 2.10 illustrates the results from evaluation metrics of the PAPR of three typical waveforms.

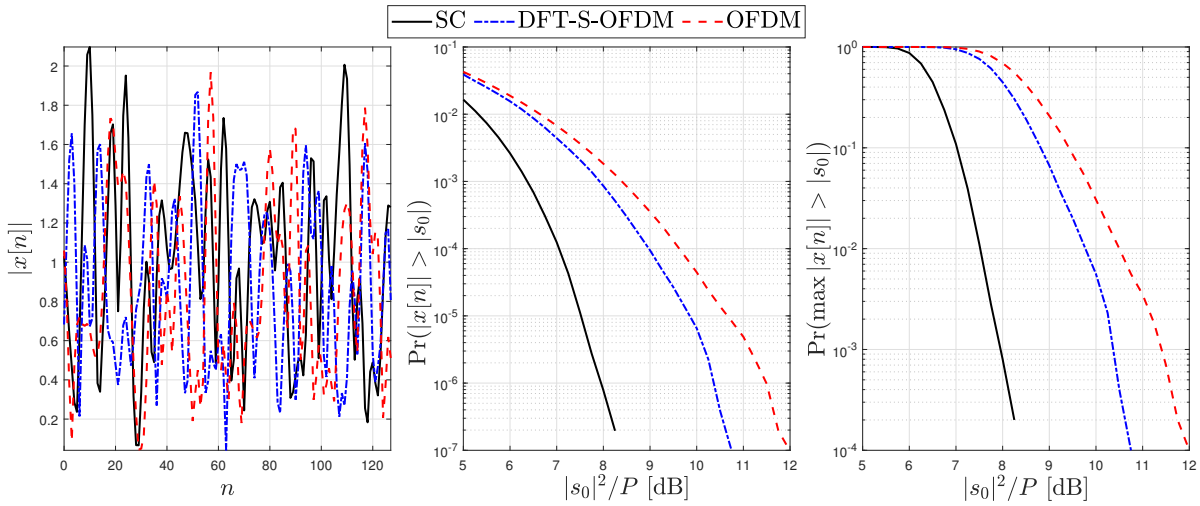

Figure 2.10: PAPR evaluation for different waveforms using QPSK, SC (Dirichlet pulse), DFT-S-OFDM ($M = 32$, $K = 16$ subbands), and OFDM ($N = 1024$, 512 active subcarriers).

Spectral and energy efficiency

The spectral efficiency is measured by the achievable bit rate per bandwidth, and it is measured by bits/s/Hz. It depends on the bandwidth B, the symbol rate per subcarrier $R_{s,k} = \frac{1}{N_k T_s} = \frac{F_s}{N_k}$, and the number of information bits per symbol $L_{b,k}$. Thus,

$$\eta_{\text{SE}} = \sum_{k=0}^{K-1} \frac{F_s}{B N_k} L_{b,k} = \frac{R_b}{B}. \tag{2.44}$$

This ideal efficiency is the upper bound that can be achieved by a single transmitter. The actual efficiency is affected by the realistic channel distortion and ISI, which reduce the ability of the receiver to correctly detect the data. For instance, considering AWGN, the achievable reliable data rate is influenced by the SNR. When the noise power is high, the system needs to increase the signal power, which increases the energy consumption. The energy efficiency measures the number of bits that can be correctly exchanged per Joule. Therefore, for normalized pulses, a given bit rate R_b and power P

$$\eta_{\text{EE}} = \frac{R_b}{P} = \frac{\sum_{k=0}^{K-1} \frac{F_s}{N_k} L_{b,k}}{\sum_{k=0}^{K-1} E_s/N_k}, \text{ where } \sum_n |g_k[n]|^2 = 1. \tag{2.45}$$

The waveform design needs to consider adding pilot signals for synchronization and channel estimation. Moreover, to improve other performance metrics, the data rate per subcarrier might be altered by using silent symbols and guard bands. Furthermore, to tackle the channel frequency selectivity, using cyclic prefix (CP) or zero-padding (ZP) consumes time resources, whereas guard bands are needed to deal with time selectivity. All these requirements affect the spectral and energy efficiency.

Implementation complexity

Linear waveforms utilize linear operations, in form of filtering and precoding. These can be performed by means of matrix multiplication in general. At the receiver, other operations to be considered in form of equalization, channel estimation and demodulation. To optimize the performance of one or two metrics, there could be several possible linear transforms. For instance, as discussed in (2.36), the PSD is influenced by the matrix $\boldsymbol{G}\boldsymbol{\Lambda}^{(R)2}\boldsymbol{G}^H$. Therefore, by using the same data rate per subcarrier $N_k = N$, and using the pulse matrix $\bar{\boldsymbol{G}} = \boldsymbol{G}\boldsymbol{\Lambda}^{(R)}$ instead, the achieved multicarrier system is given in the form

$$x[n] = \sum_{k=0}^{K-1} \sum_{i} d_{k,i} \bar{g}_k[n - iN], \tag{2.46}$$

and accordingly, $\boldsymbol{x}_i = \bar{\boldsymbol{G}}\boldsymbol{d}_i$, $\boldsymbol{x}_i[n] = \sum_{k=0}^{K-1} d_{k,i}\bar{g}_k[n]$. Hence, $x[n] = \sum_i \boldsymbol{x}_i[n - iN]$. Such a waveform that fulfills certain PSD can be implemented in blocks and the blocks are added and overlapped in a simple way. Additionally, any unitary transform of the data achieves the same PSD. However, some matrices may have special structure that involves transforms with fast algorithms such as discrete Fourier transform (DFT) and Walsh-Hadamard. Moreover, special structuring allows simplified channel equalization at the receiver. For example, using CP per block allows single tap frequency-domain (FD) equalization. On the other hand, it might be required from application or channel perspective, that each subcarrier needs to have variable data rate. By fixing the symbol rate, the number of bits per symbol is an alternative way to control the data rate. The problem that arises here is the PAPR when more bits per symbols are required, e.g. with high order QAM. Another approach of controlling the data rate while preserving the PAPR is to set the symbol rate per subcarrier to a maxima and rely on upsampling for rate matching. Thus, the PSD changes depending on the actual used symbol rate.

2.4 Wireless channel

The RF signal reaches the receiver via one or multiple paths depending on the propagation environment, which contains reflectors, scattered and obstacles. The received signal power decays because of the distance between the transmitter and receiver, the characteristic of

the medium material, the carrier frequency, and the pattern of the antennas. First, the channel between a pair of transmit and receive antennas is considered, and the model is extended to MIMO channel.

2.4.1 Line-of-sight propagation

The simplest form of wireless channel is the line of sight (LoS), where the received signal reaches the receiver via a single direct path. Consider a free space transmission, where the transmit antenna is located at the point $(0, d_{y_0})$, and the receive antenna moves along the x-axis with a speed v_x, thus, $d(t) = \sqrt{d_x(t)^2 + d_{y_0}^2}$ as depicted in Figure 2.11. The

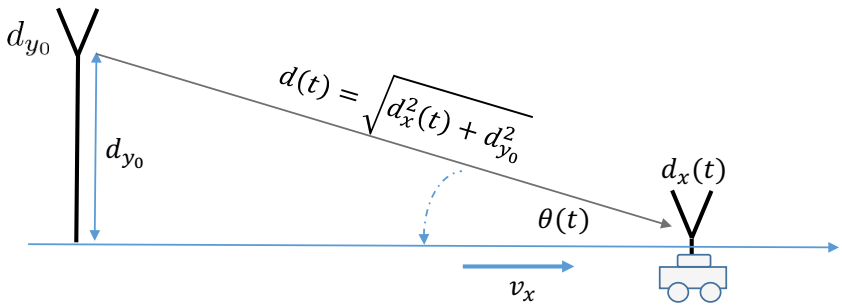

Figure 2.11: LoS channel model.

received signal is delayed by $\tau_1(t) = \frac{d(t)}{c}$, where, $c = 3 \times 10^8$ m/sec is the speed of light. Moreover, the received signal is attenuated by the path loss $\rho_1(t)$. There are several models to describe realistic path loss based on the geometry and the material of the propagation environment[AWC+05]. The simplified path loss relation is given by

$$\rho_1(t) = \sqrt{\frac{P_r}{P_t}} = \sqrt{G_0} \left(\frac{d_0}{d(t)} \right)^{\gamma/2}, \tag{2.47}$$

where P_t and P_r are the transmit and receive power, respectively, d_0 is a reference distance, $\sqrt{G_0}$ is the gain at the distance d_0, which depends on the antennas and the transmission parameters, and γ is the path loss exponent. Obstacles between transmitter and receiver may absorb the signal, which is known as shadowing. The shadowing is modeled with log-normal, i.e. the dB distribution has a Gaussian distribution $\psi_{\text{dB}}(t) = \mathcal{N}(\mu_{\psi_{\text{dB}}}(t), \sigma^2_{\psi_{\text{dB}}}(t))$. Thus, the path gain can be expressed as

$$20 \log_{10} (\rho_1(t)) = 10 \log_{10} (G_0) - 10\gamma \log_{10} \left(\frac{d(t)}{d_0} \right) + \psi_{\text{dB}}(t). \tag{2.48}$$

This model results from the assumption that the signal suffers from independent random attenuates $\bar{\rho}_n$. Thus, the overall attenuation $\rho_1 = \prod_{n=1}^{N} \bar{\rho}_n$, is a random variable of corresponds to the product of a large number of positive random variables. And based on the central limit theorem the distribution of $\log(\rho_1) = \sum_{n=1}^{N} \log(\gamma_n)$ is Gaussian. Both, the

shadowing and path loss effects happen within a large variation of the distance between the transmitter and receiver.

The wireless baseband channel (2.8) in the case of single path can be expressed as

$$h(\tau, t) = \rho_1(t)e^{j\phi_1(t)}\delta(\tau - \tau_1(t)), \tag{2.49}$$

where $\phi_1(t) = -2\pi f_c \tau_1(t)$. The change with time is influenced by the relative speed between the transmitter and receiver. The distance $d(t) = \sqrt{d_x(t)^2 + d_{y_0}^2}$ can be approximated using the first order Taylor approximation as

$$d(t) \approx d_0 + \cos(\theta) v_x t, \tag{2.50}$$

where $d_0 = \sqrt{d_x(0)^2 + d_{y_0}^2}$, $\cos(\theta) = \frac{d_x(0)}{d_0}$. This approximation is valid for a short time interval Δt, whose length depends on the speed v_x, and the angle of arrival. When $\frac{\cos(\theta) v_x t}{d0} << 1$, which means that either the speed is slow or the time interval is short, then, $\rho_1(t) \approx \rho_1(0) = \rho_1$, and $\tau_1(t) = \frac{d(t)}{c} \approx \frac{d_0}{c} = \tau_1$. This indicates that the gain and delay can be fixed. In addition, $-\phi_1(t) = 2\pi f_c \frac{d(t)}{c} \approx 2\pi f_c \tau_1 + 2\pi f_{D_1} t$, where $f_{D_1} = \frac{f_c v_x \cos(\theta)}{c}$ is the Doppler shift. As a result, for a short transmission interval Δt , the equivalent wireless channel can be expressed in the form

$$h(\tau, t) = e^{-j2\pi f_{D_1} t} \rho_1 e^{-j2\pi f_c \tau_1} \delta(\tau - \tau_1) = h_1(t)\delta(\tau - \tau_1). \tag{2.51}$$

Therefore, this channel is defined by a delay τ_1 and complex gain $h_1(t)$, which is fully defined by the parameters ρ_1, τ_1, and f_{D_1}. After certain time, these parameters need to be updated considering new initial distance, changes of the speed, and variant shadowing.

2.4.2 Multi path and fading process

The signal reaches the receiver through different path due to reflection and scattering. In rich scattering area, the path associated with a certain delay is a superposition of signals reflected from many scatters. The path delays can be clustered with P resolvable paths, each associated with a delay τ_p such that

$$h(\tau, t) = \sum_{p=1}^{P} \sum_{i} \rho_{p,i}(t) e^{-j\phi_{p,i}(t)} \delta(\tau - \tau_p) = \sum_{p=1}^{P} h_p(t)\delta(\tau - \tau_p), \tag{2.52}$$

Based on the central limit theorem, and assuming that $\rho_{p,i}(t)$ and $\phi_{p,i}(t)$ are independent, then $h_p(t) = \sum_i \rho_{p,i}(t) e^{-j\phi_{p,i}(t)} = \rho_p(t) e^{-j\phi_p(t)}$ can be modeled as a stationary complex Gaussian random process, where the amplitude $\rho_p(t)$ is Rayleigh-distributed with average power $\mathcal{P}_l$, and $\phi_p(t)$ has uniform distribution

$$f_{\rho_p(t)}(x) = \frac{2x}{\mathcal{P}_p} e^{-x^2/\mathcal{P}_p}, \ x > 0. \tag{2.53}$$

When one path is stronger than the others, e.g. LoS, the distribution becomes Rician,

$$f_{\rho_p(t)}(x) = \frac{1}{2\mathcal{P}_{p,\mathrm{NL}}} \exp\left[-\frac{x^2+\alpha_p^2}{\mathcal{P}_{p,\mathrm{NL}}}\right] I_0\left(\frac{\alpha_p x}{2\mathcal{P}_{p,\mathrm{NL}}}\right), \; x > 0. \tag{2.54}$$

Here, α_p^2 is the power of the strongest path and $\mathcal{P}_{p,\mathrm{NL}}$ the power of other paths. Moreover, $I_0(\cdot)$ is modified Bessel function of the 0-th order. The overall power is $\mathcal{P}_p = \mathcal{P}_{p,\mathrm{NL}} + \alpha_p^2$.

The variation of the fading process $h_p(t)$ in each channel path is studied with respect to the correlation function $A_{h_p}(\Delta t) = \mathrm{E}\left[h_p(t)h_p^*(t-\Delta t)\right]$.

Jakes's Model. This model is based on uniform scattering environment [DBC93]. It assumes the received signal associated with certain delay reaches the receiver from I multipath components with equal power and uniformly distributed angles θ_i. For a short observation interval,

$$h_p(t) = \sum_{i=1}^{I} \rho_{p,i} e^{-j2\pi f_c \tau_p} e^{-j2\pi \cos(\theta_i) f_{D_p} t}, \tag{2.55}$$

where $f_{D_p} = v_{x_p} \frac{f_c}{c}$ is the maximum Doppler frequency, and $\mathrm{E}\left[|\rho_{p,i}|^2\right] = \mathcal{P}_p/I$. For $I \to \infty$,

$$A_{h_p}(\Delta t) = \mathrm{E}\left[h_p(t)h_p^*(t-\Delta t)\right] = \mathcal{P}_p J_0(2\pi f_{D_p}\Delta t). \tag{2.56}$$

Here, $\mathcal{P}_p$ is the average power received in the p-th path, and $J_0(x) = \frac{1}{\pi}\int_0^\pi e^{-jx\cos(\theta)} d\theta$ is Bessel function of the 0-th order. The power spectral density is given by

$$A_{H_p}(f_d) = \left\{ \begin{array}{ll} \frac{\mathcal{P}_p}{\pi f_{D_p}} \frac{1}{\sqrt{1-(f_d/f_{D_p})^2}}, & |f_d| \le f_{D_p} \\ 0, & \text{elsewhere} \end{array} \right\}. \tag{2.57}$$

This can be used to generate the fading process by filtering a Gaussian noise with a filter whose frequency response is equal to $A_{H_p}(f_d)$. Figure 2.12 illustrates an example of one delay path with different Doppler frequencies.

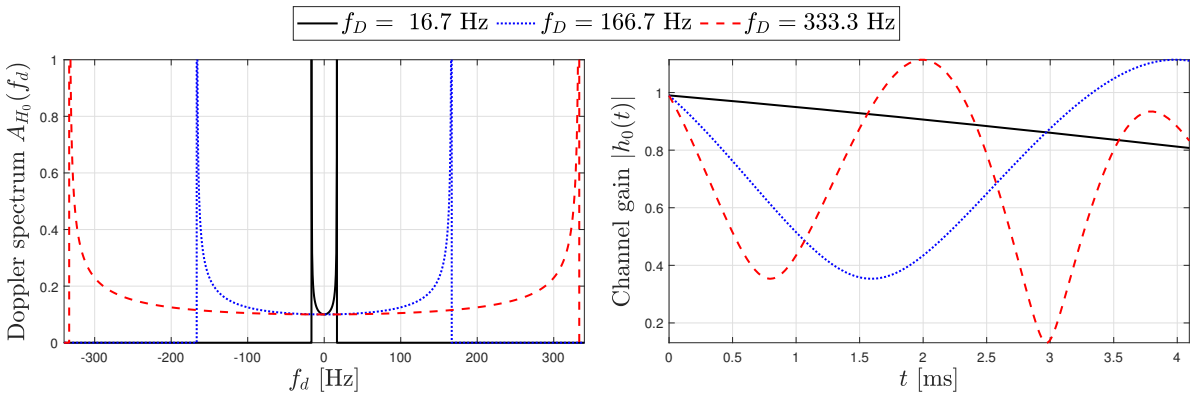

Figure 2.12: Jakes's model for a single path with $f_c = 3$ GHz, $v_x \in \{6, 60, 120\}$ Kmph.

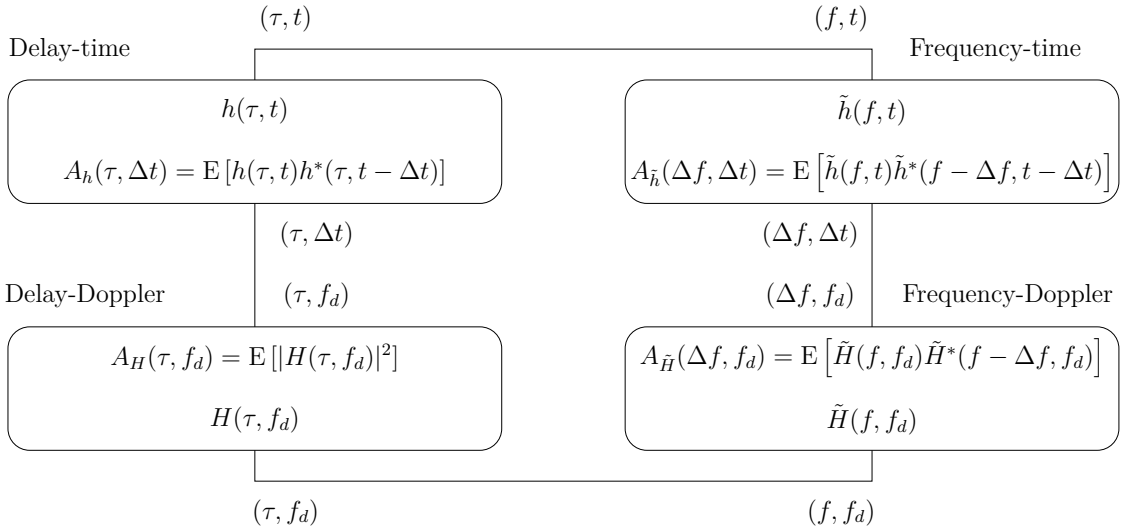

Figure 2.13: Channel representation.

2.4.3 General baseband statistical channel model

In general, the channel coefficient at the delay τ expressed as $h(\tau, t)$ can be modeled with a complex random process [Gol05]. Assuming the channel gains at any delays τ_1 and τ_2 are jointly wide-sense stationary (WSS), the autocorrelation function can be expressed as

$$A_h(\tau_1, \tau_2, \Delta t) = \mathrm{E}\left[h(\tau_1, t)h^*(\tau_2, t - \Delta t)\right]. \tag{2.58}$$

In the case of uncorrelated scattering (US), the processes $h(\tau_1, t)$ and $h(\tau_2, t)$ are uncorrelated for $\tau_1 \neq \tau_2$. Therefore,

$$A_h(\tau_1, \tau_2, \Delta t) = A_h(\tau_1, \Delta t)\delta(\tau_1 - \tau_2). \tag{2.59}$$

The autocorrelation function $A_h(\tau, \Delta t)$ for each delay can be evaluated independently. The channel considering the phase noise $\bar{h}(\tau, t) = e^{-j\phi_{\mathrm{rx}}(t)}h(\tau, t)e^{j\phi_{\mathrm{tx}}(t-\tau)}$ follows the same model. This is because the phase noise can be modeled with stationary process [DMR00] which is independent of $h(\tau, t)$, such that $\phi_x(t) = 2\pi f_x t$, where f_x is uniformly distributed random variable. Accordingly, the phase noise process $e^{j\phi_x(t)}$ is stationary process.

Power delay profile and Doppler spectrum. The power delay profile (PDP) represents the average power corresponding to each delay, i.e. $P_h(\tau) = A_h(\tau, 0) = \mathrm{E}\left[|h^*(\tau, t)|^2\right]$. Due to the WSS property, the average power is independent of time. The *delay spread* $\tau_{\max}$ is the shortest delay, where $P_h(\tau) \approx 0$, $\tau \geq \tau_{\max}$. Another definition is the rms delay spread computed as $\tau_{\mathrm{rms}} = \mathrm{var}(\tau)$ using the delay distribution function $f_h(\tau) = \frac{P_h(\tau)}{\int_\tau P_h(\tau)d\tau}$. Nevertheless, to completely define the fading process, it is necessary to provide the autocorrelation $A_h(\tau, \Delta t)$ which is related to the delay-Doppler correlation by the Fourier transform (FT)

$$A_H(\tau, f_d) = \int_{\Delta t} A_h(\tau, \Delta t)e^{-j2\pi f_d \Delta t} = \mathrm{E}\left[H(|\tau, f_d)|^2\right], \tag{2.60}$$

where $H(\tau, f_d) = \int_t h(\tau, t)e^{-j2\pi f_d t}dt$ denotes the delay-Doppler channel coefficients, which are uncorrelated with respect to τ and f_d. Thus, $A_H(\tau, f_d)$ represents the delay-Doppler profile, from which the PDP $P_h(\tau)$ and Doppler spectrum $S_h(f_d)$ are computed as

$$P_h(\tau) = A_h(\tau, 0) = \int_{f_d} A_H(\tau, f_d)df_d, \quad S_h(f_d) = \int_{\tau} A_H(\tau, f_d)d\tau. \tag{2.61}$$

The *Doppler spread* $f_{d_{\max}}$ is the smallest frequency that satisfies $S_h(f_d) \approx 0, |f_d| > f_{d_{\max}}$.

Coherence bandwidth and coherence time. The instantaneous frequency response of the channel is given by $\tilde{h}(f, t) = \int_{\tau} h(\tau, t)e^{-j2\pi f\tau}d\tau$ which is also a stationary random process, for both time and frequency. The autocorrelation is given by

$$A_{\tilde{h}}(\Delta f, \Delta t) = \mathrm{E}\left[\tilde{h}(f, t)\tilde{h}^*(f - \Delta f, t - \Delta t)\right] = \int_{\tau} A_h(\tau, \Delta t)e^{-j2\pi\tau\Delta f}d\tau. \tag{2.62}$$

This autocorrelation determines how the channel frequency-time response correlates over time and frequency. For a given frequency f, the *coherence time* T_c is the time interval where the channel response $\tilde{h}(f, t)$ is highly correlated. For instance, $\frac{\tilde{A}_h(0,\Delta t)}{\tilde{A}_h(0,0)} \geq \frac{1}{2}$, $\Delta t \leq T_c$. Therefore, the channel frequency response can be considered invariant with time such that $\forall f,\ \tilde{h}(f, t) \approx \tilde{h}(f, t_0),\ t \in [t_0, t_0 + \Delta t]$. The coherence time is related to the Doppler spectrum by the approximation $T_c \approx \frac{\alpha_T}{f_{d_{\max}}}$, where α_T is a constant depends on the criteria used to define T_c. Some common values are 1, $\frac{9}{16\pi}$, or $\sqrt{\frac{9}{16\pi}}$ depending on the correlation threshold [Rap02]. Thus, the higher the Doppler spread the shorter is the coherence time. During the coherence time, it is sufficient to estimate the channel in the frequency domain.

For a given time t, the coherence bandwidth B_c is defined by the frequency band on which the channel is flat, for instance using the correlation criteria $\frac{\tilde{A}_h(\Delta f,0)}{\tilde{A}_h(0,0)} \geq \frac{1}{2}$, $\Delta f \leq B_c$. Therefore, $\forall t$, $\tilde{h}(f, t) \approx \tilde{h}(f_0, t),\ f \in [f_0, f_0 + \Delta f]$. The coherence bandwidth is related to the PDP by the approximation $B_c \approx \frac{\alpha_B}{\tau_{\mathrm{rms}}}$, where τ_{rms} is the rms delay spread, and α_B is a constant that has value such as $\frac{1}{5}$ or $\frac{1}{50}$ depending on the correlation threshold [Rap02]. As a result, within the coherence bandwidth, it is enough to estimate the channel in the time domain. Considering both the coherence time, and the coherence bandwidth, one channel coefficient is sufficient to estimate the channel within an interval T_c, B_c.

Block fading. In many systems the data are transmitted in blocks over short frames. When the duration of the frame is clearly shorter than the coherence time, the channel variation over time is negligible. Thus, for transmitting the i-th frame, $h(\tau, t) \approx h_i(\tau)$. The transmission of the next frame occurs after a time larger than the coherence time, with $h_{i+1}(\tau)$ is independent of $h_i(\tau)$. This model is known as block fading, where the multipath channel becomes time-invariant within the transmission interval.

2.4.4 MIMO channel

The MIMO channel is defined by the channel $h_{n_r n_t}(\tau, t)$ between the n_r-th receive and n_t-th transmit antenna [KSP+02]. The spatial correlation assuming joint WSS is given by

$$\mathrm{E}\left[h_{n_{r_1} n_{t_1}}(\tau_1, t) h^*_{n_{r_2} n_{t_2}}(\tau_2, t - \Delta t)\right] = A_h[n_{r_1}, n_{r_2}, n_{t_1}, n_{t_2}](\tau_1, \tau_2, \Delta t). \tag{2.63}$$

When the delays are uncorrelated in all channel pairs, it is sufficient to study the correlation $A_h[n_{r_1}, n_{r_2}, n_{t_1}, n_{t_2}](\tau, \Delta t)$. This is a 4-D array parameterized with $(\tau, \Delta t)$. Using the delay-Doppler correlation, $A_H[n_{r_1}, n_{r_2}, n_{t_1}, n_{t_2}](\tau, f_d)$ can be evaluated for each pair (τ, f_d). By reformulating this matrix to 2-D matrix of size $N_R N_T \times N_R N_T$ such that

$$\boldsymbol{A}_H(\tau, f_d) = \mathrm{E}\left[\boldsymbol{h}(\tau, f_d)\boldsymbol{h}^H(\tau, f_d)\right], \ \boldsymbol{h}(\tau, f_d) = \mathrm{vec}\left\{\boldsymbol{H}(\tau, f_d)\right\}, \tag{2.64}$$

where $\boldsymbol{H}(\tau, f_d) \in \mathbb{C}^{N_R \times N_T}$, $\boldsymbol{H}[n_r - 1, n_t - 1](\tau, f_d) = H_{n_r n_t}(\tau, f_d)$ is the MIMO channel. In the case of independent correlation between the transmit and receive antennas, $A_H[n_{r_1}, n_{r_2}, n_{t_1}, n_{t_2}](\tau, f_d) = A_{H,\mathrm{rx}}[n_{r_1}, n_{r_2}](\tau, f_d) A_{H,\mathrm{tx}}[n_{t_1}, n_{t_2}](\tau, f_d)$, which results in the Kronecker model $\boldsymbol{A}_H = \boldsymbol{A}_{H,\mathrm{tx}} \otimes \boldsymbol{A}_{H,\mathrm{rx}}$. In the case of Rayleigh fading

$$\boldsymbol{H}(\tau, f_d) = \boldsymbol{A}^{1/2}_{H,\mathrm{rx}}(\tau, f_d)\boldsymbol{H}_0\left(\boldsymbol{A}^{1/2}_{H,\mathrm{tx}}(\tau, f_d)\right)^T, \tag{2.65}$$

where the elements of $\boldsymbol{H}_0 \in \mathbb{C}^{N_R \times N_T}$ are independent and identically distributed (i.i.d.) zero-mean circularly-symmetric complex Gaussian (ZMCSCG) with unit variance.

Multipath MIMO fading channel generation. Assuming that each subchannel is multipath channel, $h_{n_r n_r}(\tau, t) = \sum_{p=1}^{P} h_{n_r n_r, p}(t)\delta(\tau - \tau_p)$, with identical delays and Doppler spectrum. The MIMO channel matrix of the p-th path can be expressed by the matrix $\boldsymbol{H}_p(t)$ where $\boldsymbol{H}_p(t)[n_r, n_t] = h_{n_r n_t, p}(t)$. This matrix can be generated by first generating complex random channel $\boldsymbol{H}_{0,p}(t)$ at the time instance t whose elements are independent and drawn according to the fading distribution either Rayleigh or Rician. A spacial Hermitian correlation matrix $\boldsymbol{R}_p = \boldsymbol{R}_p^{\mathrm{H}}$, $\boldsymbol{R}_p \in \mathbb{C}^{N_R N_T \times N_R N_T}$, is applied such that

$$\bar{\boldsymbol{H}}_p(t) = \mathrm{unvec}_{N_R \times N_T}\left\{\boldsymbol{R}_p^{1/2}\mathrm{vec}\left\{\boldsymbol{H}_{0,p}(t)\right\}\right\}. \tag{2.66}$$

Afterwards, the corresponding Doppler filter is used on the elements of $\bar{\boldsymbol{H}}_p(t)$ such that $\boldsymbol{H}_p(t) = A_{h_p}(t) * \bar{\boldsymbol{H}}_p(t)$. For block fading channel, and when the spatial correlation matrix is independent of the path and follows Kronecker model,

$$\boldsymbol{H}_p = \boldsymbol{R}^{1/2}_{\mathrm{rx}}\boldsymbol{H}_{0,p}\left(\boldsymbol{R}^{1/2}_{\mathrm{tx}}\right)^T, \ \boldsymbol{R}_p = \boldsymbol{R}_{\mathrm{tx}} \otimes \boldsymbol{R}_{\mathrm{rx}}. \tag{2.67}$$

2.5 Summary

This chapter has provided a general overview of wireless communication system with a specific focus on RF systems. The input to the system is information bit streams of one or

multiple users. The baseband signal processing employs bit processing to add forward error correction (FEC), and then maps the bits to data symbols selected from a finite set. These symbols are used to modulate deterministic pulses in order to generate discrete baseband signal that are forwarded to one or multiple RF chains. The term MIMO-waveform is used as a generic model, which considers joint multicarrier and MIMO precoding

$$x_{n_t}[n] = \sum_{m_t=1}^{N_T} \sum_{k=0}^{K-1} \sum_{i} d_{m_t,k,i} g_{n_t,m_t,k}[n - iN_{m_t,k}].$$

The RF chains convert the discrete baseband signals to analog bandpass ones to be transmitted over the wireless channel. Taking into account the RF impairments, and the wireless channel, the received baseband signal at the output of a receiver RF chain is

$$y_{n_r}[n] = \sum_{n_t=1}^{N_T} \sum_{l} h_{n_r n_t}[l, n] x_{n_t}[n - l] + z_{n_r}[n] + v_{n_r}[n].$$

This generic model includes interference term $z_{n_r}[n]$, which depends on the hardware, AWGN $v_{n_r}[n]$ corresponding to the thermal noise, and the effective channel impulse response $h_{n_r n_t}[l, n]$, which is in general time-variant. The goal of the receiver is to correctly extract the information data from the received signal. The receiver design is mainly influenced by the waveform and the effective channel, where proper waveform can reduce the effects of the hardware impairments and simplify the relation with the channel. Accordingly, this chapter has introduced a brief overview on wireless channel models, with specific focus on multipath channel and its relation to the effective channel including the RF transmit and receive LP filters. For SISO channel, this is given by

$$h[l, n] \approx \sum_{p=1}^{P} h_p(nT_s) \int_{\alpha} \beta_{\mathrm{rx}}(\alpha) \beta_{\mathrm{tx}}(lT_s - \alpha - \tau_p) d\alpha. \tag{2.68}$$

In addition, the chapter introduces the waveform aspects, including single carrier, multicarrier and MIMO waveforms. It has also presented several waveform metrics including OOB, PAPR, spectral effectively, latency, and implementation complexity. The later can be handled by means of waveform design with finite pulses, which enables digital implementation using matrix operations and fast transforms. The reminder of the thesis is based on finite pulse to develop generic block-based multicarrier waveforms framework.

Chapter 3

Generic Block-based Waveforms

In realistic scenarios, the pulse shapes used in modulation need to be of finite length. For instance, an ideal well-localized pulse in the frequency domain is non-causal and has infinite length. A practical approximation is achieved by cutting the signal on the edges and shifting it to make it causal. The finite pulses are realized as FIR filters in conventional multicarrier waveforms, which are applied on continuous streams of data. This inspires the idea of filter bank implementation used in filtered multi tone (FMT) [CEO99], and filter bank multicarrier (FBMC)[Far11]. Other systems consider block-based transmission, where independent short signals are generated from finite-length data blocks using linear transforms such as discrete Fourier transform (DFT) in orthogonal frequency division multiplexing (OFDM) [Pra04] or circular filtering as in generalized frequency division multiplexing (GFDM) [MMG+14]. This allows exploiting additional block-based processing such as precoding and cyclic prefix (CP) insertion to simplify the channel equalization. However, non-continuous transition between blocks increases the out-of-band (OOB), which requires additional processing to smooth the transitions such as windowing [SA11] and filtering [ZIX+18].

This chapter provides a general model for representing and processing all type of linear waveforms in block-based form. First, in Section 3.1, the linear filtering approach is reformulated by means of adding an overlap, where individual blocks are generated and multiplexed. In addition, a general description of multicarrier waveforms is developed taking into account variable data rate per subcarrier. Based on that model, Section 3.2 introduces different processing techniques and their impact on the waveform. Afterwards, a structured representation of modulator and demodulator is presented in Section 3.3 introducing the concept of core block. Moreover, the equivalent received core block signal model is derived in a simplified matrix form. Furthermore, the structure is extended to serve for multiple-input, multiple-output (MIMO) scenarios. Finally, Section 3.4 is dedicated for detection and general receiver design. The contributions of this chapter are partially reported in the publications [DNM+18], [NEM+19], [NCF18], and [NCF19b].

3.1 Block-based waveform formulation

Consider a single-carrier waveform, with the causal pulse $g[n]$ of length N_t samples, i.e. $g[n] = 0, n \notin \{0, \cdots, N_t - 1\}$. In order to derive the block structure, this work proposes a reformulation of single-carrier waveforms, as have been reported in [NEM+19], such that

$$x[n] = \sum_i d_i g[n - iN] = \sum_i x_i[n - iN], \text{ where } x_i[n] = d_i g[n]. \tag{3.1}$$

Here, $x_i[n]$ of length N_t samples defines the modulated block corresponding to the i-th symbol. The normalized symbol rate $1/N$, which means there is a new data symbol every N samples. In this case, the symbol rate is equivalent to the block rate, as one block contains only one data symbol. The successive blocks overlap by $N_\mathrm{o} = N_t - N$ samples, which means that the last N_o samples from the $(i-1)$-th block needs to be superimposed with first N_o samples of $x_i[n]$, as shown in Figure 3.1.

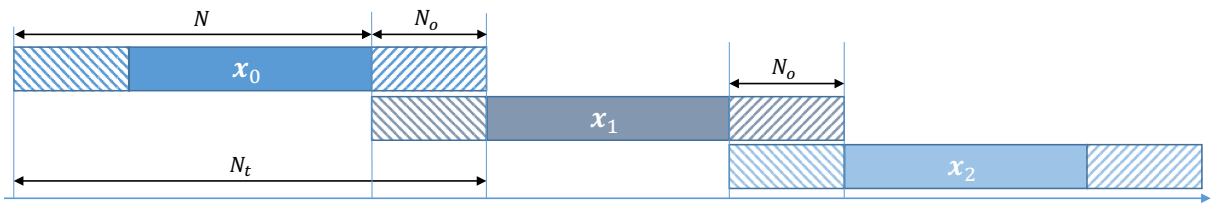

Figure 3.1: Overlapping between blocks.

From implementation perspective, the last N_o samples of of $x_{i-1}[n]$ need to be stored. Thereafter, these samples are added to the first N_o samples of $x_i[n]$ prior to transmission. This stage is denoted as *block multiplexing*. The block overlapping introduces inter-block interference (IBI), which needs handling by the receiver. On the other hand, the blocks can be separated at the transmitter by means of zero-padding (ZP). This leads to the definition of block spacing parameter $N_s = N + N_\mathrm{zp}$. Therefore, the block multiplexing becomes with the form

$$x_\mathrm{zp}[n] = \sum_i x_i[n - iN_s] = \sum_i d_i g[n - iN_\mathrm{s}].$$

With that, the power spectral density (PSD) of $x_\mathrm{zp}[n]$ is scaled by $\frac{N}{N_s}$, i.e. $S_{x_\mathrm{zp}}(\nu) = \frac{N}{N_\mathrm{s}} S_x(\nu)$, which is equivalent to reducing the symbol power.

In multicarrier, if the pulse shapes are all of the same length, and the same symbol rate, such that

$$x[n] = \sum_k \sum_i d_{k,i} g_k[n - iN] = \sum_i x_i[n - iN],$$

then the modulation block is given by

$$x_i[n] = \sum_k \sum_i d_{k,i} g_k[n] \Rightarrow \boldsymbol{x}_i = \boldsymbol{G}\boldsymbol{d}_i. \tag{3.2}$$

Here, $\boldsymbol{d}_i \in \mathbb{C}^{K\times 1}$ defines the data vector, $\boldsymbol{x}_i \in \mathbb{C}^{N_t\times 1}$ the modulation block vector, and $\boldsymbol{G} \in \mathbb{C}^{N_t\times K}$, $\boldsymbol{G}[n,k] = g_k[n]$, denotes the modulation matrix. Therefore, the implementation of linear multicarrier waveform can be achieved using matrix multiplication. However, in generic multicarrier, the symbols rate may differ among the subcarriers, which makes the block presentation difficult, especially at the overlapping stage. In the next section, the generic multicarrier is reformulated with a common symbol rate.

3.1.1 Variable-rate multicarrier

In generic multicarrier, the symbol rate $\frac{1}{N_k}$ can vary in each subcarrier, such that

$$x[n] = \sum_k \sum_j d_{k,j} g_k[n - jN_k].$$

Although the pulse shapes can also have different pulse length $N_{t,k}$, it is possible to have a common pulse length $N_t = \max_k(N_{t,k})$, where $g_k[n] = 0, n = N_{t,k}, \cdots N_t - 1$. The block implementation can be apparently achieved per a group of subcarriers with the same symbol rate. However, the block multiplexing is complicated as it needs to consider different overlapping per subcarrier. Thus, to simplify the overlapping, this work proposes to reformulate the system with a common symbol rate. This rate is selected according to the least common divider of N_k denoted as N, where $N_k = NP_k$ and P_k is positive integer. In other words, the common rate $1/N$ corresponds to the maximum rate, and $1/N_k = \frac{1}{P_k}\frac{1}{N}$ is smaller by P_k times. As a result,

$$x[n] = \sum_k \sum_j d_{k,j} g_k[n - jP_kN] = \sum_k \sum_i \left[\sum_j d_{k,j}\delta[i - jP_k]\right] g_k[n - iN].$$

Let $\bar{d}_{k,i} = \sum_j d_{k,j}\delta[p - jP_k]$, i.e. $\bar{d}_{k,i}$ is the upsampled version of $d_{k,j}$ by factor P_k. In this representation, the symbol rate of $\bar{d}_{k,i}$ is $1/N$, but some symbols are null to match the required rate. As a result,

$$x[n] = \sum_k \sum_i \bar{d}_{k,i} g_k[n - iN].$$

With this representation, the block implementation is achieved by $\boldsymbol{x}_i = \boldsymbol{G}\bar{\boldsymbol{d}}_i$. The entries of $\bar{\boldsymbol{d}}_i$ are filled at each block according to the symbol rate of the subcarriers as illustrated in Figure 3.2. Following the derivation of Section 2.3.2, the PSD is influenced by the symbol rates, and it is given by

$$S_x(\nu) = \frac{1}{N_t} \sum_{n=0}^{N_t-1} \sum_{q=0}^{N_t-1} \left(\boldsymbol{G}\boldsymbol{\Lambda}^{(R)2}\boldsymbol{G}^{\mathrm{H}}\right)[n,q] e^{-j2\pi[n-q]\nu}, \tag{3.3}$$

where $\boldsymbol{\Lambda}^{(R)2}[k,k] = \frac{N_t}{N_k} E_{s,k} = P_k E_{s,k}$ is the power and rate allocation matrix.

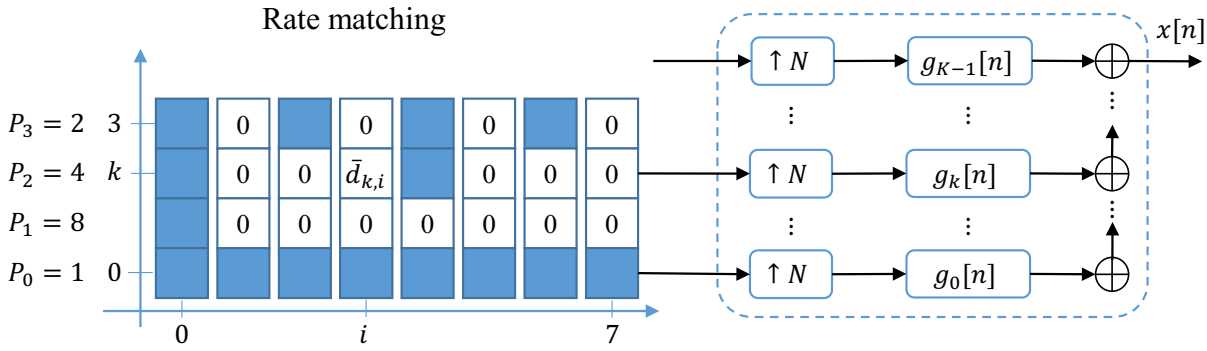

Figure 3.2: Reformulation of multicarrier with variable rates.

Active subcarriers and data mapping. The set of active subcarriers is denoted by the set $\mathcal{K}_{on}$, where $d_k = 0,\ k \notin \mathcal{K}_{on}$. One reason to have non-active subcarriers is to make the matrix $\boldsymbol{G}$ with favorable structure. For instance, in OFDM, the number of subcarrier $K = N$ is used to enable DFT processing, while the edge subcarriers are not used as they do not belong to the allocated band. Another reason is to enable multiple access, where a user signal is allocated to a partial set of the subcarriers. Note that in the case of variable rate, the subcarriers can be temporary inactive in terms of generating the block $\boldsymbol{x}_i$ for some indexes i. Accordingly, a data mapping function is used to map the data to the active subcarriers.

3.1.2 General block-based multicarrier model

The term subcarrier has been historically used in the context of frequency-division multiple access (FDMA) [SB08]. However, it is used in this chapter to refer to the signal corresponding to a given pulse shape. Accordingly, in multicarrier, the number of subcarriers is a generic parameter K that denotes the number of distinguished pulses. The subcarrier index may refer to certain structure, such as time, frequency or space. To clarify that, recall the MIMO waveform proposed in Section 2.3.3 (2.38)

$$x_{n_t}[n] = \sum_{m_t=1}^{N_T} \sum_{k=0}^{K-1} \sum_{i} d_{m_t,k,i} g_{n_t,m_t,k}[n - iN].$$

Here n_t refers to the transmitter chain. Nevertheless, the block representation is given by $\boldsymbol{x}_{n_t,i} = \boldsymbol{G}_{n_t}\boldsymbol{d}_i$. Here, $\boldsymbol{d}_i \in \mathbb{C}^{KN_T \times 1}$, $\boldsymbol{d}_i\,[k + n_t K] = d_{n_t,k,i}$, and $\boldsymbol{G}_{n_t} \in \mathbb{C}^{N_t \times KN_T}$, $\boldsymbol{G}_{n_t}[n, k + n_t K] = g_{n_t,m_t,k}[n]$. In general, this corresponds to multicarrier system with $N_T K$ subcarriers, with two subindexes. The concept can be extended to any number of indexes. In Chapter 5, the waveforms are introduced in details from perspective of time-frequency design, where the concept of time subsymbol is defined. Table 3.1 lists the main parameters of generic block-based waveform.

Table 3.1: Summary of block-based waveform parameters.

Number of subcarriers	K, with $\mathcal{K}_{on}$ active subcarriers.
Symbol rates	$1/N_k$ per subcarrier, $1/N$ common rate.
Block length	N_t corresponds to the longest pulse.
Modulation matrix	$\boldsymbol{G} \in \mathbb{C}^{N_t \times K}$. For MIMO, a modulation matrix $\boldsymbol{G}_{n_t}$ per chain.
Block spacing	N_s, $N_\mathrm{s} > N_\mathrm{t}$ (ZP), $N < N_\mathrm{s} < N_T$ (overlap and rate reduction).
Overlapping	$N_\mathrm{o} = N_t - N_\mathrm{s}$, $N_\mathrm{o} < 0$ corresponds to ZP.
Input data vector	$\boldsymbol{d}_i \in \mathbb{C}^{K \times 1}$. Allocation is performed based on $1/N_k$, $\mathcal{K}_{on}$.
Modulated block	$\boldsymbol{x}_i = \boldsymbol{G}\boldsymbol{d}_i \in \mathbb{C}^{N_\mathrm{t} \times 1}$, $x[n] = \sum_i \boldsymbol{x}_i[n - iN_s]$.

3.2 Waveform processing techniques

Waveform design aims at designing the pulses represented by the matrix $\boldsymbol{G}$ according to the metrics that has been discussed in Section 2.3.4. The techniques used to generate waveforms include filtering, windowing, ZP, and CP/cyclic suffix (CS) are introduced on top of an initial multicarrier signal $x[n] = \sum_i \boldsymbol{x}_i[n - iN] = \sum_k \sum_i d_{i,k} g_k[n - N]$ with non-overlapped blocks, i.e. $N = N_\mathrm{t} = N_\mathrm{s}$, as reported in [DNM+18].

3.2.1 Linear and circular filtering

The filtering can be used to confine the waveform in the allowed bandwidth, and it is also used to model linear multipath channel operation. Let $f[l]$ be a causal finite impulse response (FIR) filter with L_f non-zero samples.

Filtering at the transmitter

The output of the filter can be expressed as

$$x^{(f)}[n] = \sum_{l=0}^{L_f-1} f[l]x[n-l] = \sum_i \sum_{l=0}^{L_f-1} f[l]\boldsymbol{x}_i[n-l-iN] = \sum_i \boldsymbol{x}_i^{(f)}[n-iN], \tag{3.4}$$

where $\boldsymbol{x}_i^{(f)}[n] = \sum_{l=0}^{L_f-1} f[l]\boldsymbol{x}_i[n-l]$ denotes the i-th filtered block of size $N_t = N + L_f - 1$ samples. Thus, the filtered signal $x^{(f)}[n]$ results from the multiplexing of the filtered blocks with overlapping by $N_\mathrm{o} = L_f - 1$ samples. To get rid of the overlapping while preserving the block rate, the last and first $L_f - 1$ of $\boldsymbol{x}_i^{(f)}[n]$ are added together, as shown in Figure 3.3. This process is known as tail biting [MW86], it looks like the head of the block bites the tail. This operation produces a block $\boldsymbol{x}_i^{(\mathrm{c})} \in \mathbb{C}^{N \times 1}$ defined such that

$$\boldsymbol{x}_i^{(\mathrm{c})}[n] = \left\{ \begin{array}{ll} \boldsymbol{x}_i^{(f)}[n] + \boldsymbol{x}_i^{(f)}[N+n], & 0 \le n < L_f - 1 \\ \boldsymbol{x}_i^{(f)}[n], & L_f - 1 \le n < N \end{array} \right\}. \tag{3.5}$$

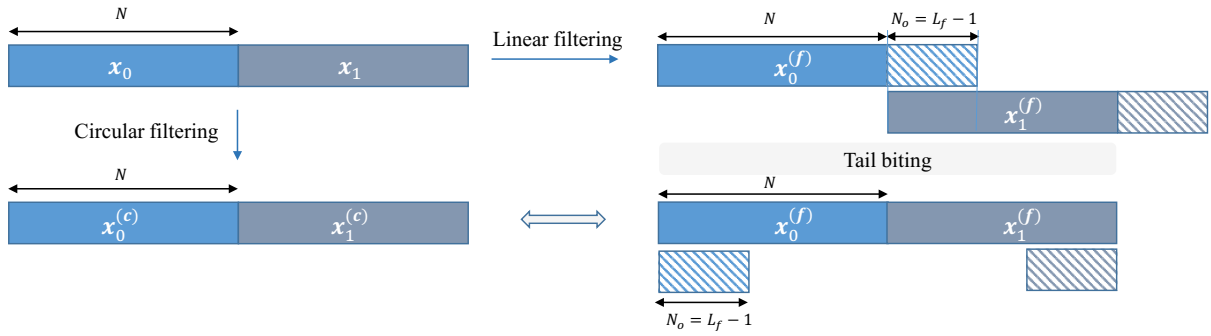

Figure 3.3: Linear and circular filtering.

Note that, for $n = 0, \cdots, L_f - 1$,

$$\boldsymbol{x}_i^{(c)}[n] = \sum_{l=0}^{L_f-1} f[l]\boldsymbol{x}_i[n-l] + \sum_{l=0}^{L_f-1} f[l]\boldsymbol{x}_i[n+N-l] = \sum_{l=0}^{L_f-1} f[l]\boldsymbol{x}_i[\langle n-l\rangle_N].$$

In addition, for $L_f \leq n < N$ $\boldsymbol{x}_i[\langle n-l\rangle_N] = \boldsymbol{x}_i[n-l]$. As a result, the tail biting is essentially another name of circular filtering, which is given by

$$\boldsymbol{x}_i^{(c)}[n] = \sum_{l=0}^{L_f-1} f[l]\boldsymbol{x}_i[\langle n-l\rangle_N]. \tag{3.6}$$

The circular convolution can be implemented by means of N-DFT because

$$\tilde{\boldsymbol{x}}_i^{(c)} = \boldsymbol{F}_N\boldsymbol{x}_i^{(c)} = \sqrt{N}\left[\boldsymbol{F}_N\boldsymbol{f}_N\right] \odot \left[\boldsymbol{F}_N\boldsymbol{x}_i\right], \quad \text{where } \boldsymbol{F}_N[n,q] = \frac{1}{\sqrt{N}}e^{-j2\pi\frac{qn}{N}}.$$

Here, $\boldsymbol{f}_N = [\boldsymbol{f}^T, \boldsymbol{0}_{N-L_f}^T]^T \in \mathbb{C}^{N\times 1}$. Similar approach can be used for the implementation of linear convolution, however, the N_t-DFT needs to be used to generate $\tilde{\boldsymbol{x}}_i^{(f)}$,

$$\tilde{\boldsymbol{x}}_i^{(f)} = \boldsymbol{F}_{N_\mathrm{t}}\boldsymbol{x}_i^{(f)} = \sqrt{N_\mathrm{t}}\left[\boldsymbol{F}_{N_t}\boldsymbol{f}_{N_\mathrm{t}}\right] \odot \left[\boldsymbol{F}_{N_\mathrm{t}}\boldsymbol{x}_{i,N_\mathrm{t}}\right].$$

Here, $\boldsymbol{x}_{i,N_t} = [\boldsymbol{x}_i^T, \boldsymbol{0}_{N_o}^T]^T$. The DFT size in this case depends on the filter length. As a results, block overlapping is the essential difference between circular and linear filtering. In both cases, the resulting pulse shapes are defined by

$$\begin{aligned} g_k^{(f)}[n] &= f[n] * g_k[n] = \sum_{l=0}^{L_f-1} f[l]g[n-l], \\ g_k^{(c)}[n] &= f[n] \circledast g_k[n] = \sum_{l=0}^{L_f-1} f[l]g[\langle n-l\rangle_N] \end{aligned} \tag{3.7}$$

Multipath channel filter

By transmitting $x[n]$ through frequency selective channel, which acts as a filter with the impulse response $h[l]$, $0 \leq l < L$, where L is the number of discrete paths, the received

block is $\boldsymbol{y}_i[n] = \sum_{l=0}^{L-1} h[l]\boldsymbol{x}_i[n-l]$ that has $N+L-1$ samples. Thus, the received signal $y[n] = \sum_i \boldsymbol{y}_i[n-iN]$ is multiplexed with overlap $N_o = L-1$, and the channel introduces IBI on the first and last $L-1$ samples.

In order to prevent the interference, ZP can be used at the transmitter [MZG+02], such that the block spacing becomes $N_s = N+L-1$, i.e. $x^{(zp)} = \sum_i \boldsymbol{x}_i[n-N_\mathrm{s}]$, and using N_s-DFT, the frequency-domain (FD) block becomes

$$\tilde{\boldsymbol{y}}_i^{(zp)} = \sqrt{N_\mathrm{s}}\,[\boldsymbol{F}_{N_\mathrm{s}}\boldsymbol{h}_{N_\mathrm{s}}] \odot [\boldsymbol{F}_{N_\mathrm{s}}\boldsymbol{x}_{i,N_\mathrm{s}}]\,.$$

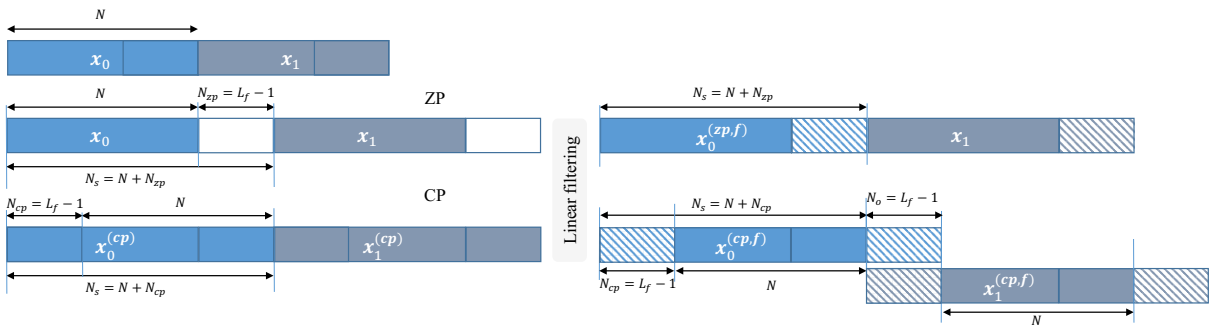

Figure 3.4: Block with ZP/CP.

Tail biting technique can be used to convert the linear convolution with the channel to a circular one, such that

$$\boldsymbol{y}_i^{(\mathrm{c})}[n] = \left\{ \begin{array}{ll} \boldsymbol{y}_i^{(zp)}[n] + \boldsymbol{y}_i^{(zp)}[N+n], & 0 \le n < L-1 \\ \boldsymbol{y}_i^{(zp)}[n], & L-1 \le n < N \end{array} \right\}.$$

As a result $\boldsymbol{y}_i^{(\mathrm{c})}[n] = \sum_{l=0}^{L-1} h[l]\boldsymbol{x}_i[\langle n-l\rangle_N]$. In this approach, the DFT size is fixed to N. However, the drawback is that the noise variance at the indexes $0 \le n < L$ is duplicated when considering additive white Gaussian noise (AWGN) channel. An alternative IBI free solution without noise enhancement can be achieved using CP, at the cost of spending energy on transmitting guard samples. The ZP and CP concepts are illustrated in Figure 3.4

Cyclic prefix

The circular convolution w.r.t. channel can be maintained at the receiver by means of adding CP of length larger than the channel delay, $N_\mathrm{cp} \ge L-1$, such that

$$\boldsymbol{x}_i^{(\mathrm{cp})}[n] = \boldsymbol{x}_i[\langle n - N_\mathrm{cp}\rangle_N >],\ n = 0 \cdots N_t - 1. \tag{3.8}$$

Thus, $x_i^{(\mathrm{cp})}[n] = \sum_i \boldsymbol{x}^{(\mathrm{cp})}[n - iN_s]$, and $y^{(\mathrm{cp})}[n] = \sum_{l=0}^{L-1} h[l]x^{(\mathrm{cp})}[n-l]$. The i-th received block can be extracted by removing the CP, which is equivalent to

$$\begin{aligned} \boldsymbol{y}_i[n] &= y^{(\mathrm{cp})}[n + N_{\mathrm{cp}} + iN_{\mathrm{s}}] = \sum_{l=0}^{L-1} h[l]\boldsymbol{x}_i^{(\mathrm{cp})}[n + N_{\mathrm{cp}} - l], n = 0, \cdots N-1 \\ &= \sum_{l=0}^{L-1} h[l]\boldsymbol{x}_i[\langle n - l\rangle_N]. \end{aligned} \tag{3.9}$$

Adding a CP allows the FD equalization using N-DFT without any enhancement of the noise while $\tilde{\boldsymbol{y}}_i = \sqrt{N}\tilde{\boldsymbol{h}}_N \odot \tilde{\boldsymbol{x}}_i$. Note that, when linear filtering is applied at the transmitter, the effective channel becomes $h_e[l] = h[l] * f[l]$ of delay $L_e = L + L_f - 1$, and thus, to maintain the circular filtering it is required that $N_{\mathrm{cp}} \geq L_e - 1$.

Windowing

The equivalent pulses after adding the CP can be expressed as

$$g_k^{(\mathrm{cp})}[n] = g_k[\langle n - N_{\mathrm{cp}}\rangle_N] = U_{N_t}[n]g_k^{(p)}[n - N_{\mathrm{cp}}], \tag{3.10}$$

where $U_{N_t}[n] = 1,\ 0 \leq n < N_t$ is a rectangular window, and $g_k^{(p)}[n]$ is periodic pulse with period N generated from $g_k[n]$. Thus, the modulated block can be expressed by means of windowed periodic block,

$$\boldsymbol{x}_i^{(\mathrm{cp})}[n] = U_{N_t}[n] \sum_k d_{k,i} g_k^{(p)}[n]. \tag{3.11}$$

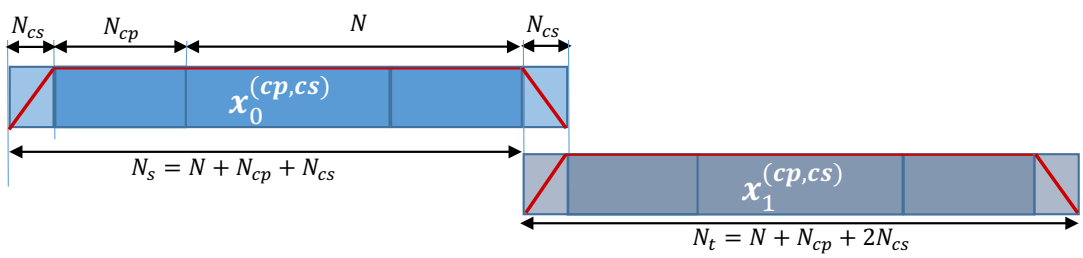

Figure 3.5: Block with ZP/CP.

In general, the window does not need to be rectangular. Windowing can be applied to control smooth transitions between blocks in order to reduce the OOB emissions. To keep the advantages of CP in FD equalization, the window function $w_{N_t}[n]$ is designed to be flat in the middle over $N + N_{\mathrm{cp}}$ samples, and the other ends are designed to decrease the side lobes in comparison with the rectangular window. For that, the CP is extended by N_{cs} samples, and a CS with N_{cs} is added. Thus, the block length becomes $N_{\mathrm{t}} = N + N_{\mathrm{cp}} + 2N_{\mathrm{cs}}$. Moreover, to reduce the overhead, overlapping by $N_{\mathrm{o}} = N_{\mathrm{cs}}$ is introduced between the windowed blocks such that the block spacing is $N_s = N + N_{\mathrm{cp}} + N_{\mathrm{cs}}$, as depicted in

Figure 3.5. This approach is known as windowing and overlapping (WOLA) [ZMSR16]. The windowed multicarrier signal can be expressed in the form

$$x^{(\mathrm{cp,\ w})}[n] = \sum_i \sum_k d_{k,i} w_{N_t}[n - iN_\mathrm{s}] g_k^{(p)}[n - iN_\mathrm{s}] = \sum_i \sum_k d_{k,i} g_k^{(\mathrm{cp,\ w})}[n - iN_\mathrm{s}]. \tag{3.12}$$

Here, $g_k^{(\mathrm{cp,\ w})}[n]$ is the equivalent pulse of the k-th subcarrier. In the next section, all these techniques are integrated in a form of generic multicarrier framework.

3.3 Generic multicarrier modem structure

Considering all the linear operations involved in generating the pulses, this thesis proposes a generic linear block-based multicarrier framework as reported in [NCF18]. The overall linear transform corresponds to a matrix whose columns represent the finite pulse shapes.

3.3.1 Modulator

In section 3.1.2, it is shown that the waveform can be generated from the data symbols by means of three functions:

1. Data mapping: the data carried by the i-th block are mapped to $\boldsymbol{d}_i^{(\mathrm{on})} \in \mathbb{C}^{K\times 1}$ taking into account the active subcarriers and the symbol rate per subcarrier.
2. Block generation: the pulse shaping matrix $\boldsymbol{G} \in \mathbb{C}^{N_\mathrm{t}\times K}$ is used to generate the block $\boldsymbol{x}_i^{(t)} = \boldsymbol{G}\boldsymbol{d}_i^{(\mathrm{on})} \in \mathbb{C}^{N_\mathrm{t}\times 1}$.
3. Block Multiplexing: the data block are multiplexed with spacing N_s such that, the waveform signal is $x[n] = \sum_i \boldsymbol{x}_i^{(t)}[n - iN_\mathrm{s}]$.

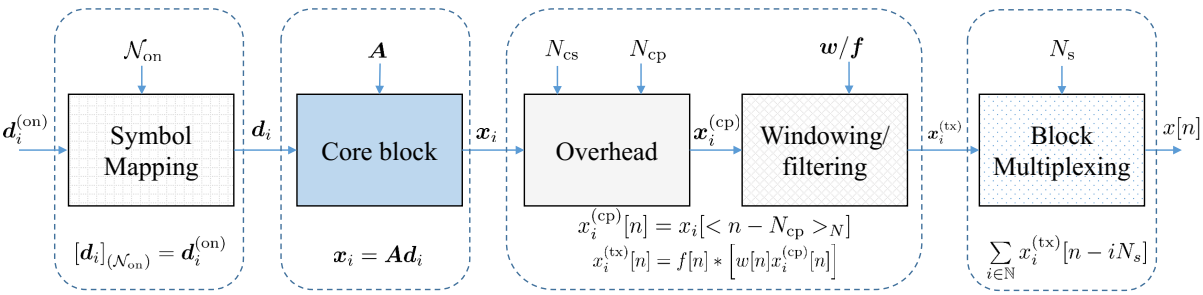

Figure 3.6: Multicarrier waveforms generator stages.

All the techniques presented in Section 3.2 can be used in the design of waveforms, and they are implicitly included in the matrix $\boldsymbol{G}$. Accordingly, the pulse length can be expressed as $N_\mathrm{t} = L_f + N_\mathrm{cp} + 2N_\mathrm{cs} + N$, where N defines the *core block* size. Assuming $N \geq K$, which can always be fulfilled by means of zero padding, either by including

non-active subcarriers for tall matrix or by increasing the pulse length by means of additional zeros for fat matrix. Thus, $\boldsymbol{G}$ can be expressed in the form

$$\boldsymbol{G} = \boldsymbol{G}^{(t)}\boldsymbol{A}\boldsymbol{\Pi}^{(d)}, \quad \text{where } \boldsymbol{G}^{(t)} = \boldsymbol{T}^{(\mathrm{F})}\boldsymbol{D}^{(\mathrm{W})}\boldsymbol{R}^{(\mathrm{cp/cs})}. \tag{3.13}$$

This works defines $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ as the *core modulation*, $\boldsymbol{\Pi}^{(d)} \in \mathbb{R}^{N\times K}$ is the allocation matrix used to map K data symbols to N subcarriers. In this case $N > K$, there will be $N - K > 0$ non-active subcarriers. The exact set can be determined by the design. Moreover, $\boldsymbol{R}^{(\mathrm{cp/cs})} \in \mathbb{R}^{[N+N_{\mathrm{cp}}+2N_{\mathrm{cs}}]\times N}$ is the matrix representing the transform of adding CP/CS. If no overhead is added, it becomes identity matrix. $\boldsymbol{D}^{(\mathrm{W})}$ is diagonal matrix of size $[N + N_{\mathrm{cp}} + N_{\mathrm{cs}}]$, which expresses the windowing. Finally, $\boldsymbol{T}^{(\mathrm{F})} \in \mathbb{C}^{N_T\times[N+N_{\mathrm{cp}}+N_{\mathrm{cs}}]}$ is a Toeplitz matrix generated from the filtering. The block diagram of general multicarrier waveform is illustrated in Fig. 3.6. A summary of the modulation steps is provided in Table. 3.2. From complexity perspective, a generic linear modulation with $\boldsymbol{G}$ of size $N_{\mathrm{t}}\times K$

Table 3.2: Multicarrier waveform generation steps

Stage	Function	Parameters	Related parameters
Symbol mapping	$\boldsymbol{d}_i\,[\mathcal{N}_{on}] = \boldsymbol{d}_i^{(\mathrm{on})}$	$\mathcal{N}_{on}$	active symbols set
Core block	$\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{d}_i$	$\boldsymbol{A} \in \mathbb{C}^{N\times N}$	modulation matrix
CP/CS overhead	$\boldsymbol{x}_i^{(\mathrm{cp})}[n] = \boldsymbol{x}_i[\langle n - N_{\mathrm{cp}}\rangle_N]$	$N_{\mathrm{cs}},\ N_{\mathrm{cp}}$	CP, CS length
Windowing	$\boldsymbol{x}_i^{(\mathrm{w})}[n] = w[n]\cdot\boldsymbol{x}_i^{(\mathrm{cp})}[n]$	$w[n]$	windowing function
Filtering	$\boldsymbol{x}_i^{(\mathrm{t})}[n] = f[n] * \boldsymbol{x}_i^{(\mathrm{w})}[n]$	$f[n]$	impulse response, $N_{\mathrm{t}} = N + N_{\mathrm{cp}} + N_{\mathrm{f}}$
Block multiplexing	$x[n] = \sum_{i\in\mathbb{Z}} \boldsymbol{x}_i^{(\mathrm{t})}[n - iN_{\mathrm{s}}]$	N_{s}	block spacing with overlapping $N_{\mathrm{o}} = N_{\mathrm{t}} - N_{\mathrm{s}}$

requires $N_T K$ complex multiplications, and a storage of size $N_{\mathrm{t}}K$ complex numbers. However, considering the presented structure, the implementation can be more efficient. For instance, the CP insertion can performed using a copy function, the number of complex multiplications and memory size decreases to NK. The filtering can be realized by means of tapped delay lines, and the windowing requires N_{t} multiplication at maximum. The main challenge is the design of the core modulation matrix $\boldsymbol{A}$ to have a special structure that enables the fast implementation while maintaining the performance metrics. The core block is defined by

$$\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{d}_i, \text{ with } \boldsymbol{d}_i = \boldsymbol{\Pi}^{(d)}\boldsymbol{d}_i^{(\mathrm{on})}. \tag{3.14}$$

Let $\mathcal{N}_{on}$ be the set of active subcarriers in $\boldsymbol{A}$, then the allocation can be performed by means of indexing, such that

$$\boldsymbol{d}_i\,[\mathcal{N}_{on}] = \boldsymbol{d}_i^{(\mathrm{on})}. \tag{3.15}$$

Note that, the allocation with respect to the symbol rate per subcarrier is implicitly included in $\boldsymbol{d}_i^{(\mathrm{on})}$. In the sequel of this theses, the core modulation will be discussed in

details. The pulse shape under consideration is $g_q[n] = \boldsymbol{A}[n,q]$, where q is a generic subcarrier index, and

$$\boldsymbol{x}_i[n] = \sum_{q=0}^{N-1} \boldsymbol{d}_i[q] g_q[n] = \sum_{q \in \mathcal{N}_{on}} \boldsymbol{d}_i[q] g_q[n]. \tag{3.16}$$

3.3.2 Demodulator

Synchronization techniques are used to allow the detection of a transmitted frame at the receiver. For example, in preamble-based synchronization, a predefined sequence is used on the front of a frame of several modulated blocks to detect the start of the frame. After the preamble, a header signal is used to provide information about the number of blocks and other modulation parameters.

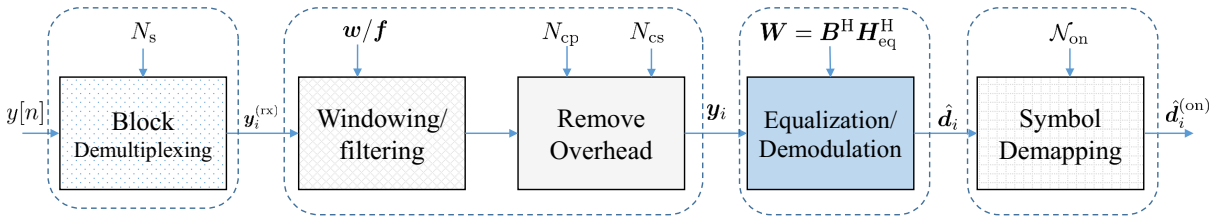

Figure 3.7: Multicarrier waveforms demodulator.

The synchronized signal is forwarded to the receiver, as shown in Figure 3.7, where the inverse operations of the modulator are performed. Based on the waveform parameters, the block demultiplexing extracts the modulated blocks. Depending on the block spacing and the number of discrete channel taps, the extracted block can be interfered with the previous and next block. Let $\boldsymbol{H}_i$ be the matrix representing the linear discrete channel. The received block of N_t samples can be represented by the vector

$$\boldsymbol{y}_i^{(\mathrm{r})} = \boldsymbol{H}_i \boldsymbol{x}_i^{(\mathrm{t})} + \boldsymbol{z}_i^{(r)} + \boldsymbol{v}_i^{(r)}, \tag{3.17}$$

where, $\boldsymbol{z}_i^{(r)}$ is the IBI and $\boldsymbol{v}_i^{(r)}$ the additive noise. Further linear processing can be performed in a form of matrix $\boldsymbol{G}^{(r)} \in \mathbb{C}^{N \times N_t}$, which involves filtering and widowing, in addition to removing the CP/CS. Hence, the obtained core block, $\boldsymbol{y}_i = \boldsymbol{G}^{(r)} \boldsymbol{y}_i^{(\mathrm{r})}$ can be expressed as

$$\boldsymbol{y}_i = \boldsymbol{H}_i^{(e)} \boldsymbol{x}_i + \boldsymbol{z}_i + \boldsymbol{v}_i, \tag{3.18}$$

where $\boldsymbol{H}_i^{(e)} = \boldsymbol{G}^{(r)} \boldsymbol{H}_i \boldsymbol{G}^{(t)} \in \mathbb{C}^{N \times N}$ is the effective channel matrix, $\boldsymbol{z}_i$ the IBI, and $\boldsymbol{v}_i$ the additive noise. It is convenient to design the waveforms to force the channel $\boldsymbol{H}^{(e)}$ to become circular, and $\boldsymbol{z}_i = \boldsymbol{0}$. This is feasible, when the channel variation is negligible within the block and the CP is selected to be longer than the filter length plus the channel length. The goal of the receiver is to estimate the data $\boldsymbol{d}_i^{(\mathrm{on})}$ from the received

block $\boldsymbol{y}_i = \boldsymbol{H}_i^{(e)} \boldsymbol{A}\boldsymbol{\Pi}^{(d)} \boldsymbol{d}_i^{(\mathrm{on})} + \boldsymbol{z}_i + \boldsymbol{v}_i$, as will be discussed in details in Section 3.4. In linear receiver, a matrix denoted as $\boldsymbol{W}_i^{\mathrm{H}}$ is used, such that

$$\hat{\boldsymbol{d}}^{(\mathrm{on})} = \boldsymbol{W}_i^{\mathrm{H}} \boldsymbol{y}_i = \boldsymbol{W}_i^{\mathrm{H}} \boldsymbol{H}_i^{(e)} \boldsymbol{A}\boldsymbol{\Pi}^{(d)} \boldsymbol{d}_i^{(\mathrm{on})} + \boldsymbol{W}_i^{\mathrm{H}} \left(\boldsymbol{z}_i + \boldsymbol{v}_i\right).$$

This thesis identify three options to design the receiver matrix as discussed in [NCF19b]:

1. Joint channel equalization and demodulation: in this approach the joint matrix $\boldsymbol{W}_i^{\mathrm{H}}$ is calculated based on the joint modulation and channel matrix $\boldsymbol{H}_i^{(e)} \boldsymbol{A}\boldsymbol{\Pi}^{(d)}$. The demodulator needs to calculate the receiver matrix for each new channel.
2. Channel equalization then demodulation: the receiver matrix can be decomposed such that $\boldsymbol{W}_i^{\mathrm{H}} = \boldsymbol{B}^{\mathrm{H}} \boldsymbol{H}_{i,\mathrm{eq}}^{\mathrm{H}}$, where $\boldsymbol{B}^{\mathrm{H}} \in \mathbb{C}^{N \times N}$ is the demodulator matrix, and $\boldsymbol{H}_{i,\mathrm{eq}}^{\mathrm{H}}$ the equalization matrix, which is variant depending on the channel. The channel equalization is performed first, which depends on the structure of the channel. Thereafter, the demodulator is applied and finally the allocated samples are extracted by demapping using $\boldsymbol{\Pi}^{(d)T}$.
3. Demodulation then channel equalization: here the demodulation and demapping are applied first such that

$$\begin{aligned} \bar{\boldsymbol{d}}^{(\mathrm{on})} &= \boldsymbol{\Pi}^{(d)T} \boldsymbol{B}^{\mathrm{H}} \boldsymbol{y}_i \\ &= \boldsymbol{\Pi}^{(d)T} \boldsymbol{B}^{\mathrm{H}} \boldsymbol{H}_i^{(e)} \boldsymbol{A}\boldsymbol{\Pi}^{(d)} \boldsymbol{d}_i^{(\mathrm{on})} + \boldsymbol{\Pi}^{(d)T} \boldsymbol{B}^{\mathrm{H}} (\boldsymbol{z}_i + \boldsymbol{v}_i). \end{aligned}$$

   The equalization matrix $\bar{\boldsymbol{H}}_{i,\mathrm{eq}} \in \mathbb{C}^{\mathcal{N}_{on} \times \mathcal{N}_{on}}$ is designed based on the mixed channel matrix $\boldsymbol{\Pi}^{(d)T} \boldsymbol{B}^{\mathrm{H}} \boldsymbol{H}_i^{(e)} \boldsymbol{A}\boldsymbol{\Pi}^{(d)}$.

In the second and third options, the demodulation applied on a block $\boldsymbol{y}_i$ is given by

$$\hat{\boldsymbol{d}}_i[q] = \sum_{n=0}^{N-1} \gamma_q^*[n] \boldsymbol{y}_i[n], \ \gamma_q[n] = \boldsymbol{B}[n, q]. \tag{3.19}$$

3.3.3 MIMO waveform processing

For the n_t-th transmitter chain, $\boldsymbol{x}_{n_t,i}^{(t)} = \boldsymbol{G}_{n_t} \boldsymbol{d}_i^{(\mathrm{on})} \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$, $\boldsymbol{d}_i^{(\mathrm{on})} \in \mathbb{C}^{N_T K \times 1}$ where $N_T K$ is the total number of active carriers, and $\boldsymbol{G}_{n_t} = \boldsymbol{G}_{n_t}^{(t)} \boldsymbol{A}_{n_t} \boldsymbol{\Pi}_{n_t}^{(d)}$. Here, $\boldsymbol{A}_{n_t} \in \mathbb{C}^{N \times N_T N}$ is a fat matrix defined by $\boldsymbol{A}_{n_t} = [\boldsymbol{A}_{n_t,1} \cdots \boldsymbol{A}_{n_t,N_T}]$, and $\boldsymbol{G}_{n_t}^{(t)} \in \mathbb{C}^{N_{\mathrm{t}} \times N}$ represents the additional processing. The allocation matrix $\boldsymbol{\Pi}_{n_t}^{(d)} = \mathrm{Bd}\{\boldsymbol{\Pi}_{n_t,1}^{(d)} \cdots \boldsymbol{\Pi}_{n_t,N_T}^{(d)}\} \in \mathbb{C}^{N_T N \times N_T K}$ defines how the symbols are mapped to the n_t-th chain. As a result, the core block can be written as

$$\boldsymbol{x}_{n_t,i} = \sum_{m_t=1}^{N_T} \boldsymbol{A}_{n_t,m_t} \boldsymbol{d}_{n_t,m_t,i}, \ \boldsymbol{d}_{n_t,m_t i} = \boldsymbol{\Pi}_{n_t,m_t}^{(d)} \boldsymbol{d}_{m_t,i}^{(\mathrm{on})}. \tag{3.20}$$

In the most general case, each core matrix $\boldsymbol{A}_{n_t,m_t}$ is associated with different allocation matrix $\boldsymbol{\Pi}_{n_t,m_t}^{(d)}$ as illustrated in Figure 3.8. In special case, when $\boldsymbol{\Pi}_{n_t,m_t}^{(d)} = \boldsymbol{\Pi}_{m_t}^{(d)}$, and thus,

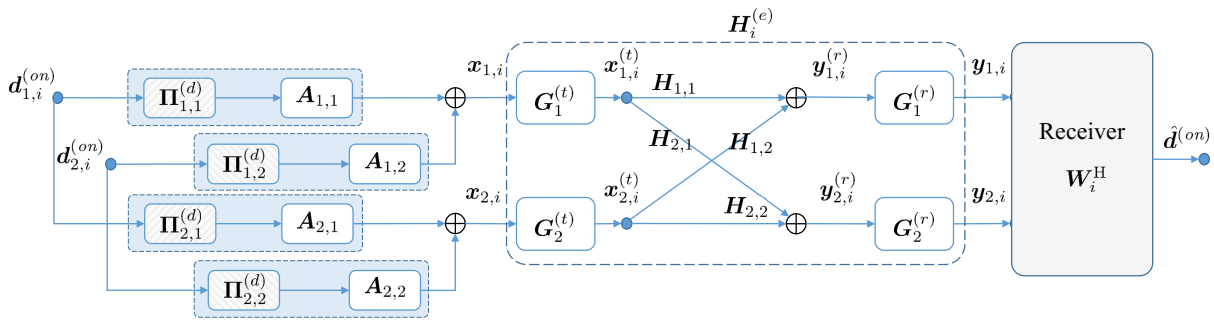

Figure 3.8: 2×2 MIMO waveform example.

$\mathbf{\Pi}_{n_t}^{(d)} = \mathbf{\Pi}^{(d)}$, the core block can be expressed as

$$\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{d}_i, \;\; \boldsymbol{d}_i = \mathbf{\Pi}^{(d)}\boldsymbol{d}_i^{(\mathrm{on})},$$

where $\boldsymbol{A} \in \mathbb{C}^{N_T N \times N_T N}$, $\boldsymbol{A}[(n_t-1)N+n, (m_t-1)N+q] = \boldsymbol{A}_{n_t,m_t}[n,q]$ is the MIMO core modulation matrix and $\mathbf{\Pi}^{(d)} \in \mathbb{R}^{N_T N \times N_T K}$ is the allocation matrix, and $\boldsymbol{x}_i = \left[\boldsymbol{x}_{1,i}^T, \cdots, \boldsymbol{x}_{N_T,i}^T\right]^T \in \mathbb{C}^{N_T N \times 1}$ is the MIMO core modulation block. The allocation $\mathbf{\Pi}^{(d)}$ can be chosen as a general allocation matrix. Thus, the processing of MIMO starts by generating the core block, then splitting this block for different transmitter chains. Note here, the vector $\boldsymbol{x}_i$ is arranged by stacking the vectors of the transmitter chains after each other, i.e by varying the time index first then the transmitter index. Different ordering, i.e. permutation, can also be used such as stacking with respect to varying the transmitter index first then the time index. After the core block modulation, further processing is performed on each chain with the matrix $\boldsymbol{G}_{n_t}^{(t)}$.

Received signal. Following similar steps in Section 3.3.2, the received signal on the n_r-th antenna is

$$\boldsymbol{y}_{n_r,i} = \sum_{n_t=1}^{N_T} \boldsymbol{H}_{n_r,n_t,i}^{(e)} \boldsymbol{x}_{n_t,i} + \boldsymbol{z}_{n_r,i} + \boldsymbol{v}_{n_r,i} \tag{3.21}$$

where $\boldsymbol{H}_{n_r,n_t,i}^{(e)} = \boldsymbol{G}_{n_r}^{(r)} \boldsymbol{H}_{n_r,n_t} \boldsymbol{G}_{n_t}^{(t)} \in \mathbb{C}^{N \times N}$. By stacking the vectors, the equivalent signal can be fit in (3.18), $\boldsymbol{y}_i = \boldsymbol{H}_i^{(e)}\boldsymbol{x}_i + \boldsymbol{z}_i + \boldsymbol{v}_i$, where $\boldsymbol{H}_i^{(e)} \in \mathbb{C}^{N_R N \times N_T N}$, is defined such that $\boldsymbol{H}_i^{(e)}[(n_r-1)N+n, (n_r-1)N+q] = \boldsymbol{H}_{i,n_r,n_t}^{(e)}[n,q]$ and $\boldsymbol{y}_i = \left[\boldsymbol{y}_{1,i}^T, \cdots, \boldsymbol{y}_{N_R,i}^T\right]^T \in \mathbb{C}^{N_R N \times 1}$. This allows the representation of core modulation in a unified form. However, the channel structure depends on the number of transmitter and receiver chains.

3.4 Detection

This section focuses on the detection techniques without considering a specific structure of the channel or modulation matrix. As introduced in (3.18), the received block follows a

linear model. Assuming that each block carries independent packet of data, and there is no IBI, i.e. $\boldsymbol{z}_i = \boldsymbol{0}$. Moreover, for simplicity, the index i is dropped and with notation abuse $\boldsymbol{d}$ replaces $\boldsymbol{d}^{(\mathrm{on})}$. The effective modulation matrix is defined by $\boldsymbol{A}^{(\mathrm{on})} = \boldsymbol{A}\boldsymbol{\Pi}^{(d)} \in \mathbb{C}^{N\times K}$ and the joint channel and modulation matrix $\boldsymbol{H} = \boldsymbol{H}_i^{(e)}\boldsymbol{A}^{(\mathrm{on})} \in \mathbb{C}^{N\times K}$. Accordingly, the signal model can be written in the form

$$\boldsymbol{y} = \boldsymbol{H}^{(e)}\boldsymbol{A}^{(\mathrm{on})}\boldsymbol{d} + \boldsymbol{v} = \boldsymbol{H}\boldsymbol{d} + \boldsymbol{v}. \tag{3.22}$$

3.4.1 Maximum-likelihood detection

Let K_b be the number of data bits per packet. Hence, there are 2^{K_b} different possible packets, which require a vector code book $\mathcal{S}_d = \{\boldsymbol{d}_0, \cdots, \boldsymbol{d}_{2^{K_b}-1}\}$ to map any packet to a unique data symbol vector. The maximum a posteriori (MAP) detection aims at finding a code word $\hat{\boldsymbol{d}}_{\mathrm{MAP}} \in \mathcal{S}_d$ that maximizes the probability $\Pr(\boldsymbol{d}|\boldsymbol{y})$, which is formulated as

$$\hat{\boldsymbol{d}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{d}\in\mathcal{S}_d} \log\left(\Pr(\boldsymbol{d}|\boldsymbol{y})\right) = \arg\max_{\boldsymbol{d}\in\mathcal{S}_d} \log\left(f_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{d})\Pr(\boldsymbol{d})\right) \tag{3.23}$$

using the relation $\Pr(\boldsymbol{d}|\boldsymbol{y}) = f_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{d})\Pr(\boldsymbol{d}_n)\frac{1}{f_{\boldsymbol{y}}(\boldsymbol{y})}$. When the code words are uniformly distributed, i.e. $\Pr(\boldsymbol{d}_n) = 1/2^{K_b}$, the MAP detection becomes similar to maximum likelihood detection (MLD), which is given by

$$\hat{\boldsymbol{d}}_{\mathrm{MLD}} = \arg\max_{\boldsymbol{d}\in\mathcal{S}_d} \log\left(f_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{d})\right). \tag{3.24}$$

For AWGN channel, where $\mathrm{E}\left[\boldsymbol{v}\boldsymbol{v}^{\mathrm{H}}\right] = \sigma^2\boldsymbol{I}_N$, the MLD is equivalent to

$$\hat{\boldsymbol{d}}_{\mathrm{MLD}} = \arg\min_{\boldsymbol{d}\in\mathcal{S}_d} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}\|^2. \tag{3.25}$$

A unitary transform with $\boldsymbol{Q} \in \mathbb{C}^{N\times N}$, $\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q} = \boldsymbol{I}_N$ preserves the distance, such that $\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}\|^2 = \|\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{y} - \boldsymbol{Q}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{d}\|^2$. Applying that using the QR decomposition defined by

$$\boldsymbol{H} = [\boldsymbol{Q}_1\ \boldsymbol{Q}_0]\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}_{N-K\times K}\end{bmatrix} = \boldsymbol{Q}_1\boldsymbol{R},\quad \boldsymbol{Q} = [\boldsymbol{Q}_1\ \boldsymbol{Q}_0],\ \boldsymbol{Q}_1 \in \mathbb{C}^{N\times K},\quad \boldsymbol{Q}_0 \in \mathbb{C}^{N\times(N-K)}. \tag{3.26}$$

where $\boldsymbol{R} \in \mathbb{C}^{K\times K}$ is an upper triangular matrix, then

$$\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}\|^2 = \|\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{y} - \begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}_{N-K\times K}\end{bmatrix}\boldsymbol{d}\|^2 = \|\boldsymbol{Q}_1^{\mathrm{H}}\boldsymbol{y} - \boldsymbol{R}\boldsymbol{d}\|^2 + \|\boldsymbol{Q}_0^{\mathrm{H}}\boldsymbol{y}\|^2, \tag{3.27}$$

As a result, the MLD problem in AWGN can be solved using $\bar{\boldsymbol{y}} = \boldsymbol{Q}_1^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{R}\boldsymbol{d} + \bar{\boldsymbol{v}}$, where $\bar{\boldsymbol{v}} = \boldsymbol{Q}_1^{\mathrm{H}}\boldsymbol{v}$ and thus, $\mathrm{E}\left[\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}^H\right] = \sigma^2\boldsymbol{I}_K$. As a result,

$$\hat{\boldsymbol{d}}_{\mathrm{MLD}} = \arg\min_{\boldsymbol{d}\in\mathcal{S}_d} \sum_{k=0}^{K-1}\left|\bar{\boldsymbol{y}}[k] - \sum_{l=k}^{K-1}\boldsymbol{R}[k,l]\boldsymbol{d}[l]\right|^2. \tag{3.28}$$

Nevertheless, solving this problem requires trying all possibilities, and the complexity scales exponentially with increase of packet length. In a special case, when $\boldsymbol{d}[k] \in \mathcal{S}_k$, $\mathcal{S}_d = \mathcal{S}_0 \times \mathcal{S}_1 \times \cdots \times \mathcal{S}_{K-1}$, and $\{\boldsymbol{d}[k]\}$ are independent, (3.28) can help in reducing the complexity of MLD using sphere decoding (SD) approach [HV05], or to implement successive interference cancellation (SIC), such as V-BLAST [WFGV98]. Moreover, if $\boldsymbol{R} = \boldsymbol{\Lambda}$ is diagonal, then $\hat{\boldsymbol{d}}_{\text{MLD}} = \arg\min_{\boldsymbol{d}\in\mathcal{S}_d} \sum_{k=0}^{K-1} |\bar{\boldsymbol{y}}[k] - \boldsymbol{\Lambda}[k,k]\boldsymbol{d}[k]|^2$, and thus

$$\hat{\boldsymbol{d}}_{\text{MLD}}[k] = \arg\min_{\boldsymbol{d}[k]\in\mathcal{S}_k} |\bar{\boldsymbol{y}}[k] - \boldsymbol{\Lambda}[k,k]\boldsymbol{d}[k]|^2 . \tag{3.29}$$

Conventional bit mapping and channel coding. To simplify the mapping from bits to a code vector, first, a finite constellation set $\mathcal{M}_c$ with $M_c = 2^{L_c}$ elements is used for each vector element, i.e. $\mathcal{S}_k = \mathcal{M}_c, k = 0, \cdots, K-1$. Each data symbols can equally encode L_c bits, and therefore, the set $\mathcal{M}_c^K$ contains 2^{K_c} elements with $K_c = KL_c$ bits per vector, where $K_c \geq K_b$. If $K_b < K_c$, which means a code rate $R_c = \frac{K_b}{K_c} < 1$, then $\mathcal{S}_d \subset \mathcal{M}_c^K$. Thus, not all $\boldsymbol{d} \in \mathcal{M}_c^K$ are valid mapping vectors. However, the output of the MAP detection considering all the entries of $\mathcal{M}_c^K$ can be used as intermediate step in the search for the closest valid vector.

In conventional linear channel coding in Galois field, the encoded bits are represented in the vector $\boldsymbol{c} = \mu(\boldsymbol{b}) \in \mathbb{F}_2^{K_c\times 1}$, where $\boldsymbol{b} \in \mathbb{F}_2^{K_b\times 1}$ denotes the data bit vector and $\mu(\cdot)$ the encoder function. In discrete hard bit-flipping channel, the perturbed code word is $\bar{\boldsymbol{c}} = \boldsymbol{c} + \boldsymbol{e}$, where $\boldsymbol{e} \in \mathbb{F}_2^{K_c\times 1}$. In this case the MLD decoding is equivalent to finding the nearest valid codeword that minimizes the Hamming distance. Given a continuous noisy channel, the noisy codeword $\boldsymbol{y}_c = \boldsymbol{c} + \boldsymbol{v}_c$ is of continuous values, which needs to be considered in the MLD. Thereby, instead of providing the hard erroneous codeword to the decoder, soft information is used in form of bit log-likelihood ratio (LLR) defined by

$$\text{LLR}_l = \boldsymbol{\lambda}[l] = \log_2\left(\frac{\Pr(\boldsymbol{c}[l]=1|\boldsymbol{y}_c)}{\Pr(\boldsymbol{c}[l]=0|\boldsymbol{y}_c)}\right), \ \boldsymbol{\lambda} \in \mathbb{R}^{K_c\times 1}. \tag{3.30}$$

The hard decision is given by $\boldsymbol{c}[l] = 1$ when $\boldsymbol{\lambda}[l] \geq 0$ and $\boldsymbol{c}[l] = 0$ if $\boldsymbol{\lambda}[l] < 0$.

LLR approximation. The mapping of code bits to a symbol is achieved by selecting a unique symbol from the constellations set based on L_c bit sequences. This can be expressed as $\boldsymbol{d}[k] = \mathcal{M}_c\{\Gamma(\boldsymbol{c}[kL_c], \cdots, \boldsymbol{c}[kL_c + L_c - 1])\}$. Here, $\Gamma : \mathbb{F}_2^{L_c} \longrightarrow \left\{0, \cdots, 2^{L_c} - 1\right\}$ is bit to integer mapping injective function. Consequently, the probability distribution of the symbol vectors is $\Pr(\boldsymbol{d}) = \Pr(\boldsymbol{c}[0], \cdots, \boldsymbol{c}[K_c - 1])$, and thus,

$$\Pr(\boldsymbol{c}[l] = c|\boldsymbol{y}) = \sum_{q,\boldsymbol{c}[l]=c} \Pr(\boldsymbol{d}_q|\boldsymbol{y}) = \sum_{q,\boldsymbol{c}[l]=c} \frac{f_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{d}_q)}{f_{\boldsymbol{y}}(\boldsymbol{y})} \Pr(\boldsymbol{d}_q), \tag{3.31}$$

where $\boldsymbol{d}_q \in \mathcal{M}_c^K,\ q = 0, \cdots, 2^{K_c} - 1$. For uniformly distributed codewords,

$$\begin{aligned}\boldsymbol{\lambda}[l] &= \log\left(\frac{\sum_{q,\boldsymbol{c}[l]=1} e^{-\frac{1}{\sigma^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{d}_q\|^2}}{\sum_{q,\boldsymbol{c}[l]=0} e^{-\frac{1}{\sigma^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{d}_q\|^2}}\right) \approx \log\left(\frac{\max_{q,\boldsymbol{c}[l]=1} e^{-\frac{1}{\sigma^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{d}_q\|^2}}{\max_{q,\boldsymbol{c}[l]=0} e^{-\frac{1}{\sigma^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{d}_q\|^2}}\right) \\ &\approx \min_{q,\boldsymbol{c}[l]=0}\{\frac{1}{\sigma^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{d}_q\|^2\} - \min_{q,\boldsymbol{c}[l]=1}\{\frac{1}{\sigma^2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{d}_q\|^2\}.\end{aligned} \tag{3.32}$$

The bit probability can be determined from the LLRs as

$$P(\boldsymbol{c}[l]=1|\boldsymbol{y}) = \frac{e^{\boldsymbol{\lambda}[l]}}{e^{\boldsymbol{\lambda}[l]}+1},\ P(\boldsymbol{c}[l]=0|\boldsymbol{y}) = 1 - P(\boldsymbol{c}[l]=1|\boldsymbol{y}). \tag{3.33}$$

Given the probabilities of estimated encoded bits, the probability of vector is

$$\Pr(\hat{\boldsymbol{d}} = \boldsymbol{d}_q) = \prod_{l=0}^{K_c-1} P(\hat{\boldsymbol{c}}[l] = c_{q,l}), \tag{3.34}$$

where $\boldsymbol{d}_q$ is the codeword corresponding to the bit sequence $(c_{q,0}, \cdots, c_{q,K_c-1})$.

3.4.2 Linear detection

Solving the MLD for general $\boldsymbol{H} \in \mathbb{C}^{N\times K}$ is of exponential complexity using an exhaustive search for all 2^{KL_c} possibilities. An approximate solution with SD reduces the complexity. However, even if the MLD problem is solved efficiently, the computation of the LLRs is not an easy task. In a special case, when the triangular matrix of the QR decomposition (3.28) is diagonal, $\boldsymbol{R} = \boldsymbol{\Lambda}$, the MLD is reduced to computing the MLD of K independent scalar channels using the linear transform $\bar{\boldsymbol{y}} = \boldsymbol{Q}_1^{\mathrm{H}}\boldsymbol{y}$ following the signal model

$$\bar{\boldsymbol{y}}[k] = \boldsymbol{\Lambda}[k,k]\boldsymbol{d}[k] + \bar{\boldsymbol{v}}[k]. \tag{3.35}$$

The LLRs can be computed from each symbol independently

$$\boldsymbol{\lambda}[kL_c + l] \approx \min_{q,\boldsymbol{c}[l]=1}\left\{\frac{1}{\sigma^2}|\bar{\boldsymbol{y}}[k] - \boldsymbol{\Lambda}[k,k]d_q|^2\right\} - \min_{q,\boldsymbol{c}[l]=0}\left\{\frac{1}{\sigma^2}|\bar{\boldsymbol{y}}[k] - \boldsymbol{\Lambda}[k,k]d_q|^2\right\}. \tag{3.36}$$

Here $d_q \in \mathcal{M}_c$, $q = 0 \cdots, 2^{L_c} - 1$. In this case, the linear transform with $\boldsymbol{Q}_1^{\mathrm{H}}$ decouples the symbols with no inter-symbol interference (ISI). In other cases, the transform may exploit certain structure of $\boldsymbol{H}$ to reduce the complexity. For instance, when $\boldsymbol{H}$ is decomposed into block diagonal matrices, the MLD and LLR computation can be performed on smaller sub channels independently.

The idea of linear transform can be generalized. Let $\boldsymbol{W}^{\mathrm{H}} \in \mathbb{C}^{M\times N}$, which is applied such that $\bar{\boldsymbol{y}} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{d} + \boldsymbol{W}^{\mathrm{H}}\boldsymbol{v}$. Therefore, defining $\boldsymbol{w}_k = \boldsymbol{W}[:,k]$ and $\boldsymbol{h}_k = \boldsymbol{H}[:,k]$, the k-th channel is given by

$$\begin{aligned}\bar{\boldsymbol{y}}[k] = \boldsymbol{w}_k^{\mathrm{H}}\boldsymbol{y} &= \boldsymbol{w}_k^{\mathrm{H}}\boldsymbol{h}_k\boldsymbol{d}[k] + \sum_{q\neq k}^{k-1}\boldsymbol{w}_k^{\mathrm{H}}\boldsymbol{h}_q\boldsymbol{d}[q] + \boldsymbol{w}_k^{\mathrm{H}}\boldsymbol{v}, \\ &= \boldsymbol{\alpha}[k]\,\boldsymbol{d}[k] + \bar{\boldsymbol{z}}[k] + \bar{\boldsymbol{v}}[k], \\ \boldsymbol{\alpha} &= \mathrm{diag}\left\{\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\right\},\ \bar{\boldsymbol{v}} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{v},\ \bar{\boldsymbol{z}} = \left(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H} - \boldsymbol{\Lambda}^{(\boldsymbol{\alpha})}\right)\boldsymbol{d}, \boldsymbol{\Lambda}^{(\boldsymbol{\alpha})} = \mathrm{diag}\{\boldsymbol{\alpha}\}.\end{aligned} \tag{3.37}$$

Linear transform design

This work proposes the design of $\boldsymbol{w}_k$ by maximizing the signal-to-interference-plus-noise ratio (SINR) in order to minimize the symbol error rate (SER) per channel. Without loss of generality, let $\mathrm{E}\left[\boldsymbol{d}\boldsymbol{d}^{\mathrm{H}}\right] = \boldsymbol{I}_N$ and $\mathrm{E}\left[\boldsymbol{v}\boldsymbol{v}^{\mathrm{H}}\right] = \sigma^2\boldsymbol{I}_M$, the SINR per channel is given by

$$\boldsymbol{\rho}\left[k\right] = \frac{\boldsymbol{w}_k^{\mathrm{H}}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{w}_k}{\boldsymbol{w}_k^{\mathrm{H}}\left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} - \boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]\boldsymbol{w}_k}. \tag{3.38}$$

The solution is determined using the derivation $\frac{\partial\boldsymbol{\rho}[k]}{\boldsymbol{w}_k^{\mathrm{H}}} = 0$, which leads to the relation

$$\begin{aligned}
\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{w}_k &= \boldsymbol{\rho}\left[k\right]\left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} - \boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]\boldsymbol{w}_k \\
(\boldsymbol{\rho}\left[k\right]+1)\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{w}_k &= \boldsymbol{\rho}\left[k\right]\left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]\boldsymbol{w}_k \\
\underbrace{\left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1}\boldsymbol{h}_k}_{\boldsymbol{u}_k}\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{w}_k &= \frac{\boldsymbol{\rho}\left[k\right]}{\boldsymbol{\rho}\left[k\right]+1}\boldsymbol{w}_k.
\end{aligned} \tag{3.39}$$

One solution of $\boldsymbol{u}_k\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{w}_k = \frac{\boldsymbol{\rho}[k]}{\boldsymbol{\rho}[k]+1}\boldsymbol{w}_k$ is achieved when $\boldsymbol{w}_k = \boldsymbol{u}_k$, and thus, $\frac{\boldsymbol{\rho}[k]}{\boldsymbol{\rho}[k]+1} = \boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{u}_k$. Thereby, a linear transform that maximizes the SINR per channel is given by

$$\boldsymbol{W} = [\boldsymbol{u}_0, \cdots, \boldsymbol{u}_{K-1}] = \left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1}\boldsymbol{H}. \tag{3.40}$$

Moreover, because $\boldsymbol{\alpha}\left[k\right] = \boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{u}_k = \frac{\boldsymbol{\rho}[k]}{\boldsymbol{\rho}[k]+1} < 1$, then

$$\boldsymbol{\rho}\left[k\right] = \frac{\boldsymbol{\alpha}\left[k\right]}{1-\boldsymbol{\alpha}\left[k\right]},\ \boldsymbol{\alpha} = \mathrm{diag}\left\{\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\right\} = \mathrm{diag}\left\{\boldsymbol{H}^{\mathrm{H}}\left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1}\boldsymbol{H}\right\} \in \mathbb{R}^{+K\times 1}. \tag{3.41}$$

In fact, this proposed solution is exactly the linear minimum mean square error (LMMSE), which is conventionally derived by minimizing the mean squared error (MSE) $\|\mathrm{E}\left[\boldsymbol{W}^{\mathrm{H}}\boldsymbol{y} - \boldsymbol{d}\right]\|^2$. Note that LMMSE is a biased estimator when considering $\hat{\boldsymbol{d}}_{\mathrm{bias}} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{y}$ [Kay93]. The bias can be corrected by elementwise division $\hat{\boldsymbol{d}}_{\mathrm{unbias}} = \hat{\boldsymbol{d}}_{\mathrm{bias}} \oslash \boldsymbol{\alpha}$, as introduced in [HLH17] under the name component-wise conditionally unbiased (CWCU) LMMSE.

Alternative form of LMMSE. Let $\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Sigma}_h\boldsymbol{U}^{\mathrm{H}}$ be the singular value decomposition (SVD) of $\boldsymbol{H}$, where $\boldsymbol{V} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{U}^{\mathrm{H}} \in \mathbb{C}^{K\times K}$. For $N \geq K$, $\boldsymbol{\Lambda}_K = \boldsymbol{\Sigma}_{\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{\Sigma}_h \in \mathbb{R}^{K\times K}$ is diagonal matrix, and $\boldsymbol{\Lambda}_N = \boldsymbol{\Sigma}_{\boldsymbol{h}}\boldsymbol{\Sigma}_h^{\mathrm{H}} \in \mathbb{R}^{N\times N}$ is diagonal with K non-zero entries, $\boldsymbol{\Lambda}_K[n,n] = 0,\ n > K-1$, and $\boldsymbol{\Lambda}_N[n,n] = \boldsymbol{\Lambda}_K[n,n],\ n \leq K-1$. Therefore,

$$\begin{aligned}
\boldsymbol{W}_{\mathrm{LMMSE}} = \left[\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1}\boldsymbol{H} &= \boldsymbol{V}[\boldsymbol{\Lambda}_N + \sigma^2\boldsymbol{I}_N]^{-1}\boldsymbol{\Sigma}_h\boldsymbol{U}^{\mathrm{H}} \\
&= \boldsymbol{V}\boldsymbol{\Sigma}_h\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}[\boldsymbol{\Lambda}_K + \sigma^2\boldsymbol{I}_K]^{-1}\boldsymbol{U}^{\mathrm{H}} \\
&= \boldsymbol{H}\left[\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma^2\boldsymbol{I}_K\right]^{-1}.
\end{aligned} \tag{3.42}$$

Accordingly, the gain can be computed from $\boldsymbol{\alpha} = \mathrm{diag}\left\{\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W}_{\mathrm{LMMSE}}\right\}$, with

$$\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W}_{\mathrm{LMMSE}} = \boldsymbol{U}\boldsymbol{\Lambda}_K[\boldsymbol{\Lambda}_K + \sigma^2\boldsymbol{I}_K]^{-1}\boldsymbol{U}^{\mathrm{H}}. \tag{3.43}$$

When $\boldsymbol{H}$ has certain structure, such that $\boldsymbol{U}$ and $\boldsymbol{V}$ are fixed, the computation of the LMMSE filter can be achieved by computing the simple inverse of the diagonal matrix $[\boldsymbol{\Lambda}_K + \sigma^2 \boldsymbol{I}_K]^{-1}$.

Equal-SINR criterion. The LLR of each stream depends on $\boldsymbol{\rho}[k]$ and $\boldsymbol{\alpha}[k]$, where

$$\boldsymbol{\lambda}[kL_c + l] \approx \boldsymbol{\rho}[k] \min_{q,\boldsymbol{c}[l]=1} \left\{ \left| \frac{1}{\boldsymbol{\alpha}[k]} \bar{\boldsymbol{y}}[k] - d_q \right|^2 \right\} - \boldsymbol{\rho}[k] \min_{q,\boldsymbol{c}[l]=0} \left\{ \left| \frac{1}{\boldsymbol{\alpha}[k]} \bar{\boldsymbol{y}}[k] - d_q \right|^2 \right\}.$$

For fixed SINR, $\boldsymbol{\rho}[k] = \rho, \forall k = 0 \cdots, K-1$, which in the case of LMMSE leads to equal gain, such that $\boldsymbol{\alpha} = \alpha \mathbf{1}_K$ and $\alpha = \frac{\rho}{1+\rho}$, the LLR computation can be reduced to

$$\boldsymbol{\lambda}[kL_c + l] \approx \rho \min_{q,\boldsymbol{c}[l]=1} \left\{ \left| \frac{1}{\alpha} \bar{\boldsymbol{y}}[k] - d_q \right|^2 \right\} - \rho \min_{q,\boldsymbol{c}[l]=0} \left\{ \left| \frac{1}{\alpha} \bar{\boldsymbol{y}}[k] - d_q \right|^2 \right\}.$$

In this case, the exact knowledge of ρ is not required. Achieving equal gain is feasible if the unitary matrix $\boldsymbol{U}$ fulfills, for any diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{C}^{K \times K}$, the condition

$$\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}[k,k] = \sum_{q=0}^{K-1} \boldsymbol{\Lambda}[q,q] \left|\boldsymbol{U}[q,k]\right|^2 = \text{const}, \quad \text{and} \quad \sum_{q=0}^{K-1} \left|\boldsymbol{U}[q,k]\right|^2 = 1. \tag{3.44}$$

One solution is obtained when $\left|\boldsymbol{U}[q,k]\right|^2 = \frac{1}{K}, \forall (q,n)$, i.e. the entries of $\boldsymbol{U}$ are of equal amplitude. Therefore, $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}[k,k] = \frac{1}{K}\mathrm{trace}\left\{\boldsymbol{\Lambda}\right\}$. As an example of that, the normalized K-DFT matrix. Accordingly, the gain achieved by LMMSE is

$$\alpha = \frac{1}{K}\mathrm{trace}\left\{\boldsymbol{\Lambda}_K[\boldsymbol{\Lambda}_K + \sigma^2 \boldsymbol{I}_K]^{-1}\right\}. \tag{3.45}$$

In fact, any equalization matrix in the form

$$\boldsymbol{W} = \boldsymbol{V}\boldsymbol{\Sigma}_h \boldsymbol{\Lambda}^{(w)} \boldsymbol{U}^{\mathrm{H}}, \ \boldsymbol{\Lambda}^{(w)} \in \mathbb{C}^{K \times K} \text{ is diagonal}, \tag{3.46}$$

with $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Lambda}_K \boldsymbol{\Lambda}^{(w)} \boldsymbol{U}^{\mathrm{H}}$ and $\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Lambda}_K \boldsymbol{\Lambda}^{(w)2} \boldsymbol{U}^{\mathrm{H}}$, achieves equal gain

$$\alpha = \frac{1}{K}\mathrm{trace}\left\{\boldsymbol{\Lambda}_K \boldsymbol{\Lambda}^{(w)}\right\}, \ \rho = \frac{\alpha^2}{\frac{1}{K}\mathrm{trace}\left\{\boldsymbol{\Lambda}_K \boldsymbol{\Lambda}^{(w)2}\left[\boldsymbol{\Lambda}_K + \sigma^2 \boldsymbol{I}_K\right]\right\} - \alpha^2}. \tag{3.47}$$

For instance, matched filter (MF), zero-forcing (ZF) and LMMSE all achieve equal gain. However, LMMSE achieves the highest SINR. When the value of σ^2 is not available, the LMMSE can be replaced by $\boldsymbol{\Lambda}^{(w)} = [\boldsymbol{\Lambda}_K + \gamma \boldsymbol{I}_K]^{-1}$, where $\gamma > 0$ is used for tuning.

Correlated Data. A correlated data vector with mean $\mathrm{E}[\boldsymbol{d}] = \boldsymbol{\mu}_d$ and covariance matrix $\boldsymbol{R}_d = \mathrm{E}\left[(\boldsymbol{d} - \boldsymbol{\mu}_d)(\boldsymbol{d} - \boldsymbol{\mu}_d)^{\mathrm{H}}\right] \in \mathbb{C}^{K \times K}$ can be expressed using zero-mean and uncorrelated data vector $\boldsymbol{d}_0$, such that $\boldsymbol{d} = \boldsymbol{R}_d^{1/2}\boldsymbol{d}_0 + \boldsymbol{\mu}_d$. The corresponding received

block $\boldsymbol{y}_0 = \boldsymbol{y} - \boldsymbol{H}\mu_d = \boldsymbol{H}\boldsymbol{R}_d^{1/2}\boldsymbol{d}_0 + \boldsymbol{v}$. Thus, the LMMSE w.r.t $\boldsymbol{d}_0$, and the channel matrix $\boldsymbol{H}\boldsymbol{R}_d^{1/2}$ can be derived using (3.42) as

$$\begin{aligned}\boldsymbol{W}_{\text{LMMSE0}}^{\text{H}} &= \boldsymbol{R}_d^{1/2}\boldsymbol{H}^{\text{H}}\left[\boldsymbol{H}\boldsymbol{R}_d\boldsymbol{H}^{\text{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1} \\ &= \left[\boldsymbol{R}_d^{1/2}\boldsymbol{H}^{\text{H}}\boldsymbol{H}\boldsymbol{R}_d^{1/2} + \sigma^2\boldsymbol{I}_K\right]^{-1}\boldsymbol{R}_d^{1/2}\boldsymbol{H}^{\text{H}} \\ &= \boldsymbol{R}_d^{-1/2}\left[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \sigma^2\boldsymbol{R}_d^{-1}\right]^{-1}\boldsymbol{H}^{\text{H}}.\end{aligned} \tag{3.48}$$

By dividing over the gain $\boldsymbol{\alpha}_0 = \text{diag}\left\{\boldsymbol{W}_{\text{LMMSE0}}^{\text{H}}\boldsymbol{H}\boldsymbol{R}_d^{1/2}\right\} = \text{diag}\left\{\boldsymbol{R}_d^{1/2}\boldsymbol{W}_{\text{LMMSE0}}^{\text{H}}\boldsymbol{H}\right\}$,

$$\hat{\boldsymbol{d}}_0 = \left[\boldsymbol{W}_{\text{LMMSE0}}^{\text{H}}\boldsymbol{y}_0\right] \oslash \boldsymbol{\alpha}_0 = \boldsymbol{d}_0 + \boldsymbol{v}_0,\ \text{E}\left[|\boldsymbol{v}_0[k]|^2\right] = \frac{1}{\boldsymbol{\rho}_0[k]} = \frac{1-\boldsymbol{\alpha}_0[k]}{\boldsymbol{\alpha}_0[k]}. \tag{3.49}$$

Here, $\boldsymbol{v}_0$ denotes the total noise plus interference. Therefore, $\boldsymbol{d}$ can be obtained such that

$$\hat{\boldsymbol{d}} = \boldsymbol{R}_d^{1/2}\hat{\boldsymbol{d}}_0 + \boldsymbol{\mu}_d = \boldsymbol{d} + \boldsymbol{R}_d^{1/2}\boldsymbol{v}_0. \tag{3.50}$$

When $\boldsymbol{R}_d = \boldsymbol{\Lambda}_d$ is diagonal, i.e. the symbols $\boldsymbol{d}[k]$ are independent but with different variance, the noise variance is scaled by $\boldsymbol{\Lambda}_d[k,k] = \text{E}\left[|\boldsymbol{d}_0[k] - \boldsymbol{\mu}_d[k]|^2\right]$, and hence,

$$\frac{1}{\boldsymbol{\rho}[k]} = \boldsymbol{\Lambda}_d[k,k]\frac{1-\boldsymbol{\alpha}_0[k]}{\boldsymbol{\alpha}_0[k]} = \frac{1}{\left(\boldsymbol{H}^{\text{H}}\left[\boldsymbol{H}\boldsymbol{\Lambda}_d\boldsymbol{H}^{\text{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1}\boldsymbol{H}\right)[k,k]} - \boldsymbol{\Lambda}_d[k,k]. \tag{3.51}$$

In general the overall LMMSE solution can be expressed as

$$\begin{aligned}\boldsymbol{W}_{\text{LMMSE}}^{\text{H}} = \boldsymbol{R}_d^{1/2}\boldsymbol{W}_{\text{LMMSE0}}^{\text{H}} &= \boldsymbol{R}_d\boldsymbol{H}^{\text{H}}\left[\boldsymbol{H}\boldsymbol{R}_d\boldsymbol{H}^{\text{H}} + \sigma^2\boldsymbol{I}_N\right]^{-1} \\ &= \left[\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \sigma^2\boldsymbol{R}_d^{-1}\right]^{-1}\boldsymbol{H}^{\text{H}},\end{aligned} \tag{3.52}$$

which can also be directly derived from minimizing $\text{E}\left[\|\boldsymbol{W}^{\text{H}}\boldsymbol{y} - \boldsymbol{d}\|^2\right]$. Note that, the gain fulfills $\boldsymbol{\alpha} = \text{diag}\left\{\boldsymbol{W}_{\text{LMMSE}}^{\text{H}}\boldsymbol{H}\right\} = \boldsymbol{\alpha}_0$.

General linear receiver

For a general receiver with a matrix $\boldsymbol{W}^{\text{H}}$ considering data vector $\boldsymbol{d}$ with covariance matrix $\boldsymbol{R}_d$ and mean $\boldsymbol{\mu}_d$, using $\bar{\boldsymbol{y}} = \boldsymbol{W}^{\text{H}}\boldsymbol{H}(\boldsymbol{d} - \boldsymbol{\mu}_d) + \boldsymbol{W}^{\text{H}}\boldsymbol{v}$

$$\begin{aligned}&\bar{\boldsymbol{y}}[k] = \boldsymbol{\alpha}[k](\boldsymbol{d}[k] - \boldsymbol{\mu}_d[k]) + \bar{\boldsymbol{z}}[k] + \bar{\boldsymbol{v}}[k] \\ &\Longrightarrow \hat{\boldsymbol{d}}[k] = \frac{1}{\boldsymbol{\alpha}[k]}\bar{\boldsymbol{y}}[k] + \boldsymbol{\mu}_d[k] = \boldsymbol{d}[k] + \frac{\bar{\boldsymbol{z}}[k] + \bar{\boldsymbol{v}}[k]}{\boldsymbol{\alpha}[k]}, \\ &\boldsymbol{\alpha} = \text{diag}\left\{\boldsymbol{W}^{\text{H}}\boldsymbol{H}\right\},\ \bar{\boldsymbol{v}} = \boldsymbol{W}^{\text{H}}\boldsymbol{v},\ \bar{\boldsymbol{z}} = \left(\boldsymbol{W}^{\text{H}}\boldsymbol{H} - \boldsymbol{\Lambda}^{(\boldsymbol{\alpha})}\right)(\boldsymbol{d} - \boldsymbol{\mu}_d)\end{aligned} \tag{3.53}$$

where, $\boldsymbol{\Lambda}^{(\boldsymbol{\alpha})} = \text{diag}\{\boldsymbol{\alpha}\}$ is a diagonal matrix. The SINR of the k-th channel is given by

$$\boldsymbol{\rho}[k] = \frac{\boldsymbol{\alpha}^2[k]}{\boldsymbol{\eta}^2[k] + \boldsymbol{\sigma}^2[k]} \tag{3.54}$$

where $\boldsymbol{\eta}^2$ and $\boldsymbol{\sigma}^2 \in \mathbb{R}^{+K\times 1}$ denote the vectors of interference and noise power, respectively,

$$\begin{aligned}\boldsymbol{\eta}^2[k] &= \mathrm{E}\left[|\bar{\boldsymbol{z}}[k]|^2\right] \Longrightarrow \boldsymbol{\eta}^2 = \mathrm{diag}\left\{\left(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H} - \boldsymbol{\Lambda}^{(\alpha)}\right)\boldsymbol{R}_d\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W} - \boldsymbol{\Lambda}^{(\alpha)}\right)\right\} \\ \boldsymbol{\sigma}^2[k] &= \mathrm{E}\left[|\bar{\boldsymbol{v}}[k]|^2\right] \Longrightarrow \boldsymbol{\sigma}^2 = \mathrm{diag}\left\{\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}\right\}.\end{aligned} \tag{3.55}$$

It can be seen that when $\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}$ is not diagonal, the interference vector $\boldsymbol{\eta}^2 \neq \boldsymbol{0}_K$. Otherwise, for instance using ZF, i.e. $\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H} = \boldsymbol{I}_K$, then $\boldsymbol{\eta}^2 = \boldsymbol{0}_K$. When $\boldsymbol{R}_d = \boldsymbol{\Lambda}_d$ is diagonal and using $\boldsymbol{W}_{\mathrm{LMMSE}}$, (3.52), then $\boldsymbol{\eta}^2[k] + \boldsymbol{\sigma}^2[k] = \boldsymbol{\alpha^2}[k]\,\boldsymbol{R}_d[k,k]\frac{1-\boldsymbol{\alpha}[k]}{\boldsymbol{\alpha}[k]}$, as shown in (3.51), which corresponds to the maximum SINR.

3.4.3 Iterative detection

This thesis introduces the concept of iterative detection in a generic model and highlight the common special cases. Without loss of generality, channel matrix is considered square $\boldsymbol{H} \in \mathbb{C}^{K\times K}$. This can be achieved, for instance, using QR decomposition (3.27). The data symbols $\boldsymbol{d}[k]$ are zero-mean independent and identically distributed (i.i.d.) selected from a constellation set $\mathcal{M}_c$, and $\boldsymbol{v} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_K)$. The iterative solution towards MLD solution is obtained by starting from a solution $\boldsymbol{d}^{(i)}$ at the i-th iteration, and determining the solution in the next iteration by $\boldsymbol{d}^{(i+1)} = \boldsymbol{d}^{(i)} + \boldsymbol{\Delta}_d^{(i)}$ where $\boldsymbol{\Delta}_d^{(i)}$ is calculated to get a smaller distance difference $\Delta^{(i)}$, such that

$$\begin{aligned}\Delta^{(i)} &= \|\boldsymbol{y} - \boldsymbol{H}\left[\boldsymbol{d}^{(i)} + \boldsymbol{\Delta}_d^{(i)}\right]\|^2 - \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}^{(i)}\|^2 < 0 \\ &= \|\boldsymbol{H}\boldsymbol{\Delta}_d^{(i)}\|^2 - 2\mathrm{Re}\left\{\boldsymbol{\Delta}_d^{(i)H}\boldsymbol{H}^{\mathrm{H}}\left[\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}^{(i)}\right]\right\}.\end{aligned} \tag{3.56}$$

By choosing $\boldsymbol{\Delta}_d^{(i)} = \boldsymbol{B}^{(i)}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{\Delta}_y^{(i)}$, with $\boldsymbol{B}^{(i)} = \boldsymbol{B}^{(i)H}$ where $\boldsymbol{\Delta}_y^{(i)} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}^{(i)}$, then

$$\Delta^{(i)} = \boldsymbol{\Delta}_y^{(i)H}\left(\left(\boldsymbol{H}\boldsymbol{B}^{(i)}\boldsymbol{H}^{\mathrm{H}}\right)^2 - 2\boldsymbol{H}\boldsymbol{B}^{(i)}\boldsymbol{H}^{\mathrm{H}}\right)\boldsymbol{\Delta}_y^{(i)}. \tag{3.57}$$

Using the SVD $\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Sigma}_h\boldsymbol{U}^{\mathrm{H}}$, $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}_h^2\boldsymbol{U}^{\mathrm{H}}$ and by choosing similar structure for $\boldsymbol{B}^{(i)}$ such that $\boldsymbol{B}^{(i)} = \boldsymbol{U}\boldsymbol{\Sigma}_b^{(i)}\boldsymbol{U}^{\mathrm{H}}$, then $\boldsymbol{B}^{(i)}\boldsymbol{H}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{\Sigma}_b^{(i)}\boldsymbol{\Sigma}_h\boldsymbol{V}^{\mathrm{H}} = \boldsymbol{W}^{(i)H}$, and

$$\Delta^{(i)} = \boldsymbol{\Delta}_y^{(i)H}\boldsymbol{V}\left(\boldsymbol{\Sigma}_h^4\boldsymbol{\Sigma}_b^{(i)2} - 2\boldsymbol{\Sigma}_h^2\boldsymbol{\Sigma}_b^{(i)}\right)\boldsymbol{V}^{\mathrm{H}}\boldsymbol{\Delta}_y^{(i)}.$$

To guarantee that $\Delta^{(i)} < 0$, it is sufficient that

$$\boldsymbol{\Sigma}_h^2[k,k]\boldsymbol{\Sigma}_b^{(i)}[k,k] \leq 2. \tag{3.58}$$

One solution is obtained when $\boldsymbol{\Sigma}_b^{(i)} = \alpha_h^{(i)}\boldsymbol{I}_K,\ \alpha_h^{(i)} > 0$, and thereby, $\boldsymbol{W}^{(i)H} = \sqrt{\alpha_h^{(i)}}\boldsymbol{H}^{\mathrm{H}}$, where $\alpha_h^{(i)} \leq \frac{2}{\boldsymbol{\Sigma}_h^2[k,k]}$. This is the solution corresponding to the Gradient descent. However, the general solution of $\boldsymbol{\Sigma}_b^{(i)2}$ depends also on $\boldsymbol{\Delta}_y^{(i)}$. Note that $\Delta^{(i)}$ is minimized, when $\boldsymbol{\Sigma}_b^{(i)2} = \boldsymbol{\Sigma}_h^{-2}$, i.e. $\boldsymbol{W}^{(i)H} = \boldsymbol{U}\boldsymbol{\Sigma}_h^{-1}\boldsymbol{V}^{\mathrm{H}}$, which corresponds to the least squares (LS) solution that minimizes $\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}\|^2$ when $\boldsymbol{d} \in \mathbb{C}^{K\times 1}$ is continuous vector. However, in this problem $\boldsymbol{d} \in \mathcal{S}_d \subset \mathcal{M}_c^K$ is discrete, and therefore, after finding $\boldsymbol{\Delta}_d^{(i)}$, the solution $\hat{\boldsymbol{d}}^{(i+1)} = \boldsymbol{d}^{(i)} + \boldsymbol{\Delta}_d^{(i)}$

may not belong to $\mathcal{S}_d$. Thus, an additional operation is required to map the obtained value to a valid codeword, such that $\hat{\boldsymbol{d}}^{(i+1)} = \Gamma_{\mathcal{S}_d}(\hat{\boldsymbol{d}}^{(i+1)})$, where $\Gamma_{\mathcal{S}_d}$ represents a demapping function. As a result, the iterative algorithm involves two steps

$$\begin{aligned} \hat{\boldsymbol{d}}^{(i+1)} &= \boldsymbol{d}^{(i)} + \boldsymbol{W}^{(i)H}\left(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}^{(i)}\right), \\ \boldsymbol{d}^{(i+1)} &= \Gamma_{\mathcal{S}_d}(\hat{\boldsymbol{d}}^{(i+1)}). \end{aligned} \tag{3.59}$$

The first is related to equalization, and the second is the demapping.

General iterative PIC receiver

This receiver employs equalization and soft decoding. It performs parallel interference cancellation (PIC) in three steps, as illustrated in Figure 3.9.

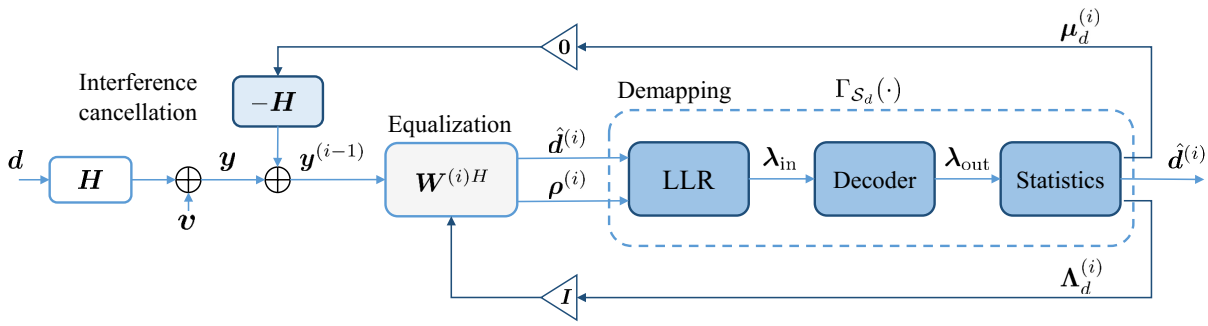

Figure 3.9: General iterative PIC receiver.

1. *Interference cancellation:* given $\boldsymbol{\mu}_d^{(i-1)}$ and $\boldsymbol{H}$, the signal to be equalized is

$$\boldsymbol{y}^{(i-1)} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{\mu}_d^{(i-1)} = \boldsymbol{H}\left(\boldsymbol{d} - \boldsymbol{\mu}_d^{(i-1)}\right) + \boldsymbol{v}. \tag{3.60}$$

   This requires the computation of $\boldsymbol{H}\boldsymbol{\mu}_d^{(i-1)}$.

2. *Linear equalization:* a linear receiver with matrix $\boldsymbol{W}^{(i)}$ is applied on $\boldsymbol{y}^{(i-1)}$. Considering $\boldsymbol{R}_d = \boldsymbol{\Lambda}_d^{(i-1)}$, the equalized data $\hat{\boldsymbol{d}}^{(i)}$ and SINR $\boldsymbol{\rho}^{(i)}$ vector are computed using (3.53) and (3.54), respectively. These samples are forwarded to the demapper to find a valid codeword.

3. *Demapping:* the demapper employs soft-decoding, which requires the computation of the LLRs. For that, the demapper also needs the SINR values $\boldsymbol{\rho}^{(i)}$. The decoder provides LLRs, from which the bit and symbol probabilities are calculated using (3.33) and (3.34), respectively. The demapper then calculates the probabilities $\Pr(\boldsymbol{d}^{(i)}[k] = d_q)$, $d_q \in \mathcal{M}_c$, where the hard decision is the value with maximum probability. Using these probabilities, the statistical values are computed such that $\boldsymbol{\mu}_d^{(i)} = \mathrm{E}\left[\boldsymbol{d}^{(i)}\right]$ and $\boldsymbol{\Lambda}_d^{(i)}[k,k] = \mathrm{E}\left[|\boldsymbol{d}^{(i)}[k] - \boldsymbol{\mu}_d^{(i)}[k]|^2\right]$. The demapping steps can be summarized by

$$\left(\boldsymbol{d}^{(i)},\ \boldsymbol{\mu}_d^{(i)},\ \boldsymbol{\Lambda}_d^{(i)}\right) = \Gamma_{\mathcal{S}_d}(\hat{\boldsymbol{d}}^{(i)}, \boldsymbol{\rho}^{(i)}) \tag{3.61}$$

These steps are repeated at each iteration. The initial values are, $\boldsymbol{y}^{(0)} = \boldsymbol{y}$, $\boldsymbol{\mu}_d^{(0)} = \boldsymbol{0}_K$, and $\boldsymbol{\Lambda}_d^{(0)} = \boldsymbol{I}_K$. The distance difference $\Delta_d^{(i)} = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}^{(i+1)}\|^2 - \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{d}^{(i)}\|^2$ is evaluated. The iteration stops when $\Delta_d^{(i)} > 0$ and the decision is $\boldsymbol{d}^{(i)}$. When $-\Delta_d^{(i)} < \epsilon$, where ϵ is a threshold, the algorithm stops and the decision is $\boldsymbol{d}^{(i+1)}$.

As an example, LMMSE-PIC employs LMMSE equalization [SFS11], [MZF18]. At its i-th iteration, $\boldsymbol{R}_d = \boldsymbol{\Lambda}_d^{(i)}$ and the exact LMMSE matrix is calculated from (3.52) as

$$\boldsymbol{W}_{\text{LMMSE}}^{(i)H} = \boldsymbol{U}\left[\boldsymbol{\Sigma}_h^2 + \sigma^2 \boldsymbol{U}^{\text{H}} \boldsymbol{\Lambda}_d^{(i)-1} \boldsymbol{U}\right]^{-1} \boldsymbol{U}^{\text{H}} \boldsymbol{H}^{\text{H}}, \tag{3.62}$$

This requires computing the inverse $\left[\boldsymbol{\Sigma}_h^2 + \sigma^2 \boldsymbol{U}^{\text{H}} \boldsymbol{\Lambda}_d^{(i)-1} \boldsymbol{U}\right]^{-1}$. However, if $\boldsymbol{\Lambda}_d^{(i)-1} = \lambda_i \boldsymbol{I}_K$, the matrix to be inverted becomes diagonal. Therefore, by replacing $\boldsymbol{\Lambda}_d^{(i)-1}$ with $\lambda_i \boldsymbol{I}_K$, where $\lambda_i = \frac{1}{K}\text{trace}\left\{\boldsymbol{\Lambda}_d^{(i)}\right\}$, the updating matrix becomes

$$\boldsymbol{W}^{(i)H} = \boldsymbol{U}\left[\boldsymbol{\Sigma}_h^2 + \frac{\sigma^2}{\lambda_i}\boldsymbol{I}_K\right]^{-1} \boldsymbol{U}^{\text{H}} \boldsymbol{H}^{\text{H}} = \boldsymbol{U}\boldsymbol{\Sigma}_b^{(i)} \boldsymbol{U}^{\text{H}} \boldsymbol{H}^{\text{H}}. \tag{3.63}$$

Noting that $\boldsymbol{\Sigma}_h^2[k,k]\boldsymbol{\Sigma}_b^{(i)}[k,k] = \frac{\boldsymbol{\Sigma}_h^2[k,k]}{\boldsymbol{\Sigma}_h^2[k,k] + \frac{\sigma^2}{\lambda_i}} \leq 1$, the convergence condition (3.59) is fulfilled. When $\frac{\sigma^2}{\lambda_i} << \Sigma_{h,\min}^2$, $\Sigma_{h,\min}^2 = \min_k \boldsymbol{\Sigma}_h^2[k,k]$, ZF equalizer $\boldsymbol{W}^{(i)H} = \boldsymbol{H}^{-1}$ can be used. This corresponds to the interference region, where the interference in higher than the noise. On the contrary, when the variance of the estimated symbols is very small, such that $\frac{\sigma^2}{\lambda_i} >> \Sigma_{h,\max}^2$, then $\boldsymbol{W}^{(i)H} = \frac{\lambda_i}{\sigma^2}\boldsymbol{H}^{\text{H}}$. As a result, changing $\boldsymbol{W}^{(i)}$ should follow the need of the current iteration not necessary based on systematic approach.

3.4.4 Numerical examples and insights

This section aims at demonstrating the detection techniques for different modulation schemes considering the received signal model $\boldsymbol{y} = \boldsymbol{H}^{(e)}\boldsymbol{A}\boldsymbol{d} + \boldsymbol{v}$, where $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ is the core modulation matrix, $\boldsymbol{d} \in \mathcal{M}_c^N$ data symbol vector, $\boldsymbol{H}^{(e)} \in \mathbb{C}^{N\times N}$ equivalent channel, and $\boldsymbol{v} \in \mathbb{C}^{N\times 1}$ AWGN vector with variance N_0. The data symbols are uncorrelated with $\boldsymbol{R}_d = E_s \boldsymbol{I}_N$, and the modulation matrix is normalized such that $\text{trace}\left\{\boldsymbol{A}^{\text{H}}\boldsymbol{A}\right\} = N$. Thereby, $\rho = \frac{E_s}{N_0}$ defines the signal-to-noise ratio (SNR) per symbol. The channel is considered known at the receiver and fixed during the transmission of a frame of K_b information bits, which are encoded with channel coding of rate R_c.

SISO system

In single-input, single-output (SISO) systems that employ CP in time-invariant multipath channel, the matrix $\boldsymbol{H}^{(e)}$ becomes circular and generated from the channel discrete impulse response $h[l]$, $l = 0, \cdots, L-1$, $L < N$, such that $\boldsymbol{H}^{(e)}[n,q] = h[\langle n-q\rangle_N]$. Therefore, $\boldsymbol{H}^{(e)} = \boldsymbol{F}_N \boldsymbol{\Lambda}^{(\tilde{h})} \boldsymbol{F}_N^{\text{H}}$, where $\boldsymbol{\Lambda}^{(\tilde{h})} = \text{diag}\left\{\tilde{\boldsymbol{h}}\right\}$, $\tilde{\boldsymbol{h}}[n] = \sum_{l=0}^{L-1} h[l] e^{-j2\pi\frac{nl}{N}}$. The channel taps are assumed uncorrelated with Rayleigh distribution, i.e. $\text{E}\left[h[l_1]h^*[l_2]\right] = \mathcal{P}_l \delta[l_1 - l_2]$, where

$\mathcal{P}_l$, is the power of the l-th tap, then $|\tilde{\boldsymbol{h}}[n]|$ has Rayleigh distribution with the power $\mathrm{E}\left[|\tilde{\boldsymbol{h}}[n]|^2\right] = \sum_{l=0}^{L-1} \mathrm{E}\left[|h[l]|^2\right] = \sum_{l=0}^{L-1} \mathcal{P}_l = \mathcal{P}$.

The equivalent signal model after normalized N-DFT becomes $\tilde{\boldsymbol{y}} = \boldsymbol{\Lambda}^{(\tilde{h})}\tilde{\boldsymbol{A}}\boldsymbol{d}+\tilde{\boldsymbol{v}}$, with $\tilde{\boldsymbol{A}} = \boldsymbol{F}_N\boldsymbol{A}$ is the frequency-domain modulation matrix, and $\tilde{\boldsymbol{v}} = \boldsymbol{F}_N\boldsymbol{v}$, $\mathrm{E}\left[\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^H\right] = N0\boldsymbol{I}_N$. This practical case belongs to precoded OFDM system as reported in [NCF19c]. In particular, $\tilde{\boldsymbol{A}}$ is orthogonal and $\boldsymbol{R}_d = E_s\boldsymbol{I}_N$. Thus, the LMMSE filter is given by

$$\boldsymbol{W}_{\mathrm{LMMSE}}^{\mathrm{H}} = \tilde{\boldsymbol{A}}^{\mathrm{H}}\left[\boldsymbol{\Lambda}^{(\tilde{h})}\boldsymbol{\Lambda}^{(\tilde{h})H} + \frac{N_0}{E_s}\boldsymbol{I}_N\right]^{-1}\boldsymbol{\Lambda}^{(\tilde{h})\mathrm{H}} = \tilde{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{\Lambda}^{(\tilde{h}_{\mathrm{eq}})H}. \tag{3.64}$$

Note that, this linear detection belongs to the case, where frequency-domain channel equalization is performed first using $\boldsymbol{\Lambda}^{(\tilde{h}_{\mathrm{eq}})} = \mathrm{diag}\left\{\tilde{\boldsymbol{h}}_{\mathrm{eq}}\right\}$, $\tilde{\boldsymbol{h}}_{\mathrm{eq}}[n] = \frac{\tilde{\boldsymbol{h}}[n]}{|\tilde{\boldsymbol{h}}[n]|^2+\frac{\sigma^2}{E_s}}$. Therefore,

$$\boldsymbol{\alpha}[n] = \sum_{q=0}^{N-1}|\tilde{\boldsymbol{A}}[n,q]|^2\frac{|\tilde{\boldsymbol{h}}[q]|^2}{|\tilde{\boldsymbol{h}}[q]|^2+\frac{N_0}{E_s}}, \text{ and } \boldsymbol{\rho}[n] = \frac{\boldsymbol{\alpha}[n]}{1-\boldsymbol{\alpha}[n]}. \tag{3.65}$$

In addition, assuming the interference is resolved, which is denoted as the genie case, the achieved SNR corresponding to the MF is given by

$$\boldsymbol{\rho}_{\mathrm{genie}}[n] = \frac{E_s}{N_0}\sum_{q=0}^{N-1}|\tilde{\boldsymbol{A}}[n,q]|^2|\tilde{\boldsymbol{h}}[q]|^2. \tag{3.66}$$

This genie case is introduced in this work to provide upper bound performance benchmark for iterative interference cancellation receivers.

OFDM vs DFT-spreading. When $\tilde{\boldsymbol{A}} = \boldsymbol{I}_N$, which is the case of OFDM, the data symbols can be separated without interference, i.e. $\tilde{\boldsymbol{y}}[n] = \tilde{\boldsymbol{h}}[n]\boldsymbol{d}[n] + \tilde{\boldsymbol{v}}[n]$. As a result, $\boldsymbol{\rho}_{\mathrm{OFDM}}[n] = \frac{E_s}{N_0}|\tilde{\boldsymbol{h}}[n]|^2$, $\mathrm{E}\left[\boldsymbol{\rho}_{\mathrm{OFDM}}[n]\right] = \mathcal{P}$ and the distribution of $|\tilde{\boldsymbol{h}}[n]|$ is Rayleigh. The single-carrier (SC) is obtained when $\tilde{\boldsymbol{A}} = \boldsymbol{F}_N$, which leads to high ISI, and it fulfills the equal-gain criterion $|\tilde{\boldsymbol{A}}[n,q]|^2 = \frac{1}{N}$. Therefore, the genie SNR is given by

$$\boldsymbol{\rho}_{\mathrm{genie,\ SC}}[n] = \frac{E_s}{N_0}\frac{1}{N}\sum_{q=0}^{N-1}|\tilde{\boldsymbol{h}}[q]|^2 = \frac{E_s}{N_0}\sum_{l=0}^{L-1}|h[l]|^2 = \rho_{\mathrm{genie,\ SC}}. \tag{3.67}$$

The distribution of $\rho_{\mathrm{genie,\ SC}}$ depends on the channel power delay profile (PDP) through the sum $\sum_{l=0}^{L-1}|h[l]|^2$, which corresponds to full frequency diversity. When the channel taps are independent and Rayleigh-distributed, the distribution of $\rho_{\mathrm{genie,\ SC}}$ is Chi-square with $2L$ degrees of freedom, and also $\mathrm{E}\left[\rho_{\mathrm{genie,\ SC}}\right] = \mathcal{P}$. This can be seen as maximum ratio combining (MRC) of the signal received from L independent paths.

As a trade-off between interference free OFDM and full spreading SC, partial spreading can be applied on a group of symbols of size M such that the received signal is decoupled into K independent streams, with $N = KM$. Each subsymbol can be independently generated by a modulation matrix $\tilde{\boldsymbol{A}}_k \in \mathbb{C}^{M\times M}$, such that

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{\Lambda}^{(\tilde{h}_k)}\tilde{\boldsymbol{A}}_k\boldsymbol{d}_k + \tilde{\boldsymbol{v}}_k, \tag{3.68}$$

where $\boldsymbol{\Lambda}^{(\tilde{h}_k)} = \text{diag}\left\{\tilde{\boldsymbol{h}}_k\right\}$, $\tilde{\boldsymbol{h}}_k[m] = \tilde{\boldsymbol{h}}[n_{k,m}]$, and $n_{k,m}$ is a mapping index. Using DFT spreading $\tilde{\boldsymbol{A}}_k = \boldsymbol{F}_M$, the equal gain can be achieved within each subsymbol. Therefore,

$$\rho_{k,\text{genie}} = \frac{E_s}{N_0}\frac{1}{M}\sum_{m=0}^{M-1}|\tilde{\boldsymbol{h}}_k[m]|^2, \text{ and } \mathrm{E}[\rho_{k,\text{genie}}] = \mathcal{P}. \tag{3.69}$$

This concept has been reported in [NCF19c], and more details on the design will be discussed in the context of time-frequency multicarrier in Chapter 5.

The distribution of $\rho_{k,\text{genie}}$ depends on $\tilde{\boldsymbol{h}}_k$. Two spreading schemes are considered. The first corresponds to contiguous spreading, where $n^{(1)}_{k,m} = m + kM$, and hence $\boldsymbol{A} = \boldsymbol{I}_K \otimes \boldsymbol{F}_M$. The SNR is given by,

$$\begin{aligned}\rho^{(1)}_{k,\text{genie}} &= \frac{E_s}{N_0}\frac{1}{M}\sum_{m=0}^{M-1}|\tilde{\boldsymbol{h}}[m+kM]|^2 \\ &= \frac{E_s}{N_0}\frac{1}{M}\sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}h[l_1]h^*[l_2]e^{-j2\pi\frac{k(l_1-l_2)}{K}}\sum_{m=0}^{M-1}e^{-j2\pi\frac{m(l_1-l_2)}{N}}.\end{aligned} \tag{3.70}$$

Under the assumption of uncorrelated Rayleigh fading multi tap,the distribution of $\rho^{(1)}_{k,\text{genie}}$ is independent of k, but it mainly depends on M.

The other is interleaved spreading with $n^{(2)}_{m,k} = k + mK$, and $\boldsymbol{A} = \boldsymbol{F}_M \otimes \boldsymbol{I}_K$. The SNR is

$$\begin{aligned}\rho^{(2)}_{k,\text{genie}} &= \frac{E_s}{N_0}\frac{1}{M}\sum_{m=0}^{M-1}|\tilde{\boldsymbol{h}}[k+mK]|^2 \\ &= \frac{E_s}{N_0}\frac{1}{M}\sum_{l_1=0}^{L-1}\sum_{l_2=0}^{L-1}h[l_1]h^*[l_2]e^{-j2\pi\frac{k(l_1-l_2)}{N}}\sum_{m=0}^{M-1}e^{-j2\pi\frac{m(l_1-l_2)}{M}}.\end{aligned} \tag{3.71}$$

When $M \geq L$, then $\sum_{m=0}^{M-1} e^{-j2\pi\frac{m(l_1-l_2)}{M}} = M\delta[l_2 - l_1]$, and hence, $\rho^{(2)}_{k,\text{genie}} = \frac{E_s}{N_0}\sum_{l=0}^{L-1}|h[l]|^2$, which achieves full frequency diversity.

WHT spreading. The equal-gain spreading can be also achieved using any orthogonal matrices that fulfills $|\tilde{\boldsymbol{A}}[q,n]|^2 = \frac{1}{N}$. In particular Walsh-Hadamard transform (WHT) matrix $\boldsymbol{\Omega}_N \in \mathbb{R}^{N\times N}$ defined such that

$$\boldsymbol{\Omega}_N = \boldsymbol{\Omega}_2 \otimes \boldsymbol{\Omega}_{N/2},\ \boldsymbol{\Omega}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}. \tag{3.72}$$

WHT does not require complex multiplications, and it can be performed with sum and scaling. This allows complexity reduction of the equalization, especially when exploiting the LMMSE approximation given by (3.63).

Mutual information evaluation

The theoretical modulation performance is evaluated by means of the average mutual information $\mathrm{E}\left[I(\tilde{\boldsymbol{h}})\right]$ measured in bit/Symbol. For i.i.d. Gaussian symbols with power E_s,

the theoretical rate without prior knowledge of the channel at the transmitter is given by

$$\begin{aligned} I_{\text{Gaussian}}(\tilde{\boldsymbol{h}}) &= \frac{1}{N}\log_2\left(\det\left(\frac{E_s}{N_0}\tilde{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{\Lambda}^{(\tilde{h})\mathrm{H}}\boldsymbol{\Lambda}^{(\tilde{h})}\tilde{\boldsymbol{A}}+\boldsymbol{I}_N\right)\right) \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\log_2\left(\frac{E_s}{N_0}|\tilde{\boldsymbol{h}}[n]|^2+1\right), \text{ [bit/symbol].} \end{aligned} \tag{3.73}$$

This theoretical rate is independent of the selected orthogonal modulation matrix $\tilde{\boldsymbol{A}}$, and $\mathrm{E}\left[I_{\text{Gaussian}}(\tilde{\boldsymbol{h}})\right]$ is equivalent to the ergodic capacity of Rayleigh flat fading channel with power $\mathcal{P}$. In practice, the data symbols are selected from a finite constellation set $\mathcal{M}_c = \{d_0, \cdots, d_{M_c-1}\}$ with uniform distribution, i.e. $P(d = d_m) = \frac{1}{M_c}$. The achievable rate for a single stream $y = d + v$, where v is AWGN, $\mathrm{E}\left[|d|^2\right] = 1$ and $\mathrm{E}\left[|v|^2\right] = \frac{1}{\rho}$, can be computed from the mutual information $I(y;d)$ as

$$\begin{aligned} I(\mathcal{M}_c,\rho) &= \int_y \sum_{m=0}^{M_c-1} f(y,d_m)\log_2\left(\frac{f(y,d_m)}{f(y)p(d_m)}\right) \\ &= \log_2(M_c) - \log_2(e) - \frac{1}{M_c\pi\rho}\int_y \sum_{m=0}^{M_c-1} e^{-\frac{|y-d_m|^2}{\rho}}\log_2\left(\sum_{m=0}^{M_c-1} e^{-\frac{|y-d_m|^2}{\rho}}\right)dy. \end{aligned} \tag{3.74}$$

The average rate is influenced by the distribution of ρ, which depends on the modulation matrix and the constellation set. The maximum achievable rate is limited to $\log_2(M_c)$.

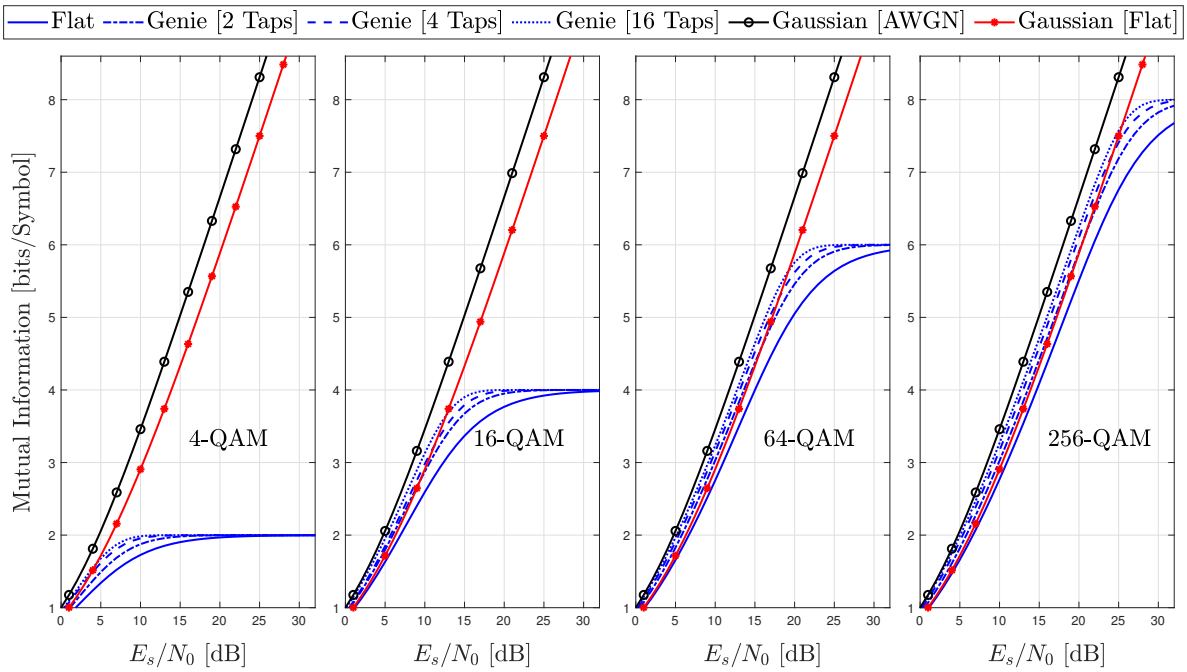

Figure 3.10: Average mutual information for QAM mapping in SISO system with frequency selective channel under the assumption of perfect interference cancellation (Genie) and uniform power delay profile with L taps.

Figure 3.10 illustrates the impact of channel, constellation, and modulation on the achievable rate. It gives insight on the design parameters with respect to the channel and

SNR region. These parameters are namely, the QAM order, the code rate, and modulation matrix. In this evaluation, the PDP consists of L taps uniformly distributed with average power $\mathcal{P} = 1$, and thus, $\mathcal{P}_l = \frac{1}{L}$. Different order QAM constellation are used. The case of Gaussian flat, is equivalent to the theoretical mutual information using Gaussian symbols, which is independent of the modulation matrix. The flat case corresponds to OFDM, which is independent of the number of taps, whereas the other genie cases represent spreading with full diversity gain. In fact, the genie case is an upper bound for the possibly achieved mutual information with QAM mapping, and this bound can exceed the theoretical flat fading with Gaussian symbols. The exact theoretical achievable rate requires the computation of $f(\boldsymbol{y}|\boldsymbol{d}_q)$, for all the N^{M_C} combinations of $\boldsymbol{d}_q \in \mathcal{M}_c^N$, which is extremely complex. A low-complexity approximation is out of the scope and this theses considers the genie upper bound instead.

It can be seen, that the gap between the theoretical mutual information with Gaussian symbols and the one with QAM symbols at sufficiently high SNRs can be reduced in multipath channel using spreading with iterative interference cancellation receiver. With the increase of the number of uncorrelated channel taps, the gap decreases due to the exploitation of frequency diversity by means of spreading. Note that, with the increase of L, the genie also approaches the AWGN as a result of larger degrees of freedom of the Chi-square distribution of $\rho_{\mathrm{SC},genie}$. However, the ISI also increases with the increase of L, which degrades the performance of interference cancellation process. Moreover, using high QAM order at lower SNRs has also the potential of approaching the theoretical rate using OFDM. The drawback in this case is the increase of peak-to-average power ratio (PAPR). For instance, at 5 dB, SC using 4-QAM with proper interference cancellation may achieve the theoretical rate with very low PAPR compared to OFDM using 256-QAM.

In fact, there are several options to achieve the same performance. In relation to the waveforms, it is sufficient to perform partial interleaved spreading with $M \geq L$. For DFT spreading, the larger is the spearheading towards SC, the lower is the PAPR, but the equalization becomes computationally more complex. Other spreading matrices, such WHT, reduce the complexity at the cost of worsening the PAPR. OFDM with very high QAM and low code rate is able to approach the theoretical rate. However, the PAPR increases leading to high interference because of the non-linearity of power amplifier (PA). Additionally, the computation complexity of the exact LLR increases or they become unreliable when calculated with low-complexity approximation. The optimal parameter selection is subject to intensive simulation, which is out of the scope of this work. Next, some typical configurations are used to evaluate the detection performance.

SER, coded BER and FER evaluation

In this simulation, a frame of $K_p = 256$ bits is encoded using basis 1/2 convolutional encoder $(171, 133)$, and code rate of $\{1/2, 2/3, 3/4, 5/6\}$ with $\{4, 16, 64, 256\}$-QAM,

respectively. The code rates larger than 1/2 are generated with puncturing. This encoder serves as an example for this simulation because it is involved in standards, e.g. IEEE 802.11, and widely used in literature. The evaluation of other channel encoding schemes is beyond the scope of this thesis. The encoded bits are interleaved such that $\boldsymbol{c}_{\text{int}}[p + qL_c] = \boldsymbol{c}[q + pK_s]$, where $L_c = \log_2(M_c)$, and K_s is the number of symbols in the frame. Gray mapping is employed prior to the QAM mapping. The block length $N = 64$, and the channel is block fading and static within the duration of the frame. An exponential channel delay profile is used as an example with $L = 10$ taps, and $\mathcal{P}_l = \frac{\exp\left(-\frac{l}{\gamma}\right)}{\sum_{l=0}^{L-1}\exp\left(-\frac{l}{\gamma}\right)}$. Here, the channel frequency selectivity increases with the increase of γ. For frame validation, the cyclic redundancy check (CRC) of polynomial (1021) is used. The simulation parameters are listed in Table 3.3.

Table 3.3: Simulation parameters.

Multipath channel PDF	$\mathcal{P}_l = \frac{\exp\left(-\frac{l}{\gamma}\right)}{\sum_{l=0}^{L-1}\exp\left(-\frac{l}{\gamma}\right)}$, $L = 10$, $\gamma = \{1, 3\}$.
MCS (M_c-QAM, code rate R_c)	$(4, 1/2)$, $(16, 2/3)$, $(64, 3/4)$, $(256, 5/6)$.
Block length and active set	$N = 64$, $\mathcal{N}_{on} = \{0, \cdots N-1\}$.
Bit mapping and channel encoder	Gray mapping, convolutional code $(171, 133)$.
Frame parameters	$K_b = 64$ information bits, K_s symbols, CRC(1021)
Interleaver	$\boldsymbol{c}_{\text{int}}[p + qL_c] = \boldsymbol{c}[q + pK_s]$, L_c bits/symbol

Uncoded SER. After equalization, the SER per channel realization is given by

$$\text{SER}(\boldsymbol{\rho}) = \frac{1}{N}\sum_{n=0}^{N-1}\frac{2(M_c-1)}{M_c}Q\left(\sqrt{\frac{3}{M_c^2-1}\boldsymbol{\rho}[n]}\right), \text{ where, } Q(x) = \int_{-\infty}^{-x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt. \tag{3.75}$$

The average SER w.r.t channel, $\text{SER} = \text{E}[\text{SER}(\boldsymbol{\rho})]$, depends on the distribution of $\boldsymbol{\rho}[n]$. Figure 3.11 shows the SER for OFDM and SC using LMMSE equalization and genie with perfect ISI cancellation under low and high frequency selectivity. The average SER for OFDM is independent of the delay profile, whereas the SER decreases with high selectivity for the genie aided interference cancellation. The performance of LMMSE depends on the QAM order, the channel selectivity, and the SNR. In particular, the SER decreases at high SNR, and starting from a threshold depending on the QAM order, it becomes lower than that of OFDM.

Higher channel selectivity ($\gamma = 3$) in comparison with low selectivity ($\gamma = 1$) is beneficial, when spreading is employed, as can be seen in the genie case. Figure 3.12 shows a comparison with different spreading length for both contiguous and interleaved case. As can be seen, the SER decreases with the increase of the spreading length starting from $M = 1$ in OFDM to maximum spreading $M = N$ in SC. Moreover, the interleaved

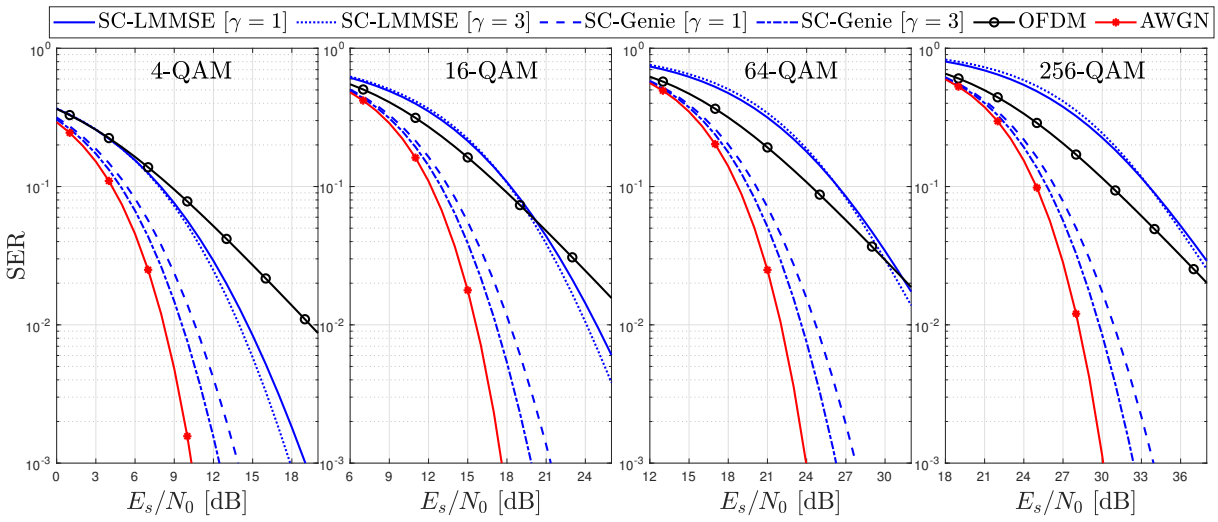

Figure 3.11: SER in different frequency selective channel. Low ($\gamma = 1$), and high ($\gamma = 3$) .

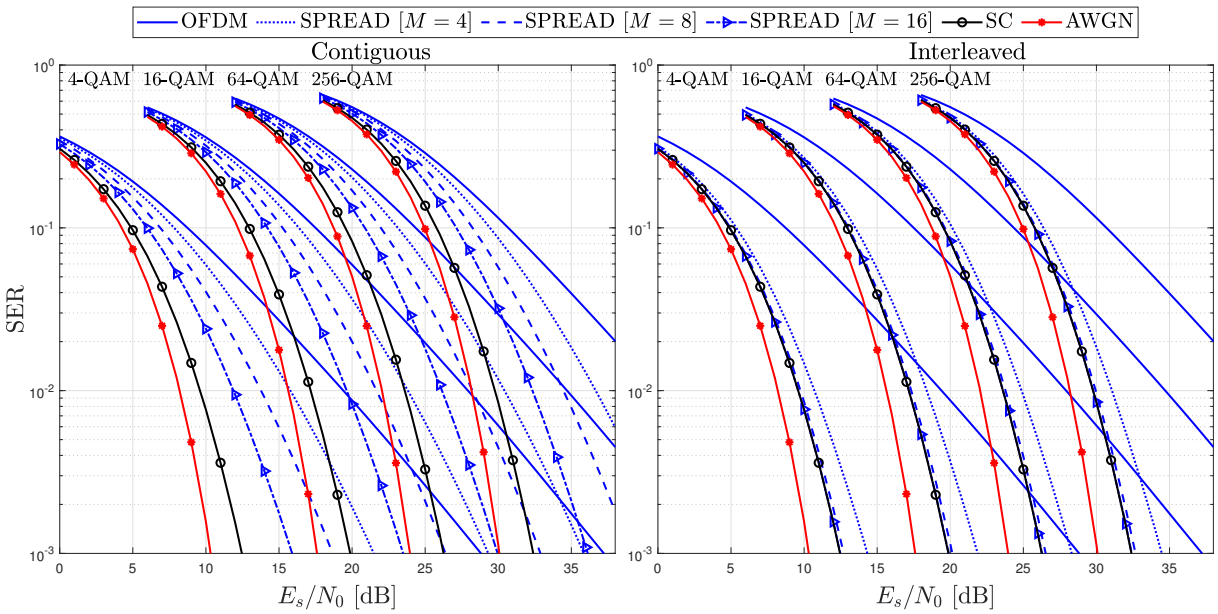

Figure 3.12: SER in different spreading with perfect ISI cancellation ($\gamma = 3$).

spreading reaches the SC performance with smaller M in comparison with the contiguous one. This because the channel gains in the interleaved case are less correlated than in the contiguous case, which allows gaining better frequency diversity. For $M = 16 > L$, the same performance as fully spread SC is achieved following (3.71).

In relation with the information rate evaluation, Figure 3.10, it can be seen that the spreading with perfect interference cancellation, converges faster to the error free transmission. For instance, 2 bit/symbols is attained without coding by employing 4-QAM at 15 dB and lower for high selectivity, whereas it requires 25 dB for OFDM. Nevertheless, as a result of high ISI, it can be observed, first, the performance of LMMSE degrades at low SNR with the increase of the selectivity. Second, the efforts required to resolve the

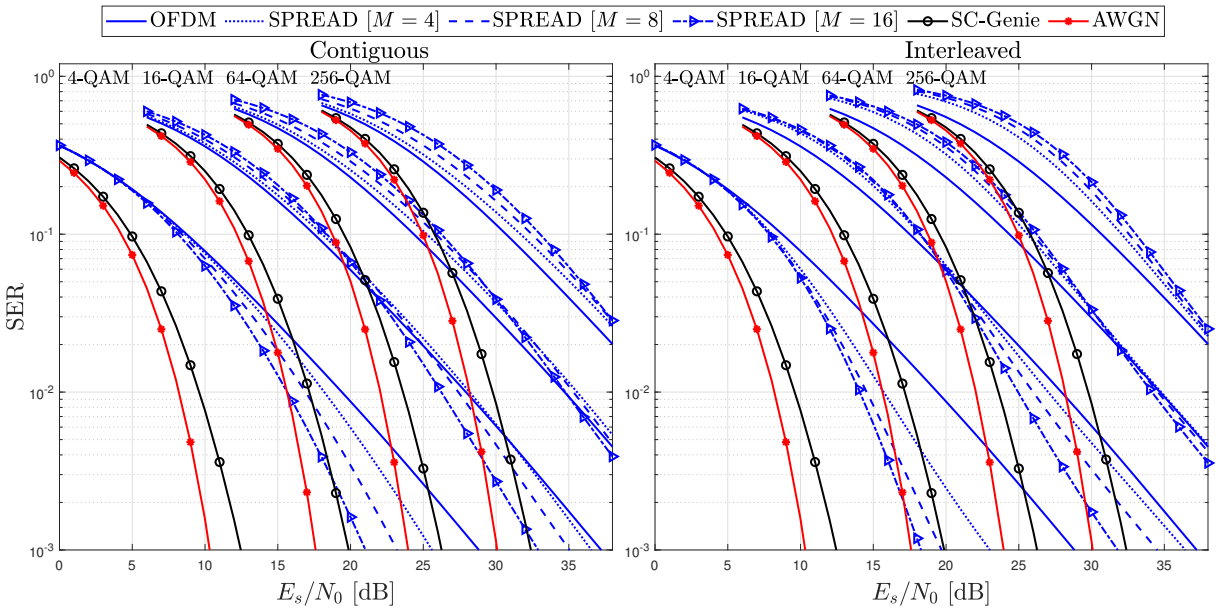

Figure 3.13: SER for different spreading with LMMSE equalization ($\gamma = 3$).

interference increases with the increase of the QAM order, which can be seen from the larger gap between the genie and LMMSE. The LMMSE performance is also superior in the interleaved case compared to the contiguous one, especially at high SNRs and lower QAM orders, as illustrated in Figure 3.12. Moreover, at high order QAM, lower spreading helps in reducing the ISI at lower SNR.

Coded BER and FER. The coded bit error rate (BER) shows different behavior when soft decoding is used. Figure 3.14 illustrates the coded BER for SC and OFDM with and without interleaving. The interleaver is applied to spread the coded bits over the QAM symbols. First, the BER of SC-LMMSE is worse than that of OFDM, especially at high order QAM because of the low reliable LLR approximation. The interference cancellation receiver, namely LMMSE-PIC helps in reducing the BER significantly when high order QAM is used. However, it is remarkable, that with the increase of the QAM order the interference cancellation is more challenging. Second, the interleaving influences the BER, with a significant increase of the BER for SC-LMMSE and slight decrease in the case of OFDM at high SNRs. This improvement in OFDM is due to the spreading of the encoded bits over uncorrelated subcarriers in addition to the implicit spreading of the information bits via the encoder. Third, as expected, higher selective channel allows harvesting frequency diversity. This is notable by a gain of about 3 dB in the case of high selective channel ($\gamma = 3$) compared to lower selectivity ($\gamma = 1$). Nevertheless, the average BER curves do not give an idea about the distribution of the bit errors within the frame. When the error nature is burst, the frame error rate (FER) becomes higher although the BER can be low. On the contrary, the BER can be high, but the erroneous bits can be localized leading to lower FER.

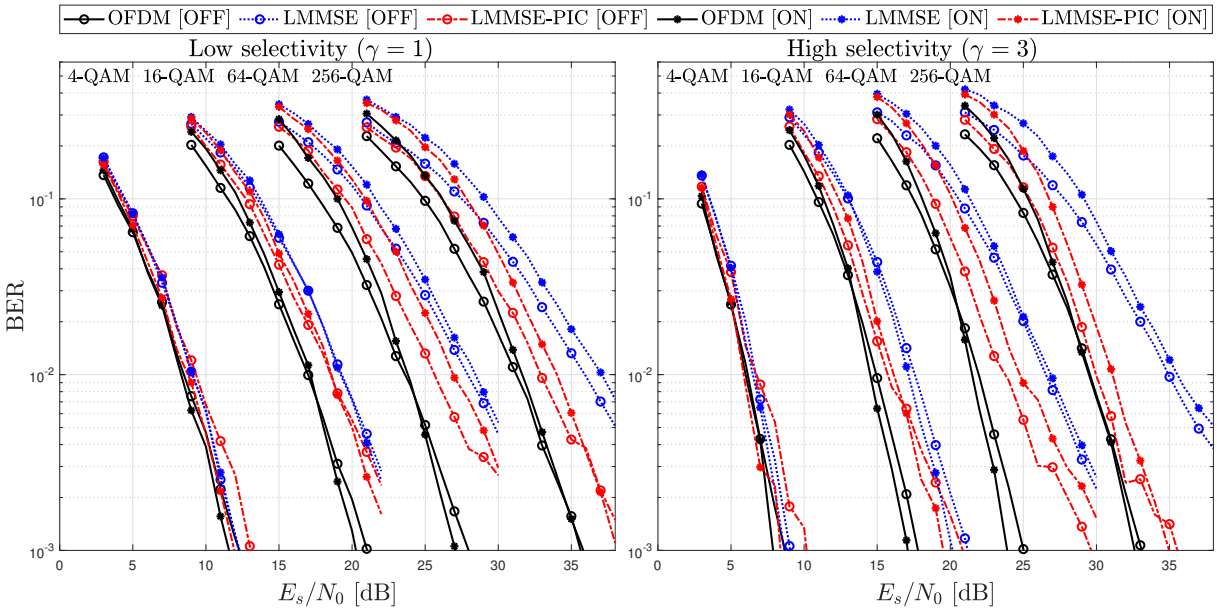

Figure 3.14: BER comparison of SC and OFDM with interleaving [on] and non-interleaving [off] for different selectivity; high selectivity $\gamma = 3$ and low-selectivity $\gamma = 1$.

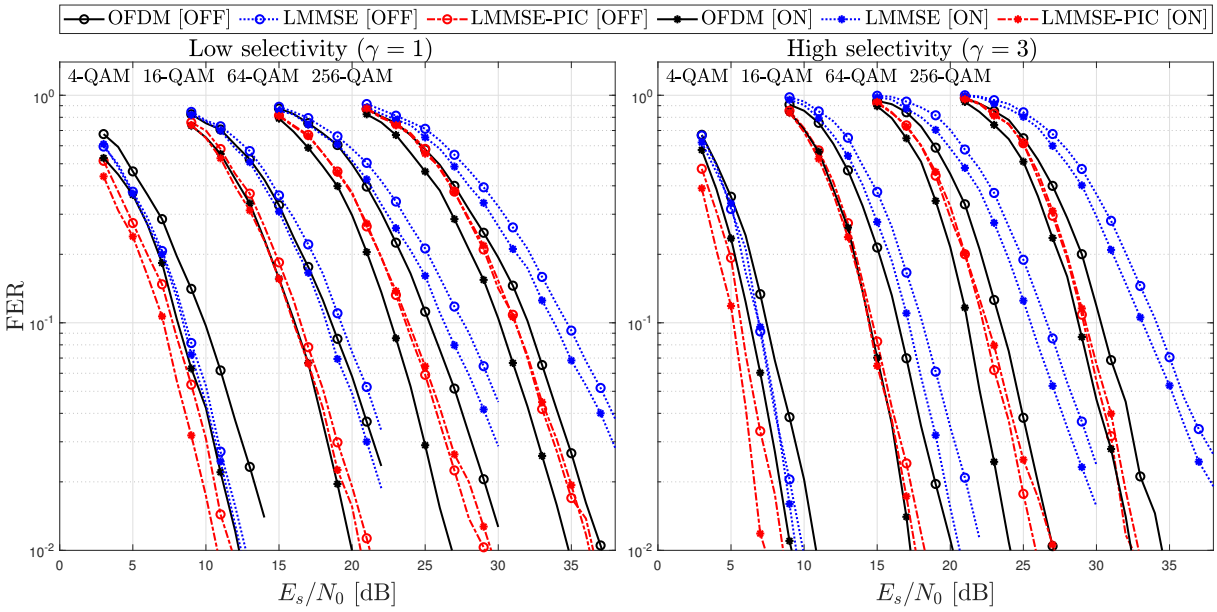

Figure 3.15: FER comparison of SC and OFDM with interleaving [on] and non-interleaving [off] for different selectivity; high selectivity $\gamma = 3$ and low-selectivity $\gamma = 1$.

The FER is illustrated in Figure 3.15, where a CRC is used to validate the decoded frame and terminate the iteration of LMMSE-PIC. It can be shown, that interleaving is essential to tackle the burst error. In particular, the FER performance of OFDM significantly improves by 3 dB gain, and the improvement of SC-LMMSE is about 1 dB with the employed interleaver. The study and evaluation of optimal interleaving schemes are out of the scope of this thesis. However, OFDM still outperforms SC-LMMSE by 2 dB employing 16-QAM up to 4 dB with 256-QAM. Moreover, the high selectivity of the

channel brings benefits due to the multipath diversity. In this case, the spreading with interference cancellation significantly outperforms OFDM at 4-QAM and becomes closer at higher order QAMs. The FER depends also on the frame length, i.e. the codeword length. Evaluating the frame length is also out of the scope of this work.

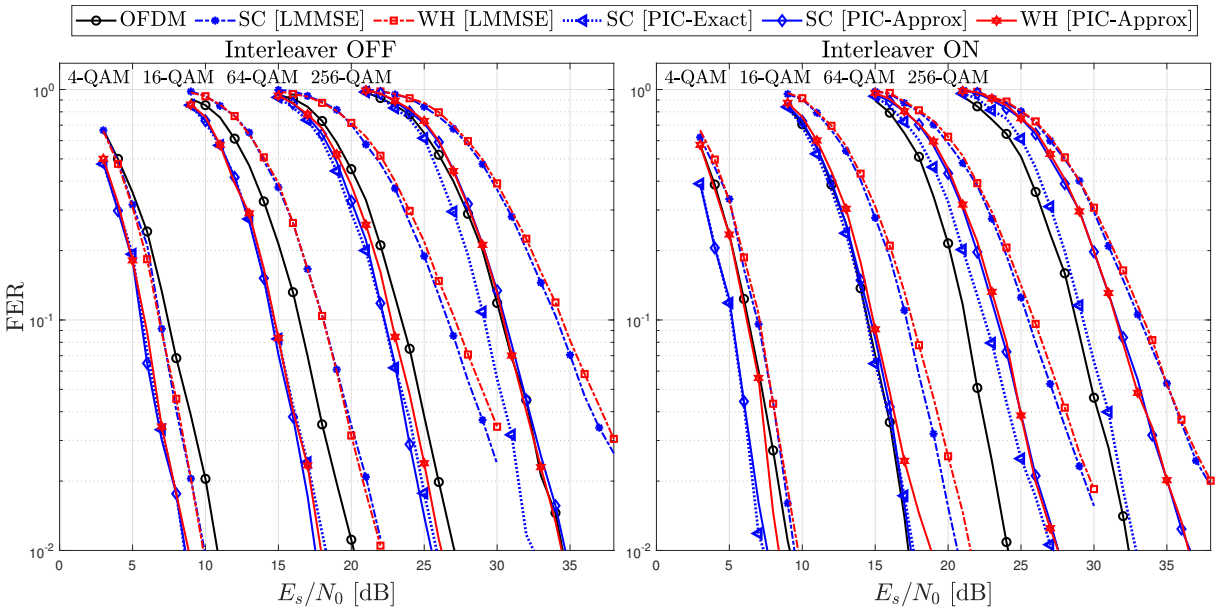

Figure 3.16: FER comparison of WHT and SC in high selective channel ($\gamma = 3$).

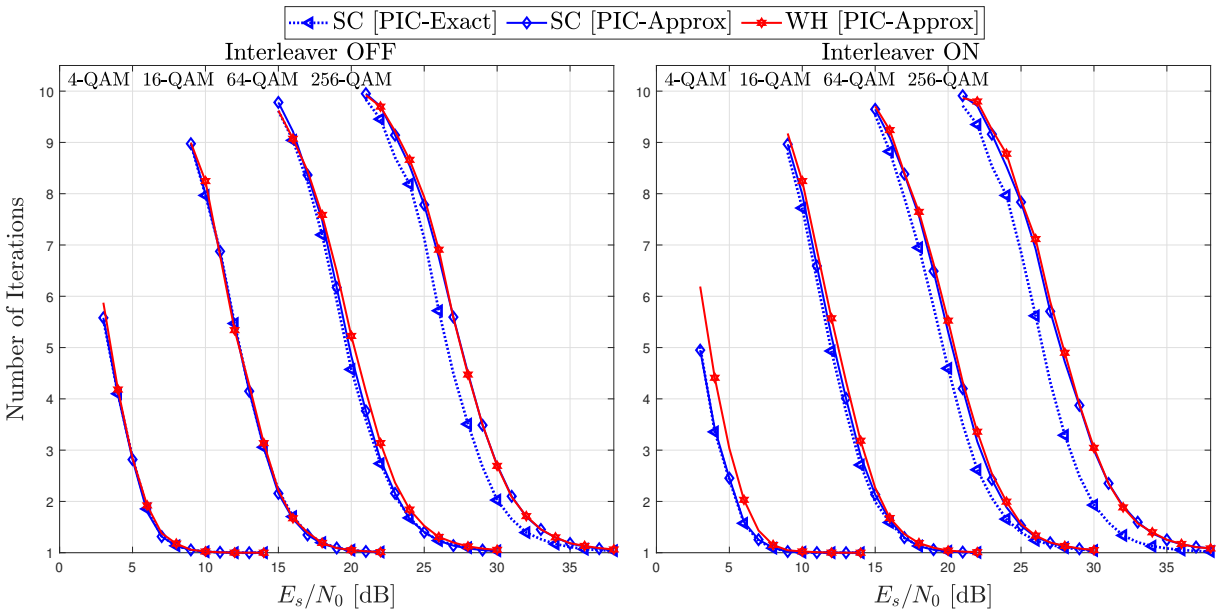

Figure 3.17: Number of iterations of WHT and SC in high selective channel ($\gamma = 3$).

Low-complexity approaches. In this configuration, the performance of SC and WHT is compared. Both achieve full frequency diversity due to their equal-gain properties.

However, WHT has the advantage of low-complexity when LMMSE-PIC approximation is considered. Figure 3.16 illustrates the FER performance of SC and WHT employing the LMMSE-PIC approximation (3.63) and compares it to the exact LMMSE-PIC applied on SC. Both modulation perform similar with LMMSE equalization. The performance of approximated LMMSE-PIC depends on the interleaving and the QAM order. For instance, SC outperforms WHT at 4-QAM with the used interleaver, but both achieve the same FER when the interleaver is off. Moreover, the exact SC-LMMSE-PIC outperforms the approximated version only at high QAM orders, which can bee observed for 256-QAM. As seen from Figure 3.17, the approximate LMMSE-PIC requires more iterations than the exact one in average to converge. This can be explained as a compensation for the approximation. Note that, the simulation considers maximum of 10 iterations. Furthermore, at high SNRs, the convergence is guaranteed with less than 2 iterations in average, i.e. one iteration for LMMSE and another one for the interference cancellation.

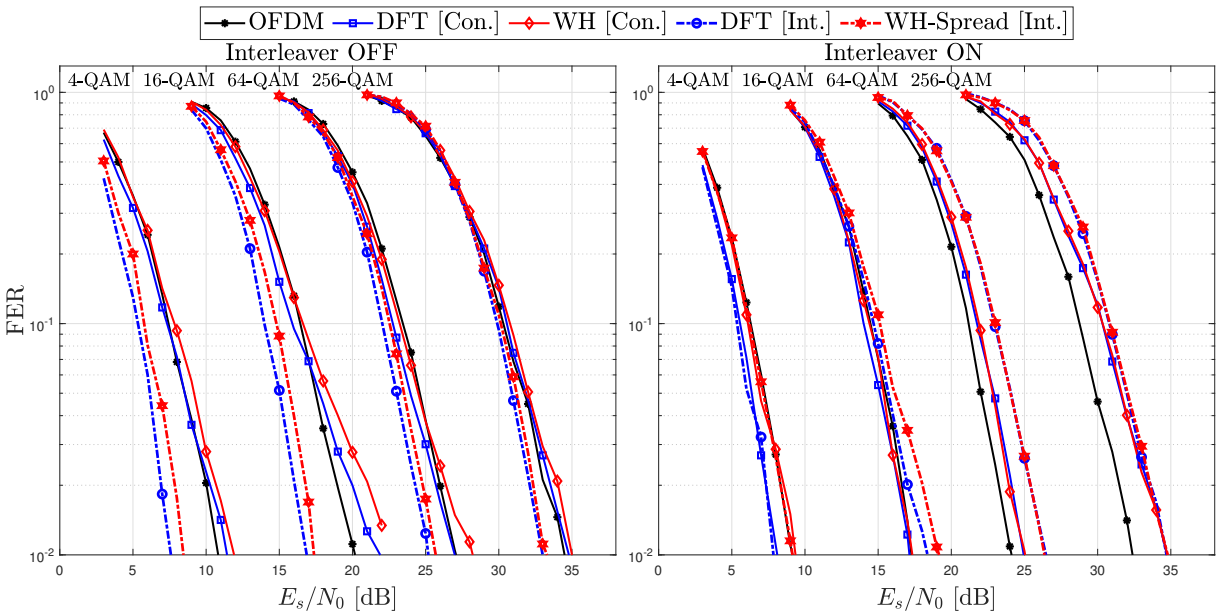

Figure 3.18: FER with different spreading matrices. Here, $M = 8$, $N = 64$ and $\gamma = 3$.

Partial spreading. Considering the high selectivity case ($\gamma = 3$), a spreading value $M = 8$ is sufficient to achieve high diversity gain in the interleaved case, as shown in Figure 3.13. In this example, WHT and DFT spreading are applied in contiguous and interleaved spreading. The FER performance is shown in Figure 3.18. It can be seen that the interleaved spreading outperforms the contiguous one without bit interleaving. However, DFT achieves slightly lower FER than WHT spreading. This can be explained by the fact that DFT is a complex-valued transform. It spreads both I and Q components of the data jointly over the I and Q components of the modulated block, whereas WHT is real-valued and thus the I and Q components of the data are spread individually. With the employed bit interleaving, different behaviors are observed depending on the QAM

order. Using 4-QAM, DFT spreading achieves similar performance in both contiguous and interleaved spreading and outperforms WHT spreading. At the higher 64 and 256-QAM orders, both WHT and DFT perform similarly, with lower FER achieved by the contiguous spreading. The contiguous spreading is very closer to OFDM as it is performed within a coherence bandwidth interval, where the channel gain can be approximated by a constant gain. Therefore, the ISI within the subblock is smaller than the case of interleaved spreading, in which the channel gains are uncorrelated.

MIMO system

This section studies MIMO systems with N_T transmit and N_R receive antennas. Assuming a perfect channel at the receiver and IBI-free by employing CP, the received signal at the n_r-th receive antenna according to (3.21) becomes,

$$\boldsymbol{y}_{n_r} = \sum_{n_t=1}^{N_T} \boldsymbol{H}^{(e)}_{n_r,n_t} \boldsymbol{x}_{n_t} + \boldsymbol{v}_{n_r} \tag{3.76}$$

The channel matrix $\boldsymbol{H}^{(e)}_{n_r,n_t,i} \in \mathbb{C}^{N\times N}$ is circular and generated from the corresponding multipath $h_{n_r,n_t}[l]$ channel between the pair (n_r, n_t). In this simulation, the channels are assumed spatially uncorrelated with the same PDP. Applying DFT and defining the diagonal channel matrix $\boldsymbol{\Lambda}^{(\tilde{h}_{n_r,n_t})} = \text{diag}\left\{\tilde{\boldsymbol{h}}_{n_r,n_t}\right\}$, the equivalent signal model after N-DFT becomes $\tilde{\boldsymbol{y}}_{n_r} = \sum_{n_t=1}^{N_T} \boldsymbol{\Lambda}^{(\tilde{h}_{n_r,n_t})} \tilde{\boldsymbol{x}}_{n_t} + \boldsymbol{v}_{n_r}$, where $\tilde{\boldsymbol{x}}_{n_t} = \sum_{m_t=1}^{N_T} \tilde{\boldsymbol{A}}_{n_t,m_t} \boldsymbol{d}_{m_t}$ is the N-DFT of the samples transmitted on the n_t-th antenna. By stacking the received samples in each antenna, a large dimensional linear model $\tilde{\boldsymbol{y}} = \boldsymbol{\Lambda}\tilde{\boldsymbol{A}}\boldsymbol{d} + \tilde{\boldsymbol{v}}$ is obtained, where $\tilde{\boldsymbol{y}} \in \mathbb{C}^{N_R N\times 1}$, $\tilde{\boldsymbol{A}} \in \mathbb{C}^{N_T N\times N_T N}$, $\boldsymbol{\Lambda} \in \mathbb{C}^{N_R N\times N_T N}$, and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}^{(\tilde{h}_{1,1})} & \dots & \boldsymbol{\Lambda}^{(\tilde{h}_{1,N_T})} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Lambda}^{(\tilde{h}_{N_R,1})} & \dots & \boldsymbol{\Lambda}^{(\tilde{h}_{N_R,N_T})} \end{bmatrix}, \ \tilde{\boldsymbol{A}} = \begin{bmatrix} \tilde{\boldsymbol{A}}_{1,1} & \cdots & \tilde{\boldsymbol{A}}_{1,N_T} \\ \vdots & \vdots & \vdots \\ \tilde{\boldsymbol{A}}_{N_T,1} & \cdots & \tilde{\boldsymbol{A}}_{N_T,N_T} \end{bmatrix}. \tag{3.77}$$

An alternative way to represent the signal model is by using N flat fading MIMO channel $\{\boldsymbol{H}_n \in \mathbb{C}^{N_R\times N_T},\ n = 0\cdots N-1\}$, such that

$$\bar{\boldsymbol{y}}_n = \boldsymbol{H}_n \bar{\boldsymbol{x}}_n + \bar{\boldsymbol{v}}_n,\ \boldsymbol{H}_n[n_r, n_t] = \tilde{\boldsymbol{h}}_{n_r,n_t}[n]\,. \tag{3.78}$$

where $\bar{\boldsymbol{y}}_n[n_r] = \tilde{\boldsymbol{y}}_{n_r}[n]$, $\bar{\boldsymbol{x}}_n[n_t] = \tilde{\boldsymbol{x}}_{n_t}[n]$.

LMMSE equalization. For orthogonal matrix $\tilde{\boldsymbol{A}}$, and when $N_T = N_R$ the LMMSE equalizer can be computed similar to (3.64) as $\boldsymbol{W}^{\text{H}}_{\text{LMMSE}} = \tilde{\boldsymbol{A}}^{\text{H}} \boldsymbol{\Lambda}^{(\text{eq})H}$, where

$$\boldsymbol{\Lambda}^{(\text{eq})\text{H}} = \begin{bmatrix} \boldsymbol{\Lambda}^{(\tilde{h}_{\text{eq},1,1})} & \dots & \boldsymbol{\Lambda}^{(\tilde{h}_{\text{eq},1,N_T})} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Lambda}^{(\tilde{h}_{\text{eq},N_R,1})} & \dots & \boldsymbol{\Lambda}^{(\tilde{h}_{\text{eq},N_R,N_T})} \end{bmatrix}, \ \boldsymbol{\Lambda}^{(\tilde{h}_{\text{eq},n_r,n_t})} = \text{diag}\left\{\tilde{\boldsymbol{h}}_{\text{eq},n_r,n_t}\right\}. \tag{3.79}$$

By defining $\boldsymbol{H}_n^{(\mathrm{eq})} \in \mathbb{C}^{N_R \times N_T}$ such that $\boldsymbol{H}_n^{(\mathrm{eq})}[n_r, n_t] = \tilde{\boldsymbol{h}}_{\mathrm{eq},n_r,n_t}[n]$, then

$$\boldsymbol{H}_n^{(\mathrm{eq})} = \left[\boldsymbol{H}_n \boldsymbol{H}_n^{\mathrm{H}} + \frac{N_0}{E_s} \boldsymbol{I}_{N_T}\right]^{-1} \boldsymbol{H}_n^{\mathrm{H}}. \tag{3.80}$$

As a result, the LMMSE equalization is realized by means of LMMSE channel equalization and MF demodulation. The equalization is performed per sample on N small-scale MIMO signals of size $N_R \times N_T$. The complexity depends on N_T. When $N_R > N_T$, the LMMSE equalization can be computed after the MF channel equalization, which is known as MRC [AH09]. The obtained signal is equivalent to $N_T \times N_T$ MIMO but the entries of $\boldsymbol{H}_n$ have a distribution different from the Rayleigh one.

SINR per symbol. The achieved SINR of LMMSE is calculated by first computing the gain $\boldsymbol{\alpha} = \mathrm{diag}\left\{\tilde{\boldsymbol{A}}^{\mathrm{H}} \boldsymbol{\Gamma} \tilde{\boldsymbol{A}}\right\}$, where $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}^{(\mathrm{eq})\mathrm{H}} \boldsymbol{\Lambda}^{(\mathrm{eq})}$, which is given by

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{1,1} & \cdots & \boldsymbol{\Gamma}_{1,N_T} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Gamma}_{N_R,1} & \cdots & \boldsymbol{\Gamma}_{N_R,N_T} \end{bmatrix}, \ \boldsymbol{\Gamma}_{n_r,n_t} = \sum_{m_t=1}^{N_T} \boldsymbol{\Lambda}^{(\tilde{h}_{\mathrm{eq},n_r,m_t})} \boldsymbol{\Lambda}^{(\tilde{h}_{n_r,m_t})}. \tag{3.81}$$

As a result, $\boldsymbol{\Gamma}_{n_r,n_t}[n,n] = \left(\boldsymbol{H}_n^{(\mathrm{eq})} \boldsymbol{H}_n\right)[n_r, n_t]$. In the case of OFDM, $\tilde{\boldsymbol{A}} = \boldsymbol{I}_{N_T N}$, and thus, $\boldsymbol{\alpha}[n] = \boldsymbol{\Gamma}[n,n]$. The case of SC is achieved when $\tilde{\boldsymbol{A}} = \boldsymbol{I}_{N_T} \otimes \boldsymbol{F}_N$, i.e. each antenna employs SC, and therefore, $\boldsymbol{\alpha}[n + (n_t - 1)N] = \frac{1}{N} \sum_{n_r=1}^{N_R} \mathrm{trace}\{\boldsymbol{\Gamma}_{n_r,n_t}\} = \alpha_{n_t}$. Accordingly, the data symbols $\boldsymbol{d}_{n_t}$ achieve equal gain. For identical PDP and uncorrelated channels, the SINRs per transmit antenna have identical distribution. In the genie case, the channel gain is equivalent to MRC, and it is given by

$$\begin{aligned} \boldsymbol{\rho}_{\mathrm{OFDM},genie}[n] &= \frac{E_s}{N_0} \sum_{n_r=1}^{N_R} |\tilde{\boldsymbol{h}}_{n_r,n_t}[n]|^2, \\ \boldsymbol{\rho}_{\mathrm{SC},genie}[n] &= \frac{E_s}{N_0} \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n_r=1}^{N_R} |\tilde{\boldsymbol{h}}_{n_r,n_t}[n]|^2 = \frac{E_s}{N_0} \sum_{l=0}^{L-1} \sum_{n_r=1}^{N_R} |\boldsymbol{h}_{n_r,n_t}[l]|^2. \end{aligned} \tag{3.82}$$

Mutual information. MIMO system can be used to harvest spacial diversity by employing multiple receive antennas, or to increase the system capacity by means of spatial multiplexing. The transmit rate can be increased by increasing the number of transmit antennas and using a proper receiver. The mutual information for different schemes is depicted in Figure 3.19. As can be seen, employing more receive antennas with MRC increases the achieved rate by increasing the average SINR and improving its distribution. For instance, at 15 dB, the achievable rate increases form 4 bits/symbol to 6 and 7 bits/symbols with $N_R = 2, 4$, respectively. However, the gap between the theoretical achievable rate and the one achieved by QAM mapping is large, especially at low QAM order. This gap can be decreased in the case of multipath channel by means of spreading using SC with a receiver that is able to resolve the interference.

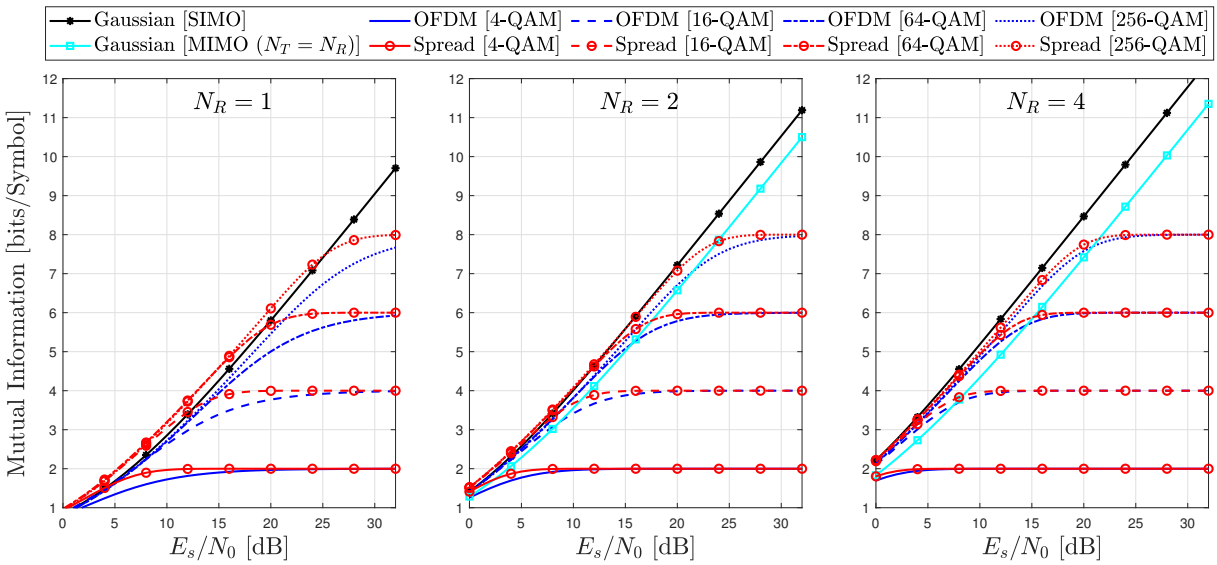

Figure 3.19: Mutual information for different MIMO settings and QAM order. The spread case is calculated for the exponential channel with $\gamma = 3$ and assuming perfect interference cancellation (genie).

The achievable rate can be scaled by the number of transmit antennas, which requires higher SNR to achieve the same number of bits/symbol in comparison with MRC. In other words, at the same SNR, the achievable rate by MIMO is not exactly N_T times the one achieved by using a single transmit antenna. Assuming perfect ISI cancellation, OFDM with QAM can achieve the theoretical rate without a need of spreading when the number of antennas is large enough. This is because the distribution of the OFDM SNR obtained by MRC is Chi-square of order $2N_R$, whereas for the spread case it is $2LN_R$. The spreading is more beneficial with lower number of antennas.

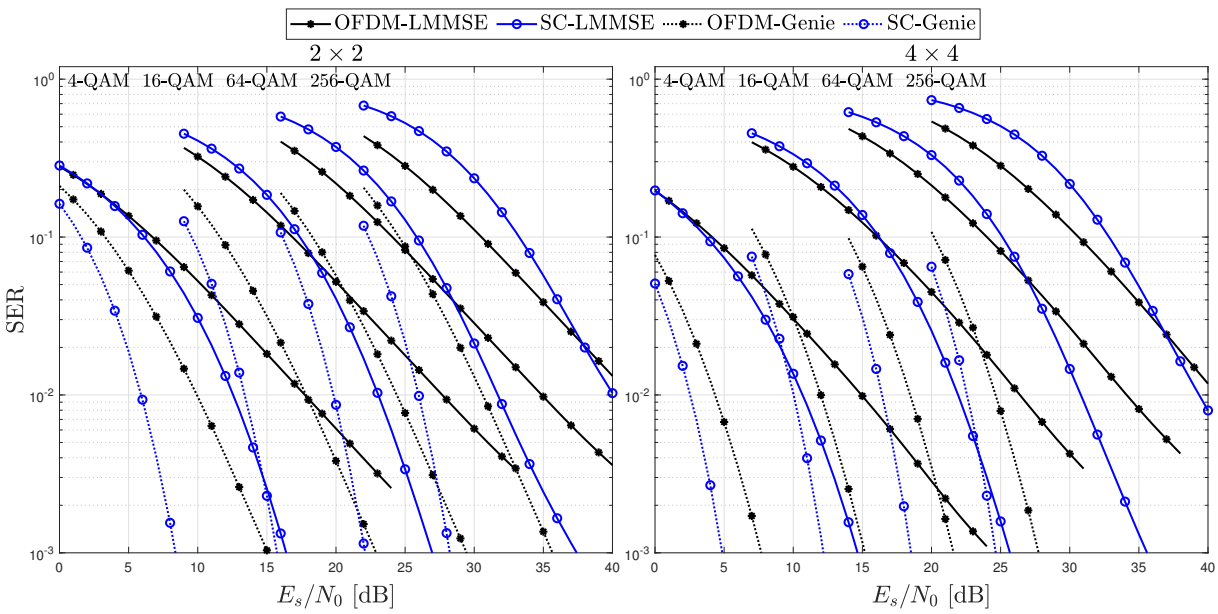

Figure 3.20: SER for MIMO with uncorrelated channels ($\gamma = 3$) .

SER and FER. Figure 3.20 illustrates the SER for LMMSE equalization compared to the genie receiver with perfect ISI cancellation. The results indicate a critical need of interference cancellation for the MIMO system to operate efficiently. There is a foreseen gain of several dBs of interference cancellation receiver over the LMMSE, and this gain increases with the increase of N_T. Moreover, the gain of spreading decreases with the increase of the number of antennas, from 6 dB with $N_T = 2$ to 3 dB for $N_T = 4$ in the genie case. This confirms that the spreading has less impact on the performance when large number of receivers is employed. Nevertheless, in practical case and especially with high order QAM, the interference cancellation largely deviates form the genie case. The

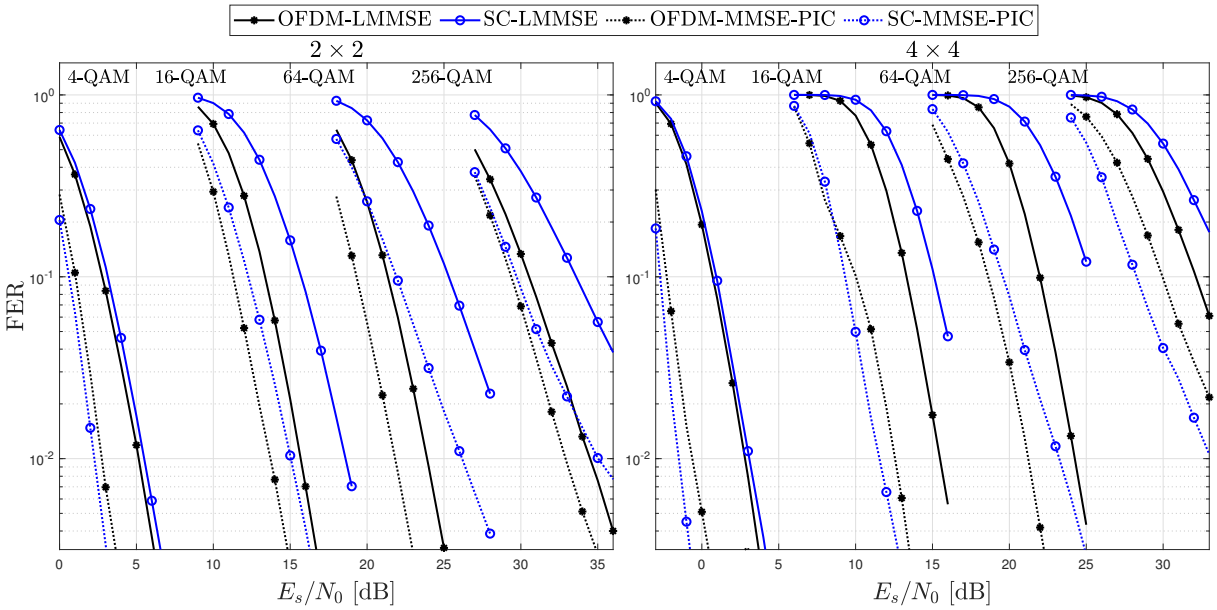

Figure 3.21: FER for MIMO with uncorrelated channels and exponential PDP($\gamma = 3$), $N = 64$ and, payload size $= K_p = L_c R_c N_T N$.

coded FER is shown in Figure 3.21, for a frame length $K_p = L_c R_c N_T N$ bits that fits in one block and with employing the bit interleaving given in Table 3.3. As shown, the gain of using LMMSE-PIC over linear equalization is significant for both OFDM and SC. The spreading with SC in the case of 2×2 achieves slightly lower FER with 4-QAM, which is practically the used mapping at low SNRs. At higher SNR, where high order QAM are used, OFDM outperforms SC by 2 dB in these settings. In the 4×4 case, a gain of 3 dB is achieved by LMMSE compared to the 2×2 scheme due to the increased spacial diversity with the doubled receive antennas. LMMSE-PIC in all modulation orders achieves much higher gain over LMMSE. The spreading with SC outperforms OFDM in the 4×4 setting with $4, 16$ and 256-QAM, whereas OFDM still achieves lower FER at 64-QAM. This variation in performance is affected by the used interleaver. Although the SER evaluation indicates that the gain of SC is larger in the 2×2 setting, the FER performance gain of SC is only observed with 4-QAM. This is because the ISI of OFDM increases with the increase of N_T, and the FER depends on the frame length and interleaving.

3.5 Summary

This chapter has introduced a general block-based waveform model. In this model, considering the general MIMO system with N_T transmit and N_R receive chains, a core data block $\boldsymbol{x}_i \in \mathbb{C}^{N_T N \times 1}$ is generated from a data vector $\boldsymbol{d}_i^{(\mathrm{on})} \in \mathbb{C}^{N_T K \times 1}$ using a modulation matrix $\boldsymbol{A} \in \mathbb{C}^{N_T N \times N_T N}$ and a mapping matrix $\boldsymbol{\Pi}^{(d)} \in \mathbb{R}^{N_T N \times N_T K}$ with

$$\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{\Pi}^{(d)}\boldsymbol{d}_i^{(\mathrm{on})}$$

The block can be split into N_T blocks $\boldsymbol{x}_{n_t,i} \in \mathbb{C}^{N \times 1}$. The transmitted block is obtained after additional processing, such as filtering, CP/CS, and windowing, which are expressed in the matrix $\boldsymbol{G}_{n_t}^{(t)} \in \mathbb{C}^{N_\mathrm{t} \times N}$ such that, $\boldsymbol{x}_{n_t,i}^{(t)} = \boldsymbol{G}_{n_t}^{(t)}\boldsymbol{x}_{n_t,i} \in \mathbb{C}^{N_\mathrm{t} \times 1}$. The waveform is generated by multiplexing the blocks with block spacing N_s, $x_{n_t}[n] = \sum_i \boldsymbol{x}_{n_t,i}^{(t)}[n - iN_\mathrm{s}]$. The value $N_\mathrm{o} = N_\mathrm{t} - N_\mathrm{s}$ defines the overlapping $N_\mathrm{o} > 0$ or spacing $N_\mathrm{o} \leq 0$ between blocks. Assuming a multipath channel, and a sufficiently large CP is used to prevent IBI as well as perfect synchronization, the received block at the n_r-th chain after applying the receiver matrix $\boldsymbol{G}_{n_r}^{(r)} \in \mathbb{C}^{N \times N_\mathrm{t}}$ is $\boldsymbol{y}_{n_r,i} = \sum_{nt=1}^{N_T} \boldsymbol{H}_{n_r,n_t,i}^{(e)}\boldsymbol{x}_{n_t,i} + \boldsymbol{v}_{n_r,i}$. And therefore, by stacking the blocks from the receive antennas, the received block is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i^{(e)}\boldsymbol{x}_i + \boldsymbol{v}_i = \boldsymbol{H}_i^{(e)}\boldsymbol{A}\boldsymbol{\Pi}^{(d)}\boldsymbol{d}_i^{(\mathrm{on})} + \boldsymbol{v}_i.$$

This model corresponds to a generic linear model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{d} + \boldsymbol{v}$, $\boldsymbol{H} \in \mathbb{C}^{N \times K}$. In linear detection with matrix $\boldsymbol{W}^\mathrm{H}$, K parallel channels are obtained in the form

$$\bar{\boldsymbol{y}} = \boldsymbol{W}^\mathrm{H}\boldsymbol{y} = \boldsymbol{\alpha} \odot \boldsymbol{d} + \bar{\boldsymbol{z}} + \bar{\boldsymbol{v}},\ \boldsymbol{\alpha} = \mathrm{diag}\left\{\boldsymbol{W}^\mathrm{H}\boldsymbol{H}\right\}.$$

It is shown that the LMMSE matrix achieves the maximum SINR per channel, where

$$\boldsymbol{W}_\mathrm{LMMSE} = \left[\boldsymbol{H}\boldsymbol{H}^\mathrm{H} + \sigma^2\boldsymbol{I}_N\right]^{-1}\boldsymbol{H} = \boldsymbol{H}\left[\boldsymbol{H}^\mathrm{H}\boldsymbol{H} + \sigma^2\boldsymbol{I}_K\right]^{-1},\ \boldsymbol{\rho}[k] = \frac{\boldsymbol{\alpha}[k]}{1 - \boldsymbol{\alpha}[k]}.$$

Considering the SVD of the channel $\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Sigma}_h\boldsymbol{U}^\mathrm{H}$, the equal SINR, and thus, equal gain can be achieved if the $|\boldsymbol{U}[k,q]|^2 = \frac{1}{K}$ with any receive matrix $\boldsymbol{W} = \boldsymbol{V}\boldsymbol{\Sigma}_h\boldsymbol{\Lambda}^{(w)}\boldsymbol{U}^\mathrm{H}$. The chapter has also introduced general iterative PIC receiver, that performs linear equalization, followed by demapping by means of a soft-in and soft-out decoder. The structure of $\boldsymbol{H}$ plays an important role in the detection complexity.

In particular scenarios, when CP is used in a frequency selective channel, $\boldsymbol{H} = \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{F}_N^H\right)\boldsymbol{\Lambda}\left(\boldsymbol{I}_{N_T} \otimes \boldsymbol{F}_N\right)\boldsymbol{A}$, the processing can be performed on the transformed model, $\tilde{\boldsymbol{y}} = \left(\boldsymbol{I}_{N_R} \otimes \boldsymbol{F}_N\right)\boldsymbol{y} = \boldsymbol{\Lambda}\tilde{\boldsymbol{A}}\boldsymbol{d} + \tilde{\boldsymbol{v}}$, where $\tilde{\boldsymbol{A}} = \left(\boldsymbol{I}_{N_T} \otimes \boldsymbol{F}_N\right)\boldsymbol{A}$. When $\tilde{\boldsymbol{A}}$ is orthogonal, the LMMSE can be computed as

$$\boldsymbol{W}_\mathrm{LMMSE}^\mathrm{H} = \tilde{\boldsymbol{A}}^\mathrm{H}\left[\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\mathrm{H} + \sigma^2\boldsymbol{I}_{N_T N}\right]^{-1}\boldsymbol{\Lambda}^\mathrm{H}$$

which shows that the equalization can be performed by means of channel equalization followed by demodulation. The channel equalization complexity is inherited from the

inverse which requires the inversion of N matrices of size $N_T \times N_T$. The demodulation depends on the structure of $\tilde{\boldsymbol{A}}$. The numerical evaluation of the mutual information and the simulation of FER show the need of adapting the modulation design to achieve certain reliability with affordable complexity. The choice of modulation depends on the channel status, hardware and spectral requirements, the frame parameters, modulation coding scheme (MCS), and bit processing techniques including channel coding and interleaving. In nutshell,

- LMMSE equalization maximizes the SINR per data symbol. Orthogonal modulation matrices allows the realization of LMMSE by means of channel equalization and MF demodulation. The theoretical ergodic channel capacity assuming Gaussian input is identical for any orthogonal modulation. Nevertheless, the achievable rate with coded modulation at certain SNR depends on the orthogonal modulation matrix structure, and the QAM order.
- In ideal hardware; perfect synchronization and channel state information (CSI) knowledge, OFDM in SISO systems allows the separation of the data symbols without interference. The gap between the actual mutual information and channel capacity increases significantly by increasing the SNR. Any linear modulation can be considered precoded-OFDM, where the frequency-domain modulation matrix can be seen as precoding or data spreading over multiple subcarriers. The spreading has the potential of harvesting frequency diversity at the cost of induced ISI. Assuming genie perfect interference cancellation, the diversity gain depends on the spreading structure, and channel selectivity.
- SC spreads the data equally over the OFDM subcarriers, and it fulfills the equal-SINR criterion per symbols. The SNR of genie-SC is equivalent to MRC of the channel delay paths. With the increase of the channel selectivity, the mutual information achieved by the genie-SC approaches the theoretical capacity. However, in reality, the ISI increases and degrades the performance of the iterative receiver. WHT spreading achieve similar performance as SC with lower complexity as it does not require complex multiplications in comparison with SC, which requires DFT at the equalizer. Sparse spreading over sufficient number of uncorrelated subcarriers allows harvesting full frequency diversity with lower complexity. For instance, using interleaved DFT or WHT spreading. Spreading over subcarriers in the same coherence bandwidth achieve performance similar to OFDM with the advantage of reduced PAPR.
- Channel coding and bit interleaving implicitly perform spreading. The FER performance is influenced by the interleaver design, the MCS, in addition to the frame length. OFDM with high order QAM and low code rate is able to approach the theoretical channel capacity in the same way SC with higher code rate and lower QAM can achieve using interference cancellation. However, SC allows low PAPR, which might be beneficial over the low-complexity equalization of OFDM.

- In MIMO systems with spacial multiplexing, the main equalization efforts are put on the LMMSE channel equalization, and this depends on the number of the transmit and receive antennas. Iterative interference cancellation is essential even for OFDM. With the increased number of the receive antennas and when the MIMO channels are uncorrelated, the spreading in frequency is less important assuming genie interference cancellation. However, in practice it has been shown by simulation, spreading still plays an important role.

The modulation design will be investigated in details in Chapter 6 and Chapter 7. The generic model developed in this chapter will be exploited in the context of generic multiple access in the next chapter.

Chapter 4

Generic Multiple Access Schemes

The radio spectrum is divided by regulation into frequency channels with certain bandwidth that can be used by different wireless communication systems. The transmission over these channels must obey rules to avoid interference to other channels [ITU16]. The goal of multiple access (MA) schemes is to enable the sharing of one or more channels, which are available in the system, among multiple users in a way to increase the utilization efficiency and fulfill quality of service (QoS) requirements. There are two main categories of MA schemes, which are adopted in wireless communications from wired networks. The first category is *packet-based*, where one user gains the access to the channel for transmitting a packet of data within a short time interval. This approach is used in some standard, such as IEEE 802.11 (WIFI) [CWKS97]. To coordinate the access, multiple access control (MAC) protocols are employed, e.g. ALOHA, carrier-sense multiple access (CSMA) and its variants, and token ring. The other category is *channel-based* MA, where the signals from different users are multiplexed in time, frequency and space. This scheme is used in cellular networks such as Long Term Evolution (LTE) [GRM+10] and fifth generation (5G) [CQC+18].

This chapter focuses on channel-based MA schemes for infrastructure networks. The goal is to provide a unified representation of different MA schemes in literature and develop a flexible MA framework, with the consideration of waveforms as a potential MA tuning parameter. This framework is able to support different access technologies, hardware constrains, traffic classes, and channel conditions, by means of optimizing the waveform parameters for each individual link. In Section 4.1, conventional and emerging MA schemes are represented in relation to the baseband model discussed in Section 2.3. A unified model for different MA based on block-based multicarrier systems is presented in Section 4.2. As a particular case of the framework, a generalization of the conventional frequency-division multiple access (FDMA) schemes used in LTE and 5G networks is developed. Most of the contributions of this chapter have been published in [NCF19a] and [NEM+19].

4.1 Basic multiple access and multiplexing schemes

To serve the communication between multiple users, the network capacity needs to be shared. This capacity depends on the physical layer (PHY) resources, such as channel bandwidth, coverage, and spatial reuse. It also depends on upper layers resources concerning processing and routing. The basic idea of MA schemes is to provide mechanisms for efficient distribution and utilization of available resources to achieve certain QoS measure, e.g. increasing the overall network throughput. The design and implementation of MA schemes is influenced by the network topology. In ad hoc networks [NM05], all distributed nodes participate in the implementation of MA procedures and the routing of data. On the contrary, infrastructure networks follow a centralized approach, where a dedicated base stations (BS) node acts as the MA coordinator and performs the routing between communicating nodes within its coverage. This chapter focuses on the infrastructure network architecture.

4.1.1 Infrastructure network system model

As illustrated in Figure 4.1, the investigated network in this thesis consists of nodes equipped with wireless access technologies denoted as user equipments (UEs). The UEs are distributed over a large area and allowed to transmit with limited power. In order to provide radio access within a certain area, an access point denoted as BSs is deployed to serve the UE within that area. The BSs communicate with each other through a radio access network (RAN) controller, which plays a role of router and gateway to the core network. The interconnection between the BSs and RAN controller can be realized using wired technologies such as optical fibers. The implementation of radio MA procedures can be centralized on the RAN controller or distributed on the BSs.

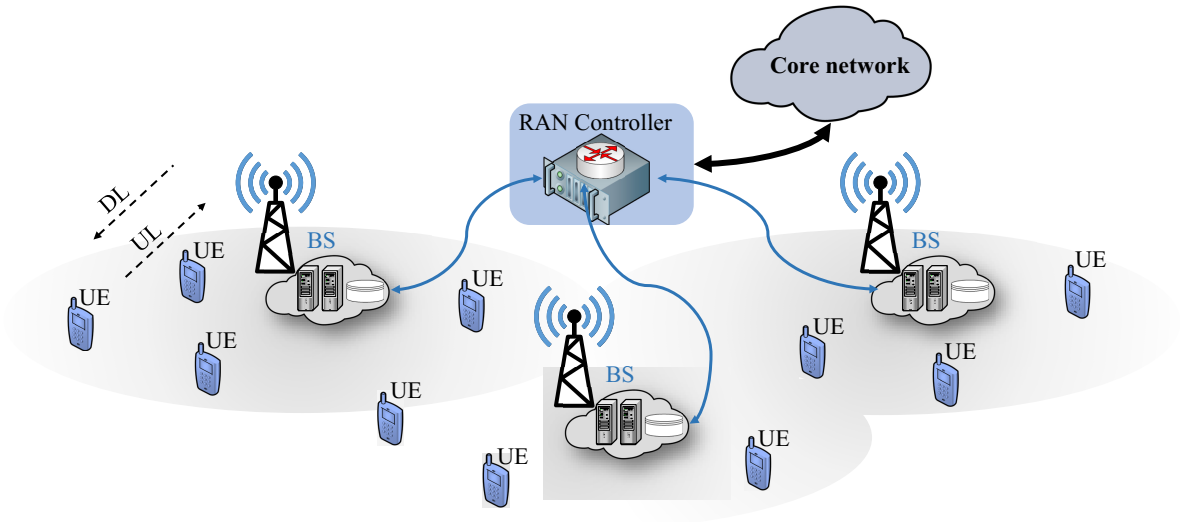

Figure 4.1: Wireless network infrastructure.

The initial access to the network can be achieved by means of the MAC protocol. Once communication links are established, the BSs multiplex the UEs signals in the downlink (DL) and demultiplex them in the uplink (UL). This thesis focuses on the air interface MA, where several UEs share the radio resources provided by one BS. This work proposes a system model of generic single-input, single-output (SISO) multiuser scheme considering U is the number of UEs served by one BS, as depicted in Figure 4.2.

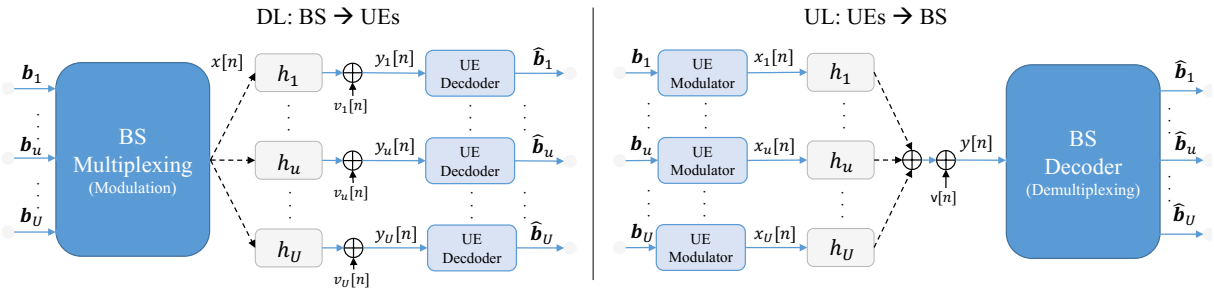

Figure 4.2: Generic SISO infrastructure multiuser system model.

Down-link. In the DL, the BS transmits data to the UEs. Let $\boldsymbol{b}_u$, the data to be transmitted to the u-th user, the BS generates a signal $x[n] = \text{Mux}\,(\boldsymbol{b}_1, \cdots \boldsymbol{b}_U)$, where $\text{Mux}\,(\cdot)$ is a generic multiplexing function. This multiplexing can also be seen as joint user modulation. Each user receives the transmitted signal through the wireless channel, where h_u denotes the channel impulse response (CIR) and $y_u[n]$ the received signal at the u-th UE. This signal is decoded to obtain the corresponding data $\boldsymbol{b}_u$.

Up-link. The UL is the transmission direction from the UEs to the BS. The u-th user generates a signal $x_u[n]$ carrying the data $\boldsymbol{b}_u$. The received signals from all UEs superimpose at the BS in the signal $y[n] = \sum_{u=1}^{U} y_u[n]$. The BS decodes this signal to obtain the data corresponding to all users. The decoding can be seen as demultiplexing of the UEs information.

4.1.2 Duplex schemes

The duplex mode determines the requirements of resources for the network in the UL and DL. In conventional TV broadcast, the communication happens in one direction from the station to the terminal. This type is called simplex communication. In contrast, duplex concerns bidirectional communication. The duplex mode defines whether a device can transmit and receive simultaneously (full-duplex) or separately (half-duplex). The duplex mode influences the complexity of the transceiver design. The simplified form is the half-duplex, where a device can only transmit or receive at a given time. Thus, time-division duplexing (TDD) is physically half-duplex but allows bidirectional

communications by assigning different time slots for the UL and DL. This mode enables the use of a single frequency channel and simplify the design of radio frequency (RF) frontend as there is no self-interference. However, it requires strict time synchronization and sufficient guard time between the UL and DL slots to mitigate the interference caused by the channel delay. On the contrary, frequency-division duplexing (FDD) allows simultaneous transmission and reception, where one frequency channel is used for the UL and another for the DL. To avoid interference and other RF problems, the separation between the frequency channels needs to be sufficiently large. Moreover, space-division duplexing (SDD) is a common choice to achieve full duplex in wired communication by using one wire for the incoming signal and another for the transmitted one. The same concept can be applied in wireless communications with massive antenna arrays and beam forming generating virtual wires. The same crosswalk problems in wired full-duplex arise in wireless SDD [ST11]. Full-duplex is related to the use of the same frequency channel while relying on interference cancellation to separate the UL and DL streams. This technique can potentially increase the network capacity conditioned by a successful interference cancellation. Full duplex is a problem mainly handled with RF processing. Nevertheless, digital interference cancellation can be considered alongside [CJS+10].

4.1.3 Common multiplexing and multiple access schemes

Multiplexing is performed in the DL as the BS combines the signals of multiple users in one signal transmitted over the shared medium. Multiple access is related to the UL, where multiple UEs are allowed to transmit on the same medium. The received signal at the BS in the UL can be seen as one multiplexed signal, whereas the received signal at each UE in the DL contains information related to the targeted UE in addition to interference from the signals targeting other users. The multiplexing is relatively an easy task, but the complexity arises at the demultiplexing in order to detect the information from a combined signal. In general, the BS in the UL can decode the signals from all UEs jointly, and the UEs in the DL may needs to decode the combined signal and extract its data. However, this increases the computational complexity of the receivers, especially at the UEs, which contrasts the requirements of low cost and low energy consumption. The main goal of multiplexing scheme is to provide a mechanism for separating the UEs signals properly considering low-complexity detection, while maintaining efficient uses of the resources.

The multiplexing of the UEs signals at the BS in the UL is linear, and it is a particular case of DL multiplexing, such that $x[n] = \sum_{u=1}^{U} x_u[n]$. Accordingly, the received signal through a linear time variant (LTV) channel with delay L_u samples can be expressed as

$$y_u[n] = \sum_{l=0}^{L_u-1} h_u[l,n]x_u[n-l] + z_u^{(\mathrm{DL})}[n] + v[n], \tag{4.1}$$

where $z_u^{(\mathrm{DL})}[n] = \sum_{v=1,v\neq u}^{U} \sum_{l=0}^{L_u-1} h_u[l,n]x_v[n-l]$ is the inter-user interference (IUI). In the UL, the received signal is given by

$$y[n] = \sum_{l=0}^{L_u-1} h_u[l,n]x_u[n-l] + z_u^{(\mathrm{UL})}[n] + v_u[n]. \tag{4.2}$$

Here, $z_u^{(\mathrm{UL})}[n] = \sum_{v=1,v\neq u}^{U} \sum_{l=0}^{L_v-1} h_v[l,n]x_v[n-l]$, which considers individual UE channels.

Orthogonal multiple access

Orthogonal multiple access refers to the case when IUI can be removed, and thus, the UEs signals can be demultiplex without loosing information. Figure 4.3 illustrates the basic orthogonal schemes commonly used in SISO system, namely, time-division multiple access (TDMA), FDMA and a combination of both.

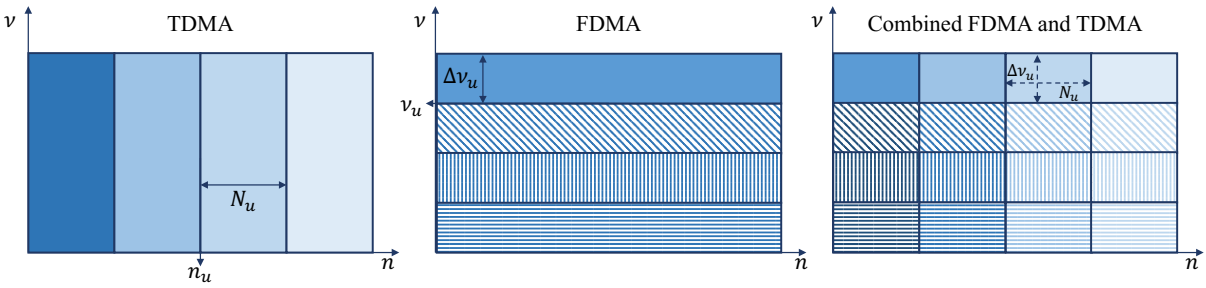

Figure 4.3: Basic orthogonal SISO MA schemes.

TDMA. Each UE is assigned a time slot in the UL while the others are silents. In the DL the UEs needs to listen and decode the signal within the assigned DL time slot. This scheme is suitable when the channel is frequency flat, such that $y_u[n] = h_u[n]x_u[n] + z_u[n] + v[n]$. Let N_u be the time slot duration in samples and n_u the starting of the slot, then $z_u[n] = 0, n = n_u, \cdots n_u + N_u - 1$. Therefore, the u-th signal can be demultiplex using a periodic rectangular window. This scheme is adopted from integrated services digital network (ISDN) [Sta95] to Groupe Spécial Mobile (GSM) [MPFBH92]. In multipath channel, interference arises because of the channel delay, where tail samples from previous slots interfere with head samples of the next slot. To overcome that, a sufficient guard time interval is required.

FDMA. In a time flat channel, $y_u[n] = \sum_{l=0}^{L_u-1} h_u[l]x_u[n-l] + z_u[n] + v[n]$, and therefore, the frequency-domain (FD) signal is $\tilde{y}_u(\nu) = \tilde{h}_u(\nu)\tilde{x}_u(\nu) + \tilde{z}_u(\nu) + \tilde{v}(\nu)$. If each user is assigned a bandwidth $\Delta\nu_u$ starting from ν_u, then $\tilde{z}_u(\nu) = 0, \nu \in [\nu_u, \nu_u + \Delta\nu_u]$, filtering can be used to separate the signals. Moreover, down-sampling is used to reduce the number of samples used for detection, as in filter bank multicarrier (FBMC) [Far11]. However, this approach requires the signals from adjacent bands to have low out-of-band (OOB) and

adequate guard band to ensure free IUI. This scheme is the basic idea behind the design of multicarrier systems [SB08].

A combination of TDMA and FDMA is practically used, which is the proper choice to deal with time-variant channels. In this approach, a user is assigned to a subband of bandwidth $\Delta\nu_u$ and a time slot of length N_u samples. When the time slot is smaller than the coherence time of the channel, and the allocated band is smaller than the coherence bandwidth, the equivalent channel becomes flat in time and frequency.

On top of that, space division multiple access (SDMA) can separate the users based on spatial properties by means of directive antennas, e.g. sectorized antennas and antenna arrays. This allows reusing the same time and frequency resources. Moreover, multi user (MU)-multiple-input, multiple-output (MIMO) provides another means of SDMA in the DL based on the CIR knowledge at the BS [SPSH04].

MU-MIMO. The basic idea is to precode the signal across the transmitted antennas, such that each user receives an IUI free signal. For instance, assuming a flat fading channel, and a BS with N_T antennas and $U \leq N_T$ users with a single antenna each. Let $\boldsymbol{x} \in \mathbb{C}^{U\times 1}$ be the sample vector corresponding to the users, and $\boldsymbol{C}^{N_T\times U}$ be a precoding matrix. The received sample vector $\boldsymbol{y} \in \mathbb{C}^{U\times 1}$ can be expressed as $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{C}\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{H} \in \mathbb{C}^{U\times N_T}$ is the MIMO channel and $\boldsymbol{v} \in \mathbb{C}^{U\times 1}$ is the additive noise. When $\boldsymbol{H}$ is known by the BS, using $\boldsymbol{C}$ such that $\boldsymbol{H}\boldsymbol{C} = \boldsymbol{I}_U$ results in $\boldsymbol{y}[u] = \boldsymbol{x}[u] + \boldsymbol{v}[u]$. The precoding matrix is given by $\boldsymbol{C} = \left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}$, which is known as zero-forcing (ZF) precoding. Accordingly, the N_T antennas at the BS can be used to serve up to $U = N_T$ users. In the UL when $N_R \geq U$, the received signal at the BS is $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}$. Here, $\boldsymbol{H} \in \mathbb{C}^{N_R\times U}$ is the UL channel. In this case, the BS need to decode the users jointly using MIMO detection techniques.

Radio resources and scheduling

Two or three basic MA schemes can be combined together. For instance a user can transmit within an allocated time slot, with certain bandwidth in a given direction or location. This leads to the definition of radio resources, which include time, frequency, and space. These resources can be shared between users. Because the varieties of requirements and considering the dynamicity of channel, number of users, it is required to consider resource management. This concerns the optimal distribution of resources in order to meet the requirements while reducing other costs, in particular, energy consumption. For instance, if the channel of a certain user goes into deep fading at certain frequency band, the user should be reallocated to another band. Thus, the resource allocation needs to consider the channel state information (CSI) when available, the power budget, and the QoS. Resource management can be formulated as optimization problems with different objectives and constraints. Changing the allocation map needs to be done periodically based on a scheduling interval. This map defines which user is using which combination

of time slots, frequency bands, and spatial resources [GRM+10]. In more generic view, each scheduling interval can be seen as a signal block with finite length, used to carry the information of multiple users. This concept will be explained in details in Section 4.2.

Non-orthogonal multiple access

Unlike orthogonal multiple access, non-orthogonal multiple access (NOMA) [LQE+17] enables multiple users to use the same resources, at the same time, using the same band, and the same spatial resources. The multiuser detection is performed jointly relying on interference cancellation techniques considering different levels of received signals. For instance, let $y[n] = h_1 x_1[n] + h_2 x_2[n] + v[n]$ with $|h_1| > |h_2|$, the detection starts by assuming $x_2[n]$ as noise and detecting the information in $x_1[n]$. Then, the estimated $x_1[n]$ is subtracted from $y[n]$ to decode $x_2[n]$. The main difference between orthogonal and non-orthogonal MA is in the way the signals can be separated without interference or loss of information, and it is not only about utilizing the same resources. As an example, in code-division multiple access (CDMA) [PSM82], multiple users transmit on the same frequency channel at the same time. However, when the channel is flat, it becomes orthogonal scheme as the users can be separated with orthogonal spreading codes. However, when the channel is frequency selective, the separation can be achieved after ZF channel equalization, which, however, enhances the noise.

NOMA aims at increasing the network throughput by multiple use of the same resources. For example, when a subband is used by multiple users the spectral efficiency increases. This idea is similar to spectrum compression applied in faster-than-Nyquist signaling (FTN) modulation discussed in Section 2.3.1. In fact, increasing the network throughput is a general problem, where the BS needs to solve an optimization problem to instruct the UEs, which parameters to use. Conventionally, the main parameters are coding rate, digital mapping order, and power loading. In this thesis, waveform parameters will be considered as part of MA scheme by introducing a generic MA framework suitable for combining different schemes, as will be discussed in the next section.

4.2 General multicarrier-based multiple access

This thesis proposes a general MA framework based on generic multicarrier as an extension of the generalized frequency division multiple access (GFDMA) system reported in [NEM+19]. Consider a scheduling interval of N_t samples, and U users each with $\boldsymbol{b}_u$ information bits to be transmitted in the DL. The information bits are mapped to a block signal $\boldsymbol{x}^{(t)} \in \mathbb{C}^{N_\mathrm{t} \times 1}$. When linear multiplexing is used, this multiplexed block can be expressed as $\boldsymbol{x}^{(t)} = \sum_{u=1}^{U} \boldsymbol{x}_u^{(t)}$, where $\boldsymbol{x}_u^{(t)} \in \mathbb{C}^{N_\mathrm{t} \times 1}$ is the block corresponding to the u-th user. One approach to generate the u-th user block is to use generic block-based modulator. After data processing as illustrated in 2.6, an uncorrelated data vector $\boldsymbol{d}_u \in \mathbb{C}^{N_u \times 1}$ with

uniform power is obtained, which is then used to modulate a set of finite length pulses such that

$$\boldsymbol{x}_u^{(t)}[n] = \sum_{k=0}^{N_u-1} \boldsymbol{d}_u[k]\, g_{k,u}^{(t)}[n] \Longrightarrow \boldsymbol{x}^{(t)}[n] = \sum_{u=1}^{U} \sum_{k=0}^{N_u-1} \boldsymbol{d}_u[k]\, g_{k,u}^{(t)}[n]. \tag{4.3}$$

Thus, the multiplexed block is generated using a generic multicarrier. Each user is assigned a unique subset of the pulses of length N_u. In other words, this turns the resources allocation to a set of pulses among the users. Here, the number of pulses $N_U = U \sum_{u=1}^{U} N_u$, depends on the user allocation, but there are no restrictions on the pulses.

4.2.1 Design with fixed set of pulses

The users may generate their pulses from a predefined set of pulses $g_q^{(t)}[n], q = 0 \cdots N-1$, using linear combination represents by $\boldsymbol{P}_u \in \mathbb{C}^{N \times N_u}$, such that

$$g_{k,u}^{(t)}[n] = \sum_{q=0}^{N-1} g_q^{(t)}[n] \boldsymbol{P}_u[q,k] \Longrightarrow \boldsymbol{x}_u^{(t)}[n] = \sum_{q=0}^{N-1} g_q^{(t)}[n] \sum_{k=0}^{N_u-1} \boldsymbol{P}_u[q,k] \boldsymbol{d}_u[k].$$

Therefore, the block is generated from the precoded data $\boldsymbol{d}_u^{(a)} = \boldsymbol{P}_u \boldsymbol{d}_u \in \mathbb{C}^{N \times N_u}$ as

$$\boldsymbol{x}_u^{(t)}[n] = \sum_{q=0}^{N-1} g_q^{(t)}[n] \boldsymbol{d}_u^{(a)}[q] = \sum_{q \in \mathcal{N}_u} g_q^{(t)}[n] \boldsymbol{d}_u^{(a)}[q]. \tag{4.4}$$

Here, $\mathcal{N}_u$ is the set of active indexes, where $\mathcal{N}_u = \{q, \boldsymbol{d}_u^{(a)}[q] \neq 0\}$. In this context, $\boldsymbol{P}_u$ denotes the allocation matrix, which can represent simple mapping and power loading, such that $\boldsymbol{d}_u^{(a)}[\mathcal{N}_u] = \boldsymbol{\Lambda}_u^{(p)} \boldsymbol{d}_u$, with $|\mathcal{N}_u| = N_u$, and $\boldsymbol{\Lambda}_u^{(p)} \in \mathbb{R}^{+N_u \times N_u}$. Moreover, using $|\mathcal{N}_u| > N_u$ means that more pulses are used to achieve diversity as in the case of MIMO. On the other hand, the multiple users can use the same pulses, with different precoding matrices, which can be used for NOMA or MU-MIMO.

Pulse design. The design of pulses $\{g_q^{(t)}[n]\}$ influences the MA scheme, where the pulses can be deigned to realize TDMA, FDMA, CDMA, etc. The first goal of the pulse design is to reduce the complexity of the allocation. For instance, reusing some bands can be achieved by designing overlapped pulses. The pulses can be obtained from core pulses $\{g_q[n]\}$ of length N samples, and additional processing including appending cyclic prefix (CP) to the beginning of the pulse, as discussed in Section 3.3. Here, the concept of inserting CP is extended to include a CP within the pulse shape. The internal CPs can act as guard intervals for TDMA, or they can be used to simplify channel equalization of LTV channel, especially, for relatively long scheduling interval. This can be seen as dividing the scheduling block into subblocks with CPs as in LTE [GRM+10]. A more flexible approach, where the CP size and number of subblocks can be different per user as in 5G [LSH+17]. Similarly, a guard interval can be used when designing for FDMA. More details on time-frequency analysis of the scheduling block will be presented in Chapter 5.

Based on that, the core scheduling block $\boldsymbol{x}_u \in \mathbb{C}^{N\times 1}$ can be expressed as

$$\boldsymbol{x}_u[n] = \sum_{q\in\mathcal{N}_u} \boldsymbol{d}[q]g_q[n],\ \boldsymbol{x}_u^{(t)} = \boldsymbol{G}_u^{(t)}\boldsymbol{x}_u,$$

where $\boldsymbol{G}_u^{(t)} \in \mathbb{C}^{N_\mathrm{t}\times N}$ represents the additional linear processing on top.

4.2.2 Computational model

This section proposes the processing on the core scheduling blocks. Let $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ be the core modulation matrix, where $\boldsymbol{A}[n,q] = g_q[n]$. Each user is assigned an allocation matrix $\boldsymbol{P}_u$, such that

$$\boldsymbol{x}_u = \boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u = \boldsymbol{A}\boldsymbol{d}_u^{(a)} = \boldsymbol{A}_u\boldsymbol{d}_u. \tag{4.5}$$

In the DL, the multiplexed signal is $\boldsymbol{x} = \sum_{u=1}^{U}\boldsymbol{x}_u = \boldsymbol{A}\sum_{u=1}^{U}\boldsymbol{d}_u^{(a)}$. The allocation vector is $\boldsymbol{d} = \sum_{u=1}^{U}\boldsymbol{d}_u^{(a)}$ is generated, then core modulation with $\boldsymbol{A}$ is applied, where $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d}$. The core block received by each user, as derived in (3.20), is given by

$$\boldsymbol{y}_u = \boldsymbol{H}_u\boldsymbol{x} + \boldsymbol{v}_u = \boldsymbol{H}_u\boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u + \boldsymbol{H}_u\boldsymbol{A}\sum_{v=1,v\neq u}^{U}\boldsymbol{P}_v\boldsymbol{d}_v + \boldsymbol{v}_u, \tag{4.6}$$

where $\boldsymbol{H}_u$ is the equivalent channel, $\boldsymbol{v}_u$ is additive noise. The term $\boldsymbol{z}_u^{(\mathrm{DL})} = \boldsymbol{H}_u\boldsymbol{A}\sum_{v=1,u\neq v}^{U}\boldsymbol{P}_v\boldsymbol{d}_v$ denotes the received interference. Here, the inter-block interference (IBI) is assumed zero, which can be fulfilled by means of a CP per scheduling block.

In the UL, the u-th user generates the core block $\boldsymbol{x}_u = \boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u$, and the received signal at the BS is given by

$$\boldsymbol{y} = \boldsymbol{H}_u\boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u + \sum_{v=1,v\neq u}^{U}\boldsymbol{H}_v\boldsymbol{A}\boldsymbol{P}_v\boldsymbol{d}_v + \boldsymbol{v}. \tag{4.7}$$

The UL interference $\boldsymbol{z}_u^{(\mathrm{UL})} = \sum_{v=1,v\neq u}^{U}\boldsymbol{H}_v\boldsymbol{A}\boldsymbol{P}_v\boldsymbol{d}_v$ depends on the individual channels, which include any synchronization misalignment.

Orthogonality. The orthogonality is fulfilled when users can be separated with orthogonal matrices $\{\boldsymbol{U}_u \in \mathbb{C}^{N\times N_u}\}$ where $\boldsymbol{U}_u^\mathrm{H}\boldsymbol{U}_u = \boldsymbol{I}_{N_u}$ such that

$$\bar{\boldsymbol{y}}_u = \boldsymbol{U}_u^\mathrm{H}\boldsymbol{H}_u\boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u + \bar{\boldsymbol{v}}. \tag{4.8}$$

This implies that the additive white Gaussian noise (AWGN), $\bar{\boldsymbol{v}} = \boldsymbol{U}_u^\mathrm{H}\boldsymbol{v}$, is not enhanced and the interference is eliminated. This depends on the channel, modulation matrix, and allocation. The orthogonality enables the per-user detection without any loss of information. When the orthogonality is not fulfilled, it might be possible to remove the interference in some cases using linear ZF receivers that yields noise enhancement. However, it is not necessary to cancel the interference while joint user detection methods

exploit the information contained in the interference. This is feasible in the UL, where the BS has access to the transmission parameters and CSI for all UEs. In the DL, UEs might not have access to information related to other users. Therefore, fulfilling orthogonality is the task of BS, which can be achieved by means of precoding based on the channel, as in MU-MIMO.

Generic FDMA

Using the normalized N-discrete Fourier transform (DFT), where $\tilde{\boldsymbol{A}} = \boldsymbol{F}_N \boldsymbol{A}$, then

$$\boldsymbol{x}_u = \boldsymbol{F}_N^{\mathrm{H}} \tilde{\boldsymbol{A}} \boldsymbol{P}_u \boldsymbol{d}_u = \boldsymbol{F}_N^{\mathrm{H}} \tilde{\boldsymbol{x}}_u. \tag{4.9}$$

where $\tilde{\boldsymbol{x}}_u = \tilde{\boldsymbol{A}} \boldsymbol{P}_u \boldsymbol{d}_u$ can be seen as FD modulated block. In the case $\tilde{\boldsymbol{A}} \boldsymbol{P}_u = \boldsymbol{S}_u \tilde{\boldsymbol{A}}_u$, where $\tilde{\boldsymbol{A}}_u \in \mathbb{C}^{N_u \times N_u}$, and $\boldsymbol{S}_u \in \mathbb{R}^{N \times N_u}$ is a mapping matrix related to $\mathcal{N}_u$, with $|\mathcal{N}_u| = N_u$, then

$$\boldsymbol{x}_u = \boldsymbol{F}_N^{\mathrm{H}} \boldsymbol{S}_u \tilde{\boldsymbol{A}}_u \boldsymbol{d}_u = \boldsymbol{F}_N^{\mathrm{H}} \boldsymbol{S}_u \bar{\boldsymbol{d}}_u, \ \bar{\boldsymbol{d}}_u = \tilde{\boldsymbol{A}}_u \boldsymbol{d}_u. \tag{4.10}$$

This corresponds to MA in which the modulation becomes OFDM and the mapping $\tilde{\boldsymbol{x}}_u [\mathcal{N}_u] = \bar{\boldsymbol{d}}_u$. In particular SISO case, where all channels are time-invariant frequency selective and the allocation sets do not overlap, i.e. $\cap_{u=1}^{U} \mathcal{N}_u = \{\Phi\}$, the unitary matrix $\boldsymbol{U}_u = \boldsymbol{S}_u^T \boldsymbol{F}_N$, fulfills the orthogonality and hence,

$$\bar{\boldsymbol{y}}_u = \boldsymbol{\Lambda}_u \tilde{\boldsymbol{A}}_u \boldsymbol{d}_u + \bar{\boldsymbol{v}}, \tag{4.11}$$

where $\boldsymbol{\Lambda}_u \in \mathbb{C}^{N_u \times N_u}$ is the diagonal channel matrix. The assumption on channels in the UL implies that all UEs are perfectly synchronized. In the asynchronous case, the equivalent channels become time variant as a result of the frequency offset, and thus, IUI arises. Reducing the IUI can be attained using sufficient guard interval, or by employing precoding (FD modulation) with low OOB properties. The special case, where $\tilde{\boldsymbol{A}}_u = \boldsymbol{I}_{N_u}$ is OFDMA [SLK97], whereas SC-FDMA is obtained when $\tilde{\boldsymbol{A}}_u = \boldsymbol{F}_{N_u}$ [MLG06b].

Combined FDMA and TDMA. A combination with TDMA is simply designed by using multiple FDMA blocks. With that, the precoded data can be spread over time and frequency. Let I be the number of FDMA blocks within the scheduling interval. The FD modulated block is stacked in a matrix $\tilde{\boldsymbol{X}}_u \in \mathbb{C}^{N \times I}$. As illustrated in Figure 4.4, after generating the precoded vector $\bar{\boldsymbol{d}}_u$, it is mapped to $\tilde{\boldsymbol{X}}_u$ using the data allocation set $\mathcal{N}_u^{(d)}$, which contains pairs (n, i) such that $\tilde{\boldsymbol{X}}_u[n, i] = \bar{\boldsymbol{d}}_u [k]$ and $\mathcal{N}_u^{(d)}(k) = (i, n)$. FDMA facilitates orthogonal multiplexing of pilots for the purpose of channel estimation. The mapping indexes of the pilot vector $\boldsymbol{p} \in \mathbb{C}^{P_u \times 1}$ are represented by the set $\mathcal{N}_u^{(p)}$. After the resource grid $\tilde{\boldsymbol{X}} = \sum_{u=1}^{U} \tilde{\boldsymbol{X}}_u$ is created, N-IDFT is performed per column to generates the TDMA subblocks. A CP guard interval is then added per subblock to overcome the interference between the TDMA slots.

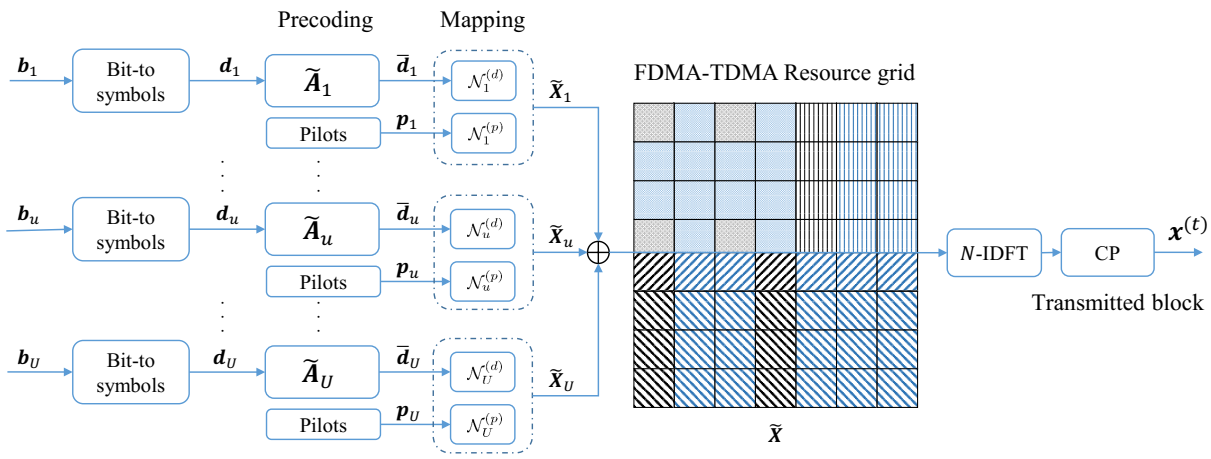

Figure 4.4: Generalized FDMA-TDMA.

4.2.3 Asynchronous multiple access

In asynchronous MA, the users apply coarse synchronization with the BS before transmitting in the UL. This causes time offset (TO) and frequency offset (FO) between the users, which introduces IUI. The TO can be compensated by using sufficiently large CP, and the IUI cased by FO can be reduced by using sufficient guard band. However, the use of large guard band reduces the spectral efficiency. The width of the guard band depends on the OOB of the waveform. To evaluate the IUI that results from the residual FO, the generic FDMA model is used in SISO configuration. Let $h_u[l], l = 0, \cdots, L_u - 1$ be the frequency selective channel between the u-th user and the BS. A CP is appended per each user such that $N_{\text{cp}} \geq \max_u L_u - 1$, and therefore,

$$\boldsymbol{H}_u = \boldsymbol{\Phi}_u \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_u)} \boldsymbol{F}_N, \ \boldsymbol{\Phi}_u = \text{diag}\left\{ e^{j2\pi \frac{n\nu_u}{N}} \right\}_{n=0}^{N-1}, \tag{4.12}$$

where ν_u is the FO and $\tilde{\boldsymbol{h}}_u \in \mathbb{C}^{N\times 1}$ is the frequency-domain channel response. As a result, the received signal at the BS can be expressed as

$$\boldsymbol{y} = \boldsymbol{\Phi}_u \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_u)} \boldsymbol{S}_u \tilde{\boldsymbol{A}}_u \boldsymbol{d}_u + \sum_{v=1, v\neq u}^{U} \boldsymbol{\Phi}_v \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_u)} \boldsymbol{S}_v \tilde{\boldsymbol{A}}_v \boldsymbol{d}_v + \boldsymbol{v}. \tag{4.13}$$

After performing N-DFT and deallocation,

$$\begin{aligned} \tilde{\boldsymbol{y}}_u &= \boldsymbol{S}_u^T \boldsymbol{F}_N \boldsymbol{\Phi}_u \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_u)} \boldsymbol{S}_u \tilde{\boldsymbol{A}}_u \boldsymbol{d}_u + \boldsymbol{S}_u^T \boldsymbol{F}_N \sum_{v=1, v\neq u}^{U} \boldsymbol{\Phi}_v \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_v)} \boldsymbol{S}_v \tilde{\boldsymbol{A}}_v \boldsymbol{d}_v + \boldsymbol{S}_u^T \boldsymbol{F}_N \boldsymbol{v} \\ &= \tilde{\boldsymbol{H}}_u \tilde{\boldsymbol{A}}_u \boldsymbol{d}_u + \sum_{v=1, v\neq u}^{U} \boldsymbol{z}_{u,v} + \boldsymbol{v}_u, \end{aligned} \tag{4.14}$$

where $\boldsymbol{z}_{u,v} = \boldsymbol{S}_u^T \boldsymbol{F}_N \boldsymbol{\Phi}_v \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_v)} \boldsymbol{S}_v \tilde{\boldsymbol{A}}_v \boldsymbol{d}_v$ is the interference from the v-th user. Assuming uncorrelated users, $\text{E}\left[\boldsymbol{z}_{u,v_1} \boldsymbol{z}_{u,v_2}^H\right] = \boldsymbol{0}_{N_u \times N_u}, \ v_1 \neq v_2$, and

$$\boldsymbol{R}_{z_{u,v}} = \text{E}\left[\boldsymbol{z}_{u,v} \boldsymbol{z}_{u,v}^H\right] = \boldsymbol{S}_u^T \boldsymbol{F}_N \boldsymbol{\Phi}_v \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_v)} \boldsymbol{S}_v \tilde{\boldsymbol{A}}_v \boldsymbol{R}_{d_v} \tilde{\boldsymbol{A}}_v^{\text{H}} \boldsymbol{S}_v^T \boldsymbol{\Lambda}^{(\tilde{h}_v)\text{H}} \boldsymbol{F}_N \boldsymbol{\Phi}_v^{\text{H}} \boldsymbol{F}_N^{\text{H}} \boldsymbol{S}_u, \tag{4.15}$$

where $\boldsymbol{R}_v = \mathrm{E}\left[\boldsymbol{d}_v \boldsymbol{d}_v^{\mathrm{H}}\right]$ is the correlation matrix of the v-th user data. The IUI depends on the precoding matrices by the term $\tilde{\boldsymbol{A}}_v \boldsymbol{R}_{d_v} \tilde{\boldsymbol{A}}_v^{\mathrm{H}}$. For example, with full allocation and independent and identically distributed (i.i.d.) uncorrelated data, i.e $\boldsymbol{R}_{d_v} = E_{s,v} \boldsymbol{I}_{N_u}$, using any type of orthogonal precoding leads to the same results. To control the IUI, orthogonal matrices can be used, while controlling $\boldsymbol{R}_{d_v}$. For example by turning one ore more data symbols.

Relation to PSD. Let $\boldsymbol{G}_v = \boldsymbol{F}_N^H \boldsymbol{\Lambda}^{(\tilde{h}_v)} \boldsymbol{S}_v \tilde{\boldsymbol{A}}_v \in \mathbb{C}^{N \times N_u}$,

$$\boldsymbol{R}_{z_{u,v}}[p,p] = \mathrm{E}\left[\left|\boldsymbol{z}_{u,v}[q]\right|^2\right] = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n=0}^{N-1} \left(\boldsymbol{G}_v \boldsymbol{R}_{d_v} \boldsymbol{G}_v^{\mathrm{H}}\right)[n,q] e^{-j2\pi[n-q]\frac{\mathcal{N}_u\{p\}-\nu_v}{N}}. \tag{4.16}$$

Accordingly, the interference power at certain frequency bin corresponds to the power spectral density (PSD) of the modulation matrix $\boldsymbol{G}_v$ sampled at that bin with a shift ν_v, as illustrated in Figure 4.5. Here, the first user is assumed perfectly synchronized, whereas the second user is offset by $\nu_2 = 0.25$. The interfering samples result from the sampling of the PSD of the second user. When both users are synchronized, the sampling of the PSD is at the zero crossing points. The level of interference depends on the PSD of the waveform whose matrix modulation is $\boldsymbol{A}_v = \boldsymbol{F}_N^{\mathrm{H}} \boldsymbol{S}_v \tilde{\boldsymbol{A}}_v$. A proper design of $\tilde{\boldsymbol{A}}_v \boldsymbol{R}_{d_v} \tilde{\boldsymbol{A}}_v^{\mathrm{H}}$ for low OOB is essential. The right side of Figure 4.5 illustrates a significant reduction of the IUI by using $\tilde{\boldsymbol{A}}_v = \boldsymbol{F}_{N_v}$ and turning off the first data symbol. Namely, $\boldsymbol{d}_2[0] = 0$, and the other data symbols are i.i.d. and uncorrelated, such that $\boldsymbol{R}_{d_2} = E_s \mathrm{diag}\left\{[0\ \boldsymbol{1}_{N_2-1}^T]\right\}$, where E_s is the energy per symbol. Note that in this case, it can be shown that when N_2

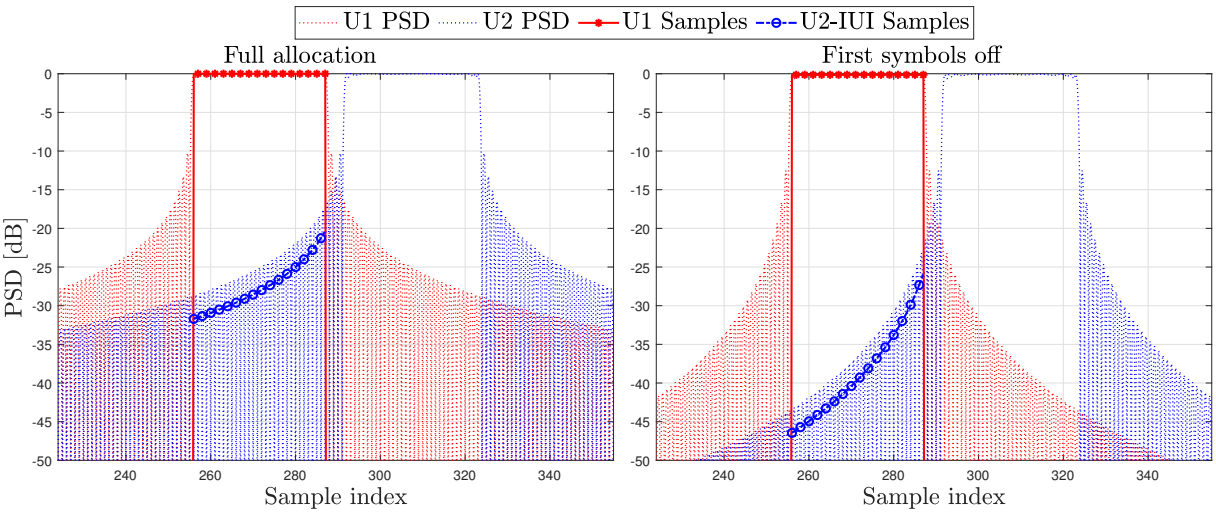

Figure 4.5: IUI using full allocation with orthogonal matrix. The right side employs DFT or Walsh-Hadamard transform (WHT) matrix with null symbol.

is radix-2, the low complexity WHT given in (3.72) can achieve the same PSD, because

$$\boldsymbol{F}_{N_2} \mathrm{diag}\left\{[0\ \boldsymbol{1}_{N_2-1}^T]\right\} \boldsymbol{F}_{N_2}^{\mathrm{H}} = \boldsymbol{\Omega}_{N_2} \mathrm{diag}\left\{[0\ \boldsymbol{1}_{N_2-1}^T]\right\} \boldsymbol{\Omega}_{N_2}^T.$$

Equalization. If the u-th user is perfectly synchronized, i.e $\nu_u = 0$, then $\tilde{\boldsymbol{H}}_u = \boldsymbol{\Lambda}_u$ is diagonal, where $\boldsymbol{\Lambda}_u[q,q] = \tilde{\boldsymbol{h}}_{\boldsymbol{u}}[\mathcal{N}_u\{q\}]$. The interference correlation matrix $\sum_{v=1,v\neq u}^{H} \boldsymbol{R}_{z_{u,v}}$ can be considered in the linear minimum mean square error (LMMSE) equalization and iterative receivers as discussed in Section 3.4. For low complexity, only the diagonal elements can be used. However, by minimizing the interference using a guard band, even the interference can be ignored. On the other hand, when $\nu_u = 0$, the full $\tilde{\boldsymbol{H}}_u$ needs to be considered in the equalization, or approximated by a diagonal matrix and additional self interference for low complexity equalization as reported in [NCF19a]. The first case is equivalent to correcting the FO followed by N-DFT for each user. Therefore, the IUI needs to be computed using the relative FO given by $\Delta\nu_{u,v} = \nu_v - \nu_u$.

Numerical example. This example considers perfect time synchronization, and evaluates the interference of two users assigned the same bandwidth $N_1 = N_2$. The user under consideration is assumed perfectly synchronized, and the offset of the second user is $\Delta\nu_{1,2} = \nu_2 - \nu_1$. Both users transmit i.i.d. uncorrelated symbols with the same energy per symbol E_s. Thus, the average signal-to-interference ratio (SIR) is evaluated by $\text{SIR}_{u,v} = \frac{1}{I_{u,v}}$, where $I_{u,v} = \frac{1}{N_v}\text{trace}\left\{\boldsymbol{R}_{z_{u,v}}\right\}$. Considering contiguous allocation, $\mathcal{N}_1 = n_1 + \{0, \cdots, N_1 - 1\}$, and $\mathcal{N}_2 = n_2 + \{0, \cdots, N_2 - 1\}$, where $n_2 = n_1 + N_1 + \Delta N_g$, and ΔN_g is the guard band in number of frequency bins. Different waveforms are evaluated by using different precoding matrices, namely $\boldsymbol{I}_{N_2}$, which corresponds to OFDM, $\boldsymbol{F}_{N_2}$ which is equivalent to DFT-spread-OFDM, in addition to WHT precoding with $\boldsymbol{\Omega}_{N_2}$. Moreover, the configurations consider turning off several symbols. This is equivalent to increasing the guard band in the case of OFDM.

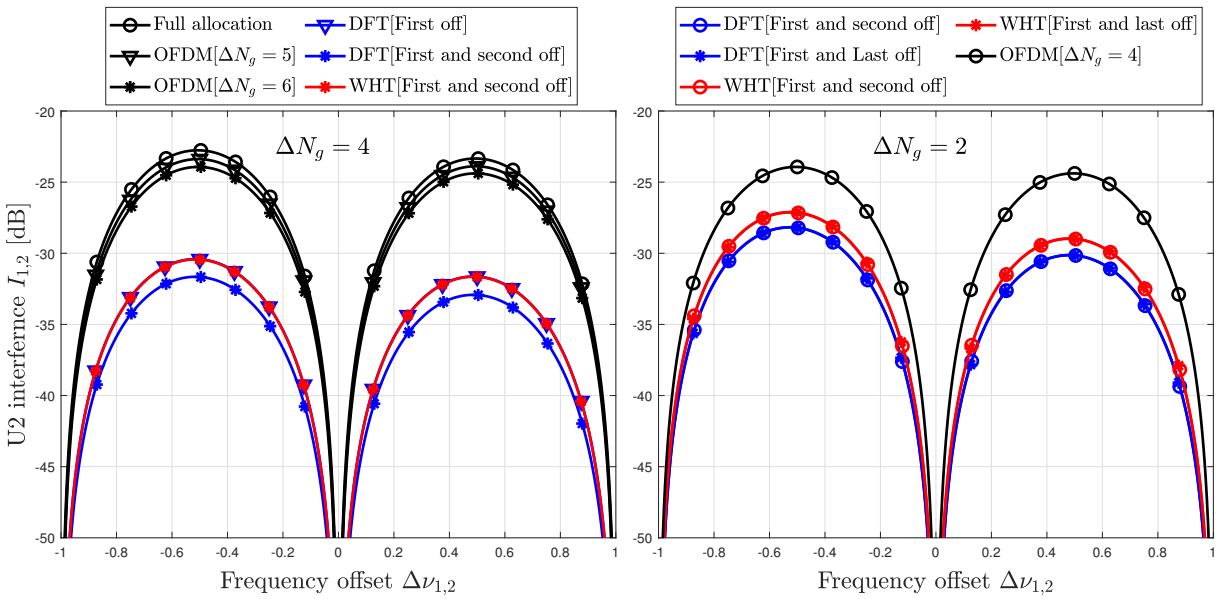

Figure 4.6: Interference of user 2 to user 1 with different waveforms settings ΔN_g is the guard band in number of frequency bins. For OFDM it is explicitly specified by $[\Delta N_g = x]$, and it is fixed for DFT and WHT.

Figure 4.6 depicts the interference $I_{2,1}$ from the second user to the first user, using $N = 1024$, $N_1 = N_2 = 32$. It can be seen that using full allocation with $\boldsymbol{I}_{N_2}$ or any other orthogonal matrix requires larger guard band to reduce the IUI. A slight decrease is achieved when increasing the guard band by one frequency sample. A reduction of 10 dB is achieved, by using $\boldsymbol{F}_{N_2}$ or $\boldsymbol{\Omega}_{N_2}$ with the first symbol is off. In this case, as shown in Figure 4.5, the OOB is significantly reduced. Moreover, turning off one more subsymbols in the case of DFT, reduces the IUI, whereas no additional improvement is achieved through WHT. By decreasing the guard band from 4 to 2 samples, and switching off two symbols, the same spectral efficiency is preserved in comparison with OFDM with 4 guard samples. However, the gain of using DFT is about 4 dB in reducing the IUI at $\Delta\nu_{2,1} = 0.5$. Two cases achieve similar performance, either by turning off the first and last or turning off the first and second symbols. Similarly, in the case of WHT turning off the second or the last symbol does not bring a significant improvement. Therefore, a trade-off between the guard band and the guard symbols is required as well as considering the complexity of the precoding. In particular, WHT precoding with one guard symbol can significantly tackle the sensitivity of orthogonal frequency division multiplexing (OFDM) with a very low complexity in comparison to DFT precoding.

4.3 Summary

In this chapter, a general model for MA has been presented in the context of block-based multicarrier. Given a set of pulses represented by a modulation matrix $\boldsymbol{A} \in \mathbb{C}^{N\times N}$, each user with $\boldsymbol{d}_u \in \mathbb{C}^{N_u}$ data symbols is assigned an allocation matrix $\boldsymbol{P}_u \in \mathbb{C}^{N\times N_u}$. This matrix changes dynamically and it provides options for precoding, power loading and mapping. The core block for each user is generated as $\boldsymbol{x}_u = \boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u$. In the DL, the BS multiplexes the blocks in a block $\boldsymbol{x} = \sum_{u=1}^{U} \boldsymbol{x}_u$. Additional processing, such as inserting CP within the block produces the transmitted block $\boldsymbol{x}^{(t)} = \boldsymbol{G}^{(t)}\boldsymbol{x}$. Assuming linear channels, each user receives a core block $\boldsymbol{y}_u = \boldsymbol{H}_u\boldsymbol{A}\boldsymbol{P}_u\boldsymbol{d}_u + \boldsymbol{z}_u^{(\mathrm{DL})} + \boldsymbol{v}_u$, where $\boldsymbol{z}_u^{(\mathrm{DL})} = \boldsymbol{H}_u\boldsymbol{A}\sum_{v=1,v\neq u}^{U} \boldsymbol{P}_v\boldsymbol{d}_v$ is the IUI. A similar signal model is obtained in the UL at the BS when per-user detection is considered. Here, the UL interference is $\boldsymbol{z}_u^{(\mathrm{UL})} = \sum_{v=1,v\neq u}^{U} \boldsymbol{H}_v\boldsymbol{A}\boldsymbol{P}_v\boldsymbol{d}_v$. The main difference in the UL interference is the effect of the different user channels, which implicitly include synchronization misalignment. This model can be used to study different MA schemes considering the channel model, modulation matrix and allocation matrices structures. A particular case denoted as generalized FDMA corresponds to $\boldsymbol{A} = \boldsymbol{F}_N^{\mathrm{H}}$, and $\boldsymbol{P}_u = \boldsymbol{S}_u\tilde{\boldsymbol{A}}_u$. Here, $\tilde{\boldsymbol{A}}_u \in \mathbb{C}^{N_u\times N_u}$ can be seen as precoding or FD modulation matrix and $\boldsymbol{S}_u \in \mathbb{R}^{N\times N_u}$ is a mapping matrix associated with the set $\mathcal{N}_u \in \{0, \cdots, N-1\}$, $|\mathcal{N}_u| = N_u$ defined by $\boldsymbol{S}_u[\mathcal{N}_u(q), q] = 1, q = 0 \cdots, N_u - 1$. Orthogonality is obtained for frequency selective channels when the allocation sets do not overlap and the synchronization is ideal. As special cases, OFDMA is obtained without any precoding, i.e. $\boldsymbol{A}_u = \boldsymbol{I}_{N_u}$, and SC-FDMA is achieved using DFT precoding, where

$\boldsymbol{A}_u = \boldsymbol{F}_{N_u}$. Using multiple FDMA blocks within the scheduling interval provides TDMA access. However, the conventional FDMA-TDMA resources grid require a CP per subblock to simplify the equalization at the cost of increased overhead. The study of asynchronous MA shows the requirement of different precoding to reduces the IUI while preserving high spectral efficiency. In particular, waveforms with low OOB are required when the frequency offset is high. While this chapter provides generic computation model, the next chapter provides design parameters based on the time-frequency selectivity of the channel and studies the interference in the absence of subblock CP.

Chapter 5

Time-Frequency Analysis of Multicarrier

In conventional time-frequency modulation, the term subcarrier is associated with frequency subcarrier. The original idea behind multicarrier comes from frequency division multiplex (FDM) [SB08], where the pulse shapes $\{g_k[n]\}$ are designed to occupy disjoint narrow bands. This concept is extended to allow overlapping among the subcarriers as in filtered multi tone (FMT) [CEO99]. This chapter focuses on the time-frequency structure of general multicarrier and the relation to the multicarrier pulses. First, Section 5.1 focuses on the time-frequency representation of multicarrier, where the modulation block is expressed by means of subsymbols and subcarriers and relation to Zak transform. Then, Section 5.2 discuses the generic time-frequency modulation from the perspective of data symbols spreading. Finally, the numerology design of the subcarrier and subsymbol in linear time variant (LTV) channel as well as the derivation of corresponding signal model are presented in Section 5.3. The contents of this chapter are based on the works reported in [NCMF18], [NZMF17], [NCF19b], [BNCF20] and [NLCF20].

5.1 General time-frequency representation

This thesis introduces a time-frequency structure of waveform as reported in [NCMF18]. Consider a multicarrier system with K generic subcarriers employing the pulses $\{g_k[n]\}$, as presented in Section 2.3.2 and Section 3.3, the multicarrier signal can be expressed as

$$x[n] = \sum_{k=0}^{K-1} \sum_{i} d_{k,i} g_k[n - iK] = \sum_{i} x_i[n - iK], \text{ where, } x_i[n] = \sum_{k=0}^{K-1} d_{k,i} g_k[n]. \tag{5.1}$$

Accordingly, the blocks $\{x_i[n]\}$ are generated for each time index i and then overlapped with sample spacing K to get $x[n]$.

Pulse shapes with time index. By focusing on a frame of M blocks, the multicarrier signal can be reformulated using the indexing $i = m + bM$ as

$$x[n] = \sum_b \sum_{m=0}^{M-1} x_{m+bM}[n - [m + bM]K] = \sum_b x_b[n - bN], \tag{5.2}$$

where $N = MK$ and $x_b[n] = \sum_{m=0}^{M-1} x_{m+bM}[n - mK]$ is the signal corresponding to a frame, and $x[n]$ is generated by multiplexing the frames with spacing N. Furthermore,

$$x_b[n] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} d_{k,m+bM} g_k[n - mK] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} d_{k,m,b} g_{k,m}[n], \tag{5.3}$$

where $g_{k,m}[n] = g_k[n - mK]$, $d_{k,m,b} = d_{k,m+bM}$. This produces a multicarrier system with $N = MK$ pulse shapes represented by means of two indexes (k, m). In this representation, the m-th index refers to the subsymbol index, M is the number of subsymbols, $x_b[n]$ is the core modulated block. Assuming the length of $g_{k,0}[n]$ is M_0K, the subsymbols overlap with $(M_0 - 1)K$ samples, which is also the overlapping between $x_b[n]$, as illustrated in Figure 5.1. In this specific case, the pulses along the time index are produced by time shift. However, the concept can be generalized to use arbitrary $g_{k,m}[n]$.

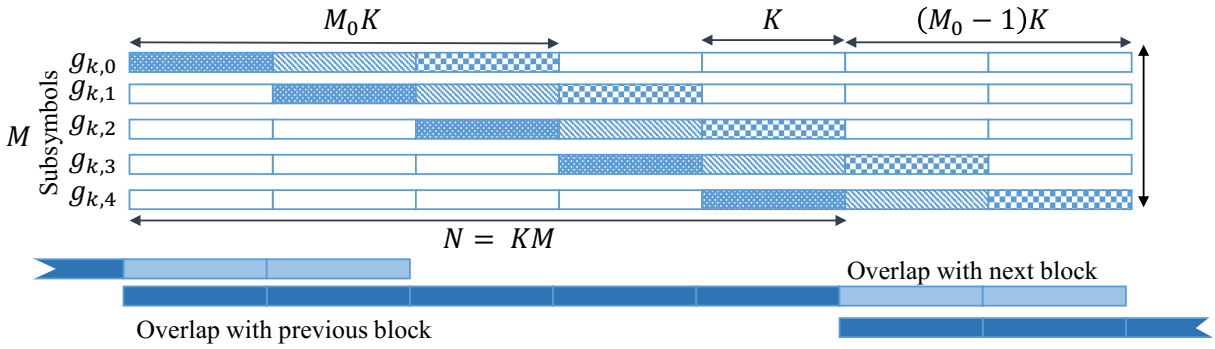

Figure 5.1: Time-frequency modulation.

5.1.1 Block representation

The general time-frequency modulation considers N pulse shapes each of length N samples, which are mapped to K *subcarriers* and M *subsymbols*, where $N = KM$. Thus, the q-th pulse shape can be mapped to subcarrier and subsymbol indexes $k = 0, \cdots, K-1$ and $m = 0, \cdots, M-1$ using a mapping function $q \overset{\pi}{\longleftrightarrow} (k, m)$. There are several possible mappings, which technically perform permutation. A simple mapping is given by $q = k + mK$. As a result, the core pulse corresponding to the (k, m)-th subcarrier-subsymbol is denoted as $g_{k,m}[n] = \boldsymbol{A}[n, k + mK]$, where $\boldsymbol{A} \in \mathbb{C}^{N \times N}$ is the core modulation matrix, as discussed in Section 3.3. The sets of active subcarriers and subsymbols are expressed as $\mathcal{K}_{on}$ and $\mathcal{M}_{on}$, respectively. Moreover, the term active symbol refers to the set $\mathcal{N}_{on}$,

$$\mathcal{N}_{on} \subset \{n = k + mK, (k, m) \in \mathcal{K}_{on} \times \mathcal{M}_{on}\}. \tag{5.4}$$

Note, that $\mathcal{N}_{on}$ may contain less elements than $|\mathcal{K}_{on} \times \mathcal{M}_{on}|$. For notation simplicity, the block index b is dropped from the modulated block, i.e. $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d}$. In relation to the multiple access (MA) discussed in Section 4.2, $\boldsymbol{x}$ is equivalent to the time-frequency scheduling block. This core block can be expressed in different forms. For example, the 2D representation simplifies the insertion of internal cyclic prefixs (CPs). In conventional vector representation, the time-domain (TD) modulated block vector is expressed as

$$\boldsymbol{x}[n] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} d_{k,m} g_{k,m}[n], \; d_{k,m} = \boldsymbol{d}[k + mK]. \tag{5.5}$$

The frequency-domain (FD) block vector is computed using the normalized N-DFT,

$$\tilde{\boldsymbol{x}} = \boldsymbol{F}_N \boldsymbol{x}, \; \tilde{\boldsymbol{x}}[s] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} d_{k,m} \tilde{g}_{k,m}[s]. \tag{5.6}$$

Thus, $\tilde{\boldsymbol{A}} = \boldsymbol{F}_N \boldsymbol{A}$ is the FD modulation matrix. Note that, $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{A}}\boldsymbol{d}$ and $\boldsymbol{x} = \boldsymbol{F}_N^H \tilde{\boldsymbol{x}}$ can be seen as the core block corresponding to the data vector $\tilde{\boldsymbol{x}}$ and the modulation matrix $\boldsymbol{F}_N^H$, which is essentially orthogonal frequency division multiplexing (OFDM). Thereby, this thesis considers any linear modulation is technically precoded OFDM, as reported in [NCF19c]. Nevertheless, this precoding is required to achieve certain waveform properties, and the design and implementation of the precoding matrix $\tilde{\boldsymbol{A}}$ depends on the waveform as will be discussed in details in Chapter 6.

2D-block representation. To provide a time-frequency structure, the entries of the modulation block $\boldsymbol{x}$ can be rearranged in a matrix $\boldsymbol{X} = \text{unvec}_{K\times M}\{\boldsymbol{x}\} \in \mathbb{C}^{K\times M}$, as reported in [NCMF18]. Using the notations $\boldsymbol{D}[k,m] = d_{k,m}$,

$$\boldsymbol{X}[q,p] = \boldsymbol{x}[q + pK], \; \boldsymbol{G}_{k,m}[q,p] = g_{k,m}[q + pK], \tag{5.7}$$

$$\boldsymbol{X}[q,p] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \boldsymbol{D}[k,m] \boldsymbol{G}_{k,m}[q,p]. \tag{5.8}$$

Similarly, the FD representation is given by

$$\tilde{\boldsymbol{X}}[p,q] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \boldsymbol{D}[k,m] \tilde{\boldsymbol{G}}_{k,m}[p,q]. \tag{5.9}$$

where $\tilde{\boldsymbol{X}} = \text{unvec}_{M\times K}\{\tilde{\boldsymbol{x}}\} \in \mathbb{C}^{M\times K}$ is the FD block, and

$$\tilde{\boldsymbol{X}}[p,q] = \tilde{\boldsymbol{x}}[p + qM], \; \tilde{\boldsymbol{G}}_{k,m}[p,q] = \tilde{g}_{k,m}[p + qM]. \tag{5.10}$$

Demodulator. As demonstrated in Section 3.3.2, the demodulator has similar structure and can be formulated as

$$\boldsymbol{D}[k,m] = \sum_{q=0}^{K-1} \sum_{p=0}^{M-1} \boldsymbol{Y}[q,p] \boldsymbol{\Gamma}_{k,m}^*[q,p], \tag{5.11}$$

where $\boldsymbol{\Gamma}^*_{k,m}[q,p] = \gamma^*_{k,m}[q+pM]$. The demodulator matrix $\boldsymbol{B}^H$ is designed with respect to the modulation matrix. Nevertheless, for some structures the modulation matrix has similar structure of the modulation matrix, which makes the implementation of the demodulator similar to the modulator. For instance, if $\boldsymbol{G}_{k,m}[q,p] = \boldsymbol{A}^{(K)}[q,k]\boldsymbol{A}^{(M)}[m,p]$, and thus, $\boldsymbol{X} = \boldsymbol{A}^{(K)}\boldsymbol{D}\boldsymbol{A}^{(M)}$, the demodulation matrices can be derived in similar form $\boldsymbol{\Gamma}_{q,p}[k,m] = \boldsymbol{B}_K[k,q]\boldsymbol{B}_M[p,m]$, and $\boldsymbol{D} = \boldsymbol{B}^*_k\boldsymbol{Y}\boldsymbol{B}^*_M$.

5.1.2 Relation to Zak transform

The TD and FD blocks are related via a unitary linear transform. The concrete form is presented here by means of the definition of N-discrete Fourier transform (DFT),

$$\begin{aligned}\tilde{\boldsymbol{X}}[p,q] &= \frac{1}{\sqrt{N}}\sum_{k=0}^{K-1}\sum_{m=0}^{M-1}\boldsymbol{x}[k+mK]e^{-j2\pi\frac{(p+qM)(k+mK)}{N}}\\ &= \frac{1}{\sqrt{N}}\sum_{k=0}^{K-1}\sum_{m=0}^{M-1}\boldsymbol{X}[k,m]e^{-j2\pi\frac{pm}{M}}e^{-j2\pi\frac{qk}{K}}e^{-j2\pi\frac{pk}{N}}\\ &= \frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}\boldsymbol{X}\boldsymbol{F}_M[k,p]\boldsymbol{\Phi}[k,p]e^{-j2\pi\frac{qk}{K}}.\end{aligned}$$

$$\text{Therefore, } \tilde{\boldsymbol{X}} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\odot[\boldsymbol{F}_M\boldsymbol{X}^{\mathrm{T}}]\right)\boldsymbol{F}_K,\ \boldsymbol{\Phi}[k,m] = e^{-j2\pi\frac{km}{N}}. \tag{5.12}$$

As a result, the time-frequency relation is given by

$$\tilde{\boldsymbol{X}}\boldsymbol{F}_K^H = \boldsymbol{\Phi}^{\mathrm{T}}\odot[\boldsymbol{F}_M\boldsymbol{X}^{\mathrm{T}}]. \tag{5.13}$$

The matrix $\boldsymbol{Z}^{(\boldsymbol{x})}_{M,K} = \boldsymbol{F}_M\boldsymbol{X}^{\mathrm{T}}$ is known as the Zak transform [BH97], which is invertible. The k-th row of $\boldsymbol{x}_k^{\mathrm{T}} = \boldsymbol{X}[k,:]$ results from the sampling of $\boldsymbol{x}$ by a factor K starting from the k-th sample. Thus, the matrix $\boldsymbol{X}^{\mathrm{T}}$ is the polyphase matrix with respect to the sampling factor K, as illustrated in Figure 5.2. This transform couples the subsymbols and provides a time-frequency relation. On the other hand, $\bar{\boldsymbol{Z}}^{(\tilde{\boldsymbol{x}})}_{K,M} = \boldsymbol{F}_K^H\tilde{\boldsymbol{X}}^{\mathrm{T}}$ is the Zak

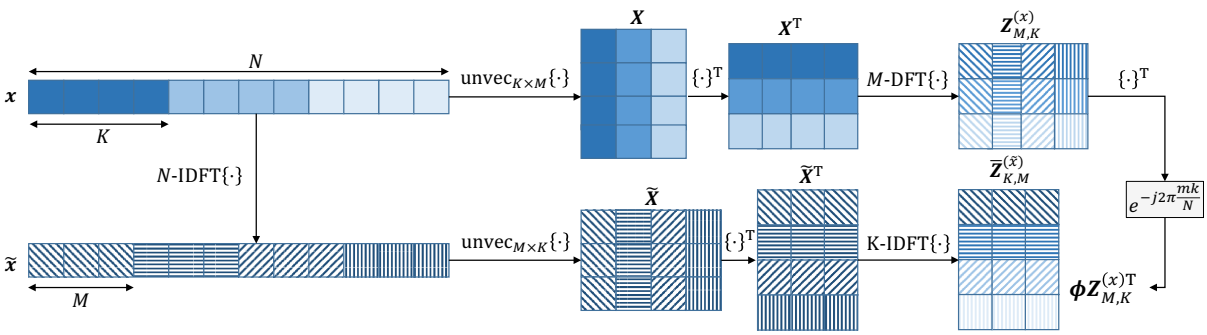

Figure 5.2: Zak transform.

transform on the frequency domain vector $\tilde{\boldsymbol{x}}$. Here, the transform couples the subcarriers. Hence, (5.13) can be expressed by means of Zak transforms as

$$\{\bar{\boldsymbol{Z}}^{(\tilde{\boldsymbol{x}})}_{K,M}\}^{\mathrm{T}} = \boldsymbol{\Phi}^{\mathrm{T}}\odot\boldsymbol{Z}^{(\boldsymbol{x})}_{M,K},\ \boldsymbol{\Phi}[k,m] = e^{-j2\pi\frac{km}{N}}. \tag{5.14}$$

Because the entries of the matrix $\boldsymbol{\Phi}$ are all of unit amplitude, using one of the Zak transforms is enough to analyze the time-frequency relation. The modulated block in the time-frequency domain using Zak transform can reformulated as

$$\begin{aligned} \boldsymbol{Z}_{M,K}^{(\boldsymbol{x})}[p,q] &= \sum_{k=0}^{K-1}\sum_{m=0}^{M-1} \boldsymbol{D}[k,m]\boldsymbol{Z}_{M,K}^{(\boldsymbol{g}_{k,m})}[p,q], \\ \bar{\boldsymbol{Z}}_{K,M}^{(\tilde{\boldsymbol{x}})}[q,p] &= \sum_{k=0}^{K-1}\sum_{m=0}^{M-1} \boldsymbol{D}[k,m]\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{\boldsymbol{g}}_{k,m})}[q,p]. \end{aligned} \tag{5.15}$$

Given that the Zak transform in either domains is computed, generating the signal in the other domain requires simple operations with element-wise multiplications and DFT transform,

$$\begin{aligned} \boldsymbol{x} &= \mathrm{vec}\left\{\{\boldsymbol{F}_M^H \boldsymbol{Z}_{M,K}^{(\boldsymbol{x})}\}^{\mathrm{T}}\right\} = \mathrm{vec}\left\{\left[\boldsymbol{\Phi}\odot \bar{\boldsymbol{Z}}_{K,M}^{(\tilde{\boldsymbol{x}})}\right]\boldsymbol{F}_M^H\right\} \\ \tilde{\boldsymbol{x}} &= \mathrm{vec}\left\{\{\boldsymbol{F}_K \bar{\boldsymbol{Z}}_{K,M}^{(\tilde{\boldsymbol{x}})}\}^{\mathrm{T}}\right\} = \mathrm{vec}\left\{\left[\boldsymbol{\Phi}\odot \boldsymbol{Z}_{M,K}^{(\boldsymbol{x})}\right]\boldsymbol{F}_K\right\}. \end{aligned}$$

OFDM time-frequency block. OFDM is commonly expressed by a vector $\tilde{\boldsymbol{x}} = \boldsymbol{d}$. Here, an alternative 2D block is used. Noting that $\tilde{\boldsymbol{X}} = \boldsymbol{D}^{\mathrm{T}}$ and $\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{\boldsymbol{x}})} = \boldsymbol{F}_K^H\boldsymbol{D}$, then $\boldsymbol{x} = \mathrm{vec}\left\{\left[\boldsymbol{\Phi}\odot(\boldsymbol{F}_K^H\boldsymbol{D})\right]\boldsymbol{F}_M^H\right\}$, and

$$\boldsymbol{X}[q,p] = \frac{1}{\sqrt{N}}\sum_{k=0}^{K-1}\sum_{m=0}^{M-1}\boldsymbol{D}[k,m]e^{j2\pi\frac{pm}{M}}e^{j2\pi\frac{qk}{K}}e^{j2\pi\frac{qm}{N}}. \tag{5.16}$$

Which shows that pulses according to (5.8) are $\boldsymbol{G}_{k,m}[q,p] = \frac{1}{\sqrt{N}}e^{j2\pi\frac{pm}{M}}e^{j2\pi\frac{qk}{K}}e^{j2\pi\frac{qm}{N}}$. This representation is useful in the implementation of OFDMA, where a partial set of contiguous subcarriers each of size M is used. The size of K and M can be chosen with respect to the allocated resources.

5.2 Time-frequency spreading

Considering the time and frequency selectivity of wireless channels, spreading the information allows gaining diversity and resistance to deep fading in order to improve the link reliability. The main concept has been introduced with numerical examples in Section 3.4.4 focusing on frequency selective channels. This can be extended to time-selective and doubly-dispersive channels using similar detection technique. This section focuses on the spreading design, whereas Section 5.3 presents the equivalent received signal models in different channels.

Consider a block of $N = MK$ samples and a sampling frequency F_s. The block duration is $T = \frac{N}{F_s}$, which is composed of M subsymbols, each has a duration $\Delta t = \frac{T}{M}$. Moreover, the frequency band is divided into K subcarriers of bandwidth $\Delta f = \frac{F_s}{K}$, this corresponds to M frequency bins of resolution $\frac{F_s}{N}$. The pulse shapes can be designed to span a width of several subcarriers and a length of several subsymbols. The m-th column in $\boldsymbol{X}$ represents a TD subsymbol with K samples.

Decoupled subsymbols. When the pulses $g_{k,m}[n]$ are localized within the corresponding subsymbol duration, i.e. $g_{k,m}[n] = 0, n \notin \{mM, \cdots (m+1)M\}$, the subsymbols are decoupled. Therefore, $\boldsymbol{G}_{k,m}[q,p] = \boldsymbol{A}_m[q,k]\delta[m-p]$, $\boldsymbol{A}_m \in \mathbb{C}^{K\times K}$, and

$$\boldsymbol{X}[q,p] = \sum_{k=0}^{K-1} \boldsymbol{D}[k,p]\boldsymbol{A}_p[q,k] \Rightarrow \boldsymbol{X}[:,p] = \boldsymbol{A}_p\boldsymbol{D}[:,p].$$

This means that the modulation matrix $\boldsymbol{A}$ is block diagonal with M matrices of size $K\times K$. Thus, each data subsymbol defined by the columns of $\boldsymbol{D}$, i.e. $\boldsymbol{d}_m = \boldsymbol{D}[:,m] \in \mathbb{C}^{K\times 1}$, is used to generate an independent modulated subsymbol, and all the subsymbols are multiplexed in one block. In other words, the symbol $d_{k,m}$ is located within the m-th subsymbol, and depending on the sub-modulation matrix $\boldsymbol{A}_m$, $d_{k,m}$ might span part or the whole band width. For instance, if $\boldsymbol{A}_m = \boldsymbol{I}_K$, i.e. the data symbols are fed directly to the digital-to-analog (DAC), as discussed in Section 2.3.1, and thus, the symbol $d_{k,m}$ is spread over the whole bandwidth. Moreover, when $\boldsymbol{A}_m = \boldsymbol{F}_K^H$, then each subsymbols is an OFDM symbol of length K. In this case, $d_{k,m}$ is mapped to the k-th subcarrier, and transmitted during the m-th subsymbol.

Decoupled subcarriers. In the frequency domain, the q-th column of $\tilde{\boldsymbol{X}}$ represents a subband of M frequency samples. The subband can be separated when the modulation pulses $\tilde{g}_{k,m}[s]$ contains non-zero samples only within the corresponding subcarrier, which can be expressed by $\tilde{\boldsymbol{G}}_{k,m}[p,q] = \tilde{\boldsymbol{A}}_k[p,m]\delta[k-q]$, $\tilde{\boldsymbol{A}}_k \in \mathbb{C}^{M\times M}$. Thereby,

$$\tilde{\boldsymbol{X}}[p,q] = \sum_{m=0}^{M-1} \boldsymbol{D}[q,m]\tilde{\boldsymbol{A}}_q[p,m] \Rightarrow \tilde{\boldsymbol{X}}[:,q] = \tilde{\boldsymbol{A}}_q\boldsymbol{D}^{\mathrm{T}}[:,q].$$

Accordingly, each data subcarrier defined by the row of $\boldsymbol{D}$, i.e. $\tilde{\boldsymbol{d}}_k = \boldsymbol{D}^{\mathrm{T}}[:,m] \in \mathbb{C}^{K\times 1}$. The FD modulation matrix can be determined from $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{A}}\mathrm{vec}\left\{\boldsymbol{D}\right\} = \tilde{\boldsymbol{A}}\boldsymbol{\Pi}_{\boldsymbol{K},\boldsymbol{M}}\mathrm{vec}\left\{\boldsymbol{D}^T\right\}$, where $\boldsymbol{\Pi}_{K,M}$ is permutation matrix. Thus, $\tilde{\boldsymbol{A}}\boldsymbol{\Pi}_{\boldsymbol{K},\boldsymbol{M}}$ is a block diagonal matrix constructed from K matrices of size $M\times M$. The symbol $d_{k,m}$ is transmitted over the k-th subcarrier, whereas the time allocation depends on the structure of $\tilde{\boldsymbol{A}}_k$. When $\tilde{\boldsymbol{A}}_k = \boldsymbol{I}_M$, which in this case corresponds to N-OFDM, $d_{k,m}$ is spread over the block duration, with $g_{k,m}[n] = e^{j2\pi\frac{(k+mK)n}{N}}$. In the case $\tilde{\boldsymbol{A}}_k = \boldsymbol{F}_M$, the symbols $d_{k,m}$ are spread with $g_{k,m}[n] = e^{j2\pi\frac{kn}{K}}e^{j\pi\frac{n(M-1)}{K}}\frac{\sin(\pi nM/N)}{sin(\pi n/N)}$, which is localized within few subsymbols.

General spreading. Further spreading can be achieved when $g_{k,m}[n]$ spans more than one subsymbol duration and $\tilde{g}_{k,m}[s]$ is spread over multiple subcarriers. One approach to perform spreading is by mean of subsymbols coupling, which can be achieved using $\boldsymbol{G}_{k,m}[q,p] = \boldsymbol{A}^{(K)}[q,k]\boldsymbol{A}^{(M)}[m,p]$, where $\boldsymbol{A}^{(K)} \in \mathbb{C}^{K\times K}$ and $\boldsymbol{A}^{(M)} \in \mathbb{C}^{M\times M} \neq I_M$. Thus,

$$\boldsymbol{X}[q,p] = \sum_{k=0}^{K-1}\sum_{m=0}^{M-1} \boldsymbol{D}[k,m]\boldsymbol{A}^{(K)}[q,k]\boldsymbol{A}^{(M)}[m,p] \Longrightarrow \boldsymbol{X} = \boldsymbol{A}^{(K)}\boldsymbol{D}\boldsymbol{A}^{(M)}. \tag{5.17}$$

This thesis considers $\boldsymbol{A}^{(K)}$ as the subsymbol modulation matrix, and $\boldsymbol{A}^{(M)}$ as the subsymbol coupling, As reported in [BNCF20]. As a result, the modulation matrix can be calculated as

$$\boldsymbol{x} = \mathrm{vec}\{\boldsymbol{X}\} = (\boldsymbol{A}^{(M)T} \otimes \boldsymbol{A}^{(K)})\boldsymbol{d} \Longrightarrow \boldsymbol{A} = (\boldsymbol{A}^{(M)T} \otimes \boldsymbol{A}^{(K)}). \tag{5.18}$$

For $\boldsymbol{A}^{(M)} = \boldsymbol{F}_M^H$ and $\boldsymbol{A}^{(K)} = \boldsymbol{I}_K$, the combination $\bar{d}_{k,m} = \frac{1}{\sqrt{M}} \sum_{p=0}^{M-1} d_{k,p} e^{j2\pi \frac{pm}{M}}$ is guaranteed to be spread over the whole bandwidth and within the duration of the m-th subsymbol. And while every symbol $d_{k,p}, p = 0 \cdots M-1$ is presented in the m-th subsymbol, it can be concluded that $d_{k,m}$ is spread over time and frequency. This spreading is reinvented as orthogonal time frequency space (OTFS) [H$^+$17] waveform. Note that $\boldsymbol{A} = \boldsymbol{F}_M^H \otimes \boldsymbol{I}_K$ is in fact it is a permutation of the block diagonal matrix $\boldsymbol{I}_K \otimes \boldsymbol{F}_M^H$, which is known as DFT-spread-OFDM [DFCS04]. This relation has been reported in [NCMF18]. Replacing the DFT spreading with Walsh-Hadamard transform (WHT) is an alternative solution reported in [BNCF20]. Note that, for achieve high spreading in time and frequency, both $\boldsymbol{A}$ and $\boldsymbol{F}_N\boldsymbol{A}$ need to be dense. This is not fulfilled with OTFS and it variants. However, the required level of spreading depends on the selectivity of the channel for harvesting full diversity. For instance, in frequency selective channel, to obtain full frequency diversity, it is sufficient to apply sparse interleaved spreading with DFT or WHT of size larger or equal to the number of delay taps, as discussed in Section 3.4.4. The next section gives more insights w.r.t. to LTV channels.

5.3 Time-frequency block in LTV channel

The design of the block dimensions is influenced by the channel characteristics. Let $h[l,n]$ be the equivalent discrete LTV which is observed at the input of the demodulator after receiver filtering and windowing. To guarantee free interference between the blocks, a sufficiently large CP is appended to the block $\boldsymbol{x}$. Therefore, the received block in the TD and FD ignoring the additive noise for simplicity are given by

$$\boldsymbol{Y}[k,m] = \sum_{l=0}^{L-1} h[l, k+mK]\boldsymbol{x}[\langle k+mK-l\rangle_N],$$
$$\tilde{\boldsymbol{Y}}[p,q] = \sum_{v=0}^{V-1} \tilde{H}[p+qM-v, v]\tilde{\boldsymbol{x}}[\langle p+qM-v\rangle_N],$$

where L is the maximum delay spread in time samples, and V is the maximum Doppler shift in number of frequency samples, and

$$\tilde{H}(s,v) = \frac{1}{N}\sum_{l=0}^{L-1}\sum_{n=0}^{N-1} h[l,n]e^{-j2\pi\frac{sl+nv}{N}}.$$

Here, the coarse delay and frequency offset (FO) of the oscillators are compensated by the synchronization. If the channel is flat in frequency, i.e. $L = 1$, the TD block becomes

$\boldsymbol{Y}[k,m] = h[k,m]\boldsymbol{X}[k,m]$, and when the channel is flat in time, i.e. $V = 1$, the FD block is $\tilde{\boldsymbol{Y}}[p,q] = \tilde{h}[p,q]\tilde{\boldsymbol{X}}[p,q]$. The variation of the channel within the block depends on the delay and Doppler spread, or equivalently the coherence time and coherence bandwidth.

5.3.1 Subcarrier and subsymbol numerology

This thesis introduces a design approach for the number of subcarriers K and subsymbols M based on the channel selectivity and receiver sensitivity. For a mobile device with relative maximum velocity $v_{\max}$ and carrier frequency f_c, the Doppler spread using Jakes's model is $f_{D,\max} = 2\frac{f_c}{c}v_{\max}$. Referring to Section 2.4, the coherence time can be approximated by $T_C \approx \frac{\alpha_T}{f_{D,\max}}$. The path gain variation depends on the relative change of the distance $\rho_l(t) = \rho_l(0)\frac{d_l(0)}{d_l(t)}$, where $d_l(t) = d_l(0) + v_l t$, $d_l(0)$ and v_l are the initial path length and velocity, respectively. Intuitively, when $v_l t << d(0)$, this variation can be neglected. This corresponds to a time interval that fulfills $t << \frac{d_l(0)}{v_l} = \frac{\tau_l c f_c}{c f_{D_l}} = \tau_l f_c \frac{1}{f_{D_l}}$. In practice, the distance between the transmitter and receiver is multiple of the wavelength λ, i.e. $\frac{d_l(0)}{\lambda} = \tau_l f_c = d_l(0)\frac{f_c}{c} >> 1$. Hence, the path gain becomes invariant for a duration smaller than $\frac{1}{f_{D_l}}$, which is satisfied when $\Delta t << T_C$.

The delay of a maximum path length $d_{\max}$ is $\tau_{\max} = \frac{d_{\max}}{c}$. Moreover, $d_{\max}$ depends on the transmit power P_t, transmit and receive antenna gains G_t and G_r, sensitivity of the receiver denoted as the minimum receive power $P_{r,\min}$, and f_c. This determined from Friis equation as [Bal05]

$$P_r(d) = \left(\frac{\lambda}{4\pi d}\right)^2 G_t G_r P_t \Longrightarrow d_{\max} = \sqrt{\frac{P_t G_t G_t}{P_{r,\min}}}\frac{c}{4\pi f_c}. \tag{5.19}$$

The coherence bandwidth is computed from the rms delay as $B_C \approx \frac{\alpha_B}{\tau_{\text{rms}}}$, where the frequency response of the channel is flat within a bandwidth smaller than B_C. In multipath channel, $\tau_{\text{rms}} \leq \tau_{\max}$, and therefore, $B_C \approx \frac{\alpha_B}{\tau_{\max}}$ is a pessimistic approximation. In practice, the delay spread depends on the difference between the paths considering the synchronization at the receiver. However, B_C will be used in the next discussion for simplicity. The significant path depends on the sensitivity of the receiver to achieve certain signal-to-noise ratio (SNR). This sensitivity is given by the minimum received signal power $P_{r,min}$ at the output of the analog processing module, and it is given by [MWQP19]

$$P_{r_{\min}}[\text{dBm}] = \text{SNR}_{\text{dB}} + N_0[\text{dBm}] + \text{NF}_{\text{dB}} + 10\log_{10}(F_s), \tag{5.20}$$

where N_0 is the thermal noise power, NF_{dB} receiver noise figure, and F_s is the sampling frequency. The typical value of the noise power is $N_0 = -174dBm$, and the noise figure NF_{dB} depends on the quality of the receiver, and it ranges from 6 to 15 in commercial systems.

Parameter design for time-invariant channel. For a sampling frequency F_s, the number of samples corresponding to the coherence time is $K_c = F_s T_C$. The number of channel delay taps is $L = L_c + L_f$, where $L_c \approx F_s \tau_{\max}$ is the channel delay spread in samples, and L_f is the additional delay taps result from the filtering. When $K_c >> L$, the subsymbol length K can be chosen as $L < K << K_c$, whereas the number of subsymbols is flexible. Thereby, the delay-time channel gain is approximated as $h[l, k+mK] \approx h[l, m]$. This design allows inserting CP of size $K_{\text{cp}} \geq L-1$ per subsymbol given that the condition $2L - 1 < K + K_{\text{cp}} << K_c$ is fulfilled. The ratio between the coherent time and the delay spread is given by

$$\frac{L_c}{K_c} = \frac{1}{\alpha_T} f_{D,\max} \tau_{\max} = \frac{2}{\alpha_T} \frac{f_c v_{\max}}{c} \frac{d_{\max}}{c} = \frac{2}{\alpha_T} \frac{1}{4\pi c} \sqrt{\frac{P_t G_t G_t}{P_{r,\min}}} v_{\max}. \tag{5.21}$$

For example, given $G_t = G_r = 3$ dB, $P_{r,\min} = -80$ dBm, $P_t = 10$ dBm, and $\alpha_T = \frac{9}{16\pi}$, then $\frac{L_c}{K_c} = 5.2 \times 10^{-5} v_{\max}$, where $v_{\max}$ is given by kmph. Therefore, this design is feasible when K_c is large enough depending on F_s and T_c. The value of K_c can be calculated as

$$K_c = F_s T_c = \alpha_T \frac{F_s}{f_{D,\max}} = \frac{\alpha_T}{2} \frac{F_s}{f_c} \frac{c}{v_{\max}}. \tag{5.22}$$

Accordingly, K_c increases with the increase of the sampling frequency (bandwidth) and decreases with the increase of the velocity. As an example, for $\frac{F_s}{f_c} = 0.001$, then $K_c \approx 9.7 \times 10^4 \frac{1}{v_{\max}}$, and for $v_{\max} = 500$ kmph, $K_c = 194$, whereas $L_c = 5$ and $\frac{K_c}{L_c} = 0.025$.

Parameter design for time-variant channel. In the case, where K_c is small because of high mobility and small bandwidth, L_C becomes also small, which implies relatively large coherence bandwidth. Let M_c be the number of frequency samples corresponding to the coherence bandwidth, which can be expressed as

$$M_c = N \frac{B_C}{F_s} = \alpha_B \frac{N}{\tau_{\max} F_s} = \alpha_B \frac{N}{L_c}. \tag{5.23}$$

Moreover, $V_c = N \frac{f_{D,\max}}{F_s} = \alpha_T \frac{N}{K_c}$ denotes the number of frequency bins representing the Doppler spread, and therefore,

$$\frac{V_c}{M_c} = \frac{\alpha_T}{\alpha_B} \frac{L_c}{K_c}. \tag{5.24}$$

If windowing is used, then $V = V_c + V_w$ is the total number of Doppler samples, with V_w is the number of additional samples result form the windowing spread. Based on that, the number of subcarriers M is chosen such that $V < M << M_c$, and thus, the frequency-Doppler channel gain can be approximated as $\tilde{H}[p + qM, v] \approx \tilde{H}[q, v]$. This design requires that M_C is large enough, which can be controlled by increasing N. The number of subcarrier is calculated from $K = \frac{N}{M}$. For example, given that $\frac{F_s}{f_c} = 0.0001$ and $v_{\max} = 200$ kmph, then $K_c = 48$ and $L_c = 1$. With $\alpha_B = \frac{1}{50}$, and $N = 8000$, $M_c = 320$ and $V_c = 30$. Using $M = 150$ requires $K = 53 > K_c$, and thereby, the channel is time-variant within the subsymbol.

Parameter design for doubly dispersive channel. This design is suitable when both conditions, $V < M << M_c$ and $L < K << K_c$, are fulfilled, the frequency-time channel gain can can be approximated as $\tilde{h}[p + qM, k + mK] \approx \tilde{h}[q, m]$. This implies $M << M_c = MK\frac{B_C}{F_s}$, and thus $1 << K\frac{B_C}{F_s} = \alpha_B \frac{K}{L_c}$. For example, given that $\frac{F_s}{f_c} = 0.0001$ and $v_{\max} = 50$ kmph, then $K_c = 194$ and $L_c = 1$. With $\alpha_B = \frac{1}{50}$, and $N = 8000$, $M_c = 320$ and $V_c = 8$. Using $M = 150$ requires $K = 53 < K_c$.

Practical example

Consider extended ITU Veh. A with the power delay profile (PDP) listed in Table 5.1 [SMF05]. The Doppler spectrum for each tap follows Jakes's model, with maximum Doppler frequency given by $f_{D,\max} = \frac{v_{\max}}{cf_c}$. Ideal transmit and receive filters are used, such that the discrete channel following (2.68) is given by

$$h[l, n] \approx \sum_{p=1}^{P} h_p(nT_s)\mathrm{sinc}(l - \tau_p F_s).$$

Table 5.1: Extended ITU Veh. A PDP

Excess tap delay τ_p [ns]	0	30	150	310	370	710	1090	1730	2510
Relative power $\mathcal{P}_p$ [dB]	0	−1.5	−1.4	−3.6	−0.6	−9.1	−7.0	−12.0	−16.9

The receiver is assumed to be synchronized at the first tap. The actual received power on the l-th tap, denoted as $P_{r,l}$, depends on the distance between the transmitter and receiver. In this example, it is assumed that $P_{r,l} \geq P_{r,\min}$.

Table 5.2: Coherence time in samples for different velocity and $f_c = 5$ GHz.

		$v_{\max}$ [Kmph]					
		6	60	120	240	360	480
F_s [MHz]	L	K_C					
10	27	128915	12891	6445	3222	2148	1611
20	41	257831	25783	12891	6445	4297	3222
40	72	515662	51566	25783	12891	8594	6445
80	145	1031324	103132	51566	25783	17188	12891

Figure 5.3 illustrates the discrete channel delay taps for different sampling frequencies where all discrete channel taps with gain less by −17 dB than the maximum one are ignored. The number of significant taps increases with the increase of the sampling frequency. In the case of $F_s = 10$ MHz, the effective channel length $L = 27$, which is

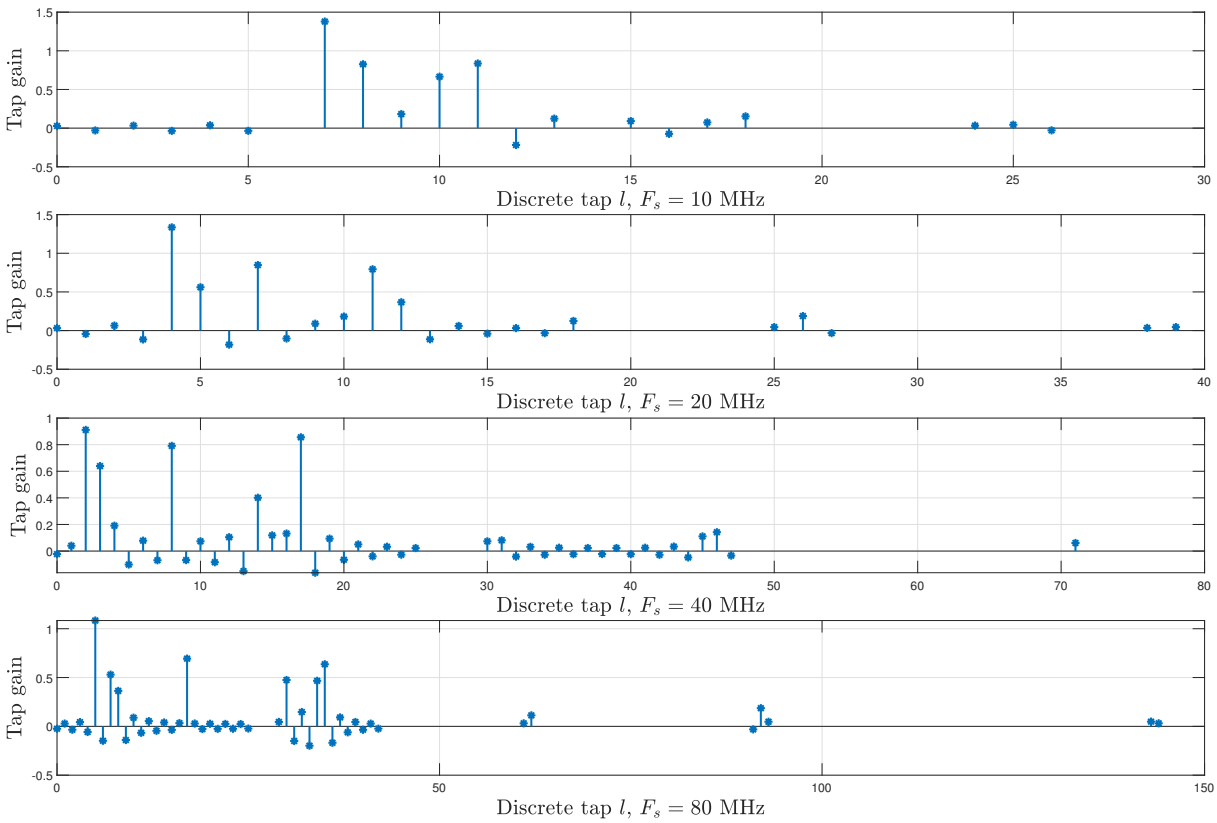

Figure 5.3: Discrete channel delay profile for extended ITU Veh. A.

larger than the propagation delay $L_c = \tau_{\max} F_s = 25$. This is because of the additional delays by the filtering. In the other cases, the number of significant taps is less than the one equivalent to the channel delay. For instance, with $F_s = 40$ MHz, $L = 40$ samples, whereas $L_c = 50$. The coherence time in number of samples is listed in Table 5.2 for different velocity s at carrier frequency $f_c = 5$ GHz. In all the practical cases, the coherence time in samples is significantly larger than the channel length. For instance, at $v_{\max} = 480$ Kmph, $\frac{L}{K_c} = 0.02$. For a design with $K = 256$, it is guaranteed that the channel is static within the subsymbol. Moreover, even by adding a CP of length 32 samples, the channel can be still considered static within the subsymbols. The approximation error is illustrated in Figure 5.4. In this approximation, the effective channel is obtained by the average $h_e[l,k] = \frac{1}{K}\sum_{q=0}^{K-1} h[l,q]$, $k = 0, \cdots, K-1$. The normalized error is given by

$$E_h = \frac{\sum_{l=0}^{L-1} \sum\limits_{k=0}^{K-1} \mathrm{E}\left[|h[l,k]|^2\right]}{\sum_{l=0}^{L-1} \sum\limits_{k=0}^{K-1} \mathrm{E}\left[|\epsilon[l,k]|^2\right]}, \text{ where } \epsilon[l,k] = h_e[l,k] - h[l,k]. \tag{5.25}$$

As expected, the higher is K the larger is the approximation error. The approximation becomes worse with the increase of the velocity, i.e. the decrease of the coherence time. For instance, with $K = 256$ and $v_{\max} = 480$ Kmph, the approximation error is less than -22 dB. However, This error is negligible when low order QAM mapping is used.

On the other hand, with slower velocity, the channel can be static over multiple subsymbols

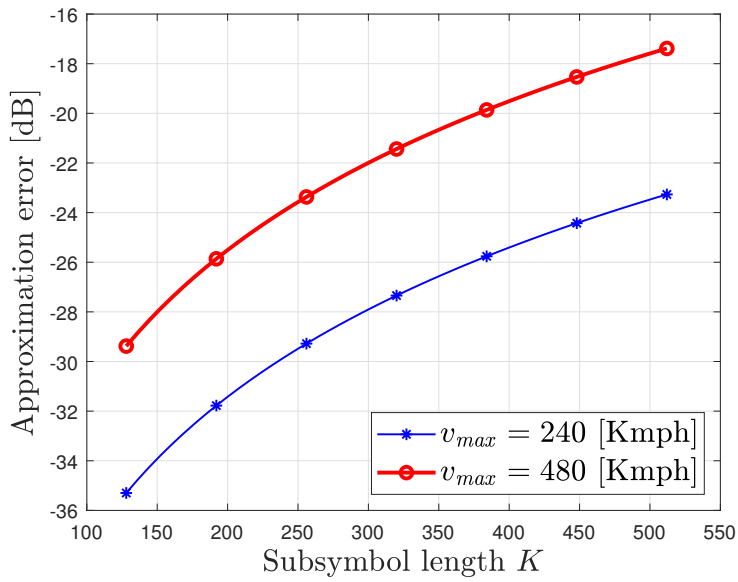

Figure 5.4: Static channel approximation error within the subsymbol vs number of subcarriers using extended ITU Veh. A at $f_c = 5$ GHz, $F_s = 10$ MHz.

and only one CP per block is required. In this case, M is chosen to be smaller than M_c. For example, at a velocity less than 120 Kmph and $F_s = 10$ MHz, with the use of $N = 4096$, the coherence bandwidth in number of samples can be computed from (5.23) as $M_c = 7$. Thus, by choosing $M = 4$ then $K = 1024$. In addition, with the decrease of the velocity to 60 Kmph, M can be doubled to to 8.

5.3.2 Processing based on the time-domain signal

Here the condition $L < K << K_c$ is considered and thus, $h[l, k + mK] \approx h[l, m]$. The received block in the time domain ignoring the additive noise, is given by

$$\boldsymbol{Y}[k, m] = \sum_{l=0}^{L-1} h[l, m]\boldsymbol{x}[\langle k + mK - l\rangle_N].$$

Noting that

$$\boldsymbol{x}[\langle k - l + mK\rangle_N] = \left\{ \begin{array}{ll} \boldsymbol{X}[\langle k - l\rangle_K, \langle m - 1\rangle_M], & k < l \\ \boldsymbol{X}[\langle k - l\rangle_K, m], & k \geq l \end{array} \right\}.$$

By computing the K-DFT of the columns, i.e. $\boldsymbol{F}_K\boldsymbol{Y}$,

$$\begin{aligned} [\boldsymbol{F}_K\boldsymbol{Y}]\,[q, m] &= \frac{1}{\sqrt{K}} \sum_{l=0}^{L-1} h[l, m] \sum_{k=l}^{K-1} e^{-j2\pi\frac{kq}{K}} \boldsymbol{X}[\langle k - l\rangle_K, m] \\ &+ \frac{1}{\sqrt{K}} \sum_{l=0}^{L-1} h[l, m] \sum_{k=0}^{l-1} e^{-j2\pi\frac{kq}{K}} \boldsymbol{X}[\langle k - l\rangle_K, \langle m - 1\rangle_M] \\ &= \frac{1}{\sqrt{K}} \sum_{l=0}^{L-1} h[l, m] \sum_{k=0}^{K-1} e^{-j2\pi\frac{kq}{K}} \boldsymbol{X}[\langle k - l\rangle_K, m] + \boldsymbol{Z}[q, m] \\ &= \tilde{h}[q, m][\boldsymbol{F}_K\boldsymbol{X}][q, m] + \boldsymbol{Z}[q, m], \end{aligned} \tag{5.26}$$

where $\tilde{h}[q,m] = \sum_{l=0}^{L-1} h[l,m]e^{-j2\pi\frac{ql}{K}}$, and $\boldsymbol{Z}[q,m]$ is inter-subsymbol interference given by

$$\boldsymbol{Z}[q,m] = \frac{1}{\sqrt{K}} \sum_{l=0}^{L-1} h[l,m] \sum_{k=0}^{l-1} e^{-j2\pi\frac{kq}{K}} \left(\boldsymbol{X}[\langle k-l\rangle_K, \langle m-1\rangle_M] - \boldsymbol{X}[\langle k-l\rangle_K, m]\right). \tag{5.27}$$

This interference can be eliminated using different ways For instance, when $\boldsymbol{X}[\langle k-l\rangle_K, \langle m-1\rangle_M] = \boldsymbol{X}[\langle k-l\rangle_K, m]$, $\forall k < l$. This condition can be satisfied if $\boldsymbol{X}[K-l,m] = 0, \forall m$ and $l = 1 \cdots L-1$. This means, that zero-padding (ZP) between subsymbols is required. The ZP can be achieved approximately by the design of the waveform as reported in [NZMF17]. Another option is to use a unique word of length $L-1$ samples, where $\boldsymbol{X}[K-l,m] = u[m]$. Moreover, the inter-symbol interference (ISI) can be avoided by using a CP per subsymbol such that $\boldsymbol{X}^{(\mathrm{cp})}[k,m] = \boldsymbol{X}[\langle k-K_{\mathrm{cp}}\rangle_K, m]$, the subsymbols duration becomes $K+K_{\mathrm{cp}}$, and the receiver needs to remove the CPs for each subsymbols. In a special case, each subsymbol can apply independent modulation with length K. Nevertheless, the block modulation allow spreading over time and frequency for harvesting time and frequency diversities.

The time flat channel is a special case, where $\tilde{h}[q,m] = \tilde{h}[q]$. Therefore, only one channel gain is required per subcarrier, $[\boldsymbol{F}_K\boldsymbol{Y}][q,m] = \tilde{h}[q][\boldsymbol{F}_K\boldsymbol{X}][q,m] + \boldsymbol{Z}[q,m]$. However, this case can also be processed with interference free using frequency domain block with N-DFT. Let $\tilde{\boldsymbol{H}} \in \mathbb{C}^{M\times K}$, where $\tilde{\boldsymbol{H}}[p,q] = \tilde{h}[p+qM]$, then $\tilde{\boldsymbol{Y}}[p,q] = \tilde{\boldsymbol{H}}[p,q]\tilde{\boldsymbol{X}}[p,q]$.

5.3.3 Processing based on the frequency-domain signal

In the case where $V < M << M_c$, the channel becomes $\tilde{H}[m+qM,v] \approx \tilde{H}[q,v]$. The frequency domain received block is

$$\tilde{\boldsymbol{Y}}[p,q] = \sum_{v=0}^{V-1} \tilde{H}[q,v]\tilde{\boldsymbol{x}}[\langle p-v+qM\rangle_N].$$

Following the similar steps in (5.26),

$$\boldsymbol{F}_M^H\tilde{\boldsymbol{Y}}[m,q] = \tilde{h}[q,m][\boldsymbol{F}_M^H\tilde{\boldsymbol{X}}][m,q] + \tilde{\boldsymbol{Z}}[m,q]. \tag{5.28}$$

where $\tilde{h}[q,m] = \sum_{v=0}^{V-1} \tilde{H}[q,v]e^{j2\pi\frac{mv}{M}}$ and $\tilde{\boldsymbol{Z}}[q,m]$ is inter-subcarrier interference given by

$$\tilde{\boldsymbol{Z}}[m,q] = \frac{1}{\sqrt{M}} \sum_{v=0}^{V-1} \tilde{H}[q,v] \sum_{p=0}^{v-1} e^{j2\pi\frac{mp}{M}} \left(\tilde{\boldsymbol{X}}[\langle p-v\rangle_M, \langle q-1\rangle_K] - \tilde{\boldsymbol{X}}[\langle p-v\rangle_M, q]\right). \tag{5.29}$$

Similar approaches of ISI avoidance can be applied to get rid of inter-carrier interference (ICI). The conventional way is to use a guard band. In addition, using FD-CP is relevant here, where the subcarrier length becomes $M+M_{\mathrm{CP}}$. The CP is removed after performing the large DFT transform.

The frequency flat channel is a special case where $\tilde{h}[q,m] = h[m]$. And therefore, only one channel gain per subsymbol is required, $\boldsymbol{F}_M^H \tilde{\boldsymbol{Y}}[m,q] = h[m][\boldsymbol{F}_M^H \tilde{\boldsymbol{X}}][m,q] + \tilde{\boldsymbol{Z}}[m,q]$. This case can be handled in the time domain $\boldsymbol{Y}[k,m] = \boldsymbol{H}[k,m]\boldsymbol{X}[k,m]$, where $\boldsymbol{H} \in \mathbb{C}^{M\times K}$, $\boldsymbol{H}[k,m] = h[k+mK]$.

5.3.4 Unified signal model

For unified presentation, this thesis introduces a generic signal model for all aforementioned cases. The received signal includinge can be expressed as

$$\boldsymbol{F}\boldsymbol{Y}_0 = \boldsymbol{H}_0 \odot [\boldsymbol{F}\boldsymbol{X}_0] + \boldsymbol{Z}_0 + \boldsymbol{V}_0 \in \mathbb{C}^{K_0\times M_0}, \tag{5.30}$$

where $\boldsymbol{X}_0 \in \mathbb{C}^{K_0\times M_0}$ is the modulated block, $\boldsymbol{F} \in \mathbb{C}^{K_0\times K_0}$ a transform depends on the selectivity. Moreover, $\boldsymbol{H}_0$ is the channel gain matrix, $\boldsymbol{Z}_0$ interference matrix, and $\boldsymbol{V}_0$ is the additive noise matrix. Table 5.3 summarizes the exact form in all related options.

Table 5.3: General received block in time variant channel.

Model	K_0	M_0	$\boldsymbol{F}$	Signal $\boldsymbol{X}_0$	Channel gain $\boldsymbol{H}_0$
Time flat	M	K	$\boldsymbol{I}_M$	FD, $\tilde{\boldsymbol{X}}$	$\boldsymbol{H}_0[p,q] = \frac{1}{N}\sum_{l=0}^{L-1} h[l]e^{-j2\pi\frac{l[p+qM]}{N}}$
Frequency flat	K	M	$\boldsymbol{I}_K$	TD, $\boldsymbol{X}$	$\boldsymbol{H}_0[k,m] = \frac{1}{N}\sum_{v=0}^{V-1} \tilde{H}[v]e^{j2\pi\frac{v[k+mK]}{N}}$
Low time-selectivity	K	M	$\boldsymbol{F}_K$	TD, $\boldsymbol{X}$	$\boldsymbol{H}_0[q,m] = \sum_{l=0}^{L-1} h[l,m]e^{-j2\pi\frac{lq}{K}}$
Low frequency-selectivity	M	K	$\boldsymbol{F}_M^H$	FD, $\tilde{\boldsymbol{X}}$	$\boldsymbol{H}_0[m,q] = \sum_{v=0}^{V-1} \tilde{H}[q,v]e^{j2\pi\frac{vm}{M}}$

The interference $\boldsymbol{Z}_0$ represents the inter-subsymbol or inter-subcarrier interference in addition to the error caused by the channel approximation. For instance, there is interference resulting from the assumption of static channel withing the subsymbol, such that $h[l,k+mK] = h[l,m] + \epsilon[k,m]$. The channel matrix can be computed from a sparse matrix parameters, such that

$$\boldsymbol{H}_0[q,m] = \frac{1}{\sqrt{M_0K_0}} \sum_{l=0}^{L_0-1}\sum_{v=0}^{V_0-1} h_0[l,v]e^{-j2\pi\frac{ql}{K_0}}e^{j2\pi\frac{mv}{M_0}}, \tag{5.31}$$

where $L_0 << K_0, V_0 << M_0$. This model allows channel estimation using few reference samples and using interpolation between them. The interpretation of the parameters w.r.t. the subscript M_0, K_0 depends on the channel mode and the processing domain.

MIMO received signal model. Let $\boldsymbol{X}_{0_{n_t}} \in \mathbb{C}^{N\times N}$ be the transmitted signal from the n_t-th antenna, $\boldsymbol{H}_{0_{n_r,n_t}} \in \mathbb{C}^{N\times N}$ is the channel between the n_t-th transmitter and n_r-th

receiver. The domain and form of the signal and channel follows the settings in Table 5.3. The received signal at the n_r-th antenna can be expressed as

$$\boldsymbol{F}\boldsymbol{Y}_{0_{n_r}} = \sum_{n_t=1}^{N_T} \left(\boldsymbol{H}_{0_{n_r,n_t}} \odot [\boldsymbol{F}\boldsymbol{X}_{0_{n_t}}] + \boldsymbol{Z}_{0_{n_r,n_t}} \right) + \boldsymbol{V}_{0_{n_r}}. \tag{5.32}$$

In order to perform channel equalization, a vector $\boldsymbol{y}_n \in \mathbb{C}^{N_R \times 1}$ is constructed by stacking the samples at the index $n = k_0 + m_0 K_0$, such that $\boldsymbol{y}_n [n_r] = [\boldsymbol{F}\boldsymbol{Y}_{0_{n_r}}][k_0, m_0]$. As a result,

$$\boldsymbol{y}_n = \boldsymbol{H}_n \bar{\boldsymbol{x}}_n + \boldsymbol{z}_n + \boldsymbol{v}_n, \tag{5.33}$$

where $\bar{\boldsymbol{x}}_n \in \mathbb{C}^{N_T \times 1}$, $\bar{\boldsymbol{x}}_n [n_t] = \left(\boldsymbol{F}\boldsymbol{X}_{0_{n_t}} \right) [k_0, m_0]$ corresponds to the transmitted samples, $\boldsymbol{H}_n \in \mathbb{C}^{N_R \times N_T}$, $\boldsymbol{H}_n[n_r, n_t] = \boldsymbol{H}_{0_{n_r,n_t}}[k_0, m_0]$ is the multiple-input, multiple-output (MIMO) channel at the n-th sample, $\boldsymbol{z}_n \in \mathbb{C}^{N_R \times 1}$, $\boldsymbol{z}_n [n_r] = \sum_{n_t=1}^{N_T} \boldsymbol{Z}_{0_{n_r,n_t}}[k_0, m_0]$ denotes the interference, and $\boldsymbol{v}_n \in \mathbb{C}^{N_R \times 1}$ is the additive noise.

The presented signal models in this section belong to a general linear system and the detection techniques of Section 3.4 can be directly applied. In particular, the vectorization of (5.30) results in $\boldsymbol{y}_0 = \boldsymbol{\Lambda}^{(h_0)} \tilde{\boldsymbol{A}}_0 \boldsymbol{d} + \boldsymbol{z}_0 + \boldsymbol{v}_0$, where $\boldsymbol{\Lambda}^{(h_0)}$ is diagonal channel matrix, and $\tilde{\boldsymbol{A}}_0 = \boldsymbol{I}_{M_0} \otimes \boldsymbol{F}\boldsymbol{A}_0$ is the precoding matrix composed of a modulation matrix $\boldsymbol{A}_0$ and channel-related transform $\boldsymbol{I}_{M_0} \otimes \boldsymbol{F}$. This model is similar to the one presented and simulated in Section 3.4.4, and the introduced precoding design approaches based on the channel statistic can be exploited for $\tilde{\boldsymbol{A}}_0$. Extensive link-level simulation is required to evaluate different waveform designs, channel models, and deployment scenarios, which is out of the scope of this thesis.

5.4 Summary

This chapter has represented the time-frequency modulation in form of subcarriers and subsymbols. In general linear modulation, the modulated block $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d}$ of size $N = MK$ can be represented using M subsymbols each of duration K samples in a block $\boldsymbol{X} = \text{unvec}_{K \times M} \{\boldsymbol{x}\}$. The frequency domain block, $\tilde{\boldsymbol{X}} = \text{unvec}_{M \times K} \{\tilde{\boldsymbol{x}}\}$ where $\tilde{\boldsymbol{x}} = \boldsymbol{F}_N \boldsymbol{x}$ focuses on the subcarrier structure. Both blocks are related by means of Zak transform $\tilde{\boldsymbol{X}} \boldsymbol{F}_K^H = \boldsymbol{\Phi}^{\mathrm{T}} \odot [\boldsymbol{F}_M \boldsymbol{X}^{\mathrm{T}}]$, $\boldsymbol{\Phi}[k, m] = e^{-j2\pi \frac{km}{N}}$. The choice of M and K depends on the selectivity of the channel. For low time selectivity, where the coherence time is large, K is chosen such that the channel is static within the subsymbol duration. Therefore, the number of subsymbols can be flexible. In this case, the processing requires K-DFT on the received block to get a simplified relation with the channel. On the contrary, with high time selectivity that mainly arises when using narrow bandwidth, the coherence bandwidth is relatively large. This enables, for sufficiently large N, choosing M such that the subcarrier width is smaller than the coherence bandwidth. However, the processing requires using the frequency domain block obtained by an N-DFT and then M-DFT to obtain the channel.

On top of that, a general signal model is derived $\boldsymbol{F}\boldsymbol{Y}_0 = \boldsymbol{H}_0 \odot [\boldsymbol{F}\boldsymbol{X}_0] + \boldsymbol{Z}_0 + \boldsymbol{V}_0$, which shows simple elementwise multiplication relation with the channel. A straightforward extension to a MIMO model has been also presented. Based on this model, the modulation matrix can be designed to enable simple equalization by the decoupling of subsymbols or subcarriers, or to harvest diversity by means of spreading. These design criteria and other aspects will be further investigated in the next chapter.

Chapter 6

Generalized waveforms based on time-frequency shifts

In common multicarrier systems, such as filtered multi tone (FMT) [CEO99], the carrier pulses are generated by means of frequency shifts of a finite-length prototype pulse $g[n]$. As discussed in 5.1, the multicarrier can be reformulated to include time shifts leading to the definition of subsymbol. This allows the representation of multicarrier by means of frequency-time pulses corresponding to a pair of subcarrier and subsymbol. Accordingly, orthogonal frequency division multiplexing (OFDM) [SCC93] can be described by a single subsymbol and multiple subcarriers employing frequency shifts on a rectangular pulse, whereas single-carrier (SC) [NLF07] consists of one subcarrier and multiple subsymbols. Moreover, generalized frequency division multiplexing (GFDM) is a general case that uses multiple subcarriers and subsymbols [MMG+14]. Although GFDM employs circular filtering, it can be used to represent FMT relying on switching off some subsymbols and controlling the overlap between blocks [NEM+19].

The common trend among the aforementioned multicarrier systems is that the time and frequency shifts are performed with equal steps. The previous chapter suggested a generic design of the time-frequency pulses. This chapter introduces a special case, where a prototype pulse is used, but the time and frequency shift can be done arbitrary. The general idea is presented in Section 6.1 demonstrating how multicarrier systems can be interpreted from different perspectives. In Section 6.2, the general case of equally-spaced shifts is revised in the context of Gabor frame. It is shown that GFDM is a special case and it can be used to represent a general Gabor frame by means of multiple prototype pulses. Due to the important role of GFDM modulation in developing a flexible multicarrier framework, Section 6.3 is dedicated to provide thorough details on GFDM representations and modem structure. In addition, an alternative interpretation of GFDM is discussed, which paves the way to extended GFDM framework, which will be discussed in the next chapter. The contents of this chapter are partially based on the works reported in [NMZF17], [NEM+19], [NCMF18], [NCF18] and [NCF19b].

6.1 General time-frequency shift

This thesis presents generic time and frequency shifts applied on a given prototype pulse $g[n]$ of length N samples in order to generate the time-frequency pulses $\{g_{k,m}[n]\}$. The special well-known cases are derived accordingly. In order to preserve the length N, circular time shift is used instead of linear shift. Let $n_{k,m}, s_{k,m} \in \{0, \cdots, N-1\}$ be the number of samples corresponding to the time and frequency shift, respectively, then

$$g_{k,m}[n] = g[\langle n - n_{k,m} \rangle_N] e^{j2\pi \frac{[n-n_{k,m}]s_{k,m}}{N}}. \tag{6.1}$$

Relation to analog representation. $g[n]$ can be seen as the discrete signal obtained from the sampling of one period of a band-limited periodic signal $g_p(t)$ of period $T = \frac{N}{F_s}$ using the sampling frequency F_s. This can be formulated as

$$g[n] = g_T(\frac{n}{F_s}),\ g_T(t) = w_T(t) g_p(t),\ g_p(t) = \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} \tilde{g}[s] e^{j2\pi \frac{st}{T}}. \tag{6.2}$$

where $w_T(t)$ is rectangular window of length $T = \frac{F_s}{N}$, and $\tilde{g}[s]$ corresponds to the Fourier series coefficients and it is the N-discrete Fourier transform (DFT) of $g[n]$. Accordingly, $g_{k,m}[n]$ results form the sampling of the analog pulse

$$g_{T,k,m}(t) = w_T(t) g_p(t - \Delta t_{k,m}) e^{j2\pi(t-\Delta t_{k,m})\Delta f_{k,m}} = g_T(\langle t - \Delta t_{k,m} \rangle_T) e^{j2\pi(t-\Delta t_{k,m})\Delta f_{k,m}}.$$

Here, $\Delta t_{k,m} = \frac{n_{k,m}}{F_s}$ and $f_{k,m} = \frac{s_{k,m}}{N} F_s$ are the time and frequency shifts applied on the periodic pulse. Therefore, the circular time shift results from the time periodicity, whereas the circular frequency shift is because the spectrum of the discrete signal is periodic. Using the N-DFT, the frequency-domain (FD) pulse is obtained as

$$\tilde{g}_{k,m}[s] = \tilde{g}[\langle s - s_{k,m} \rangle_N] e^{-j2\pi \frac{s n_{k,m}}{N}}.$$

6.1.1 Time-frequency shift design

For a given prototype pulse, a set $\mathcal{G}_g = \{g_{v,u}, (u,v) \in \{0, \cdots, N-1\}^2\}$ of N^2 different possible pulses result from the time and frequency shift, where

$$g_{v,u}[n] = g[\langle n - u \rangle_N] e^{j2\pi \frac{[n-u]v}{N}},\ \tilde{g}_{v,u}[s] = \tilde{g}[\langle s - v \rangle_N] e^{-j2\pi \frac{su}{N}}, \tag{6.3}$$

A time-frequency modulation can be obtained by choosing a subset of N different pulses from $\mathcal{G}_g$. Therefore, there are $\prod_{q=0}^{N-1}(N^2 - q) = \frac{N^2!}{N!(N^2-N)!}$ possible combinations to formulate the modulation matrix $\boldsymbol{A} \in \mathbb{C}^{N \times N}$, $\boldsymbol{A}[k + mK, n] = g_{k,m}[n]$. The are several criteria of selecting the shift value. For instance, orthogonality, time-frequency spreading, and implementation complexity. The arbitrary shift belongs to generic linear modulation. However, the realization requires lower storage, where only the N complex samples of the pulse shape in addition to $2N$ integers to store the shift values. Nevertheless, in general, the pulse shaping still requires N^2 complex multiplications. In some particular cases, the implementation can be achieved with low-complexity transforms.

Frequency shift. In this case, the subset is given by $g_{v,0}[n] = g[n]e^{j2\pi\frac{nv}{N}}$. Thereby, $s_{k,m}$ is mapped to v and $n_{k,m} = 0$. One possible mapping is given by $s_{k,m} = m + kM$, while the other mapping options correspond to permutation of the same set of pulses. As a result,

$$g_{k,m}[n] = g[n]e^{j2\pi\frac{nm}{N}}e^{j2\pi\frac{nk}{K}} \Longrightarrow \boldsymbol{A} = \boldsymbol{\Lambda}^{(g)}\boldsymbol{F}_N^{\mathrm{H}}, \boldsymbol{\Lambda}^{(g)} = \frac{1}{\sqrt{N}}\mathrm{diag}\left\{\{g[n]\}_{n=0}^{N-1}\right\}.$$

This modulation is windowed OFDM [AKR16], where $g[n]$ plays the role of window.

Time shift. Here, the set of pure time shift is chosen, i.e. $g_{0,u}[n] = g[\langle n-u\rangle_N]$, or equivalently $\tilde{g}_{u,0}[s] = \tilde{g}[s]e^{-j2\pi\frac{su}{N}}$. Thus, $s_{k,m} = 0$, and by using the mapping $n_{k,m} = k + mK$, the obtained pulses are

$$\tilde{g}_{k,m}[n] = \tilde{g}[s]e^{-j2\pi\frac{sk}{N}}e^{-j2\pi\frac{sm}{M}},\ g_{k,m}[n] = g[\langle n-k-mK\rangle_N].$$

Here, $\tilde{\boldsymbol{A}} = \boldsymbol{\Lambda}^{(\tilde{g})}\boldsymbol{F}_N, \boldsymbol{\Lambda}^{(\tilde{g})} = \frac{1}{\sqrt{N}}\mathrm{diag}\left\{\{\tilde{g}[s]\}_{s=0}^{N-1}\right\}$, which is equivalent to filtered SC using circular filtering with the filter $g[n]$.

Time and frequency shifts. All other cases employ time and frequency shift, and the result depends on the selected pulses. For example, the choice $n_{k,m} = k$ and $s_{k,m} = m$

$$g_{k,m}[n] = g[\langle n-k\rangle_N]e^{j2\pi\frac{nm}{N}}.$$

For prototype pulse defined as a train of Dirac pulses as $g[n] = \sum_{m=0}^{M-1}\delta[n-mK]$,

$$g_{k,m}[q+pK] = g[q+pK-k]e^{j2\pi\frac{[q+pK-k]m}{N}} = \delta[q-k]e^{j2\pi\frac{pm}{M}}.$$

Therefore, using the 2-D block representation in Section 5.1.1

$$\boldsymbol{X}[q,p] = \sum_{m=0}^{M-1}\boldsymbol{D}[q,m]e^{j2\pi\frac{pm}{M}} \Rightarrow \boldsymbol{X} = \boldsymbol{D}\boldsymbol{F}_M^{\mathrm{H}}.$$

This waveform is orthogonal time frequency space (OTFS) [NCMF18]. Another special case of time-frequency shift is related to Gabor frame, which will be introduced in Section 6.2.

6.1.2 Relation between the shifted pulses

The inner product between the pulses g_{u_1,v_1} g_{u_2,v_2} is given by

$$\begin{aligned}\langle g_{v_1,u_1}, g_{v_2,u_2}\rangle &= \sum_{n=0}^{N-1} g_{v_1,u_1}[n]g^*_{v_2,u_2}[n] = \sum_{n=0}^{N-1} g[n]g^*[\langle n-\Delta u\rangle_N]e^{j2\pi\frac{nv_1}{N}}e^{-j2\pi\frac{[n-\Delta u]v_2}{N}} \\ &= e^{j2\pi\frac{\Delta uv_1}{N}}\sum_{n=0}^{N-1} g[n]g^*[\langle n-\Delta u\rangle_N]e^{-j2\pi\frac{[n-\Delta u]\Delta v}{N}} = e^{j2\pi\frac{\Delta uv_1}{N}}\langle g, g_{\Delta v,\Delta u}\rangle,\end{aligned} \tag{6.4}$$

where $\Delta u = u_2 - u_1$, $\Delta v = v_2 - v_1$. The function $A(\Delta v, \Delta u) = \langle g, g_{\Delta v, \Delta u} \rangle$, known as the ambiguity function [MCC98] associated with the pulse $g[n]$, helps identifying the orthogonality of candidate pulses obtained from time and frequency shifts. For example in the case of rectangular pulse, $A(\Delta v, \Delta u) = N\delta[\Delta v]$, which means, that the relative frequency shift must be larger than 0 in order to get orthogonal pulses. Another example, for the pulse $g_c[n] = e^{j\pi \frac{n^2}{N}}$, which is known as the Chirp pulse,

$$\begin{aligned}\langle g_c, g_{c,\Delta v, \Delta u} \rangle &= \sum_{n=0}^{N-1} e^{j\pi \frac{n^2}{N}} e^{-j\pi \frac{(n-\Delta u)^2}{N}} e^{-j2\pi \frac{[n-\Delta u]\Delta v}{N}} \\ &= N e^{j2\pi \frac{\Delta u^2}{N}} \delta[\Delta u - \Delta v].\end{aligned}$$

Therefore, the relative shift between the pulses must satisfy $\Delta u \neq \Delta v$ for any two pulses. This can be achieved, for example, by fixing the time shift to 0 and varying the frequency shift, or vice versa. In the first case $g_v[n] = e^{j\pi \frac{n^2}{N}} e^{j2\pi \frac{nv}{N}}$, which is equivalent to windowed OFDM, where the window is $g[n] = e^{j\pi \frac{n^2}{N}}$, whereas the second case $g_u[n] = e^{j\pi \frac{(n-u)^2}{N}}$ is equivalent to filtered SC.

The ambiguity function is related to the time-frequency distribution with Wigner transform defined for analog signal as

$$A(f, \tau) = \int_{-\infty}^{\infty} g(t) g^*(t - \tau) e^{j2\pi f t}, \tag{6.5}$$

where an indication of the localization of the pulse is obtained. Accordingly, with respect to the discrete pulses, the rectangular pulse is localized in the frequency, and spread in time, whereas the delta pulse $g[n] = \delta[n]$ is localized in time but spread in frequency. On the other hand, the Chirp pulse is spread equally in time and frequency.

6.2 Time-frequency shift in Gabor frame

A particular case of the shift design is when using uniform shifts in both domains, where $n_{m,k} = mP$ in the time-domain (TD) and $s_{k,m} = kQ$ in the frequency, so that

$$g_{k,m}[n] = g[\langle n - mP \rangle_N] e^{j2\pi [n - mP] k \frac{Q}{N}}. \tag{6.6}$$

The derived pulses belong to Gabor time-frequency lattices [D+94], which are defined as $g_{k\alpha, m\beta}[n] = g[\langle n - m\beta \rangle_N] e^{j2\pi n k \alpha}$ with $\alpha = Q/N$, and $\beta = P$. In order to uniquely demodulate the data symbols $d_{k,m}$ from a given $x[n]$, Wexler-Raz duality condition [WR90] must be satisfied. Thus, there exists a pulse $\gamma[n]$ that attains $\langle \gamma, g_{k/\beta, m/\alpha} \rangle = \alpha\beta\delta_{k,0}\delta_{m,0}$, where $g_{k/\beta, m/\alpha}$ is the dual Gabor lattice, δ_{ij} refers to Kronecker delta, and $\langle \cdot \rangle$ denotes the inner product. Furthermore, γ is the demodulator prototype pulse, such that $d_{k,m} = \langle x, h_{k\alpha, m\beta} \rangle$. It is rigorously proven in [Jan94] that Wexler-Raz duality condition cannot be fulfilled if $\alpha\beta > 1$. In other words, the condition $QP \leq N$ is required to obtain a

non-singular modulation matrix $\boldsymbol{A}$, as reported in [NCF18]. Based on that, the choice of $P = \Delta t F_s$ and $Q = T\Delta f$ is influenced by

1. $QP \leq N$, which is necessary but not sufficient to achieve Wexler-Raz duality condition.
2. $\Delta f K \leq F_s = \frac{N}{T}$, which is a necessary, but not sufficient condition for the signal to have a bandwidth $B \leq F_s$.

Consequently, the design needs to fulfill the conditions

$$Q \leq M,\ P \leq K.$$

The case where $Q = M$, and $P = K$ corresponds to to critically sampled system, which is known as GFDM system [MMG+14].

6.2.1 Conventional GFDM

The time-frequency pulses corresponding to GFDM are given by

$$g_{k,m}[n] = g[\langle n - mK \rangle_N] e^{j2\pi \frac{nk}{K}}. \tag{6.7}$$

This design of the shifts enables low-complexity implementation, as will be discussed in Section 6.3. The other cases, where $PQ < N$, i.e. an over-sampled system, can be redesigned to satisfy the condition $PQ = N$ by managing the sampling frequency, the block length, and properly defining the active sets, as reported in [NCF18].

Time-domain reformulation. Let $Q = T\Delta f = M$ and $P = \Delta t F_s = \frac{K}{L}$, where L is a positive integer. The set of active subcarriers $\mathcal{K}_{on}$ is adjusted to the available bandwidth $B \leq F_s$. Considering $PQ = N$, the modulation pulses can be redefined as

$$g_{k,mL+l}[n] = g[\langle n - (mL+l)P \rangle_N] e^{j2\pi[n-(mL+l)P]k\frac{Q}{N}} = g(\left\langle n - mK - l\tfrac{K}{L} \right\rangle_N) e^{j2\pi \frac{nk}{K}}.$$

Accordingly L prototype pulses can be defined as

$$g^{(l)}[n] = g[\left\langle n - l\tfrac{K}{L} \right\rangle_N],\ l = 0 \cdots, L-1.$$

These prototype pulses are achieved with different shift of the original prototype pulse. Each of these prototype pulses is used to generate a subset of the pulses with the subsymbol spacing K and the set

$$\mathcal{M}_{on}{}^{(l)} = \{mL + l,\ l \leq mL + l < M\}.$$

The final block is obtained as a superposition of L GFDM blocks,

$$x[n] = \sum_{l=0}^{L-1} x^{(l)}[n],\ \text{where } x^{(l)}[n] = \sum_{k \in \mathcal{K}_{on}} \sum_{m \in \mathcal{M}_{on}{}^{(l)}} d_{k,m} g^{(l)}[\langle n - mK \rangle_N] e^{j2\pi \frac{nk}{K}}. \tag{6.8}$$

Frequency-domain reformulation. An alternative reformulation can be achieved in the FD with respect to $\tilde{g}_{k,m}[q] = \tilde{g}[\langle q - kQ \rangle_N] e^{-j2\pi \frac{qmP}{N}}$. The design is adjusted such that $P = \Delta t F_s = K$ and $Q = \frac{M}{L}$ resulting in the prototype pulses

$$\tilde{g}^{(l)}[q] = \tilde{g}[\left\langle q - l\tfrac{M}{L} \right\rangle_N], \; l = 0 \cdots, L-1,$$

for the sets $\mathcal{K}_{on}{}^{(l)} = \{kL + l, \; l \leq kL + l < K\}$.

GFDM with multiple prototype pulses. The reformulation of general Gabor frame leads to propose the definition of multiple prototype pulses in order to reduce the implementation complexity. This approach can be generalized to design a waveform with L independent prototype pulses $\{g^{(l)}[n]\}$. Each prototype pulse is associated with the sets $\mathcal{K}_{on}{}^{(l)}$ and $\mathcal{M}_{on}{}^{(l)}$, as shown in Fig. 6.1. Thus, for $(k, m) \in \mathcal{K}_{on}{}^{(l)} \times \mathcal{M}_{on}{}^{(l)}$

$$g_{k,m}[n] = g^{(l)}(\langle n - mK \rangle_N) e^{j2\pi \frac{nk}{K}}. \tag{6.9}$$

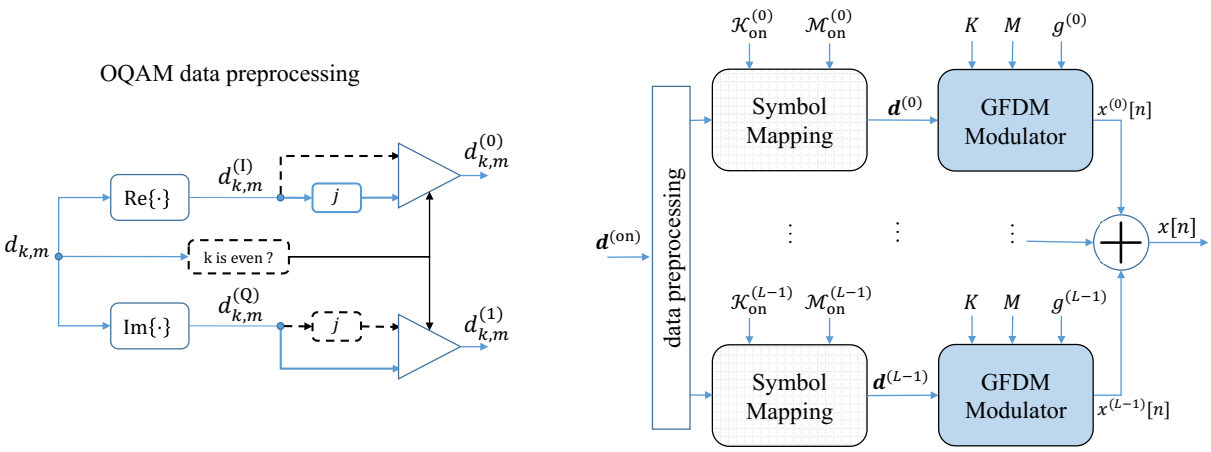

Figure 6.1: Core block with multiple prototype pulses.

The input data symbol can be preprocessed prior to modulation, for instance to produce offset quadrature amplitude modulation (OQAM) [G+15]. This design allows the realization of wide range of multicarrier waveforms. For example, to generate filter bank multicarrier (FBMC), where $Q = M$ and $P = K/2$. First, the complex QAM data symbols $d_{k,m} = d_{k,m}^{(I)} + j d_{k,m}^{(Q)}$ are split into two OQAM precoded streams, $d_{k,m}^{(0)} = \theta_{0,k} d_{k,m}^{(I)}$, $d_{k,m}^{(1)} = \theta_{1,k} d_{k,m}^{(Q)}$, where

$$\theta_{0,k} = \left\{ \begin{array}{ll} j, & k \text{ is even} \\ 1, & k \text{ is odd} \end{array} \right\}, \theta_{1,k} = \left\{ \begin{array}{ll} 1, & k \text{ is even} \\ j, & k \text{ is odd} \end{array} \right\}.$$

The streams are then fed to two GFDM modulators with the parameters $g^{(0)}[n] = g[n]$, $g^{(1)}[n] = g[< n - K/2 >]$, $\mathcal{K}_{on}{}^{(0)} = \mathcal{K}_{on}{}^{(1)}$, and $\mathcal{M}_{on}{}^{(0)} = \mathcal{M}_{on}{}^{(1)}$. The output is a

superposition of both GFDM blocks, i.e. $x[n] = x^{(0)}[n] + x^{(1)}[n]$. It is worth noting that OQAM modulation can also be represented using the proposed multiple-input, multiple-output (MIMO) waveform approach presented in section 2.3.3.

In the remaining of this chapter, GFDM modulation will be discussed as a common building block for wide range of time-frequency based waveforms. These waveforms can be defined by a single prototype pulse shape, the number of subcarriers and subsymbols, and the active subcarrier and subsymbols sets, as listed in Table 6.1.

6.3 GFDM modulation

The pulses in GFDM are generated by means of a prototype pulse $g[n]$, such that

$$g_{k,m}[n] = g[\langle n - mK \rangle_N] e^{j2\pi \frac{nk}{K}}, \ \tilde{g}_{k,m}[s] = \tilde{g}[\langle s - kM \rangle_N] e^{-j2\pi \frac{sm}{M}}. \tag{6.10}$$

Not that normalized DFT is considered, where $\tilde{g}[s] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} g[n] e^{-j2\pi \frac{ns}{N}}$. The pulses have similar formulation in the TD and FD. First, the TD structure is considered and then used to derive the concrete FD form.

Table 6.1: GFDM based waveform parameters

N. of subcarriers	K
N. of subsymbols	M
Active subcarriers	$\mathcal{K}_{on} \subset \{0, \cdots K-1\}$
Active subsymbols	$\mathcal{M}_{on} \subset \{0, \cdots M-1\}$
Active symbols	$\mathcal{N}_{on} = \{k + mK, (k,m) \in \mathcal{K}_{on} \times \mathcal{M}_{on}\}$
Prototype pulse	$g[n]$
Pulse shapes	$g_{k,m}[n] = g[\langle n - mK \rangle_N] e^{j2\pi \frac{nk}{K}}$
GFDM block	$x[n] = \sum_{k \in \mathcal{K}_{on}} \sum_{m \in \mathcal{M}_{on}} d_{k,m} g_{k,m}[n]$

6.3.1 Filter bank representation

The TD-GFDM block can be expressed as

$$x[n] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \boldsymbol{D}[k,m] g[\langle n - mK \rangle_N] e^{j2\pi \frac{k}{K} n}. \tag{6.11}$$

This equation can be seen as a circular convolution with prototype pulse shape g and frequency upconversion, as seen in Figure 6.2,

$$x[n] = \sum_{k=0}^{K-1} \left[g[n] \circledast \sum_{m=0}^{M-1} \boldsymbol{D}[k,m] \delta[n - mK] \right] e^{j2\pi \frac{k}{K} n}.$$

The circular filtering was introduced in the original GFDM publication [FKB09] as an approach for digital implementation of classical filter bank. The circular convolution results from linear filtering with tail biting to reduce the overhead due to the filter tail and allow adding cyclic prefix (CP). This representation corresponds to the conventional GFDM design, where the focus is on specific type of prototypes that span a spacing of two subcarriers, as reported in [NMZF17].

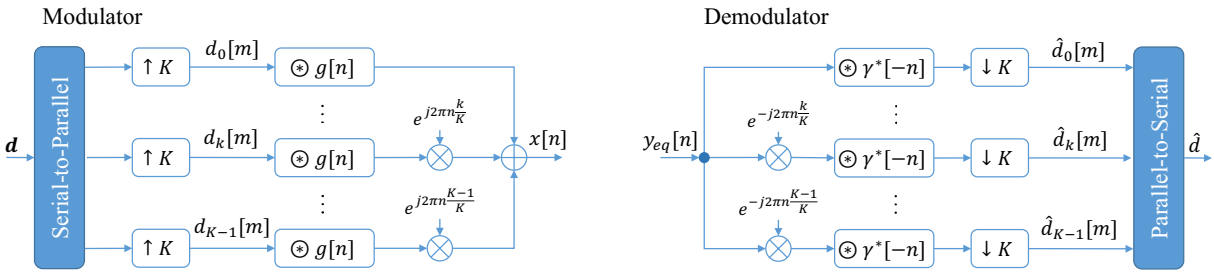

Figure 6.2: GFDM modem with circular convolution representation.

6.3.2 Block representation

This section presents the GFDM modulation from different perspectives. The block representation follows the general model discussed in Section 5.1.1. The corresponding block pulses in the TD are given by

$$\boldsymbol{G}_{k,m}[q,p] = g[\langle q+(p-m)K\rangle_N]e^{j2\pi\frac{[q+pK]k}{K}} = \boldsymbol{G}[q,\langle p-m\rangle_M]e^{j2\pi\frac{gk}{K}}. \tag{6.12}$$

Therefore, the TD block is given by

$$\boldsymbol{X}[q,p] = \sum_{k=0}^{K-1}\sum_{m=0}^{M-1}\boldsymbol{D}[k,m]\boldsymbol{G}[q,\langle p-m\rangle_M]e^{j2\pi\frac{gk}{K}} = \sqrt{K}\sum_{m=0}^{M-1}[\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}][q,m]\boldsymbol{G}[q,\langle p-m\rangle_M]. \tag{6.13}$$

Thus, the m-th row of the TD block results from circular convolution between the m-th row of $\boldsymbol{G}$ and the m-th row of the matrix $\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}$. As a result,

$$[\boldsymbol{X}\boldsymbol{F}_M][q,p] = \sqrt{KM}[\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M][q,p][\boldsymbol{G}\boldsymbol{F}_M][q,p].$$

Therefore, using the expression of Zak transform $\boldsymbol{Z}_{M,K}^{(g)} = \boldsymbol{F}_M\boldsymbol{G}^{\mathrm{T}}$.

$$\boldsymbol{X}\boldsymbol{F}_M = \sqrt{N}\{\boldsymbol{Z}_{M,K}^{(g)}\}^{\mathrm{T}} \odot [\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M], \tag{6.14}$$

Similarly, the FD block can be expressed by means of the circular convolution

$$\tilde{\boldsymbol{X}}[p,q] = \sqrt{M}\sum_{k=0}^{K-1}[\boldsymbol{D}\boldsymbol{F}_M][k,p]\tilde{\boldsymbol{G}}[p,\langle q-k\rangle_K]. \tag{6.15}$$

Recalling, from the relation between the Zak transforms in the TD and FD (5.14)

$$\begin{aligned}\tilde{\boldsymbol{X}}\boldsymbol{F}_K^{\mathrm{H}} &= \boldsymbol{\Phi}^{\mathrm{T}} \odot [\boldsymbol{X}\boldsymbol{F}_M]^{\mathrm{T}} = \sqrt{N}\boldsymbol{\Phi}^{\mathrm{T}} \odot \{\boldsymbol{Z}_{M,K}^{(g)}\} \odot [\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M]^{\mathrm{T}} \\ &= \sqrt{N}\{\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{g})}\}^{\mathrm{T}} \odot [\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M]^{\mathrm{T}}, \quad \{\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{g})}\} = \boldsymbol{F}_K^{\mathrm{H}}\tilde{\boldsymbol{G}}^{\mathrm{T}}.\end{aligned} \tag{6.16}$$

6.3.3 GFDM matrix structure

The TD block can be presented in a vector form $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d}$, where $\boldsymbol{d} = \text{vec}\{\boldsymbol{D}\}$ and $\boldsymbol{A}[n, m + kM] = g_{k,m}[n]$. The structure of the GFDM modulation matrix can be derived by the vectorization of (6.14) and (6.16). To do so, the following notation are defined. For a matrix $\boldsymbol{Z} \in \mathbb{C}^{Q\times P}$,

$$\begin{aligned} \text{vec}\{\boldsymbol{F}_Q\boldsymbol{Z}\} = \boldsymbol{U}_{P,Q}\text{vec}\{\boldsymbol{Z}\}\,,\ \boldsymbol{U}_{P,Q} = \boldsymbol{I}_P \otimes \boldsymbol{F}_Q \\ \boldsymbol{\Pi}_{Q,P} : \text{vec}\left\{\boldsymbol{Z}^{\mathrm{T}}\right\} = \boldsymbol{\Pi}_{Q,P}\text{vec}\{\boldsymbol{Z}\}\,, \end{aligned} \tag{6.17}$$

$\boldsymbol{U}_{P,Q}$ is unitary matrix and $\boldsymbol{\Pi}_{Q,P}$ permutation matrix known as the commutative matrix of size $QP \times QP$ defined by $\boldsymbol{\Pi}_{Q,P}[p + qP, q + pQ] = 1$. Note that $\boldsymbol{\Pi}_{Q,P}^{-1} = \boldsymbol{\Pi}_{P,Q}$. By vectorizing both sides of the TD relation, $\boldsymbol{X}\boldsymbol{F}_M = \sqrt{N}\left[\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M\right] \odot \{\boldsymbol{Z}_{M,K}^{(g)}\}^{\mathrm{T}}$

$$\begin{aligned} \text{vec}\{\boldsymbol{X}\boldsymbol{F}_M\} = \boldsymbol{\Pi}_{M,K}\text{vec}\left\{\boldsymbol{F}_M\boldsymbol{X}^{\mathrm{T}}\right\} = \boldsymbol{\Pi}_{M,K}\boldsymbol{U}_{K,M}\text{vec}\left\{\boldsymbol{X}^{\mathrm{T}}\right\}, \\ \text{vec}\left\{\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M\right\} = \boldsymbol{\Pi}_{M,K}\text{vec}\left\{\boldsymbol{F}_M\boldsymbol{D}^{\mathrm{T}}\boldsymbol{F}_K^{\mathrm{H}}\right\} = \boldsymbol{\Pi}_{M,K}\boldsymbol{U}_{K,M}\text{vec}\left\{\boldsymbol{D}^{\mathrm{T}}\boldsymbol{F}_K^{\mathrm{H}}\right\}. \end{aligned}$$

Let $\boldsymbol{\Lambda}^{(g)} = \sqrt{N}\text{diag}\left\{\text{vec}\left\{\{\boldsymbol{Z}_{M,K}^{(g)}\}^{\mathrm{T}}\right\}\right\}$, and because $\boldsymbol{x} = \boldsymbol{A}\text{vec}\{\boldsymbol{D}\}$, then

$$\boldsymbol{A} = \underbrace{\boldsymbol{\Pi}_{M,K}\boldsymbol{U}_{K,M}^{\mathrm{H}}}_{\boldsymbol{V}_t}\boldsymbol{\Lambda}^{(g)}\underbrace{\boldsymbol{U}_{K,M}\boldsymbol{\Pi}_{K,M}\boldsymbol{U}_{M,K}^{\mathrm{H}}}_{\boldsymbol{U}_t}. \tag{6.18}$$

The matrices $\boldsymbol{V}_t$ and $\boldsymbol{U}_t$ are unitary matrices, and therefore, the condition of $\boldsymbol{A}$ is influenced by the diagonal elements of $\boldsymbol{\Lambda}^{(g)}$. Another decomposition formula can be achieved using different order of the vectorization,

$$\text{vec}\left\{\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D}\boldsymbol{F}_M\right\} = \boldsymbol{U}_{M,K}^{\mathrm{H}}\text{vec}\{\boldsymbol{D}\boldsymbol{F}_M\} = \boldsymbol{U}_{M,K}^{\mathrm{H}}\boldsymbol{\Pi}_{M,K}\text{vec}\left\{\boldsymbol{F}_M\boldsymbol{D}^{\mathrm{T}}\right\}$$

$$\boldsymbol{A} = \boldsymbol{\Pi}_{M,K}\boldsymbol{U}_{K,M}^{\mathrm{H}}\boldsymbol{\Pi}_{M,K}\boldsymbol{\Lambda}^{(g)}\boldsymbol{U}_{M,K}^{\mathrm{H}}\boldsymbol{\Pi}_{M,K}\boldsymbol{U}_{K,M}\boldsymbol{\Pi}_{K,M}.$$

The FD block is accordingly $\tilde{\boldsymbol{x}} = \boldsymbol{F}_N\boldsymbol{A}\boldsymbol{d}$. Thus, $\tilde{\boldsymbol{A}} = \boldsymbol{F}_N\boldsymbol{A}$ represents the FD modulation matrix. The structure of the FD matrix is obtained, from $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{A}}\boldsymbol{\Pi}_{M,K}\text{vec}\left\{\boldsymbol{D}^{\mathrm{T}}\right\}$, which can be derived from $\tilde{\boldsymbol{X}}\boldsymbol{F}_K^{\mathrm{H}} = \sqrt{N}[\boldsymbol{F}_M\boldsymbol{D}^{\mathrm{T}}\boldsymbol{F}_K^{\mathrm{H}}] \odot \{\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{g})}\}^{\mathrm{T}}$,

$$\tilde{\boldsymbol{A}} = \underbrace{\boldsymbol{\Pi}_{K,M}\boldsymbol{U}_{M,K}}_{\boldsymbol{V}_f}\boldsymbol{\Lambda}^{(\tilde{g})}\underbrace{\boldsymbol{U}_{M,K}^{\mathrm{H}}\boldsymbol{\Pi}_{M,K}\boldsymbol{U}_{K,M}}_{\boldsymbol{U}_f}\boldsymbol{\Pi}_{K,M}, \tag{6.19}$$

where $\boldsymbol{\Lambda}^{(\tilde{g})} = \sqrt{N}\text{diag}\left\{\text{vec}\left\{\{\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{g})}\}^{\mathrm{T}}\right\}\right\}$. The permutation $\boldsymbol{\Pi}_{K,M}$ is equivalent to vectorization with respect to the rows. Any matrix that fulfills one of those structures is called GFDM matrix. As a result, the condition of the modulation matrix is determined from the TD or FD Zak transform of $g[n]$ given by $z_{m,k} = \boldsymbol{Z}_{M,K}^{(g)}[m, k]$ and $\bar{z}_{k,m} = \bar{\boldsymbol{Z}}_{K,M}^{(\tilde{g})}[k, m]$, respectively. For instance, a real symmetric pulse produces singular matrix when both K and M are even numbers [NMZF17].

6.3.4 GFDM demodulator

The demodulator performs the inverse operations of the modulator. It is represented by the matrix $\boldsymbol{B} \in \mathbb{C}^{N\times N}$, which needs, under ideal conditions, to fulfill the condition $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}_N$. For instance, given that $\boldsymbol{A}$ is non-singular, the zero-forcing (ZF) demodulator matrix is $\boldsymbol{B}_{\mathrm{zf}} = \boldsymbol{A}^{-H} = \boldsymbol{V}_t\{\boldsymbol{\Lambda}^{(g)H}\}^{-1}\boldsymbol{U}_t$, which is also a GFDM matrix. In general, a GFDM demodulation matrix can be defined by $\boldsymbol{B} = \boldsymbol{V}_t\boldsymbol{\Lambda}^{(\gamma)}\boldsymbol{U}_t$, where $\gamma[n]$ is the demodulator pulse defined such that

$$\boldsymbol{\Lambda}^{(\gamma)} = \sqrt{N}\mathrm{diag}\left\{\mathrm{vec}\left\{\{\boldsymbol{Z}^{(\gamma)}_{M,K}\}^{\mathrm{T}}\right\}\right\},\ \boldsymbol{B}[n, m+kM] = \gamma[\langle n - mK\rangle_N]e^{j2\pi\frac{nk}{K}}. \tag{6.20}$$

The demodulated symbols are given by $\mathrm{vec}\left\{\hat{\boldsymbol{D}}\right\} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{y}_{\mathrm{eq}}$, where $\boldsymbol{y}_{\mathrm{eq}} \in \mathbb{C}^{N\times 1}$ is the received block after channel equalization. The demodulation can be expressed as a filter bank with circular convolution, as depicted in Figure 6.2 such that

$$\hat{\boldsymbol{D}}[k,m] = \sum_{n=0}^{N-1} y_{\mathrm{eq}}[n]\gamma^*[\langle n - mK\rangle_N]e^{-j2\pi n\frac{k}{K}} = \left[y_{\mathrm{eq}}[n]e^{-j2\pi n\frac{k}{K}}\right] \circledast \gamma^*(-n)|_{n=mK}. \tag{6.21}$$

By means of 2D block using $\boldsymbol{\Gamma}_{k,m}[q,p] = \boldsymbol{\Gamma}[q, \langle m-p\rangle_M]e^{j2\pi\frac{qk}{K}}$, and $\boldsymbol{\Gamma}[q,p] = \gamma[q-pK]$,

$$\begin{aligned}\hat{\boldsymbol{D}}[k,m] &= \sum_{q=0}^{K-1}\sum_{p=0}^{M-1} \boldsymbol{Y}_{\mathrm{eq}}[q,p]\boldsymbol{\Gamma}^*[q,\langle m-p\rangle_M]e^{-j2\pi\frac{qk}{K}},\\ [\boldsymbol{F}_K^{\mathrm{H}}\hat{\boldsymbol{D}}][q,m] &= \sqrt{K}\sum_{p=0}^{M-1} \boldsymbol{Y}_{\mathrm{eq}}[q,p]\boldsymbol{\Gamma}^*[q,\langle m-p\rangle_M].\end{aligned} \tag{6.22}$$

Similarly the FD demodulator can be expressed using $\tilde{\boldsymbol{\Gamma}}[q,p] = \tilde{\gamma}[q-pK]$ as

$$[\boldsymbol{F}_M\hat{\boldsymbol{D}}^{\mathrm{T}}][p,q] = \sqrt{M}\sum_{k=0}^{K-1} \tilde{\boldsymbol{Y}}_{\mathrm{eq}}[p,k]\tilde{\boldsymbol{\Gamma}}^*[p,\langle q-k\rangle_K]. \tag{6.23}$$

Demodulator pulse design. The GFDM demodulator type is defined by the relation between $\boldsymbol{\Lambda}^{(g)}$ and $\boldsymbol{\Lambda}^{(\gamma)H}$, which influences the product $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{U}_t^{\mathrm{H}}\boldsymbol{\Lambda}^{(\gamma)H}\boldsymbol{\Lambda}^{(g)}\boldsymbol{U}_t$. For instance, the ZF satisfies $\boldsymbol{\Lambda}^{(\gamma)H}_{\mathrm{zf}} = \boldsymbol{\Lambda}^{(g)-1}$, whereas the matched simply uses $\gamma[n] = g[n]$. The choice of the demodulator depends on the channel. Consider the signal model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{d}+\boldsymbol{v}$., where $\boldsymbol{v}$ is additive noise. The goal is to design a pulse shape that minimizes the mean squared error (MSE), as reported in [NCF19b],

$$\mathrm{MSE}(\boldsymbol{\Lambda}^{(\gamma)}) = \mathrm{E}\left[\|\boldsymbol{B}^{\mathrm{H}}\boldsymbol{y} - \boldsymbol{d}\|^2\right]. \tag{6.24}$$

This is equivalent to find linear minimum mean square error (LMMSE) on the model

$$\underbrace{\boldsymbol{V}_t^{\mathrm{H}}\boldsymbol{y}}_{\bar{\boldsymbol{y}}} = \boldsymbol{\Lambda}^{(g)}\underbrace{\boldsymbol{U}_t\boldsymbol{d}}_{\bar{\boldsymbol{d}}} + \underbrace{\boldsymbol{V}_t^{\mathrm{H}}\boldsymbol{v}}_{\bar{\boldsymbol{v}}} \tag{6.25}$$

under the constraint of diagonal matrix. Let $\boldsymbol{R}_{\bar{v}} = \mathrm{E}\left[\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}^{\mathrm{H}}\right]$ and $\boldsymbol{R}_{\bar{d}} = \mathrm{E}\left[\bar{\boldsymbol{d}}\bar{\boldsymbol{d}}^{\mathrm{H}}\right]$. The optimal LMMSE matrix $\boldsymbol{\Psi}^{(\mathrm{rx})}$, where $\hat{\bar{\boldsymbol{d}}} = \boldsymbol{\Psi}^{(\mathrm{rx})H}\boldsymbol{y}$, can be expressed as

$$\boldsymbol{\Psi}^{(\mathrm{rx})H} = \left(\boldsymbol{\Lambda}^{(g)H}\boldsymbol{R}_{\bar{v}}^{-1}\boldsymbol{\Lambda}^{(g)} + \boldsymbol{R}_{\bar{d}}^{-1}\right)^{-1}\boldsymbol{\Lambda}^{(g)H}\boldsymbol{R}_{\bar{v}}^{-1}. \tag{6.26}$$

However, $\boldsymbol{\Psi}^{(\mathrm{rx})H}$ is not necessary diagonal matrix. If $\boldsymbol{R}_{\bar{v}}$ and $\boldsymbol{R}_{\bar{d}}$ are diagonal, an exact diagonal LMMSE is obtained. For example in additive white Gaussian noise (AWGN) channel $\boldsymbol{R}_v = \sigma^2 \boldsymbol{I}_N$ with uncorrelated data $\boldsymbol{R}_d = E_s \boldsymbol{I}_N$, the LMMSE demodulator is obtained from

$$\boldsymbol{\Lambda}^{(\gamma)H} = \left(\boldsymbol{\Lambda}^{(g)H} \boldsymbol{\Lambda}^{(g)} + \frac{\sigma^2}{E_s} \boldsymbol{I}_N \right)^{-1} \boldsymbol{\Lambda}^{(g)H}. \tag{6.27}$$

$\boldsymbol{R}_{\bar{v}}$ can be non-diagonal for example after ZF channel equalization, and $\boldsymbol{R}_{\bar{d}}$ can be non-diagonal in case $\boldsymbol{d}$ is not fully allocated. To force diagonal matrix in this case, the LMMSE results is obtained under diagonal constraint as

$$\boldsymbol{\Lambda}^{(\gamma)H} = \left(\boldsymbol{\Lambda}^{(g)H} \boldsymbol{\Lambda}^{(g)} + \boldsymbol{\Lambda}_{\bar{\boldsymbol{d}}}^{-1} \boldsymbol{\Lambda}_{\bar{\boldsymbol{v}}} \right)^{-1} \boldsymbol{\Lambda}^{(g)H}, \tag{6.28}$$

where $\boldsymbol{\Lambda}_{\bar{d}} = \mathrm{diag}\left\{\boldsymbol{R}_{\bar{d}}\right\} = \mathrm{diag}\left\{\boldsymbol{U}_t \boldsymbol{R}_d \boldsymbol{U}_t^{\mathrm{H}}\right\}, \ \boldsymbol{\Lambda}_{\bar{v}} = \mathrm{diag}\left\{\boldsymbol{R}_{\bar{v}}\right\} = \mathrm{diag}\left\{\boldsymbol{V}_t^{\mathrm{H}} \boldsymbol{R}_v \boldsymbol{V}_t\right\}.$

The demodulator pulse can also be computed in the FD in similar way from the model $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{A}}\boldsymbol{d} + \tilde{\boldsymbol{v}}$ using the FD decomposition $\tilde{\boldsymbol{A}} = \boldsymbol{V}_f \boldsymbol{\Lambda}^{(\tilde{g})} \boldsymbol{U}_f$, such that

$$\boldsymbol{\Lambda}^{(\tilde{\gamma})H} = \left(\boldsymbol{\Lambda}^{(\tilde{g})H} \boldsymbol{\Lambda}^{(\tilde{g})} + \boldsymbol{\Lambda}_{\bar{\boldsymbol{d}}}^{-1} \boldsymbol{\Lambda}_{\bar{\boldsymbol{v}}} \right)^{-1} \boldsymbol{\Lambda}^{(\tilde{g})H}. \tag{6.29}$$

where, $\boldsymbol{\Lambda}_{\bar{d}} = \mathrm{diag}\left\{\boldsymbol{U}_f \boldsymbol{R}_d \boldsymbol{U}_f^{\mathrm{H}}\right\}, \ \boldsymbol{\Lambda}_{\bar{\tilde{v}}} = \mathrm{diag}\left\{\boldsymbol{V}_f^{\mathrm{H}} \boldsymbol{R}_{\tilde{v}} \boldsymbol{V}_f\right\}.$

The evaluation of these design approaches are reported in [NCF19b] and [NCF19a].

6.3.5 Alternative interpretation of GFDM

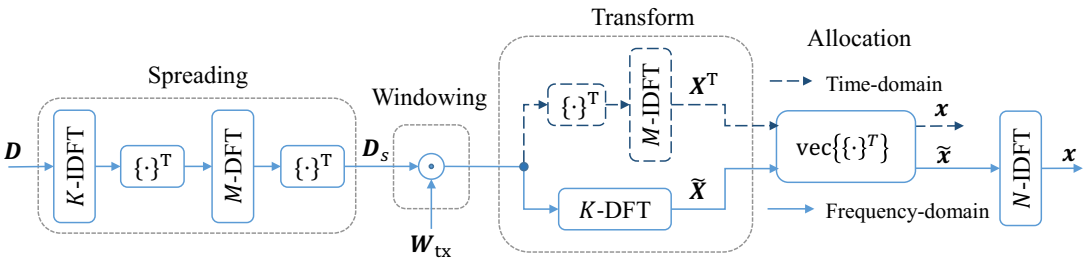

Figure 6.3: GFDM modulator

This work introduces an alternative representation of GFDM, which is reported in [NCMF18]. Based on (6.14) and (6.16) the GFDM modulation can be divided into four main modules as depicted in Figure 6.3.

Data spreading. The spreading is achieved by applying DFT on the rows and IDFT on the columns of $\boldsymbol{D}$, producing the precoded matrix

$$\boldsymbol{D}_s = \boldsymbol{F}_K^{\mathrm{H}} \boldsymbol{D} \boldsymbol{F}_M. \tag{6.30}$$

Windowing. The spread data matrix $\boldsymbol{D}_s$ is element-wise multiplied with a transmitter windowing matrix $\boldsymbol{W}_{\text{tx}} \in \mathbb{C}^{K\times M}$, which is generated based on the prototype pulse shape,

$$\boldsymbol{X}_0 = \boldsymbol{W}_{\text{tx}} \odot \boldsymbol{D}_s \in \mathbb{C}^{K\times M}. \tag{6.31}$$

The entries of $\boldsymbol{W}_{\text{tx}}$ depends on the implementation domain,

$$\boldsymbol{W}_{\text{tx}}^{(\text{TD})} = \sqrt{N}\{\boldsymbol{Z}_{M,K}^{(\boldsymbol{g})}\}^{\text{T}} = \sqrt{N}\boldsymbol{G}\boldsymbol{F}_M \quad , \boldsymbol{W}_{\text{tx}}^{(\text{FD})} = \sqrt{N}\bar{\boldsymbol{Z}}_{K,M}^{(\tilde{\boldsymbol{g}})} = \sqrt{N}\boldsymbol{F}_K^{\text{H}}\tilde{\boldsymbol{G}}^{\text{T}}. \tag{6.32}$$

Transformation. The matrix $\boldsymbol{X}_0$ can be seen as FD blocks of length M in rows, or as TD symbols of length K in columns. Thus, a final M-DFT or K-IDFT is applied to obtain the block samples, as follows

$$\boldsymbol{X}^{\text{T}} = \boldsymbol{F}_M^{\text{H}}\boldsymbol{X}_0^{\text{T}}, \quad \tilde{\boldsymbol{X}}^{\text{T}} = \boldsymbol{F}_K\boldsymbol{X}_0 \ . \tag{6.33}$$

Sample allocation. The final vector is achieved by allocating the generated samples to the corresponding indexes. Namely,

$$\boldsymbol{x} = \text{vec}\left\{\boldsymbol{X}\right\}, \quad \tilde{\boldsymbol{x}} = \text{vec}\left\{\tilde{\boldsymbol{X}}\right\}. \tag{6.34}$$

Additionally, N-IDFT is required for the FD implementation to obtain $\boldsymbol{x} = \boldsymbol{F}_N^{\text{H}}\tilde{\boldsymbol{x}}$.

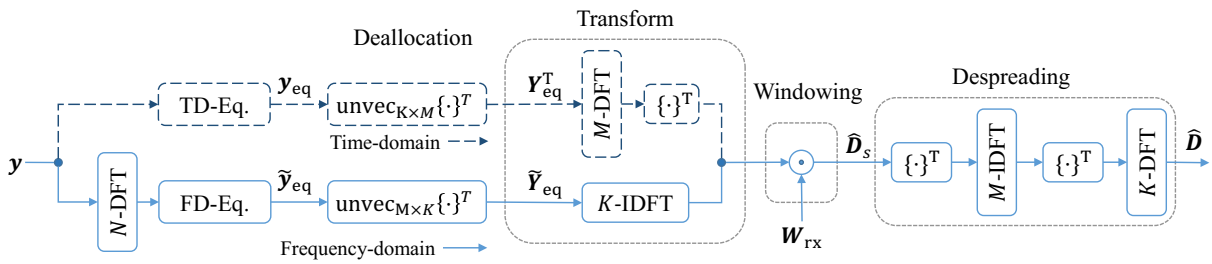

Figure 6.4: GFDM demodulator modulator

The demodulator performs the inverse steps, as illustrated in Figure 6.4. First, the matrix $\boldsymbol{Y}_0 \in \mathbb{C}^{K\times M}$ is constructed from the TD or FD equalized signal $\boldsymbol{y}_{\text{eq}}$ or $\tilde{\boldsymbol{y}}_{\text{eq}}$ as

$$\boldsymbol{Y}_{\text{eq}} = \text{unvec}_{K\times M}\left\{\boldsymbol{y}_{\text{eq}}\right\}, \quad \tilde{\boldsymbol{Y}}_{\text{eq}} = \text{unvec}_{M\times K}\left\{\tilde{\boldsymbol{y}}_{\text{eq}}\right\}. \tag{6.35}$$

Afterwards, transformation is performed as

$$\boldsymbol{Y}_0^{(\text{TD})} = \left[\boldsymbol{F}_M\boldsymbol{Y}_{\text{eq}}^{\text{T}}\right]^{\text{T}}, \quad \boldsymbol{Y}_0^{(\text{FD})} = \boldsymbol{F}_K^{\text{H}}\tilde{\boldsymbol{Y}}_{\text{eq}}^{\text{T}}. \tag{6.36}$$

Then a receive window $\boldsymbol{W}_{\text{rx}}$ is applied to $\boldsymbol{Y}$ followed by despreading, so that

$$\hat{\boldsymbol{D}} = \boldsymbol{F}_K\left(\boldsymbol{W}_{\text{rx}} \odot \boldsymbol{Y}_0\right)\boldsymbol{F}_M^{\text{H}}. \tag{6.37}$$

6.3.6 Orthogonal modulation and GFDM spreading

In order to achieve an orthogonal modulation matrix, it is necessary that $|\bar{z}_{k,m}| = |z_{m,k}| =$ const. Thus, different orthogonal waveforms can be obtained by designing the phases $\phi_{m,k} = \angle(z_{m,k})$ in the TD or $\bar{\phi}_{k,m} = \angle(\bar{z}_{k,m})$ in the FD. Two corner cases arise;

Subsymbol separation. The TD case when $\phi_{m,k} = \phi_k$ means that $\boldsymbol{W}_{\text{tx}}[k,m] = w_k$. As a result, $\boldsymbol{X}_0 = \boldsymbol{\Lambda}^{(w)}\boldsymbol{D}_s$, with $\boldsymbol{\Lambda}^{(w)} = \text{diag}\left\{\{w\}_{k=0}^{K-1}\right\}$. Therefore, $\boldsymbol{X} = \boldsymbol{X}_0\boldsymbol{F}_M^H = \boldsymbol{\Lambda}^{(w)}\boldsymbol{F}_K^H\boldsymbol{D}$. This results in a frame of M consecutive windowed OFDM symbols. In other words, $\boldsymbol{D}[:,m]$ can be seen as FD symbols mapped to the m-th OFDM symbol of length K. In this case, there is clear separation of the signals corresponding to each subsymbol.

Subcarrier separation. The other case is in the FD when $\bar{\phi}_{k,m} = \bar{\phi}_m$ and therefore, $\boldsymbol{W}_{\text{tx}}[k,m] = \tilde{w}_m$. Moreover, $\boldsymbol{X}_0 = \boldsymbol{D}_s\boldsymbol{\Lambda}^{(\tilde{w})}$, $\boldsymbol{\Lambda}^{(w)} = \text{diag}\left\{\{\tilde{w}\}_{m=0}^{M-1}\right\}$. Thereby, $\tilde{\boldsymbol{X}} = \boldsymbol{X}_0^{\text{T}}\boldsymbol{F}_K = \boldsymbol{\Lambda}^{(w)}\boldsymbol{F}_M\boldsymbol{D}^{\text{T}}$, which is equivalent to SC-FDMA. Therefore, $\boldsymbol{D}[k,:]$ can be seen as TD symbols that use the k-th subchannel of bandwidth M.

In all other cases, the subsymbols and subcarriers overlap, i.e. the data are spread over the time and frequency domains, Fig. 6.5.

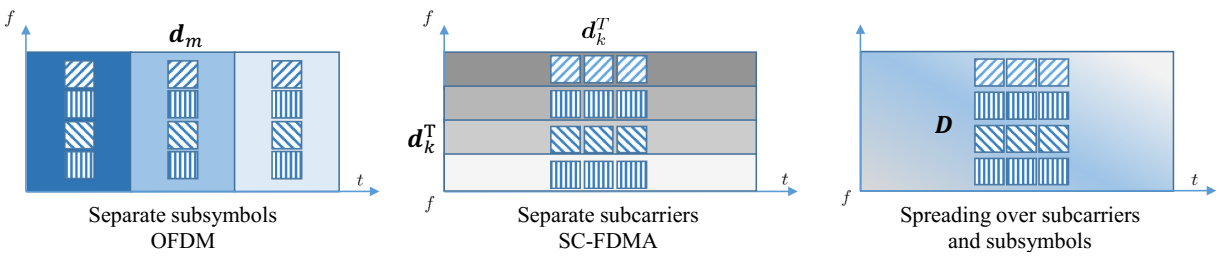

Figure 6.5: Orthogonal GFDM.

6.4 Summary

This chapter has introduced waveforms that are generated from time and frequency shifts of subcarrier of a prototype pulse, where $g_{k,m}[n] = g[\langle n - n_{k,m}\rangle_N]e^{j2\pi\frac{[n-n_{k,m}]s_{k,m}}{N}}$. A wide range of waveforms can be generated depending on the shifts $n_{k,m}$ and $s_{k,m}$. The ambiguity function $A(\Delta u, \Delta v) = \sum_{n=0}^{N-1} g[n]g^*[\langle n - \Delta u\rangle_N]e^{-j2\pi\frac{[n-\Delta u]\Delta v}{N}}$ gives information about the orthogonality of the shifted pulses and indicates the time and frequency spreading. A special case belongs to Gabor frame, where $n_{k,m} = mP$ and $s_{k,m} = kQ$ with $Q \leq M$ and $P \leq K$. GFDM modulation with $P = K$ and $Q = M$ is a special case of Gabor frame. It is shown that Gabor frame can be implemented with multiple GFDM modulators using different prototype pulses. A generalization of GFDM with L prototype pulses is given

by $x[n] = \sum_{l=0}^{L-1} x^{(l)}[n] = \sum_{k\in\mathcal{K}_{on}} \sum_{m\in\mathcal{M}_{on}{}^{(l)}} d_{k,m} g^{(l)}[\langle n - mK\rangle_N] e^{j2\pi\frac{nk}{K}}$, where the active subcarrier and subsymbol sets are used to control the waveform. Thus, GFDM plays an important role in the realization of flexible waveforms. This flexibility is achieved by means of changing the number of subcarriers K, number of subsymbols M, prototype pulse $\boldsymbol{g}$, in addition to the sets of active subcarriers and subsymbols. Among the different representation of GFDM, the 2-D block gives insight on the time-frequency structure. Given that $\boldsymbol{G} = \text{unvec}_{K\times M}\{\boldsymbol{g}\}$, $\tilde{\boldsymbol{G}} = \text{unvec}_{M\times K}\{\tilde{\boldsymbol{g}}\}$, $\tilde{\boldsymbol{g}} = \boldsymbol{F}_N \boldsymbol{g}$, the modem operation can be performed in four steps as summarized in Table 6.2.

Table 6.2: The four steps of conventional GFDM modulator.

	Modulation		Demodulation	
	TD	FD	TD	FD
Spread.	$\boldsymbol{D}_s = \boldsymbol{F}_K^{\text{H}} \boldsymbol{D} \boldsymbol{F}_M$		$\hat{\boldsymbol{D}} = \boldsymbol{F}_K \hat{\boldsymbol{D}}_s \boldsymbol{F}_M^{\text{H}}$	
Wind.	$\boldsymbol{W}_{\text{tx}} = \sqrt{N}\boldsymbol{G}\boldsymbol{F}_M$	$\boldsymbol{W}_{\text{tx}} = \sqrt{N}\boldsymbol{F}_K^{\text{H}}\tilde{\boldsymbol{G}}^{\text{T}}$	$\boldsymbol{W}_{\text{rx}}$ generated from $\boldsymbol{W}_{\text{tx}}$	
	$\boldsymbol{X}_0 = \boldsymbol{W}_{\text{tx}} \odot \boldsymbol{D}_s$		$\hat{\boldsymbol{D}}_s = \boldsymbol{W}_{\text{rx}} \odot \boldsymbol{Y}_0$	
Trans.	$\boldsymbol{X}^{\text{T}} = \boldsymbol{F}_M^{\text{H}} \boldsymbol{X}_0^{\text{T}}$	$\tilde{\boldsymbol{X}}^{\text{T}} = \boldsymbol{F}_K \boldsymbol{X}_0$	$\boldsymbol{Y}_0^{\text{T}} = \boldsymbol{F}_M \boldsymbol{Y}_{\text{eq}}{}^{\text{T}}$	$\boldsymbol{Y}_0 = \boldsymbol{F}_K^{\text{H}} \tilde{\boldsymbol{Y}}_{\text{eq}}^{\text{T}}$
Alloc.	$\boldsymbol{x} = \text{vec}\{\boldsymbol{X}\}$	$\tilde{\boldsymbol{x}} = \text{vec}\left\{\tilde{\boldsymbol{X}}\right\}$	$\text{unvec}_{K\times M}\{\boldsymbol{y}_{\text{eq}}\}$	$\text{unvec}_{M\times K}\{\tilde{\boldsymbol{y}}_{\text{eq}}\}$

Orthogonal GFDM modulation is achieved when the elements of $\boldsymbol{W}_{\text{tx}}$ are of equal amplitude, whereas the phases determines the time and frequency spreading.

Inspired from the GFDM structure, the next chapter will present generic low-complexity implementation architectures, where the DFT blocks are replaced by generic unitary transforms. One particular case will be the extended GFDM framework, which enhances the conventional GFDM with additional flexibility.

Chapter 7

Modulation Framework: Architectures and Applications

Digital signal processing algorithms can be implemented by means of software and hardware approaches. On the one hand, software implementation e.g. using digital signal processor (DSP), provides full flexibility, but the throughput and latency depends on the complexity in terms of number of operations, as well as the processor computation power. On the other hand, hardware implementation, e.g. using application-specific integrated circuit (ASIC), allows for a high throughput and low latency by parallelization, but it lacks the support of high flexibility because of the limitation of hardware resources. A compromise solution is based on hardware and software co-design. In this approach, the implementation of certain functions can be performed by hardware while others can be executed by a processor. A flexible hardware and software architecture can be realized by means of reprogrammed logic such as field programmable gate array (FPGA).

This chapter focuses on the implementation of core modem functions, with the scope of providing programmable hardware architectures that allow a wide range of waveform processing. In particular, Section 7.1 presents two generic low-complexity architectures inspired from the conventional generalized frequency division multiplexing (GFDM) system, which has been presented in Section 6.3; the first is based on unitary transforms, and the other is inspired from the circular convolution. Additionally, an extension of the architecture is presented for multiple-input, multiple-output (MIMO) waveforms. As a special case of the generic model, an extended GFDM framework is presented in details in Section 7.2. This framework adds more flexibility on top of the conventional GFDM framework by means of bypass functions and memory indexing. A numerical analysis is performed for realistic FPGA implementation. This flexibility enables the realization of new waveforms. Section 7.3 presents applications of the developed framework including the realization of generalized frequency-division multiple access (FDMA) and enhancement of existing orthogonal frequency division multiplexing (OFDM) systems. The results of this chapter are reported in [NCF18], [LNF19] and [DNM$^+$18].

7.1 Modem architectures

A generic core block modulator targets the realization of the linear transform $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d}$, where $\boldsymbol{A} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{d} \in \mathbb{C}^{N\times 1}$. The demodulator also performs linear transform using demodulator matrix $\boldsymbol{B}$ such that $\hat{\boldsymbol{d}} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{y}_{\mathrm{eq}}$, where $\boldsymbol{y}_{\mathrm{eq}} \in \mathbb{C}^{N\times 1}$ is the received block after channel equalization. Without the consideration of any special structure of $\boldsymbol{A}$, the modulation requires N^2 complex multiplications (CMs), and a storage of N^2 complex coefficients to store $\boldsymbol{A}$. If one multiplication is performed in each clock cycle, the latency of the modulation is N^2 clock cycles, and a constant throughput can be achieved with proper pipelining. The latency can be scaled down by increasing the number of multipliers and parallel memory access. However, this scaling is limited by the available hardware resources and it requires more control overheads to manage the parallel processing. By exploiting a special structure of the modulation matrix, the complexity can be significantly reduced. For example, OFDM employs $\boldsymbol{F}_N^{\mathrm{H}}$ as a modulation matrix, which can be realized with a complexity of $\frac{N}{2}\log_2(N)$ CMs and a memory to store N complex values. Nevertheless, this complexity can only be achieved for radix-2 parameters, i.e. $N = 2^x$, which limits the flexibility of radix-2 fast Fourier transform (FFT) function. Moreover, any sparsity in $\boldsymbol{A}$ can be exploited. For instance, if $\boldsymbol{A}$ is block diagonal with K matrices, each of size $M \times M$ and $N = KM$, the modulation complexity can be reduced to KM^2.

7.1.1 General modulation matrix structure

In order to define the modulation architecture, this thesis proposes the term modulation based on the singular value decomposition (SVD) composition of the modulation matrix. The SVD decomposition of the time-domain (TD) matrix can be expressed as

$$\boldsymbol{A} = \boldsymbol{V}_t\boldsymbol{\Lambda}^{(\mathrm{tx})}\boldsymbol{U}_t, \tag{7.1}$$

where $\boldsymbol{V}_t, \boldsymbol{U}_t \in \mathbb{C}^{N\times N}$ are unitary matrices, whereas $\boldsymbol{\Lambda}^{(\mathrm{tx})} \in \mathbb{C}^{N\times N}$ is diagonal. By fixing $\boldsymbol{U}_t$ and $\boldsymbol{V}_t$ to a certain structure, and varying $\boldsymbol{\Lambda}^{(\mathrm{tx})}$, a class of modulation matrices is obtained. For example, the GFDM modulation is achieved using the structure in Section 6.3.3.

The demodulation matrix follows the same structure, i.e. $\boldsymbol{B} = \boldsymbol{V}_t\boldsymbol{\Lambda}^{(\mathrm{rx})}\boldsymbol{U}_t$. The demodulator design requires the computation of $\boldsymbol{\Lambda}^{(\mathrm{rx})}$. For instance, by following the approach presented in Section 6.3.4. Note that, even though this design has been discussed in the context of GFDM, it is valid for the general case as well.

The frequency-domain (FD) modulation matrix $\tilde{\boldsymbol{A}} = \boldsymbol{F}_N\boldsymbol{A}$ can be expressed in the form

$$\tilde{\boldsymbol{A}} = \boldsymbol{V}_f\tilde{\boldsymbol{\Lambda}}^{(\mathrm{tx})}\boldsymbol{U}_f. \tag{7.2}$$

Because the normalized discrete Fourier transform (DFT) matrix $\boldsymbol{F}_N$ is unitary, the singular values of $\tilde{\boldsymbol{A}}$ and $\boldsymbol{A}$ are the same. Therefore, $\tilde{\boldsymbol{\Lambda}}^{(\mathrm{tx})}\tilde{\boldsymbol{\Lambda}}^{\mathrm{H}(\mathrm{tx})} = \boldsymbol{\Pi}_{\mathrm{TD}}\boldsymbol{\Lambda}^{(\mathrm{tx})}\boldsymbol{\Lambda}^{\mathrm{H}(\mathrm{tx})}\boldsymbol{\Pi}_{\mathrm{TD}}^{\mathrm{T}}$,

where $\boldsymbol{\Pi}_{\mathrm{TD}}$ is a permutation matrix. An obvious choice of the unitary matrices in the FD is $\boldsymbol{U}_f = \boldsymbol{U}_t$ and $\boldsymbol{V}_f = \boldsymbol{F}_N \boldsymbol{V}_t$, and thus $\tilde{\boldsymbol{\Lambda}}^{(\mathrm{tx})} = \boldsymbol{\Lambda}^{(\mathrm{tx})}$. However, considering other decomposition is significant, e.g. to provide similar TD and FD implementation as in the case of GFDM, Section 6.3.3.

7.1.2 Run-time flexibility

The run-time flexibility can be achieved by changing a set of parameters without changing the whole architecture. For a certain modulation class, the unitary transform structure is preserved whereas the entries of $\boldsymbol{\Lambda}^{(\mathrm{tx})}$ and the block length N can be used as the modulation parameters. The allowed values of N depends on the unitary transforms. For instance, in the case of single-carrier (SC) with circular convolution

$$\boldsymbol{A}_c = \frac{1}{N} \boldsymbol{F}_N^{\mathrm{H}} \boldsymbol{\Lambda}^{(\mathrm{tx})} \boldsymbol{F}_N, \tag{7.3}$$

the flexibility of the transform is achieved by changing the DFT size, which, in general, allows $N_{\max}$ options, where $N_{\max}$ is the maximum supported size. If the DFT is implemented by radix-2 FFT, only $\log_2(N_{\max})$ options are possible. In GFDM,

$$\boldsymbol{A} = \underbrace{\boldsymbol{\Pi}_{M,K} \boldsymbol{U}_{K,M}^{\mathrm{H}}}_{\boldsymbol{V}_t} \boldsymbol{\Lambda}^{(g)} \underbrace{\boldsymbol{U}_{K,M} \boldsymbol{\Pi}_{K,M} \boldsymbol{U}_{M,K}^{\mathrm{H}}}_{\boldsymbol{U}_t}, \tag{7.4}$$

$\boldsymbol{U}_{M,K} = \boldsymbol{I}_M \otimes \boldsymbol{F}_K$, $\boldsymbol{U}_{K,M} = \boldsymbol{I}_K \otimes \boldsymbol{F}_M$, and $\boldsymbol{\Pi}_{K,M}$ is a matrix equivalent to the transform $\boldsymbol{x}_o = \boldsymbol{\Pi}_{K,M} \boldsymbol{x}_i$ with $\boldsymbol{x}_o[m + kM] = \boldsymbol{x}_i[k + mK]$. This structure allows changing N as well as changing K, M, where $N = KM$. The SC case is a special case achieved with $M = 1$, which makes GFDM a more general class of modulation.

Bypass flexibility. Another type of flexibility arises from the number of different unitary transforms that a structure allows. This can be achieved by bypassing the function corresponding to some of the transforms. For instance, let $\boldsymbol{U}_t = \boldsymbol{U}^{(1)} \boldsymbol{U}^{(2)} \cdots \boldsymbol{U}^{(L_u)}$ and $\boldsymbol{V}_t = \boldsymbol{V}^{(1)} \boldsymbol{V}^{(2)} \cdots \boldsymbol{V}^{(L_v)}$, where $\boldsymbol{U}^{(l)}$ and $\boldsymbol{V}^{(l)}$ are unitary transforms, there are $2^{(L_u+L_v)}$ additional options results from controlling the active transforms. Accordingly, GFDM enables the implementation of 32 modulation options as it is composed of 5 unitary matrices.

Transform flexibility. This type of flexibility is relevant when it is possible to change the transforms in run-time, e.g. switching between DFT and inverse discrete Fourier transform (IDFT) or by customizing the permutation matrices.

Architecture flexibility. Further degree of freedom is related to the flexibility of the architecture design, where certain functions can be completely replaced. This type can also be realized on the run-time, at the cost of using more resources, which implies larger software code, or more logic consumption in the case of hardware implementation.

7.1.3 Generic GFDM-based architecture

In this thesis, the wide range of options motivate the implementation of the GFDM as a flexible framework, as reported in [NCF18]. Inspired from this architecture, this section aims at presenting a more generic architecture, where GFDM becomes a special case. The 2-D GFDM modulation block presented in Section 6.3.2, can be rewritten in the form

$$\begin{aligned} \boldsymbol{X}^{\mathrm{T}} &= \sqrt{N}\boldsymbol{F}_M^{\mathrm{H}}\left([\boldsymbol{F}_M\boldsymbol{G}^{\mathrm{T}}]\odot[\boldsymbol{F}_M(\boldsymbol{F}_K^{\mathrm{H}}\boldsymbol{D})^{\mathrm{T}}]\right) \\ \tilde{\boldsymbol{X}}^{\mathrm{T}} &= \sqrt{N}\boldsymbol{F}_K\left([\boldsymbol{F}_K^{\mathrm{H}}\tilde{\boldsymbol{G}}^{\mathrm{T}}]\odot[\boldsymbol{F}_K^{\mathrm{H}}(\boldsymbol{F}_M\boldsymbol{D}^{\mathrm{T}})^{\mathrm{T}}]\right). \end{aligned} \tag{7.5}$$

The vector blocks are given by $\boldsymbol{x} = \mathrm{vec}\{\boldsymbol{X}\}$ and $\tilde{\boldsymbol{x}} = \mathrm{vec}\left\{\tilde{\boldsymbol{X}}\right\}$. A general model to describe both GFDM TD and FD equations can be written as

$$\boldsymbol{X}_0^{\mathrm{T}} = \boldsymbol{U}_3\left(\boldsymbol{W}_{\mathrm{tx}}\odot[\boldsymbol{U}_2(\boldsymbol{U}_1\boldsymbol{D}_0)^{\mathrm{T}}]\right),\ \boldsymbol{x}_0 = \mathrm{vec}\{\boldsymbol{X}_0\}. \tag{7.6}$$

In addition, for TD to FD conversion an additional transform is required to perform $\tilde{\boldsymbol{x}}_0 = \boldsymbol{U}_4\boldsymbol{x}_0$, when necessary. Each matrix $\boldsymbol{U}_x$ is unitary of size N_x, with $N_2 = N_3$. The matrix $\boldsymbol{W}_{\mathrm{tx}}$ of size $N_2 \times N_1$ is the modulator window matrix, $\boldsymbol{D}_0$ represents the data matrix of size $N_1 \times N_2$, $\boldsymbol{X}_0$ is the block matrix of size $N_1 \times N_2$ and $\boldsymbol{x}_0$ is the vector of size $N_1N_2 \times 1$. The overall modulation considering generic unitary matrices can be decomposed similar to the GFDM matrix, as discussed in Section 6.3.3, but with different unitary matrices. The elements of the window corresponds to the singular values. Namely,

$$\boldsymbol{A} = \underbrace{\boldsymbol{\Pi}_{N_2,N_1}\boldsymbol{U}^{(3)}_{N_1,N_2}}_{\boldsymbol{V}_t}\boldsymbol{\Lambda}^{(\mathrm{tx})}\underbrace{\boldsymbol{U}^{(2)}_{N_1,N_2}\boldsymbol{\Pi}_{N_1,N_2}\boldsymbol{U}^{(1)}_{N_2,N_1}}_{\boldsymbol{U}_t}. \tag{7.7}$$

where $\boldsymbol{U}^{(3)}_{N_1,N_2} = \boldsymbol{I}_{N_1}\otimes\boldsymbol{U}_3$, $\boldsymbol{U}^{(2)}_{N_1,N_2} = \boldsymbol{I}_{N_1}\otimes\boldsymbol{U}_2$, and $\boldsymbol{U}^{(1)}_{N_2,N_1} = \boldsymbol{I}_{N_2}\otimes\boldsymbol{U}_1$ are unitary matrices, and $\boldsymbol{\Lambda}^{(\mathrm{tx})} = \mathrm{diag}\{\mathrm{vec}\{\boldsymbol{W}_{\mathrm{tx}}\}\}$. The demodulator performs the inverse operations on the equalized block $\boldsymbol{y}_0$ using the receiver window $\boldsymbol{W}_{\mathrm{rx}}$. When $\boldsymbol{U}_4$ is enabled, $\boldsymbol{y}_0 = \boldsymbol{U}_4^{\mathrm{H}}\tilde{\boldsymbol{y}}_0$. First, the vector is reshaped in the matrix $\boldsymbol{Y}_0 = \mathrm{unvec}_{N_1\times N_2}\{\boldsymbol{y}_0\}$, then the demodulated data is computed as $\hat{\boldsymbol{D}}_0 = \boldsymbol{U}_1^{\mathrm{H}}\hat{\tilde{\boldsymbol{D}}}_0^{\mathrm{T}}$, where

$$\hat{\tilde{\boldsymbol{D}}}_0 = \left[\boldsymbol{U}_1\hat{\boldsymbol{D}}_0\right]^{\mathrm{T}} = \boldsymbol{U}_2^{\mathrm{H}}\left(\boldsymbol{W}_{\mathrm{rx}}\odot\left[\boldsymbol{U}_3^{\mathrm{H}}\boldsymbol{Y}_0^{\mathrm{T}}\right]\right). \tag{7.8}$$

Figure 7.1 illustrates the modem architecture, which includes four unitary transforms, two blocks to perform transpose and vectorization, and one element-wise multiplication. All matrix operation are performed column-wise. The transpose requires storing the transformation of all columns before proceeding to the next transform. The transpose is equivalent to writing a matrix in columns and reading in rows, thus, it involves memory indexing. The modulation and demodulation window are constant matrices that can be recalculated and stored. The unitary transform are generic and can be of any type. The run-time flexibility of this architecture follows the options presented in Section 7.1.2. In particular, while the permutation is performed by memory read and write, it can

be flexibly adjusted without changing the architecture. Accordingly, a further flexible modulation can be expressed using general flexible permutation $\boldsymbol{P}$ and $\boldsymbol{\Pi}$ as

$$\boldsymbol{A}_0 = \underbrace{\boldsymbol{U}_4 \boldsymbol{P} \boldsymbol{U}^{(3)}_{N_1,N_2}}_{\boldsymbol{V}_t} \boldsymbol{\Lambda}^{(\mathrm{tx})} \underbrace{\boldsymbol{U}^{(2)}_{N_1,N_2} \boldsymbol{\Pi} \boldsymbol{U}^{(1)}_{N_2,N_1}}_{\boldsymbol{U}_t}. \tag{7.9}$$

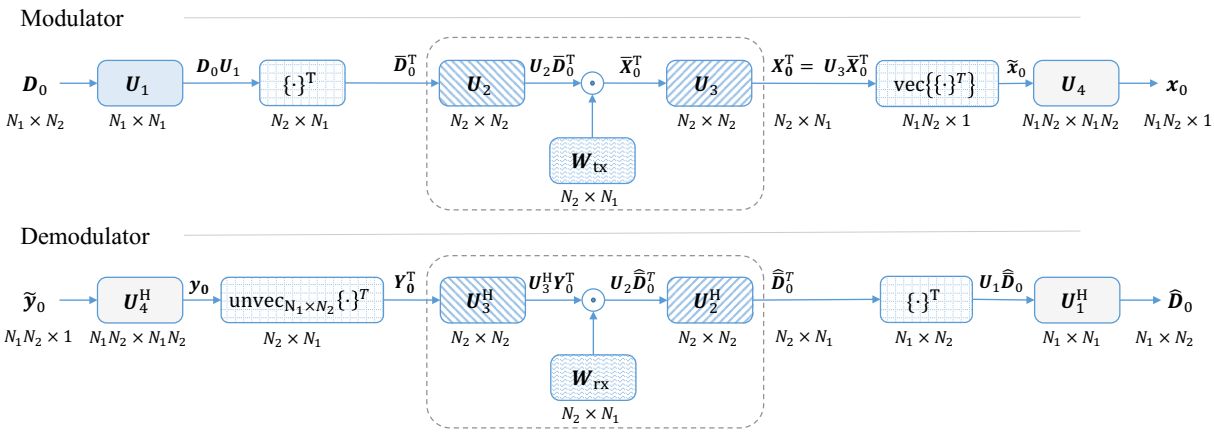

Figure 7.1: Generic GFDM-based modem structure.

7.1.4 Flexible parallel multiplications architecture

The highlighted transforms in Figure 7.1 perform the operations,

$$\boldsymbol{X}_0^{\mathrm{T}} = \boldsymbol{U}_3 \left([\boldsymbol{U}_2 \bar{\boldsymbol{D}}_0^{\mathrm{T}}] \odot \boldsymbol{W}_{\mathrm{tx}} \right), \ \hat{\bar{\boldsymbol{D}}}_0^{\mathrm{T}} = \boldsymbol{U}_2^{\mathrm{H}} \left([\boldsymbol{U}_3 \boldsymbol{Y}_0^{\mathrm{T}}] \odot \boldsymbol{W}_{\mathrm{rx}} \right), \text{where } \bar{\boldsymbol{D}}_0 = \boldsymbol{U}_1 \boldsymbol{D}_0. \tag{7.10}$$

Note that for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{N\times 1}$ and a unitary transform defined by $\boldsymbol{U} \in \mathbb{C}^{N\times N}$

$$\bar{\boldsymbol{y}} = \boldsymbol{U}(\boldsymbol{h} \odot \boldsymbol{x}) = \boldsymbol{U}\mathrm{diag}\{\boldsymbol{h}\}\boldsymbol{x} = \underbrace{\left(\boldsymbol{U}\mathrm{diag}\{\boldsymbol{h}\}\boldsymbol{U}^{\mathrm{H}}\right)}_{\boldsymbol{H}} (\boldsymbol{U}\boldsymbol{x}).$$

Therefore, for $\boldsymbol{U}_2^{\mathrm{H}} = \boldsymbol{U}_3$, $\boldsymbol{X}_0^{\mathrm{T}}[:, n_1] = \underbrace{\left(\boldsymbol{U}_3\mathrm{diag}\{\boldsymbol{W}_{\mathrm{tx}}[:, n_1]\}\boldsymbol{U}_3^{\mathrm{H}}\right)}_{\boldsymbol{C}_0^{(n_1)}} \bar{\boldsymbol{D}}_0^{\mathrm{T}}[:, n_1]$, and thus,

$$\boldsymbol{X}_0[n_1, n_2] = \sum_{m_2=0}^{N_2-1} \bar{\boldsymbol{D}}_0[n_1, m_2]\boldsymbol{C}_0^{(n_1)}[n_2, m_2] = \sum_{m_2=0}^{N_2-1} \bar{\boldsymbol{D}}_0^{(m_2)}[n_1, n_2]\boldsymbol{G}_0^{(m_2)}[n_1, n_2]. \tag{7.11}$$

Here, $\boldsymbol{G}_0^{(m_2)}[n_1, n_2] = \boldsymbol{C}_0^{(n_1)}[n_2, m_2]$, and $\bar{\boldsymbol{D}}_0^{(m_2)}[n_1, n_2] = \bar{\boldsymbol{D}}_0[n_1, m_2]$. This representation allows parallel computation as illustrated in Figure 7.2. The architecture consists of N_2 memory blocks to store the filter matrices $\boldsymbol{G}^{(m_2)} \in \mathbb{C}^{N_1\times N_2}$ and other N_2 memory blocks to store $\bar{\boldsymbol{D}}_0^{(m_2)} \in \mathbb{C}^{N_1\times N_2}$. The data memory writing is customized with the indexing unit. These memories are read sequentially and fed to N_2 parallel complex multipliers and accumulation. In the case of GFDM with DFT/IDFT transforms, $\boldsymbol{C}_0^{(n_1)}$ is a circular matrix defined by $\boldsymbol{C}_0^{(n_1)}[n_2, m_2] = \sqrt{N_1}\boldsymbol{G}_0[n_1, \langle n_2 - m_2\rangle_{N_2}]$ and therefore, $\boldsymbol{G}_0^{(m_2)}[n_1, n_2] = \boldsymbol{G}_0[n_1, \langle n_2 - m_2\rangle_{N_2}]$. Here, $\boldsymbol{G}_0 = \mathrm{unvec}_{N_1\times N_2}\{\boldsymbol{g}_0\}$ corresponds to the prototype pulse.

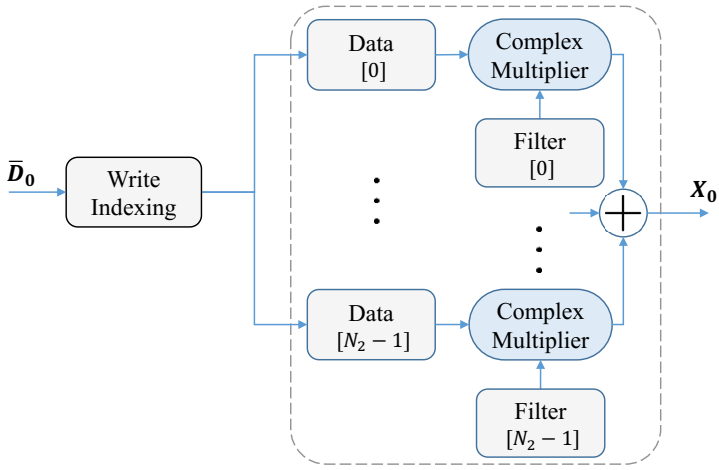

Figure 7.2: Parallel convolution architecture.

Extended flexibility. Any generic form of $\boldsymbol{C}_0^{(n_1)} = \boldsymbol{U}_3^{(n_1)} \mathrm{diag}\left\{\boldsymbol{W}_{\mathrm{tx}}[:, n_1]\right\} \boldsymbol{U}_3^{(n_1)H}$ can be used, where the unitary matrices change with n_1 as well. In all these cases $N_2^2 N_1$ coefficients need to be stored. The overall modulation matrix still have the same singular values as in (7.7),

$$\boldsymbol{A} = \underbrace{\boldsymbol{\Pi}_{N_2,N_1} \boldsymbol{U}^{(3)}_{N_1,N_2}}_{\boldsymbol{V}_t} \boldsymbol{\Lambda}^{(\mathrm{tx})} \underbrace{\boldsymbol{U}^{(2)}_{N_1,N_2} \boldsymbol{\Pi}_{N_1,N_2} \boldsymbol{U}^{(1)}_{N_2,N_1}}_{\boldsymbol{U}_t},$$

but with different orthogonal transforms, where $\boldsymbol{U}^{(3)}_{N_1,N_2}$ is block diagonal with the matrices $\{\boldsymbol{U}_3^{(n_1)}\}$ and $\boldsymbol{U}^{(2)}_{N_1,N_2} = \boldsymbol{U}^{(3)H}_{N_1,N_2}$.

Realization of GFDM with sparse pulses.

In the case of sparse pulse matrix, where $\boldsymbol{G}_0$ contains only few non-zero columns, such that $\boldsymbol{G}_0[:, n_2] = \boldsymbol{0}_{N_1}, n_2 \notin \mathcal{N}_2$. The convolution can be expressed as

$$\begin{aligned}\boldsymbol{X}_0[n_1, n_2] &= \sum_{m_2=0}^{N_2-1} \bar{\boldsymbol{D}}_0[n_1, m_2] \boldsymbol{G}_0[n_1, \langle n_2 - m_2 \rangle_{N_2}] = \sum_{m_2=0}^{N_2-1} \bar{\boldsymbol{D}}_0[n_1, \langle n_2 - m_2 \rangle_{N_2}] \boldsymbol{G}_0[n_1, m_2] \\ &= \sum_{m_2 \in \mathcal{N}_2} \bar{\boldsymbol{D}}_0^{(m_2)}[n_1, n_2] \boldsymbol{G}_0^{(m_2)}[n_1, n_2].\end{aligned}$$

To realize this, using the flexible indexing, the data and filter memories are set to $\bar{\boldsymbol{D}}_0^{(m_2)}$, and $\boldsymbol{G}_0^{(m_2)}$, respectively, where

$$\bar{\boldsymbol{D}}_0^{(m_2)}[n_1, n_2] = \bar{\boldsymbol{D}}_0[n_1, \langle n_2 - m_2 \rangle_{N_2}], \ \boldsymbol{G}_0^{(m_2)}[n_1, n_2] = \boldsymbol{G}_0[n_1, m_2].$$

Here, only the required memories and multipliers need to be activated.

7.1.5 MIMO waveform architecture

The general MIMO waveform is presented in Section 3.3.3, $\boldsymbol{x}_{n_t} = \sum_{m_t=0}^{N_T-1} \boldsymbol{A}_{n_t,m_t} \boldsymbol{d}_{n_t,m_t}$. It requires using N_{t}^2 independent modulators. The conventional MIMO systems, where the

transmitted signal on each antenna is generated independently $\boldsymbol{x}_{n_t} = \boldsymbol{A}_{n_t}\boldsymbol{d}_{n_t}$, is a special case that requires N_t independent modems. A compromise of both cases is achieved when $\boldsymbol{d}_{n_t,m_t} = \boldsymbol{d}_{m_t}$, i.e. the active data symbols $\boldsymbol{d}^{\mathrm{on}}_{m_t}$ are allocated to all transmit antennas similarly. Considering all the involved modulation matrices are from the same modulation class, where $\boldsymbol{A}_{n_t,m_t} = \boldsymbol{V}_{n_t}\boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}\boldsymbol{U}_{m_t}$, then

$$\boldsymbol{x}_{n_t} = \boldsymbol{V}_{n_t}\sum_{m_t=0}^{N_T-1}\boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}\bar{\boldsymbol{d}}_{m_t},\ \bar{\boldsymbol{d}}_{m_t} = \boldsymbol{U}_{m_t}\boldsymbol{d}_{m_t}. \tag{7.12}$$

To implement this, first $\bar{\boldsymbol{d}}_{m_t}$ is computed and then forwarded to N_T parallel multipliers to generate the vectors $\boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}\bar{\boldsymbol{d}}_{m_t}$. The values are stored in N_T memories with respect to the index n_t, i.e. $\bar{\boldsymbol{x}}_{n_t} = \bar{\boldsymbol{x}}_{n_t} + \boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}\bar{\boldsymbol{d}}_{m_t}$. After that, the transform $\boldsymbol{x}_{n_t} = \boldsymbol{V}_m\bar{\boldsymbol{x}}_{n_t}$ is computed. Using the unitary matrices according to (7.9)

$$\boldsymbol{U}_{m_t} = \boldsymbol{U}^{(2)}_{N_1,N_2}\boldsymbol{\Pi}\boldsymbol{U}^{(1)}_{N_2,N_1}, \boldsymbol{V}_{n_t} = \boldsymbol{U}_4\boldsymbol{P}\boldsymbol{U}^{(3)}_{N_1,N_2},$$

the modem architecture in Section 7.1.3 can be modified by adding additional complex multipliers and memories to allow the implementation of the MIMO waveform, as shown in Figure 7.3. The data are fed to the spreading unit, which performs the

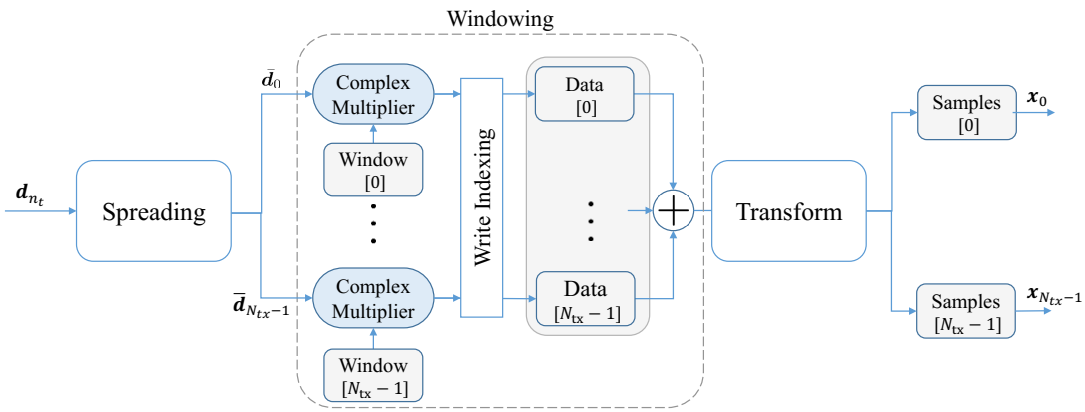

Figure 7.3: Flexible MIMO architecture.

transforms with $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ on the input $\boldsymbol{d}_{m_t}$ and the result is $\bar{\boldsymbol{d}}_{m_t} = \boldsymbol{U}_{m_t}\boldsymbol{d}_{m_t}$. This is then forwarded to N_T parallel multipliers to perform the windowing. The modulation windows are stored in N_T memory blocks, and each memory contains N_TN coefficients mapped to $\boldsymbol{W}_{n_t} \in \mathbb{C}^{N\times N_T}$, where $\boldsymbol{W}_{n_t}[n,m_t] = \boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}[n,n]$. The intermediate results $\boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}\bar{\boldsymbol{d}}_{m_t}$ are stored in N_T block memories corresponding to $\bar{\boldsymbol{D}}_{m_t} \in \mathbb{C}^{N\times N_T}$, where $\bar{\boldsymbol{D}}_{\alpha(n_t)}[n,\beta(m_t)] = \boldsymbol{W}_{n_t}[n,m_t]\bar{\boldsymbol{d}}_{m_t}[n]$. The indexing $(m_t,n_t) \to (\alpha(n_t),\beta(m_t))$ is applied to enable parallel writing, and to allow parallel reading such that $\bar{\boldsymbol{x}}_{n_t}[n] = \sum_{m_t=0}^{N_T-1}\bar{\boldsymbol{D}}_{m_t}[n,n_t]$. These memories introduce latency N_TN cycles, which is required to collect all the spread samples. Finally, $\bar{\boldsymbol{x}}_{n_t}$ are transformed with $\boldsymbol{V}_{n_t}$ and stored in a memory corresponding to $\boldsymbol{x}_{n_t} = \boldsymbol{V}_{n_t}\bar{\boldsymbol{x}}_{n_t}$ prior to transmission, which can take place after finishing all the N_T vectors. This memories also induces N_TN delay cycles.

The proposed architecture allows the implementation of different MIMO schemes via the programming of the indexing and the window memory. For instance, in spacial diversity, the data vector $\boldsymbol{d}_0$ is the input to be transmitted through different chains, such that $\boldsymbol{x}_{n_t} = \boldsymbol{V}_{n_t}\boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,0}\bar{\boldsymbol{d}}_0$. In this case, the data memory can be written as $\bar{\boldsymbol{D}}_{n_t}[n, n_t] = \boldsymbol{W}_{n_t}[n, 0]\bar{\boldsymbol{d}}_0[n]$, and the read is triggered after N samples.

Matrix representation. The MIMO vector $\boldsymbol{x} \in \mathbb{C}^{N_T N \times 1}$ can be expressed as

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{\Pi}^{\mathrm{T}}_{N_T}\boldsymbol{\Psi}^{(\mathrm{tx})}\boldsymbol{\Pi}_{N_T}\boldsymbol{U}\boldsymbol{d} = \boldsymbol{A}\boldsymbol{d}. \tag{7.13}$$

where $\boldsymbol{V} = \boldsymbol{I}_{N_T} \otimes \left(\boldsymbol{U}_4\boldsymbol{P}\boldsymbol{U}^{(3)}_{N_1,N_2}\right)$, $\boldsymbol{U} = \boldsymbol{I}_{N_T} \otimes \left(\boldsymbol{U}^{(2)}_{N_1,N_2}\boldsymbol{\Pi}\boldsymbol{U}^{(1)}_{N_2,N_1}\right)$, $\boldsymbol{\Pi}_{N_T}$ is the cumulative matrix defined such that $\mathrm{vec}\left\{\{\mathrm{unvec}_{N\times N_T}\{\boldsymbol{x}\}\}^{\mathrm{T}}\right\} = \boldsymbol{\Pi}_{N_T}\boldsymbol{x}$, $\boldsymbol{\Psi}^{(\mathrm{tx})} = \mathrm{Bd}\{\boldsymbol{\Psi}^{(\mathrm{tx})}_0, \cdots \boldsymbol{\Psi}^{(\mathrm{tx})}_{N-1}\}$ is a block diagonal matrix with N diagonal matrices, where $\boldsymbol{\Psi}^{(\mathrm{tx})}_n[n_t, m_t] = \boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}[n, n]$. The demodulator matrix $\boldsymbol{B}$ has similar structure when $N_R = N_T$, and it can be determined from $\boldsymbol{\Psi}_{(\mathrm{tx})}$. Thus, the demodulator is implemented in similar way. For instance if $\boldsymbol{\Psi}_{(\mathrm{tx})}$ is Hermitian, then $\boldsymbol{\Lambda}^{(\mathrm{rx})}_{n_t,m_t} = \boldsymbol{\Lambda}^{(\mathrm{tx})}_{n_t,m_t}$.

For other MIMO combination, the receiver matrix structure is adjusted according to the equivalent modulator matrix. For example, in multiple-input single-output (MISO) the demodulator matrix is equivalent to single-input, single-output (SISO).

7.2 Extended GFDM framework

For conventional GFDM, the unitary transforms are performed with DFT/IDFT matrices, as listed in Table 7.1. The TD modulation window is generated from the prototype pulse defined by the vector $\boldsymbol{g} \in \mathbb{C}^{N\times 1}$, where $\boldsymbol{G} = \mathrm{unvec}_{K\times M}\{\boldsymbol{g}\}$, and the FD modulation uses $\tilde{\boldsymbol{G}} = \mathrm{unvec}_{M\times K}\{\tilde{\boldsymbol{g}}\}$, $\tilde{\boldsymbol{g}} = \boldsymbol{F}_N\boldsymbol{g}$.

Table 7.1: GFDM unified representation parameters

	N_1	N_2	$\boldsymbol{X}_0$	$\boldsymbol{U}_3$	$\boldsymbol{W}_{\mathrm{tx}}$	$\boldsymbol{U}_2$	$\boldsymbol{D}_0$	$\boldsymbol{U}_1$	$\boldsymbol{U}_4$
TD	K	M	$\boldsymbol{X}$	$\boldsymbol{F}^{\mathrm{H}}_M$	$\sqrt{N}\boldsymbol{F}_M\boldsymbol{G}^{\mathrm{T}}$	$\boldsymbol{F}_M$	$\boldsymbol{D}$	$\boldsymbol{F}^{\mathrm{H}}_K$	$\boldsymbol{I}_N$
FD	M	K	$\tilde{\boldsymbol{X}}$	$\boldsymbol{F}_K$	$\sqrt{N}\boldsymbol{F}^{\mathrm{H}}_K\tilde{\boldsymbol{G}}^{\mathrm{T}}$	$\boldsymbol{F}^{\mathrm{H}}_K$	$\boldsymbol{D}^{\mathrm{T}}$	$\boldsymbol{F}_M$	$\boldsymbol{F}^{\mathrm{H}}_N$

From a hardware perspective, the unitary matrices need to be easily realized and flexible, such as using radix-2 FFT, Walsh-Hadamard transform. Moreover, using the same type of matrices for $\boldsymbol{U}_1$ and $\boldsymbol{U}_4$, and for $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ enables symmetric architecture, given that the transpose and vectorization can be achieved by means of memory and indexing. The symmetry allows using the same hardware core as modulator and demodulator.

This thesis considers the hardware architecture using flexible FFT blocks, as reported in [NCF18] and [LNF19]. In this design, which is illustrated in Figure 7.4, 4 flexible FFT

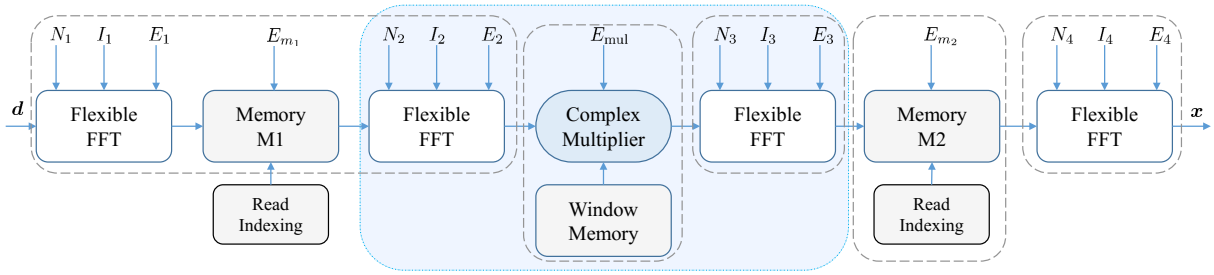

Figure 7.4: Unified architecture for TD and FD GFDM processing.

cores with the configuration parameter N_x for the DFT size and I_x to set the core in the direct $[D]$ or inverse $[I]$ mode. Additionally, each FFT block can be enabled $[E]$ or disabled $[D]$ with the parameter E_x, such that the disabled block forwards the samples to the next stage. This is equivalent to set the corresponding unitary matrix to identity. Moreover, two memory blocks are used to store the result of the previous transform and it is read by customized indexing unit. The modulator window is stored in a memory as part of the modulator configuration. This memory is always read incrementally and written with respect to the implementation domain. Furthermore, one high throughput complex multiplier is used to perform the element-wise multiplication. All the building blocks can be disabled as well. This symmetrical architecture can be used for demodulation. Table 7.2 summarizes the configuration parameters of the modulator and demodulator.

Table 7.2: Configuration parameters for conventional GFDM modem.

	Modulation		Demodulation	
	TD	FD	TD	FD
$[N_1, N_2, N_3, N_4]$	$[K, M, M, -]$	$[M, K, K, N]$	$[N, M, M, K]$	$[-, K, K, M]$
$[I_1, I_2, I_3, I_4]$	$[I, D, I, -]$	$[D, I, D, I]$	$[I, D, I, D]$	$[-, I, D, I]$
$[E_1, E_2, E_3, E_4]$	$[E, E, E, D]$	$[E, E, E, E]$	$[E, E, E, E]$	$[D, E, E, E]$
Window	$\sqrt{N}\boldsymbol{F}_N\boldsymbol{G}^{\mathrm{T}}$	$\sqrt{N}\boldsymbol{F}_K^{\mathrm{H}}\tilde{\boldsymbol{G}}$	$\sqrt{N}\boldsymbol{F}_M\boldsymbol{\Gamma}^{\mathrm{T}}$	$\sqrt{N}\boldsymbol{F}_K^{\mathrm{H}}\tilde{\boldsymbol{\Gamma}}$
M1	$\{\cdot\}^{\mathrm{T}}$	$\{\cdot\}^{\mathrm{T}}$	$\mathrm{unvec}_{K\times M}\{\cdot\}^{\mathrm{T}}$	$\mathrm{unvec}_{M\times K}\{\cdot\}^{\mathrm{T}}$
M2	$\mathrm{vec}\left\{\{\cdot\}^{\mathrm{T}}\right\}$	$\mathrm{vec}\left\{\{\cdot\}^{\mathrm{T}}\right\}$	$\{\cdot\}^{\mathrm{T}}$	$\{\cdot\}^{\mathrm{T}}$

Extended flexibility. For a given realization of the unitary matrices, the intrinsic flexibility in this architecture comes from the choice of the modulation window, and the setting of the transform dimensions. In addition, this design provides more degrees of freedom as follows,

1. Flexible spreading: it is achieved by switching one or both the spreading matrices. This is equivalent to replacing the spreading matrix by identity matrix. It can also be seen as GFDM precoding [MMG+16].

2. Bypassing windowing and/or transform: this flexibility can be exploited to reduce the latency when generating specific waveforms, such as OFDM and DFT-spread-OFDM.

3. Flexible sample mapping: instead of transmitting $\boldsymbol{x}_0 = \text{vec}\left\{\boldsymbol{X}_0^{\text{T}}\right\}$, other type of mapping, such as $\boldsymbol{x}_0 = \text{vec}\left\{\boldsymbol{X}_0\right\}$, which essentially produces orthogonal time frequency space (OTFS) signal, as reported in [NCMF18]. On the other hand, the flexible allocation allows direct integration of FD-GFDM within an OFDM system. For that, the additional IDFT block used in the FD modulation needs to be flexible.

4. Extended design flexibility: this can be achieved by replacing one or more DFT/IDFT blocks with other unitary transforms, such as Walsh-Hadamard transform (WHT) which does not require any CMs. This can also be provided as a run-time reconfiguration option.

7.2.1 Architectures complexity and flexibility analysis

This analysis targets the architecture of Figure 7.4, where the unitary transforms are performed with DFT/IDFT using radix-2 parameters, M, K, N, which enables the implementation with radix-2 FFT. Considering the input signal to the demodulator is in the FD, as illustrated in Figure 7.5 the complexity analysis involves the modulator, demodulator, and the N-FFT transform at the equalizer. The design is evaluated with respect to the implementation of the convolution, either with *FFT-based* or *direct* using the convolution in Figure 7.2. Depending on the configuration of the modulator-demodulator, there are three possible realizations. Namely, TD-FD, TD-TD, and FD-FD. In this scenario, the FFT-based architecture can perform TD-FD with any parameters, whereas the convolution architecture is limited by the number of memory blocks, thus the selected domain needs to be aligned with the smallest parameter.

Figure 7.5: Complexity evaluation model.

7.2.2 Number of multiplications

The N-FFT requires $\frac{N}{2}\log_2(N)$ CMs for $N > 2$, and for $N = 2$, the FFT is performed with addition only. The FFT-based TD-FD architecture performs $3M$ times K-FFT/IFFT and $3K$ times M-FFT/IFFT operations. Therefore, the overall transforms require $3\frac{MK}{2}\log_2(K) + 3\frac{KM}{2}\log_2(M) + \frac{N}{2}\log_2(N) = 2N\log_2(N)$ CMs. This is twice the

number of CMs required by OFDM of length N. Moreover, $2N$ multiplications are needed for the product with the modulation/demodulation window. The direct modulation is achieved with M times K-IFFT transforms using TD modulator, K times M-IDFT with FD demodulator. Thus, $\frac{MK}{2}\log_2(K)+\frac{KM}{2}\log_2(M)+\frac{N}{2}\log_2(N) = N\log_2(N)$ CMs for the transforms. Moreover, the TD direct convolution requires MN and the FD requires KN CMs. If only TD or FD modem is applied, two N-FFT/IFFT transforms with $N\log_2(N)$ CMs are required to transform the signal into the corresponding domain. In addition, the direct TD-TD and FD-FD require $2[\frac{MK}{2}\log_2(K)+MN]$, and $2[\frac{KM}{2}\log_2(M)+KN]$ CMs, respectively. The special case of the direct FD-FD, where the complexity can be reduced by the consideration of the sparsity of the FD prototype pulse, reduces the number of CMs for the convolution to $2LN$. This special case is useful for the processing of conventional GFDM waveform. However, the receiver pulse may overlap with more than L subcarriers, e.g. when zero-forcing demodulator is applied with non-orthogonal modulation matrix. Table 7.3 lists the overall number of CMs for the different type of modem realizations. The

Table 7.3: Total number of CMs considering FD equalization.

Implementation method	Total number of CMs
FFT-based, TD-FD	$2N\log_2(N)+2N$
Direct, TD-FD	$N\log_2(N)+[K+M]N$
Direct, TD-TD	$N\log_2(N)+N\log_2(K)+2MN$
Direct, FD-FD	$N\log_2(N)+N\log_2(M)+2KN$
Direct (L overlap), FD-FD	$N\log_2(N)+N\log_2(M)+2LN$

number of CMs required by the FFT-based TD-FD design depends only on $N = KM$ but not on the individual K and M. The complexity is approximately 2 times the complexity of OFDM. On the contrary, the complexity of the direct TD-FD implementation depends on the sum $K + M$, which in most cases requires more CMs than the one based on FD or TD only, as can be seen in Figure 7.6. For $M < K$, it is more efficient to use the direct TD-TD modem, while in the case of $K < M$ using the direct FD-FD modem is more efficient. Thus, the direct convolutional structure can be switched depending on the smaller parameter. Compared with the FFT-based TD-FD, the complexity of the direct TD-TD is slightly higher with less than 2-times up to $M < 16$, $N = 1024$. For $M = 16$, it is 2.1 times higher and For $K = M = 32$, the complexity is 3.6 times higher. For $M > 32$, the modem can be switched to the FD-FD mode. In the special direct FD-FD with $L = 2$, the complexity increases with M. For $M = N/4$, this modem has the same complexity as the FFT-based TD-FD, and it is about 23% lower for $M = 8$ and $N = 1024$.

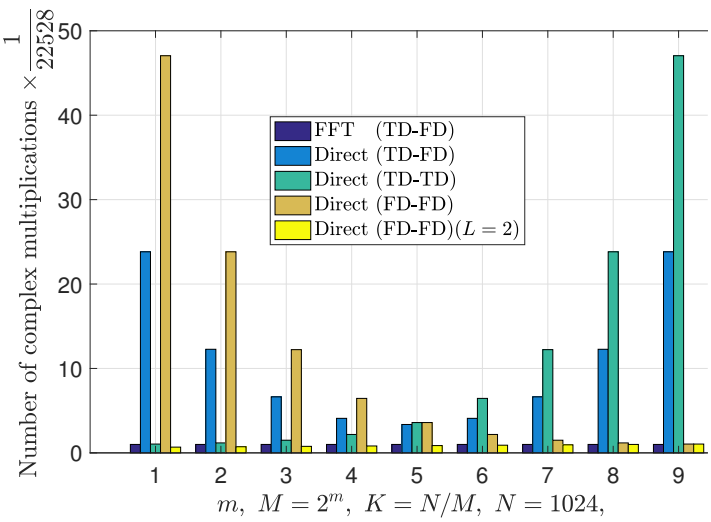

Figure 7.6: Number of CMs, where $N = 1024$ and $M = 2^m, m = 1 \cdots, 9$.

7.2.3 Hardware analysis

Based on the complexity analysis, the most efficient setups are considered. Namely, FFT-based TD-FD, the direct TD-TD and the direct FD-FD, as shown in Figure 7.7. The architectures are compared in terms of flexibility, the required resources, and the modulation/demodulation latency, which also includes the FD equalization transform. Both direct architectures are equivalent, but they differ in the placement of the N-IFFT block. Let $N_{\max}$ be the maximum supported FFT length and $L_{\max}$ the number of parallel multiplication chains in the direct modem.

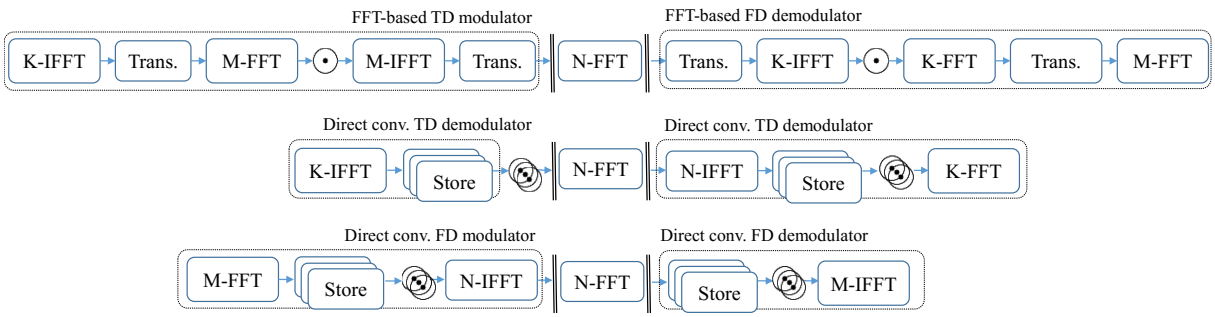

Figure 7.7: Efficient modem realizations.

Flexibility. The FFT-based architecture supports all combinations of radix-2 K and M with $KM \leq N_{\max}$. The direct architecture supports all combinations that additionally satisfy $\min(K, M) \leq L_{\max}$. For instance, if $N_{\max} = 2048$ and $L_{\max} = 16$, then $(K, M) \in \{(32, 32), (32, 64), (64, 32)\}$ cannot be supported. Here, it is assumed that the switching between direct TD-TD and direct FD-FD is realized. If the direct architecture is implemented with N-DFT internet protocole (IP) cores instead of FFT cores, then

the allowable combinations are additionally influenced by the design of the DFT core. The FFT-based design can provide additional flexibility by customizing the indexing of the two memories, disabling some blocks. However, only the convolution is feasible with the internal FFT transforms, whereas the convolution architecture can be used to realize other transforms, benefiting from the parallel memories.

Resource consumption. For real-time FPGA implementation, the resource consumption is considered in terms of number of consumed FFT and complex multiplier IP cores, and the read-and-write (RW) and read-or-write (R/W) random-access memmory (RAM) blocks. Each block RAM can fit up to $N_{\max}$ complex samples. The RW-RAMs are constructed to enable read and write at the same time for the purpose of pipelining. Table 7.4 lists the number of required resources. The FFT-based design mainly

Table 7.4: Number of required resources.

Implementation	FFT	Mult.	(RW)-RAM	R/W-RAM
FFT-based	7	2	4	2
Direct	4	$2L_{\max}$	$2L_{\max}$	$2L_{\max}$

saves memory and multiplier resources. Only 2 complex multipliers, 2 R/W-RAM blocks are required to store the modulation/demodulation windows, and 4 RW-RAM blocks to perform the transpose. The direct architecture consumes more complex multipliers and memory blocks depending on $L_{\max}$. The FFT-based modem requires 3 more FFT IP cores. The overall resource consumption depends on the design and implementation of the IP cores and the parallel multiplications in the direct architecture. On the other hand, the direct modem requires more resources for control and routing, which is not considered in this evaluation.

Latency. This evaluation considers the latency of the modulation/demodulation and the FD transform at the equalizer. The latency is a measure of the delay in number of cycles between the first input data symbol at the modulator and the last output symbol at the demodulator. Considering both designs incorporate proper pipelining. Each N-FFT/IFFT requires N cycles to load the samples and P_M cycles to perform the transform. The memory storage of N samples requires N cycles, and the multiplier delay is denoted as T_{m}. Moreover, reading all the demodulated symbols requires N cycles. From Figure 7.7, it can be seen that the FFT-based implementation requires $6N$ delay cycles corresponding to 4 memory transposes, the equalizer N-FFT load and the unload of the demodulated samples. Moreover, $3(M+K)$ delay cycles to load the data in 3 time M-FFT/IFFT and 3 time K-FFT/IFFT blocks. The processing delay is $P_N + 3(P_K + P_M) + 2T_{\mathrm{m}}$ for the FFT/IFFT transforms and multipliers. The overall latency is given by

$$T_{\mathrm{FFT}} = 6N + 3(K+M) + P_N + 3\left(P_K + P_M\right) + 2T_{\mathrm{m}}. \tag{7.14}$$

The direct modem requires $3N$ delay cycles for intermediate storage and unloading the demodulated samples. $2N + 2K$ are required for loading the samples to the FFT/IFFT blocks. The processing delay is $2P_N + 2P_K + 2T_m$ for the transforms and the parallel multipliers. Therefore, the overall delay of the direct TD-TD modem is

$$T_{\text{Direct-TD}} = 5N + 2K + 2P_N + 2P_K + 2T_{\text{m}}. \tag{7.15}$$

Similarly for the FD-FD modem by replacing K by M,

$$T_{\text{Direct-FD}} = 5N + 2M + 2P_N + 2P_M + 2T_{\text{m}}. \tag{7.16}$$

The latency is influenced by the FFT size and the configuration of the FFT IP cores. The higher the FFT size, the higher the latency. It is worth mentioning that there is a trade-off between the consumed resources and latency.

Table 7.5: Xilinx FFT IP processing latency in cycles.

N	8	16	32	64	128	256	512	1024	2048
P_N	57	110	126	177	241	387	643	1170	2194

Numerical example. In this example, the complex multiplier has a latency $T_m = 12$ cycles and the FFT blocks are realized based on an Xilinx-FFT IP core, which is configured in the pipelined mode with 3 complex multipliers. The processing delay of the FFT block is listed in Table 7.5[1]. The difference between the FFT-based and the direct TD-TD is given by

$$\Delta_{\text{FFT, D-TD}} = N + K + 3M + P_K + 3P_M - P_N. \tag{7.17}$$

As shown in Figure 7.8 and listed in Table 7.6, the latency difference mainly depends

Table 7.6: Latency evaluation.

	$M = 16$				$M = 8$					
N	256	512	1024	2048	64	128	256	512	1024	2048
K	16	32	64	128	8	16	32	64	128	256
$T_{\text{Direct-TD}}$	2330	4186	7966	15390	828	1398	2394	4352	8222	15938
$\Delta_{\text{FFT, D-TD}}$	373	405	473	601	147	208	222	305	418	692
Increase %	16.0	9.6	5.9	3.9	17.7	14.9	9.3	7.0	5.0	4.3

on the size of N. It can be observed that the FFT-based architecture requires additional latency that decreases with the increase of N. For $M = 16$, the additional delay decreases from 16.0% at $N = 256$ to 3.9% at $N = 2048$, and for $M = 16$, it decreases from 17.7%

[1] In the data sheet of the FFT IP core, the latency includes the number of cycles required to load the input and unload the output, i.e. $2N$ more cycles.

at $N = 64$ to 4.3% at $N = 2048$. Therefore, the direct design is more appropriate for low latency requirement, especially, with smaller block length. However, this gain is at the cost of significantly increased resources consumption.

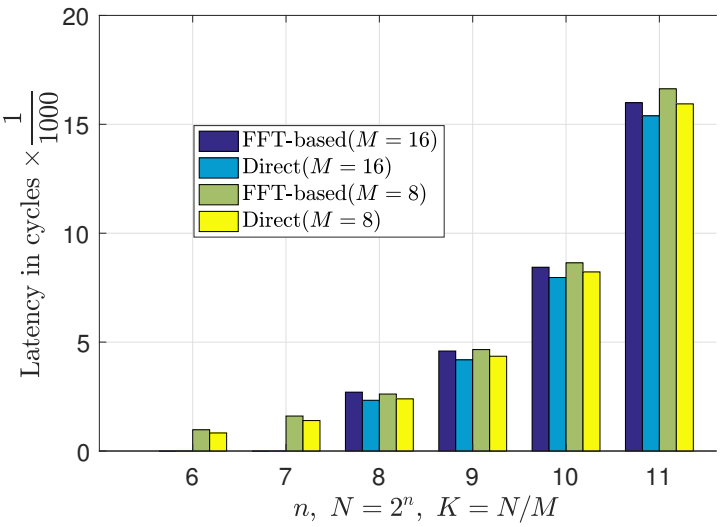

Figure 7.8: Latency evaluation, where M is fixed, and $N = 2^n, n = 6, \cdots, 11$.

7.3 Applications of the extended GFDM framework

As discussed in Section 6.2, the GFDM modem is a core building block for waveforms based on time-frequency shift, where one or multiple modulators are used to generate the waveform. Using multiple modems can be a logical term, where a single core is reused, or by replicating the core several times to reduce latency. The developed modem core can be also used as standalone waveform generator or integrated within other systems to provide additional features, such as integrating with OFDM. With respect to MIMO, this core can be used to generate the samples for each antenna and to perform precoding. It can be used in the design of flexible multiple access to provide a customized waveform for each user. To exploit the extended GFDM modem, a linear problem needs to be formulated to have the generic matrix form as in (7.7)

$$\boldsymbol{A}_0 = \boldsymbol{U}_4 \boldsymbol{S} \boldsymbol{U}^{(3)}_{N_1,N_2} \boldsymbol{\Lambda}^{(\mathrm{tx})} \boldsymbol{U}^{(2)}_{N_1,N_2} \boldsymbol{\Pi} \boldsymbol{U}^{(1)}_{N_2,N_1}. \tag{7.18}$$

Given $\boldsymbol{D}_0 \in \mathbb{C}^{N_1 \times N_2}$, $\boldsymbol{A}_0 \in \mathbb{C}^{N_0 \times N_0}$, $N_0 = N_1 \times N_2$. Thus, $\boldsymbol{x}_0 = \boldsymbol{A}_0 \boldsymbol{d}_0$, $\boldsymbol{d} = \mathrm{vec}\{\boldsymbol{D}_0\}$. The matrix $\boldsymbol{U}^{(1)}_{N_2,N_1}$ is a block diagonal matrix generated by placing N_2 unitary matrices $\boldsymbol{U}_1 \in \mathbb{C}^{N_1 \times N_1}$, i.e. $\boldsymbol{U}^{(1)}_{N_2,N_1} = \boldsymbol{I}_{N_2} \otimes \boldsymbol{U}_1$. This matrix corresponds to performing the transform on the columns of $\boldsymbol{D}_0$, such that $\bar{\boldsymbol{D}}_0 = \boldsymbol{U}_1 \boldsymbol{D}_0$. Thus, $\bar{\boldsymbol{d}}_0 = \boldsymbol{U}^{(1)}_{N_2,N_1} \boldsymbol{d}_0$. The matrix $\boldsymbol{\Pi}$ is a permutation matrix of size $N \times N$, and it is realized by memory and indexing. The matrices $\boldsymbol{U}^{(2)}_{N_1,N_2}$ and $\boldsymbol{U}^{(3)}_{N_1,N_2}$ are block diagonal unitary matrices, consisting of N_1 unitary

matrices of size $N_2 \times N_2$, and $\boldsymbol{\Lambda}^{(\mathrm{tx})} = \mathrm{diag}\left\{\mathrm{vec}\left\{\boldsymbol{W}_{\mathrm{tx}}\right\}\right\}$, $\boldsymbol{W}_{\mathrm{tx}} \in \mathbb{C}^{N_2 \times N_1}$. The form of these two unitary matrices depends on the available architecture. In case of a single transform,

$$\boldsymbol{U}^{(2)}_{N_2,N_1} = \boldsymbol{I}_{N_1} \otimes \boldsymbol{U}_2,\ \boldsymbol{U}^{(3)}_{N_2,N_1} = \boldsymbol{I}_{N_1} \otimes \boldsymbol{U}_3.$$

If the parallel memory architecture of Section 7.1.4 is used then, the block matrices can be any unitary matrices $\{\boldsymbol{U}_2^{(n_1)}\}$such that

$$\boldsymbol{U}^{(2)}_{N_2,N_1} = \mathrm{Bd}(\{\boldsymbol{U}_2^{(n_1)}\}_{n_1=0}^{N_1-1}),\ \boldsymbol{U}^{(3)}_{N_2,N_1} = \mathrm{Bd}(\{\boldsymbol{U}_3^{(n_1)}\}_{n_1=0}^{N_1-1}).$$

In all cases, the matrix $\boldsymbol{C}_0 = \boldsymbol{U}^{(3)}_{N_1,N_2}\boldsymbol{\Lambda}^{(\mathrm{tx})}\boldsymbol{U}^{(2)}_{N_1,N_2}$ is block diagonal, with N_1 matrices defined by

$$\boldsymbol{C}_0^{(n_1)} = \boldsymbol{U}_3^{(n_1)}\boldsymbol{\Lambda}_{n_1}^{(\mathrm{tx})}\boldsymbol{U}_2^{(n_1)},\ \boldsymbol{\Lambda}_{n_1}^{(\mathrm{tx})}\mathrm{diag}\left\{\boldsymbol{W}_{\mathrm{tx}}[:,n_1]\right\}. \tag{7.19}$$

Finally, $\boldsymbol{S}$ is a permutation matrix of size $N_0 \times N_0$, such that $\boldsymbol{x}_0 = \boldsymbol{P}\bar{\boldsymbol{x}}_0$, with

$$\bar{\boldsymbol{x}}_0 = \bar{\boldsymbol{A}}_0\boldsymbol{d}_0,\ \bar{\boldsymbol{A}}_0 = \boldsymbol{C}_0\boldsymbol{\Pi}\boldsymbol{U}^{(1)}_{N_2,N_1}. \tag{7.20}$$

The matrix $\bar{\boldsymbol{A}}_0$ produces the transform samples, and $\boldsymbol{S}$ places the samples in the final order. The matrix $\boldsymbol{U}_4 \in \mathbb{C}^{N_0 \times N_0}$ provides a final transform to the desired domain.

7.3.1 Generalized FDMA

In generic FDMA introduced in Section 4.2.2, each user can use a matrix $\boldsymbol{A}_u \in \mathbb{C}^{N_u \times N_u}$ then to map to the final vector. Using the developed framework, with $\boldsymbol{U}_4 = \boldsymbol{F}_N^{\mathrm{H}}$, and $\boldsymbol{x}_0 = \tilde{\boldsymbol{x}}$ the final vector, each user matrix is set to the size $N_u = M_u K_u$, $N_1 = M_u$ and $N_2 = K_u$ such that $\tilde{\boldsymbol{A}}_u = \boldsymbol{S}_u\bar{\boldsymbol{A}}_0$. The required modification is to add user allocation matrix $\boldsymbol{S}_u \in \mathbb{R}^{N \times N_u}$ and overlap function to perform the function $\tilde{\boldsymbol{x}}\left[\mathcal{N}_u\right] = \tilde{\boldsymbol{x}}\left[\mathcal{N}_u\right] + \tilde{\boldsymbol{x}}_u$. This can be combined with the allocation matrix in the same memory. Thus, no additional memory is required for the design. For the downlink (DL) realization, parallel small-size modulators can be used at the base stations (BS), or one single modulator can be reused. Depending on the implementation, the allocation mechanism works to write the samples in a memory corresponding to $\tilde{\boldsymbol{x}}$, which is then read sequentially to the final DFT block. To enable the overlapping, the previous value is read first then added to the new values, and the result is written back. The demodulator for uplink (UL) at the BS can also be implemented with parallel small-size demodulators or reusing a single demodulator IP core. The deallocation unit is responsible for feeding the samples to each user.

Control overhead and practical considerations. In theory, each user can use any matrix parameters, namely, the type and size of the unitary transforms and the diagonal matrix coefficients. Moreover, any allocation pattern can be used. Exchanging arbitrary coefficients of the diagonal matrix is not practical, and this can be solved by using predefined functions to generate them. Considering different type of unitary matrices

are available, this type needs to be conveyed in the control. For the extended transform (7.18), it is required to convey the types of $2N_1$ unitary matrices. Thus, by limiting the transform to fixed matrices $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ and $\boldsymbol{U}_3$ only 3 values are required. The transform size needs to be conveyed for N_1 and N_2 parameters. However, the number of possible combination can be limited, for example radix-2 combinations. This reduces the amount of control overhead. Finally, conveying the information of allocation can be achieved by logical indexes, using a series of N bits. However, this amount is very high, and can be reduced by setting a group of allocated indexes. For instance, by defining the active set as a union of subsets $\mathcal{N}_{on} = \cup_{n_a=1}^{N_a} \mathcal{N}_{n_a}$. The sets, $\mathcal{N}_{n_a}$ contains predefined indexes. These sets can overlap and their elements can be contiguous or spread. In this case, only N_a bits need to be sent in the control. Considering also that N is configurable parameter to set the block length, several options need to be conveyed. Additional waveform parameter, such as cyclic prefix (CP) length, or overlapping are also part of the control overhead. As a result, the waveform control header needs to include the following information,

1. Waveform options: to determine the selected unitary transforms and how to generate the modulator window, either by a generator function or using look-up memory.
2. Waveform dimensions: an option to determine the block parameters M_u, K_u.
3. Allocation set indexes: defined by binary values, where the value 1 denotes an allocated set. Note that this set is related to the sample allocation not the data symbol allocation.
4. Further processing options: to convey information about the block multiplexing including, CP, padding or overlapping, and the length of the core block.

Note that the physical layer header needs to contain data processing control information such as the active data set and the modulation coding scheme (MCS).

The required waveform control overhead depends on the degree of flexibility the system offers. To reduce the overhead traffic, the waveform parameter can be set occasionally, such as when a user connects to the network, or based on channel requirements. On the other hand, to reduce the latency required by writing the configuration memories, the involved matrix entries can be pre-calculated and stored, which however increases the required memory resources. Additionally, providing different hardware precoding, requires additional resources. To get rid of that, some parameters can be fixed. For instance, if the diagonal matrix elements are constant, and $\boldsymbol{U}_2 = \boldsymbol{U}_3^{\mathrm{H}}$, then $\boldsymbol{C}_0$ is identity, and thus, all the required memories and transforms can be disabled. Moreover, if $\boldsymbol{U}_1$ is fixed to DFT or identity matrix, only one bit is required to determine $\boldsymbol{U}_1$.

Extension to MIMO-FDMA. Using the FD implementation, the output MIMO samples computed in the vectors $\{\tilde{\boldsymbol{x}}_{u,n_t} \in \mathbb{C}^{N_u \times 1}\}$, are mapped to the MIMO vectors

$\{\tilde{\boldsymbol{x}}_{n_t} \in \mathbb{C}^{N\times 1}\}$. In General, a user samples can be spread over one or multiple antennas with the same or different FD sample indexes. However, to preserve orthogonality, the same FD allocation is used on all antennas. For control, it is required to communicate the MIMO allocation in addition to the waveform parameters.

7.3.2 Enhancement of OFDM system

As reported in [NCF19c], GFDM can be integrated in OFDM by inserting a flexible small-size FD-GFDM modulator to precode the input data of certain subcarriers. At the receiver side, channel equalization is required in addition to the FD-GFDM demodulator, as illustrated in Figure 7.9. One application of this model is to improve the spectral

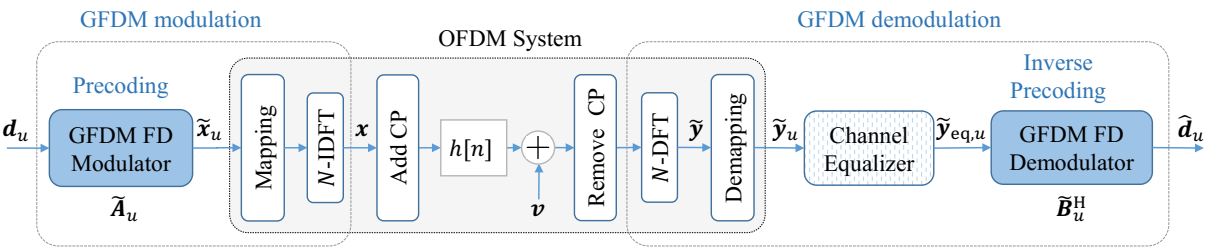

Figure 7.9: Precoded OFDM system model.

utilization of OFDM by reducing the required guard band. Instead of disabling a large number of OFDM subcarriers to produce low out-of-band (OOB), GFDM can be used on the edge subcarriers with proper configurations to provide low side lobes. On the other hand, to preserve low-complexity, the middle subcarriers can keep using plain OFDM via bypassing of the GFDM precoding. This feature can also be used to support asynchronous multiple access, which allows using fewer guard subcarriers between adjacent users. This model can also be used to introduce different type of orthogonal precoding, using the extended flexibility features. The precoding can be used, e.g., to spread the data over larger bandwidth to harvest frequency diversity.

7.4 Summary

In this chapter, modulation classes are defined based on the SVD decomposition of the core modulation matrix $\boldsymbol{A} = \boldsymbol{V}_t\boldsymbol{\Lambda}^{(\mathrm{tx})}\boldsymbol{U}_t$. In a specific modulation class, the unitary matrices $\boldsymbol{V}_t$ and $\boldsymbol{U}_t$ are fixed, whereas the diagonal matrix $\boldsymbol{\Lambda}^{(\mathrm{tx})}$ is flexible. Allowing flexibility in the unitary matrices enables the implementation of several modulation classes with one architecture. Inspired from the GFDM DFT-based architecture, a generic modulation framework is obtained by replacing the DFT by generic unitary transforms such that,

$$\boldsymbol{X}_0^{\mathrm{T}} = \boldsymbol{U}_3\left(\boldsymbol{W}_{\mathrm{tx}} \odot [\boldsymbol{U}_2(\boldsymbol{U}_1\boldsymbol{D}_0)^{\mathrm{T}}]\right), \ \boldsymbol{x}_0 = \boldsymbol{U}_4\mathrm{vec}\left\{\boldsymbol{X}_0\right\}.$$

where $\boldsymbol{U}_x \in \mathbb{C}^{N_x \times N_x}$ are unitary matrices, $\boldsymbol{W}_{\text{tx}} \in \mathbb{C}^{N_2 \times N_1}$, $N_2 = N_3$ and $N_4 = N_1 N_2$. Additional flexibility is added by disabling one or more transforms, and changing the permutation involved in the transpose and vectorization. In a special case, where $\boldsymbol{U}_3 = \boldsymbol{U}_2^{\text{H}}$, and using $\boldsymbol{G}_0^{(m_2)}[n_1, n_2] = \left(\boldsymbol{U}_3 \text{diag}\left\{\boldsymbol{W}_{\text{tx}}[:, n_1]\right\} \boldsymbol{U}_3^{\text{H}}\right)[n_2, m_2]$, the modulation can be realized with parallel multiplication architecture with N_2 complex multipliers as

$$\boldsymbol{X}_0[n_1, n_2] = \sum_{m_2=0}^{N_2-1} \bar{\boldsymbol{D}}_0^{(m_2)}[n_1, n_2] \boldsymbol{G}_0^{(m_2)}[n_1, n_2], \ \bar{\boldsymbol{D}}_0^{(m_2)}[n_1, n_2] = (\boldsymbol{U}_1 \boldsymbol{D}_0)[n_1, m_2].$$

This architecture allows customizing the unitary matrices $U_3^{(n_1)}$ for each column of $\boldsymbol{D}_0$.

The extension to MIMO waveform is achieved by changing the transformation architecture to include multiple N_T windows. The general MIMO modulation matrix becomes

$$\boldsymbol{A} = \left(\boldsymbol{I}_{N_T} \otimes \left(\boldsymbol{U}_4 \boldsymbol{P} \boldsymbol{U}_{N_1,N_2}^{(3)}\right)\right) \boldsymbol{\Pi}_{N_T}^{\text{T}} \boldsymbol{\Psi}^{(\text{tx})} \boldsymbol{\Pi}_{N_T} \left(\boldsymbol{I}_{N_T} \otimes \left(\boldsymbol{U}_{N_1,N_2}^{(2)} \boldsymbol{\Pi} \boldsymbol{U}_{N_2,N_1}^{(1)}\right)\right),$$

where $\boldsymbol{U}_{N_1,N_2}^{(3)} = \boldsymbol{I}_{N_1} \otimes \boldsymbol{U}_3$, $\boldsymbol{U}_{N_1,N_2}^{(2)} = \boldsymbol{I}_{N_1} \otimes \boldsymbol{U}_2$, $\boldsymbol{U}_{N_2,N_1}^{(1)} = \boldsymbol{I}_{N_2} \otimes \boldsymbol{U}_1$ are unitary transforms, $\boldsymbol{\Pi}_{N_T}$ is defined such that $\text{vec}\left\{\left\{\text{unvec}_{N \times N_T}\{\boldsymbol{x}\}\right\}^{\text{T}}\right\} = \boldsymbol{\Pi}_{N_T} \boldsymbol{x}$, $\boldsymbol{\Psi}^{(\text{tx})} = \text{Bd}\{\boldsymbol{\Psi}_0^{(\text{tx})}, \cdots \boldsymbol{\Psi}_{N-1}^{(\text{tx})}\}$ is block diagonal window matrix, $\boldsymbol{\Psi}_n \in \mathbb{C}^{N_T \times N_T}$, and the permutation matrices $\boldsymbol{P}$ and $\boldsymbol{\Pi}$ corresponds to flexible memory indexing. For the SISO case, where $N_T = 1$,

$$\boldsymbol{A} = \boldsymbol{U}_4 \boldsymbol{P} \boldsymbol{U}_{N_1,N_2}^{(3)} \boldsymbol{\Lambda}^{(\text{tx})} \boldsymbol{U}_{N_1,N_2}^{(2)} \boldsymbol{\Pi} \boldsymbol{U}_{N_2,N_1}^{(1)}.$$

The extended GFDM framework is a special practical case, where $\boldsymbol{U}_x = \boldsymbol{F}_{N_x}$ or $\boldsymbol{U}_x = \boldsymbol{F}_{N_x}^{\text{H}}$, which allows controlling the direction of DFT transform. A low-complexity hardware implementation can be obtained using radix-2 FFT implementation. Other low-complexity transforms can be integrated, such as WHT. Finally, this framework can be integrated in OFDM system and can be used for the generalization of FDMA.

Chapter 8

Conclusions and Future Works

This thesis has provided an overview of radio frequency (RF) wireless systems, with specific focus on the role of waveforms. A generic multiple-input, multiple-output (MIMO) signal model for digital baseband processing is introduced. The relation between the transmitted waveform and received discrete signal is defined by a channel function. This function is modeled as a linear time variant (LTV) system, in additional to an interference and additive noise terms. The linear part results form the multipath propagation, analog filtering, and it implicitly involves carrier frequency offset (CFO). The interference term represents the power amplifier (PA) non-linearity, antenna coupling, and IQ imbalance, whereas the noise consists of thermal and quantization noise. This model allows the study and evaluation of the system from different perspectives. In this work, the linear waveform design aspects are investigated highlighting the relation to hardware impairment, the impact on the receiver performance, and the design of modem architectures.

Summary of the results

The results and observations are summarized as follows:

- ▷ A proper deign of the waveform allows the relaxing of the hardware constraints. In particular, low out-of-band (OOB) plays an important role in achieving inter-symbol interference (ISI)-free conversion with relaxed low-pass (LP) filter design. With that, the discrete channel delay length is mainly influenced by the sampling frequency and the wireless channel delay. Moreover, the low peak-to-average power ratio (PAPR) reduces the non-linear distortion caused by the PA. The PAPR needs to be evaluated on the analog baseband signal obtained after the digital-to-analog (DAC) conversion and LP filtering. Nevertheless, the low PAPR measured at the discrete samples can be used as an indication.
- ▷ This work focuses on linear waveforms generated by linear combination of data symbols. The single-carrier (SC) waveform is obtained by linear filtering of the input

data symbols that are modeled with discrete-time random process. This is equivalent to pulse shaping, where the symbols modulate a deterministic pulse. When the symbol rate is smaller than a given bandwidth, Nyquist filter with ISI-free recovery is feasible. The maximum achievable rate is limited to 1 symbol/sec/Hz. The other case is denoted as faster-than-Nyquist signaling (FTN), where the achievable rate is larger than 1 symbol/sec/Hz. In general multicarrier waveforms, the data symbols are split into K streams with lower symbol rates. Each stream applies a different pulse shaping, and the streams are superimposed to formulate the waveform. This thesis extend the multicarrier concept to MIMO waveforms, where KN_T data symbols are used at each chain to generate a different multicarrier waveform with KN_T pulses. This allows the integration of both MIMO precoding and conventional waveforms. The concept is applicable, e.g., to the 2×2 real-valued MIMO obtained from IQ components, in order to design OQAM-FBMC.

- ▷ Waveform design with finite-length pulses is practical for implementation and low-latency transmissions. A generic block-based waveform model is developed to elaborate all waveform design techniques. The modulated block is created from K input data symbols using a modulation matrix $\boldsymbol{G} = \boldsymbol{G}^{(t)}\boldsymbol{A}\boldsymbol{\Pi}^{(d)} \in \mathbb{C}^{N_\mathrm{t} \times K}$, where $\boldsymbol{A} \in \mathbb{C}^{N \times N}, N \geq K$ is the core modulation matrix, $\boldsymbol{\Pi}^{(d)} \in \mathbb{R}^{N \times K}$ allocation matrix, and $\boldsymbol{G}^{(t)} \in \mathbb{C}^{N_\mathrm{t} \times N}$ corresponds to addition waveform design techniques, including windowing, filtering, and cyclic prefix (CP)/cyclic suffix (CS). The transmitted blocks are multiplexed with spacing N_s, where $N_\mathrm{o} = N_\mathrm{t} - N_\mathrm{s}$ determine the overlapping ($N_\mathrm{o} > 0$) or zero-padding (ZP) ($N_\mathrm{o} < 0$). The power spectral density (PSD) depends on $\boldsymbol{G}\boldsymbol{\Lambda}^{(R)2}\boldsymbol{G}^H$, where $\boldsymbol{\Lambda}^{(R)2}$ is the power and rate allocation matrix. Thus, there are many modulation options to obtain the same PSD. At the the demodulator, the received block are demultiplexed, and a receiver matrix $\boldsymbol{G}^{(r)}$ is applied. The effective linear channel is obtained by the overall transform of the receiver, discrete channel, and transmitter matrix. The received core block in relation to the transmitted data symbols follows a generic linear model, $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{d} + \boldsymbol{z} + \boldsymbol{v}$, where $\boldsymbol{H}$ is a channel-related matrix. Additional interference term corresponds to inter-block interference (IBI), hardware impairments, and additive noise. The IBI can be eliminated by using a guard interval such as ZP or CP.
- ▷ This thesis has reviewed the detection techniques of general linear system, which can be applied to different system configurations including MIMO and multiuser systems. It is shown that linear minimum mean square error (LMMSE) is optimal linear receiver in terms of maximizing the achieved signal-to-interference-plus-noise ratio (SINR) per symbol. However, the LMMSE performance is poor in comparison to advanced iterative receivers such LMMSE-PIC.
- ▷ Flexible core modulation design is essential for tuning the system performance depending on the channel. To demonstrate the need of flexible core modulation, this work performs analytical and simulation evaluation in single-input, single-output (SISO) and MIMO systems, for frequency selective channel considering CP

guard interval, ideal hardware, and perfect channel state information (CSI) and synchronization, i.e. $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{d} + \boldsymbol{v}$. The following results are obtained:

- The achievable rate with coded modulation at certain signal-to-noise ratio (SNR) depends on the orthogonal modulation matrix structure, and the QAM order. In ideal conditions, orthogonal frequency division multiplexing (OFDM) in SISO systems allows the separation of the data symbols without interference. This feature is behind the choice of OFDM in many standards. However, the gap between the actual mutual information and channel capacity increases significantly by increasing the SNR.
- Other linear waveforms can be technically considered as precoded OFDM performing data spreading over the subcarriers. The spreading over frequency has the potential of harvesting frequency diversity at the cost of induced ISI. The evaluation of the mutual information assuming genie perfect interference cancellation, shows that the diversity gain depends on the spreading structure and channel selectivity. Full frequency diversity gain is achieved by interleaved spreading over uncorrelated subcarriers. The length of spreading needs to be larger or equal to the number of significant channel taps. On the other hand, contiguous spreading within a number of subcarriers smaller than the coherence bandwidth of the channel, allows similar performance as OFDM with the benefit of reducing the PAPR. Orthogonal modulation matrices with equal amplitude entries achieve equal SINR per symbol, and thus, full frequency diversity. Among them, the discrete Fourier transform (DFT) and Walsh-Hadamard transform (WHT). Spares spreading with DFT enjoys low complexity equalization, whereas as full spreading reduces the PAPR.
- Channel coding and bit interleaving implicitly perform spreading. The frame error rate (FER) performance is influenced by the interleaver design, the coding and modulation, in addition to the frame length. OFDM with high order QAM and low code rate is able to approach the theoretical channel capacity in the same way SC with higher code rate and lower QAM can achieve by employing interference cancellation. However, SC allows low PAPR, which might be beneficial in comparison with the low-complexity equalization of OFDM.
- In MIMO systems with spacial multiplexing, iterative interference cancellation is essential even for OFDM. With the increased number of the receive antennas and when the MIMO channels are uncorrelated, the spreading in frequency is less important assuming genie interference cancellation. However, it has been shown by simulation that spreading, in practice, is still beneficial.
- Considering the input conditions; channel model, SNR level, hardware constraints, and frame length, the performance indicators; reliability, throughput, and latency, intensive simulation is required to find the optimal coding, QAM order, interleaving, and modulation matrix.

- ▷ A generic flexible multiple access (MA) framework is developed with the consideration of waveforms as a tuning parameter to achieve certain performance with affordable complexity. In this model, each uses is allocated to a set of pulses. This allows the realization of different MA schemes including non-orthogonal multiple access (NOMA). Orthogonality is achieved when the users are separated without inter-user interference (IUI). A generalized frequency-division multiple access (FDMA) frameworks is developed in form of OFDM precoding, where the FDMA schemes used in Long Term Evolution (LTE) and fifth generation (5G) are special cases. This provides low-cost enhancement of OFDM, e.g., precoding with low OOB can be applied on the edge subcarriers instead of turning them entirely off. The impact of the waveform (precoding) is studied in terms of asynchronous MA. It is shown that simple design with DFT precoding and employing one guard data symbols is able to significantly reduce the IUI of OFDM resulting from frequency offset (FO) while preserving the same spectral efficiency. Moreover, the WHT precoding achieves similar performance. However, DFT outperforms WHT when using more guard symbols. Using multiple FDMA blocks within the scheduling interval provides time-division multiple access (TDMA).
- ▷ To show the time-frequency structure, the modulated block is represented using M subsymbols each of K sample. The frequency-domain block focuses on the subcarrier structure. Both blocks are related by means of Zak transform. A design approach is proposed for the selection of M and K depending on the channel selectivity. For low selectivity in time, K is chosen such that the channel is static within the subsymbol duration. Therefore, the number of subsymbols can be flexible. In this case, the processing requires K-DFT on the received block to get a simplified relation with the channel. With high time selectivity and sufficiently large $N = MK$, M is selected to obtain subcarrier width smaller than the coherence bandwidth. However, the frequency-domain processing requires N-DFT and M-DFT. On top of that, a general signal model is derived, which shows simple elementwise multiplication relation with the channel. This model is mathematically similar to the case of frequency selective channel. A straightforward extension to a MIMO model has been also presented. Accordingly, the core modulation matrix can be designed to enable simple equalization, or to harvest diversity by means of spreading.
- ▷ A special case of time-frequency modulation results when the pulses are generated from a prototype pulse by means of time and frequency shift. The shift can be arbitrary in general. A particular case with equally-spaced shift corresponds to Gabor frame. The relation of Gabor frame to conventional generalized frequency division multiplexing (GFDM) is studied in details showing the role of GFDM modulation structure in developing flexible multicarrier framework. This thesis proposes an extension to GFDM by employing multiple prototype pulses. One application of that is to realize over-sampled Gabor frame. GFDM has been reviewed and represented in different ways, proposing additional degrees of freedom. Based

on the extended flexibility, it is shown that orthogonal time frequency space (OTFS) results form a permutation of GFDM.

- ▷ This work develops the concept of modulation classes based on the singular value decomposition (SVD) of the core modulation matrix. A class of modulation is defined by a fixed structure of the left and right singular vectors, whereas the singular values are programmable parameters. The flexibility of a modulation class is studied in terms of the number of reconfigurable parameters other than the singular values. GFDM is one of the modulation classes that employs DFT transforms. This thesis designs programmable hardware architectures that allow wide range of waveform processing. Specifically, extended GFDM framework is implemented, which has more degrees of freedom added by means of simple bypass functions and memory indexing or by replacing the DFT transforms, e.g., with WHT. The developed hardware architecture is evaluated in terms of resource consumption, latency and throughput, showing the impact of different designs. Finally, this framework can be integrated in OFDM systems to be used for generalized FDMA.

In summary, the practical linear waveforms are block-based and their main building function is the core modulator. Certain modulation structures can be exploited to reduce the modem complexity. The core block is obtained from a matrix-vector product of complexity N^2. In the most flexible waveform, the whole matrix is parameterized. However, this requires a memory to store N^2 complex sample. Low-complexity modem design targets reducing the number of parameters to be stored, and decreasing the number of required complex multiplications. The main performance metrics to be considered in selecting a specific modulation scheme are: the overall modem low-complexity, low PAPR, reliability, resistance to synchronization misalignment, and low-OOB. The framework developed in this work provides a powerful platform for the evaluation of low-complexity waveforms, not only via simulation but also in real-time transmission.

Future work

The following related future work and open research topics are recognized:

- ▷ In this work, the mutual information evaluation considers the genie perfect interference cancellation. However, this upper bound does not provide accurate analysis that allows the explanation of different behaviors, Thus, the modulation performance needs to be evaluated based on the data vector. The evaluation of the conditional distribution for all combination is computationally exhaustive problem, and therefore, an appropriate approximation is required.
- ▷ Solving the maximum likelihood detection (MLD) w.r.t. data symbols for encoded transmission does not help in the detection without the computation of the

log-likelihood ratios (LLRs), which is a complex problem. Alternatively, this thesis employs iterative receivers based on linear equalization and employing the soft output from the decoder. Finding an appropriate approximation of the LLRs with low-complexity MLD algorithms, such as sphere decoding (SD), is an open issue.

- ▷ The flexible multicarrier framework developed in this work provides a wide range of parameters that can be exploited to optimize the waveforms based on the channel status, the performance requirements in terms of latency, reliability, considering hardware constraints, and low-complexity processing. The best configurations is subject to intensive simulation. Another potential candidate for solving this problem is machien learning (ML) techniques.
- ▷ It has been shown that different combinations of modulation coding scheme (MCS) and modulation matrices can achieve the same performance. For instance, in the numerical example Section 3.4.4, it is shown that SC with 4-quadrature amplitude modulation (QAM) and 1/2 code rate at low SNR might achieve the same performance as OFDM with 256-QAM and code rate 1/8. Also, the performance is influenced by the selected interleaver. This is an open research topic concerns the evaluation of the waveforms with different MCS, different types of channel coding, and interleaving. That is to answer the question, which is better high order modulation with low code rate or the opposite.
- ▷ The current baseband design considers the split of waveform to three blocks; 1) bit processing that produces encoded bits $\boldsymbol{c}$ out of information bits $\boldsymbol{b}$, i.e., $\boldsymbol{c} = \mu(\boldsymbol{b})$, 2) the digital mapping maps the encoded bits to discrete symbols $\boldsymbol{d} = \gamma(\boldsymbol{c})$, 3) linear waveform, where $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{d}$. The number of possible code words is 2^{K_b}, and thus, there are 2^{K_b} valid blocks. The exact MLD is computed by finding the valid block that minimizes the distance. Linear waveforms can be seen as special cases of non-linear modulation, defined by the function $\boldsymbol{x} = f(\boldsymbol{b})$. One approach is by using a codebook stored in a memory. The bit sequence can be used for memory indexing. This approach works for relatively short frames. There are several open question regarding the codebook and receiver design. As the function $f(\cdot)$ is in general non-linear, ML approaches such as deep neural network (DNN) can be exploited. Note that, index modulation [BAPP13] is also a spacial case.
- ▷ End-to-end training focuses on the design of the overall baseband transceiver under different channels. In this approach, instead of independent training, the transmitter and receiver are jointly trained to find the appropriate design.
- ▷ For the emerging applications in joint communications and sensing, it is essential to study the waveform design optimized for localization.
- ▷ Flexible waveforms require exchanging the configuration parameters. One approach proposed in this thesis is based on a list of predefined configurations to minimize the control overhead. First, it is important to determine the best set of parameters, that can serve for different purposes. Second, this design requires proof-of-concept implementation to check its validity in real-time reconfiguration.

- ▷ Considering the available radio and computational resources, joint communications and computations targets the optimization of resource consumption to achieve application requirements. For instance, the latency budget can be distributed among the application processing, such as compression, and data transmission. A large bandwidth enables sending raw data without investing a lot of computation resources on compression, and vice versa. Other trade-off solutions might be also possible. Among the different schemes, the one that minimizes the energy consumption is preferable. This idea can be also employed to the baseband processing design, e.g., to obtain the optimal waveform, bandwidth, and encoding.
- ▷ As discussed in this work, FTN can achieve a nominal rate larger than $\frac{F_s}{B} L_c$ bit/sec/Hz, using a symbol rate F_s, bandwidth B, and L_c bit per symbol. Using low-resolution DAC and analog-to-digital (ADC) converters with very high sampling rate is a potential solution for ultra high data rate. Another approach is to consider using high order QAM, low sampling rate and, aggregated subcarriers. Comparing both approaches in terms of reliability, and energy consumption is a subject for future work.

Publications

Book chapters

[1] **A. Nimr**, S. Ehsanfar, N. Michailow, M. Danneberg, D. Zhang, H. D. Rodrigues, L. L. Mendes, and G. Fettweis, *Generalized Frequency Division Multiplexing: A Flexible Multicarrier Waveform.* Springer International Publishing, 2019, pp. 93–163, ISBN: 9783319920900.

[2] **A. Nimr**, Z. Li, M. Chafii, and G. Fettweis, *Generalized Frequency Division Multiplexing: Unified Multicarrier Framework.* Wiley, 2020, pp. 63–82, ISBN: 9781119652458

Journal publications

[3] **A. Nimr**, M. Chafii, and G. P. Fettweis, "Unified Low Complexity Radix-2 Architectures for Time and Frequency-Domain GFDM Modem," *IEEE Circuits and Systems Magazine*, vol. 18, no. 4, pp. 18–31, Fourthquarter 2018.

[4] **A. Nimr**, M. Matthe, D. Zhang, and G. P. Fettweis, "Optimal radix-2 FFT compatible filters for GFDM," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1497–1500, 2017.

[5] R. Bomfin, **A. Nimr**, M. Chafii, and G. Fettweis, "A robust and low-complexity walsh-hadamard modulation for doubly-dispersive channels," *IEEE Communications Letters*, pp. 1–1, 2020.

[6] M. Danneberg, **A. Nimr**, M. Matthé, and G. P. Fettweis, "Network Slicing for Industry 4.0 Applications–Initial RAN Testbed Results," *Transport*, vol. 2018, 2017.

[7] A. K. Gizzini, M. Chafii, **A. Nimr**, and G. Fettweis, "Deep learning based channel estimation schemes for IEEE 802.11 p standard," *IEEE Access*, vol. 8, pp. 113 751–113 765, 2020.

[8] W. Njima, M. Chafii, **A. Nimr**, and G. Fettweis, "Deep learning based data recovery for localization," *IEEE Access*, vol. 8, pp. 175 741–175 752, 2020.

[9] J. S. Ferreira, H. D. Rodrigues, A. A. Gonzalez, **A. Nimr**, M. Matthe, D. Zhang, L. L. Mendes, and G. Fettweis, "GFDM frame design for 5g application scenarios," *Journal of Communication and Information Systems*, vol. 32, no. 1, 2017.

Conference publications

[10] **A. Nimr**, M. Chafii, and G. P. Fettweis, "Precoded-OFDM within GFDM Framework," in *VTC Spring*. IEEE, 2019, pp. 1–5.

[11] **A. Nimr**, M. Chafii, and G. P. Fettweis, "Practical GFDM-based linear receivers," in *SCC*. ITG, 2019.

[12] **A. Nimr**, M. Chafii, and G. P. Fettweis, "Low-complexity transceiver for GFDM systems with partially allocated subcarriers," in *WCNC*. IEEE, 2019.

[13] **A. Nimr**, M. Chafii, M. Matthe, and G. P. Fettweis, "Extended GFDM framework: OTFS and GFDM comparison," in *GLOBECOM*. IEEE, 2018, pp. 1–6.

[14] **A. Nimr**, D. Zhang, A. B. Martinez, and G. P. Fettweis, "A study on the physical layer performance of GFDM for high throughput wireless communication," in *EUSIPCO*. IEEE, 2017, pp. 638–642.

[15] Z. Li, **A. Nimr**, and G. Fettweis, "Implementation and Performance Measurement of Flexible Radix-2 GFDM Modem," in *2019 IEEE 2nd 5G World Forum (5GWF)*, 2019, pp. 130–134.

[16] M. Danneberg, **A. Nimr**, N. Michailow, S. Ehsanfar, M. Matthe, A. B. Martinez, D. Zhang, and G. P. Fettweis, "Universal Waveforms Processor," in *EuCNC*. IEEE, 2018, pp. 357–362.

[17] A. Fadel, **A. Nimr**, H. . L. Chiang, M. Chafii, and B. Cousin, "Cross-Layer Multi-User Selection in 5G Heterogeneous Networks Based on Hybrid Beamforming Optimization for Millimeter-Wave," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–7.

[18] J. Zhang, **A. Nimr**, K. Naskovska, and M. Haardt, "Enhanced tensor based semi-blind estimation algorithm for relay-assisted MIMO systems," in *LVA/ICA*, ser. Lecture Notes in Computer Science, vol. 9237. Springer, 2015, pp. 64–72.

[19] J. Zhang, **A. Nimr**, and M. Haardt, "Joint design of multi-tap filters and power control for FBMC/OQAM based two-way decode-and-forward relaying systems in highly frequency selective channels," in *ICASSP*. IEEE, 2015, pp. 2439–2443.

[20] M. Danneberg, Z. Li, P. Kühne, **A. Nimr**, S. Ehsanfar, M. Chafii, and G. Fettweis, "Real-time waveform prototyping," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* IEEE, 2019, pp. 1–5.

[21] M. Danneberg, R. Bomfin, **A. Nimr**, Z. Li, and G. Fettweis, "USRP-Based Platform for 26/28 GHz mmWave Experimentation," in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2020, pp. 1–6.

[22] M. Danneberg, R. Bomfin, S. Ehsanfar, **A. Nimr**, Z. Lin, M. Chafii, and G. Fettweis, "Online wireless lab testbed," in *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, 2019, pp. 1–5.

Bibliography

[AAB+06] F. Agnelli, G. Albasini, I. Bietti, A. Gnudi, A. Lacaita, D. Manstretta, R. Rovatti, E. Sacchi, P. Savazzi, F. Svelto, et al. Wireless multi-standard terminals: system analysis and design of a reconfigurable RF front-end. *IEEE Circuits and Systems Magazine*, 6(1):38–59, 2006.

[ABB+07] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, et al. Survey of channel and radio propagation models for wireless MIMO systems. *EURASIP Journal on Wireless Communications and Networking*, 2007(1):019070, 2007.

[AES+16] Y. Akai, Y. Enjoji, Y. Sanada, R. Kimura, and R. Sawai. GFDM with Different Subcarrier Bandwidths. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5, Sep. 2016.

[AH09] K. S. Ahn and R. W. Heath. Performance analysis of maximum ratio combining with imperfect channel estimation in the presence of cochannel interferences. *IEEE Transactions on Wireless Communications*, 8(3):1080–1085, 2009.

[AIM10] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[AJM15] J. Abdoli, M. Jia, and J. Ma. Filtered OFDM: A new waveform for future wireless systems. In *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 66–70, June 2015.

[AKR16] C. An, B. Kim, and H. Ryu. Design of W-OFDM and nonlinear performance comparison for 5G waveform. In *2016 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1006–1009, Oct 2016.

[All15] N. Alliance. 5G white paper. *Next generation mobile networks, white paper*, 1, 2015.

[AWC+05] V. Abhayawardhana, I. Wassell, D. Crosby, M. Sellars, and M. Brown. Comparison of empirical propagation path loss models for fixed wireless access systems. In *2005 IEEE 61st Vehicular Technology Conference*, volume 1, pages 73–77. IEEE, 2005.

[AWM14] A. Asadi, Q. Wang, and V. Mancuso. A survey on device-to-device communication in cellular networks. *IEEE Communications Surveys & Tutorials*, 16(4):1801–1819, 2014.

[Bal05] C. A. Balanis. Antenna theory: Analysis and design. 2005.

[BAPP13] E. Başar, . Aygölü, E. Panayırcı, and H. V. Poor. Orthogonal Frequency Division Multiplexing With Index Modulation. *IEEE Transactions on Signal Processing*, 61(22):5536–5549, Nov 2013.

[BEH12] D. Boswarthick, O. Elloumi, and O. Hersent. *M2M communications: a systems approach.* John Wiley & Sons, 2012.

[BH97] H. Bolcskei and F. Hlawatsch. Discrete Zak transforms, polyphase transforms, and applications. *IEEE Transactions on Signal Processing*, 45(4):851–866, April 1997.

[Bin90] J. A. C. Bingham. Multicarrier modulation for data transmission: an idea whose time has come. *IEEE Communications Magazine*, 28(5):5–14, May 1990.

[BK79] B. R. Burleson and S. L. Kline. Habermas' theory of communication: A critical explication. *Quarterly Journal of Speech*, 65(4):412–428, 1979.

[BNCF20] R. Bomfin, A. Nimr, M. Chafii, and G. Fettweis. A robust and low-complexity walsh-hadamard modulation for doubly-dispersive channels. *IEEE Communications Letters*, pages 1–1, 2020.

[BPW+18] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, Č. Stefanović, et al. Towards massive connectivity support for scalable mMTC communications in 5G networks. *IEEE access*, 6:28969–28992, 2018.

[BS01] J. Bray and C. F. Sturman. *Bluetooth 1.1: connect without cables.* pearson Education, 2001.

[BTS+13] G. Berardinelli, F. M. L. Tavares, T. B. Sørensen, P. Mogensen, and K. Pajukoski. Zero-tail DFT-spread-OFDM signals. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 229–234, Dec 2013.

[CEO99] G. Cherubini, E. Eleftheriou, and S. Olcer. Filtered multitone modulation for VDSL. In *Seamless Interconnection for Universal Services. Global Telecommunications Conference. GLOBECOM'99. (Cat. No.99CH37042)*, volume 2, pages 1139–1144 vol.2, Dec 1999.

[CJS+10] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel, full duplex wireless communication. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, MobiCom '10, page 1–12, New York, NY, USA, 2010. Association for Computing Machinery.

[CKFM18] D. Carrillo, S. Kumar, G. Fraidenraich, and L. L. Mendes. Bit Error Probability for MMSE Receiver in GFDM Systems. *IEEE Communications Letters*, 22(5):942–945, May 2018.

[CMP99] E. Costa, M. Midrio, and S. Pupolin. Impact of amplifier nonlinearities on OFDM transmission system performance. *IEEE Communications Letters*, 3(2):37–39, 1999.

[CQC+18] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann. Modulation and Multiple Access for 5G Networks. *IEEE Communications Surveys Tutorials*, 20(1):629–646, 2018.

[CS17] P. Chen and B. Su. Filter optimization of out-of-band radiation with performance constraints for GFDM systems. In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, July 2017.

[CSH17] P. Chen, B. Su, and Y. Huang. Matrix Characterization for GFDM: Low Complexity MMSE Receivers and Optimal Filters. *IEEE Transactions on Signal Processing*, 65(18):4940–4955, Sep. 2017.

[CT65] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[CTC91] J. S. Chow, J. C. Tu, and J. M. Cioffi. A discrete multitone transceiver system for HDSL applications. *IEEE Journal on Selected Areas in Communications*, 9(6):895–908, Aug 1991.

[CWKS97] B. P. Crow, I. Widjaja, J. G. Kim, and P. T. Sakai. IEEE 802.11 wireless local area networks. *IEEE Communications magazine*, 35(9):116–126, 1997.

[D+94] I. Daubechies, , et al. Gabor time-frequency lattices and the Wexler-Raz identity. *Journal of Fourier Analysis and Applications*, 1(4):437–478, 1994.

[D+15] M. Danneberg et al. Flexible GFDM implementation in FPGA with support to run-time reconfiguration. In *IEEE VTC Fall*, pages 1–2, 2015.

[DB18] K. David and H. Berndt. 6G vision and requirements: Is there any need for beyond 5G? *IEEE Vehicular Technology Magazine*, 13(3):72–80, 2018.

[DBC93] P. Dent, G. E. Bottomley, and T. Croft. Jakes fading model revisited. *Electronics letters*, 29(13):1162–1163, 1993.

[Dd17] J. T. Dias and R. C. de Lamare. Unique-Word GFDM Transmission Systems. *IEEE Wireless Communications Letters*, 6(6):746–749, Dec 2017.

[DFCS04] R. Dinis, D. Falconer, Chan Tong Lam, and M. Sabbaghian. A multiple access scheme for the uplink of broadband wireless systems. In *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04.*, volume 6, pages 3808–3812 Vol.6, Nov 2004.

[DKU98] J. V. Draper, D. B. Kaber, and J. M. Usher. Telepresence. *Human factors*, 40(3):354–375, 1998.

[DMG+15] M. Danneberg, N. Michailow, I. Gaspar, M. Matthé, D. Zhang, L. L. Mendes, and G. Fettweis. Implementation of a 2 by 2 MIMO-GFDM Transceiver for Robust 5G Networks. In *2015 International Symposium on Wireless Communication Systems (ISWCS)*, pages 236–240. IEEE, 2015.

[DMK+12] R. Datta, N. Michailow, S. Krone, M. Lentmaier, and G. Fettweis. Generalized Frequency Division Multiplexing in cognitive radio. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2679–2683, Aug 2012.

[DML+17] M. Dohler, T. Mahmoodi, M. A. Lema, M. Condoluci, F. Sardis, K. Antonakoglou, and H. Aghvami. Internet of skills, where robotics meets AI, 5G and the Tactile Internet. In *2017 European Conference on Networks and Communications (EuCNC)*, pages 1–5. IEEE, 2017.

[DMLF12] R. Datta, N. Michailow, M. Lentmaier, and G. Fettweis. GFDM Interference Cancellation for Flexible Cognitive Radio PHY Design. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, Sep. 2012.

[DMR00] A. Demir, A. Mehrotra, and J. Roychowdhury. Phase noise in oscillators: A unifying theory and numerical methods for characterization. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(5):655–674, 2000.

[DMR19] W. D. Dias, L. L. Mendes, and J. J. P. C. Rodrigues. Low complexity GFDM receiver for Frequency-Selective Channels. *IEEE Communications Letters*, pages 1–1, 2019.

[DNM+18] M. Danneberg, A. Nimr, N. Michailow, S. Ehsanfar, M. Matthé, A. Martinez, D. Zhang, and G. Fettweis. Universal waveforms processor. In *2018 European Conference on Networks and Communications (EuCNC)*, pages 357–362, 2018.

[Far11] B. Farhang-Boroujeny. OFDM Versus Filter Bank Multicarrier. *IEEE Signal Processing Magazine*, 28(3):92–112, May 2011.

[FBA94] G. Fettweis, A. S. Bahai, and K. Anvari. On multi-carrier code division multiple access (mc-cdma) modem design. In *Proceedings of IEEE Vehicular Technology Conference (VTC)*, pages 1670–1674. IEEE, 1994.

[FBM07] F. Farrukh, S. Baig, and M. Mughal. Performance comparison of DFT-OFDM and wavelet-OFDM with zero-forcing equalizer for FIR channel equalization. In *2007 International Conference on Electrical Engineering*, pages 1–5. IEEE, 2007.

[FC12] R. Fiorelli and M. Colombo. ST7580 power line communication system-on-chip design guide. *Application note AN4068, STMicroelectronics*, 2012.

[Fet14] G. P. Fettweis. The tactile internet: Applications and challenges. *IEEE Vehicular Technology Magazine*, 9(1):64–70, 2014.

[FGZ+17] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Y. Li, and X. Chen. Faster-than-nyquist signaling: An overview. *IEEE Access*, 5:1925–1940, 2017.

[FKB09] G. Fettweis, M. Krondorf, and S. Bittner. GFDM - Generalized Frequency Division Multiplexing. In *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, pages 1–4, April 2009.

[FMD16] A. Farhang, N. Marchetti, and L. E. Doyle. Low-Complexity Modem Design for GFDM. *IEEE Transactions on Signal Processing*, 64(6):1507–1518, March 2016.

[FSCH04] M. Feng, S.-C. Shen, D. C. Caruth, and J.-J. Huang. Device technologies for RF front-end circuits in next-generation wireless communications. *Proceedings of the IEEE*, 92(2):354–375, 2004.

[G+15] I. Gaspar et al. Frequency-Shift Offset-QAM for GFDM. *IEEE Communications Letters*, 19(8):1454–1457, Aug 2015.

[GMM+14] I. Gaspar, L. Mendes, M. Matthé, N. Michailow, A. Festag, and G. Fettweis. LTE-compatible 5G PHY based on generalized frequency division multiplexing. In *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, pages 209–213, Aug 2014.

[GMM+15] I. Gaspar, M. Matthé, N. Michailow, L. Leonel Mendes, D. Zhang, and G. Fettweis. Frequency-Shift Offset-QAM for GFDM. *IEEE Communications Letters*, 19(8):1454–1457, Aug 2015.

[GMN+13] I. Gaspar, N. Michailow, A. Navarro, E. Ohlmer, S. Krone, and G. Fettweis. Low Complexity GFDM Receiver Based on Sparse Frequency Domain Processing. In *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pages 1–6, June 2013.

[GMP17] D. Gaspar, L. Mendes, and T. Pimenta. GFDM BER Under Synchronization Errors. *IEEE Communications Letters*, 21(8):1743–1746, Aug 2017.

[GNP06] W. A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: Half a century of research. *Signal processing*, 86(4):639–697, 2006.

[Gol05] A. Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[GRM+10] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas. LTE-advanced: next-generation wireless broadband technology. *IEEE wireless communications*, 17(3):10–22, 2010.

[GWX+11] X. Gao, W. Wang, X. Xia, E. K. S. Au, and X. You. Cyclic Prefixed OQAM-OFDM and its Application to Single-Carrier FDMA. *IEEE Transactions on Communications*, 59(5):1467–1480, May 2011.

[H+17] R. Hadani, , et al. Orthogonal time frequency space modulation. In *IEEE WCNC*, pages 1–6. IEEE, 2017.

[Hay94] S. Haykin. *An introduction to analog and digital communication*. John Wiley, 1994.

[HL91] J. F. Helard and B. Le Floch. Trellis coded orthogonal frequency division multiplexing for digital video transmission. In *IEEE Global Telecommunications Conference GLOBECOM '91: Countdown to the New Millennium. Conference Record*, pages 785–791 vol.2, Dec 1991.

[HLH17] M. Huemer, O. Lang, and C. Hofbauer. Component-wise conditionally unbiased widely linear MMSE estimation. *Signal Processing*, 133:227–239, 2017.

[HP99] S. Hara and R. Prasad. Design and performance of multicarrier CDMA system in frequency-selective Rayleigh fading channels. *IEEE transactions on vehicular technology*, 48(5):1584–1595, 1999.

[HP17] R. P. Hebbar and P. G. Poddar. Generalized frequency division multiplexing for acoustic communication in underwater systems. In *2017 International Conference on Circuits, Controls, and Communications (CCUBE)*, pages 86–90, Dec 2017.

[HSL17] S. Han, Y. Sung, and Y. H. Lee. Filter design for generalized frequency-division multiplexing. *IEEE Transactions on Signal Processing*, 65(7):1644–1659, April 2017.

[HV05] B. Hassibi and H. Vikalo. On the sphere-decoding algorithm I. Expected complexity. *IEEE transactions on signal processing*, 53(8):2806–2818, 2005.

[ITU06] CCIT/ITU-T 50 Years of Excellence. *International Telecommunication Union*, 2006.

[ITU16] ITU Radio Regulations. *International Telecommunication Union*, 2016.

[Jan94] A. Janssen. Signal analytic proofs of two basic results on lattice expansions. *Applied and Computational Harmonic Analysis*, 1(4):350–354, 1994.

[JBS93a] P. Jung, P. W. Baier, and A. Steil. Advantages of CDMA and spread spectrum techniques over FDMA and TDMA in cellular mobile radio applications. *IEEE Transactions on Vehicular Technology*, 42(3):357–364, 1993.

[JBS93b] P. Jung, P. W. Baier, and A. Steil. Advantages of CDMA and spread spectrum techniques over FDMA and TDMA in cellular mobile radio applications. *IEEE Transactions on Vehicular Technology*, 42(3):357–364, Aug 1993.

[JE97] D. A. Johns and D. Essig. Integrated circuits for data transmission over twisted-pair channels. *IEEE Journal of Solid-State Circuits*, 32(3):398–406, 1997.

[JYL+99] R. Joshi, P. Yang, H.-C. Liu, K. Kindsfater, K. Cameron, D. Gee, H. Vu, G. Gorman, S. Tsai, A. Hung, et al. A 52 Mb/s universal DSL transceiver IC. In *1999 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC. First Edition (Cat. No. 99CH36278)*, pages 250–251. IEEE, 1999.

[KAC90] S. Kasturia, J. T. Aslanis, and J. M. Cioffi. Vector coding for partial response channels. *IEEE Transactions on Information Theory*, 36(4):741–762, July 1990.

[Kay93] S. M. Kay. *Fundamentals of statistical signal processing.* Prentice Hall PTR, 1993.

[KB13] P. Kryszkiewicz and H. Bogucka. Out-of-band power reduction in NC-OFDM with optimized cancellation carriers selection. *IEEE communications letters*, 17(10):1901–1904, 2013.

[KH00] D. Kivanc and Hui Liu. Subcarrier allocation and power control for OFDMA. In *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers (Cat. No.00CH37154)*, volume 1, pages 147–151 vol.1, Oct 2000.

[KL14] E. D. Knapp and J. T. Langill. *Industrial Network Security: Securing critical infrastructure networks for smart grid, SCADA, and other Industrial Control Systems.* Syngress, 2014.

[KMB17] A. Kumar, M. Magarini, and S. Bregni. Improving GFDM Symbol Error Rate Performance using "Better than Nyquist" Pulse Shaping Filters. *IEEE Latin America Transactions*, 15(7):1244–1249, 2017.

[KSP+02] J.-P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen. A stochastic MIMO radio channel model with experimental validation. *IEEE Journal on selected areas in Communications*, 20(6):1211–1226, 2002.

[KST04] R. Koetter, A. C. Singer, and M. Tuchler. Turbo equalization. *IEEE signal processing magazine*, 21(1):67–80, 2004.

[LAB95] B. Le Floch, M. Alard, and C. Berrou. Coded orthogonal frequency division multiplex [TV broadcasting]. *Proceedings of the IEEE*, 83(6):982–996, June 1995.

[LDL19] K. Liu, W. Deng, and Y. Liu. Theoretical Analysis of the Peak-to-Average Power Ratio and Optimal Pulse Shaping Filter Design for GFDM Systems. *IEEE Transactions on Signal Processing*, 67(13):3455–3470, July 2019.

[LHJ17] H. Lee, M. Han, and J. Jang. Trends in the Telepresence Technologies and Services Beyond 5G. *Electronics and Telecommunications Trends*, 32(5):20–29, 2017.

[LHS18] C. Lin, Y. Huang, and B. Su. Prototype Filter Design in GFDM Systems in Presence of PA Nonlinearity. In *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pages 1–5, Nov 2018.

[LK17] B. Lim and Y. Ko. SIR Analysis of OFDM and GFDM Waveforms With Timing Offset, CFO, and Phase Noise. *IEEE Transactions on Wireless Communications*, 16(10):6979–6990, Oct 2017.

[LK19] B. Lim and Y. Ko. Multiuser Interference Cancellation for GFDM With Timing and Frequency Offsets. *IEEE Transactions on Communications*, 67(6):4337–4349, June 2019.

[LLL+10] Q. Li, G. Li, W. Lee, M.-i. Lee, D. Mazzarese, B. Clerckx, and Z. Li. MIMO techniques in WiMAX and LTE: a feature overview. *IEEE Communications magazine*, 48(5):86–92, 2010.

[LNF19] Z. Li, A. Nimr, and G. Fettweis. Implementation and Performance Measurement of Flexible Radix-2 GFDM Modem. In *2019 IEEE 2nd 5G World Forum (5GWF)*, pages 130–134, 2019.

[LQE+17] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo. Nonorthogonal multiple access for 5G and beyond. *Proceedings of the IEEE*, 105(12):2347–2381, Dec 2017.

[LS08] H. Lin and P. Siohan. A new transceiver system for the OFDM/OQAM modulation with Cyclic Prefix. In *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–5, Sep. 2008.

[LS+13] S. G. Lee, P. Sathikh, et al. A framework for effective human-to-machine communication for artificial interactive systems. In *DS 75-7: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 7: Human Behaviour in Design, Seoul, Korea, 19-22.08. 2013*, pages 367–376, 2013.

[LS14] H. Lin and P. Siohan. Multi-carrier modulation analysis and WCP-COQAM proposal. *EURASIP Journal on Advances in Signal Processing*, 2014(1):79, May 2014.

[LSH+17] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei. 5G new radio: Waveform, frame structure, multiple access, and initial access. *IEEE communications magazine*, 55(6):64–71, 2017.

[LSL03] E. G. Larsson, P. Stoica, and J. Li. Orthogonal space-time block codes: Maximum likelihood detection for unknown channels and unstructured interferences. *IEEE Transactions on Signal Processing*, 51(2):362–372, 2003.

[Lu17] Y. Lu. Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, 6:1–10, 2017.

[LW16] D. W. Lin and P. Wang. On the Configuration-Dependent Singularity of GFDM Pulse-Shaping Filter Banks. *IEEE Communications Letters*, 20(10):1975–1978, Oct 2016.

[LYY10] T. Liu, C. Yang, and L. Yang. Joint Transmitter–Receiver Frequency-Domain Equalization in Generalized Multicarrier Code-Division Multiplexing Systems. *IEEE Transactions on Vehicular Technology*, 59(8):3786–3797, Oct 2010.

[LZZ+19] F. Li, K. Zheng, L. Zhao, H. Zhao, and Y. Li. Design and Performance of a Novel Interference-Free GFDM Transceiver With Dual Filter. *IEEE Transactions on Vehicular Technology*, 68(5):4695–4706, May 2019.

[Maz75] J. E. Mazo. Faster-than-nyquist signaling. *The Bell System Technical Journal*, 54(8):1451–1462, 1975.

[MBT19] A. Mohammadian, M. Baghani, and C. Tellambura. Optimal Power Allocation of GFDM Secondary Links With Power Amplifier Nonlinearity and ACI. *IEEE Wireless Communications Letters*, 8(1):93–96, Feb 2019.

[MCC98] Minsheng Wang, A. K. Chan, and C. K. Chui. Linear frequency-modulated signal detection using radon-ambiguity transform. *IEEE Transactions on Signal Processing*, 46(3):571–586, 1998.

[Men96] M. Mendelsohn. The media and interpersonal communications: The priming of issues, leaders, and party identification. *The Journal of Politics*, 58(1):112–125, 1996.

[MF16] M. Matthé and G. Fettweis. Conjugate root offset-QAM for orthogonal multicarrier transmission. *EURASIP Journal on Advances in Signal Processing*, 2016(1):41, Apr 2016.

[MGZF15] M. Matthe, I. Gaspar, D. Zhang, and G. Fettweis. Reduced Complexity Calculation of LMMSE Filter Coefficients for GFDM. In *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, pages 1–2, Sep. 2015.

[MH81] K. Murota and K. Hirade. GMSK modulation for digital mobile radio telephony. *IEEE Transactions on communications*, 29(7):1044–1050, 1981.

[MKLF12] N. Michailow, S. Krone, M. Lentmaier, and G. Fettweis. Bit error rate performance of generalized frequency division multiplexing. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, Sep. 2012.

[MLG06a] H. G. Myung, J. Lim, and D. J. Goodman. Peak-to-average power ratio of single carrier FDMA signals with pulse shaping. In *2006 IEEE 17th international symposium on personal, indoor and mobile radio communications*, pages 1–5. IEEE, 2006.

[MLG06b] H. G. Myung, J. Lim, and D. J. Goodman. Single carrier FDMA for uplink wireless transmission. *IEEE Vehicular Technology Magazine*, 1(3):30–38, Sep. 2006.

[MMF97] H. Meyr, M. Moeneclaey, and S. Fechtel. *Digital communication receivers: synchronization, channel estimation, and signal processing.* John Wiley & Sons, Inc., 1997.

[MMF14a] M. Matthé, L. L. Mendes, and G. Fettweis. Generalized Frequency Division Multiplexing in a Gabor Transform Setting. *IEEE Communications Letters*, 18(8):1379–1382, Aug 2014.

[MMF14b] M. Matthe, L. L. Mendes, and G. Fettweis. Space-Time Coding for Generalized Frequency Division Multiplexing. In *European Wireless 2014; 20th European Wireless Conference*, pages 1–5, May 2014.

[MMF15] M. Matthé, L. L. Mendes, and G. Fettweis. Asynchronous multi-user uplink transmission with generalized frequency division multiplexing. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 2269–2275, June 2015.

[MMG+14] N. Michailow, M. Matthé, I. S. Gaspar, A. N. Caldevilla, L. L. Mendes, A. Festag, and G. Fettweis. Generalized Frequency Division Multiplexing for 5th Generation Cellular Networks. *IEEE Transactions on Communications*, 62(9):3045–3061, Sep. 2014.

[MMG+15] M. Matthe, L. L. Mendes, I. Gaspar, N. Michailow, D. Zhang, and G. Fettweis. Multi-user time-reversal STC-GFDMA for future wireless networks. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):132, May 2015.

[MMG+16] M. Matthé, L. Mendes, I. Gaspar, N. Michailow, D. Zhang, and G. Fettweis. Precoded GFDM transceiver with low complexity time domain processing. *EURASIP Journal on Wireless Communications and Networking*, 2016(1):138, May 2016.

[MMGF14] M. Matthé, N. Michailow, I. Gaspar, and G. Fettweis. Influence of pulse shaping on bit error rate performance and out of band radiation of Generalized Frequency Division Multiplexing. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 43–48, June 2014.

[MMM+15] N. Michailow, L. Mendes, M. Matthé, I. Gaspar, A. Festag, and G. Fettweis. Robust WHT-GFDM for the Next Generation of Wireless Networks. *IEEE Communications Letters*, 19(1):106–109, Jan 2015.

[Mor10] L. Morency. Modeling human communication dynamics [social sciences]. *IEEE Signal Processing Magazine*, 27(5):112–116, 2010.

[MPFBH92] M. Mouly, M.-B. Pautet, and T. Foreword By-Haug. *The GSM system for mobile communications*. Telecom publishing, 1992.

[MW86] H. Ma and J. Wolf. On tail biting convolutional codes. *IEEE Transactions on Communications*, 34(2):104–111, February 1986.

[MWQP19] Y. Ma, X. Wang, Z. Quan, and H. V. Poor. Data-driven measurement of receiver sensitivity in wireless communication systems. *IEEE Transactions on Communications*, 67(5):3665–3676, 2019.

[MZF16] M. Matthé, D. Zhang, and G. Fettweis. Sphere-decoding aided SIC for MIMO-GFDM: Coded performance analysis. In *2016 International Symposium on Wireless Communication Systems (ISWCS)*, pages 165–169. IEEE, 2016.

[MZF18] M. Matthé, D. Zhang, and G. Fettweis. Low-Complexity Iterative MMSE-PIC Detection for MIMO-GFDM. *IEEE Transactions on Communications*, 66(4):1467–1480, 2018.

[MZG+02] B. Muquet, Zhengdao Wang, G. B. Giannakis, M. de Courville, and P. Duhamel. Cyclic prefixing or zero padding for wireless multicarrier transmissions? *IEEE Transactions on Communications*, 50(12):2136–2148, Dec 2002.

[NCF18] A. Nimr, M. Chafii, and G. P. Fettweis. Unified Low Complexity Radix-2 Architectures for Time and Frequency-Domain GFDM Modem. *IEEE Circuits and Systems Magazine*, 18(4):18–31, Fourthquarter 2018.

[NCF19a] A. Nimr, M. Chafii, and G. Fettweis. Low-Complexity Transceiver for GFDM systems with Partially Allocated Subcarriers. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2019.

[NCF19b] A. Nimr, M. Chafii, and G. Fettweis. Practical GFDM-based linear receivers. In *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, pages 1–6. VDE, 2019.

[NCF19c] A. Nimr, M. Chafii, and G. Fettweis. Precoded-OFDM within GFDM Framework. In *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, pages 1–5, 2019.

[NCMF18] A. Nimr, M. Chafii, M. Matthe, and G. Fettweis. Extended GFDM Framework: OTFS and GFDM Comparison. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2018.

[NEM+19] A. Nimr, S. Ehsanfar, N. Michailow, M. Danneberg, D. Zhang, H. D. Rodrigues, L. L. Mendes, and G. Fettweis. Generalized frequency division multiplexing: A flexible multicarrier waveform. In *Multiple Access Techniques for 5G Wireless Networks and Beyond*, pages 93–163. Springer, 2019.

[NLCF20] A. Nimr, Z. Li, M. Chafii, and G. Fettweis. *Generalized Frequency Division Multiplexing: Unified Multicarrier Framework*, pages 63–82. Wiley, 2020.

[NLF07] B. Ng, C. Lam, and D. Falconer. Turbo frequency domain equalization for single-carrier broadband wireless systems. *IEEE Transactions on Wireless Communications*, 6(2):759–767, Feb 2007.

[NLLNK09] H. Nguyen-Le, T. Le-Ngoc, and C. C. Ko. RLS-based joint estimation and tracking of channel response, sampling, and carrier frequency offsets for OFDM. *IEEE Transactions on Broadcasting*, 55(1):84–94, 2009.

[NLM13] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. The multicell multiuser mimo uplink with very large antenna arrays and a finite-dimensional channel. *IEEE Transactions on Communications*, 61(6):2350–2361, 2013.

[NLZ+19] Z. Na, J. Lv, M. Zhang, B. Peng, M. Xiong, and M. Guan. GFDM Based Wireless Powered Communication for Cooperative Relay System. *IEEE Access*, 7:50971–50979, 2019.

[NM05] M. J. Neely and E. Modiano. Capacity and delay tradeoffs for ad hoc mobile networks. *IEEE Transactions on Information Theory*, 51(6):1917–1937, 2005.

[NMZF17] A. Nimr, M. Matthé, D. Zhang, and G. Fettweis. Optimal Radix-2 FFT Compatible Filters for GFDM. *IEEE Communications Letters*, 21(7):1497–1500, 2017.

[NZMF17] A. Nimr, D. Zhang, A. B. Martinez, and G. P. Fettweis. A study on the physical layer performance of GFDM for high throughput wireless communication. In *EUSIPCO*, pages 638–642. IEEE, 2017.

[OBC16] E. Ozturk, E. Basar, and H. A. Cirpan. Generalized frequency division multiplexing with index modulation. In *2016 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6, Dec 2016.

[OMH+09] P. Odling, T. Magesacher, S. Host, P. O. Borjesson, M. Berg, and E. Areizaga. The fourth generation broadband concept. *IEEE Communications Magazine*, 47(1):62–69, 2009.

[OP98] T. Ojanpera and R. Prasad. *Wideband CDMA For Third Generation Mobile Communications: Universal Personal Communications.* Artech House, Inc., 1998.

[Pal88] J. C. Palais. *Fiber optic communications.* Prentice Hall Englewood Cliffs, 1988.

[PGNB04] 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*, 92(2):198–218, 2004.

[PR80] A. Peled and A. Ruiz. Frequency domain data transmission using reduced computational complexity algorithms. In *ICASSP '80. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 964–967, April 1980.

[Pra04] R. Prasad. *OFDM for wireless communications systems.* Artech House, 2004.

[PSM82] R. Pickholtz, D. Schilling, and L. Milstein. Theory of spread-spectrum communications - a tutorial. *IEEE Transactions on Communications*, 30(5):855–884, May 1982.

[Rap02] T. S. Rappaport. Wireless communications–principles and practice, (the book end). *Microwave Journal*, 45(12):128–129, 2002.

[RASPK11] S. Rajagopal, S. Abu-Surra, Z. Pi, and F. Khan. Antenna array design for multi-Gbps mmWave mobile broadband communication. In *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011*, pages 1–6. IEEE, 2011.

[Rei98] U. Reimers. Digital video broadcasting. *IEEE Communications Magazine*, 36(6):104–110, June 1998.

[SA11] A. Sahin and H. Arslan. Edge windowing for ofdm based systems. *IEEE Communications Letters*, 15(11):1208–1211, 2011.

[SAD+16] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis. 5G-enabled tactile internet. *IEEE Journal on Selected Areas in Communications*, 34(3):460–473, 2016.

[SAIZ18] S. A. A. Shah, E. Ahmed, M. Imran, and S. Zeadally. 5G for vehicular communications. *IEEE Communications Magazine*, 56(1):111–117, 2018.

[Sal67] B. Saltzberg. Performance of an Efficient Parallel Data Transmission System. *IEEE Transactions on Communication Technology*, 15(6):805–811, December 1967.

[SAZ07] S. R. Saunders and A. Aragón-Zavala. *Antennas and propagation for wireless communication systems*. John Wiley & Sons, 2007.

[SB08] M. Schwartz and C. Batchelor. The origins of carrier multiplexing: Major George Owen Squier and AT T. *IEEE Communications Magazine*, 46(5):20–24, May 2008.

[SCC93] K. Sistanizadeh, P. Chow, and J. Cioffi. Multi-tone transmission for asymmetric digital subscriber lines (ADSL). In *Proceedings of ICC '93 - IEEE International Conference on Communications*, volume 2, pages 756–760 vol.2, May 1993.

[Sch04] P. Schniter. Low-complexity equalization of OFDM in doubly selective channels. *IEEE Transactions on Signal processing*, 52(4):1002–1011, 2004.

[SFS11] C. Studer, S. Fateh, and D. Seethaler. ASIC Implementation of Soft-Input Soft-Output MIMO Detection Using MMSE Parallel Interference Cancellation. *IEEE Journal of Solid-State Circuits*, 46(7):1754–1765, 2011.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[SKJ95] H. Sari, G. Karam, and I. Jeanclaude. Transmission techniques for digital terrestrial TV broadcasting. *IEEE Communications Magazine*, 33(2):100–109, Feb 1995.

[SLK97] H. Sari, Y. Levy, and G. Karam. An analysis of orthogonal frequency-division multiple access. In *GLOBECOM 97. IEEE Global Telecommunications Conference. Conference Record*, volume 3, pages 1635–1639 vol.3, Nov 1997.

[SMF05] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen. Extension of the ITU channel models for wideband (OFDM) systems. In *VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference, 2005.*, volume 1, pages 392–396, 2005.

[SOSF16] Z. Sharifian, M. J. Omidi, H. Saeedi-Sourck, and A. Farhang. Linear Precoding for PAPR Reduction of GFDMA. *IEEE Wireless Communications Letters*, 5(5):520–523, Oct 2016.

[SPSH04] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt. An introduction to the multi-user MIMO downlink. *IEEE communications Magazine*, 42(10):60–67, 2004.

[Spu00] C. E. Spurgeon. *Ethernet: the definitive guide*. " O'Reilly Media, Inc.", 2000.

[SSA16] R. K. Saha, P. Saengudomlert, and C. Aswakul. Evolution toward 5G mobile networks-A survey on enabling technologies. *Engineering Journal*, 20(1):87–119, 2016.

[ST11] D. Senaratne and C. Tellambura. Beamforming for space division duplexing. In *2011 IEEE International Conference on Communications (ICC)*, pages 1–5, June 2011.

[Sta95] W. Stallings. *ISDN and Broadband ISDN with Frame Relay and ATM*. Prentice-Hall, 1995.

[SU09] A. Soysal and S. Ulukus. Optimality of beamforming in fading MIMO multiple access channels. *IEEE Transactions on Communications*, 57(4):1171–1183, 2009.

[TA18] M. Towliat and S. M. J. Asgari Tabatabaee. GFDM Interference Mitigation Without Noise Enhancement. *IEEE Communications Letters*, 22(5):1042–1045, May 2018.

[TCVdP+05] J. Tubbax, B. Come, L. Van der Perre, S. Donnay, M. Engels, H. De Man, and M. Moonen. Compensation of IQ imbalance and phase noise in OFDM systems. *IEEE Transactions on Wireless Communications*, 4(3):872–877, 2005.

[TD18] S. Tiwari and S. S. Das. Low-Complexity Joint-MMSE GFDM Receiver. *IEEE Transactions on Communications*, 66(4):1661–1674, AFpril 2018.

[TG13] A. M. Tonello and M. Girotto. Cyclic block FMT modulation for broadband power line communications. In *2013 IEEE 17th International Symposium on Power Line Communications and Its Applications*, pages 247–251, March 2013.

[THX+17] N. Tang, S. He, C. Xue, Y. Huang, and L. Yang. IQ Imbalance Compensation for Generalized Frequency Division Multiplexing Systems. *IEEE Wireless Communications Letters*, 6(4):422–425, Aug 2017.

[Tom98] L. Tomba. On the effect of Wiener phase noise in OFDM systems. *IEEE Transactions on communications*, 46(5):580–583, 1998.

[TPZ05] S. C. Thompson, J. G. Proakis, and J. R. Zeidler. The effectiveness of signal clipping for PAPR and total degradation reduction in OFDM systems. In *GLOBECOM'05. IEEE Global Telecommunications Conference, 2005.*, volume 5, pages 5–pp. IEEE, 2005.

[TSB15] S. Tiwari, S. Sekhar Das, and K. K. Bandyopadhyay. Precoded generalised frequency division multiplexing system to combat inter-carrier interference: performance analysis. *IET Communications*, 9(15):1829–1841, 2015.

[TSC18] C. Tai, B. Su, and P. Chen. Optimal filter design for GFDM that minimizes PAPR under performance constraints. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, April 2018.

[vAM+99] R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford. New high-rate wireless LAN standards. *IEEE Communications Magazine*, 37(12):82–88, Dec 1999.

[vB09] J. van de Beek and F. Berggren. N-continuous OFDM. *IEEE Communications Letters*, 13(1):1–3, January 2009.

[VT14] A. Varghese and D. Tandur. Wireless requirements and challenges in industry 4.0. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pages 634–638. IEEE, 2014.

[VWS+13] V. Vakilian, T. Wild, F. Schaich, S. ten Brink, and J. Frigon. Universal-filtered multi-carrier technique for wireless systems beyond LTE. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 223–228, Dec 2013.

[WE71] S. Weinstein and P. Ebert. Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform. *IEEE Transactions on Communication Technology*, 19(5):628–634, October 1971.

[Wei09] S. B. Weinstein. The history of orthogonal frequency-division multiplexing [History of Communications]. *IEEE Communications Magazine*, 47(11):26–35, November 2009.

[WFGV98] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela. V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel. In *1998 URSI international symposium on signals, systems, and electronics. Conference proceedings (Cat. No. 98EX167)*, pages 295–300. IEEE, 1998.

[Wic97] I. J. Wickelgren. The facts about FireWire [serial communication bus]. *IEEE spectrum*, 34(4):19–25, 1997.

[WJK+14] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Ten Brink, I. Gaspar, N. Michailow, A. Festag, et al. 5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications. *IEEE Communications Magazine*, 52(2):97–105, 2014.

[WMSL18] Z. Wang, L. Mei, X. Sha, and V. C. M. Leung. BER Analysis of WFRFT Precoded OFDM and GFDM Waveforms With an Integer Time Offset. *IEEE Transactions on Vehicular Technology*, 67(10):9097–9111, Oct 2018.

[WR90] J. Wexler and S. Raz. Discrete Gabor expansions. *Signal Processing*, 21(3):207 – 220, 1990.

[WS73] T. Walzman and M. Schwartz. Automatic equalization using the discrete frequency domain. *IEEE Transactions on Information Theory*, 19(1):59–68, January 1973.

[WS18] H. Wang and R. Song. Low Complexity ZF Receiver Design for Multi-User GFDMA Uplink Systems. *IEEE Access*, 6:28661–28667, 2018.

[WSW04] C. Waldschmidt, S. Schulteis, and W. Wiesbeck. Complete RF system model for analysis of compact MIMO arrays. *IEEE Transactions on Vehicular Technology*, 53(3):579–586, 2004.

[WXXL16] P. Wei, X. Xia, Y. Xiao, and S. Li. Fast DGT-Based Receivers for GFDM in Broadband Channels. *IEEE Transactions on Communications*, 64(10):4331–4345, Oct 2016.

[ZFF17] D. Zhang, A. Festag, and G. P. Fettweis. Performance of generalized frequency division multiplexing based physical layer in vehicular communications. *IEEE Transactions on Vehicular Technology*, 66(11):9809–9824, Nov 2017.

[ZIX+18] L. Zhang, A. Ijaz, P. Xiao, M. M. Molu, and R. Tafazolli. Filtered OFDM Systems, Algorithms, and Performance Analysis for 5G and Beyond. *IEEE Transactions on Communications*, 66(3):1205–1218, 2018.

[ZK02] X. Zhu and J. M. Kahn. Free-space optical communication through atmospheric turbulence channels. *IEEE Transactions on communications*, 50(8):1293–1300, 2002.

[ZMMF17] D. Zhang, M. Matthe, L. L. Mendes, and G. Fettweis. A study on the link level performance of advanced multicarrier waveforms under MIMO wireless

communication channels. *IEEE Transactions on Wireless Communications*, 16(4):2350–2365, 2017.

[ZMSR16] R. Zayani, Y. Medjahdi, H. Shaiek, and D. Roviras. WOLA-OFDM: A Potential Candidate for Asynchronous 5G. In *2016 IEEE Globecom Workshops (GC Wkshps)*, pages 1–5, Dec 2016.