Spectrally and Energy Efficient Wireless Communications: Signal and System Design, Mathematical Modelling and Optimisation

Xinyue Liu

A thesis submitted for the degree of Doctor of Philosophy

of

University College London

Information & Communication Systems Research Group

Department of Electronic and Electrical Engineering

University College London

February, 2023

Declaration

I, Xinyue Liu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Xinyue Liu

Abstract

This thesis explores engineering studies and designs aiming to meeting the requirements of enhancing capacity and energy efficiency for next generation communication networks. Challenges of spectrum scarcity and energy constraints are addressed and new technologies are proposed, analytically investigated and examined.

The thesis commences by reviewing studies on spectrally and energy-efficient techniques, with a special focus on non-orthogonal multicarrier modulation, particularly spectrally efficient frequency division multiplexing (SEFDM). Rigorous theoretical and mathematical modelling studies of SEFDM are presented. Moreover, to address the potential application of SEFDM under the 5th generation new radio (5G NR) heterogeneous numerologies, simulation-based studies of SEFDM coexisting with orthogonal frequency division multiplexing (OFDM) are conducted. New signal formats and corresponding transceiver structure are designed, using a Hilbert transform filter pair for shaping pulses. Detailed modelling and numerical investigations show that the proposed signal doubles spectral efficiency without performance degradation, with studies of two signal formats; uncoded narrow-band internet of things (NB-IoT) signals and unframed turbo coded multi-carrier signals. The thesis also considers using constellation shaping techniques and SEFDM for capacity enhancement in 5G system. Probabilistic shaping for SEFDM is proposed and modelled to show both transmission energy reduction and bandwidth saving with advantageous flexibility for data rate adaptation. Expanding on constellation shaping to improve performance further, a comparative study of multidimensional modulation techniques is carried out. A four-dimensional signal, with better noise immunity is investigated, for which metaheuristic optimisation algorithms are studied, developed, and conducted to optimise bit-to-symbol mapping. Finally, a specially designed machine learning technique for signal and system design in physical layer communications is proposed, utilising the application of autoencoder-based end-to-end learning. Multidimensional signal modulation with multidimensional constellation shaping is proposed and optimised by using machine learning techniques, demonstrating significant improvement in spectral and energy efficiencies.

Impact Statement

This research contributes to the advancement of communication networks by addressing the challenges of spectrum scarcity and energy constraints. Through rigorous theoretical and mathematical modelling, the thesis presents various spectrally and energy-efficient techniques with a special focus on signal and system design and optimisation. The proposed designs are expected to have an impact on future cellular networks (beyond 5G and 6G), internet of things (IoT), satellite broadcast systems (DVB-S2, DVB-S2X) and millimetre wave communication systems.

The analysis and techniques introduced in this thesis are shown to significantly improve the spectral and energy efficiency for wireless communication systems in different use scenarios through various techniques, including pulse shape filtering, constellation shaping, non-orthogonal signalling, and optimisation via metaheuristic algorithms as well as end-to-end machine learning. Such innovations in communication systems design, mathematical modelling, software simulation, hardware implementation, and experimental verification will pave the way for new research and possible implementation in future commercial communication systems.

The work funded by Cisco led to the research on system transmission for next-generation IoT. The proposed signal design and corresponding transceiver structure using Hilbert transform filter pair for shaping pulses are shown to have overall performance improvement with increased spectral efficiency. This particular work, having considered full system design and implementation, is expected to lead to further research and development efforts in the area of high frequency, high bit rate and spectrally efficient systems. Interest in developing this work into future local area network (LAN) standards has already been discussed by industry.

Collaboration with and funding from NEC Corporation-Japan and NEC laboratories Europe (NLE) led to the work on signal design via constellation shaping and multidimensional modulations.

The transmitter, detector and channel estimator designs, were mainly studied for fixed wireless systems; however, they may well be utilised in various wired and wireless systems such as 5G and beyond, DVB-S2, VLC, optical fibre systems and fast-DSL. The successful application of SEFDM in these systems makes it a serious signal contender in future networks and paves the way for new research and possible implementation in future commercial communication systems.

The promising implementation of spectrally and energy efficient techniques discussed further establishes SEFDM as a strong signal candidate for future networks, including 6G, opening up avenues for further research and development, and potential integration in upcoming commercial communication systems.

In this research, a new application of the use of machine learning and autoencoding led to the design of a bespoke autoencoder-based communication system, which uses deep learning techniques to optimise system performance through the optimisation of information theoretic metrics. This work will have an impact on the development of new approaches to optimise communication systems and to integrate the use of machine learning in communication links and their design. The impact will go beyond simple links and it is hoped that the techniques developed will have applications in wider networks, both wired and wireless, fixed and mobile.

Finally, the comparative studies of the application of four heuristic and meta-heuristic optimisation methods to the complex problem of choosing an optimum binary mapping will hopefully be useful not only to communications engineers but also to operational research specialists aiming to apply heuristics to various complex problems ranging from technology to finance.

Much of the research and experimental work presented in this thesis is of transformational nature, taking mathematical and theoretical concepts into engineering designs. The work was well received by the communications engineering community when submitted and presented at six leading IEEE international conferences. The work has also resulted in two patents with industrial exploitation potential.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Professor Izzat Darwazeh for his great supervision, generous support and friendship throughout my PhD and my MSc. Izzat's knowledge and experience have been invaluable in guiding my academic and personal development. All the early morning meetings and the conference travels are unforgettable memories I hold dear and will always cherish. The journey would not have been possible without your guidance and encouragement and I am forever grateful for your belief in me.

I would like to thank Mr Eisaku Sasaki from NEC Corporation, Japan and Dr Nader Zein from NEC Laboratories Europe, for the funding the research project. I am grateful for their fruitful discussions, unwavering support and invaluable feedback during the project. The collaboration leads to two proceedings and two patents which are included in Chapters 4, 5 and 6.

The research was carried out in the EEE department at University College London. I am grateful that I've been surrounded by brilliant researchers and supporting friends. I would like to offer my gratitude to my second supervisor Dr. Ryan Gremmenos for his guidance and support. I would like to offer my gratitude to Dr Tongyang Xu, for the technical discussions and collaborative work presented in Chapter 2. I would like to thank Dr Hedaia Ghannam who finished her PhD when I first joined the group, for all the suggestions and help in my research work and for being a shining role model and supportive friend. To my dearest friend, Dr Amany Kassem, for being my incredible friend and ”twin sister” in the lab. The memories we have shared are etched in my heart and will forever motivate me; I also want to thank Dr Temitope Odedeyi for all the inspiring words and fun discussions.

I am also grateful to my lab friends at UCL; to Maryam Habibollahi for providing supportive and stimulating living experience throughout my study journey; to Afroditi Papadaki who offers me sincere friendship and empowering energy since I first came to London; to dear Nagham Osman, for her company for the late night work in the lab. To my colleagues from the MPEB 708 office (Dr Silvia Rossi, Dr Sephora Majiheurem, Maria, Pedro, Fan, Aya, Alan and all the others) and not (Joshua and Noora), thank you for all the laughter during coffee and lunch.

I would like to express my immense gratitude to my family. To my mother, thank you for always supporting me and encouraging me to stay curious and keep learning. To my father, thank you for sharing your love of art and literature and being my best friend and greatest supporter. To my grandmothers, for setting incredible examples of independent and strong women and for your endless love and support. To my grandfathers for generously sharing your knowledge and life experience with me and for being pillars of strength in my life. Thank you for your love and belief in me. I feel blessed that you are always my shelter in life.

Special thanks to my dearest Nour Khaled and Imogen May. Thank you for being my chosen family in London. Your friendships have always been my constant source of motivation. In particular, my heartfelt gratitude to my friend always despite the distance, Mengya Wu, for sending all the love across the ocean. I want to thank my artistic friends I am honoured to have in my life; to Majka Zelmanova, for your friendship and sweetness; to Sofia Bongiovanni and Anna Andreopoulou, *gratias pro omnibus bonis temporibus et risibus*; to Baihui Cheng and Huihui Li, for uniqueness and boldness; to Jessica González, for our passion for dance.

My life in London and the work of this thesis were partly funded by UCL, NEC -Japan and Europe and Cisco-USA. I am grateful for the funding received, which enabled the research work of this thesis to take place and progress over a period extending more than three years.

Contents

List of Figures

List of Tables

List of Symbols & Operators

α	Bandwidth compression factor
β	Roll-off factor
η	Spectral efficiency
ρ	Oversampling factor
σ^2	Variance (second moment)
Λ	Correlation matrix
Φ	Modulator matrix
$\mathbb{Z}$	Set of integer number
$\mathbb{E}[\cdot]$	Expectation operator
$\Re\{\cdot\}$	Real part of a complex number
$\Im\{\cdot\}$	Imaginary part of a complex number
$\lvert\cdot\rvert$	Absolute value of a number
$\mathscr{O}$	The order of complexity
Min	Minimum
Max	Maximum
$Q(\cdot)$	Gaussian tail function
$[\cdot]^H$	Hermitian operator
$\mathscr{H}$	Hilbert transform operator
$\mathscr{F}$	Fourier transform operator
$\otimes$	Convolution operator
I	In-phase component
Q	Quadrature component

List of Abbreviations

2G	2nd Generation
3G	3rd Generation
3GPP	3rd Generation Partnership Project
4G	4th Generation
5G	5th Generation
5G NR	5G New Radio
6G	6th Generation
Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
AIR	Achievable Information Rates
AM	Amplitude Modulation
AMC	Adaptive Modulation And Coding
APSK	Amplitude Phase Shift Keying
ASK	Amplitude-shift Keying
ASR	Automatic Speech Recognition
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BICM	Bit-interleaved Coded-modulation
BLER	Block Error Rate
BPSK	Binary Phase Shift Keying
BRGC	Binary Reflected Gray Code
BSA	Binary Switching Algorithm

BWP	Bandwidth Part
CAP	Carrierless Amplitude And Phase
CCDF	Complementary Cumulative Distribution Function
CCDM	Constant Composition Distribution Matching
CDMA	Code Division Multiple Access
CM	Coded Modulation
CP	Cyclic Prefix
CRC	Cyclic Redundancy Check
CSI	Channel State Information
CV	Computer Vision
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DL	Deep Learning
DM	Distribution Matching
DM-OFDM	Dual-mode OFDM
DM-RS	Demodulation Reference Signals
DMT	Discrete Multi-tone
DoF	Degrees Of Freedom
DSB	Double Side-band Modulation
DSP	Digital Signal Processing
DVB-S2	Digital Video Broadcasting-satellite-2nd Generation
EMBB	Enhanced Mobile Broadband
FBMC	Filter Bank Multi-carrier
FCT	Fast Cosine Transform
FDM	Frequency Division Multiplexing

FEC	Forward Error Correction
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FOFDM	Fast-OFDM
FTN	Faster Than Nyquist
GA	Genetic Algorithm
GFDM	Generalized Frequency Division Multiplexing
GI	Guard Interval
GS	Geometric Shaping
GSM	Global System For Mobile Communications
HOM	High-order Modulation
HT	Hilbert Transform
IC	Interference Cancellation
ICI	Inter-carrier Interference
ID	Iterative Detector
IDCT	Inverse Discrete Cosine Transform
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute Of Electrical And Electronics Engineers
IFFT	Inverse Fast Fourier Transform
IM	Index Modulation
InvDM	Inverse Distribution Matching
IoT	Internet Of Things
ISI	Inter-symbol Interference
ITU	International Telecommunications Union
LDPC	Low-density Parity-check

LLR	Log Likelihood Ratio
LoS	Line-of-sight
LP	Linear Programming
LPWAN	Low Power Wide Area Network
LS	Least Square
LTE	Long-term Evolution
LUT	Look-up Table
M-ASK	M-ary Amplitude-shift Keying
MBB	Mobile Broadband
MCM	Multi-carrier Modulation
MF	Matched Filtering
MIMO	Multiple-input Multiple-output
ML	Maximum Likelihood
MLC	Multilevel Coding
MMSE	Minimum Mean Square Error
MMTC	Massive Machine-type Communications
MmWave	Millimeter-wave
MmWave	Millimeter-wave
MSD	Multistage Decoding
MSE	Mean Square Error
NB-IoT	Narrowband Internet-of-things
NLP	Natural Language Processing
NOMA	Non-orthogonal Multiple Access
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
PAM	Pulse Amplitude Modulation
PAPR	Peak-to-average Power Ratio
PAS	Probabilistic Amplitude Shaping
PS	Probabilistic Shaping
PSK	Phase-shift Keying
QAM	Quadrature Amplitude Modulation
QAP	Quadratic Assignment Problem
QP	Quadratic Programming
QPSK	Quadrature Phase Shift Keying
ReLU	Rectified Linear Unit
RF	Radio Frequency
RLS	Recursive Least Square
RTS	Reactive Tabu Search
SA	Simulated Annealing
SC-FDMA	Single-carrier Frequency Division Multiple Access
SCM	Single-carrier Modulation
SCS	Subcarrier Spacing
SD	Sphere Decoder
SDP	Semi-definite Programming
SEFDM	Spectrally Efficient Frequency Division Multiplexing
SER	Symbol Error Rate
SGD	Stochastic Gradient Descent
SIC	Successive Interference Cancellation

SISO	Single-input Single-output
SM	Spatial Modulation
SNR	Signal-to-noise Ratio
SRRC	Square-root Raised Cosine
SSB	Single Side-band Modulation
STBC	Space-time Block Coded
TCM	Trellis Coded Modulation
TDM	Time Division Multiplexing
TFP	Time-frequency Packing
UE	User Equipment
UFMC	Universal Filtered Multi-carrier
UMTS	Universal Mobile Telecommunications System
URLLC	Ultra-reliable Low-latency Communications
USTLD	Uncoded Space-time Labelling Diversity
VA	Viterbi Algorithm
VLC	Visible Light Communications
W-OFDM	Windowed OFDM
WiMAX	Worldwide Interoperability For Microwave Access
WLAN	Wireless Local Area Network
ZP	Zero Padding

Chapter 1

Introduction

In the information age we inhabit, the rapid development of wireless communication technology has revolutionised the way we live, work and connect with one another. With the increasing use of smart devices, high-speed internet and a plethora of digital services, wireless communication has become an indispensable aspect of modern society. Over the past few years, the Covid-19 pandemic and consequent lockdown restrictions amplified the need for mobile connectivity and necessitated individuals to turn towards digital technologies to stay connected and access crucial services, ranging from essential education, finance and healthcare to social and family activities, and entertainment. The Mobile communications organisation GSMA reported that 67% of the global population was connected to mobile services by 2021. Driven by the increasing adoption of mobile devices and emerging data-demanding applications, mobile data is expected a significant increase; to be more than triple by 2027. Despite this growth, two key challenges - the scarcity of radio spectrum and energy constraints - which have been present throughout the evolution of mobile networks, pose a significant and increasingly alarming concern. Moreover, the varying characteristics of spectrum bands exacerbate the challenges by restricting the efficient allocation and utilisation of different frequency bands regarding the requirements of services and the possible expansion to the untouched spectrum range. Consequently, to provide even more efficient and reliable communications within the limited radio frequency (RF) bands, techniques are continually evolving to improve spectral and energy efficiency as in the current 5th generation

(5G) and for the future 2030 6th generation (6G) cellular systems.

In spite of acquiring new spectrum or reframing and reusing existing spectrum, various modulation, multiplexing and multiple access technologies have been proposed and employed in today's mobile communication standards, such as code division multiple access (CDMA); high-order modulation (HOM); orthogonal frequency division multiplexing (OFDM); multiple-input multiple-output (MIMO); non-orthogonal multiple access (NOMA), to improve spectrum utilisation and reduce energy consumption. Of these technologies, OFDM has been considered to be one of the most prominent technology, offering significantly enhanced bandwidth efficiency while showing high robustness against wireless channel impairments. The orthogonality among the overlapping subcarrier, within an OFDM symbol, advantageously allows for simple symbol generation and recovery, facilitated by the transceiver design adopting discrete Fourier transform (DFT)/ fast Fourier transform (FFT) and resulting in a low-complexity implementation in practice. Significantly, OFDM is widely used in numerous standards and applications in wired [1], wireless [2–5] and optical [6] communication systems. Unfortunately, OFDM's orthogonality benefit constrains its spectral efficiency, due to the required fixed subcarrier spacings (SCS). Hence, the optimum use of RF spectrum calls for novel technologies to tackle the challenge of providing higher data rates within the available bandwidth, without sacrificing the channel tolerance and power advantages when compared to OFDM.

Since the 3rd generation partnership project (3GPP) Release 15 [3], where the new air interface 5G NR was proposed, 5G has targeted three main scenarios;enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC), to continue the development of 4th generation (4G)-long-term evolution (LTE) to provide superior performance in terms of capacity and reliability for many new use cases. Following successful commercial deployments of 5G globally, 3GPP recently started the new phase of *5G Advanced*, proposed in Release 18 to boost network capabilities for new applications, stepping closer to the era of 6G communications. Both academic and industry trends show more focus on the use of scenario-based studies on technologies for optimising spectrum use, suggesting advanced signal processing techniques such as new signalling designs from a physical layer perspective.

New waveform/signalling designs proposed, as alternatives to OFDM and its variants and credited candidates for 5G networks, have been studied extensively, theoretically and practically. The 2014 paper "What 5G will be" [7] discussed some of the strong candidates for 5G and next generation networks such as pulse shaping filter based waveforms;filter bank multi-carrier (FBMC); universal filtered multi-carrier (UFMC); generalized frequency division multiplexing (GFDM), demonstrating the improved spectral efficiency and the corresponding penalties. Non-orthogonal waveforms, aiming to achieve higher spectrum utilisation efficiency by violating the orthogonality required by OFDM, have also attracted increasing research interest worldwide [8]. Amongst the techniques, spectrally efficient frequency division Multiplexing (SEFDM) achieves higher spectral efficiency relative to OFDM by compressing the SCS, showing greater than 25% [9–13] and acceptable performance close to OFDM by using advanced detection methods. faster than Nyquist (FTN) is a similar technique focusing on time domain signal design and processing while sharing a similar non-orthogonal concept. In FTN, symbols are transmitted at higher rates than the Nyquist rate so as to reach a higher data rate whilst occupying the same bandwidth. time-frequency packing (TFP) that optimises the time and frequency spacing to maximise the overall spectral efficiency.

Different from non-orthogonal signalling, which improves the system capacity by enhancing the efficiency of spectrum utilisation, non-uniform signalling requires lower energy to transmit at the same data rate for a given constellation relative to uniform signalling. Constellation shaping has been studied over decades and used as an excellent tool to achieve higher capacity close to Shannon Limit [14–16]. Information theory teaches that an ultimate shaping gain of 1.53 dB can be achieved if the shaped constellation yields a Gaussian distribution. There are two types of constellation shaping, geometric shaping (GS) [17] and probabilistic shaping (PS) [16]; GS refers to constellations with irregular shapes, for example, non-rectangular quadrature amplitude modulation (QAM); PS transmits constellation symbols at different frequencies, such that the occurrence probabilities of constellation symbols, contrary to the case of uniform constellations, are unequal. Recently, PS has been extensively studied and applied in optical communication systems [18, 19], and was shown to provide significant energy efficiency gains and flexibility in rate adaptability.

This thesis presents an in-depth analysis of the state-of-the-art of spectrally and energy efficient technologies in wireless communications and proposes innovative and effective solutions to enhance the overall system performance, particularly in terms of the data rate, with the goal of reaching the theoretical limits set by Shannon's theorem. The proposed methods are rigorously evaluated through the development of mathematical models and their efficacy is validated through extensive simulations. In order to demonstrate their feasibility in real-world scenarios, the proposed techniques are also examined in various practical use cases, including those specified in deployed 5G networks. The theoretical and analytical findings are substantiated by evaluating the simulation results and providing a critical and comprehensive understanding of the trade-offs and limitations of the proposed techniques. Therefore, the research work presented in this thesis aims to contribute to the advancement of wireless communication technology by providing practical solutions with a theoretical support for improving the spectral and energy efficiency in wireless communication systems.

1.1 Aims and Motivations

The work in this thesis addresses the challenge of improving transmission data rates, limited by scarce spectrum resources, while maintaining good system performance and accounting for various use scenarios for next-generation wireless systems. The physical layer approaches, conventionally adopted in the current standards, constitute a major bottleneck for further enhancements of systems performance, including spectral efficiency, error performance, system complexity and energy efficiency. Furthermore, the utilisation of higher frequency bands such as millimeter-wave (mmWave) presents new system and signal design challenges, due to channel complexities and associated high signal powers required and higher power consumption of hardware components operating at such frequencies.

A variety of promising physical layer technologies are critically reviewed and discussed in the research in this thesis to demonstrate their advantages in performance improvement but also the limitations, which come along with such advantages. Amongst these techniques, some have been investigated extensively over the past two decades, such as non-orthogonal signalling SEFDM [8] and FOFDM, pulse shape filtering [20], non-uniform signalling [16] and multidimensional modulations [21]. Some are new designs recently proposed, derived from previous technologies and techniques applied to specific physical layer scenarios, such as the case of constellation shaping [22] and filtered multi-carrier modulations [23–25]. As mentioned above, it is widely acknowledged that future communication systems, as beyond-5G and 6G, will be more scenario-oriented, with increasingly complex and multi-faceted requirements [26]. Therefore, a single technique is unlikely to suffice in supporting future networks; consequently, a combination of diverse technologies may be required.

As such, new challenges arise in terms of the optimisation of the overall system designs where multiple techniques are used and operational parameters are chosen. This drives research focused on providing insights into the effects of different tuning and prototyping the optimisation methods. In this line, this thesis investigates optimisation problem and explores different solutions via conventional metaheuristic search algorithms and machine learning techniques. The main research aims to contribute to advancements of spectrally and energy efficient techniques towards the ultimate goal of bringing more potential solutions for various specific use scenarios in the envisioned 6G and future networks.

1.2 Contributions

This thesis documents research work that has resulted in a number of original theoretical and engineering design contributions towards improving the spectral and energy efficiency in wireless communications. The advantages of the reported techniques are substantiated by theoretical analysis and mathematical modellings of wireless system design and signal processing, optimisation through metaheuristic algorithm development and machine learning, computer simulations and performance investigations. The key contributions of the thesis are highlighted as follows:

- Designed the coexisting signalling model of SEFDM and OFDM under 5G NR frame and assessed the performance of the coexisting signal transmission. This study delved into three use scenarios based on the flexible numerology of SCS, with the system performance verified with and without the assistance of low-density parity-check (LDPC) coding. This work was published in [27] and appears in Chapter 2 of this thesis.

- Proposed a new signalling design which employs Hilbert transform pair that offers improved spectral efficiency without almost no error performance degradation. The mathematical model was designed and developed for signal generation and reception based on the theoretical analysis of the principles Hilbert transform pair. This work is presented in Chapter 3.

- Applied the proposed Hilbert transform (HT) filter pair based signalling technique to two special cases: *i)* turbo-coded signal formats, for which HT-SEFDM system, where turbo coding and its efficacy are verified by mathematical modelling; *ii)* internet of things (NB-IoT) for which HT-FOFDM system is modelled and evaluated through numerical simulation. This study included an investigation to show advantages in error performance and effective data rate achieved relative to OFDM. This work was published in [28], [29] and [27] and appears in Chapter 3 of this thesis.

- Applied constellation shaping to SEFDM system and designed the PS-SEFDM system with the use of successive interference cancellation (SIC) to eliminate the deliberately introduced inter-subcarrier interference. This work includes the comparative study of PS-SEFDM and LDPC-coded OFDM for various transmission rates and the numerical results verified the power advantage of PS-SEFDM and a minimal increase in system complexity. This work was published in [30] and is demonstrated in Chapter 4.

- Developed and implemented four metaheuristic algorithms that optimised the bit-to-symbol mapping for multidimensional modulation signals. Simulations of the developed algorithms demonstrated the efficacy of the mapping optimisation for a specially designed four-dimensional signal format. The numerical results showed the performance of each algorithm in terms of their complexity and convergence speed. The performance of the four-dimensional signal, using the optimised mapping, was tested and verified through simulations of a designed single-carrier modulation (SCM) system model. The work and the aforementioned results are presented in Chapter 5 of this thesis and are the basis of two patents listed in the section below.

- Proposed a new machine learning model to optimise multidimensional probabilistic shaping for multidimensional modulation signals, which is introduced and described in Chapter 5. The proposed technique utilises an autoencoder-based model structure to approximate the transceiver design of conventional communication systems. The work contains simulation studies, which showed the optimisation feasibility in terms of the machine learning complexity and the effect of key parameters on the learning performance. In addition, the study reported significantly improved achievable data rates of the optimised probabilistically shaped signals. This work was published in [31] and appears in Chapter 6 of this thesis.

1.3 Publications

The work presented in this thesis has resulted, to date, in publications in six international conference proceedings and two patents. The publications are listed below in chronological order.

- Publications

  1. **X. Liu** and I. Darwazeh, "Doubling the Rate of Spectrally Efficient

FDM Systems Using Hilbert Pulse Pairs," 2019 26th International Conference on Telecommunications (ICT), Hanoi, Vietnam, 2019, pp. 192-196

2. **X. Liu** and I. Darwazeh, "Quadrupling the Data Rate for Narrowband Internet of Things without Modulation Upgrade," 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia, 2019, pp. 1-5

3. **X. Liu**, T. Xu and I. Darwazeh, "Coexistence of Orthogonal and Non-orthogonal Multicarrier Signals in Beyond 5G Scenarios," 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 2020, pp. 1-5

4. **X. Liu**, I. Darwazeh, N. Zein and E. Sasaki, "Spectrally Efficient FDM System with Probabilistic Shaping," 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 2021, pp. 1-6

5. **X. Liu**, I. Darwazeh, N. Zein and E. Sasaki, "Probabilistic Shaping for Multidimensional Signals with Autoencoder-based End-to-end Learning," 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 2022, pp. 2619-2624

6. **X. Liu** and I. Darwazeh, "Energy and Spectrally Efficient Signalling for Next Generation IoT," 2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Porto, Portugal, 2022, pp. 202-207

- Patents

1. E. Sasaki, **X. Liu**, I. Darwazeh, and N. Zein, Japanese Patent Application No.2021-105088, filed June 24, 2021.

2. E. Sasaki, **X. Liu**, I. Darwazeh, and N. Zein. "Signal modulation apparatus and signal modulation method." U.S. Patent US 2022/0417075 A1, filed June 7, 2022 and published December 29, 2022.

1.4 Structure of the Thesis

This thesis consists of 7 chapters. Chapters 2-6 contain original work by the author and chapter 7 offers the author's summary and critical assessment of the work chapters. In addition to this introductory chapter, the remaining chapters are organised as follows.

- To contextualise this PhD research and highlight its novel contributions, Chapter 2 presents a comprehensive review of the current state-of-the-art in spectrally efficient technologies in wireless communications. The chapter starts with the introduction of the theoretical capacity upper bounds set by Shannon's limit and the key concepts related to spectral efficiency, then gives an overview of technologies aimed at increasing spectral efficiency in different mobile system generations, with a particular emphasis on multi-carrier modulation (MCM) waveform/signalling formats. This chapter delves into the fundamentals of three MCM systems: OFDM, FOFDM and non-orthogonal SEFDM and presents detailed mathematical models of their transceiver designs and interference features. To provide further insights, a model is developed to run a set of simulation based studies into various coexistence scenarios of non-orthogonal and orthogonal waveforms with flexible numerology in 5G NR. These different systems and studies provide the fundamental framework that stimulates the research in the following chapters.

- Chapter 3 proposes a novel spectrally efficient design using Hilbert filter pair, by building on the SEFDM and FOFDM signals introduced in chapter 2. The chapter commences with a brief review of waveform/signalling designs, which adopt pulse shaping filters, from which the key advantages of filtering are identified. Then, it proceeds with an overview of the theoretical principles of Hilbert transform and Hilbert filter pair. A mathematical model of the developed system design employing Hilbert pair and simulation studies are presented to prove the enhanced spectral efficiency achieved by introducing the extra level of orthogonality of the filter pair. Two signal formats are studied as proof of concept, which are a multi-carrier signal with turbo coding and an uncoded narrowband internet-of-things (NB-IoT) format based on that used in 4G-LTE and 5G systems.

- Improving energy efficiency is the focus of Chapter 4, which explores the use of probabilistic constellation shaping to SEFDM, where shaping with variable transmission rates is investigated. The design of the probabilistically shaped SEFDM system is outlined with detailed descriptions of the newly designed transceiver structure. Systems of varying coding rates and SEFDM compression factors are compared. For fair comparisons, systems with fixed achievable spectral efficiencies are modelled in additive white Gaussian noise (AWGN). Specifically, performance of SEFDM systems, with and without probabilistic shaping and with varying compression factors, are compared to OFDM. The numerical results demonstrate significant shaping gains, along with the power advantage relative to OFDM. Therefore, the new proposed systems are shown to be advantageous in energy efficiency and also show substantial bandwidth saving and a further advantage in allowing flexible data rate adaptation. The rest of the chapter includes two sections: detailing the modelling of frequency selective channels appropriate for the type of signals studied and then proposes new system designs, combining probabilistic shaping and SEFDM, of pilot based channel estimation and equalisation. The chapter presents simulation studies of the proposed scheme and shows their working through channel response studies and channel estimation error analysis.

- Chapter 5 is dedicated to achieving higher energy efficiency by multidimensional modulation techniques, for which metaheuristic optimisation algorithms are developed and used to optimise the bit-to-symbol mapping. The chapter provides an overview of existing multidimensional modulation techniques to show their advantages and main applications in wireless communications. Then, a four-dimensional signal design, which obtains the extra degree of freedom in time, is introduced and discussed in terms of its power advantages and simplicity in implementation. The challenges of bit-to-symbol mapping for multidimensional signals are addressed. Thereafter, the optimisation problem was formally set and four metaheuristic algorithms, RTS, BSA, GA and SA, are developed and operated on the same problem. Finally, a fair comparison study is conducted for the computation results, in terms of convergence speed, algorithm complexity and the bit error rate (BER) performance of optimised mappings.

- Chapter 6 continues the investigation on optimisation for multidimensional modulation signal design in chapter 5 and constellation shaping in chapter 4 by applying machine learning techniques instead of conventional metaheuristic optimisation methods. The research gaps of multidimensional probabilistic shaping and the challenge of using conventional optimisation methods are identified. A novel autoencoder-based end-to-end learning model is proposed to optimise multidimensional probabilistic shaping and is described in detail. Studies in this chapter show that with the machine learning-aided model, the optimisation is proved feasible and efficacious. Then, the system model is presented to evaluate the optimised modulation formats, which are distinguished by their implementation simplicity, higher spectral efficiency and power efficiency.

- Chapter 7 offers a summary of the work in this thesis, highlighting the significance of the proposed techniques and offering a critical look at the work presented in the thesis. Additionally, an outline for potential future work is provided.

- Three brief appendices are included with relevant mathematical formulations/definitions of Hilbert Transform, AWGN and BER and of PAPR as used in the thesis.

Chapter 2

Spectrally Efficient Systems Concepts, Applications and Designs

2.1 Introduction

This chapter presents the background of the development of spectrally efficient technologies in wireless communications throughout various generations of mobile network, with the focus of waveform/signalling schemes that are currently employed and expected to be used in beyond 5G and future 6G networks. The chapter commences with the essential Shannon-Hartley theorem that determines the maximum achievable spectral efficiency, then discusses different techniques aiming at improving the spectral efficiency in section 2.2. Section 2.3 provides an overview of the different spectrally efficient waveforms/signalling designs, highlighting the design, advantages and limitations of two important techniques: FOFDM and SEFDM. These two systems are basic systems used in the following two chapters. To address the feasibility of application of SEFDM in 5G NR and show the advantageous performance gain of SEFDM, section 2.4 studies the coexistence of the non-orthogonal SEFDM and the orthogonal OFDM in different scenarios. The coexistence studies resulted in a paper in the proceedings of the 6G Summit, which was presented online in March 2020 [27][1].

[1]X. Liu, T. Xu and I. Darwazeh, "Coexistence of Orthogonal and Non-orthogonal Multicarrier Signals in Beyond 5G Scenarios," 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 2020, pp. 1-5.

2.2 Shannon Limit and Spectral Efficiency

New technologies aimed at enhancing the system capacity have been proposed towards standardisation and commercialisation. Hereby, a key question arises: how to quantify the achievable system capacity for assessment and upper bound calculation? The answer was given by Shannon based on the ideas developed by Nyquist [32] and Hartley [33] in information theory; Nyquist proposed the concept of Nyquist rate where transmitted signal needs to be sampled at least twice the highest frequency (i.e. the bandwidth) to get rid of distortion over the transmission; Hartley constructed a quantitative measure of the data rate. Therefore, based on Shannon's theorem, the channel capacity C in presence of noise can be calculated as in equation (2.1):

$$C = B\log_2(1 + \frac{\mathrm{S}}{\mathrm{N}}), \tag{2.1}$$

where B denotes the channel bandwidth in Hz, S and N are the average signal and noise power over the bandwidth, respectively. The ratio of S and N also denotes the signal-to-noise ratio (SNR) of the signal at the receiver side. This equation indicates the theoretical upper bound of the channel capacity. In other words, it provides the upper bound on the achievable data rate over a certain bandwidth and for a given SNR. Due to the RF spectrum scarcity, it is unfeasible to rely on increasing the bandwidth B to obtain higher channel capacity. Another possible way to improve the channel capacity is to increase the transmitted signal power S, which is at the cost of power efficiency.[2]

Where capacity C gives the maximum data rate achievable over a given channel, a normalised measure of capacity is what is termed spectral efficiency, which normalises capacity to the channel bandwidth. Therefore, the term spectral efficiency refers to the data rate that is transmitted per unit channel bandwidth, thus providing a more straightforward measure of the spectrum utilisation. The spectral

[2]Modern wireless systems increasingly use multiple input multiple output (MIMO) techniques to increase capacity by utilising diversity and using more than one transmit and receive antennas. This will correspondingly increase capacity, especially when there is multipath propagation. MIMO techniques can be used on top of any existing modulation method and therefore are considered by add-on methods and are not discussed further in this thesis.

efficiency η can be given by 2.2 with the unit of bits/s/Hz[10]

$$\eta = \frac{R_b}{B}, \tag{2.2}$$

where R_b denotes the transmission bit rate of a communication system and B is the channel bandwidth. Relating this to the Shannon's capacity equation above, the maximum spectral efficiency achievable will be simply equal to C/B. These concepts are explored in the mathematical treatment below. The bit rate $R_b = R_s \times \log_2(\mathrm{M})$ can be calculated by the the ratio of symbol rate R_s and bits-per-symbol, wherein M is the cardinality order.

Based on the definition of η, for a single channel use, equation 2.3 can be derived easily by referring to equation 2.1

$$\eta \leq \log_2(1 + \frac{\mathrm{S}}{\mathrm{N}}), \tag{2.3}$$

which draws the upper bound on the spectral efficiency η for a given SNR. Or

$$\frac{S}{N} \geq 2^{\eta} - 1, \tag{2.4}$$

which on the contrary gives the lower bound of SNR for a given spectral efficiency η. Such lower bound, also termed as Shannon limit, can be normalised to SNR per bit, i.e. E_b/N_0, as given by equation 2.5

$$E_b/N_0 \geq \frac{2^{\eta} - 1}{\eta}. \tag{2.5}$$

Fig. 2.1 shows how the spectral efficiency of different modulation formats approaches Shannon limit as the SNR and the modulation order increase[3]. It is worth noting that Shannon limit is shown as the blue curve in the illustration, which is the smallest E_b/N_0 to achieve a certain error rate for a given spectral efficiency.

According to equation 2.2, higher spectral efficiency can be achieved by either increasing the transmission data rate or reducing the bandwidth required. Addition-

[3]The Shannon capacity or the upper bound is calculated at the BER of 10^{-5}.

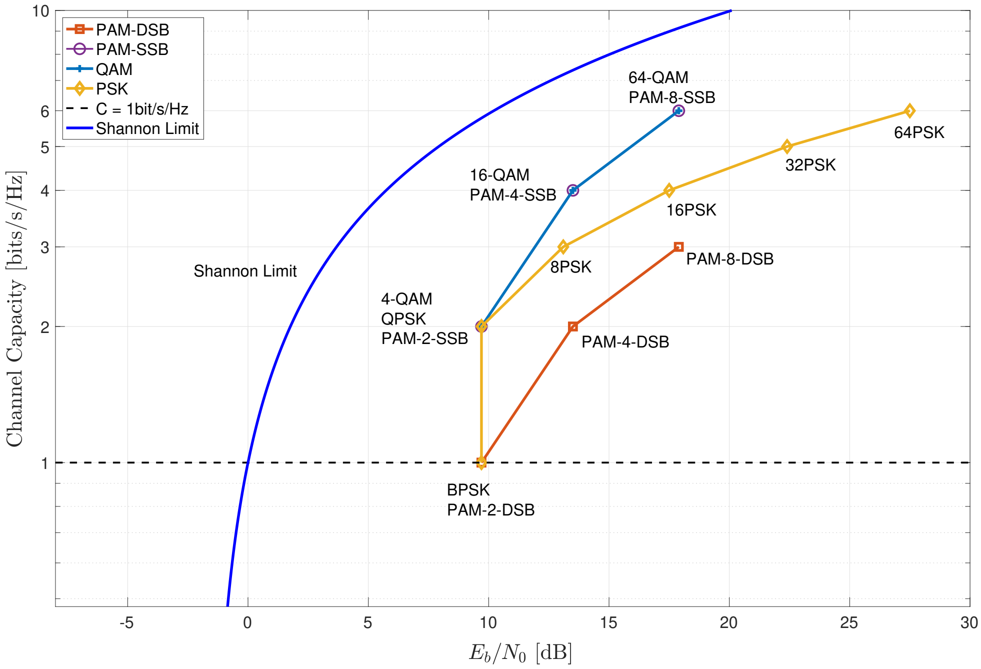


Figure 2.1: Simulated channel capacity over certain spectrum versus E_b/N_0 for different modulation schemes

ally, equation 2.3 shows that the spectral efficiency of a communication system is directly related to the required SNR. In practice, the required SNR largely depends on the channel conditions, which can vary significantly in wireless environments. In this study, wireless channel impairments are considered with the focuses of signal distortion, such as inter-symbol interference (ISI) and inter-carrier interference (ICI), caused by dispersive fading resulting from multipath propagation. [34, Section 8.2] provides a general review of the characteristics of the wireless channel. A detailed discussion of wireless channel models is presented in chapter 4.

There are several approaches to improving the spectral efficiency of communication systems. Regarding the modulation scheme, MCM techniques allow high-speed data to be transmitted in a given bandwidth to approach higher spectral efficiency. Instead of transmitting data over a single carrier as in SCM, MCM divides the channel into multiple narrower subchannels. The data carried by each subcarrier are transmitted in parallel and aggregated to provide a higher data rate than SCM. Over the years, SCM was the dominant modulation technique, which has been widely adopted in 2nd generation (2G) and 3rd generation (3G) cellular systems (global system for mobile communications (GSM), universal mobile telecommunications system (UMTS), CDMA2000). The concept of parallel transmission of MCM was initially designed for high-frequency Collins Kineplex system [35] in the early 1960s, followed by applications in both wire-line and wireless communications, where different medium-based channels are divided into non-overlapping subchannels to modulate data. A decade later, the principles of OFDM were proposed to provide more efficient bandwidth utilisation by overlapping the individual subchannels imposed by specific orthogonal rules [36, 37]. OFDM, also known as discrete multi-tone (DMT), is the most prominent MCM scheme that has been extensively researched, developed and consequently used in numerous wireless systems and standards over the past 50 years. In particular, OFDM forms part of various wireless local area network (WLAN) Institute of Electrical and Electronics Engineers (IEEE) 802.11 [4] and 802.16 [5] standards, which provide the basis of Wi-Fi and worldwide interoperability for microwave access (WiMAX), respectively. Significantly, OFDM with multiple access (OFDMA) is chosen for the 4G-LTE networks as the core of the downlink radio transmission [38]. Moreover, the 3GPP specified in favour of OFDM-based techniques, specifically cyclic prefix (CP)-OFDM for uplink and downlink and DFT-spread-OFDM downlink transmission for 5G NR [39]. A comprehensive review of OFDM can be found in [40]. To combat ISI, CP-OFDM adopts the cyclic extension with adjustable numerology supported by 5G NR. Due to the dispersive fading in multipath propagation scenarios, CP duration is chosen to be greater than the delay spread of the channel to prevent the signals from ISI. The use of CP in comparison with another type of guard interval (GI), termed zero padding (ZP), is detailed in section 2.3.1. The DFT-spread-OFDM, which can be regarded as precoded OFDM, combines the benefits of better power efficiency of SCM and flexibility for frequency resource allocation of MCM.

Despite the success of OFDM, the pursuit of higher spectral efficiency and better transmission reliability has always been the objective due to the spectrum scarcity and growing demand for higher data rates. One key advantage of OFDM is that the subcarriers are orthogonal to one another, which guarantees the symbols from suffering ICI. However, the frequency spacing between adjacent subcarriers is fixed, making it impossible to improve spectral efficiency further. To address this issue and improve spectrum utilisation, non-orthogonal signalling (waveform) techniques, with a significant increase in spectral efficiency, have been investigated by academia and industry over the past three decades. In 1975, Mazo first studied a sinc-pulse based format-FTN, which violates the Nyquist criteria by transmitting at a rate higher than the Nyquist rate [41]. In the study, despite self-induced ISI, Mazo showed that up to 25% gain can be obtained without degradation to the BER performance, provided an optimum detector is used. The identified transmission rate limit is termed Mazo limit, above which the minimum Euclidean distance remains the same, and no error penalty occurs [42]. Such concept of non-orthogonal signalling was later applied to two-dimensional modulations such as 16-QAM [43] and proceeded to extend to MCM formats [44], hence offering direct use of FTN in practical multicarrier signalling. Notwithstanding the advantageous higher spectral efficiency of FTN relative to its orthogonal counterpart, sophisticated detection algorithms are required at the receiver. The increased computation complexity hinders practical implementations of FTN due to the additional digital signal processing (DSP) requirements. Such DSP procedures include the use of trellis-structured Viterbi algorithm (VA) and BCJR algorithm decoder, the SIC and the commonly-used ISI removal method-equalisation [44]; the trellis-structured VA detects the most likely symbol; the BCJR algorithm decoder offers the likelihood; SIC subtracts the estimated soft information of ISI iteratively. The former two methods are proven to perform ISI-free transmission when the designed FTN transmits within a 42% increase in bits density. At the same time, the latter two techniques are employed when the system encounters a much more severer ISI for a significant increase in spectral efficiency.

As in OFDM, the orthogonality is the advantage and constraint at the same time. Instead of obliterating the orthogonality rule, Rodriguez and Darwazeh first proposed a semi-orthogonal signal format–FOFDM [45] in 2002, which aims to transmit at the same bit rate while occupying half the bandwidth of OFDM. A similar system termed M-ary amplitude-shift keying (M-ASK) OFDM [46] was developed theoretically by F. Xiong in 2003. By halving the frequency separation of OFDM subcarriers, FOFDM achieves doubled spectral efficiency relative to OFDM. The theoretical principle of FOFDM relies on the fact that the orthogonality is maintained when the input symbols modulated by OFDM subcarriers are real. Therefore, FOFDM delivers a similar bit error performance to OFDM when one-dimensional modulation schemes, such as binary phase shift keying (BPSK), pulse amplitude modulation (PAM) and amplitude-shift keying (ASK), are used. The aforementioned orthogonality rule is no longer valid for FOFDM with higher-dimensional modulations. As such, modulating FOFDM subcarriers with higher-order complex symbols leads to BER degradation due to the ICI [47, 48]. Despite its constraint with real input symbols, FOFDM has been used in wireless [49] and optical [50] communication systems. The concept of FOFDM motivated the invention of SEFDM, a non-orthogonal multicarrier signalling technique proposed to improve spectral efficiency by purposely relaxing the orthogonality [51]. Similar to FOFDM, the non-orthogonality of SEFDM is achieved by compressing the frequency spacing between adjacent subcarriers. In the past two decades, this technique has been developed theoretically and practically, covering various topics, which are generally described in [8]. Analogously to FTN, where ISI is introduced by interference between adjacent pulses, ICI is introduced to SEFDM subcarriers. It is worth noting that multi-carrier FTN can be seen as SEFDM's time domain counterpart. In SEFDM, the enhancement of bandwidth efficiency [52] is traded against the cost of the self-introduced ICI, potentially leading to performance degradation or increased detection complexity. Notwithstanding, compared to FTN, SEFDM shows a greater than 25% bandwidth efficiency improvement with similar error performance to OFDM, providing advanced detection methods [9, 10, 12]. Sharing similar characteristics with FTN and SEFDM, another non-orthogonal technique termed TFP improves spectral efficiency by reducing the spacing between adjacent signals in both time and frequency domains [53].

6G networks are envisioned to support extremely high data rates (1 Terabits/s) and secure URLLC (less than 0.1 ms) for high-mobility communications operating at expanded higher frequency bands [54]. In such scenarios, the conventional OFDM-based waveform suffers from severe impairment in the presence of time and frequency selectivity. In this context, a new two-dimensional modulation scheme, orthogonal time frequency space (OTFS), has been proposed to tackle the critical challenges of providing reliable transmission in a high-mobility environment. OTFS performs modulation by multiplexing QAM symbols over a new class of carriers localized in the delay-Doppler domain. Hence, the information symbols are transformed to span the bandwidth and time duration of the transmission frame; consequently, all symbols experience the same channel gain over the transmission. In the meantime, the time and frequency shifts due to the channel effects are separated, and therefore the fading effect can be substantially mitigated. OTFS has shown improved BER performance in the mmWave bands (28 GHz), where severe phase noise and high Dopplers are encountered [55]. Moreover, the OTFS transceiver can be implemented based on the conventional OFDM designs with pre-processing and post-processing blocks added [56], thus making OTFS an attractive candidate for next-generation communication systems in terms of its implementation.

2.3 Spectrally Efficient Multi-carrier Systems

Among the wireless communication systems mentioned above, the multi-carrier signalling used in the research presented in this thesis are mainly based on three spectrally efficient systems, namely the OFDM, the FOFDM and the SEFDM systems. The statistical model of each system is provided to establish the foundation of the energy and spectrally efficient techniques presented in the following chapters.

2.3.1 OFDM

The popularity of MCM system in modern wireless communications can be attributed to its potential in supporting high transmission data rates and related robustness against wireless channel impairments. Frequency division multiplexing

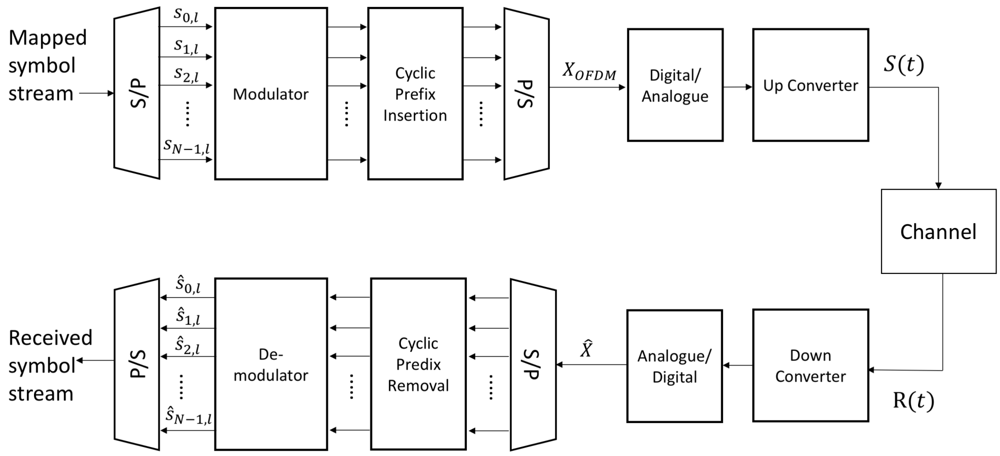


Figure 2.2: Block diagram of a typical OFDM system

(FDM) donates those advantages by transmitting a high-data-rate input stream over multiple lower-data-rate parallel sub-streams [11].

In an ordinary FDM system, a wide-band channel is split into several narrow-bands. Each refers to a subcarrier, occupied a specific range of frequencies. The sub-bands are non-overlapping with each other, and hence no interference is introduced in such scheme [10]. Compared to conventional FDM scheme, OFDM reduces the overall bandwidth by overlapping the subcarriers in a particular way such that at the peak of one subcarriers all the other subcarriers land at the zero-crossing points. In other words, the series of subcarriers are orthogonal to each other. The orthogonality not only prevents the scheme from introducing ICI but also contributes to the simplified demodulation at the receiver, where multiple carriers can be separated easily due to the mathematical properties of it [11]. Fig. 2.2 shows the mechanism of a typical OFDM system. The serial to parallel converter (S/P) groups the serial mapped symbol stream into N blocks. Each block of symbols then modulates the corresponding subcarrier, referring to the exponential functions with different frequency factor f_n in the diagram. Subsequently, N blocks of modulated carriers are multiplexed in the parallel to serial converter (P/S) and thus form the OFDM symbol. This process is commonly carried out by using IDFT, or IFFT due to its faster computation speed [10]. Having obtained the OFDM signal, the digital-to-analogue (D/A) translates the digital signal into an analogue signal. This signal is then upconverted to the desired frequency and passed through the wireless channel [10]. At the receiver, the previous processes at the transmitter are reversed to demodulate and detect the original input signal.

Based on Fig. 2.3, the OFDM subcarriers are modulated by the incoming mapped symbol streams, thus forming the OFDM signal given by:

$$x_{OFDM}(t) = \frac{1}{\sqrt{T}} \sum_{n=0}^{N-1} s_{n,l} g(t - l \cdot T) \cdot e^{j2\pi nt/T}, \tag{2.6}$$

$$n, l \in \mathbb{Z}, n \in [0, N-1]$$

where $s_{n,l}$ represents the information symbol in the l_{th} symbol period and corresponds to the n_{th} subcarrier. T is the time duration of one OFDM symbol, which can also be given by the reciprocal of the input data rate. $g(t)$ denotes the pulse shaping function. $\sqrt{T}$ is the scaling factors. In the exponential function, $j = \sqrt{-1}$ is defined as the imaginary unit.

The pulse shaping function $g(t)$ refers to a rectangular function in most common scenarios. To satisfy the orthogonality in OFDM scheme, the frequency spacing Δf_{OFDM} should equal to the reciprocal of the time period T of each unmodulated symbol, i.e. $\frac{1}{T}$ [10]. According to the frequency-time conversion, the theoretical spectra of OFDM subcarriers can be seen as multiple individual sinc pulses as depicted in Fig. 2.3.

It is evident that at the peak point of one subcarrier, the rest carriers land at the zero-crossing point, hence the orthogonality is ensured. The mathematical expres-

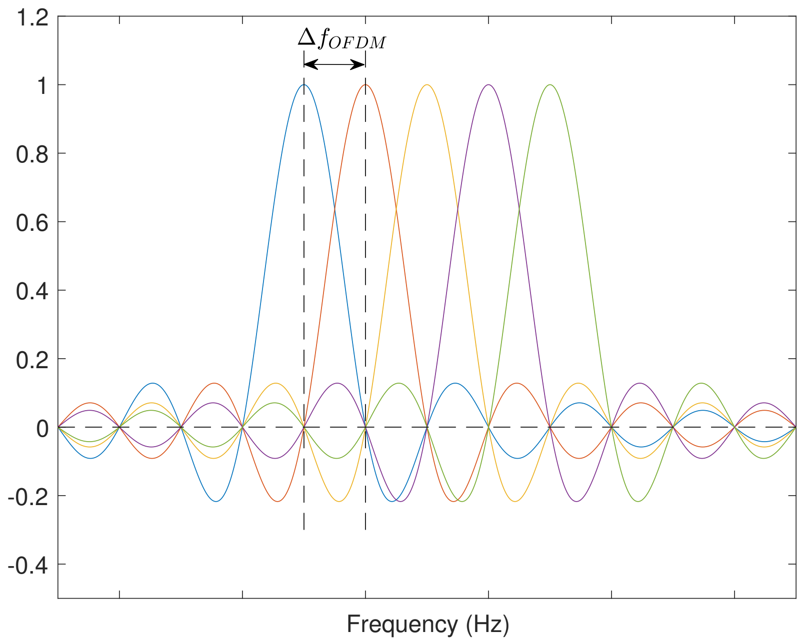


Figure 2.3: OFDM spectrum

sion of this can be given by equation 2.7 [10, 57]:

$$\int_{0}^{T} x_m(t)x_n(t) = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases} \tag{2.7}$$

Guard Interval: Cyclic Prefix or Zero Padding

As described in the previous section, OFDM signal can be demodulated without interference between the subcarriers in the idea case. However, in practical signal propagation, the channel may consist of multiple paths and could be time-dispersive. In this case, delays and attenuation are introduced to the transmitted OFDM signal when it passes through the dispersive channel. As a result, both ISI and ICI occur and consequently distort the symbols, leading to degradation to the error performance evaluated at the receiver.

To deal with the time dispersion, guard interval insertion between adjacent OFDM symbols is commonly used. Generally, a set of time domain symbols are inserted to the output of inverse fast Fourier transform (IFFT) at the transmitter and removed from each received OFDM symbol before the FFT operation at the receiver. In principle, the guard interval length should be greater than the maximum length of the time dispersion among all paths (i.e. the channel delay spread τ_d) to cover the signal corruption. This means the selection of the guard interval length is dependent on the channel, or the cell size in practical cellular communications [58]. A longer guard interval is better at controlling the time dispersion, whereas the effective data rate decreases as more redundant symbols are transmitted and thereby resulting in a higher power consumption. Therefore, flexible selections of guard interval are configured in 4G [59] and 5G [3] standards to meet the various requirements in different scenarios.

The none-zero CP and zero-valued ZP are two typical guard intervals commonly used in wireless communication systems. In a discrete-time model, the cyclic prefix is obtained by copying a section of the end of an OFDM symbol and appending that to the front of the same symbol as a prefix. The cyclic extension of the OFDM symbol makes the transmitted symbol continuously in one symbol period. This yields a significant benefit that by using single-tap equaliser at the receiver, the ISI can be eliminated completely in an OFDM system when CP is employed. CP eliminates the ISI without introducing ICI to transmitted signals. As discussed, 4G LTE standard defines 2048-points IFFT based CP-OFDM for downlink transmissions (for uplink DFT-spread OFDM is used), where two kinds of CP length are employed, normal CP for small cell size scenario and extended CP for large cell size [59]. 5G NR employs CP-OFDM for both uplink and downlink transmissions. In principle, the time-domain convolution operation corresponds to the multiplication in frequency domain. The main drawback of CP-OFDM is that the symbols can not be recovered when the channel's components located on a subcarrier are all zero.

Zero padding[4] was proposed in [60] to replace CP, where a series of zeros are padded to the OFDM symbol outputs. ZP-OFDM achieves the same spectral efficiency as CP-OFDM when the length of appended symbols are equivalent. Unlike

[4]To distinguish the two types of zero padding, hereby ZP is performed in time domain. ZP in frequency domain corresponds to zero insertion to the IFFT inputs

CP, the OFDM signal suffers ICI when ZP is used and as such filtering is adopted other than FFT to guarantee the recovery of the received signal. Nevertheless, the price paid is the increased complexity of the receiver.

2.3.2 Fast-OFDM

FOFDM is a variation of OFDM achieving a doubled spectral efficiency. The attractive feature of FOFDM is that for a given data rate, the FOFDM symbols requires only half the bandwidth of the OFDM when the same modulation is used. The reduced spectrum utilisation is achieved by reducing the frequency separation $\Delta f = 1/T$ to $1/2T$ while the multi-carrier symbol duration T remains the same. The normalised FOFDM symbol in the k^{th} signalling interval is given by [45]:

$$x_{FOFDM}(t) = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{+\infty} \sum_{n=0}^{N-1} X_{n,k} e^{\frac{j2\pi nt}{2T}}, \quad t \in [0,T], \tag{2.8}$$

where N is the total number of the subcarriers, $n \in [0, N-1]$ denotes the frequency index of the subcarrier, $X_{n,k}$ represents the symbols transmitted on the n^{th} subcarrier, which is denoted by the exponential component (i.e. $e^{j2\pi nt/2T}$).

The halved frequency separation of FOFDM results in the loss of orthogonality when compared conventional OFDM, constraining its multiplexing/demultiplexing implementation, where the standard DFT/inverse discrete Fourier transform (IDFT) operations are not feasible. To tackle the problem, one of the solutions is to use discrete cosine transform (DCT)/inverse discrete cosine transform (IDCT). More specifically, the fast cosine transform (FCT) algorithm is used to perform DCT, where cosine functions with different frequencies are used to express the real-valued subcarriers. In this case, only one-dimensional modulation schemes, such as BPSK, M-ASK and PAM, can be adopted. Taking M-ASK as an example, the output of the IDCT is expressed as [46]:[5]

$$x(t) = \sum_{k=-\infty}^{+\infty} \sum_{n=0}^{N-1} X_{n,k} \cos(2\pi nt \frac{1}{2T}), \quad t \in [0,T], \tag{2.9}$$

[5] The equation is taken from [46] without normalisation

where $X_{n,k} = A_n$ denotes the M-ASK symbols that modulate the n^{th} subcarrier, of which the amplitude A_n is in the M-ary amplitudes set. Thus, the verification of the orthogonality can be simplified in equation(2.10)

$$\int A_i A_j \cos(2\pi t \frac{i}{2T}) \cos(2\pi t \frac{j}{2T}), \mathrm{d}t = 0 \quad i \neq j. \tag{2.10}$$

The partial symmetry property of IFFT output of the real-valued input was verified in [61]. This provides another method to implement FOFDM by using the efficient IFFT/FFT operations [62]. The truncation of the IFFT output gives the FOFDM symbol. While at the receiver, the conjugate of the received symbols are generated to mend the discarded half of the input for FFT.

Fig 2.4 shows the OFDM and FOFDM signals, where parallel sinc pulses are used as time-domain representations of the 5 subcarriers. The separated real and imaginary parts of the associated subcarriers in a single signalling interval T are given. The FOFDM is shown to have a more compressed subcarrier spacing when compared to that of the OFDM. When singular dimensional modulation schemes are adopted, only the in-phase component of the subcarriers carry the data. As a result, FOFDM system using BPSK violates the orthogonality among multiple subcarriers though, the transmitted signal does not suffer the introduced ICI as proved in the following. The correlation of two arbitrary subcarriers indicates the ICI, given by:

$$\begin{aligned} \Lambda(m,n) &= \frac{1}{T}\int_0^T e^{j2\pi mt/2T} e^{j2\pi nt/2T} \\ &= \mathrm{sinc}(m-n) \\ &\quad + j\sin(\frac{\pi(m-n)}{2})\mathrm{sinc}(\frac{(m-n)}{2}). \end{aligned} \tag{2.11}$$

where $e^{j2\pi mt/2T}$ and $e^{j2\pi nt/2T}$ are the m^{th} and the n^{th} subcarrier respectively, $1/2T$ represents the subcarrier separation as stated in equation 2.11. By extracting the real and imaginary part of the correlation, there are:

$$\Re\{R_c\} = \mathrm{sinc}(m-n), \tag{2.12}$$

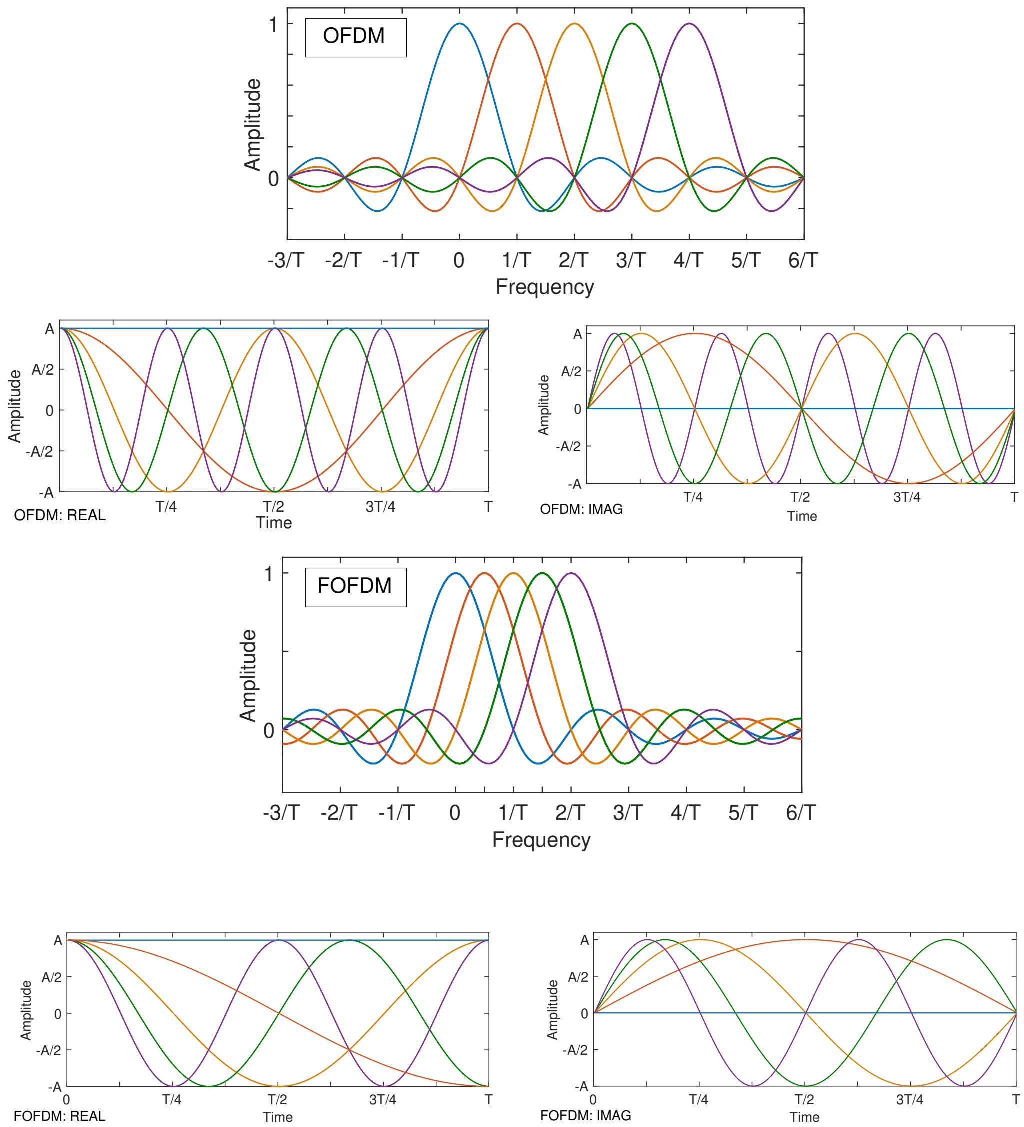


Figure 2.4: OFDM and FOFDM subcarrier representations in frequency domain and time domain. The in-phase and quadruple components of the 5 complex subcarriers are shown as cosine (left) and sine (right), respectively.

for the real part of the correlation and

$$\Im\{R_c\} = \sin(\frac{\pi(m-n)}{2})\mathrm{sinc}(\frac{(m-n)}{2}), \tag{2.13}$$

for the imaginary part, where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary part of the complex expression, respectively, and the sinc function is defined as $\mathrm{sinc}(x) = \sin(\pi x)/x$. It is clear from the equations that, for the real part, the auto-correlation (i.e. when $m = n$) is non-zero and the cross-correlation (i.e. when $m \neq n$) is zero. Even though the cross-correlation for the imaginary part is not equal to zero, the orthogonality rule is still valid since no data are carried on this dimension. Therefore, by using such signal format, the ICI can be removed for the received signal.

Table 2.1: FOFDM: Wireless and optical work in UCL

Reference	Descriptions
Jun.2002[45]	FOFDM: A proposal for doubling the data rate of OFDM
Sept.2002[62]	Effects of linear phase noise on FOFDM
Sept.2003[63]	OFDM over GSM multipath fading environment
Sept.2004[47]	Doctoral thesis on modelling and performances assessment of OFDM and FOFDM
2006[64]	Performance comparison: overlapping MC-DS-CDMA and FOFDM
2007[65]	Performance comparison: FOFDM, overlapping MC-DS-CDMA and MT-CDMA
2008[66]	Detection techniques in FOFDM
2017[50]	FOFDM in VLC
2018[49]	Experimental verification for FOFDM using for eNB-IoT
Apr.2018[67]	Test FOFDM in LOS scenario for NB-IoT
Apr.2019[68]	A co-equalisation method for MIMO FOFDM in VLC
Apr.2019[29]	Quadrupled data rate by using Hilbert filter pair
2019[69]	Duobinary 3-PSK modulation in FOFDM
2021[70]	A combined nonlinearity compensation and channel estimation for optical FOFDM
2021[71]	Experimental demonstration of optical FOFDM
2021[72]	Modeling the nonlinear and memory effects of optical FOFDM

The overview of the studies of FOFDM on wireless and optical communications in and outside UCL are listed in Table 2.1 and Table 2.2, respectively. The initial proposal has shown that theoretical BER performance can be only achieved for

Table 2.2: FOFDM: Wireless and optical work outside UCL

Reference	Descriptions
Mar.2010[73]	IDCT/DCT-based optical 4-ASK FOFDM system
Sept.2010[74]	DFT-based optical 4-ASK SSB FOFDM
Sept.2010[75]	Experimental optical FOFDM at 7.174 Gbit/s and 14.348 Gbit/s
Feb.2011[76]	Symbol synchronisation method Start-of-frame
Mar.2011[77]	A novel precise symbol synchronisation method
Sept.2011[78]	Chromatic dispersion compensation using guard interval
Oct.2011[79]	Investigation of guard interval length on chromatic dispersion compensation
Oct.2011[80]	DBPSK-based DSB FOFDM over MMF
July.2012[81]	Channel estimation and compensation for 840km transmission
Nov.2012[82]	Multiple-tap equaliser for residual frequency offset compensation
Mar.2012[83]	Comparative study of IM/DD and IM/Full-field detection
May.2013[84]	IM/FD optical FOFDM
Sept.2013[85]	Symbol synchronisation and dispersion estimation 16QAM
Sept.2013[86]	Blind maximum-likelihood estimation
Sept.2013[]	fibre nonlinearity mitigation using precoding
Jun.2014[87]	One-tap equaliser for IM/DD optical fast OFDM
July.2014[88]	Interleaved DFT-based 16QAM FOFDM
Aug.2015[89]	Experimental demonstration of IM/DD and IM/FD detection
2017[90]	Frequency oversampling to eliminate ICI
Dec.2017[91]	Performance analysis of FOFDM for VLC
Mar.2018[92]	CSI-free equalisation using OCT precoding
May.2018[93]	Orthogonal circulant matrix transform precoding for set partitioned coded QAM FOFDM
May.2018[85]	I/Q-encryption for FOFDM PON
2019[94]	FOFDM With Index Modulation for NB-IoT

FOFDM when one-dimensional modulation scheme is employed [45]. Inspired by the doubled bandwidth efficiency achieved in multi-carrier CDMA, the IDFT-based framework was applied to binary input FOFDM. The equivalent BER performance and halved bandwidth consumption can be achieved in FOFDM when compared to OFDM with high dimensional modulations such as M-PSK ans M-QAM. In addition to this, IDFT-based FOFDM shows a simpler implementation and its potential robustness in multipath fading environment.

The performance comparison between FOFDM and other multi-carrier wireless communication systems was studied in various aspects [62–65]. Attributed to its 50% increase in bandwidth efficiency, FOFDM was widely adopted in optical communications. The investigations are mainly based in a group in University College Cork and The Chinese University of Hongkong with up to 19 publications and a doctoral thesis [95] from 2010 [73] till present. Apart from the optical communications over fibre, FOFDM was introduced to visible light communications (VLC) with M-PAM modulations employed for both single-input single-output (SISO) [50] and MIMO systems [68]. The one-dimensional modulation of FOFDM, constraining its further improvement in spectral efficiency though, tends to fit into NB-IoT, which is a low power wide area network (LPWAN) technique proposed for one of the 5G scenarios. An experimental test bed based on software-defined radio was established and the feasibility of FOFDM in NB-IoT has been verified in [49, 67]. These applications of FOFDM were also researched with the use of index modulation [94]. Aside from the aforementioned work, other signal processing techniques are applied to FOFDM system, leading to a quadrupled data rate relative to FOFDM signal, which will be detailed in the next chapter.

As discussed, FOFDM, for the same spectral width of OFDM,twice the number of subcarriers can be accommodated, by halving the SCS of OFDM signal. Since each of the FOFDM subcarriers can carry the same symbol rate as in the equivalent OFDM, then FOFDM effectively doubles the spectral efficiency. Advantageously, the error performance remains the same only when one dimensional modulation formats are employed, which constraint the use of high-order modulation schemes to improve spectral efficiency. This leads to the study of feasible facilitation of SEFDM signals, which provides flexible bandwidth compression and works compatibly with high-order modulation schemes. The next section introduces SEFDM by presenting an overview of the signal design, advantages and drawbacks.

2.3.3 Spectrally Efficient FDM

The orthogonality in OFDM allows the overlapping between subcarriers [11]. In the meantime, it sets the limitation that the positions of the multiple subcarriers are strictly fixed. SEFDM extends such overlapping concept of conventional OFDM by violating the orthogonal rule among subcarriers and compressing the frequency spacing deliberately. A mathematical expression of a SEFDM signal is given by [57]:

$$x_{SEFDM}(t) = \frac{1}{\sqrt{T}} \sum_{n=0}^{N-1} s_{n,l} g(t - l \cdot T) \cdot e^{j2\pi n \alpha t/T}, \tag{2.14}$$

$$n, l \in \mathbb{Z},\ n \in [0, N-1].$$

It is worth noting that, as is compared to equation(2.6), a factor α is introduced in the SEFDM signal. This factor, termed the bandwidth compression factor, can be described as [57]:

$$\alpha = \frac{\Delta f_{SEFDM}}{\Delta f_{OFDM}}, \tag{2.15}$$

which can also be described as the reciprocal of the SEFDM symbol duration. The fraction α determines the level of bandwidth compression in comparison to the bandwidth separation of adjacent OFDM subcarriers. Thus, $\alpha = 1$ for a SEFDM signal is equivalent to an OFDM signal. Fig. 2.5 shows the spectra of a SEFDM signal. In this case, the overlapping subcarriers are non-orthogonal such that at the peak point of any subcarrier the power density of other subcarriers are not zero.

Theoretically, the overall bandwidth of the spectrum of an FDM signal consists of N frequency spacing and two of half the subcarrier band, located at the start and the end of the spectrum. The frequency range of each subcarrier is $\frac{1}{T}$, where T is the input symbol period. For the given SEFDM spectrum above, the occupied bandwidth equals to:

$$BW_{SEFDM} = (N-1)\Delta f_{SEFDM} + \frac{2}{T} = \frac{\alpha(N-1)+2}{T}. \tag{2.16}$$

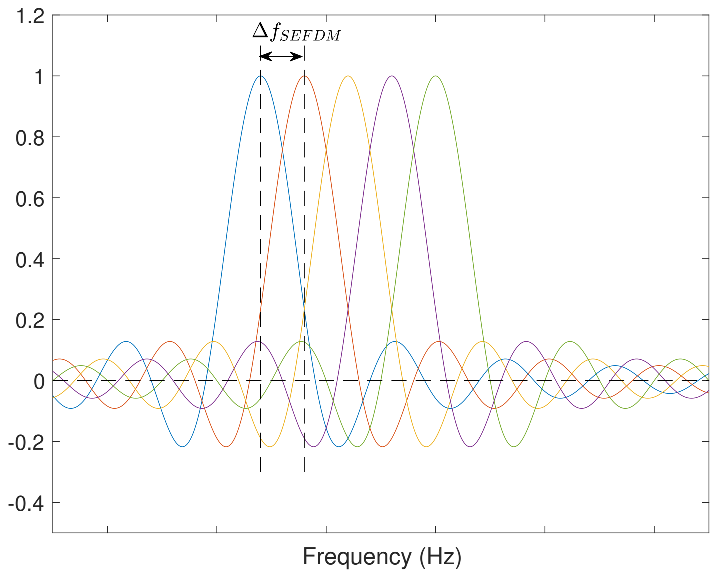


Figure 2.5: Spectra of an SEFDM signal

Recalling OFDM spectrum in 2.4, the bandwidth is given by:

$$BW_{OFDM} = (N-1)\Delta f_{OFDM} + \frac{2}{T} = \frac{(N+1)}{T}. \tag{2.17}$$

As is shown in equation 2.16 and 2.17, when the number of subcarriers N is considerably large, the ratio of the occupied bandwidth of a SEFDM signal and an OFDM signal approaches α.

Similar to OFDM, SEFDM can be generated using IFFT with appropriate modifications [96]. Relative to OFDM, SEFDM compresses the bandwidth by factor α and thus improves the spectral efficiency by $(1/\alpha - 1) \times 100\%$, which is called type-I SEFDM [97]. Another form of SEFDM termed as type-II SEFDM [97] achieves higher spectral efficiency by increasing the data rate for each subcarrier while keeping the subcarrier separation the same as OFDM.

Although the compressed frequency spacing provides the SEFDM transceiver with a more flexible and bandwidth efficient structure, the transmission performance degrades under this scheme. The SEFDM signal is not only contaminated by the noise from the channel but also suffers the additional self-created interference, termed ICI [10, 11]. In an OFDM system, the ICI mainly results from the frequency offset due to the Doppler spread in multipath channels. Besides, the phase noise associated with local oscillators and the mismatches between the transmitter and the receiver oscillators also result in frequency offset to the system [11]. While in the SEFDM system, the loss of orthogonality makes the scenario more complicated. Theoretically, the cross-correlation among all subcarriers is utilised to estimate the energy leakage during the transmission. Thereby, similar to equation 2.11, a correlation matrix is built up to describe the ICI in the SEFDM system mathematically and wherein each component can be derived by [34]

$$\Lambda(m,n) = \frac{1}{Q}\begin{cases} Q & \text{if } m = n \\ \dfrac{1-e^{j2\pi\alpha(m-n)}}{1-e^{j2\pi\alpha(m-n)/Q}} & \text{if } m \neq n \end{cases} \tag{2.18}$$

where Q denotes the number of samples per subcarrier, m and n are the indexes of two arbitrary subcarriers and α is the compression factor as mentioned above. The auto-correlation elements of Λ, i.e., the diagonal, are all of value one while the non-diagonal elements (when $m \neq n$) represent the interference between the two subcarriers. The interference power level decreases as the α increases, indicating that SEFDM with higher spectral efficiency gain encounters more severe ICI. For a detailed statistical characteristic analysis of ICI in SEFDM, readers are referred to [34] and [98] for the treatment of the introduced ICI.

A series of advanced detection algorithms have been studied to mitigate the ICI effect so as to maintain good BER performance for SEFDM with high compression level [8]. Among these techniques, sphere decoder (SD) has been applied to SEFDM and shows significant improvement in error performance within the range of poor SNR [99]. The mathematical algorithm of SD searches for the symbol sequence that is closest to the received symbols in terms of Euclidean distance. The search is performed over the surface of a sphere in a high-dimensional space so that only a small region of signal space needs to be searched, which reduces the computation burden when compared to other detection methods such as maximum likelihood (ML) and minimum mean square error (MMSE) [99]. Numerical results in Fig. 2.6 provides a good summary showing the BER for SEFDM system under

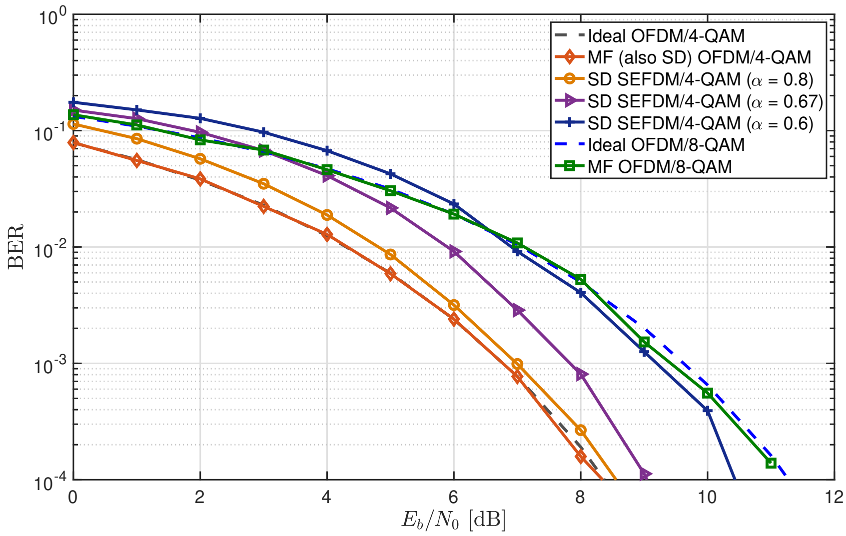


Figure 2.6: BER performance for uncoded SEFDM and OFDM systems.

4-QAM modulation with varying values of $\alpha = 0.8$, 0.67 and 0.6. The SEFDM signal is detected using the SD algorithm. It can be seen in Fig. 2.6 that the SEFDM with $\alpha = 0.6$ has similar BER as the OFDM modulated with 8-QAM. Besides, the SEFDM with $\alpha = 0.8$ obtains approximately the same BER as OFDM modulated with 4-QAM when the E_b/N_0 is higher than 5 dB.

Another benefit of SEFDM is its higher power efficiency when compared to OFDM system. Power efficiency is another important parameter when assessing the effectiveness of communication in both uplink and downlink. Such efficiency is commonly evaluated quantitatively in communication systems by the system feature peak-to-average power ratio (PAPR), of which the calculation is detailed in Appendix C. As 4G LTE sheds substantial light on OFDM in physical layer transmission, the significant drawback of having high PAPR in OFDM signal transmission has always been a problem. Techniques are developed and solutions are tested to reduce the PAPR so as to enhance the power efficiency and subsequently lower down mobile system deployment cost. Detailed investigations are provided in [100] on the PAPR performance of SEFDM relative to OFDM and a variety of PAPR reduction techniques. Some of the numerical results are reproduced in simulations, as shown in Fig. 2.7. The complementary cumulative distribution function (CCDF)

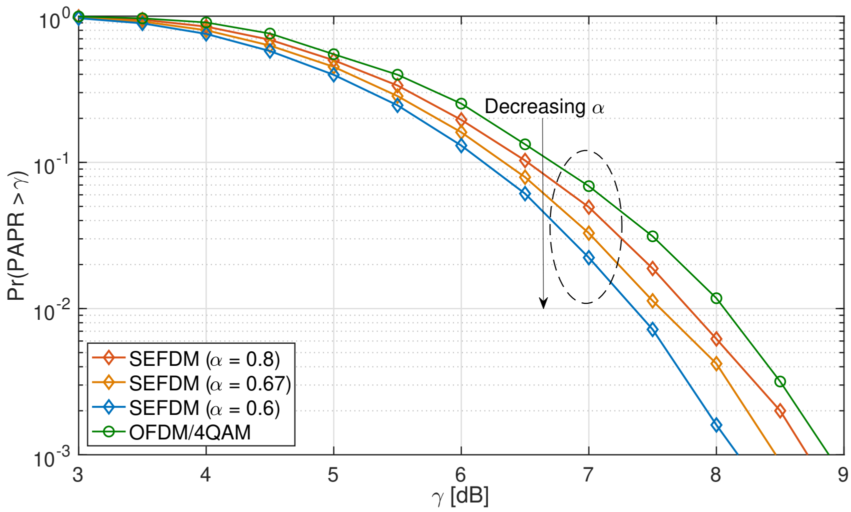


Figure 2.7: Comparison of CCDF of PAPR for OFDM and SEFDM with varying α.

of PAPR[6] of OFDM signal modulated with 4-QAM is given for comparison. This figure depicts the CCDF of the PAPR of the SEFDM/4-QAM signal exceeding a threshold of γ for different value of α. The probability increases as expected with the reduction of α, which indicates the improving power efficiency.

The aforementioned advantages of SEFDM, coupled with new efficacious techniques for channel estimation and equalisation [101–103], make SEFDM an attractive candidate for future communication systems derived from and coexisting with 5G NR. A brief survey on SEFDM is provided in [8], covering the techniques that have been applied to SEFDM including signal generation, reception, channel estimation and equalisation. In the next section, to provide insights of the advantageous spectral efficiency gains of SEFDM, we assess its coexistence with OFDM in 5G NR frame by presenting a series of simulation-based studies, where new coexistence system models are proposed and their performance is analyzed and evaluated. These studies provide valuable insights into the coexistence of non-orthogonal SEFDM and OFDM, enabling us to understand the trade-offs between spectral efficiency and error performance as well as the effects of the compression factor and the SCS to the performance with and without the use of coding scheme.

[6]The calculation of PAPR and the CCDF of PAPR are included in Appendix C

2.4 Coexistence Study of SEFDM in 5G Scenarios

To meet the requirements of the use scenarios in IMT-2020, the 3GPP proposed the new radio-access standard for 5G system, termed as 5G NR [3, 39]. One of the primary purposes of 5G NR is to support heterogeneous system implementation. Hence, higher flexibility is needed so that different services can be provided to devices using the same radio resources. As such, the feasibility of coexistence of different signalling formats, configured in the same range of time-frequency resources, needs to be studied and assessed. 5G NR maintains the OFDM waveform format (as used in 4G-LTE [59]) as specified in Release 15 [3]. To handle different scenarios, 5G NR offers higher spectrum flexibility through defining flexible numerology for its physical layer specifications. Rather than the fixed SCS of LTE, 5G NR supports five sets of SCSs of 15/30/60/120/240 kHz.

2.4.1 Scenario-based System Design

The ultimate goal of this work is to investigate the effects of coexisting signals when they are transmitted simultaneously under 5G NR. To this purpose, coexisting signalling is categorised into three main scenarios given by Table 2.3: OFDM-OFDM, OFDM-SEFDM and SEFDM-SEFDM coexistence. To simplify the problem, this study assumes two signals - signal 1 with SCS of Δf_1 and signal 2 with SCS of Δf_2 are transmitted and received by different user equipment (UE). The two signals are independent; however, they are allocated bandwidth part (BWP) without guard band. It is worth noting in Table 2.3 that the α for SEFDM signal is noted to show the reduced SCS when compared to OFDM, of which the underlying $\alpha = 1$.

More specifically, two OFDM signals of different SCS are defined in scenario-I. In order to assess the effects of applying flexible SCS, one of the OFDM signal is maintained with fixed SCS $\Delta f_1 = 15$ kHz meanwhile vary the coexisting signal SCS $\Delta f_2 = 15, 30, 60, 120, 240$ kHz. Scenario-II considers different types of waveforms such that OFDM and SEFDM signals coexist occupying the neighbouring BWPs. Scenario-III specifies the coexistence of two non-orthogonal signals; one has fixed $\alpha_1 = 0.8$ and the other's α_2 varies. To illustrate, scenario-II is used as an example in Fig. 2.8.

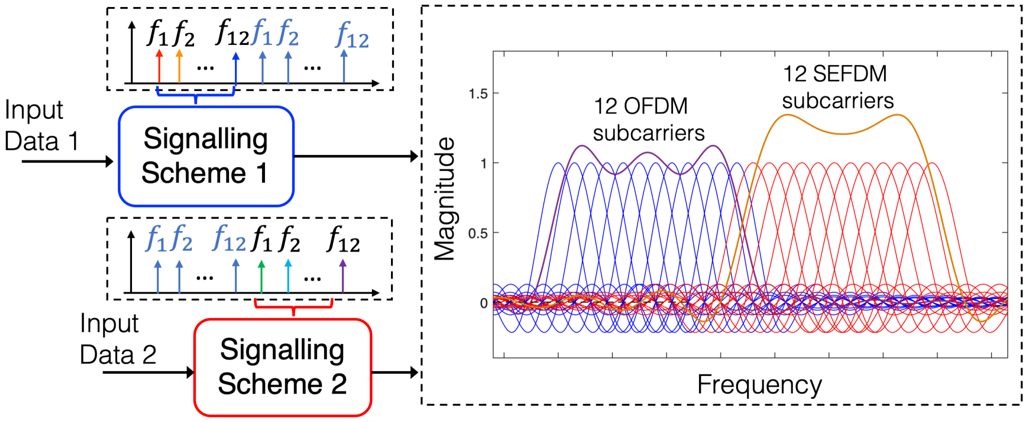


Figure 2.8: Coexisting Scenario-II OFDM-SEFDM subcarrier representation. Similar subcarrier placement with different parameters will be used for Scenario-I and Scenario-III.

Table 2.3: Three types of coexistence scenario under 5G NR numerologies

Scenario	**Signal 1**	α_1	Δf_1 [kHz]	**Signal 2**	α_2	Δf_2 [kHz]
Senario-I	OFDM	1	15	OFDM	1	15/30/60/120/240
Senario-II	OFDM	1	15	SEFDM	0.8	12
				SEFDM	0.67	10
				SEFDM	0.6	9
Senario-III	SEFDM	0.8	15	SEFDM	0.8	12
				SEFDM	0.67	10
				SEFDM	0.6	9

For simplicity, 12 subcarriers modulated with 4-QAM scheme are used for both OFDM and SEFDM signals. This follows the standard where each resource block consists of 12 subcarriers [3]. The centre carrier frequency is set to 3.5 GHz which is in the new spectrum range included in the 5G NR technology specifications. This is to match the 5G numerologies that are used for the flexible SCS settings. Table 2.4 provides the detailed specifications for the numerical simulations.

2.4.2 BER Performance Evaluation

This section presents the numerical simulation results commencing with the specifications of the system models. The BER performance for signals transmitted in three

Table 2.4: Specifications for the numerical simulated coexistence system

Parameter	**OFDM**	**SEFDM**
Centre Carrier frequency [GHz]	3.5	3.5
Sampling Frequency [MHz]	1.92	1.92
IFFT Output Size	128	128
Number of Subcarriers	12	12
Modulation Scheme	4-QAM	4-QAM
Bandwidth Compression Factor α	1	0.8, 0.67, 0.6
Data Bandwidth [kHz]	180	144, 120, 108
Data Rate [kbit/s]	180	180

different scenarios is evaluated for the uncoded systems. Besides, the block error rate (BLER) is assessed against practical SNR for the same systems with LDPC coding.

The BER performance for the single-antenna systems in the three different scenarios is examined for different values of E_b/N_0. Extensions to multi-antenna is straightforward with minor system architecture modifications [97]. For comparison, the ideal error performance for OFDM modulated with 4-QAM scheme is provided in the results in the following section. The simulations assume that the received signal is contaminated only by AWGN.

For scenario-I, Fig. 2.9 displays the power spectra for the two independent OFDM signals with normalised amplitude and frequency. Numerical simulations are carried out to assess the performance of coexisting OFDM signals with different values of SCS as listed in Table 2.3. The signal detection at the receiver employs matched filtering and the corresponding BER curves of the associated OFDM signals are presented in Fig. 2.10. Apart from the first coexisting system where $\Delta f_1 = \Delta f_2 =$ 15kHz, evident BER degradation appears in the rest of coexisting systems due to the inter-numerology interference. The signals with large SCS are robust to the inter-numerology interference while those with small subcarrier spacing are more susceptible to the interference resulting in error floors when E_b/N_0 is further increased.

Fig. 2.11 presents the error performance of the coexisting signals in scenario-II, where independent OFDM and SEFDM signals are demodulated and detected

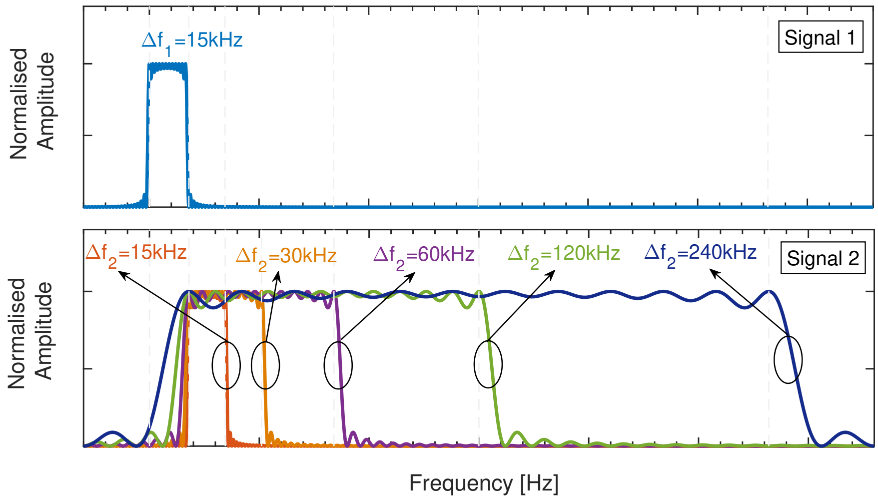


Figure 2.9: Scenario-I: power spectrum with normalised amplitude for OFDM signal 1 (top) and OFDM signal 2 with varying SCS (bottom).

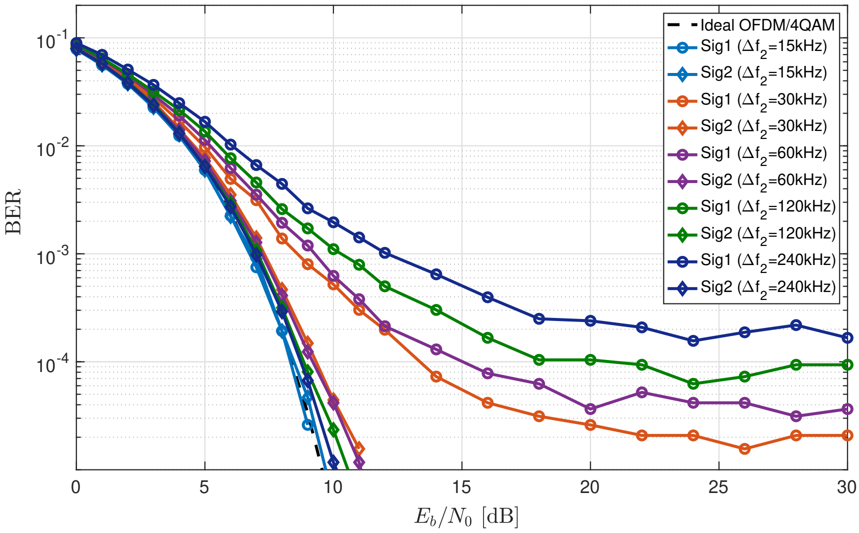


Figure 2.10: Scenario-I: BER performance for OFDM-OFDM coexisting systems with different subcarrier spacing. Matched filtering is used for signal detection.

separately using matched filtering and SD, respectively. Herein, SD is used to maintain good BER performance for SEFDM since the orthogonality is deliberately violated and consequently the self-introduced ICI compromises the signal recovery. Other advanced detection algorithms are required to maintain good BER performance. SD is chosen due to its significant improvement in error performance within the range of poor SNR in SEFDM system [99]. Since signals of different

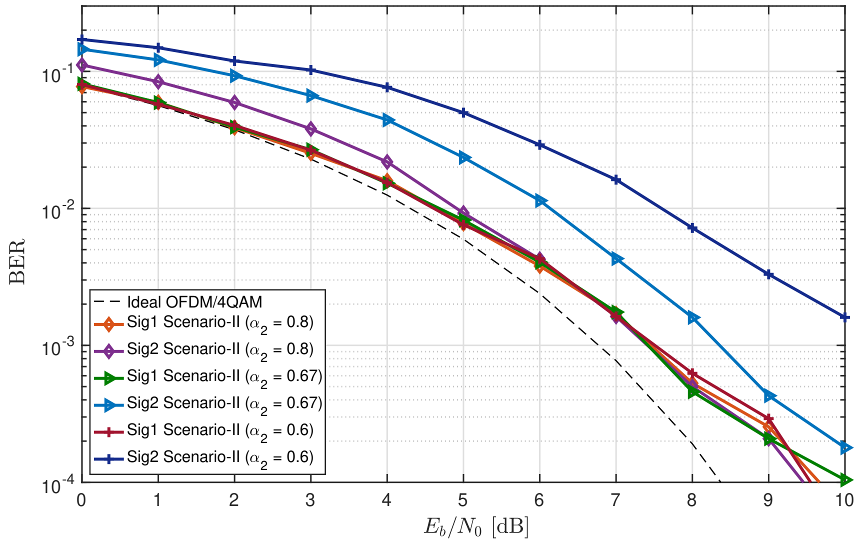


Figure 2.11: Scenario-II: BER performance for OFDM-SEFDM coexistence with different subcarrier spacing due to varying values of α for SEFDM signal. SD is used for SEFDM signal detection. For OFDM signal detection matched filtering (also SD) is adopted.

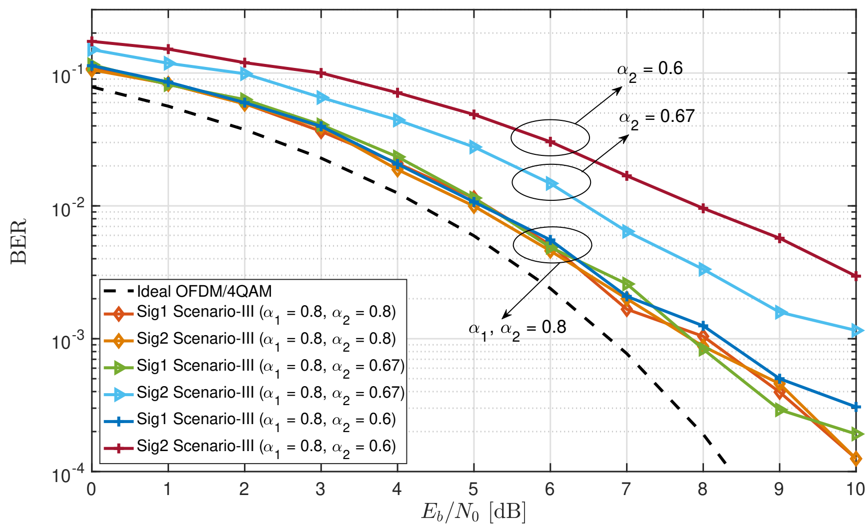


Figure 2.12: Scenario-III: BER performance for SEFDM-SEFDM coexistence. Sphere decoder is used for signal detection.

formats are transmitted in the adjacent BWPs, SEFDM signals interfere with the OFDM signals and vice versa. The BER is noise-dominant at the high-noise range, whereas it turns to interference-dominant when the noise is relatively low. Thus, it can be seen that the BER performance of OFDM signals follows the ideal values at

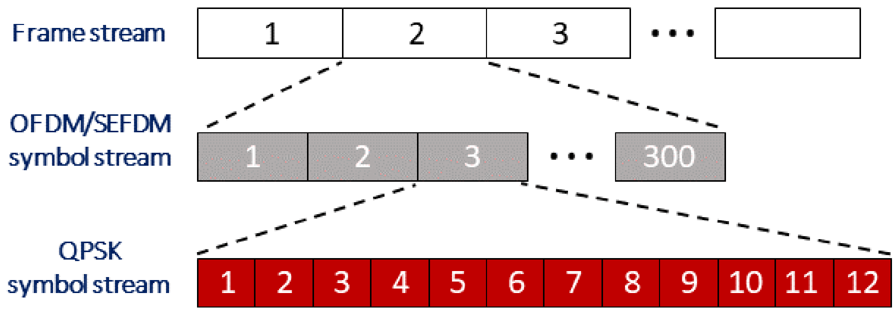


Figure 2.13: Radio frame for coexistence scenarios.

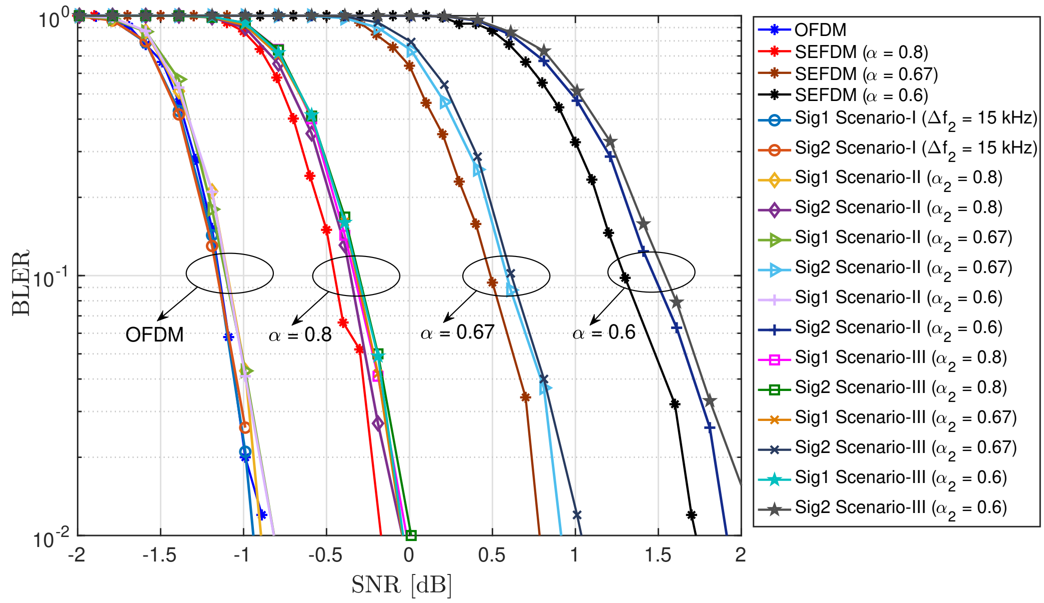


Figure 2.14: BLER performance for the coexisting systems in three scenarios using LDPC coding, N = 12, coding rate R_c = 1/3.

high-noise levels whilst slightly deviates when E_b/N_0 is larger than 3 dB.

The BER performance for coexisting SEFDM signals with varying α in scenario-III is illustrated in Fig. 2.12. Since both signal waveforms are non-orthogonal, under similar circumstances severe interference is shown in the BER degradation due to the high-level of bandwidth compression, especially when E_b/N_0 is greater than 5 dB. Consequently, LDPC is considered for the system to achieve error performance gain by performing interference cancellation in the next section.

2.4.3 Improvement using LDPC

Channel coding is commonly used in communication systems for the mitigation of channel impairments. Due to its high achievable data rate and low implementation complexity, LDPC [104] is used as standardised coding scheme for data channel for 5G NR [3]. Hence, LDPC coding is adopted for interference cancellation for the coexistence scenarios. A fixed coding rate of $R_c = 1/3$ is used for both SEFDM and OFDM signals and the 16-bit cyclic redundancy check (CRC) to assist BLER for both signals at $R_c = 1/3$. Each frame (block) is defined to have 300 OFDM/SEFDM symbols, and each multi-carrier symbol consists of 12 modulated 4-QAM symbols as shown in Fig. 2.14. There are four puncturing strategies in 5G LDPC rate matching. For simplicity, this work follows the first puncturing scheme, redundancy version RV0 [3]. For the channel decoding on the receiver side, the typical belief propagation decoding algorithm [104] is applied with 50 iterations.

To illustrate the practical advantages, SNR is used instead of theoretical E_b/N_0 for error performance assessment[7]. Fig. 2.14 shows the BLER performance for the three scenarios of the different systems with varying configurations. For each system, the performance of two independent coexisting signals is evaluated separately. It is shown that all orthogonal signals achieve the same BLER performance whilst non-orthogonal signals have less than 0.3 dB variations when employing LDPC coding, regardless of the scenarios changes. The BLER categories associate with the four different values of α, meaning an increasing bandwidth efficiency as α decreases at the cost of system performance. This leads to the conclusion that the negative effects of coexistence of orthogonal and non-orthogonal signals can be eliminated by adopting LDPC coding.

This work studies the coexistence of orthogonal and non-orthogonal multi-carrier signals under the 5G NR frame. Three typical coexistence scenarios are investigated, namely OFDM-OFDM, OFDM-SEFDM and SEFDM-SEFDM coexistence, where varying subcarrier spacings are considered under the flexible 5G NR numerology. This work reports the effects of the coexistence signalling using simu-

[7]The conversion between SNR and E_b/N_0 is specified in Appendix B

lations, where both uncoded and LDPC coding assisted systems are studied. Results show that there is minor BER degradation for uncoded coexisting systems with non-orthogonal signals. The coexistence effects are ameliorated by LDPC coding and the BLER results demonstrate only negligible degradation for coexisting signals when compared to single signalling format system. This work offers insight for future heterogeneous system implementation considering non-orthogonal signalling coexistent with 5G NR.

2.5 Conclusions

The main goal of this chapter is to provide the foundation for the work of this thesis by presenting a comprehensive overview of advanced modulation techniques aimed at achieving higher spectral efficiencies. The chapter commences with a theoretical analysis of spectral efficiency in relation to Shannon limit and channel capacity. Then a review of the development of techniques throughout generations of communication networks is provided. Three waveform/signalling techniques, OFDM, FOFDM and SEFDM, are highlighted with brief descriptions of their respective transceiver designs and basics of their system mathematical modeling. OFDM improves the spectral efficiency relative to conventional FDM system by overlapping the subcarriers without introducing ICI. FOFDM further compresses the subcarrier separation by half based on OFDM scheme while the orthogonality is only maintained when singular modulation scheme is used. SEFDM provides a more flexible signalling scheme, where the subcarrier separation is not fixed when compared to OFDM. At the meantime, the improved spectral efficiency of SEFDM relative to OFDM is at the cost of slight increased system complexity to ameliorate the effects from the self-introduced ICI. In other words, both OFDM and FOFDM may be regarded as special (or extreme) cases of SEFDM when the compression factor $\alpha = 0.5$ and 1. A commonly used α in SEFDM system is between the aforementioned values such that $0.5 < \alpha < 1$. The advantages and disadvantages of these three system are discussed in order to address the trade-offs between the spectral efficiency and the system complexity. SEFDM has been proved to more power efficient as compared to OFDM in terms of the measure of PAPR. To provide an insight of introducing applications of SEFDM to 5G and beyond communication systems, new work is reported in the chapter in studies of coexistence of OFDM and non-orthogonal SEFDM, showcasing the capability of mixed transmission of different signal formats under the flexible numerology in 5G NR. As the thesis focuses on various spectrally and energy efficient techniques, the systems discussed in this chapter serve as the foundation for the subsequent chapters, which delve into specific techniques used to address the main challenges of this research.

Chapter 3

Spectrally Efficient Design and Performance of Hilbert Filter Pair

This chapter presents the design and simulated verification of a filtering technique proposed to apply an additional layer of orthogonality using Hilbert filter pair, yielding higher spectral efficiency of MCM systems. The chapter includes a summary of the developed studies of pulse shaping filters on subcarrier level and waveform designs that have been proposed as candidates for 5G and beyond communication systems. Two studies are conducted on the performance of the new waveform design, with and without the use of channel coding, for cellular communication systems and use cases for internet of things (IoT); the proposed waveform format using Hilbert transform filter pair and the designed transceiver model are applied to the aforementioned FOFDM and SEFDM system, respectively. The developed systems are modelled and simulated in this chapter with different configurations for specific use scenarios for performance investigations. Work demonstrated in this chapter includes results presented in three conference papers [28, 29, 105] [1].

The outline of this chapter is as follows: section 3.1 provides a brief review of current spectrally efficient techniques using pulse shaping filters, highlighting

[1] *i)* X. Liu and I. Darwazeh, "Doubling the Rate of Spectrally Efficient FDM Systems Using Hilbert Pulse Pairs," 2019 26th International Conference on Telecommunications (ICT). *ii)* X. Liu and I. Darwazeh, "Quadrupling the Data Rate for Narrowband Internet of Things without Modulation Upgrade",2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring). *iii)* X. Liu et al., "Energy and Spectrally Efficient Signalling for Next Generation IoT", 2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP).

the benefits and drawbacks. Section 3.2 introduces the mathematical form of the Hilbert transform, analysing the orthogonal features of the Hilbert filter pair and its associated applications in communication systems. Based on such advantageous features, applications in two use scenarios are detailed in section 3.3 and 3.4, where the Hilbert filter pair is proposed for use with SEFDM and FOFDM. The newly designed transceiver models and corresponding systems are detailed. The advantages of simple implementation and enhanced spectral efficiency are evaluated. Finally, section 3.5 draws the conclusions.

3.1 Waveform Designs using Pulse Shaping Filters

Waveform/signalling is the most fundamental aspect of physical layer communications and has changed significantly at each generation to support the corresponding technical features. For example, in 4G LTE, CP-OFDM is used for downlink transmissions, and its low PAPR variant, DFT-s-OFDM, is used for uplink transmissions, providing efficient support for the mobile broadband (MBB) service. In contrast to LTE, 5G NR is designed to support three main scenarios as specified in IMT-2020 by International Telecommunications Union (ITU): eMBB leveraging the foundation of the Gigabit LTE (20Gb/s peak data rate); URLLC to deliver broader latency and reliability sensitive services (less than 1 ms latency); mMTC to support efficient connection for a variety of diverse services such as massive IoT ($10^6/km^2$ device density) [106]. To this end, two orthogonal waveforms are supported in 5G NR; CP-OFDM is used for uplink and downlink to simplify the overall design by having the same waveform for both directions; DFT-s-OFDM for coverage-limited single-stream transmissions in uplink [3]. Notwithstanding, these OFDM-based waveform formats suffer from large spectral side lobes due to the rectangular pulse shape in the time domain. The high side lobe leads to high out-of-band (OOB) emission; therefore, wide frequency guard bands are needed to prevent adjacent channel leakage. For example, 10% band is used as a guard band in 4G LTE and 1-3% in 5G NR to give space for signal attenuation [107]. Low-OOB emission means better frequency localisation, which relaxes the requirement on synchronisation and therefore benefits in supporting asynchronous access.

The high OOB emission mainly results from the discontinuity between the adjacent symbols. Windowing operation is one way to smooth the transition between the neighbouring symbols. This is usually performed by extending the original CP or adding a new GI. Benefit from its low implementation complexity (i.e., time domain multiplication and addition operations), though, windowed OFDM (W-OFDM) achieves only limited OOB emission suppression at the cost of the length of effective CP and therefore has reduced efficiency in multipath channels [106].

Sub-band filtering also suppresses the OOB emission. Filter OFDM [25] employs filters of arbitrary bandwidth larger than 180 kHz (i.e., one physical resource block in 4G-LTE) to reduce the OOB emission. By splitting the assigned band into sub-bands of OFDM waveform with different configurations, such as SCS and CP length, filter OFDM allows flexibility of different sub-band to be designed to suit the needs of certain services, retaining all the advantages of OFDM meanwhile achieving reduced OOB emission. UFMC [108] is another waveform format that performs filtering to the sub-bands to reduce OOB emission. Different from filter OFDM, the bandwidth of each sub-band for UFMC is fixed to 180 kHz. Moreover, instead of CP, UFMC employs a zero prefix to avoid ISI; thus, the filter length is limited to the length of the zero prefix. As a result, UFMC provides better OOB suppression than CP-OFDM. However, without CP, the cyclic convolution property of the symbol is not preserved, leading to increased computational complexity at the receiver.

Sharing the same aim to reduce the OOB emission, FBMC employs the pulse shape filter at subcarrier level rather than sub-band level as in UFMC and filter OFDM. In FBMC, a set of prototype pulse-shaping filters with a much longer pulse duration than the symbol duration is used to provide a more confined spectrum. Nevertheless, due to the longer pulse duration, the FBMC symbols are overlapped, resulting in ICI. To minimise the introduced interference, a signal synthesis structure termed offset-quadrature amplitude modulation (OQAM) implements FBMC by maintaining the signal orthogonality in the real domain [23]. The input QAM symbols are divided into two streams, denoted by their odd and even index. The two symbol streams are then split into its in-phase and quadrature components. Once the quadrature components of each stream have been exchanged, the QAM symbols are sent to the FBMC waveform synthesis and then transmitted with a time offset introduced on the even indexes stream [23, 109]. FBMC achieves higher spectral efficiency by saving on both CP and guard band. In the meantime, the smooth edge of the pulse shape of FBMC waveform makes it robust against delay spread in multipath channels without the need for CP [106].

GFDM is the other waveform format that uses subcarrier-level filtering for pulse shaping. Similar to FBMC, GFDM employs prototype filters. However, the variable pulse shaping filters break the orthogonality between multiple subcarriers; hence the GFDM waveform is non-orthogonal. Compared to the ordinary OFDM scheme, the mechanism of GFDM makes the implementation more flexible and simple [110]. In GFDM, the FDM symbol stream is grouped into a data block of a certain number of subcarriers modulated by a set of sub-symbols. The tail biting technique is adopted to provide a shortened CP in each block to reserve the circular structure [111]. On the receiver side, interference cancellation (IC) is a necessary process of GFDM scheme to cope with the inevitable self-introduced ICI [112]. This process can remove the interference from both adjacent subcarriers simultaneously. Nevertheless, like FBMC, the GFDM suffers from high complexity in receiver design resulting from the large-size FFT and IC procedure [109].

In this work, the novel signal design utilises a special pulse shaping filter pair, the Hilbert transform filter pair, to improve the spectrum utilisation efficiency and reduce the OOB emission power. The design takes advantage of the orthogonality feature of Hilbert transform pair, which are formed by two orthogonal components. Therefore, two path of independent data can be transmitted simultaneously. The mathematical model of the newly designed transceiver is provided in detail later in this chapter, as a part of the system design for performance investigations. The following section briefly introduces the basic concept and theoretical fundamental of the Hilbert transform and the Hilbert transform pair.

3.2 Hilbert Transform and Hilbert Filter Pair

Amongst existing techniques in wireless communications, Hilbert transform promises significant bandwidth efficiency enhancement due to its unique mathematical properties. This operation was initially used to define the analytic signal, which is a complex signal that specifies instantaneous amplitude and frequency [113]. In 1962, Bedrosian proposed the analytic signal representation for the SSB modulation [114], which offers generalised formulation to obtain the one-sided spectrum.

SSB modulation is a commonly used amplitude modulation (AM) in communication systems. For a real passband signal, the AM spectrum consists of a lower sideband and an upper sideband, which are uniquely related by hermitian symmetry about the carrier frequency [115]. Since one side-band is the mirror image of the other side-band, half of the spectrum is wasted due to the repetition of information. Hence, SSB can be used to suppress one of the sideband, and consequently to reduce the bandwidth occupation [116].

Among many conceptual approaches to generate SSB, Fig. 3.1 shows a commonly used structure by using the phase-shift method. The input signal $x(t)$ is a normal double side-band modulation (DSB) signal. As is suggested in the figure, two DSB signals are utilised to obtain SSB, one of them is sent through the quadrature filter $\mathscr{H}$. As detailed in Appendix A, the quadrature filter performs the function that all the positive frequency spectrum get a $+90°$ phase-shift. The quadrature carriers are used to modulate the signal $x(t)$ and $\hat{x}(t)$ respectively. After the previous process, the phase of the signal spectrum on one path is changed such that one side-band can be cancelled as shown in Fig. 3.2 [116].

Mathematically, the output of the quadrature filter is the Hilbert transform of the DSB signal, which can be expressed as [115]:

$$\hat{x}(t) = \mathscr{H}(x(t)) \tag{3.1}$$

Since one path of the signals is multiplied by the cosine function directly and an-

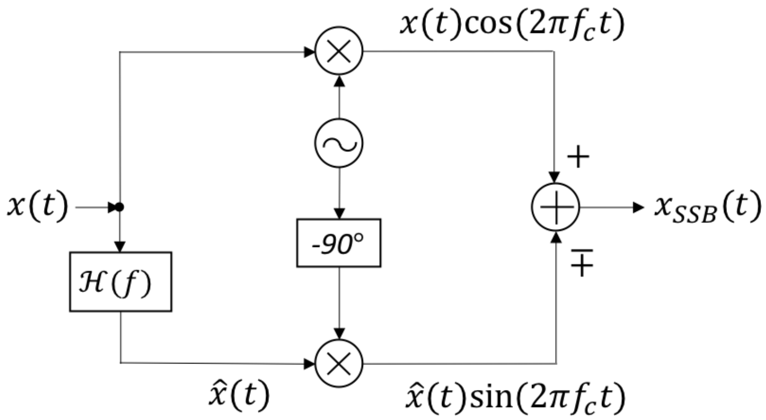


Figure 3.1: The block chart of SSB generation

other one is multiplied by a sinusoidal function, the output SSB signal can be given by [116]:

$$x_{SSB}(t) = x(t)\cos(2\pi f_c t) + \hat{x}(t)\sin(2\pi f_c t). \tag{3.2}$$

The cancellation of half the sideband can be obviously visualised in frequency domain. Thus, the process of using the phase-shift method can be explained by using the frequency representation as below [116]:

$$\begin{aligned}
X_{SSB}(f) &= X(f) * \mathscr{F}\{\cos(2\pi f_c t)\} + \hat{X}(f) * \mathscr{F}\{\sin(2\pi f_c t)\} \\
&= X(f) * \frac{1}{2}(\delta(f - f_c) + \delta(f + f_c)) + \hat{X}(f) * \frac{1}{2}j(\delta(f - f_c) - \delta(f + f_c)) \\
&= X(f) * \frac{1}{2}\delta(f - f_c) + \hat{X}(f) * \frac{1}{2}j\delta(f - f_c) \\
&= X(f) * \delta(f + f_c)
\end{aligned} \tag{3.3}$$

where $\mathscr{F}$ is the Fourier transform operator. There is no priority to keep the upper sideband or the lower sideband to carry the information. However, due to the fact that the quadrature filter is unreliable in practice, only approximation can be used with additional networks to perform the same functionality. The implementation of Hilbert transform might introduce extra distortion to the modulated signals [116]. For further information about the mathematical formulations and properties of Hilbert transform, readers are encouraged to refer to Appendix A

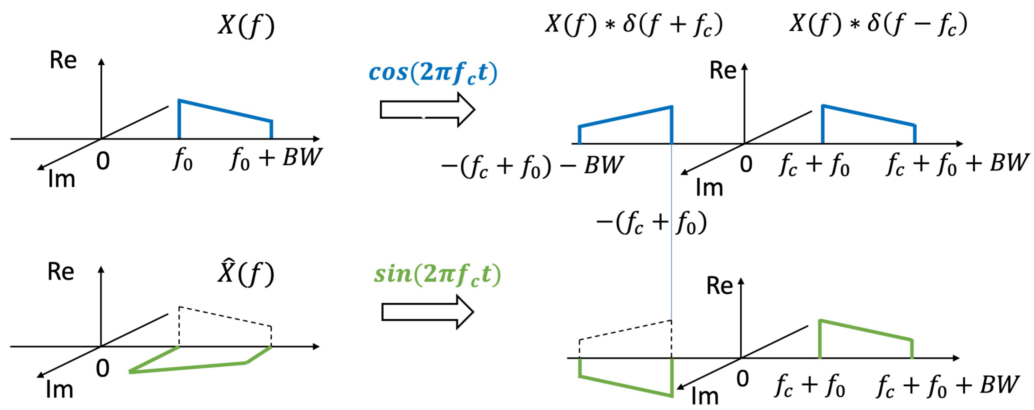


Figure 3.2: SSB modulation: baseband and passband spectrum

Due to its property of quadrature filtering, Hilbert transform is used in carrierless amplitude and phase (CAP) modulation, which has been proposed for use in optical communication systems, with a particular utilisation in visible light communications (VLC) [117, 118]. In CAP modulation, finite impulse response (FIR) filters are employed to form a Hilbert pair. The I/Q components are separated, filtered and finally added up before transmission. Inspired by the CAP modulation, the orthogonal Hilbert transform pair is adopted as pulse shaping filter pair to improve the spectral efficiency of multi-carrier signals by employing the Hilbert superposition. In the next section, two different signal formats, the turbo-coded SEFDM signal using Hilbert filter pair and FOFDM signal with BPSK and PAM-M schemes for NB-IoT use scenario, are studied and shown to offer the spectral efficiency improvement and reduced OOB emission at cost of negligible complexity increase.

3.3 Turbo-coded SEFDM with Hilbert Filter Pair

In the first system model, a new multi-carrier signal format for SEFDM system is proposed to further improve the spectral efficiency, where a Hilbert pair is utilised as pulse-shaping filters. The SRRC pulse is employed to generate the Hilbert pulse pair at the transmitter and an equivalent matched filter configuration is used to generate the receiver Hilbert pair. To verify the data rate gain of the proposed system, the mathematical models are designed and simulations are carried out with varying compression factors for SEFDM signals.

3.3.1 Modulation Scheme

Fig. 3.3 (a) depicts the block diagram of the transmitter of the proposed system. The incoming binary data stream is mapped either uncoded or with turbo coding of rate (1/3). Assuming M-QAM modulation, $2N$ complex symbols $\mathbf{S} = [s_{l,0}, s_{l,1}, \ldots, s_{l,2N-1}]$ are generated, where l represents the time sample index. Specifically, 4-QAM modulation scheme is used in the mapper block shown in Fig. 3.3 (a), indicating that complex symbols are generated with I and Q components. The symbol stream is then split into two blocks of $N = 16$ parallel lower-rate substreams vectorised as $\mathbf{S}_1 = [s_{l,0}, \ldots, s_{l,N-1}]$ and $\mathbf{S}_2 = [s_{l,N}, \ldots, s_{l,2N-1}]$

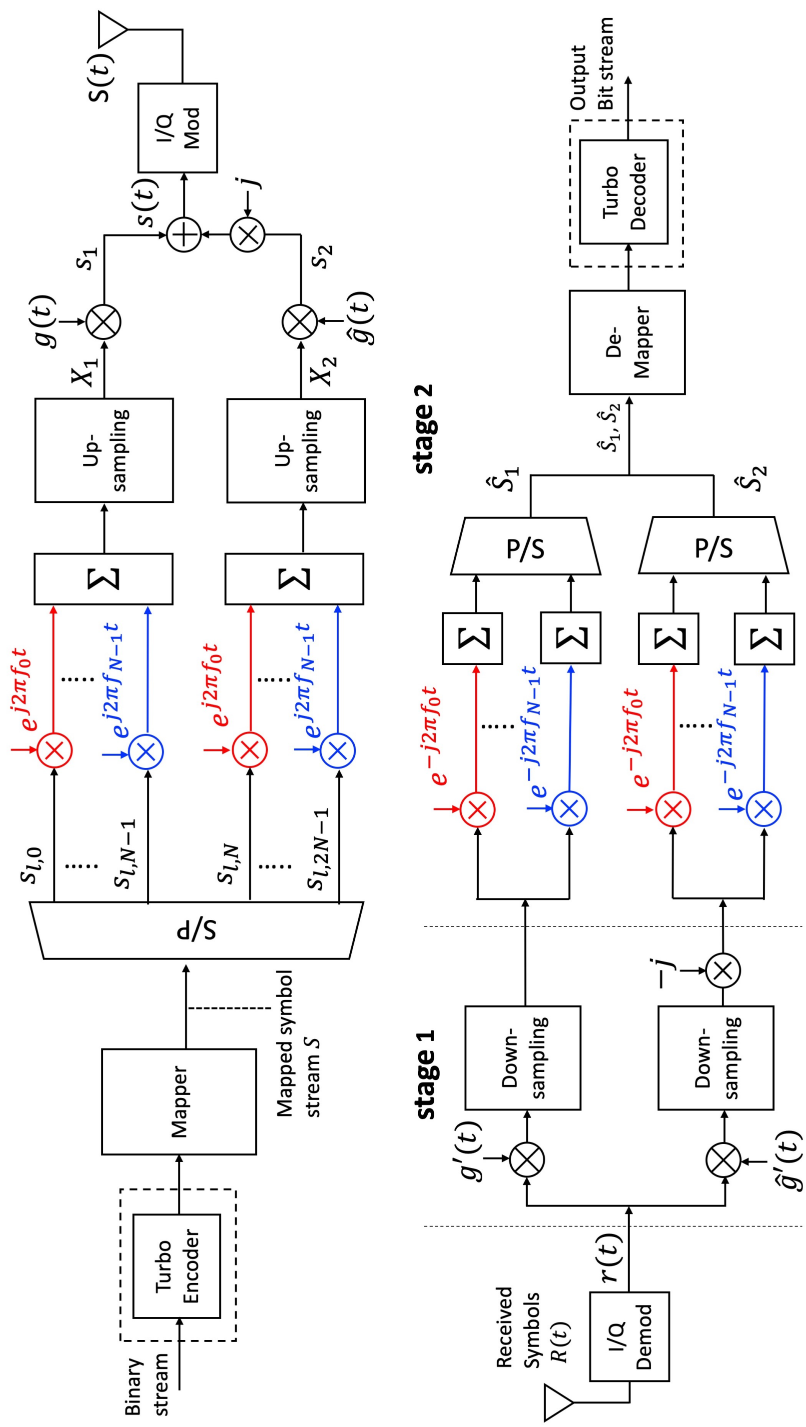
For simplicity, a bank of modulators is used here to generate the SEFDM carriers with different compression levels $(\alpha = 0.6, 0.8, 1)$ that are used for both of the separate and independent frames. Thus, $\mathbf{S}_1$ and $\mathbf{S}_2$ are modulated onto the same subcarrier frequencies and will be later separated in phase by the Hilbert pair. Before pulse shaping, both groups of signals are up-sampled via zero-padding between successive samples with the upsampling factor $q = 4$.

The first step towards simulating the multi-carrier transmitter is to generate the subcarriers. SEFDM can be seen as an extension of the OFDM system, the transmitter structures are similar. Since the SEFDM transmitter and receiver designs have been thoroughly studied and it's not the focus of the study in this thesis, for simplicity, reader are referred to [34] for details of the mathematical derivation and performance investigations of SEFDM. In this work, the SEFDM transmitter structure employs a bank of modulators running at the designed SEFDM subcarrier frequencies. The input symbol streams are modulated to the non-orthogonal subcarriers, which are generated by the parallel independent modulators.

As is explained in section 2.3.3, the frequency separation of the non-orthogonal subcarriers equals to $\frac{1}{T_{SEFDM}}$, where T_{SEFDM} denotes the signalling interval of each subcarrier. Hence, the subcarrier on the n^{th} frame can be defined as [10]

$$\Phi_{n,l} = e^{j2\pi\alpha(n-1)(l-1)/Q} \tag{3.4}$$

where $Q = \rho N$ is the number of the time samples for one SEFDM symbol, $n, l \in$



(a) Transmitter (b) Receiver

Figure 3.3: Simplified block diagram of the transceiver design of the proposed system: $g(t)$ and $\hat{g}(t)$ are from equation (6); $g'(t)$ and $\hat{g}'(t)$ are from equation (3.13); the portions in the same colour (red and blue) represent the same subcarriers

$\mathbb{Z}, n \in [0, N-1]$, α is the bandwidth compression factor. The most significant difference between the subcarriers of OFDM and SEFDM is the frequency separation Δf, which can be expressed as a function of α [10],

$$\Delta f = \frac{\alpha}{T_{SEFDM}}. \tag{3.5}$$

Similar to the DFT carrier matrix, the subcarrier matrix of SEFDM system also defines the N non-orthogonal subcarriers using the parameters and intermediate vectors above. Based on the equation (3.4), the $Q \times N$ carrier matrix can be written as [34]:

$$\Phi = \frac{1}{\sqrt{Q}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{j2\pi\alpha/Q} & e^{j4\pi\alpha/Q} & \dots & e^{j2\pi(N-1)\alpha/Q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j2\pi(Q-1)/Q} & e^{j4\pi(Q-1)/Q} & \dots & e^{j2\pi(N-1)(Q-1)/Q} \end{bmatrix} \tag{3.6}$$

where the factor $\frac{1}{\sqrt{Q}}$ is a normalization constant. Each row of the matrix corresponds to a subcarriers and each column of the matrix corresponds to a time stamp of the time vector [10].

As is mentioned in section 2.3.3, the discrete SEFDM signal is prepared by employing the periodic sampling. Based on equation (3.4), the discrete SEFDM can be represented as [10]:

$$X = \Phi \cdot S, \tag{3.7}$$

where Φ is the sampled subcarrier matrix, $S = [s_0, s_1, ..., s_{N-1}]^T$ is the input symbol vector. Consequently, the discrete-time representation for the k^{th} time sample of a single SEFDM symbol may be expressed as [119]:

$$X[k] = \frac{1}{\sqrt{Q}} \sum_{n=0}^{N-1} s_n \cdot e^{j2\pi nk\alpha/Q}, \tag{3.8}$$

where $\mathbf{s} = [s_0, s_1, ..., s_{N-1}]$ is the incoming M-ary QAM signal vector for N subcarriers and Q represents the total number of discrete-time samples in one SEFDM symbol. Given that the number of samples-per-symbol is denoted by ρ, there are $Q = \rho N$ ($\rho \geq 1, \rho \in \mathbb{Z}$) samples for the discrete time SEFDM scheme. The factor $1/\sqrt{Q}$ is for the purpose of normalisation.

3.3.2 Filtering using Hilbert Transform Pair

The Hilbert pair originates from the I/Q components of the analytic signal $f_+(t)$ given by [116]:

$$f_+(t) = f(t) + j \cdot \hat{f}(t), \tag{3.9}$$

wherein $f(t)$ is a real-valued function with continuous value t. Its Hilbert transform $\hat{f}(t)$ can be expressed using either an integration or a convolution as [116]:

$$\hat{f}(t) = \mathscr{H}\{f(t)\} = f(t) * \frac{1}{\pi t} = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(\tau)}{t-\tau} d\tau, \tag{3.10}$$

where $\mathscr{H}[\cdot]$ is the Hilbert transform operator. It is worth stating that the real function $f(t)$ and its Hilbert Transform $\hat{f}(t)$ are orthogonal:

$$\int_{-\infty}^{+\infty} f(t) \cdot \hat{f}(t) = 0. \tag{3.11}$$

To construct the orthogonal filter pair $g(t)$ and $\hat{g}(t)$, a given pulse shape $p(t)$ is multiplied with a Hilbert pair $f(t)$ and $\hat{f}(t)$. The process is shown in the equation below [120]:

$$g(t) = p(t)f(t), \quad \hat{g}(t) = p(t)\hat{f}(t). \tag{3.12}$$

Traditionally, a pair of sinusoidal carrier and co-sinusoidal carriers are used to form the Hilbert pair. Subsequently, the impulse response of the shaping pulse filters are the product of the pulse $p(t)$ and the carrier pair of frequency f_c, given by [117]:

$$g(t) = p(t)\cos(2\pi f_c t), \quad \hat{g}(t) = p(t)\sin(2\pi f_c t). \tag{3.13}$$

The SRRC pulse, a commonly-used root-Nyquist pulse, is adopted as the shape function $p(t)$ of the filter pair. It is assumed that the corresponding matched receiver

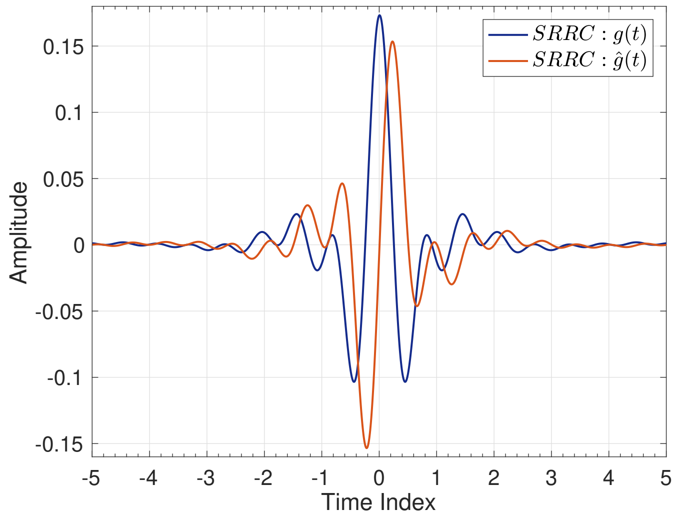


Figure 3.4: Time domain representation of the Hilbert pair formed by SRRC pulse ($\beta = 0.35$)

is used in the system. The SRRC pulse is expressed by:

$$p(t) = \frac{2\beta[\cos(\frac{(1+\beta)\pi t}{T_s}) + \sin(\frac{(1-\beta)\pi t}{T_s})(\frac{4\beta t}{T_s})^{-1}]}{\pi\sqrt{T_s}[1-(\frac{4\beta t}{T_s})^2]}, \tag{3.14}$$

where the roll-off factor $\beta \in [0,1]$ controls the excess bandwidth, T_s represents the symbol period. The filter bandwidth is equal to $B = (1+\beta)/2T_s$. Fig. 3.4 depicts the time-domain representation of the aforementioned filter pair generated by SRRC pulse with $\beta = 0.35$.

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be the up-sampled SEFDM signal on the two independent frames, and hence the process of pulse shaping can be expressed as:

$$s_1(t) = g(t) * \mathbf{X}_1, \quad s_2(t) = \hat{g}(t) * \mathbf{X}_2, \tag{3.15}$$

where $g(t)$ and $\hat{g}(t)$ are a normalised Hilbert pulse pair yielded from equation (3.13). The filters have 10 symbols length span since ideal filters of infinite length are impractical. The outputs of the two orthogonal filters are added up, after multiplying the second stream by j to give $s(t)$, which can be expressed by:

$$s(t) = s_1(t) + j \cdot s_2(t). \tag{3.16}$$

The achieved signal is then converted to RF signal by I/Q modulation and passed through the wireless channel, which, for this work is assumed to be a simple AWGN channel.

Fig. 3.3 (b) shows the block diagram of the receiver, which functionally incorporates two stages; the first separates two symbol streams and demodulates the signal in each. The second stage employs time-reversed matched filtering (MF) without the use of advanced detection methods. The matched filtering pair at the first stage can be obtained by flipping the Hilbert pulse pair in the time domain, as given by [117]:

$$g'(t) = g(-t), \quad \hat{g}'(t) = \hat{g}(-t). \tag{3.17}$$

The matched pair $g'(t)$ and $\hat{g}'(t)$ separate the received signals into two sub-streams followed by the downsampling and decimation process. On the second frame $-j$ is introduced with the negative sign to recover the signal as in the lower arm of transmitter. Once two groups of symbols are properly separated, matched filtering along with hard decision work in the manner as in conventional SEFDM or OFDM systems. The matched filtering at the second stage is equivalent to the conjugate complex of the subcarriers matrix Φ^H, where $[\cdot]^H$ is the Hermitian operator. The demodulation process can be expressed as [119]:

$$\hat{S} = \Phi^H R = \Phi^H (X + W) = \Phi^H \Phi S + \Phi^H W, \tag{3.18}$$

where $\hat{S}$ represents the demodulated signal, Φ is the Q-by-N modulator matrix, Φ^H is an N-by-Q matrix, the input X is defined by equation(3.8) and W denotes the noise vector. The correlation matrix C is constructed as [119]:

$$\Lambda = \Phi^H \Phi, \tag{3.19}$$

where ideally the correlation denotes an N-by-N unitary matrix, and consequently the received signals can be recovered if the introduced noise-associated term can be removed or suppressed. As is shown in the equation, the AWGN noise is expanded by the complex conjugate of the subcarrier matrix. Since the Hilbert pair is orthogonal, no degradation is introduced to the correlation matrix C, the proposed system is expected to have an identical BER performance when compared to the conventional SEFDM. The recovered complex symbol streams, termed as $\hat{S}_1$ and $\hat{S}_2$ are input into the demapper block serially.

In the simulation, based on the investigation in [121], 5 iterations are optimal to counteract the introduced ICI of SEFDM. A soft de-mapper is then used with the log likelihood ratio (LLR) algorithm so that soft bit and extrinsic information (from the previous iteration) are utilised to feed the turbo decoder.

3.3.3 Performance Analysis of Numerical Results

The proposed system performance is evaluated in terms of the BER and the PAPR for different α values. Tests are carried out on both conventional SEFDM and the proposed system for the purpose of comparison. The mathematical models are designed and generated in MATLAB.

Fig. 3.5 compares the normalised spectrum of the conventional SEFDM signal and the proposed transmitted signal when the compression factor $\alpha = 0.8$ and 1 (i.e., OFDM). It is known that the inherent non-orthogonality of SEFDM signal, leading to the compression of the frequency spacing between adjacent subcarriers, results in the spectral efficiency gain when compared to OFDM signal. The spectral efficiency generally describes the maximum data rate R_b that can be transmitted over a particular bandwidth B and hence can be measured by their ratio. Therefore, the smaller α is, the higher the spectral efficiency that can be achieved due to the decreased bandwidth consumption. Fig. 3.5 shows that the designed signal filtered by the Hilbert pair occupies the same bandwidth at the same carrier frequency as compared to the SEFDM. In conventional SEFDM, the frequency spectrum carries only one symbol stream. However, with the same spectrum utilisation the frequency spectrum of the proposed signal carries two independent 4-QAM symbol streams,

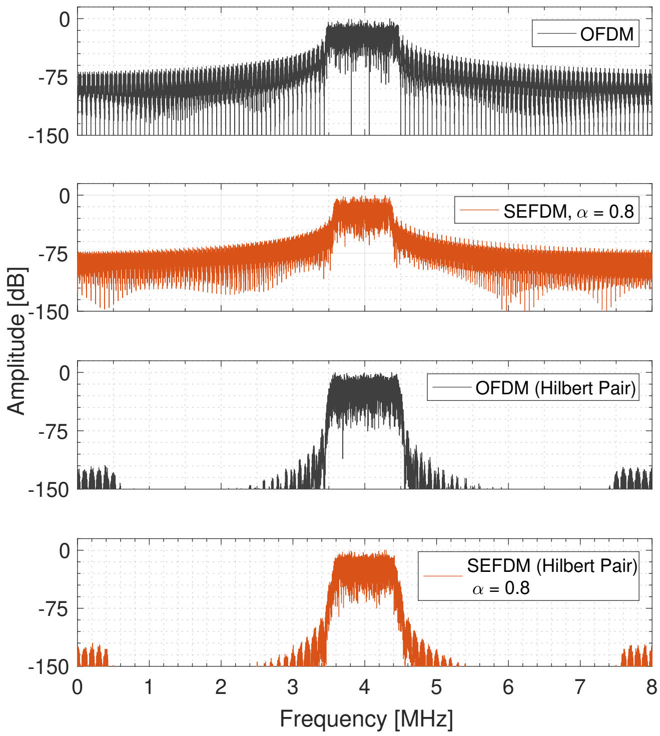


Figure 3.5: Spectrum Comparison: OFDM; SEFDM (α = 0.8); OFDM using Hilbert pulse pair; SEFDM using Hilbert pulse pair ($\alpha = 0.8$). $N = 16$ for all cases.

leading to a doubled data rate. Consequently, the spectral efficiency of the proposed signal can be expressed as $\frac{2\log_2 \mathrm{MR_s}}{\alpha \mathrm{B}}$ $bits/s/Hz$, which is twice the spectral efficiency of the SEFDM (i.e. $\frac{\log_2 \mathrm{MR_s}}{\alpha \mathrm{B}}$ $bits/s/Hz$) when using the same α. This leads to the conclusion that by using the Hilbert pulse pair, the spectral efficiency is doubled.

The BER performance of the proposed system using Hilbert pulse pair is investigated in terms of different choices of the compression factor α. Given that the SEFDM system is ill-conditioned when $\alpha \leq 0.8$ when only MF is used, the system will not be expected to lead to good BER results. To improve performance, turbo coding is added and at the receiver the decoder employs 5 iterations based on the results obtained in [121]. This is to confirm that coding is effective in enhancing the BER of the proposed system.

Fig. 3.6 demonstrates the BER performance of the signal generated from the

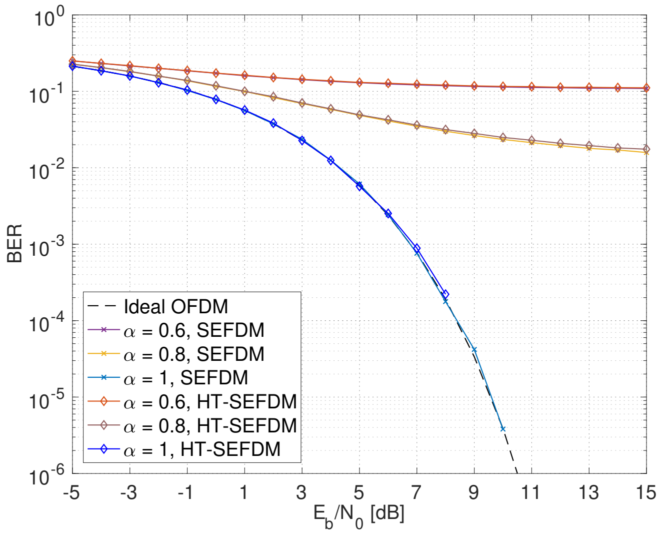


Figure 3.6: BER performance comparison of SEFDM signals with and without pulse-shaping using Hilbert SRRC pair ($\alpha = 0.6, 0.8, 1, \beta = 0.35$)

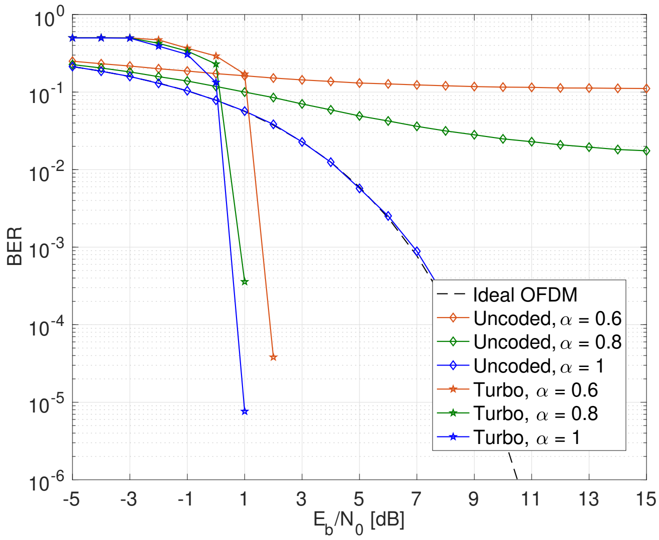


Figure 3.7: BER performance of SEFDM signals pulse-shaped by Hilbert SRRC pair with and without turbo Coding ($\alpha = 0.6, 0.8, 1, \beta = 0.35$)

proposed design given in Fig. 3.3 versus E_b/N_0. It is evident that the proposed system achieves identical BER performance compared to the conventional OFDM ($\alpha = 1$) and SEFDM ($\alpha = 0.8, 0.6$). With the reduction of the compression factor α, the error performance degrades rapidly, which is in accordance with the theoretical results. This leads to the conclusion that the use of Hilbert pulse pair as shaping pulses doubles the spectral efficiency of SEFDM without incurring error penalties.

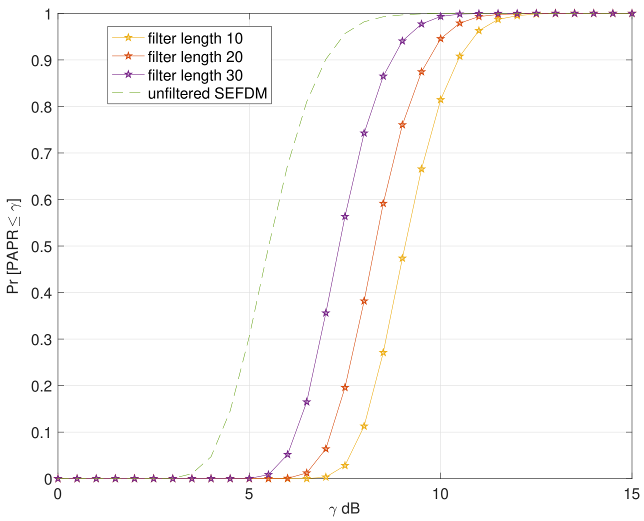


Figure 3.8: PAPR for the SEFDM system employing Hilbert SRRC pulse pair with varying values of filter length ($\alpha = 0.8, N = 16, \beta = 0.35$)

Fig. 3.7 depicts the numerical results of the precoded system with the structure in Fig. 3.3 , showing the substantial BER reduction when compared to the uncoded signal.

The PAPR is considered to estimate the power efficiency of the proposed system. The results shown in Fig. 3.8 are for the uncoded signal generated for the same specifications with varying filter length of the Hilbert pulse pair. It can be observed from Fig. 3.8 that the increasing number of filter spans has significant impact on the PAPR performance of such system. However, the observations also show that the designed signal is suspective to have lower power efficiency even for shaping pulse pairs of longer length (30 spans). This is due to the upsampling process where zero-insertion is used and subsequently power fluctuation is introduced to discrete-time signal, expanding the gap between the average and the peak power.

In conclusion, this section proposed a new signal processing method using Hilbert transform pair to enhance the spectral efficiency of SEFDM systems. A new transceiver structure is designed to generate the new signal format, employing Hilbert pulse pair as shaping pulses. The new system has the major advantage in doubling the spectral efficiency by transmitting two different complex symbols simultaneously and within the same occupied spectrum. Importantly, such doubling of spectral efficiency is achieved without degrading the BER performance. Moreover, the use of turbo coding shows its appropriateness in ameliorating the BER degradation due to high levels of interference resulting from using SEFDM signal format.

3.4 Fast-OFDM with Hilbert Filter Pair for NB-IoT

The IoT and 5G wireless technologies are evolving towards the next-generation IoT, assembling technologies such as edge computing, artificial intelligence (AI) and distributed ledger technologies (DLT) and integrating augmented and virtual reality (AR/VR) based services and applications [122]. To support demanding high bit rate applications and ubiquitous connectivity, from the physical layer aspect, new techniques with low complexity, high power efficiency, and increased capacity are necessarily needed.

In five years from now, by 2027, the number of IoT connections is forecast to exceed 30 billion worldwide, and with over half of these are envisioned to be cellular IoT connections [123]. One of the cellular IoT connections standardised in 4G-LTE [59], i.e., NB-IoT, continues to form part of the 5G standards and therefore coexist with the other 5G NR components in the same networks. This indicates a high share of NB-IoT and 5G devices and hence, sets the demand for technologies that support the LPWAN communications in the 5G context. In the case of LPWAN, narrowband signalling is adopted to have higher robustness to the severe pass loss in large-scale transmission. More specifically, NB-IoT employs orthogonal frequency division multiple access (OFDMA) for the downlink transmission and single-carrier frequency division multiple access (SC-FDMA) for its uplink, where a narrowband of 180 kHz is occupied by 12 modulated subcarriers. Due to the ever increasing demand for higher capacity and energy efficiency for ultra-massive connectivity and traffic, there have been much research and several proposals on evolving the physical layer design of IoT, by replacing the OFDM signalling with more spectrally efficient techniques.

The next generation IoT, which requires high data rate, is being considered in

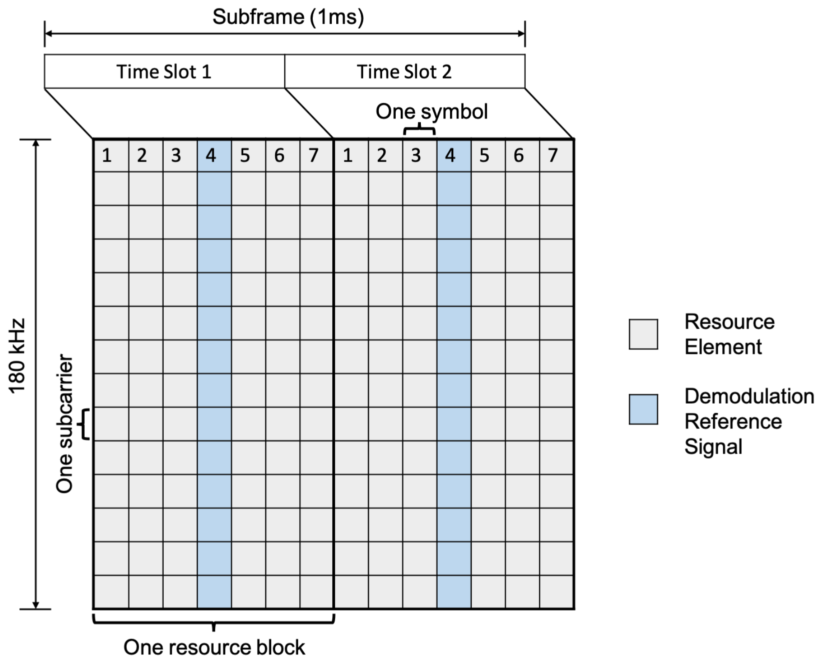


Figure 3.9: Uplink resource block definition for NB-IoT. There are 12 subcarriers for 180 kHz bandwidth and 7 symbols bundled into one time slot

beyond 5G systems and the envisioned 6G networks. As the deployment density of IoT devices grows rapidly with the exponentially increasing demands of communications and traffic, from heterogeneous networks and services, the requirements for the key performance indicators such as the spectral efficiency, energy efficiency, reliability and latency are expected to be challenges in the next generation IoT era. In the 4G standards, NB-IoT has a channel bandwidth of 200 kHz while data occupies 180 kHz, which is equivalent to one LTE resource block. As discussed previously, OFDMA is employed for downlink while SC-FDMA is used for the uplink transmission mainly due to its relatively low PAPR. $\pi/2$-BPSK and quadrature phase shift keying (QPSK) are mainly used as modulation schemes for NB-IoT. Fig. 3.9 provides the uplink resource block map for NB-IoT It is worth noting that, the demodulation reference signals (DM-RS), which is the fourth symbol in each time slot, are intended for channel estimation. In this study, the channel is assumed to be flat and hence, no symbols are adopted as pilot tones in each subframe. The investigations of the new signalling format that follows the NB-IoT specifications are detailed in the next sections.

As introduced in Chapter 2, FOFDM is a variation of OFDM achieving a doubled spectral efficiency [45]. Due to its advantage of doubling the data rate within the bandwidth-limited spectrum while maintaining the transmission reliability, FOFDM has increasingly attracted researchers' interest, consequently leading to a wide application in wireless [49] and optical communications [74], particularly arising rapidly in VLCin recent years [68] for both SISO and MIMO transmissions. FOFDM requires one-dimensional modulation, which constrains its improvement of spectral efficiency relative to two dimensional schemes, nevertheless, makes it a good fit for IoT standardised systems. In particular, the feasibility of FOFDM for NB-IoT applications has been theoretically investigated and experimentally verified in [124], where FOFDM is shown to an enhanced capacity by 200%. Moreover, for the use cases of NB-IoT, FOFDM was combined with index modulation (IM) [94] and filtering technique with HT pair [29], showing the advantages of enhanced power efficiency and quadrupled data rate relative to conventional IoT systems, respectively.

Inspired by the aforementioned work and considering the main demands of next generation IoT integrated with 5G scenarios, The second model, which is both spectrally and energy-efficient, is designed in the context of NB-IoT mobile system. The data rate can be quadrupled by combining two orthogonal techniques; the frequency orthogonal FOFDM scheme coupled with the time orthogonal Hilbert transform pair. The application of Hilbert transform filter pair to FOFDM is first tested with BPSK. The design is simulated in presence of AWGN and shown to offer enhanced data rate when requiring the same power consumption. However, using BPSK limits the maximum achievable spectral efficiency of the system. By demonstrating the advantages of using HT-FOFDM in NB-IoT scenarios, a new design is proposed by utilising the singular dimensional constellation of PAM modulation to maintain the orthogonality in FOFDM signal. Performance comparison of the proposed signals is investigated relative to the QAM modulated SC-FDMA signal under fixed spectral efficiencies.

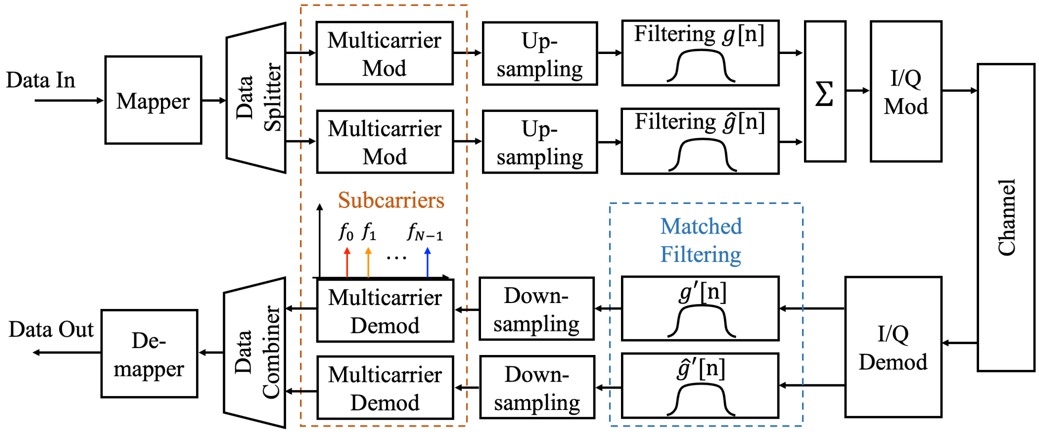


Figure 3.10: Simplified block diagram of HT-FOFDM transceiver; subcarriers occupy the same frequency range on two separate paths

3.4.1 Transceiver Architecture

To meet the requirements for the NB-IoT scenario, 12 subcarriers are used in the proposed system with the overall bandwidth of 180 kHz [125]. The system modelling focuses on the filtering structure design where Hilbert pair and its matched filter pair are used on both transmission and reception sides. Numerical expressions are given to clarify the processing procedures. The AWGN channel is considered in this case.

Fig. 3.10 illustrates the block diagram of the proposed FOFDM system using Hilbert pulse pair. The pseudo-randomly generated bit stream is input into the mapper block, where the binary bits are transferred to bipolar BPSK symbols for the first study. The data splitter splits the mapped symbol stream S equally, therefore yielding two sub-streams X_1 and X_2. On two independent paths, the symbols are processed through identical blocks. For each path, the serial symbols are paralleled. Then the N points IFFT is used to modulate these symbols to the 12 subcarriers.

Recalling the continuous-time FOFDM signal introduced in Chapter 2, the signal expression is given by:

$$x(t) = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{+\infty} \sum_{n=0}^{N-1} X_{n,k} e^{j2\pi nt/2T}, \tag{3.20}$$

where T represents the period of an OFDM symbol, the subcarrier spacing of

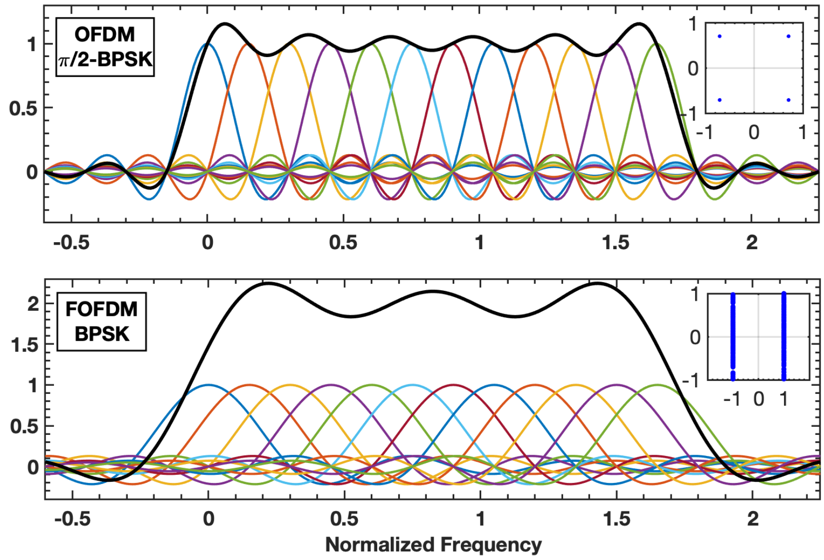


Figure 3.11: OFDM and FOFDM subcarriers representations ($N = 12$) with identical spectrum occupancy while FOFDM provides twice the data rate of OFDM

FOFDM is equal to $1/2T$ in equation (3.20), N is the total number of the subcarriers, $n \in [0, N-1]$ denotes the frequency index of the subcarrier and $X_{n,k}$ represents the symbol transmitted on the n^{th} subcarrier and the k^{th} signalling interval. Fig. 3.11 provides a simplified subcarrier representation of the FOFDM symbol in comparison to that of the OFDM symbol. The bottom figure in the same illustration shows the constellation diagram of the received FOFDM symbols when singular modulation schemes, hereby the BPSK scheme and then PAM-M for the second study, are applied to. It can be seen that no distortion occurs to the BPSK symbol as the points are located exactly on the underlying line at the in-phase value of 1 and -1.

Once the FOFDM signals are generated, the next block performs up-sampling by inserting $q-1$ zeros between the adjacent symbols, wherein $q = 4$ is the up-sampling factor. The up-sampled symbols are then filtered by the Hilbert filter pair, giving S_1 and S_2. The summation of the filtered signals on two paths output the transmitted signal $s(t)$. Herein, the Hilbert transform pair, namely the $g(t)$ and $\hat{g}(t)$, are generated also by a SRRC pulse $p(t)$ as the mathematical expression of its time response given in equation (3.14). The time domain representation of the HT filter pair used in this work is the same as shown in Fig. 3.4, with the filter length of $L_s = 10$ and the roll-off factor $\beta = 0.35$, which are chosen based on the investigations in [117].

Fig. 3.10 shows the receiver's architecture, where matched filters are employed to demodulate the received signals, followed by a hard decision detector. The matched filtering procedure consists of two stages: separate the received symbol to two streams that correspond to the output symbols on the independent transmitter paths; demodulate each symbol streams using FFT operation.

The first stage involves the correlation of the received signal with the matched filter pair, which is the time-reversed form of the Hilbert pulse pair employed at the transmitter. Note that the matched filter pair are orthogonal[2] and hence enabling the recovery of the Hilbert superposition at the transmission side. Moving to the second stage, the separated signals are then down-sampled and input into the FFT block, where demodulation take place in a more efficient way for implementation. After the hard decision, bipolar symbol streams on both paths are achieved and subsequently the serial symbols are de-mapped to give the recovered bit stream.

3.4.2 Performance Analysis of Numerical Results

Theoretically, based on the definition as mentioned in section 2.3.2, FOFDM can achieve higher data rate without error penalties when the transmitted signal occupies the same bandwidth as OFDM. The Hilbert pulse pair, attributed to its orthogonality feature, allows the superposition of two independent symbols to be separated without interference added. The results presented in Fig. 3.13 show the BER performance of different systems with respect to the varying value of E_b/N_0 regimes. Wherein, the BER curve of the proposed system is generated by employing the transceiver architecture in Fig. 3.10. A significant observation from the figure is that the FOFDM signal using the Hilbert filter pair scheme has the same BER performance as the typical FOFDM signal as well as the BPSK OFDM signal. Therefore, a compelling advantage of the FOFDM system using Hilbert filter pair can be seen that better BER performance is achieved when transmitting the same data rate.

The effective data rate is defined as the achievable bit rate in condition of a

[2] The proof of this orthogonality of matched filter pair is detailed in Appendix A

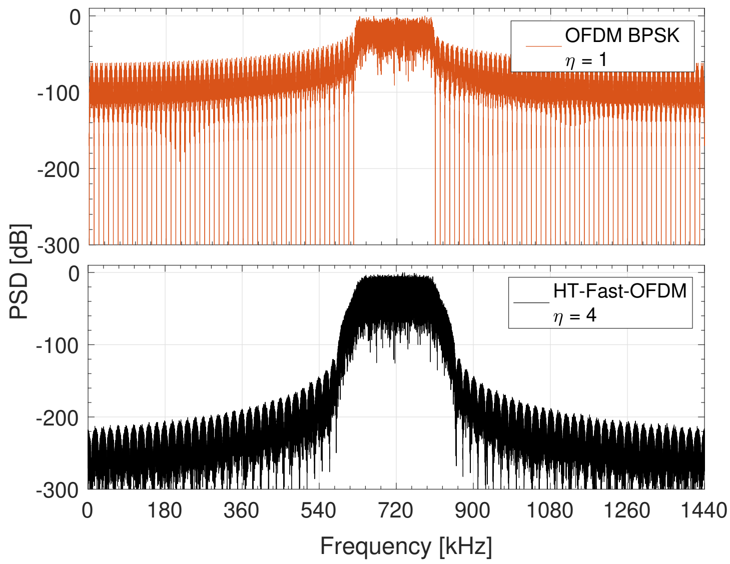


Figure 3.12: Spectrum comparison: OFDM signal for 16-QAM modulation (upper); Fast OFDM signal using Hilbert SRRC pair ($N = 12, \beta = 0.3$)

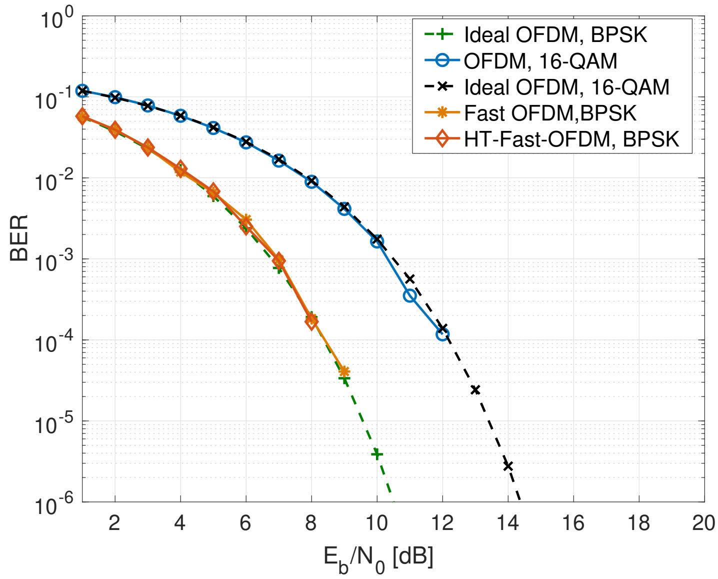


Figure 3.13: BER performance for HT-FOFDM and 16-QAM OFDM system

fixed bandwidth consumption with a certain value of BER reserved. The numerical expression of effective data rate is given by [49]:

$$R_e = (1 - BER) \times R_s \times \log_2 O \times (N_d/N), \tag{3.21}$$

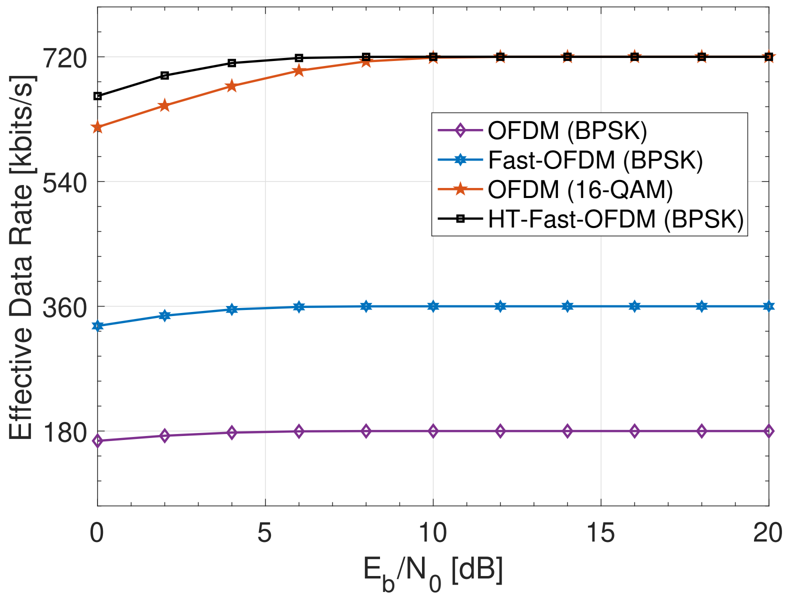


Figure 3.14: Effective Data Rate Comparison: typical OFDM system for 16QAM and BPSK modulation, FOFDM BPSK system and the proposed system.

Table 3.1: A comparison of attained performance for various systems

System	Modulation Scheme	Effective Data Rate [kbits/s]	Spectral Efficiency [bits/s/Hz]
OFDM	BPSK	180	1
FOFDM	BPSK	360	2
OFDM	16-QAM	720	4
HT-FOFDM	BPSK	720	4

where BER represents the metric bit error rate. This value is used to evaluate the data rate of the aforementioned systems. Additionally, in the equation, R_s is the symbol rate, O denotes the constellation cardinality, N_d is the number of data subcarriers and N is the total number of subcarriers. Since no cyclic prefix nor guard band is utilised in the proposed system, the ratio of N_d and N equals one. Table 3.1 compares the effective data rate of the proposed system with other three systems. It is assumed that in the simulations all the other specifications for various systems are identical.

Fig. 3.14 depicts the measured effective data rate of the systems in relation to the varying value of E_b/N_0. As expected, the proposed system transmits four times the data rate of the BPSK OFDM system as well as twice that of the FOFDM system. Since the error penalty is considered, the 16-QAM OFDM system theoretically offers the same data rate as the proposed system though, achieves lower effective data rate as it approaches but cannot reach 720 kbits/s as the value of E_b/N_0 increases.

In the second studies, to further meet the requirements for NB-IoT scenario and attain various spectral efficiencies, the 1D modulation BPSK scheme is extended to PAM-M, which allows higher spectral efficiency to be achieved. For comparison, SC-FDMA is used with QAM schemes considered. To evaluate the performance of the proposed PAM modulated HT-FOFDM signal in terms of their spectral efficiency and BER performance, Monte Carlo simulations are carried out with AWGN taken into account to obtain the same spectral efficiencies of HT-FOFDM and to show the advantages of HT-FOFDM together with PAM-M modulations, three different schemes, i.e., PAM-2, -4, and -8, are studied to give the spectral efficiency η of 4, 8 and 12 bits/s/Hz.

The HT pair is generated by the basis function with a length of $L = 10$ and roll-off factor $\beta = 0.35$. The parameters are chosen based on investigations in [117] as well as to maintain consistency with the work in [29]. The designed system employs Better-than-Nyquist pulse [126], one of the Nyquist pulse family that satisfies the first Nyquist criterion for zero ISI, due to its low OOB. The continuous form of the BTN pulse can be given by [126]:

$$p(t) = \frac{\sin[\frac{\pi \mathrm{t}}{\mathrm{T_s}}][\frac{2\pi\beta \mathrm{t}}{\mathrm{T_s}\ln(2)}\sin(\frac{\pi\beta \mathrm{t}}{\mathrm{T_s}} + 2\cos(\frac{\pi\beta \mathrm{t}}{\mathrm{T_s}}) - 1)]}{(\frac{\pi t}{\mathrm{T_s}})[(\frac{\pi\beta t}{\mathrm{T_s}\ln(2)})^2 + 1]}, \tag{3.22}$$

where $\mathrm{T_s}$ is the pulse duration. Fig. 3.15 illustrates the BTN pulse and the corresponding HT pair. After filtering, the two streams of output symbols are summed up for the preparation for I/Q modulation and then passed through the channel. Due to the narrowband property of signalling for IoT scenarios, in this work, the channel is assumed to be static and the frequency response is almost flat. Therefore, the AWGN channel is used as the channel model.

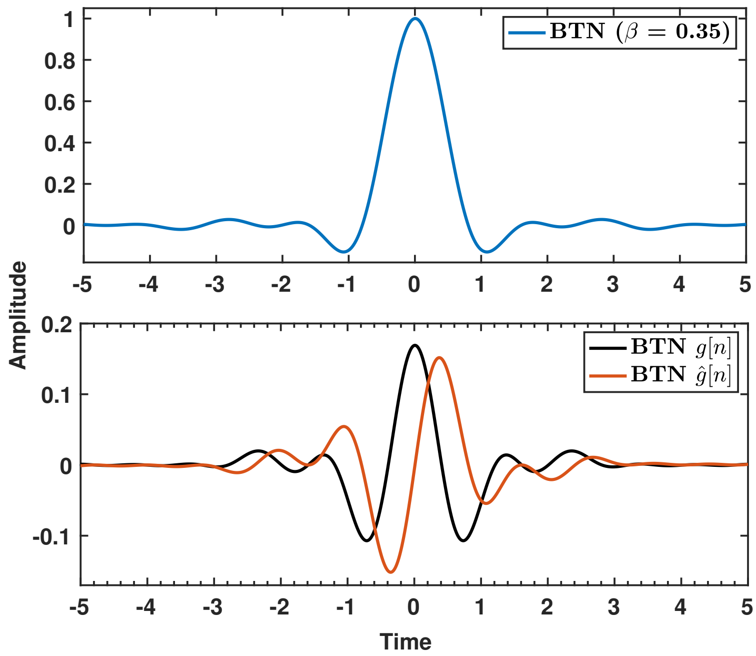


Figure 3.15: Time domain BTN pulse and Hilbert filters generated by BTN pulse

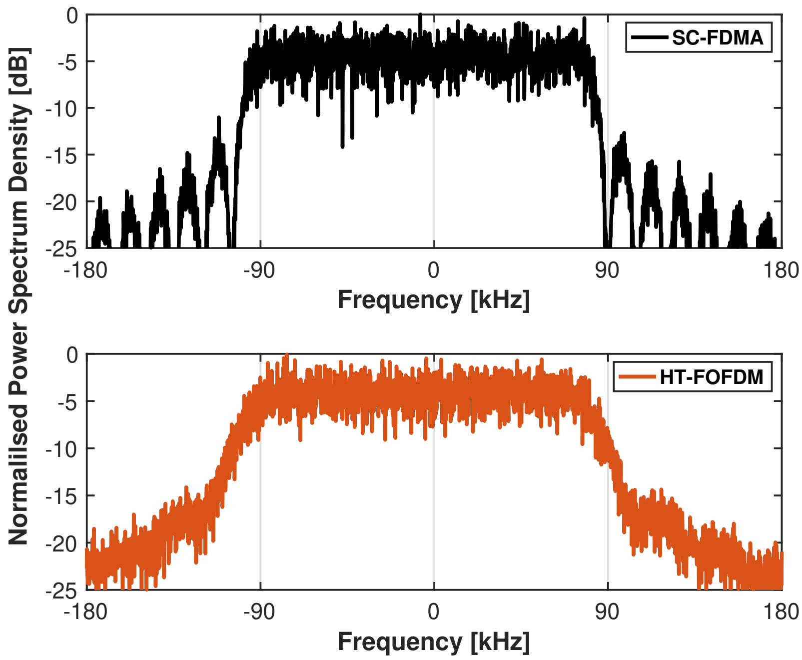
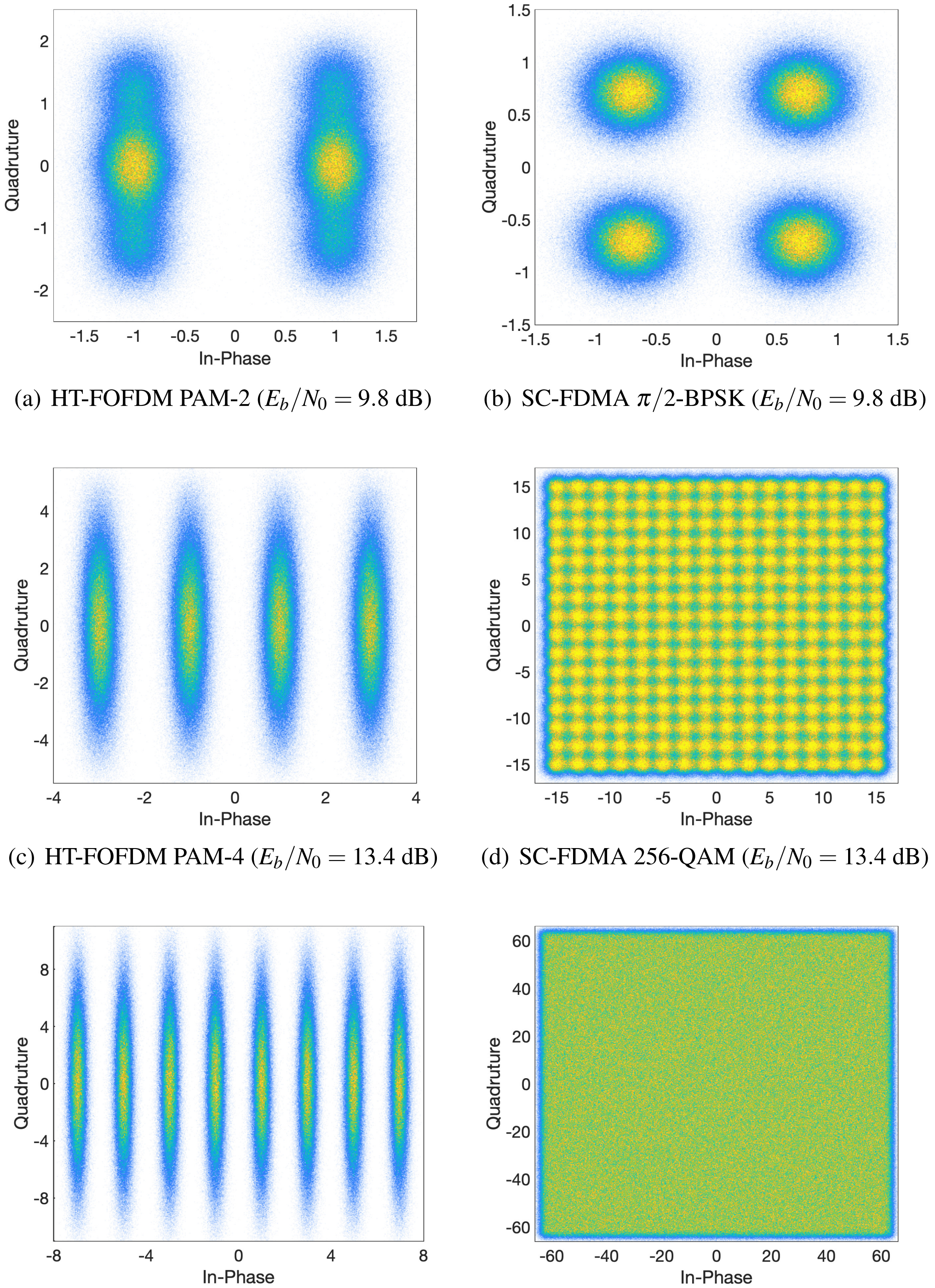


Figure 3.16: Spectrum for HT-FOFDM signal occupying the same bandwidth while transmitting at multiple times data rate of SC-FDMA

The power spectra of SC-FDMA and HT-FOFDM signals are shown in Fig. 3.16, where 12 subcarriers are modulated by $\pi/2$-BPSK for SC-FDMA and PAM-2



(a) HT-FOFDM PAM-2 ($E_b/N_0 = 9.8$ dB)

(b) SC-FDMA $\pi/2$-BPSK ($E_b/N_0 = 9.8$ dB)

(c) HT-FOFDM PAM-4 ($E_b/N_0 = 13.4$ dB)

(d) SC-FDMA 256-QAM ($E_b/N_0 = 13.4$ dB)

(e) HT-FOFDM PAM-8 ($E_b/N_0 = 17.9$ dB)

(f) SC-FDMA 4096-QAM ($E_b/N_0 = 17.9$ dB)

Figure 3.17: Constellation diagrams of received noisy HT-FOFDM and SC-FDMA signals at varying E_b/N_0

for HT-FOFDM ($M = 2$). It is worth noting that both signals occupy the same bandwidth of 180 kHz while the HT-FOFDM achieves quadrupled spectral efficiency

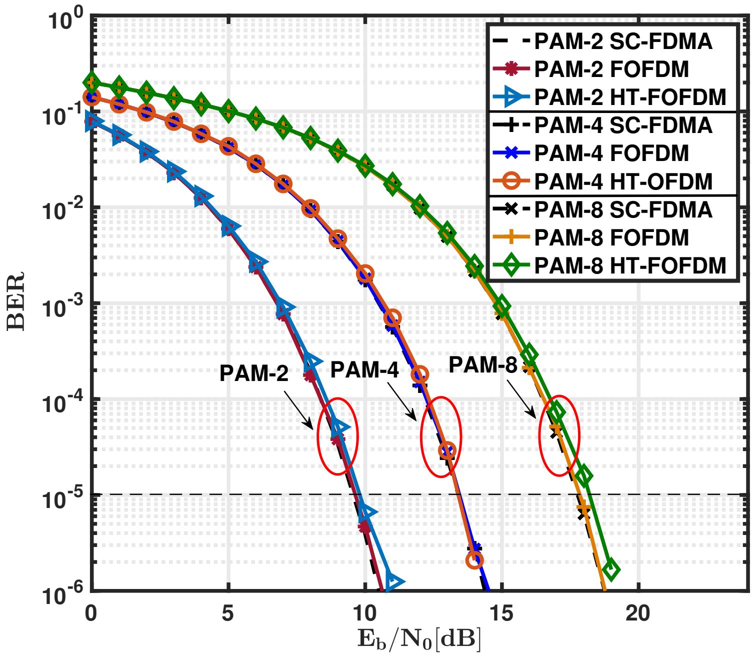


Figure 3.18: BER performance comparison for SC-FDMA, FOFDM and HT-FOFDM with varying order of PAM modulations

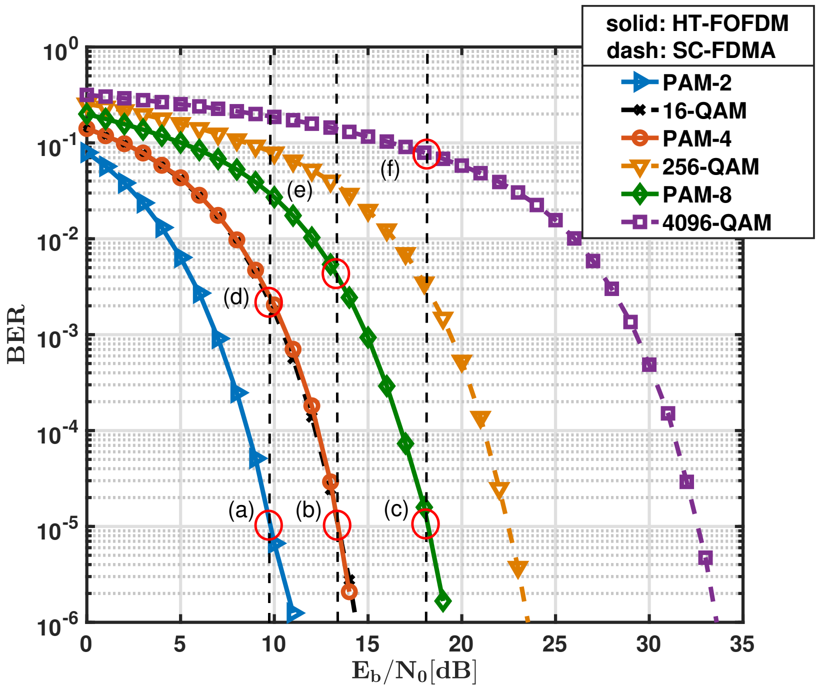


Figure 3.19: BER performance comparison for PAM-M HT-FOFDM and SC-FDMA with high-order QAM modulations that achieve the same spectral efficiency

compared to SC-FDMA.

Two comparisons of BER results are presented in Fig. 3.19 and Fig. 3.18 to show the power advantages of HT-FOFDM over the conventional signalling for IoT uplink transmission. Fig. 3.19 compares the BER performance of proposed HT-FOFDM signalling based on the system given in Fig. 3.10. The energy per bit to noise ratio $\mathrm{E_b/N_0}$ in dB is used for the signal-to-noise ratio (SNR) for performance assessment in the study. The comparison is held for three different levels of PAM modulations ($M = 2,4,8$) and three different schemes, i.e., HT-FOFDM, FOFDM and SC-FDMA. It can be seen that the combined use of waveform compressed FOFDM and HT filter pair improves the data rate without degrading the BER performance. Fig. 3.18 depicts that the proposed HT-FOFDM with 1D modulation signalling offers significant power advantages compared to the SC-FDMA with QAM modulations under the same condition of band-limitation. To show the power reduction gap in between the BER curves of HT-FOFDM and SC-FDMA, the $\mathrm{E_b/N_0}$ at BER of 10^{-5} is assessed. When achieving the spectral efficiencies $\eta = 4,8$ and 12 bits/s/Hz, PAM-M HT-FOFDM has power advantages of 3.5 dB, 9.1 dB and 14.7 dB, respectively, with respect to SC-FDMA with different QAM schemes ($M = 16,256,4096$).

Moreover, to clearly show the performance advantages of the proposed signalling scheme, the constellation diagrams of the received contaminated HT-FOFDM signals modulated by various PAM schemes and SC-FDMA with the aforementioned QAM schemes are studied. For clarity, the arrangement is held by comparing the constellations between the two schemes of the same spectral efficiencies at certain E_b/N_0 values where the performance of HT-FOFDM reaches the BER of 10^{-5}. Specifically, simulations are conducted at $E_b/N_0 = 9.8$ dB for (a) with (d), 13.4 dB for (b) with (e) and 17.9 dB for (c) with (f) as depicted in Fig. 3.18 and the associated BER versus E_b/N_0 performance of these signals are marked in Fig. 3.19. The numerical results show the more clear constellations for HT-FOFDM signals with all considered PAM schemes relative to those of SC-FDMA, as expected, indicating that better performance can be achieved for the proposed signals for the same

E_b/N_0. The main conclusion can be drawn from Fig. 3.17 and Fig. 3.19 that HT-FOFDM with 1D PAM modulation has enhanced error performance compared to SC-FDMA with the same spectral efficiency or requires less power when achieving the same BER.

3.5 Discussions and Conclusions

This chapter presents a new signalling design to increase the spectral efficiency through the employment of Hilbert transform filter pair. The use of the Hilbert filter pair introduces a layer of orthogonality, which has been proved mathematically, such that two independent data streams can be transmitted simultaneously by relying on the orthogonal components. Based on such property, novel transceiver design is proposed to implement the filtering by Hilbert transform pair to multi-carrier signals.

For proof of concept, two studies are conducted, considering different signal formats for use cases in cellular communication systems and IoT to investigate the performance of the Hilbert filtering technique as well as the new signalling method. In the first study, the Hilbert transform pair is employed as pulse shaping filters, applied to OFDM and SEFDM signals. Different compression levels are considered for the SEFDM signals to see how the Hilbert filter pair affect the performance at various ICI levels. Matched filter receiver is utilised first, for the uncoded signals, followed by the employment of turbo channel coding with its iterative decoder to ameliorate the deteriorative effects of ICI in SEFDM signals by subtracting the interference gradually. BER performance of the modelled system is obtained via computer simulations. Comparison studies have been carried out on evaluating the error performance of OFDM and SEFDM signal that occupy the same bandwidth with and without the use of Hilbert filter pair. Numerical results have demonstrated that the proposed signal achieves doubled spectral efficiency while preserving the same BER performance. For the SEFDM that suffers from high ICI, the coded signals are shown to have substantial BER reduction, which is in accordance with the theoretical analysis. Nevertheless, the proposed signal is suspective to have higher PAPR due to the upsampling process, where power fluctuation is introduced due to the zero-insertion, posing the trade-offs between the advantageous spectral efficiency gain and the required complexity increase for reducing the PAPR.

The second study aims at testing the Hilbert transform filtering technique in a specific scenario, NB-IoT, which plays an important role in 5G communications. For such scenario, FOFDM with 1D modulation schemes, first BPSK and then PAM-M with more achievable spectral efficiencies, are used in the same transceiver design-based system model. A fair comparison is held by transmitting signals at the same data rate within fixed bandwidth. A compelling advantage of FOFDM with Hilbert transform filter pair is observed relative to OFDM signals. The attractive performance is delivered by the proposed signal, which achieves better error performance when offering the same spectral efficiency or attaining significant higher spectral efficiency without distorting the BER. Same observations occur to HT-FOFDM with PAM-M schemes,showing power advantages over the conventional SC-FDMA QAM for IoT uplink transmission and effectively maintaining reliable transmission up to 2.16 Mbits/s within the narrowband of 180 kHz.

Overall, the work done in this chapter is for the purpose 'proof of concept' for the proposed filtering technique of Hilbert transform pair, the newly designed waveform format employing thus technique and the developed transceiver design. The efficacy of the new signalling method, in enhancing the spectral efficiency while maintaining the performance, has been examined via two scenario-based simulation studies, where AWGN is considered. Although there are compromises in terms of additional receiver complexity by the filtering process and the potential need for PAPR reduction mechanisms, the results of this work indicate new signalling and system designs to improve the spectral and energy efficiency for various scenarios for the next generation communication systems.

Chapter 4

Constellation Shaping for Spectrally Efficient Signalling

4.1 Introduction

Driven by the common goal of approaching the Shannon limit in communication systems as the non-orthogonal signalling presented in the last chapter, non-uniform signalling techniques have gained significant attention in academia and industry in the last two decades. Typical non-uniform approaches are based on constellation shaping as means of capacity-achieving methods by optimising the constellation design. Two existing techniques, geometric shaping [17] and probabilistic shaping [16], have been shown to offer significant shaping gains, thus requiring reduced SNR to reach a particular data rate. In geometric shaping, irregular constellation shapes are used, offering some energy gains but at the expense of modulation and detection complexity. Probabilistic shaping, however, where the occurrence probabilities of different constellation symbols are not equal and depend, by design, on the symbol located in the constellation diagram, offers good energy gains and more flexible data adaption [18]. In this thesis, the designs are based on probabilistic shaping as it achieves higher capacity for Gaussian channels by approximating the Gaussian distribution among the occurrence probability distribution of the transmitted symbols. In particular, up to 1.53 dB (i.e. $\pi e/6$) shaping gain can be achieved by probabilistic shaping as the modulation dimension approaches infinity [14].

To implement a probabilistic shaping system, a distribution matching (DM) encoder is used, which encodes the input bit stream to symbols with a target occurrence probability distribution. The first distribution matcher for noisy channels was implemented by a prefix-free code, proposed by G. Forney in 1984 [17]. Research on probabilistic shaping for non-uniform signalling mainly focused on designs of coded-modulation schemes and optimisations of the probability distribution to achieve higher shaping gains [16, 127] till the invention of the reverse concatenation architecture in [22] in 2015 was followed by the proposal of classical constant composition distribution matching (CCDM) algorithm in 2016 [128]. Afterwards, gains achieved using probabilistic shaping resulted in increased interest, with investigations emerging in various communication systems. Probabilistic shaping has been widely used in coherent optical transmission systems to improve the capacity and overcome the nonlinearity [19, 129]. There are also studies in wireless communications using probabilistic shaping for multicarrier signallings; The application of probabilistic amplitude shaping (PAS) in OFDM system results in up to 1.6 dB and 0.7 dB more energy-efficient than uniform signalling for a certain frame error rate in AWGN and frequency selective channels, respectively [130]. In another non-orthogonal FTN based system study, an overall 0.75 dB precoding and shaping gain is shown [131].

The rest of the chapter is organised as follows: section 4.2 describes the basic concepts of probabilistic shaping scheme; section 4.3 details the transceiver structure and addresses the key metric-the achievable spectral efficiency that is held the same for the performance comparison; then in section 4.3, simulation results are discussed to show the system performance; finally, section 4.5 wraps the findings and concludes the chapter. Part of this chapter's work includes results given in the author's 2021 conference paper [30] [1]

[1]X. Liu, I. Darwazeh, N. Zein and E. Sasaki, "Spectrally Efficient FDM System with Probabilistic Shaping," IEEE 94th Vehicular Technology Conference (VTC2021-Fall), 2021.

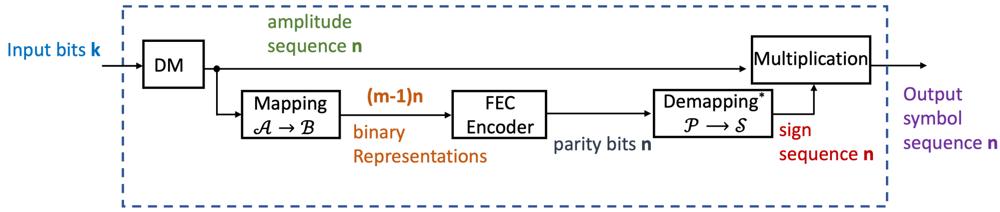
4.2 Probabilistic Shaping with Variable Transmission Rate

As pointed out in the introduction section, probabilistic shaping exploits a symmetric occurrence distribution among the constellations of transmitted symbols to maximise the data rate for a given average power. The amplitude shaping based reverse concatenation architecture introduced in 2015 [22] combines shaping and coding in one scheme. Overall, on the transmitter side, the probabilistic shaping scheme can be seen as an encoder that converts the input random binary data to a shaped sequence with a target probability distribution based on the probabilistic shaping rate R_{ps}. Such architecture cascades the distribution matcher with FEC encoding and the probabilistic shaping decoder executes the inverse operations.

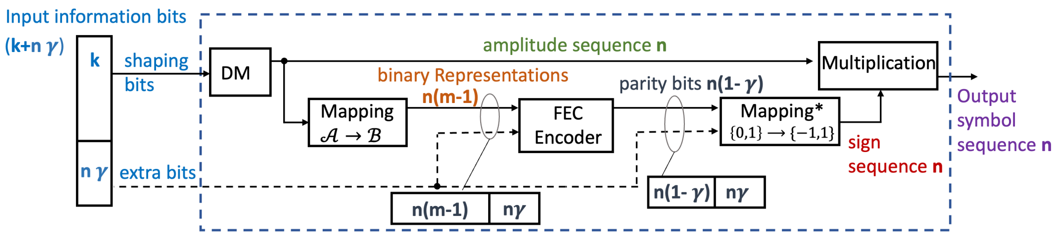
Regarding the forward error correction (FEC) coding rate, the probabilistic shaping encoder scheme has two different structures, as illustrated in Fig. 4.1. In the first case, the FEC coding rate R_c:

$$R_c = \frac{(m-1)}{m}, m = \log_2(\sqrt{M}) \tag{4.1}$$

where m is determined by the constellation cardinality M. The DM maps the k input binary bits to a stream of n amplitude symbols with a target distribution. The ratio between the number of input bits and output symbols is the DM rate that $R_{DM} = k/n$. Then, the amplitude sequences $\mathscr{A} : \{a_1, a_2, ..., a_n\}$ are mapped to their binary representations $\mathscr{B} : \{0,1\}^{m-1}$, which performs decimal to binary conversion so that each amplitude symbol is converted to $m-1$ bits. This step reuses the amplitude sequence to generate the binary inputs for the FEC encoding with coding rate R_c. Although the information bits are discarded after encoding, the same information is kept in the amplitude and thus can be recovered for the decoding at the receiver side. The $n(m-1)(1-R_c)/R_c = n$ parity bits output from the encoder are kept and mapped into sign sequence such that $\mathscr{P} \rightarrow \mathscr{S} : 0,1 \rightarrow -1,1$. Each amplitude symbol is assigned a sign by multiplying the sign sequence with the amplitude sequence, outputting the signed output symbol sequence of length n. The encoder



(a) Case 1. FEC coding rate $R_c > (m-1)/m$



(b) Case 2. FEC coding rate $R_c > (m-1)/m$

Figure 4.1: FEC encoder structures for two cases

structure of the second case is shown in Fig. 4.1 (b). Extra input bits are needed to generate the sign sequence when the FEC coding rate R_C:

$$R_c > \frac{(m-1)}{m}. \tag{4.2}$$

As shown in the figure, the extra $n\gamma$ bits are fed into the FEC encoder. Fig. 4.2 shows the structure of the binary reflected gray code (BRGC) embedded FEC encoder. The $n(m-1)$ binary representations are encoded and reused to amend the parity bits to generate the sign sequence via mapping.

As discussed previously, the distribution matcher performs shaping, setting each symbol amplitude to match the occurrence probability distribution of the transmitted symbols to a set target probability. Fixed-to-fixed length distribution matching techniques, employed regularly for probabilistic shaping, and the conventional single-composition arithmetic coding-based CCDM algorithm [128] is used in this work. With this technique, a negligible rate loss can be achieved using a relatively large input data length, as reported and thoroughly studied in an optical communications setting in [132].

As initially discussed in [128], the transmission rate is determined by the prob-

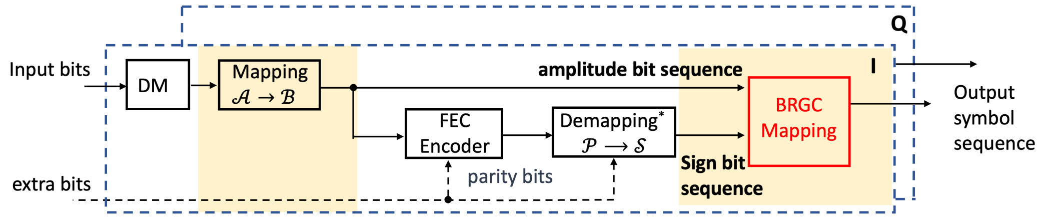
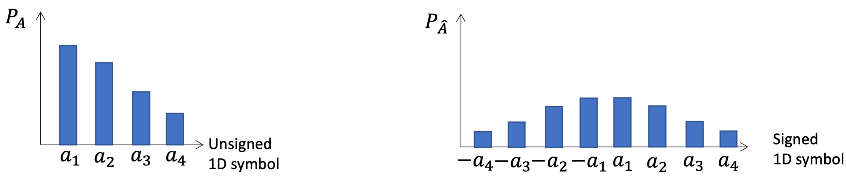


Figure 4.2: FEC encoder with BRGC embedded



(a) (b)

Figure 4.3: Probability distribution of (a) the unsigned and (b) the signed 1D symbols

ability distribution of the symbols for specified power constraints. For approaching the Shannon capacity limit of an AWGN channel, Gaussian-like distributions are considered. Specifically, the Maxwell-Boltzmann distribution is found to be the optimal distribution for the occurrence probability of transmitted symbols in AWGN channels [16]. The probability mass function of Maxwell-Boltzmann distribution is given by:

$$P_X(x) = \frac{e^{-\lambda x^2}}{\sum_{x' \in X} e^{-\lambda x'^2}}, \tag{4.3}$$

for any $x, x' \in X$ with parameter λ. Fig. 4.3 shows an example of the probability distribution of an unsigned and a signed 1D symbols. For a given base constellation and transmission data rate, the symbols can achieve optimal minimised average energy when Maxwell-Boltzmann distribution is adopted for probabilistic shaping, as discussed in section 4.3.

To clarify, the FEC coding integrated with the DM and inverse distribution matching (invDM) in this scheme is to enable data rate adaptation and to gain higher capacity. Moreover, the improved symbol error rate (SER) achieved by FEC coding also guarantees the invDM; since the invDM only works at sufficient low SER to match the shaped symbols back into the input bits. To be precise, in the initial study, when only AWGN is taken into account, the post-FEC BER and post-invDM BER are investigated to present more practical results of error performance.

4.3 Probabilistically Shaped SEFDM System

Coded-SEFDM systems have been investigated in [133] using LDPC coding, showing the advantages of bandwidth saving and less power requirement when preserving the same BER relative to OFDM. Inspired by the works on coded-SEFDM and coded-modulation-based probabilistic shaping, a non-orthogonal multicarrier SEFDM system with probabilistic shaping is proposed. Specifically, the reverse concatenation architecture [22] is employed for the probabilistic shaping encoding/decoding with the CCDM algorithm [128] and LDPC coding [134]. The simplified block diagram of the designed system is given in Fig. 4.4 by showing only the key modules of the scheme. An iterative detector (ID) is designed and employed to mitigate the interference resulting from the non-orthogonality of SEFDM [135]. Performance comparison of the proposed system is investigated relative to the OFDM scheme under fixed spectral efficiencies by adjusting other parameters such as bandwidth compression level, coding rate and probabilistic shaping rate. Theoretical analysis of the achievable spectral efficiency is provided for different modulation schemes. Simulation results show the advantages of the proposed PS-SEFDM system compared to OFDM in terms of less required power when achieving the same error performance and saved bandwidth. These matters are explored conceptually and through detailed design and mathematical modelling.

4.3.1 Transmitter Structure

The block diagram of the proposed PS-SEFDM system is depicted in Fig. 4.4. The transmitter is constructed from two parallel encoders, followed by the concatenation of an interleaver, a bit-mapper and a SEFDM modulator. The one-dimensional probabilistic shaping followed by the BRGC mapper is employed in the system model. The BRGC mapper is used to generate PAM symbols. Two parallel probabilistic shaping encoders are used to generate the in-phase and quadrature components of a two-dimensional QAM.

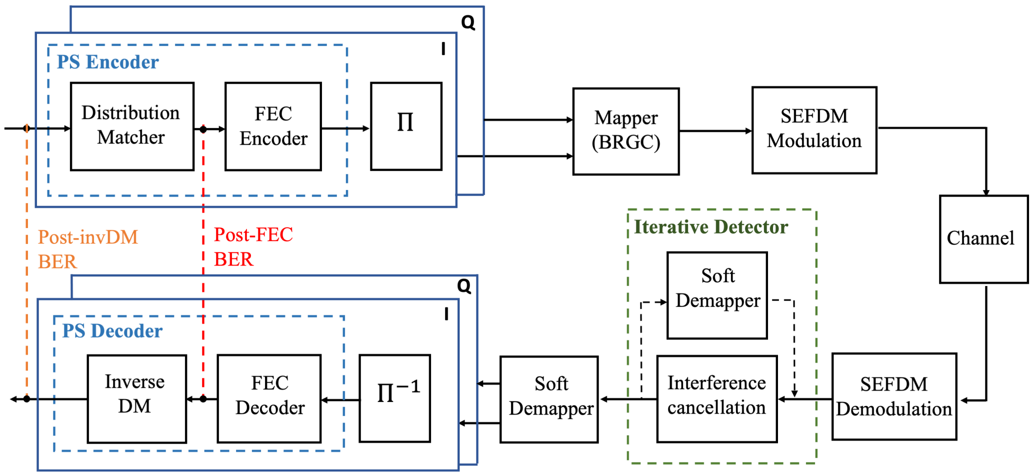


Figure 4.4: Functional simplified Block diagram of probabilistically shaped SEFDM system with iterative detection and LDPC coding.

At the transmitter, a stream of binary data stream $\mathbf{b} \in \{0,1\}$ feeds the probabilistic shaping encoder. The incoming binary sequence is framed into blocks of k bits. In the distribution matcher, each block of bit sequence is transformed into $n(m-1)$ bits corresponding to a block of amplitude sequence $\mathscr{A} \in \{0,1\}^{n(m-1)}$ with a desired occurrence probability distribution, giving the DM rate $R_{DM} = k/n$, where $m = log_2\sqrt{M}$ is determined by the constellation cardinality of M-QAM. In this work, standardized LDPC code [134] is used as the FEC, with codeword length n_c = 64,800 and coding rate $R_{fec} = (n_c - n)/n_c$. Since LDPC is a systematic binary code, in this procedure, only the n parity bits are extracted to generate the sign sequence $\mathscr{S} \in \{0,1\}^n$. Although the information bits are discarded after encoding, the same information is kept in $\mathscr{A}$ for the decoding at the receiver side. The output nm bits of the probabilistic shaping encoder is obtained by concatenating the amplitude sequence $\mathscr{A}$ and the sign sequence $\mathscr{S}$.

Then, the BRGC maps the shaped bits to PAM symbols as one of the quadrature components of M-QAM. It is worth noting that the BRGC allows the probability distribution to be reserved when the Cartesian product of the two real-valued PAM constellations is exploited. The probabilistically shaped QAM symbols are split into $N = 4$ parallel streams and modulated to N non-orthogonal SEFDM subcarriers using the IFFT-based method as mathematically detailed in [119]. The

SEFDM symbols are passed through an AWGN channel.

4.3.2 Receiver Structure

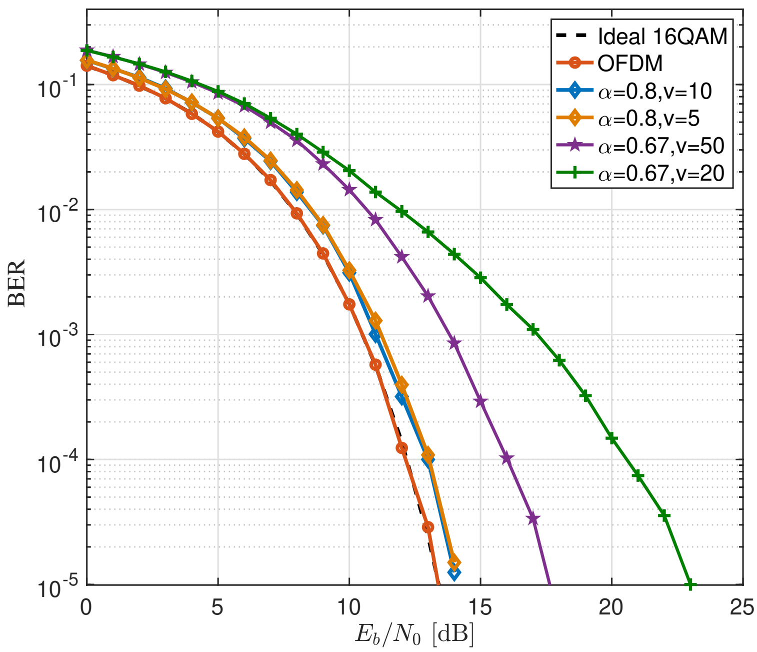


Figure 4.5: BER performance for SEFDM system using iterative detector with varying compression factor α and number of iterations.

SEFDM demodulation using modified FFT based matched filters [136] is adopted on the receiver side. However, unlike OFDM, the non-orthogonality of SEFDM with its self-introduced ICI, significant error performance degradation would result unless appropriately mitigated. To this end, the iterative detector proposed in [135] is utilised to perform the interference cancellation and to recover the transmitted symbols with minimal error rate degradation. The efficacy of iterative detector is proved that the performance degradation for SEFDM with small α can be mitigated as shown in Fig. 4.5. Compression factor α determines the bandwidth reduction in the SEFDM system. A smaller value of α leads to more severe ICI and, therefore, worse BER performance. For SEFDM signal $\alpha = 0.8$, 25% more data can be transmitted relative to OFDM within the same band. When iterations $v = 5$, the BER performance is less than 1 dB to OFDM when BER is 10^{-5}. For SEFDM signal $\alpha = 0.67$, 50% more data can be transmitted when compared to

OFDM with the same bandwidth. When iterations $v = 20$, the E_b/N_0 is 17 dB at BER is 10^{-3}. When the number of iterations increases to $v = 50$, a 3 dB advantage can be achieved when BER is at 10^{-3}.

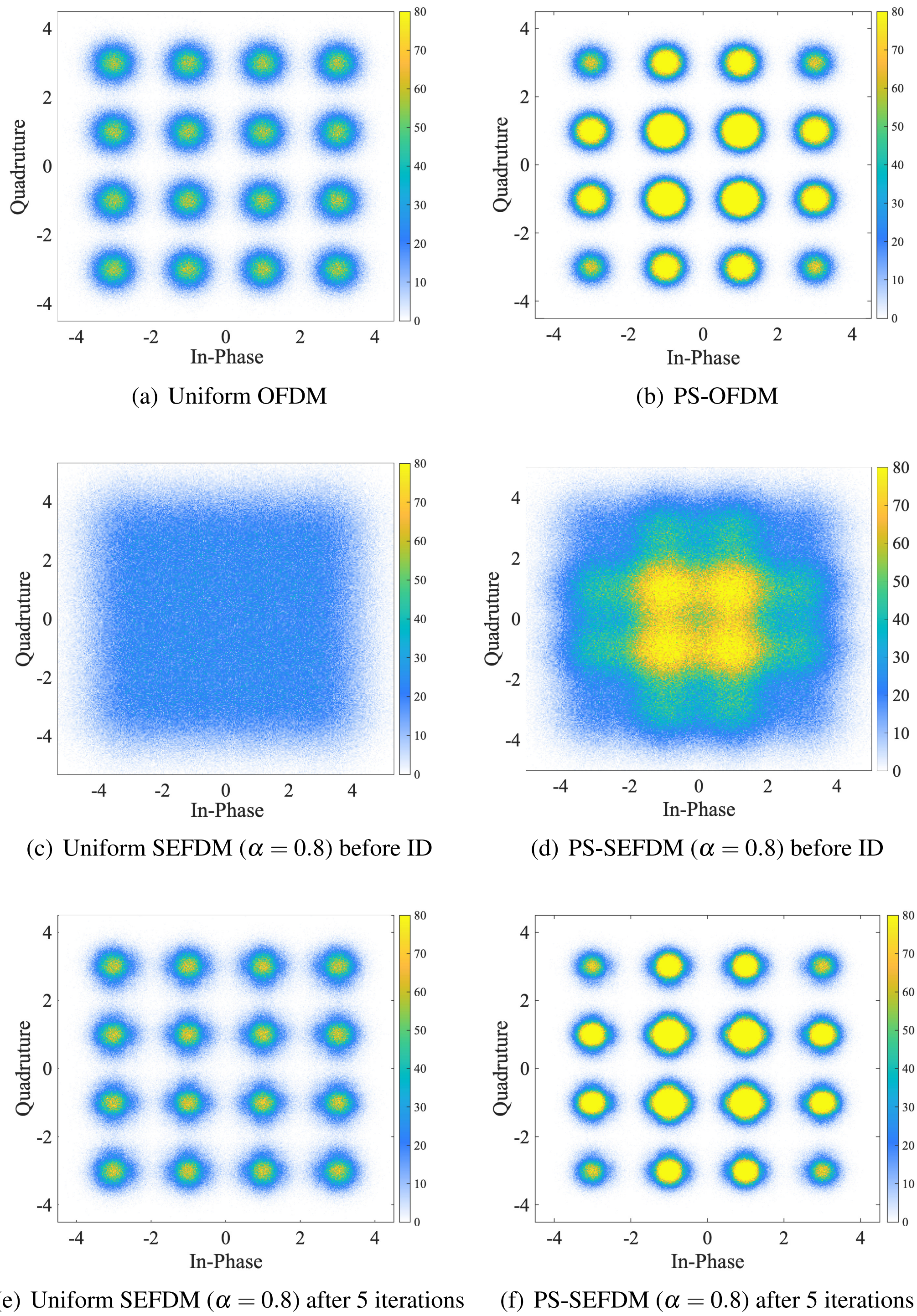
The dashed line, within the iterative detector block of Fig. 4.4, indicates the feedback process that uses the subtraction results from the last iteration to update the new ICI information, according to the number of iterations set by the model. The constellation diagrams given in Fig. 4.6 show the effectiveness of the iterative detector using SEFDM ($\alpha = 0.8$) as an example. For uniformly distributed SEFDM and PS-SEFDM, the 16-QAM constellation turns from (b) to (c) and from (e) to (f) after five iterations, respectively. The same study has been done with SEFDM ($\alpha = 0.67$), obtaining symbol recovery after 20 iterations as shown in Fig. 4.7. The output detected symbols are then passed the soft demapper to generate soft bits, i.e. the *a posteriori* LLR. After the de-interleaving, the soft bits are fed to the probabilistic shaping decoder, in particular, to the FEC decoder, so as to generate the encoded bits for the subsequent iterations of the decoding. Herein, 50 iterations are set for the LDPC decoder to achieve sufficiently low SER to make invDM perform the de-matching process.

4.3.3 Achievable Spectral Efficiency

For the proposed system, considering the PS rate R_{ps} and SEFDM compression factor α, the spectral efficiency can be expressed by:

$$\eta = \frac{log_2 M \times R_s \times R_{ps}}{\alpha \times B} \tag{4.4}$$

where M is the constellation cardinality, B represents the occupied bandwidth, R_s denotes the SEFDM symbol rate. For M-QAM symbols, the PS rate is determined by the DM rate and the FEC code rate as detailed in [132] for N-dimensional constellations. In this work, regular M-QAM is adopted. The corresponding one-dimensional constellation is used as the base constellation for probabilistic shaping. For the case of 16-QAM, $M = 16$ and therefore the associated PAM-4 constellation points are $\{\pm 1, \pm 3\}$, where the number of amplitude levels $m = 2$. Consequently,



(a) Uniform OFDM

(b) PS-OFDM

(c) Uniform SEFDM ($\alpha = 0.8$) before ID

(d) PS-SEFDM ($\alpha = 0.8$) before ID

(e) Uniform SEFDM ($\alpha = 0.8$) after 5 iterations

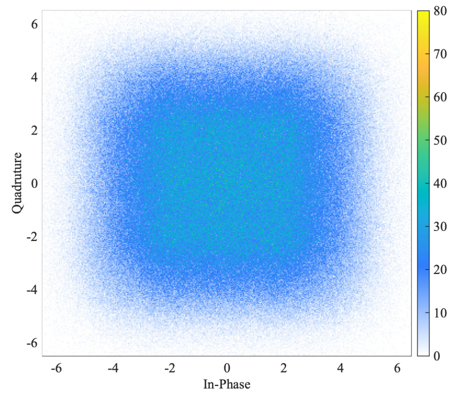
(f) PS-SEFDM ($\alpha = 0.8$) after 5 iterations

Figure 4.6: Constellation diagram for the received OFDM and SEFDM (α=0.8) 16-QAM symbols at E_b/N_0 =12 dB

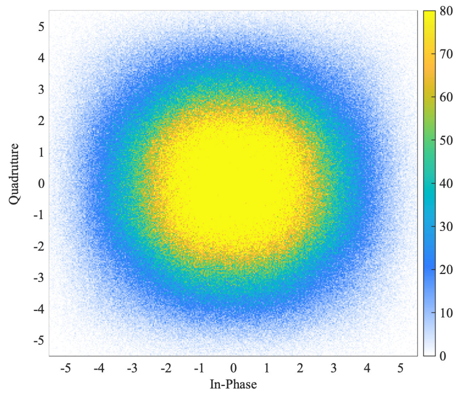
the condition of the FEC code rate R_{fec} is given by:

$$R_{fec} = \frac{n_c - n}{n_c} \geq \frac{m-1}{m}. \tag{4.5}$$

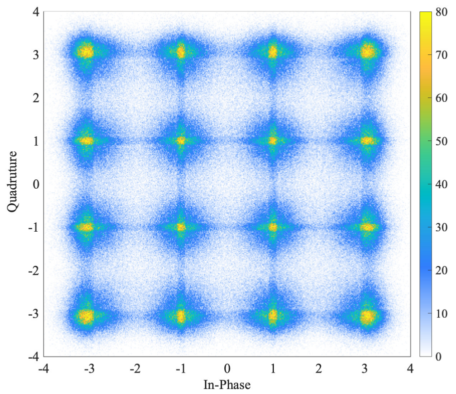
Thus, the smallest value of R_{fec} is 1/2 for 16-QAM.



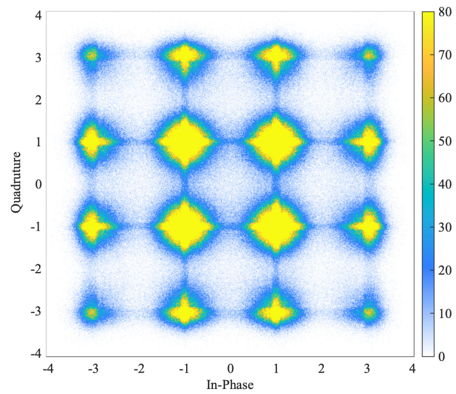
(a) Uniform SEFDM ($\alpha = 0.67$) before ID



(b) PS-SEFDM ($\alpha = 0.67$) before ID



(c) Uniform SEFDM ($\alpha = 0.67$) after 20 iterations



(d) PS-SEFDM ($\alpha = 0.67$) after 20 iterations

Figure 4.7: Constellation diagram for the received SEFDM ($\alpha = 0.67$) 16-QAM symbols at E_b/N_0 =12 dB

4.3.4 Performance Investigations

To evaluate the error performance of the proposed PS-SEFDM system, numerical simulations are carried out for both SEFDM and OFDM with and without probabilistic shaping based on complete modelling of the system blocks shown in Fig. 4.4. To compare fairly the performance of the PS-SEFDM scheme relative to OFDM as well as to uniformly distributed SEFDM, different signals achieving the same spectral efficiencies are investigated. Specifically, systems are evaluated at spectral efficiency values of 3.33, 3.75 and 4 bits/s/Hz in terms of BER performance and spectrum property; E_b/N_0 is used to measure the SNR for a fair comparison

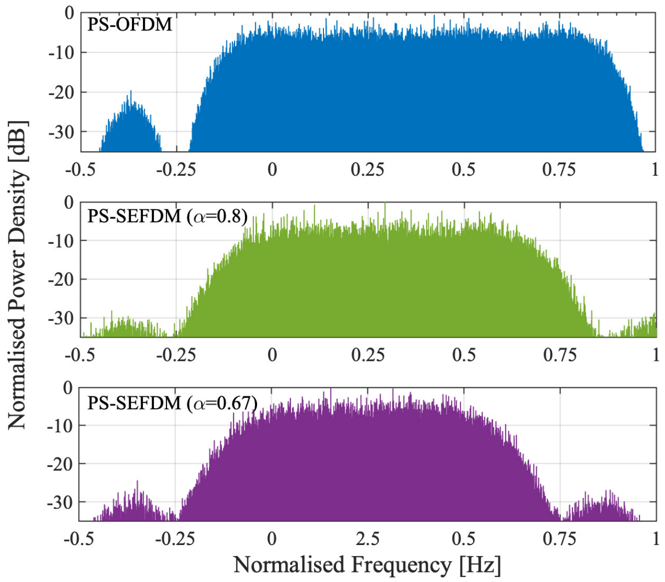


Figure 4.8: Power spectrum comparison of PS-OFDM/-SEFDM (α =0.8, 0.67).

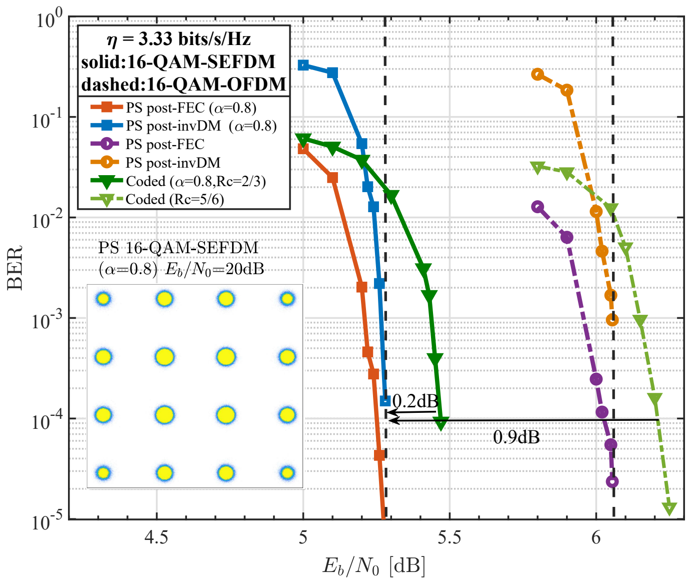


Figure 4.9: η = 3.33 bits/s/Hz: BER performance of PS-SEFDM (α = 0.8), PS-OFDM, coded-SEFDM ($\alpha = 0.8$) and coded-OFDM.

due to the use of FEC coding and bandwidth compression. The required E_b/N_0 are assessed at post-FEC BER of 10^{-4} meanwhile the post-invDM BER of 10^{-3} approximately. For spectral efficiency $\eta = 3.33$ bits/s/Hz, the comparison is held for 16-QAM SEFDM ($\alpha = 0.8$) and OFDM with and without probabilistic shaping. In this case, coded-OFDM with code rate $R_c = 5/6$ achieves the same spectral ef-

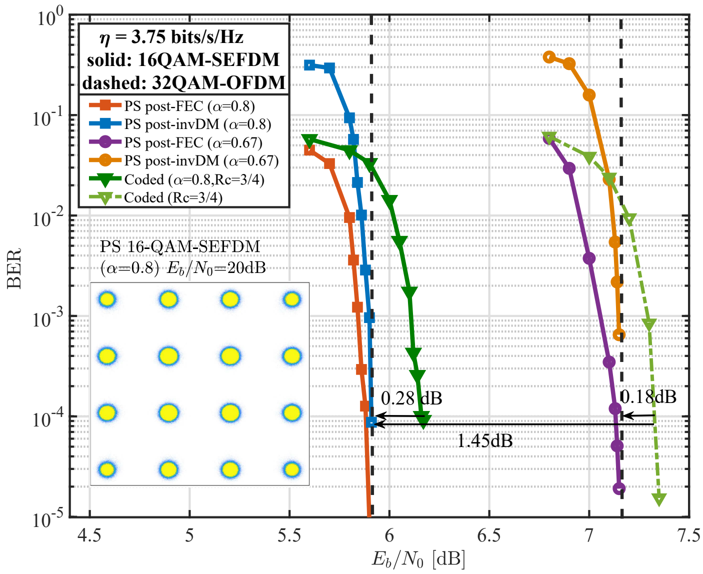


Figure 4.10: $\eta = 3.75$ bits/s/Hz: BER performance of PS-SEFDM ($\alpha = 0.8, 0.67$), coded-SEFDM ($\alpha = 0.8$) and coded-OFDM.

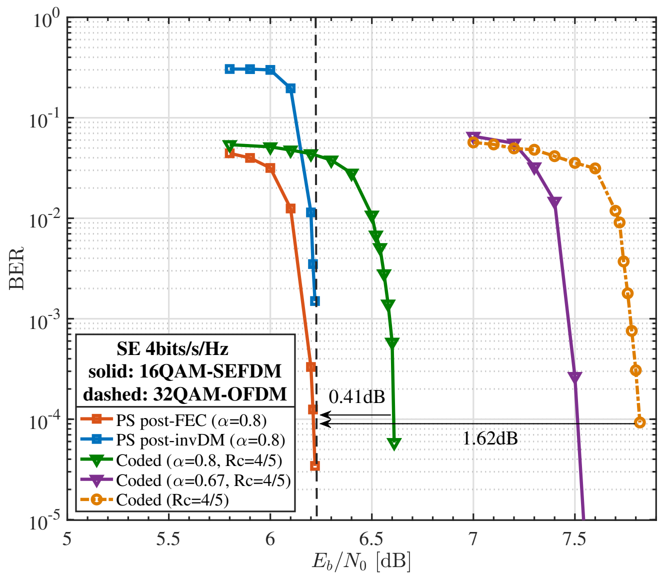


Figure 4.11: $\eta = 4$ bits/s/Hz: BER performance of PS-SEFDM ($\alpha = 0.8$), coded-SEFDM ($\alpha = 0.67, 0.8$) and coded-OFDM.

ficiency as coded-SEFDM with $R_c = 2/3$. To achieve the same spectral efficiency, the corresponding R_{ps} of the PS scheme is required to equal the R_c of the schemes without probabilistic shaping. For the $\eta = 3.75$ bits/s/Hz spectral efficiency case, the PS-16-QAM-SEFDM symbols, with two different compression factors $\alpha = 0.8$

and 0.67, are compared with the 32-QAM coded-OFDM. For the $\eta = 4$ bits/s/Hz spectral efficiency case, the PS-16-QAM-SEFDM with $\alpha = 0.8$ is compared with coded-SEFDM ($\alpha = 0.8$ and 0.67) as well as coded-OFDM.

Before assessing the BER performance, the constellation diagrams of the uniformly distributed and the probabilistically shaped symbols are first studied using a base constellation of 16-QAM. These constellation diagrams are given in Fig. 4.6 for OFDM and SEFDM ($\alpha = 0.8$) symbols and SEFDM ($\alpha = 0.67$) in Fig. 4.7. The simulations were conducted at $E_b/N_0 = 12$ dB. Our analysis reveals that, in both OFDM and SEFDM with probabilistic shaping, the symbols with lower energy are transmitted more frequently, consistent with theoretical expectations. Additionally, both the constellation diagrams of the received signals before and after applying the ID are provided in the two figures, showing the effectiveness of ID in ameliorating the distortion caused by the self-introduced ICI in SEFDM. The number of iterations $v = 5$ is used for $\alpha = 0.8$ and $v = 20$ for $\alpha = 0.67$, which are shown to be sufficient to recover the distorted signals such that the constellation symbols can be distinguished.

Simulations are conducted for various spectral efficiencies and only three groups of results are presented to illustrate the performance and advantages of using PS-SEFDM over OFDM. The first group with 3.33 bits/s/Hz in Fig. 4.9 shows the advantage of using SEFDM with probabilistic shaping and a lower coding rate relative to OFDM, specifically, 0.7 dB power advantage and 0.2 dB shaping gain. The total 0.9 dB gain of 16-QAM PS-SEFDM ($\alpha = 0.8$) can be seen when compared to 16-QAM OFDM without probabilistic shaping. In all cases, the SEFDM signal has 25% bandwidth saving relative to the OFDM, noting the spectrum of the ($\alpha = 0.8$) case of Fig. 4.8. For the second group with spectral efficiency $\eta = 3.75$ bits/s/Hz, Fig. 4.10 shows a similar advantage of PS-SEFDM when $\alpha = 0.8$. The PS-16-QAM-SEFDM achieves 1.45 dB power savings when compared to the uniformly distributed 32-QAM-SEFDM. For PS-SEFDM with $\alpha = 0.67$ (bottom spectrum of Fig. 4.8), which is lower than the Mazo limit (i.e. 0.8), there is a slight performance advantage as well as 50% bandwidth saving over the uniformly distributed

Signal Type	**Data Rate [Mbits/s]**	**Coding Rate**	**DM Rate [Pr]**	$\mathbf{E_b/N_0}$ at 10^{-5}	**Shaping Gain [dB]**
OFDM	300	3/4	-	5.33	0.37
PS-OFDM	300	4/5	9/10	4.96	
SEFDM(0.8)	375	3/4	-	6.2	0.4
PS-SEFDM(0.8)	375	4/5	9/10	5.8	
SEFDM(0.67)	400	2/3	-	7.6	0.14
PS-SEFDM(0.67)	400	3/4	5/6	7.46	

Table 4.1: Performance for multicarrier 16-QAM signals of the same bandwidth

32-QAM OFDM. Furthermore, for the third group where signals achieve spectral efficiency of 4 bits/s/Hz, Fig. 4.11 shows that PS-16-QAM-SEFDM achieves a 1.62 dB power advantage when compared to coded-OFDM. It can also be observed that PS-SEFDM can achieve 0.41 dB and 1.31 dB advantages over coded-SEFDM with $\alpha = 0.8$ and 0.67, respectively.

In addition, a set of numerical simulations are carried out for SEFDM and OFDM signals with and w/o probabilistic shaping that occupy the same bandwidth of 100 MHz. Results are provided for 16-QAM signals for systems with different α values. The results show the BER performance between the problematically shaped and uniformly distributed signals for comparison purposes. Fig. 4.12 depicts the error performance for different groups of signals that transmit at the data rate of $R_b = 300, 360, 375, 400$ bits/s/Hz. The results confirm the high transmission rate is at the cost of the error performance. Moreover, the PS signals outperform the coded signals, i.e., the uniformly distributed signals, for all the transmission rates that have been examined. In addition, the shaping gain results are concluded in Table 4.1 for the signals using the same bandwidth and α values. The shaping gains demonstrate that PS signals consistently outperform non-shaped signals regardless of the α value when the same data rates are achieved. It is worth noting that the shaping gain is not proportional to the transmission data rate. This can be explained by the different combinations of coding rates and DM rates.

To conclude, the PS-SEFDM consistently outperforms OFDM as well as non-shaped SEFDM with the same spectral efficiency. When the same BER perfor-

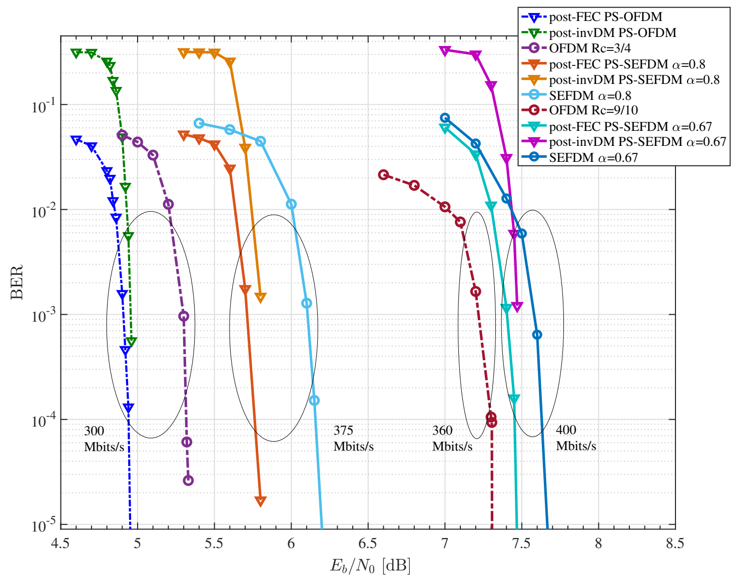


Figure 4.12: BER Performance comparison: 16-QAM signals of same bandwidth

mance is obtained, the shaping procedure introduces extra shaping gain on top of the power advantage of SEFDM using a lower coding rate [133] relative to OFDM. In addition, the PS scheme per se is limited by the DM rate considering the rate loss as well as the minimum FEC code rate as given in equation (4.5). Using SEFDM with varying compression factor α extends the flexibility of rate adaptation of the probabilistic shaping scheme.

4.4 Probabilistic Shaping in Multipath Fading Channels

The previous sections presented a new robust PS-SEFDM system with enhanced spectral efficiency. It was demonstrated that the PS-SEFDM signals offer significant power advantages over OFDM and non-shaped SEFDM signals, regardless of the transmission rates or achievable spectral efficiencies. In the investigations, systems with probabilistic shaping were tested in the presence of AWGN to examine performance improvement. However, the question arises whether the power advantages achieved by PS-SEFDM signals remain in practical propagation scenarios. To further study the performance of PS-SEFDM signals when impaired by the wireless channel effects, the wireless channel effects are first addressed with a focus on multipath fading. Two typical multipath fading channel scenarios are discussed: a static frequency selective channel model and Rummler's two-ray fading channel model.To compensate for the channel effects, channel estimation and equalisation techniques are typically used to obtain channel information and mitigate signal distortion. In this work, the PS-SEFDM systems are designed using channel estimation and equalisation techniques and test the system uses the two aforementioned multipath fading models. Similar to the performance evaluation in the previous studies, the system performance of PS-SEFDM is compared with OFDM and non-shaped SEFDM to show its maintained advantages in multipath fading channels. Overall, this section comprehensively analyses the practical considerations involved in designing PS-SEFDM systems for wireless communication.

4.4.1 Modelling and Analysis of Multipath Fading Channel

The commonly used AWGN channel was introduced in chapter 2. The following discusses several wireless channel models to approximate different practical propagation environments. They are used for evaluating the system performance in multipath fading channels in the simulation study.

In practical wireless communications, the signals do not always travel to the receiver along a single direct path, i.e. line-of-sight (LoS). Because of the varying propagation environments, major electromagnetic wave propagation mechanisms such as reflection, diffraction, and scattering are attributed to occur and subsequently impact the amplitude, phase, and delay of the transmitted signal. This leads to the conclusion that the inherent nature of the mobile radio channel constrains the performance of wireless communications. Besides, it reveals the importance of channel modelling. Communication systems are designed and developed based on the analytic channel model to compensate for wireless impairments and improve communication performance.

Wireless channels are modelled analytically based on the measurements. There are two major propagation models, distinguished by the transmitter-receiver separation distance, namely the large-scale and the small-scale propagation models.

Herein, the former model is mainly used to estimate the radio coverage range, while the latter, also known as a fading model, is used to measure the multipath environments. In this study, the following sections mainly focus on the multipath fading channel.

The channel impulse response is used to represent the mobile radio channel model. Assume there are the transmitted signal $x(t)$ and received signal $y(t)$, which are also the input and output of the channel. The output of a signal passing through a multipath fading channel can be expressed by the following equation,

$$y(t) = x(t) \otimes h(t,\tau) + w(t) \tag{4.6}$$

where $\otimes$ denotes the convolution operation, $h(t,\tau)$ is the channel impulse response wherein t is the varying time and τ represents the multipath delay, $w(t)$ is the AWGN component. The channel can be seen as a linear filter applied to the signal.

A discrete-time channel with l paths in total and the delays of $\tau_n = n$ samples for each path is considered to make the channel term more concrete. Thus, the received k^{th} samples $y[k]$ is related to the input sample stream $x[k]$ and the channel h, given by

$$y[k] = \sum_{1}^{l} h_{k,n} x[k] + w[k] \tag{4.7}$$

By vectorising the components, the system can be expressed by

$$Y = HX + W, \tag{4.8}$$

where for different types of fading channels, the matrix H is different. There are several parameters that characterise the communication channel statistically.

The multipath fading channel can be classified in two ways: slow/fast fading channel and flat/frequency selective fading channel [137]. Doppler effects lead to Doppler spread for the transmitted symbols, where the Doppler frequency shift occurs, and consequently, the range of frequencies is broadened. Coherence time measures the maximum duration over which the channel response is approximately time-invariant. This duration can be given by the reciprocal of the Doppler spread. When the symbol interval is smaller than the coherence time of the channel, the fading is termed slow fading. In contrast, fast fading occurs when the symbol interval is greater than the coherence time. The delay spread indicates the maximum delay for the transmitted signal among all paths. The associated coherence bandwidth, given by the reciprocal of delay spread, is compared with the narrowband signal. When the signal bandwidth is smaller than the coherence bandwidth, it is referred to as flat fading. Otherwise, the fading is regarded as frequency selective.

Static Frequency Selective Channel

The static frequency selective channel model indicates a time-invariant/time-independent scenario. One of the prominent features of such a channel is that the Doppler frequency is 0 Hz. The fading is a result of multipath propagation. Herein we assume that the Rayleigh process occurs in each path.

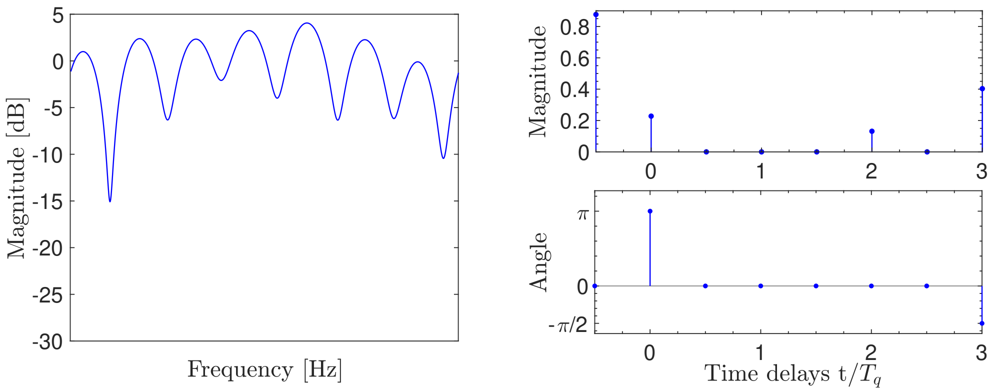
When only fading effect is accounted for, the impulse response of the static channel is given by equation (4.9), which was used in [101, 138] when evaluating the channel estimation methods for the SEFDM system.

$$\begin{aligned} h(t) =& 0.8765\delta(t) - 0.2279\delta(t-T_q) + 0.1315\delta(t-5T_q) \\ & - 0.4032e^{i\pi/2}\delta(t-7Tq), \end{aligned} \tag{4.9}$$

or its vectorised form given by

$$h = [0.8765,\ -0.2279,\ 0,\ 0,\ 0.1315,\ 0,\ 0,\ -0.4032e^{i\pi/2}]^T, \tag{4.10}$$

where δ is the Dirac delta function, $T_q = T/Q$ is the sample period, $T = N \times T_s$ is the multicarrier signal duration with N subcarriers and Q denotes sample-per-symbol. Figure 4.13 shows the channel time and frequency response.



(a) Frequency Response (b) Time Response

Figure 4.13: Frequency and time response of static frequency selective channel

Rummler's Two-ray Channel Model

In terrestrial microwave communications, two- and three-wave (or two-/three-ray) models are commonly used to describe the multipath fading channel. Rummler's model [139] was developed using three paths and was fixed later to a two-path model, which is one of the most widely used to characterise fading in LoS microwave links 2-6 GHz and 11 GHz bands [140, Chapter 4].

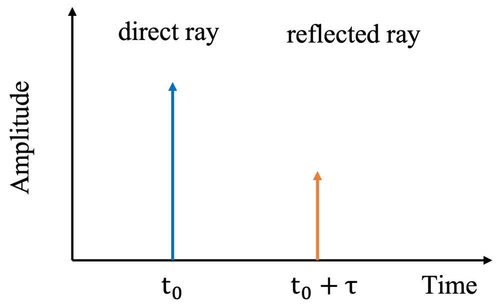
Fig. 4.14 (a) illustrates the Rummler's two-ray channel. The channel model considers two received signals from the direct and reflected ray. There are also two conditions-minimum-phase, where the direct ray arrives before the reflected ray as shown in Fig. 4.14 (a), and non-minimum-phase, where the reflected ray comes before the reflection. The minimum-phase condition is chosen to be the focus of this study. In this condition, the direct ray can be expressed by:

$$\text{direct ray} = \alpha_d \delta(t - t_0) \tag{4.11}$$

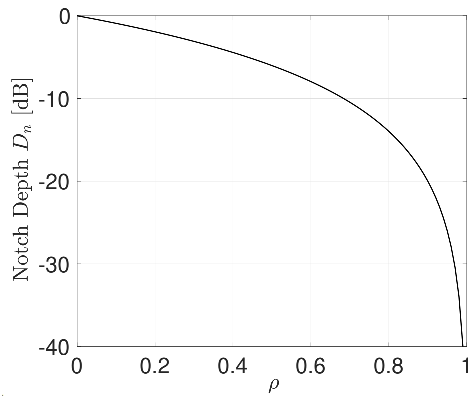
and the reflected ray is expressed by:

$$\text{reflected ray} = \alpha_r \delta(t - t_0 - \tau) \tag{4.12}$$

where α_d and α_r are the path attenuation and τ is the relative delay between the direct and the reflected rays. It is worth noting that $\tau = 6.3ns$ for Rummler's two-



(a)



(b)

Figure 4.14: Rummler's two-ray channel: (a) Illustration of the two rays (b) Relation between the notch depth and the amplitude ratio of the two rays

ray channel. Assume the direct ray phase angle to be 0 rad, the reflected ray has

$$\alpha_r = \rho \cdot \alpha_d e^{j\theta}, \tag{4.13}$$

where ρ is the amplitude ratio and θ indicates the phase difference. If the transmitted signal **s** is passed through the two-ray channel, the output signal is a combination of two signals:

$$\begin{aligned} r &= r_d + r_r \\ &= \alpha_d \theta(t - t_0) \otimes \mathbf{s} + \alpha_r \theta(t - t_0 - \tau) \otimes \mathbf{s} \\ &= \alpha_d \theta(t - t_0) \otimes \mathbf{s} + \rho \cdot \alpha_d e^{j\theta} \theta(t - t_0 - \tau) \otimes \mathbf{s}. \end{aligned} \tag{4.14}$$

The frequency response of the Rummler's propagation channel is given by:

$$H(f) = 1 - \rho \cdot e^{-j2\pi(f - f_n)\tau}, \tag{4.15}$$

wherein the notch frequency f_n corresponds to the minimum magnitude of the response. The relative notch depth D_n is determined by the amplitude ratio ρ. The relation can be given by

$$\rho = 1 - 10^{(-\frac{D_n}{20})}. \tag{4.16}$$

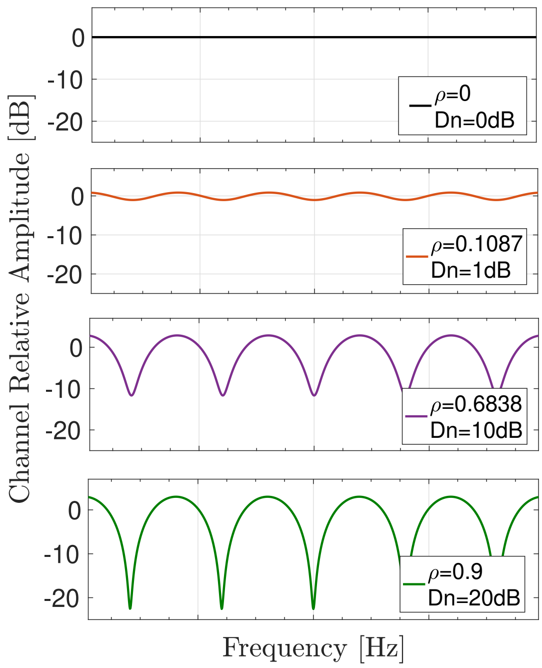


Figure 4.15: Attenuation of the modeling function of the Rummler's channel for depth notch $D_n = 0, 1, 10, 20\,\text{dB}$

which is shown by Fig. 4.14 (b). Fig. 4.15 shows the channel attenuation of Rummler's channel model with varying notch depths. In the study, the notch depths $D_n = 10$ dB is selected and two systems are investigated in simulations: coded-OFDM and PS-OFDM. The error performance of both signals is evaluated with and without the use of frequency domain channel estimation and equalisation techniques. Before demonstrating the numerical results of the comparative study, channel estimation and equalisation are briefly discussed in the next section, focusing on the technique used in this work.

4.4.2 Channel Estimation and Equalisation

The last section describes several statistical channel models that fit different signal propagation scenarios. Impairments occur to the transmitted signals as they pass through the wireless channel. To compensate for such channel effects and recover the signals, the receiver is required the acquisition of the channel state information (CSI). It is unrealistic to achieve the perfect CSI knowledge (i.e. ideal CSI) from a practical perspective. While in some studies, the ideal CSI is assumed to establish the achievable performance bounds without the disturbances of channel effects. This section will focus on the techniques for CSI acquisition by performing channel estimation, particularly the pilot-aided estimation, including both time domain and frequency domain schemes. The performance evaluation of estimators will be discussed in conjunction with channel equalisation methods.

Channel estimation has been extensively researched on OFDM. In the literature, various estimation techniques were proposed [141]. Pilot-aided channel estimation is a typical method widely used in practice. A known pilot tone or training sequence is transmitted and embedded within the data signal. Since the pilots are predefined, it is always known by the receiver. In this case, the receiver can reconstruct the channel response from the received signal and the known pilots. This can be achieved by applying algorithms to exploit the time-frequency correlation of the pilot signals [142]. These algorithms are mostly least square (LS), MMSE or ML based [143].

Unlike pilot-aided estimation, blind and semi-blind channel estimation techniques avoid using pilots and thus achieve higher throughput [144]. Existing blind and semi-blind estimations can be categorised into statistical and deterministic methods. The statistical techniques are mainly based on the statistics of signals, which require sufficient OFDM symbols to estimate the channel auto-correlation matrix accurately [145]. In general, these methods are of low complexity. However, most statistical methods suffer from slow convergence. On the other hand, the deterministic methods process the demodulated/post-DFT received signals where exhaustive search, such as ML-based approaches, is used. Though these methods are highly complex, the much faster convergence and superior performance make them attractive compared to the statistical techniques [146].

Recalling the pilot-aided estimation, a similar trade-off exists between the performance and complexity for blind/semi-blind estimation. Basic channel estimation can be carried out in either frequency or time domain. The time domain channel estimation, block-type pilot estimation, is performed by inserting pilot tones in each block period. The time domain channel can be modelled as an FIR filter for

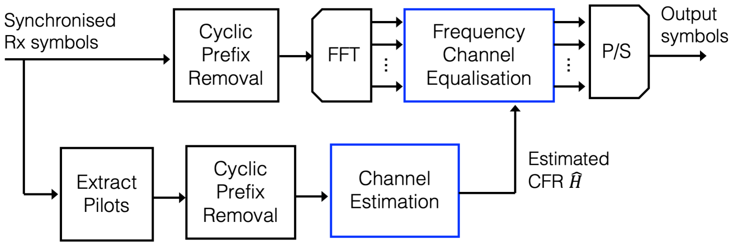


Figure 4.16: Block Diagram for Pilot-assisted Channel Estimation Equalisation in Frequency Domain

single-carrier systems. The coefficients of such FIR filters can be estimated from the received time domain symbols and subsequently transformed into the channel frequency response. Though the estimation scheme for single-carrier systems can be applied to multi-carrier systems, the estimation algorithms for OFDM are simple to implement across time or frequency domains due to the unique orthogonality property [147]. However, for non-orthogonal multi-carrier systems, taking the SEFDM system as an example, estimation of the channel is challenged by the interference effects. [148] proposed two estimation techniques for the SEFDM system, full channel estimation and partial channel estimation. For the former method, the pilot signals are conveyed by a SEFDM symbol. While in the second method, a mutually orthogonal subset from SEFDM symbols is selected to carry the pilot signals, designed to tackle the ill-conditioning of the SEFDM system that arises with the ICI effect. The estimation accuracy of both techniques is proved; however, the high complexity of matrix inversion calculation makes the implementation impractical. Nevertheless, if the channel is static, such as for radio over fibre transmission in [149], the DSP can be done offline, thus allowing the aforementioned method to be feasible. In another experimental work in [150], the recursive least square (RLS) algorithm proposed for SEFDM based VLC system. Though the time domain RLS scheme benefits from its fast convergence, it suffers high computational complexity to achieve reasonable estimation accuracy. Hence a balanced design of RLS scheme is needed considering both the performance and complexity.

The frequency domain estimation scheme processes the frequency response of received signals. For pilot-aided estimation, it is also known as comb-type channel estimation, where pilots are inserted into specific subcarriers of each OFDM symbol. To estimate the data-bearing subcarriers, interpolation is required, which can be either simple linear or lowpass interpolation. Particularly for the second interpolation, lowpass filters with cut-off frequency are used, and substantial pilots are required to match the perfect sampling of the channel [151]. In addition, the subcarriers that are chosen to carry pilot signals need to be evenly distributed to achieve optimal estimation performance [152]. For example, in IEEE standard 802.11a [153], four equally spaced subcarriers are employed in the OFDM system for WLAN network. The equidistant scattered pilot subcarriers are also employed for OQAM-FBMC systems in [154], where LS and DFT-based interpolation with de-noising are used for estimation. At the same time, over-sampled domain equalisation is considered to reduce the fading effect. [155] utilised the LS method to estimate the channel frequency response for dual polarisation coherent optical SEFDM system, where a second-order polynomial interpolator is chosen to perform the frequency domain estimation. However, the use of interpolation leads to a complexity increase as well as an accuracy reduction. To solve this problem, two simple yet robust estimation methods for SEFDM systems were proposed in [101] without using an interpolator, which provides a more accurate estimate and low complexity.

Based on the above discussion, each algorithm's estimation accuracy and computational complexity differ. The performance of channel estimators is not only dependent on the algorithms. The channel conditions and the a priori information exploited also affect the estimates. The channel estimation can be carried out in either frequency or time domain. In the case of fast-fading channels, the time domain estimator outperforms the frequency domain estimator as proved in [156]. In contrast, pilots need to update frequently enough to capture the channel changes over time for slow-fading channels. To evaluate the accuracy of a channel estimator, mean square error (MSE) is used as the measure and is given by

$$\mathrm{MSE} = \mathbb{E}[\boldsymbol{h} - \hat{\boldsymbol{h}}]^H [\boldsymbol{h} - \hat{\boldsymbol{h}}], \tag{4.17}$$

where $\mathbb{E}[\cdot]$ is the expectation operator and $[\cdot]^H$ is the Hermitian operator.

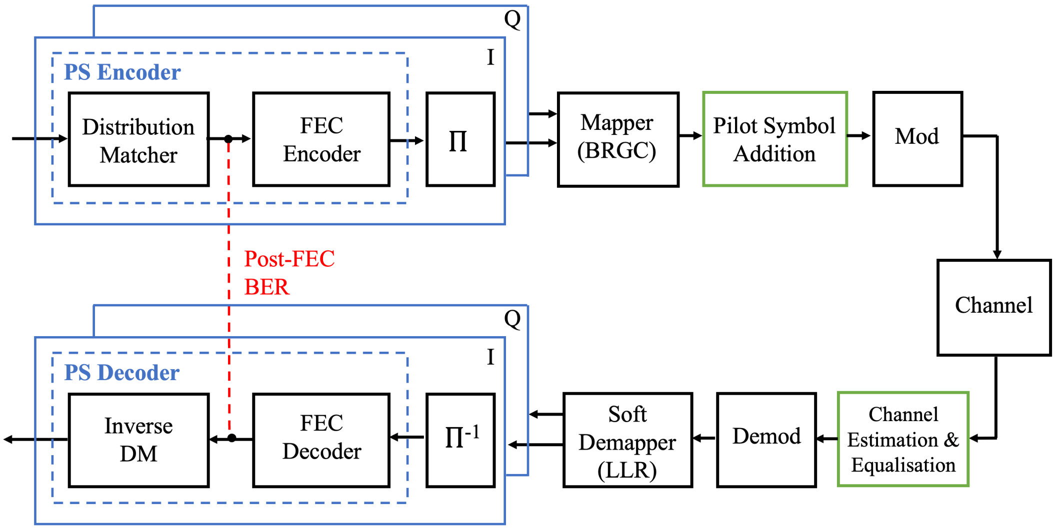


Figure 4.17: Functional simplified block diagram of PS-SEFDM system with pilot-aided channel equalisation

On the other hand, the computational complexity of the channel estimation algorithm is usually calculated in terms of the number of arithmetic operations, such as mystification and addition. Although, with this, subtraction is considered a type of addition as far as the aforementioned algorithms are concerned, sorting and compare-select operations are not taken into account for the computational complexity in this thesis.

4.4.3 Performance Investigations

A high-level system design is shown in Fig. 4.17, where the soft demapper together with an iterative detector is used, same as the system in Fig. 4.4 and the channel estimation and equalisation designs have been detailed in Fig. 4.16. In this work, the two channel models are used to test the PS-SEFDM system and the corresponding performance is assessed and compared with OFDM and coded SEFDM, achieving the same spectral efficiencies. A CP of sufficient length is added to the SEFDM symbols to avoid the ISI.

Based on the channel estimation and equalisation algorithms discussed above, we first evaluate the accuracy of the channel estimator by using the metric MSE given by equation (4.17). Simulation results in Fig. 4.18 show the MSE versus E_b/N_0 in dB for the pilot-aided channel estimation technique for SEFDM in the

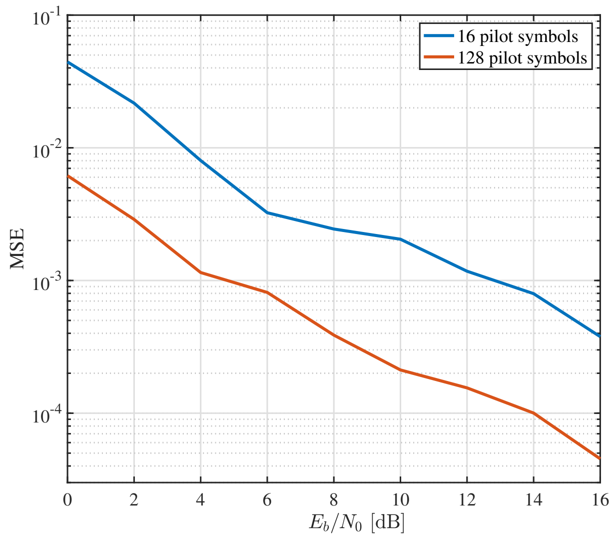
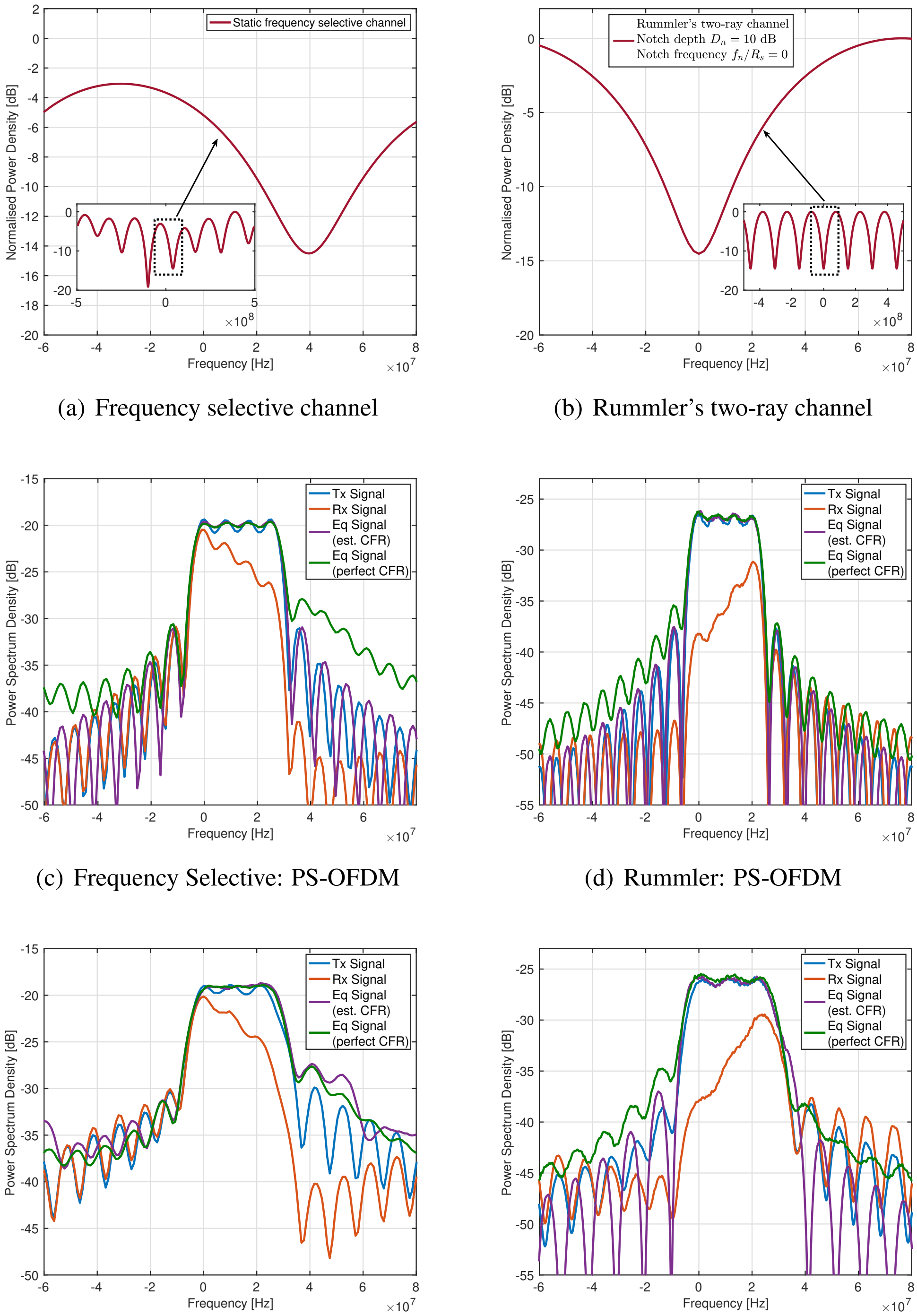


Figure 4.18: MSE results for the channel estimation using 16 and 128 pilots

static frequency selective channel with different pilot sizes ($N_{pilot} = 16, 128$). It can be seen from the figure that the estimation accuracy is affected by the channel condition such that the estimation is more accurate as the SNR is higher. Moreover, the estimation performance is reasonable for both pilot sizes, while it is more accurate when the pilot size is larger. As a result, the pilot size $N_{pilot} = 128$ is adopted for the further performance investigation of the PS-SEFDM system in multipath fading channels.

The simulation results in Fig. 4.19 show the frequency response of the static frequency selective channel, the Rummler's two-ray channel, and the power spectra of the transmitted, received and equalised signals. In the figure, Fig. 4.19 (a) and (b) show the notch in the channel band, which are obvious when comparing the transmitted and the received signal spectra. The perfect CSI and estimated CSI obtained using the pilot symbols are compared to show the effectiveness of the adopted frequency domain channel estimation and equalisation technique. The corresponding recovered signal spectra are given in Fig. 4.19 (c), (d), (e) and (f) for PS-OFDM and SEFDM.

To assess the performance of the PS-SEFDM in the static frequency selective channel and the Rummler's two-ray channel, numerical simulations are conducted



(a) Frequency selective channel

(b) Rummler's two-ray channel

(c) Frequency Selective: PS-OFDM

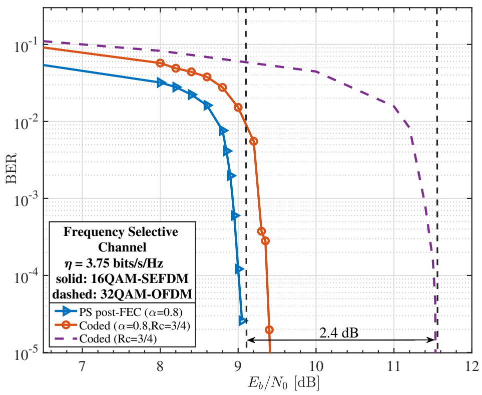
(d) Rummler: PS-OFDM

(e) Frequency selective: PS-SEFDM ($\alpha = 0.8$)

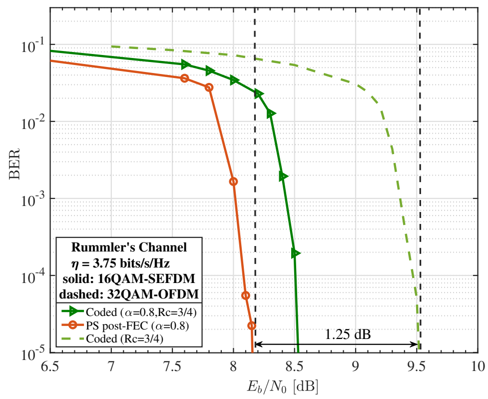
(f) Rummler: PS-SEFDM ($\alpha = 0.8$)

Figure 4.19: Frequency response of two channel models and spectra of transmitted, received and equalised PS-SEFDM and uniform OFDM signals

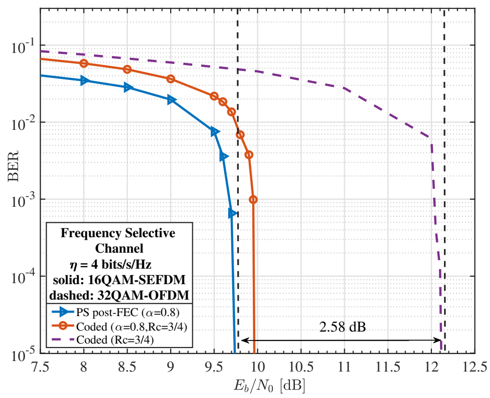
for two achievable spectral efficiencies $\eta = 3.75$ and 4 bits/s/Hz for PS-16-QAM-SEFDM ($\alpha = 0.8$) and coded 32-QAM-OFDM. For both systems, four subcarriers



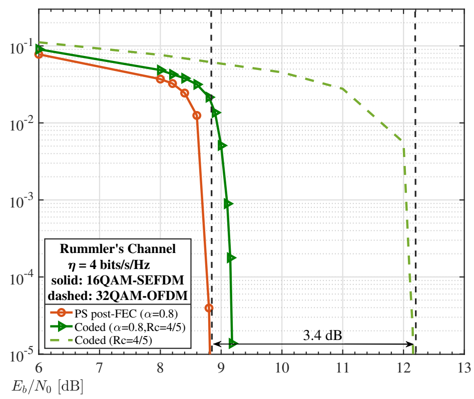
(a) Frequency Selective Channel (η=3.75 bits/s/Hz)



(b) Rummler's Channel (η=3.75 bits/s/Hz)



(c) Frequency Selective Channel (η=4 bits/s/Hz)



(d) Rummler's Channel (η=4 bits/s/Hz)

Figure 4.20: BER for PS-SEFDM in multipath fading channels

are used, and the same set of combinations of coding rates and DM rates are utilised to achieve certain transmission rates. More specifically, the coding rate is $R_c = 3/4$ for the coded OFDM signal while the coding rate is set to $R_c = 4/5$ and the DM rate is $R_{dm} = 9/10$ for the PS-SEFDM signal for the case $\eta = 3.75$ bits/s/Hz. Similarly, for the case $\eta = 4$ bits/s/Hz, the coding rate is $R_c = 4/5$ for coded OFDM while $R_c = 5/6$ and $R_{dm} = 14/15$ are used for PS-SEFDM. Fig. 4.20 (a) and (c) present the power advantages of 2.4 dB and 2.58 dB for PS-16-QAM-SEFDM over coded 32-QAM-OFDM when transmitted the static frequency selective channel, respectively. Similar power advantages of 1.25 dB and 3.4 dB achieved by PS-16-QAM-SEFDM are presented in Fig. 4.20 (b) and (d) in the Rummler's two-ray channel. These results confirm that PS-SEFDM outperforms coded-OFDM in

AWGN and multipath fading channels for various achievable spectral efficiencies.

4.5 Conclusions

The work in this chapter investigates the use of constellation shaping, specifically probabilistic shaping schemes, to improve the spectral efficiency of OFDM and SEFDM. A new system design for robust transmission of PS-SEFDM signals is proposed, which employs the conventional fixed-length single-composition distribution matching algorithm and reverse concatenation probabilistic shaping architecture coupled with LDPC coding and iterative detector for interference cancellation. A fair comparison is conducted by maintaining a fixed achievable spectral efficiency for SEFDM with and without probabilistic shaping, with varying compression factors compared to OFDM. The numerical results demonstrate significant shaping gains and the power advantage relative to OFDM. Therefore, this technique advantageously results in transmission energy saving coupled with substantial bandwidth saving substantial and allows flexible data rate adaptation, all at the expense of a limited increase in complexity.

The proposed PS-SEFDM systems are tested in multipath fading channels to verify the proposed scheme's suitability for practical systems. This chapter briefly reviews multipath fading channel models, highlighting the static frequency selective channel and Rummler's two-ray channel model, which is largely suited for fixed links. These two channel models are chosen for the performance evaluation for PS-SEFDM system by considering the multipath fading effects of different statistical features. The system design therefore employs frequency domain estimation and equalisation techniques to mitigate the fading effects, so as to maintain the reliable transmission. The numerical results show that signals with probabilistic shaping have reduced error penalty compared to uniformly distributed signals in multipath fading channels. Besides, PS-SEFDM signals benefit more from both coding and shaping gains when compared to PS-OFDM. Overall, the results demonstrate the potential of the proposed system design for achieving high spectral efficiency and robust transmission in practical communication systems.

Chapter 5

Multidimensional Modulations and Mapping Optimisation

Modulation is the key block in the design of communication systems, aiming to approach Shannon's capacity limit [157]. The modulation format sets the upper bound of the spectral efficiency for a given signal. According to Shannon–Hartley theorem, modulation formats with higher dimensionality have higher attainable SNR efficiency, which measures the power efficiency at a certain symbol rate. In terms of channel utilisation efficiency, higher dimensional modulation formats consume less power when the transmitted signals occupy a certain bandwidth. In practice, these dimensions are interpreted by the degrees of freedom (DoF) of the transmission channels. Typical two-dimensional modulation formats employ the two quadratures of the waveform, termed the real and imaginary components of complex signals. Other features of waves that have been adopted to multiple DoF are the polarisation state of electromagnetic carrier waves and the spatial and temporal dimensions in the signalling frames. The polarisation state and the spatial dimensions are commonly used in optical fibre communications due to the physical properties of optical fibres. Implementing the temporal dimension to increase the dimensionality of signal formats relies on progressing several adjacent symbols in time as an entity, similar to applying error control coding to singular modulation formats when the codes are binary. In wireless communications, more attention has been put into using temporal DoF to obtain multidimensional modulation formats for signalling. Some multidimensional signal designs are implemented and verified to achieve higher channel capacity and power efficiency, as expected to follow the theoretical results from [157].

In this chapter, we provide an overview of multidimensional modulation techniques, mainly in wireless communications, with specific attention to one special design of a four-dimensional modulation format. Different from most existing multidimensional modulation designs that expand dimensions in the time domain, such a four-dimensional modulation format processes the multiple time symbols jointly, thereby providing non-uniform constellations and requiring a design for bit-to-symbol mapping. In digital signal processing, Gray mapping offers the optimal bit labelling regarding its superior bit error performance by allocating the neighbouring constellation symbols to differ in only one bit. However, the number of neighbouring symbols for a particular symbol in the multidimensional space varies for different designs. Therefore, Gray mapping can not be directly applied to multi-dimensional modulation symbols. Furthermore, to the best of our knowledge, there has yet to be an existing study on optimising the bit-to-symbol mapping for multi-dimensional modulation formats.

Thus, this chapter contributes to filling this gap by investigating the mapping optimisation for multidimensional modulation formats. To this end, the work in this thesis focuses on the bit-to-symbol mapping problem for multidimensional modulations, specifically using the example of a four-dimensional signal. The cost function of the mapping problem is designed to suit higher-dimensional schemes, allowing us to present a generalised optimisation method. The bit-to-symbol mapping optimisation problem is formulated as a typical quadratic assignment problem (QAP), known to be an NP-hard non-convex problem to compare various optimisation methods. A brief literature review introduces the difficulty and complexity of such problem and its potential solutions. In particular, four commonly-used meta-heuristic optimisation methods are detailed, based on which algorithms are developed tailored to the specific bit-to-symbol mapping problem. These algorithms are carried out in computer simulations and are shown to provide mapping solutions with improved error performance. Beyond evaluating the BER performance of the optimised mapping results, the computational complexity and convergence pattern of each algorithms are compared. The results provide valuable insights for future algorithm selection for this study. The multidimensional signal format used in the studies in this chapter is detailed in the author's patent [1].

The outline of this chapter is: section 5.1 provides an introduction to multidimensional modulations and summarises the current techniques with a focus on their applications and advantages. Section 5.2 introduces the basic concept of the four-dimensional modulation format, 4D-256-QAM, followed by a description of the system model for multidimensional modulation system and performance evaluation for the 4D signal to address the need to design and optimising its bit-to-symbol mapping. To explore the optimisation problem, section 5.3 commences with a brief discussion of optimisation on signal design, followed by the general concepts of convex and non-convex optimisation and their common solutions, particularly metaheuristic optimisation algorithms. In section 5.4, four metaheuristic algorithms, including RTS, BSA, GA and SA, are employed for the bit-to-symbol mapping optimisation problem so as to improve the performance of multidimensional constellations potentially. The problem is formulated, and several different metaheuristic optimisation methods are introduced and derived based on the specific problem and performed with numerical simulations. A comparison study of different optimisation algorithms is provided and the performance of the optimal mapping is evaluated in terms of the error performance for the optimised mappings, the computational complexity of the algorithms and their convergence speed. Finally, section 5.5 draws the conclusions.

5.1 Multidimensional Modulation Preliminaries

In the choices of modulation format, there is always the inherent trade-off between the spectral efficiency, the noise tolerance (i.e., the error performance or the power efficiency) and the complexity of the modems. 1-dimensional (1D) modulation for-

[1] E. Sasaki, X. Liu, I. Darwazeh, and N. Zein. "Signal modulation apparatus and signal modulation method." U.S. Patent US 2022/0417075 A1.

mats such as PAM and ASK encode the message information by using one property of digital signals. The most-commonly used 2-dimensional (2D) modulation format, QAM, utilises two signal properties (i.e., amplitude and phase), where the two orthogonal in-phase and quadrature components are considered. Another typical 2D modulation scheme, Voronoi constellations [158], provide the attractive characteristic of minimising the SER at high SNR range. In such modulation format, the constellation symbols are located at a hexagonal grid [17], which reduces the average symbol energy by replacing the square boundary with a circular one. In [159], A. Gersho first proposed the term multidimensional constellation since the information data are mapped to more than two dimensions. Additional dimensions on top of the legacy modulation scheme can be achieved by using other signal resources, such as successive time slots [159] and a pair of polarisation of the optical signals [160].

The design of the multidimensional constellations can be generalised as a sphere packing problem, which is well-studied in mathematical literature [161]. G. D. Forney addressed the problem focusing on finding the densest packing algorithm to place a set of M points in an n-dimensional space (nD, $n \geq 2$), where the minimum Euclidean distance d_{min} between each possible symbol pair are of a fixed value [14]. In the same paper, a lattice-based constellation design was proposed as a typical solution to the packing problem by choosing a subset of an infinite nD lattice Λ to maximise the minimum Euclidean distance d_{min}. Similar to the 2D hexagonal constellations, lattice constructions are regular periodic structures shaped by spherical boundaries rather than cubic ones. This design places the constellation point over the grids and thus leading to simple implementation. The concept of lattice-based constellation design was developed considering different channel conditions, where signals are contaminated by Gaussian or non-Gaussian noises [162] or transmitted over Rayleigh fading channels [163]. The performance of nD modulation formats is measured by the gain regarding the baseline modulation format at the same spectral efficiency η.

Multidimensional modulation formats have been widely used due to their high

SNR efficiency, compatibility with coding and modulation schemes and low PAPR. The following section provides an overview of the current multidimensional modulation techniques, highlighting their applications, main advantages and limitations.

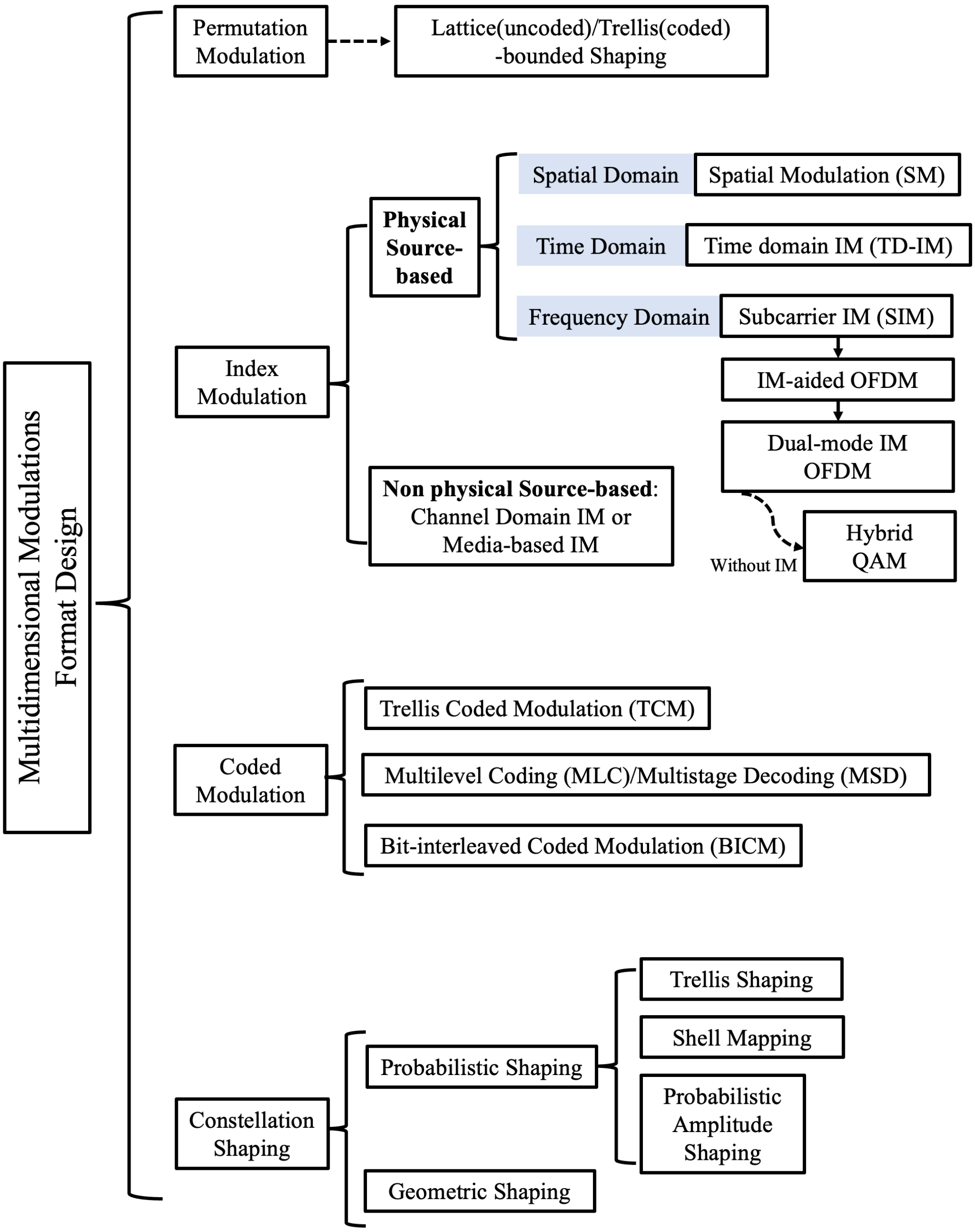


Figure 5.1: Summary of multidimensional modulation techniques

5.1.1 Current Techniques, Applications and Advantages

Figure 5.1 shows a summary of different techniques for multidimensional modulation format design, including the earliest concept of multidimensional modulation, permutation modulation, the paradigm shift to both index modulation and coded modulation, as well as the constellation shaping that have been discussed in Chapter 4.

Permutation Modulation

Permutation modulation, proposed by Slepian in 1965 [164], provides a coding format in that code words are specially selected from all distinct permutations of a prescribed sequence with and without sign changes. Each code word can be seen as a point in an n-dimensional Euclidean space, thus forming the code dictionary of M particular messages, i.e. the M code points, in the n-dimensional space. One attractive feature of permutation modulation is that the decoding over the AWGN channel is relatively simple by employing the maximum likelihood algorithm. Nevertheless, the permutation modulation offers good codes only for low dimensions. Moving on to a generalised concept, Slepian developed the concept of group codes, which other researchers studied as concluded in [165]. Nevertheless, the same drawback remains: good codes only occur in low dimensions.

Index Modulation

Applying a similar idea of permutation modulation to the mapping in space domain constructs GSM in the context of coherent MIMO. In particular, spatial modulation (SM) allows the information signal to be transmitted by a selected antenna from multiple antennas. The resource selection is generalised to the technique known as IM, where a fraction of indexed resources are activated for data transmission. These resources include physical resources such as frequency carriers/bands, time slots and antennas, or non-physical resources such as codes, channel states and signal constellations. In index modulation, additive bits (i.e. indices) are used to allocate the selected resource and to convey information implicitly so that no redundancy will be introduced under such schemes.

Due to its flexible structure, achievable spectral efficiency and low computation complexity, index modulation attracts increasing research interests in communications. Applications can be found in communication systems such as millimeter-wave (mmWave) [166], massive MIMO [167] and VLC [168]. The frequency domain IM are mostly investigated in OFDM systems, where the orthogonal subcarriers are activated to carry the data bits modulated by classic modulation schemes such as QAM, phase-shift keying (PSK) and PAM. In some system designs, the IM-aided OFDM gains higher energy efficiency [169] and even achieves enhanced spectral efficiency. Moreover, [170] provides a detailed discussion of the applications and use scenarios of IM-aided OFDM with a focus on the system models and simulation results. dual-mode OFDM (DM-OFDM) combined with IM techniques were proposed in [171], showing significant performance gains in terms of the enhanced attainable throughput and BER performance. More specifically, OFDM subcarriers are partitioned into two groups and modulated by two mapping schemes in [172]. Furthermore, the dual-mode techniques were incorporated into the time domain the IM FTN signalling [173], which results in a reduced ISI compared to conventional FTN due to the partially deactivated symbols. In general, the applications of IM technique in the spatial domain are extensively researched [174]. Spatial domain IM has shown its significant advantages over conventional MIMO schemes; except for the BER performance and energy efficiency, it achieves mitigated ICI and the synchronisation issues.

Coded Modulation

For a reliable transmission over a channel, the maximum achievable data rate, i.e. channel capacity, can be given by the well-known results stemming from Shannon's work [157, 175]. One possible solution is to increase the modulation order M ($log_2 M = k$ bits/symbol) to increase the data rate for a band-limited channel. High-order modulated symbols are prone to severe BER distortion and therefore, the method of combining the coding with high-order modulation was proposed. The entity of coding and modulation, called coded modulation, was first researched in [176] by Massey in 1974. Ungerbock proposed Trellis coded modulation (TCM)

[177] by concatenating a Trellis code, a type of convolutional code, and a constellation mapping through set-partitioning. Besides, a pragmatic TCM approach [178] allows the Viterbi algorithm to be used for the decoding, which is practical but not as effective as the set-partitioning approach. Imai and Hirakawa's multilevel coding (MLC) is another attractive technique, proposed in 1977 in [179]. Multiple levels of encoders are employed to encode a portion of the input bits each and the bit-to-symbol mapping is typically performed by set-partitioning to improve the coded modulation (CM) performance [180]. The decoding can employ SIC, also known as multistage decoding (MSD), or operates in parallel, which is more efficient at the penalty of suboptimality. Bit-interleaved coded-modulation (BICM) was introduced by Zehavi in 1992, which inserts an interleaving permutation between the coding and the modulation stage [181]. Appealing to its flexibility due to bit-level operations, BICM has attracted lots of investigations over the past few decades and is employed in communication standards such WLAN, digital video broadcasting-satellite-2nd generation (DVB-S2), etc. Good summaries and detailed work on BICM can be found in [182, 183].

Hybrid QAM

A recent technique called Hybrid QAM also introduces extra flexibility for achieving arbitrary spectral efficiency by concatenating QAM schemes of different orders. In the research on hybrid QAM, by far, two regular M-QAM schemes are assigned to different time slots within the time division multiplexing (TDM) frame. Therefore, the target spectral efficiency can be designed by allocating the time slots to the two constitutive QAM symbols in a certain ratio. This time domain hybrid QAM exhibits advantages such as low PAPR and robustness against phase noise due to its shorter symbol period. Some recent work also studied time-domain hybrid QAM with probabilistic shaping and has shown good system performance, as well as the capability of rate-adaptive transmission [184].

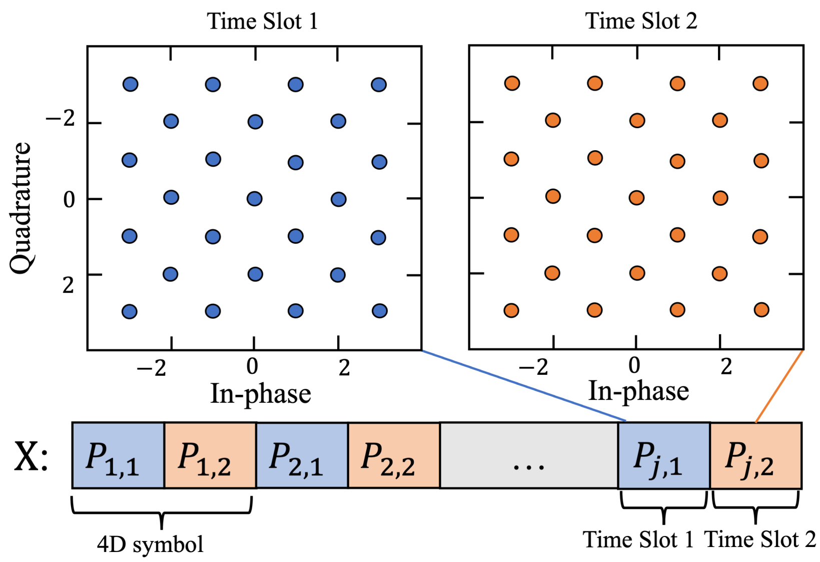


Figure 5.2: Constellation diagram of four-dimensional modulated symbol, which consists of two constitutive two-dimensional symbols allocated in successive time slots

5.2 Four-Dimensional Modulation Signal Design

5.2.1 4D-256-QAM Basic Concept

The four-dimensional modulation scheme employed here is formed by using a pair of 2D complex symbols, which are transmitted in two successive signalling intervals, as shown in Fig. 5.2. In a 4D-256-QAM scheme, each 4D symbol encodes 8 bits. Hence, it achieves 8 bits per 4D symbol, which is equivalent to 4 bits per 2D symbol of 2D-16-QAM. The I-Q coordinates of the 4D signal can be given by a 4D vector $\boldsymbol{x_j} = [x_{j,1}, x_{j,2}, x_{j,3}, x_{j,4}] \in \mathbb{Z}^4$, wherein the coordinates are either all even or all odd integers. Fig. 5.2 shows that the 2D projection diagram of 4D-256-QAM signals is identical for the two signalling time slots, consisting of 25 constellation points.

The total 256 4D symbol points are generated by permutation and manipulation of a set of coordinates, termed lead points, following the system of [159]. Table 5.1 lists the coordinates of the seven lead points and the cardinality of their associated subsets. The 4D symbols drawn from the sublattice have 8 to 23 nearest neighbours, rendering an average number of neighbours $\bar{N}_n = 14.25$, which is smaller than the 24 possible nearest neighbours in this sublattice. The reduced num-

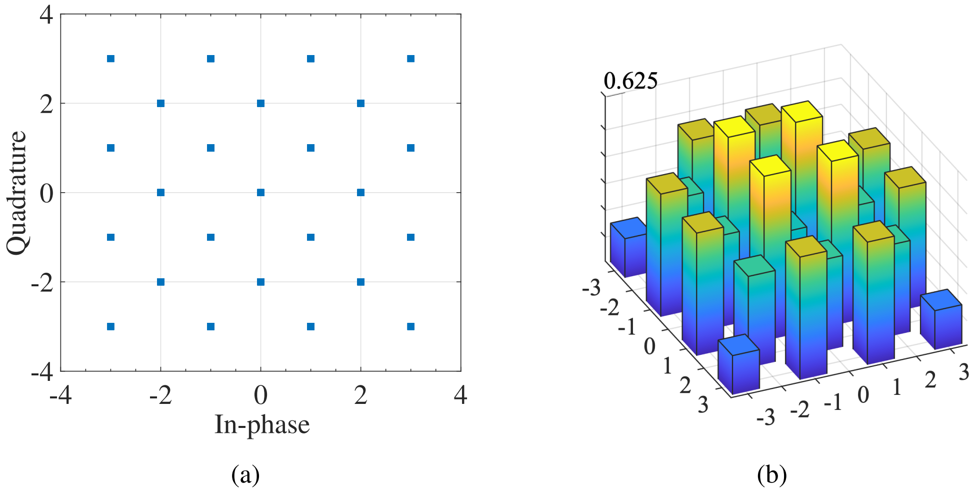


Figure 5.3: (a) Constellation diagram of the projection of the 4D-256-QAM symbol on 2D plane (b) occurrence probability distribution among 4D-256-QAM constellations

ber of neighbours of 4D-256-QAM leads to a reduced upper bound of the SER. The 4D-256-QAM design in [159] is based on a special bit-to-symbol mapping rule as well as an associated detection method.

Table 5.1: Seven Lead Points for 4D-256-QAM Constellations (from [159])

Coordinates	Quantity	Power	No.of Neighbours
1 1 1 1	16	4	23
2 0 0 0	8	4	22
2 2 0 0	24	8	22
2 2 2 0	32	12	19
2 2 2 2	16	16	15
3 1 1 1	64	12	15
3 3 1 1	96	20	8

5.2.2 Multidimensional Modulation System Model

To evaluate the performance of the 4D-256-QAM, a multidimensional modulation system is modelled. The block diagram that depicts the general multidimensional modulation system architecture is given in Fig. 5.4. For proof of concept, a system that adopts single carrier modulation with AWGN as the only channel impairment is considered. At the transmitter side, a stream of bits **b** from the source are mapped by the high-dimensional mapper based on the previously designed look-up table (LUT). Each $\log_2 M$ bits are mapped to an **n**-dimensional symbol. For instance, $M = 256$ for 4D-256-QAM and $M = 16$ for 2D-16-QAM. A shaping pulse is applied to single-carrier modulated symbols **s** before the AWGN is added. At the receiver, the received noisy symbols **y** are firstly demodulated and then demapped, also based on the LUT, to obtain $\hat{\mathbf{b}}$.

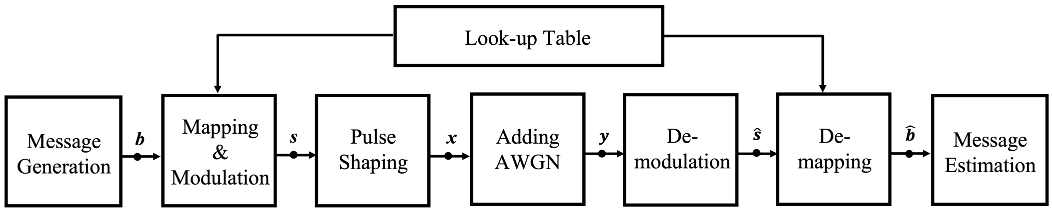


Figure 5.4: 4D-256-QAM system block diagram

5.2.3 Performance Metrics for Four-Dimensional Modulation

Fig. 5.5 shows BER and SER versus SNR of 4D-256-QAM in relation to the BER of ideal 2D-256-QAM. Both signals are generated and evaluated by the system given in Fig. 5.4. The figure confirms the performance improvement of 4D-256-QAM in a high SNR regime. More specifically, the 4D-256-QAM outperforms 2D-16-QAM when the SNR exceeds 16.4 dB or the BER is lower than 10^{-3}. For a bit error rate of 10^{-6}, the 4D modulation format gains approximately 0.6 dB when compared to the 2D regular format. It is worth noting that the 2D and 4D modulation formats have the same bits-per-symbol. Hence the corresponding baud rate and the bandwidth of the spectrum are identical. This indicates that the performance is improved by merely increasing the signal dimensionality with no complexity penalty.

Another advantage of 4D-256-QAM over 2D-16-QAM is the power efficiency. Herein, we evaluate such efficiency quantitatively by the signal feature PAPR. The calculation of PAPR is detailed in Appendix C. The PAPR of 4D-256-QAM is investigated and evaluated with numerical simulations. Fig. 5.6 depicts the peak and average power of the two signal formats. The commonly used root-Nyquist pulse, SRRC with roll-off factor $\rho = 0.1$ and filter length L = 20, is employed for the pulse shaping. It can be seen that both the peak and average power of 2D-16-QAM

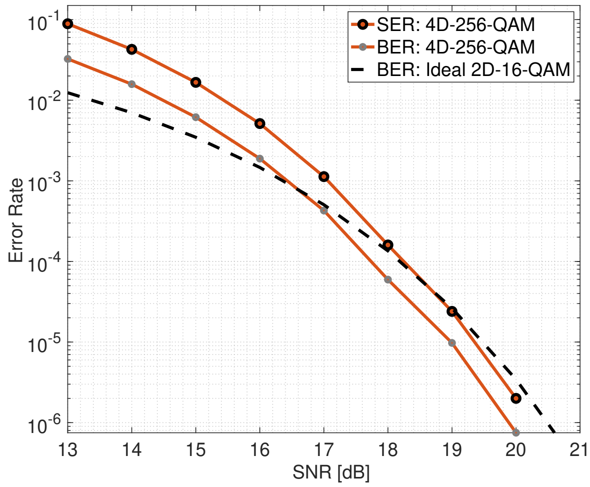
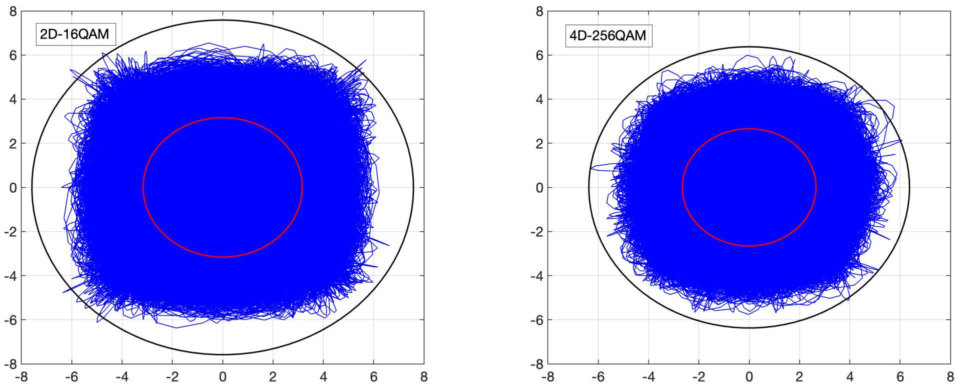


Figure 5.5: BER and SER performance for 4D-256-QAM



(a) 2D-16-QAM

(b) 4D-256-QAM

Figure 5.6: PAPR performance comparison: peak and average power illustrations

are larger than those of 4D-256-QAM. From Fig. 5.3 (b), it can be seen that the occurrence probability of each point in the constituent 2D constellations is non-equiprobable - the symbols with higher power occur more often than the ones with lower power. Due to this predetermined signalling mechanism, similar to probabilistic shaping as discussed in Chapter 4, the average and peak power of the 4D signal are lower.

A more precise method to evaluate the PAPR performance is to use CCDF to characterise the signal power distribution for a certain signalling duration. The CCDF of PAPR calculates the probability that the PAPR of the signal exceeds a

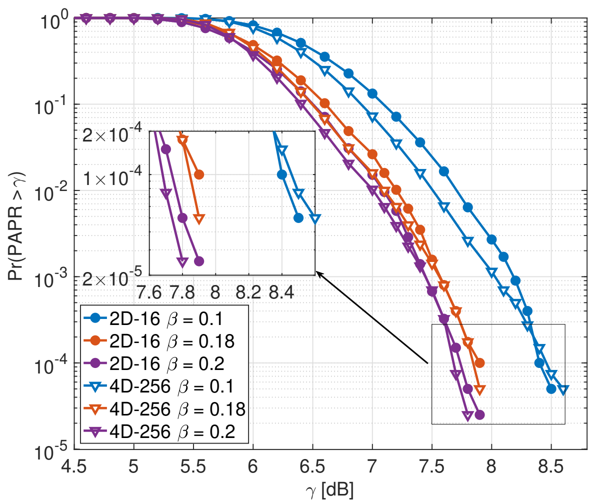


Figure 5.7: CCDF of the PAPR for 2D-16QAM and 4D-256QAM after filtering by SRRC pulse of roll-off factor $\alpha = 0.1, 0.18, 0.2$.

certain value γ that $PAPR > \gamma$. Fig 5.7 compares the CCDF of the PAPR for 2D-16-QAM and 4D-256-QAM. Firstly, the figure indicates that the probability of PAPR surpassing a certain value γ increases with the increase of the roll-off factor β for the shaping pulse for both 2D-16-QAM and 4D-256-QAM. Moreover, 4D-256-QAM shows slightly better PAPR performance than 2D-16-QAM, which is more obvious when β is smaller.

5.2.4 Bit-to-symbol Mapping Design and Optimisation

For multidimensional modulation formats, bit-to-symbol mapping involves assigning the bit vectors to the symbol set. The ultimate goal of the mapping optimisation is to minimise the BER of a given symbol set, i.e. the constellation, by finding the optimal bit-to-symbol mapping rule.

The mapping problem can be formulated as a typical quadratic assignment problem QAP, which addresses the case that assigns $\boldsymbol{N}$ facilities to $\boldsymbol{N}$ locations with a minimum cost. To simplify this process, the bit vectors are converted to decimal numbers, termed symbol indices. The bit-to-symbol mapping optimisation aims to minimise the BER for a given constellation symbol set. Search algorithms are used to find the optimal solution, which assigns the bit vectors to the symbols with minimum cost. An N-dimensional constellation contains M symbols. Each symbol encodes k bits, $k = \log_2 M$. Taking 4D-256-QAM mapping optimisation, for example, $N = 4$, $M = 256$ and each 4D QAM symbol encodes $k = 8$ bits.

The symbol set $\mathscr{S}$ contains 256 4D symbols. The ith symbol from the set $\mathscr{S}$ is denoted by $\mathbf{s}_i = [s_{i,1}, s_{i,2}, s_{i,3}, s_{i,4}] \in \mathbb{R}^4$ with $i = 1, 2, ..., M$. Therefore, the symbol set can be expressed as $\mathscr{S} = [\mathbf{s_1}^T, \mathbf{s_2}^T, ..., \mathbf{s_M}^T]^T$.

The bit vector set $\mathscr{B}$ contains 256 8-bit vectors corresponding to the 256 4D symbols, thus indicating the mapping. The bit vectors can be seen as a binary set $\mathbf{b}_i = [b_{i,1}, b_{i,2}, ..., b_{i,8}] \in \{0,1\}^8$ with $i = 1, 2, ..., M$. Herein, $\mathbf{b}_i, \mathbf{b}_j \in \mathscr{B}$ are the corresponding 8-bit vectors assigned to $\mathbf{s}_i, \mathbf{s}_j$ based on the mapping.

5.2.4.1 Cost Function Design

The ultimate goal of mapping optimisation is to minimise the BER of the multidimensional symbol. Hence, the cost function is derived with respect to the upper bound of the BER of a given symbol. The analytical upper bound of the SER is expressed as:

$$P_{\text{symbol error}} \leq Q(\sqrt{\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{4\sigma^2/N}}), \tag{5.1}$$

wherein $Q(\cdot)$ is the Gaussian tail function, σ^2 denotes the variance of the white Gaussian noise $\mathscr{N}(0, \sigma^2)$.

To convert the SER upper bound to that of BER, in this work, the cost of bit-to-symbol mapping optimisation accounts for both the Euclidean distance between an arbitrary pair of symbols and the Hamming distance between their corresponding bit vectors. The Euclidean distance between two arbitrary 4D symbol vectors $\mathbf{s}_i, \mathbf{s}_j \in \mathscr{S}$ can be given by:

$$\begin{aligned} d(\mathbf{s}_i, \mathbf{s}_j) &= \|\mathbf{s}_i - \mathbf{s}_j\| \\ &= \sqrt{\sum_{n=1}^{N} (s_{i,n} - s_{j,n})^2}, \end{aligned} \tag{5.2}$$

The Hamming distance $h(\mathbf{b}_i, \mathbf{b}_j)$ can be calculated by executing the bit-wise operation between $\mathbf{b}_i$ and $\mathbf{b}_j$. More specifically, the number of different bits is measured by counting the number of '1' in the output of XOR of $\mathbf{b}_i$ and $\mathbf{b}_j$. Combining the Hamming distance $h(\mathbf{b}_i, \mathbf{b}_j)$ and the Euclidean distance $\|\mathbf{s}_i - \mathbf{s}_j\|$, the cost function can be defined as the averaged double summation function **z**, given by:

$$\mathbf{z} = \frac{1}{M}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} h(\mathbf{b}_i,\mathbf{b}_j)Q\left(\sqrt{\frac{\|\mathbf{s}_i-\mathbf{s}_j\|^2}{4\sigma^2/N}}\right). \tag{5.3}$$

Based on this cost function, the optimisation problem is formulated to minimise **z** as shown in the next section.

5.2.4.2 Optimisation Problem Formulation

With the designed cost function, the bit-to-symbol mapping optimisation problem for multidimensional modulation symbols can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{b},\mathbf{s}} \quad & \mathbf{z} = \frac{1}{M}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} h(\mathbf{b}_i,\mathbf{b}_j)Q\left(\sqrt{\frac{\|\mathbf{s}_i-\mathbf{s}_j\|^2}{4\sigma^2/N}}\right), \\ \text{s.t.} \quad & \mathbf{b}_i,\mathbf{b}_j \in \mathscr{B}, \quad \forall i,j \\ & \mathbf{s}_i,\mathbf{s}_j \in \mathscr{S}, \quad \forall i,j \end{aligned} \tag{5.4}$$

QAP problem is a non-convex optimisation problem proven NP-hard. This indicates the optimisation problem is difficult to have solutions that will always efficiently produce the expected outcome. In the next section, the mapping optimisation problem is explored with a focus on potential solutions. Metaheuristic algorithms are proposed and developed for the problem addressed in equation (5.4).

5.3 Nonconvex Optimisation and Metaheuristic Algorithms

5.3.1 Applications of Optimisation on Signal Design

Optimisation problems can be formulated as finding a solution that minimises the criterion (i.e. the cost) among a number of candidate solutions. In signal design, optimisation techniques are applied to different aspects of the signal to meet the need of improving the signalling performance against the imperfect channel impairments, so as to reduce operational costs of its communications. Such aspects can include all the blocks in the communication system for signal processing from the modulation design in terms of the modulation parameters (amplitude, phase and frequency) to code design, filter design, power allocation, received antenna selection in MIMO system and performing demodulation and detection. A typical example is the design of FIR filter coefficients via linear programming (LP) [185]. In the design applications, the optimisations are offline. Once the optimisations are performed, the optimised results will be utilised in the applications.

All the candidate solutions constitute a search space, where a search algorithm searches in such space to find the optimal solution that reaches the minimum. A search space can be subdivided into several *basins of attraction* such that each contains an isolated minimum. Each of these minima, termed as local optimum, has no better solutions surrounding them in the sub-space. Among them, the smallest one, i.e. the global optimum, is the true minimum that a solution can reach in the optimisation problem. For a deterministic local search algorithm, the starting point is predetermined. In this case, the search is most likely to reach a local optimum, depending on the subspace that the starting point is located at. Therefore, certain mechanisms are required to prevent the searching process from ”cycling” or being trapped to local optima. A common solution is to introduce stochasticity to the search process and this is used in some optimisation algorithms as will be discussed in the following sections.

5.3.2 Convex and Non-convex Optimisation and Their Solutions

Optimisation problems can be split into two categories, convex and non-convex optimisation, depending on the mathematical features of the objective function and the constraints of the problems. In the first, convex optimisation refers to a broad class of optimisation problems such as least-square LP, quadratic programming (QP) and semi-definite programming (SDP)[186]. It is proved to be efficiently solved in theory whilst problems of practical interests have been shown to have good convex formulations or good convex formulations. Compared to convex optimisation, non-convex optimisation is seen to be less tractable and usually lack of generalised theoretical properties though, lots of practical problems with applications in signal processing and machine learning are unable to be modelled in convex formulation. Thus, non-convex optimisation has been researched extensively with the focus on the two main solutions; convex relaxation reformulates the non-convex problem as a convex problem so that optimisation can be performed efficiently; metaheuristic algorithms, which will be addressed and discussed in the next section, provide good enough solutions with reasonable efficiency at cost of optimality and accuracy.

5.3.3 Metaheuristic Optimisation Algorithms

From a perspective of the complexity of problems, as discussed in the previous sections, most of the real-world optimisation problems can not be solved in polynomial-time, termed as NP-hard. Moreover, some require exponential time and yet global optimum is not guaranteed. Hence, *exact methods*, such as tree-based algorithms, guarantee the optimality for problems with small search space though, they are not used to solve complex optimisation problems. By contrast, *heuristic methods* can provide sufficiently good solutions in a reasonable time for large-size problems. Among them, **metaheuristic algorithms**, relative to problem-specific heuristic algorithms, are general methodologies that the heuristics are guided and can be applied to most problems, finding solutions with acceptable performance. There are investigations on different optimisation techniques proposed to multidimensional constellation design for maximising the minimum Euclidean distance under certain constraints. Algorithms such as quadratic programming [187], sequential quadratic programming [188] and SA technique [189] have been used to optimise the constellation for multidimensional signals and shown considerable gains in different system scenarios. In this study, SA and other three commonly-used metaheuristic optimisation algorithms are investigated.

Binary Switching Algorithm

BSA is a local search algorithm initially proposed in [190] to perform index assignment for vector quantisation. It is the most well-known optimisation method for mapping, first used in [191] for BICM. BSA conducts iterative switching of symbol indices to optimise the index labelling for a given constellation symbol set. These

Algorithm 1: Binary Switching Algorithm

Input: n-dimensional constellation set **S** (M-by-N)
Output: optimised bit to symbol mapping set **Z** for **S**

Initialise the mapping set **B**;
for *k = 1 to number of iterations* **do**
    Calculate the individual cost;
    Sort the mapping set in a decreasing order of individual cost, also store the indices in set **B**;
    interationflag = 0 $\mathbf{Z} = \mathbf{B}$;
    for *i = 1 to M* **do**
        for *j = 1 to M* **do**
            if $j \neq i$ **then**
                switch index j and i;
                calculate the *Cost* D' for the new mapping set;
                if $D' < D$ **then**
                    update the current lowest cost $D = D'$;
                    update the mapping set **Z**;
                    interationflag = 1
                end
                if *interationflag* $== 1\&\ m == M$ **then**
                    $\mathbf{B} = \mathbf{Z}$;
                    Break
                end
            end
        end
    end
end

symbol indices are decimal numbers that were originally converted from bit vectors. Therefore, the optimisation can be simplified by processing the assignment of the index set and the symbol set. In BSA, there are the *individual cost* and the *total cost*; the former is based on the Euclidean distance between an individual symbol and all the other symbols and associated Hamming distance; the *individual cost* of all symbols adds up to obtain the *total cost*. BSA is a trajectory-based algorithm where the current search is always based on the outcome of the last iteration. BSA starts with a random assignment of the bit-to-symbol mapping. The *total cost* is calculated and stored as the best cost so far. The symbols are sorted by their *individual cost* from the highest to the lowest. The symbol with the highest *individual cost* swaps its symbol index with the second-highest-cost symbol and then swaps with the rest of the symbols in descending order. The *total cost* is calculated after each swap. If *total cost* is lower than the best cost thus far, the best cost is updated by the better *total cost* and the current iteration ends. Otherwise, the pair of indices are swapped back and the next index swap is performed until a better *total cost* relative to the best cost occurs. It is worth noting that once the best cost is updated, the permuted set after the index swap is maintained, or in other words, the reverse swap is not performed. The new mapping with the best cost so far becomes the current solution, and the next iteration commences. The index switching repeats for another iteration until no better *total cost* can be found. This is considered optimum.

Genetic Algorithm

GA is one of the typical population-based evolutionary algorithms, which explore multiple candidate solutions during the search and are guided by the heuristic rule that mimics the Darwinian theory of natural selection and heredity. It is commonly used in search problems by relying on mutation, crossover and selection operations. The GA was proposed by J.H. Holland in 1992 [192]. Due to its strong potential for obtaining global optimum, this method has been extensively investigated, developed and practically applied. A detailed review can be found in [193].

Different from the single solution-based BSA, GA simultaneously searches in multiple directions and hence maintains the diversity in the population (i,e, the candidate solutions) so as to approach to a global optimum. This is achieved by the three biologically-inspired operations - selection, crossover and mutation, altogether termed as breeding. In each cycle of breeding, two parent solutions are selected via a selection routine to perform crossover to yield the offspring. The decision of replacing the parent solutions with the offspring is made according to the cost value, herein termed as *fitness*, of the parents and the offspring. Once the offspring has worse *fitness* value than the parent solutions, the mutation operator is applied to produce new solutions. As the GA progresses through generations, the population is updated by repeating the three operations until the convergence condition is met.

Amongst the metaheuristic algorithms, GA is found to be one of the most

Algorithm 2: Genetic Algorithm

Input: n-dimensional Constellation set **S** (M-by-N)
Output: optimised bit to symbol mapping set **Z** for **S**

Initialise the populations of N random mappings $\mathbb{B}_0$;
Calculate the fitness of each $\mathbf{B}_0$ in $\mathbb{B}_0$;
Sort $\mathbb{B}_0$ in descending order depending on the fitness;
for *s = 0 to MaxGeneration-1* **do**
  for *r = 1 to total number of breeding* **do**
    Randomly select two parent solutions μ_j and μ_k, $\mu_j, \mu_k \in \mathbb{B}_s$;
    Select the crossover point λ;
    Perform partially mapped crossover on μ_j and μ_k to generate the offspring $\tilde{\mu}_j$ and $\tilde{\mu}_k$;
    if $FITNESS(\tilde{\mu}_l) < FITNESS(\mu_j)$, $l \in j,k$ **then**
      $\mu_l \leftarrow \tilde{\mu}_l$
    else
      Update the mutation rate ρ;
      Generate L_ρ mutants from $\tilde{\mu}_l$;
      for $\tilde{\mu}$ *in* L_ρ *mutants* **do**
        if $FITNESS(\tilde{\mu}) < FITNESS(\mu_l)$ **then**
          $\mu_l \leftarrow \tilde{\mu}$
        end
      end
    end
    Add μ_l to the new population $\mathbb{B}_s$
  end
end
for $\mu \in \mathbb{B}_{MaxGeneration}$ **do**
  Find out the μ with the best fitness value;
  $Z \leftarrow \mu$
end

robust methods for QAP [194]. Attributed to its effectiveness in solving complex nonlinear optimisation problems [195], GA is applied to optimise the constellation [196] and mapping design [197] for amplitude phase shift keying (APSK) with the investigation of different selection and crossover operators. Moreover, GA is implemented to minimise the error floor by optimising the constellation for BICM system in [198]. For the same QAP problem, GA has also been exploited for the optimisation for uncoded space-time labelling diversity (USTLD) mapper for space-time block coded (STBC) systems [199].

Simulated Annealing

Another typical metaheuristic algorithm, SA, is an iterative probabilistic algorithm that aims to approximate global optimisation in a large search space. Referring to the gradual temperature variation in the physical process of annealing, the SA algorithm integrates a *cooling schedule* into the heuristic search rule for the optimisation. The heuristic rule determines the transition probabilities between each current solution and neighbouring solutions. More specifically, the transition probability (or acceptance probability) is a function of the key variable *temperature*, the current and the neighbour solution, indicating the probability of accepting worse solutions. By introducing such perturbation, the search can avoid getting stuck in local optima in the solution space. The *temperature* is updated at each step as a fraction of the previous value, which is determined by the cooling rate $\alpha \in (0,1)$.

The optimisation method was first proposed in [200] based on the Metropolis algorithm and has been researched its applications to physical systems to solve high dimensional combinatorial optimisation problems [201]. A detailed description of the simulated annealing algorithm and its general use for optimisation can be found in [202]. In essence, the applications of SA for optimising signal design for communication systems have been investigated in terms of source code design [203], index assignment [204], constellation design and bit labelling [205, 206]. In this work, the modified SA algorithm is presented for the case of minimising the BER of multidimensional modulation format.

Tabu Search and Reactive Tabu Search

Tabu search is a guided local search method. The basic idea of tabu search is to forbid moves that take the search into previously visited sub-spaces. A tabu list, a short-term memory structure, stores these visited moves so that future searches can check and avoid them. However, for a complex optimisation problem with a large search space, the required tabu list is also very large, and consequently, it requires large memory for storage. To tackle this problem, one way is to specify a fixed size for the tabu list. Thus, a visited move can be removed from the list and available after a specific interval.

Algorithm 3: Simulated Annealing

Input: n-dimensional Constellation set $\mathbf{S}$ (M-by-N)
Output: optimised bit to symbol mapping set $\mathbf{Z}$ for $\mathbf{S}$

Initialise the mapping $\mathbf{B_0}$;
$\mathbf{B_{best}} \leftarrow \mathbf{B_0}$;
Set the initial temperature $\mathbf{T} \leftarrow \mathbf{T_0}$;
while *Termination condition is not met* **do**
    Generate neighbouring solutions of $\mathbf{B_{best}}$;
    $bestSolution \leftarrow \mathbf{B_{best}}$;
    for *i = 1 to MaxIteration* **do**
        for *a random solution* $\mathbf{B}$ *in* $\mathbf{B_{best}}$*'s Neighbourhood* **do**
            Calculate $\Delta = COST(\mathbf{B}) - COST(bestSolution)$;
            if $\Delta < 0$ **then**
                $bestSolution \leftarrow \mathbf{B}$
            else
                Generate a random number $r \in [0,1]$;
                if $exp(-\Delta/\mathbf{T}) > r$ **then**
                    $bestSolution \leftarrow \mathbf{B}$
                end
            end
        end
    end
    $\mathbf{B_{best}} \leftarrow bestSolution$;
    $\mathbf{T_{n+1}} \leftarrow \alpha \mathbf{T_n}$;
end
Return $\mathbf{Z} \leftarrow \mathbf{B_{best}}$

Another method is to adopt a varying tabu list size as in the RTS algorithm, which was proposed by Battiti in 1994. In RTS, an extra level of stochasticity is added to make the optimisation more robust towards finding the global optimum. RTS employs a tabu list that records the recency and the frequency of the occurrence of visited moves. RTS permits the revisit of some moves. The revisit interval of the same move, equivalent to the tabu list size, is adapted to the iteration number of the configuration automatically. Therefore, RTS does not require a predetermined tabu list size, which is initially set to a random value and will converge to its optima after a few iterations of the searching. RTS has the advantage of efficient and robust convergence in large-size optimisation problems. Similar to GA, RTS was adopted for optimising labelling diversity mapper for BICM system [207], showing significant

Algorithm 4: Reactive Tabu Search

```
Input: n-dimensional Constellation set S (M-by-N)
Output: optimised bit to symbol mapping set Z for S
Initialise the mapping set B_0;
B_best ← B_0;
Initialise the tabu list Φ;
Initialise the tabu list size l_Φ = SIZE(Φ);
for k = 1 to MaxIteration do
    Generate neighbouring solutions of B_best;
    Update tabu list size l_Φ by reaction function;
    for solution B in B_best's Neighbourhood do
        bestSolution ← B_best;
        if B NOT in tabu list Φ AND COST(B)¡COST(bestSolution) then
            bestSolution ← B
        end
    end
    if COST(bestSolution) < COST(B_best) then
        B_best ← bestSolution
    end
    Update Φ by adding bestSolution;
    if SIZE(Φ) > l_Φ then
        Remove the first element in Φ
    end
end
Return Z ← B_best
```

improvement in the asymptotic coding gain for 16-QAM and 64-QAM.

The pseudocode of the above algorithms provided have been developed in this work for optimising the bit-to-symbol mapping for multidimensional constellations. The implementation of four algorithms and, consequently, the system model that exploits the optimised mappings is provided in detail later in this chapter. The following section first gives the mathematical analysis of the mapping optimisation problem for multidimensional modulation. Then it presents the system design with a performance evaluation of the aforementioned optimisation methods.

5.4 Optimisation for Multidimensional Symbol Mapping

The bit-to-symbol mapping optimisation problem for the 4D-256-QAM is presented in section 5.2.4. The performance of mapping optimisation based on the algorithms discussed above is investigated by simulations. The experiments use an 8-core Intel Xeon CPU E5-2650 v2 @2.60GHz with 50.5GB RAM. For all four methods, the algorithms are first implemented. Once the cost converges and the optimisation operation terminates, the optimised mappings are collected and evaluated. More specifically, the optimised mapping tables are used as the LUT in the system model in Fig. 5.4.

5.4.1 BER Performance for Optimised Mappings

As the bit-to-symbol mapping only affects bit-level error performance, it is expected that with the same SER, the BER for the 4D signal with different mappings can be lower-bounded by $SER/8$. Fig. 5.8 shows the BER performance for 4D-256-QAM with the optimised mappings generated by the RTS, BSA, GA and SA. The pseudocodes and the parameters of these four algorithms are provided in section 5.3.3. The convergence costs are all minimal and close, reducing from 0.12 to 0.08036 for RTS, 0.08244 for BSA, 0.0884 for GA and 0.08324 for SA as shown in Fig. 5.9. The BER versus SNR results align with the value of the associated cost value as expected: the BER performance improves as the converged cost decreases. There are only slight differences between the four converged cost values, which is in a high SNR regime as shown by the inset plot in Fig. 5.8 depicting similar error performance at BER = 10^{-2}. A 0.5 dB SNR advantage can be achieved for all signals with optimised mappings. While in the low SNR regime, the differences are distinguished more clearly, where the 4D signal with mapping optimised by RTS has a 2.4 dB power advantage relative to the mapping without optimisation. Whereas the BER for mapping generated by GA only achieves half the power advantage when reaching the same BER. In general, the results given by Fig. 5.8 have proved the efficacy of the optimisation algorithms and the cost function design.

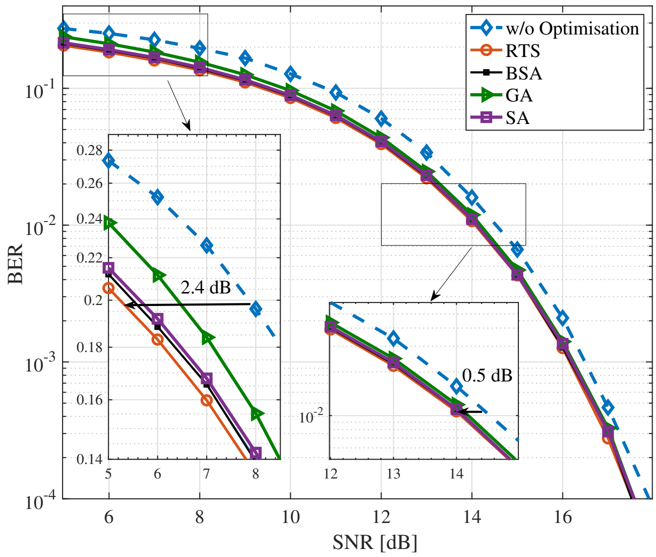


Figure 5.8: BER performance for 4D-256-QAM with bit-to-symbol mappings optimised by RTS, BSA, GA and SA.

5.4.2 Computational Complexity of the Algorithms

The Big-$\mathcal{O}$ notation is used to assess the complexity of optimisation algorithms. Usually, worst-case-based complexity is considered rather than the average performance and therefore, an asymptotic bound is adopted. For example, given a bit-to-symbol mapping problem that assigns n bit vectors to n symbols, the problem is size n. The calculation of the cost for each solution is carried out by using matrix multiplication to reduce the algorithm complexity. The main computation in all four algorithms is the generation of the neighbourhood or the set of neighbouring solutions.

In BSA and SA, swap operations are performed in each iteration during the optimisation to generate the neighbour solutions from the current solution. Considering the worst case, the complexity function $f(n)$ can be given by:

$$f(n) = \frac{1}{2}n(n-1). \tag{5.5}$$

In RTS method, three types of operations, swap, reversion and insertion, are used to generate the neighbourhood of the current solution. Therefore, the complexity function can be given by:

$$f(n) = \frac{1}{2}n(n-1) + \frac{1}{2}(n-2)(n-3) + (n-1)(n-2), \tag{5.6}$$

where the three components correspond to the number of operations for swap, reversion and insertion, respectively. The function $f(n)$ is upper bounded $g(n)$ such that $f(n) \leq c \cdot g(n)$ with positive constant c. The complexity can be given by $\mathcal{O}(g(n))$. Hence, both algorithms have the polynomial complexity of $\mathcal{O}(n^2)$. For example,

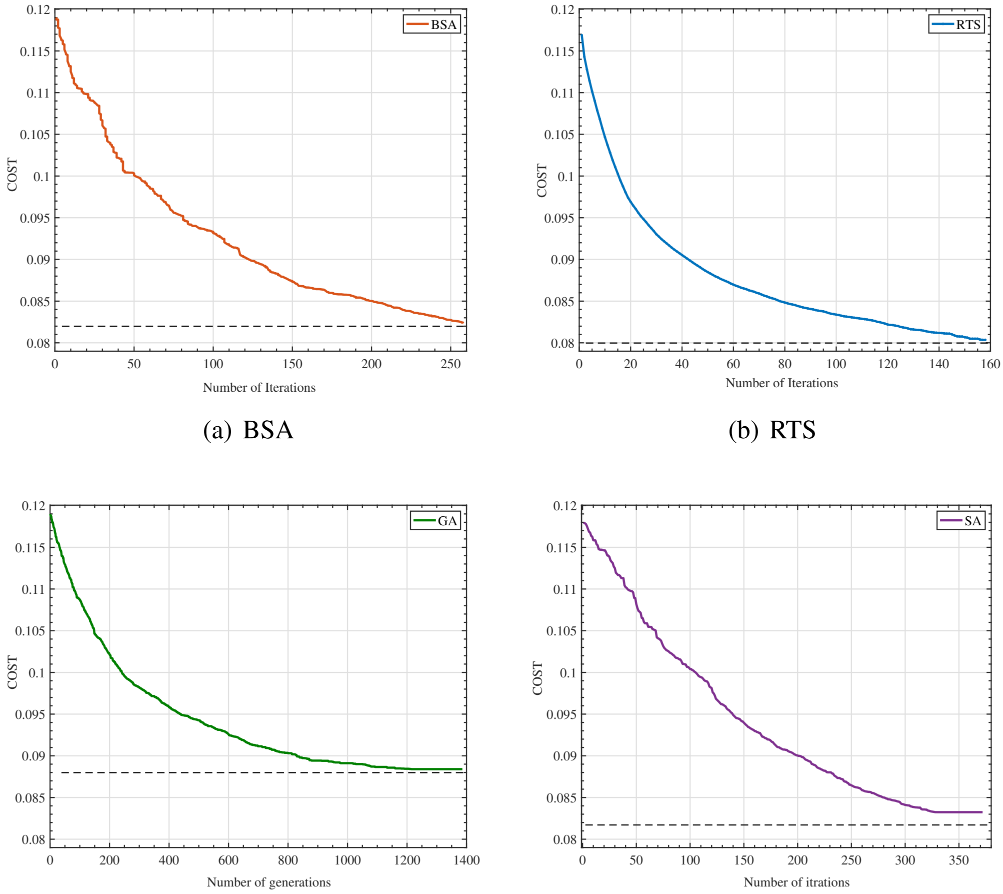
- for $n = 2^8 = 256$, $f(n) \leq c \cdot 2^{16}$,

- for $n = 2^{12} = 4096$, $f(n) \leq c \cdot 2^{24}$.

In GA, instead of generating neighbouring solutions by any of the three aforementioned operations, child solutions are generated by crossover and mutations. The equivalent number of swaps cannot be calculated for the partial crossover used in the GA algorithm. This is because the number of operations fully depends on the inherent structure of the parent solutions. For the mutation procedure, the mutation rate affects the probability of mutation. By considering the worst case, the complexity function can be given by

$$f(n) = generation \cdot population \cdot (breeding + n), \tag{5.7}$$

wherein $generation$, $population$ and $breeding$ are the number of the generation, population size and the number of breeding in GA. It is clear that GA has different algorithmic bases, making it difficult to compare fairly with other algorithms. The complexity of GA relies heavily on its parameters, which makes it challenging to compare with other algorithms. On the other hand, BSA has a higher computational complexity than SA. This is because BSA evaluates the cost of all neighbour solutions to select the best one to replace the current solution, while SA only requires the best solution found so far. Therefore, SA requires fewer calculations than BSA.



(a) BSA (b) RTS

(c) GA (d) SA

Figure 5.9: Converge speed comparison between RTS, BSA, GA and SA for bit-to-symbol mapping optimisation.

5.4.3 Convergence Speed

The numerical simulations of the algorithms are carried out in MATLAB. Since the aforementioned four algorithms are trajectory-based methods, parallel computing can not be used to reduce the computation time. The convergence curve for the four algorithms is shown in Fig. 5.9 in terms of the *total cost* versus the number of iterations or generations. Due to the non-convex nature of the mapping optimisation problem, the global optimum can not be guaranteed by using metaheuristic algorithms, which is discussed in 5.3.3. Therefore, the achieved experimental results can only provide general performance trends and cannot be reproduced unless the experiments are initialised with the same starting points and parameter settings.

Since even under the same design of the metaheuristic optimisation method, the alternative decisions regarding each aspect of a particular algorithm, such as the starting points, the neighbourhood search sequence and particularly the key parameter values, strongly impact the outcome of the optimisation. This means some implementations of the same algorithm may perform better than others. Hence, this study analyses the performance of the four algorithms with respect to the overall pattern of the convergence curve.

It can be seen from Fig. 5.9 that for BSA, RTS and SA, the cost decreases rapidly at the beginning of the optimisation and then slow down as the curve becomes less steep. Moreover, it has been observed that a sharp decay in the cost for the first 20 iterations in RTS is due to the selection of the best neighbour solution, resulting in the update of the lowest-cost solution. This can lead to the conclusion that RTS is the most greedy method and SA is less greedy. The cost for GA exhibits a similar pattern, i.e., a rapid decrease in the first 200 generations followed by a plateau after the first 400 generations. The hill climbing property is evident in all four metaheuristic algorithms.

5.5 Conclusions and Discussions

This chapter investigates another spectrally and energy-efficient technique-multidimensional modulation techniques. An overview of multidimensional modulation techniques has been presented, with specific attention to a four-dimensional modulation format, 4D-256-QAM. Unlike existing multidimensional modulation designs that expand dimensions in the time domain, this four-dimensional modulation format processes multiple time symbols jointly. One of the main challenges of multidimensional modulation is the design of its bit-to-symbol mapping. Due to the irregular constellation format of multidimensional signals, the grey coding cannot be applied directly to offer optimal mapping rule. Therefore, the work in this chapter focuses on such optimisation problem and take the 4D-256-QAM as an example to prototype the methods for optimising the mapping for multidimensional modulations. Investigations have been performed on the mapping optimisation for multidimensional modulation formats and presented a generalised optimisation method for the four-dimensional signal with the design of a cost function that suits higher-dimensional schemes.

Within the exploration of optimization challenges, the thesis engaged with foundational concepts of convex and non-convex optimization, dedicating attention to metaheuristic optimization algorithms. The implementation of four distinct metaheuristic algorithms—RTS, BSA, GA, and SA—was introduced to tackle the bit-to-symbol mapping optimization issue, aiming to bolster the performance of multidimensional constellations. Through comprehensive formulation and execution, these algorithms were evaluated through numerical simulations.

The outcomes of the comparative study involving various optimization algorithms have unequivocally demonstrated that optimal mapping holds the potential to enhance error performance and convergence speed of multidimensional constellations. A critical assessment of computational complexities showcased the efficiency and efficacy of the proposed method, underscoring its suitability for prototype implementations.

Overall, this chapter has contributed to the overarching objective of the thesis by not only investigating multidimensional modulation techniques but also by providing algorithms of optimizing bit-to-symbol mapping for such schemes. This contribution stands as a testament to the potential for achieving heightened performance and efficiency in communication systems through innovative optimization methodologies.

Chapter 6

The Application of Machine Learning for Multidimensional Signal Design

AI and ML techniques have recently aroused growing research interest in their applications in physical layer communications [208, 209]. The applications of AI/ML have been studied extensively and gained great success in domains such as computer vision (CV), natural language processing (NLP) and automatic speech recognition (ASR), where rigid or tractable mathematical models are usually unavailable. Unlike the aforementioned fields, traditional wireless communication systems have block-based architectures with concrete mathematical models for each block. The problem arises when the practical systems and environments are too complicated to be approximated using such models. For instance, in current 5G networks, scenarios for wireless communications are addressed, i.e. the eMBB, URLLC, mMTC, and for each scenario, there is a variety of models for different use cases. While using ML, the conventional models can be optimised for sophisticated configurations. In the envisioned 6G, the wireless networks will be migrated from over the top towards the AI era. Distributed AI/ML solutions are expected to support emerging scenarios and applications [26]. The optimisations by using ML in communications can be categorised into two approaches–optimising a single module in the chained block structure or optimising multiple modules jointly in the systems to improve the end-to-end performance. For the former, modules with isolated functions can be optimised individually, such as source coding [210], channel coding and decoding [211], waveform design [212], signal detection and decoding, channel equalisation [213] and estimation [214]. As for end-to-end optimisation, ML functions as a universal approximator [215] to interpret conventional communication systems without rigid analytical models [208]. Yet the lack of flexibility and capability for diverse scenarios renders ML-based models inapt for practical implementation. Comprehensive overviews of the application of ML in physical layer communications can be found in [216, 217]. This work mainly summarises the emerging theoretical and practical studies, showing their promising performance improvements as well as the constraints and future challenges of the technologies.

Previously, the advantages of constellation shaping and multidimensional modulations have been shown in Chapter 4 and in Chapter 5, respectively. To enable constellation shaping for multidimensional modulation signal design, in this chapter, a particular neural network model, autoencoder, is proposed with the use of machine learning techniques to improve the signal performance and optimise the designs. The four-dimensional modulation signal with optimised mapping presented in the last chapter is used as the case study of the application of machine learning for multidimensional constellation shaping design.

The remaining part of the chapter is organised as follows: the basic concepts of deep learning (DL), neural networks, and the specific model *autoencoder* that is used in the later sections are provided in section 6.1. Section 6.2 presents the multidimensional probabilistic shaping system, highlighting how the autoencoder model is designed to optimise the overall signal performance. The figure of merits for performance evaluation are explained in section 6.3.1. The simulation study and a deeper analysis of the performance assessment are provided in section 6.4.2, with a focus on performance improvements of the probabilistically shaped multidimensional signal obtained by end-to-end learning of the autoencoder-based model. Finally, section 6.5 wraps up the study and draws a conclusion. Part of the work in this chapter includes results given in the author's 2022 conference paper [31] [1]. The multidimensional signal format used in the study is detailed in the author's patent [2].

6.1 Preliminaries: Neural Networks, Deep Learning and Autoencoder

Modern ML techniques have demonstrated excellent superiorities for diverse applications, especially in physical layer wireless communications. Leveraging efficient utilisations of large datasets, ML techniques can solve real-world problems that are hard or impossible to be tackled by conventional explicit methods. Regarding different ways that data are used, ML is commonly categorised into three types of learning: supervised, unsupervised and reinforcement learning. Various ML algorithms have been widely used and extensively studied over the past decades, such as linear models, support vector machines, k-means clustering and expectation-maximisation [218]. Among such learning algorithms, neural networks re-emerge with tremendous success due to the development of computational resources since the 1980s [219]. Moreover, in recent years, DL, a class of ML technique based on neural networks, has shown great potential for extracting high-level information from raw data. This work focuses on the use of neural networks as well as deep learning techniques in physical layer communication system design. To ensure a good understanding of the proposed signal design in this chapter, a brief preliminary of the basic concepts of neural networks, deep learning and autoencoder is provided in this section.

6.1.1 Neural Networks

An artificial neural network is a type of computing system model for ML that mimics the biological neuron models in human brains to approximate some underlying relationship between the data. Similar to biological brains, a neural network consists of interconnected group of artificial neurons that follows a mathematical model

[1] X. Liu, I. Darwazeh, N. Zein and E. Sasaki, "Probabilistic Shaping for Multidimensional Signals with Autoencoder-based End-to-end Learning," 2022 IEEE WCNC, 2022.

[2] E. Sasaki, X. Liu, I. Darwazeh, and N. Zein. "Signal modulation apparatus and signal modulation method." U.S. Patent US 2022/0417075 A1.

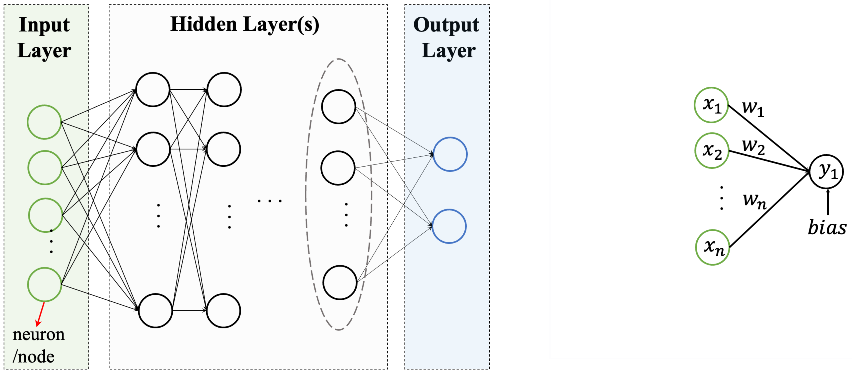
to perform computation in a connectionistic way. Typical neural networks' architecture include multiple successive layers, the input layer, the output layer and several hidden layers that are set in-between, as depicted in Fig. 6.1(a). Each layer contains a certain number of neurons/nodes, which is the basic unit of neural networks. In a dense neural network, the nodes in each layer are fully connected to nodes in the adjacent layers. Taking a single neuron in Fig. 6.1(b) for an example, the inter-connected neuron process the data by first taking all the inputs $\mathbf{x} = \{x_1,\ x_2 ... x_n\}$, assigning each input a weight $\mathbf{w} = \{w_1,\ w_2 ... w_n\}$, adding up the weighted inputs and the bias of the layer b, and passing the summation to the activation function $\sigma(\cdot)$ so as to generate the output of such neuron as follow:

$$y_1 = \sigma(\sum_{i=1}^{n} w_i x_i + b). \tag{6.1}$$

Each layer of a neural network is independent. Thus, the number of nodes per hidden layer and the activation function $\sigma(\cdot)$ for each layer can be different. The activation functions introduces nonlinearity to the neural networks, allowing the model to have the ability to approximate functions [220]. Table 6.1 provides a list of commonly used activation functions, among which the rectified linear unit (ReLU) and Softmax function are used in this work and will be elaborated with more details in section 6.4.1. With the introduced nonlinearity by the activation functions, neural networks can benefit from stacking the hidden layers on top of each other. As such, the learning depth is enlarged by having multiple hidden layers. A neural network is considered deep when it has more then two hidden layers and consequently the ML procedure, deep learning.

6.1.2 Deep Learning

The core of the learning procedure is optimisation [219, Chapter 8]. More specifically, the goal of learning (training) is to find the optimum parameter vector for a given neural network structure so as to achieve a desired function. With the layered architecture, neural networks can be established by connecting the input layer with the hidden layers to provide a computational model, of which a mathematical



(a) (b)

Figure 6.1: (a) A dense feedforward neural network architecture and (b) a single neuron

Table 6.1: List of activation functions

Name	Activation function $\sigma(u)$	Range
Linear	u	$(-\infty,\ \infty)$
tanh	$tanh(u)$	$(0,\ 1)$
ReLU	$max(0,u)$	$[0,\ \infty)$
Sigmoid	$\frac{1}{1+e^{-u}}$	$(0,\ 1)$
Softmax	$\frac{e^{u}}{\sum_j e^{u_j}}$	$(0,\ 1)$

representation can be given by:

$$\mathbf{y} = f(\mathbf{x}, \theta) \tag{6.2}$$

where $\mathbf{x}$, $\mathbf{y}$ indicate the node vector of the input layer and output layer, respectively, θ represents the learning parameter vector and $f(\cdot, \theta) : \mathbb{R}^0 \rightarrow \mathbb{R}^L$ describes the mapping of the neural network given θ with $L-1$ hidden layer. Referring to equation (6.1), the processing of lth layer of a dense neural network can be represented by:

$$f_l(\mathbf{x}_{l-1}, \theta_l) = \sigma(\sum \mathbf{w}_l \mathbf{x}_{l-1} + \mathbf{b}_l), \quad l = 1, ..., L-1 \tag{6.3}$$

where the set of parameter for this lth layer is $\theta_l = \mathbf{w}_l,\ \mathbf{b}_l$. In deep learning, $\theta = \theta_1, ..., \theta_{L-1}$ is adjusted such that the neural network can accurately map the input vector to a desired output. The loss function is established by measuring the difference between the neural network output and the target one. The loss function serves real-time evaluation of the performance of neural networks and subsequently guides the learning process. The design of the loss function is dependent on the task that the neural network is trained to perform, which is further discussed in section 6.3.

The process of minimising loss function for neural networks adopts iterative methods. The gradient descent method and its variants, such as the most used Stochastic gradient descent (SGD) [221, 222], mini-batch gradient descent and Nesterov accelerated gradient descent (momentum) [223] are simple and efficient to implement. Other optimisation algorithms have been investigated and used for neural networks. For example, AdaGrad adjusts the learning rate based on historical gradients. AdaDelta and RMSProp further improved AdaGrad. Moreover, adaptive moment estimation (Adam) combines the adaptive learning rate and momentum, known as the most commonly used optimiser for machine learning [224]. For different optimisation methods for neural networks, [225] provides a detailed state-of-art. One of the primary challenges of optimisation problems is to avoid the local optima and, therefore, to reach the global optimum. The theoretical limitations of optimisation for learning resulted from the inherent structure of neural networks per se, specified in [219]. In the convex optimisation field, the loss function needs to be designed to ensure the optimisation problem is convex. Thus, the solution can guarantee a global optimum since any local optimum is a global optimum for convex problems. However, the optimisation of neural networks is a non-convex problem. More specifically, a considerable amount of local optima with equivalent losses can exist. The theoretical limitation of local minima remains an open problem for deep learning; nevertheless, the concern has little bearing in practice [226].

6.1.3 Autoencoder

As emerged in the previous discussion, the special feedforward neural network *autoencoder* attempts to efficiently reduce the dimensionality by learning the optimum representation of the input information [227]. The simplest autoencoder structure

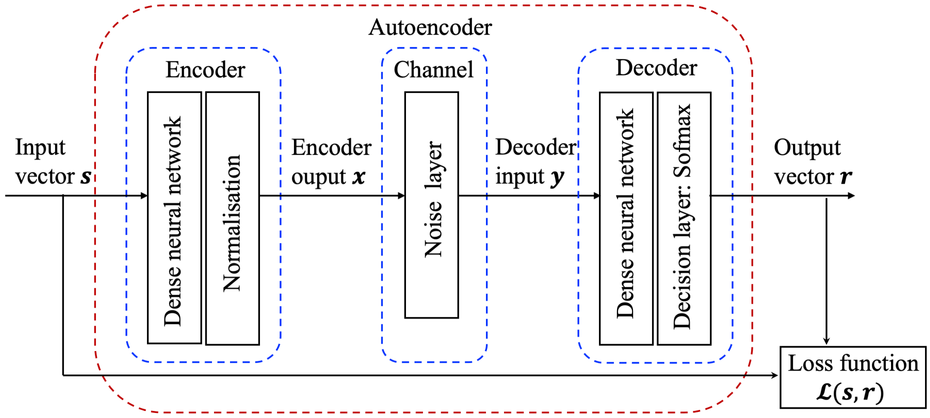


Figure 6.2: Basic autoencoder architecture

is sequentially composed of two independent neural networks, an encoder and a decoder. The encoder maps the input data to lower dimensional data and the decoder functions to reproduce the input from the encoded data. The unsupervised learning task of the autoencoder is performed to the encoder and the decoder simultaneously. In order for the encoder to learn the efficient representation, the input data are used as targets to be compared with the replicated output data from the decoder. The low dimensional data representation is usually used for feature extraction or message compressing [228] to reduce the data complexity and to improve data abstraction efficacy and classification accuracy.

Autoencoder has been introduced to optimise physical layer communication system design by inserting a channel model as a layer between the encoder and the decoder, termed latent space. In section 6.2.1, the applications of autoencoder in communication systems to optimise the system performance are reviewed. This is followed by how an autoencoder-based model is designed to learn the optimum multidimensional probabilistic shaping in the presence of the AWGN channel impairment.

6.2 Multidimensional Probabilistic Shaping

As discussed in Chapter 4, the non-uniform signalling technique, probabilistic shaping, has the advantage of achieving higher capacity while requiring lower power consumption and enabling flexible data rate adaption. Recalling the basic idea of probabilistic shaping, we have discussed that the technique allows data streams to be encoded so that symbols are transmitted with unequal probabilities. Therefore, the achievable data rate can be maximised by optimising the occurrence probabilities of such symbols. To this end, conventional probabilistic shaping is implemented by coding schemes [17] such that the input bit stream can be encoded to symbols with a target probability distribution. In section 4.2, DM structure, the reverse concatenation architecture of PAS, was discussed and the efficient implementation of CCDM for probabilistic shaping was highlighted. DM-based probabilistic shaping has been intensively investigated, with many applications emerging for various communication systems. Typical designs of probabilistic shaping schemes are mainly 1D, which can be directly applied to singular dimensional modulation formats such as PAM [132]. To construct n-dimensional (nD, $n \geq 2$) modulations with probabilistic shaping, the n-fold Cartesian product of the 1D probability distribution is used [229]. Particularly, probabilistically shaped 2D QAM can be achieved by applying 1D probabilistic shaping for the **I** and **Q** components.

However, the Cartesian product is not feasible to shape non-regular 2D constellations such as 32-QAM and 128-QAM. To tackle this problem, [230] proposed a universal probabilistic shaping scheme based on a 2D distribution matcher, which easily exploits shaping to 2D symbols. For multidimensional signals, additional dimensions for nD symbols $(n > 2)$, apart from the **I** and **Q** dimensions, can be achieved by employing successive time slots [159] or a pair of polarisation of optical signals [160]. While there have been numerous studies on geometrical shaping for multidimensional modulations [21] [231], only a few papers exist on multidimensional probabilistic shaping, with such shaping itself being applied to only 2D signals in all existing literature [232–234].

As presented in the previous section, deep learning technology has been applied to communication systems by approximating and optimising the transmitter and receiver functions using neural networks [208]. In particular, A very recent work in [235] proposes a special end-to-end learning approach employing the autoencoder to shape and optimise the probability distribution for 2D constellation symbols. In the autoencoder, two neural networks, i.e., the encoder and decoder, are jointly trained to minimise the error rate of the reconstructed symbols at the receiver. Furthermore, by adding the channel effects (i.e., channel-embedded autoencoder), the learned constellation shapes will be robust against channel impairments. More precisely, the encoder finds the optimal signal representation for a given channel and then the decoder recovers the input from the contaminated data. The learning procedure of autoencoder for 2D probabilistic shaping is leveraged on the Gumbel-Softmax trick [236], which provides a continuous and therefore differentiable approximation of the originally discrete probability distribution for the training [235].

The following section presents the nD probabilistic shaping optimisation design employing the autoencoder-based model. The implementation of end-to-end learning is introduced to the autoencoder to optimise the occurrence probability distribution of the nD signals.

6.2.1 Autoencoder-based System Architecture

A typical communication system can be interpreted as an autoencoder, in which the encoder and the decoder are jointly optimised to perform the same tasks as traditional transceivers. Fig. 6.2 shows the general architecture of the autoencoder combined with the structure of communication systems. In general, an autoencoder aims to generate optimal representations of the input data, pass these through the channel and reconstruct the initial data from the contaminated channel outputs. In the case of nD probabilistic shaping, the autoencoder-based system is designed to optimise the probability distribution of the nD constellation symbols by training in an end-to-end manner.

As shown in Fig. 6.4, the encoder outputs the nD symbols $\mathbf{x} \in X = \mathbb{C}^M$ with an occurrence probability $p_X(x)$. In the encoder, an implicit selection procedure is conducted to apply the symbol probability distribution to the finite set X of M constellation points in nD space. The Gumbel-Softmax trick [236] is used as the sampling mechanism to approximate the discrete probability distribution $p_X(x)$ using the continuous, differentiable softmax function. An $\boldsymbol{n}$-dimensional approximation

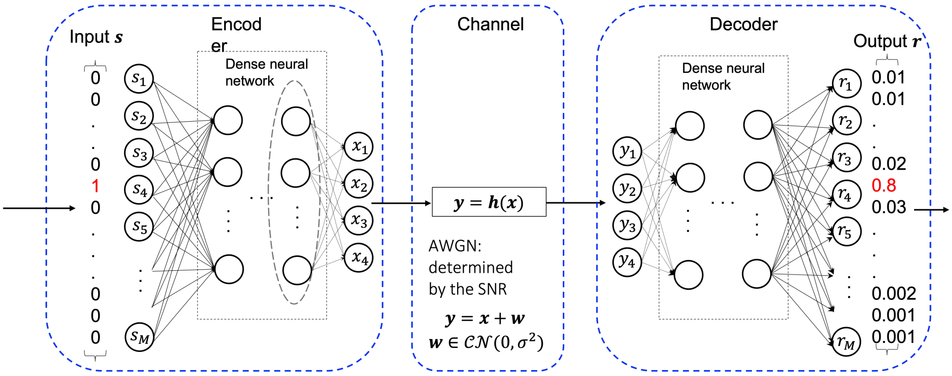


Figure 6.3: An example for autoencoder model of using one-hot vector as inputs

vector is employed in this work.

Consequently, an $\boldsymbol{n}$-dimensional approximation vector $\tilde{\mathbf{x}}$ is given by

$$\tilde{x}_i = \frac{e^{(\log(p_X(i)+g_i)/\tau)}}{\sum_{j=1}^{X} e^{(\log(p_X(j)+g_j)/\tau)}}, i = 1, \ldots, M \tag{6.4}$$

where g_i are independent and identically distributed (i.i.d.) samples drawn from standard Gumbel distribution. The softmax temperature τ determines the distribution sparsity such that the Gumbel-Softmax distribution converges to a uniform distribution as τ increases and the vector $\tilde{\mathbf{x}}$ becomes one-hot when $\tau = 0$.

Our optimisation method is set to find the optimum probability distribution of the transmitted symbols. This is based on training the autoencoder at different received SNR values and then choosing the optimum probability distribution with the best error performance.

To generate the discrete probability distribution $p_X(x)$, a softmax activation function is applied to the encoder neural network outputs (i.e., the unscaled logits $\boldsymbol{s}'$). The aforementioned Gumbel-Softmax trick is applied to obtain the approximation continuum of $p_X(x)$, which is used for end-to-end learning. While the true one-hot representation $\hat{\mathbf{s}} = [0, 0, \ldots, 1, \ldots, 0]$ of the discrete distribution is used to perform the selection among the nD constellation points. An example of autoencoder using one-hot vector is given in Fig. 6.3. More specifically, the normalised constellation $\mathbf{C}^M$ with a unity energy expectation is multiplied by the one-hot representation $\hat{\mathbf{s}}$ to give the output nD symbol, which is passed to the channel.

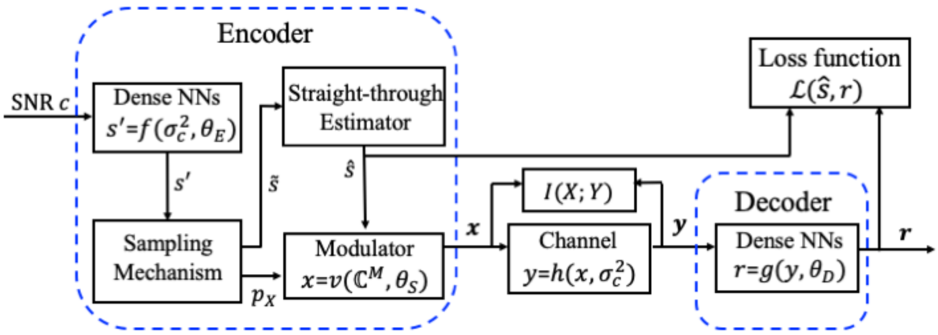


Figure 6.4: Block diagram of autoencoder based probabilistic shaping system

The channel can be described by a set of layers embedded in an autoencoder structure. In this work, AWGN is considered and the Gaussian noise layer performs $\mathbf{y} = \mathbf{x} + \mathbf{w}$, wherein for $w_i \in \mathbf{w}, w_i \sim \mathscr{N}(0, \sigma_c^2)$, $i = 1, ..., n$ and σ_c^2 denotes the noise variance determined by the given SNR c for the training. Subsequently, the noisy symbols $\mathbf{y}$ are sent to the autoencoder decoder, which performs classification to reconstruct the input message. The decoder neural network process the impaired data with several dense layers followed by a softmax layer. The softmax function outputs a probability vector that indicates the predicted probability of each element sent as in the input message.

6.2.2 Optimisation by End-to-end Learning

The end-to-end training of the autoencoder based system employs SGD. To simplify the learning of the autoencoder, a set of transformations can be used to describe the general processes:

$$\begin{aligned}
&\boldsymbol{p_X}, \hat{\boldsymbol{s}} = f(\sigma_c^2, \theta_E), \\
&\boldsymbol{x} = v(\boldsymbol{p_X}, \hat{\boldsymbol{s}}, \theta_S), \\
&\boldsymbol{y} = h(\boldsymbol{x}, \sigma_c^2), \\
&\boldsymbol{r} = g(\boldsymbol{y}, \theta_D),
\end{aligned} \tag{6.5}$$

where $f(\cdot)$ and $v(\cdot)$ construct the encoder neural network, $h(\cdot)$ denotes the channel layer and $g(\cdot)$ denotes the decoder neural network. θ_E and θ_S are the trainable parameters of the neural network for the encoder and θ_D is for the decoder. In particular, θ_S is the parameter for the sampling mechanism. The loss function $\mathscr{L}$ measures how correctly the decoder reproduces the input. The next section discusses the figures of merit for the multidimensional probabilistically shaped signal as well as how the loss function is constructed based on these figures of merit.

6.3 Figures of Merit for Optimisation

The choice of loss function is task specific. Often, the cross entropy loss is chosen for classification problems and the MSE for regression problems. Regarding the optimisation problem for multidimensional probabilistic shaping, two figures of merit are considered– the mutual information and the categorical cross entropy.

6.3.1 AIR and Mutual Information

Achievable information rates (AIR) is an asymptotic metric measuring the number of bits per symbol reliably transmitted through a channel. A typical AIR, the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$, is commonly used to provide predictions of maximum throughput for fair comparisons among different modulation schemes for a given channel. More specifically, the channel capacity can be obtained by maximising the MI over the PDF of the transmitted symbol $\boldsymbol{X}$, i.e. the distribution $P_{\boldsymbol{X}}(x)$, which can be expressed as

$$C = \max_{P_{\boldsymbol{X}}(x)} I(\boldsymbol{X};\boldsymbol{Y}) \tag{6.6}$$

where the MI between the channel input $\boldsymbol{X}$ and output $\boldsymbol{Y}$ is defined as:

$$I(\boldsymbol{X};\boldsymbol{Y}) = \mathbb{E}[\log \frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)}{P_{\boldsymbol{Y}}(y)}], \tag{6.7}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operation, $P_{\boldsymbol{Y}}(y)$ is the marginal entropy of $\boldsymbol{Y}$ and the $P_{\boldsymbol{X},\boldsymbol{Y}}(x,y)$ represents the joint PDF. which can be derived from $P_{\boldsymbol{X},\boldsymbol{Y}}(x,y) =$

$P_{\boldsymbol{X}}(x)P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)$. Hence, the mutual information can be given by:

$$\begin{aligned}
I(\boldsymbol{X};\boldsymbol{Y}) &= \mathbb{E}[\log\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)}{P_{\boldsymbol{Y}}(y)}] \\
&= \sum_{x\in\mathbb{C}^M} P_{\boldsymbol{X}}(x) \sum_{y\in Y} P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)\log\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)}{P_{\boldsymbol{Y}}(y)} \\
&= \sum_{x\in\mathbb{C}^M} P_{\boldsymbol{X}}(x) \sum_{y\in Y} P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)\log\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)}{\sum\limits_{x\in\mathbb{C}^M} P_{\boldsymbol{X}}(x)P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)}
\end{aligned} \tag{6.8}$$

where the transmitted symbol $\boldsymbol{X}$ are drawn from a discrete constellation $\mathbb{C}$ with the cardinality M. The conditional entropy $H(\boldsymbol{Y}|\boldsymbol{X})$ is expressed as:

$$H(\boldsymbol{Y}|\boldsymbol{X}) = -\sum_{x\in\mathbb{C}^M} P_{\boldsymbol{X}}(x) \sum_{y\in Y} P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x) \log(P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x)) \tag{6.9}$$

and the marginal entropy $H(\boldsymbol{Y})$ is given by

$$H(\boldsymbol{Y}) = -\sum_{y\in Y} P_{\boldsymbol{Y}}(y) \log(P_{\boldsymbol{Y}}(y)). \tag{6.10}$$

Besides, the marginal distribution $P_{\boldsymbol{Y}}(y)$ can be related to the integral of the joint PDF with respect to transmitted symbol x, given by:

$$P_{\boldsymbol{Y}}(y) = -\sum_{x\in\mathbb{C}^M} P_{\boldsymbol{X},\boldsymbol{Y}}(x,y) = -\sum_{x\in\mathbb{C}^M} P_{\boldsymbol{X}}(x)P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x). \tag{6.11}$$

The mutual information $I(\mathbf{X};\mathbf{Y})$ can be expressed as a function of entropies, given by

$$I(\mathbf{X};\mathbf{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}). \tag{6.12}$$

The typical 2D AWGN numerical channel model is given by:

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y-x)^2}{2\sigma^2}), \tag{6.13}$$

wherein, the discrete channel is assumed to be memoryless and σ^2 denotes the Gaussian noise variance. Consequently, the maximum entropy can be $H_{\max} =$

$\log_2(\sqrt{2\pi e \sigma})$. Given the $\boldsymbol{n}$-dimensional transmitted symbols $\boldsymbol{X}$ and received symbols $\boldsymbol{Y}$, the channel is characterised by a conditional probability density function (PDF) $P_{\boldsymbol{Y}|\boldsymbol{X}}$. The channel transition distribution of an $\boldsymbol{n}$-D AWGN channel is given by $\boldsymbol{n}$-dimensional Gaussian distribution:

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{d=1}^{n} \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left(\frac{-\sum_{d=1}^{n}(y_d - x_d)^2}{2\sigma^2}\right), \tag{6.14}$$

where σ_d^2 is the Gaussian noise variance for the dth dimension, x_d and y_d represent the dth element of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. For the $\boldsymbol{n}$-D i.i.d Gaussian noise, $\sigma_d^2 = \sigma^2$ is identical for all the dimensions. Therefore, the transition function for the $\boldsymbol{n}$-dimensional AWGN channel can be rewritten as:

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-\|\boldsymbol{y} - \boldsymbol{x}\|^2}{2\sigma^2}\right), \tag{6.15}$$

6.3.2 Categorical Cross Entropy

Categorical cross entropy, in contrast with binary cross entropy, is used for two probability distributions over the same set of multiple variables. The cross entropy is defined as the average number of bits that are needed to encode data with a probability p to the codewords with a probability q. Based on the concept, the categorical cross entropy between p and q is given by:

$$H(p,q) = \mathbb{E}_p[-\log(q)] = -\sum_{x\in\mathscr{X}} p(x)\log(q(x)) \tag{6.16}$$

which can be derived by a summation of the entropy of p and the Kullback-Leibler (KL) divergence $\mathrm{D_{KL}}$ as:

$$H(p,q) = H(p) + \mathrm{D_{KL}}(p\|q), \tag{6.17}$$

where the entropy $H(p) = \mathbb{E}_p[-\log(p)]$ denotes the average amount of bits that are needed to represent the symbols with probability distribution p and $\mathrm{D_{KL}}$ denotes the difference between the two probability distributions.

Neural networks that perform classification tasks usually adopt one-hot vectors for labelling and probability distribution vectors as the classification outcomes. Thus, categorical cross entropy is commonly used as the classification loss function, which assesses the difference between the target labels and the output categories.

Recall the objective of the end-to-end learning with autoencoder for probabilistic shaping in this work; the encoder learns the latent representations robust to the channel impairments; the decoder learns to reconstruct the input symbols. Herein, the representations are the probability distribution of the transmitted symbols, and the decoder performs a classification task to recover the original constellation symbols from the noisy data. Consequently, the two neural networks in the encoder and the decoder blocks are trained jointly to optimise the reconstruction. Therefore, the loss function can be derived from the difference between the transmitted symbol $\boldsymbol{x}$ and the recovered symbol $\boldsymbol{r}$. To this end, we develop the loss function as:

$$\begin{aligned}\mathscr{L}_{\boldsymbol{x},\boldsymbol{r}}(\theta_E,\theta_D) &= \mathbb{E}_{x,y}[-\log(P(r|y,\theta_D)] - H(\boldsymbol{X}) \\ &= \mathbb{E}_{y}[\mathrm{D_{KL}}(P(r|y,\theta_D)\|P(x|y))] - I(\boldsymbol{X},\boldsymbol{Y}),\end{aligned} \tag{6.18}$$

where the $p(r|y,\theta_D)$ is the empirical PDF of r given y, which estimates the true posterior PDF $p(x|y)$, $x,y \sim P_{\boldsymbol{X},\boldsymbol{Y}}(x,y)$ and the entropy of the channel input is subtracted $H(\boldsymbol{X})$ and therefore the loss function only contains the negative mutual information and the difference between the two distributions.

6.4 4D Probabilistic Shaping Optimisation

In Chapter 5, we have discussed and studied the benefits of multidimensional modulation formats, in particular, the four-dimensional signal format 4D-256-QAM [159]. The simulation results show a significant power advantage of the 4D signal when compared to regular 2D QAM signals. In this chapter, we continue the investigation using the 4D-256-QAM signal as an example of multidimensional probabilistic shaping for a few reasons: *i)* first, to verify that multidimensional probabilistic shaping can improve the performance of the non-shaped signal; *ii)* to test the efficacy of the autoencoder-based optimisation. Finally, the proposed model and end-to-end learning method are presented and simulated on the 4D signal as a prototype and can be used for other multidimensional modulation formats.

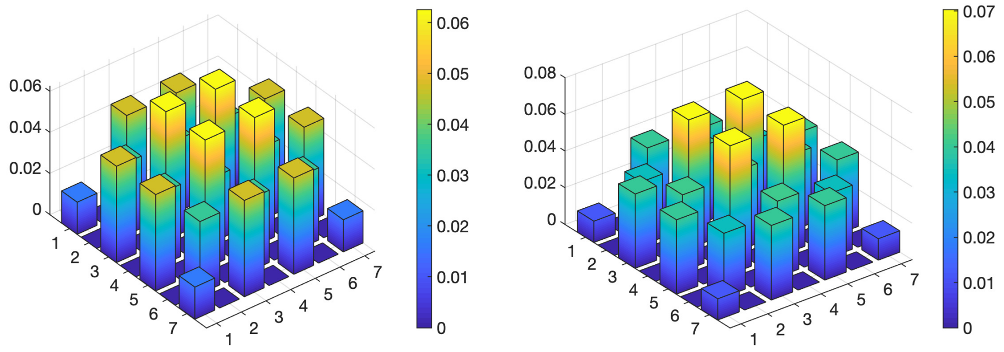
6.4.1 Autoencoder Model Design

This section presents the performance of the proposed autoencoder-based system for multidimensional probabilistic shaping for the 4D-256-QAM signal with end-to-end learning. Tensorflow framework [237] is used to implement the training. With regard to the neural networks in the autoencoder, we employ one dense layer of 128 nodes with ReLU activation function followed by a linear layer in the encoder. For the decoder, three dense layers are used; the first two layers consist of 128 nodes with ReLU activation function; the last employs softmax function to output the probability vector. The structure is given in Table 6.2 We utilise the Adam optimiser [224] for the learning of the autoencoder. In addition, the i.i.d. n-dimensional Gaussian noise channel is employed for the embedded channel in the autoencoder architecture as discussed in the section 6.2.1. Finally, we show the results of the optimised PS-4D-256-QAM in this section, including the probability distribution, the error performance of the signal in a single carrier system, the mutual information and the power efficiency in terms of PAPR.

Table 6.2: Autoencoder structure

	Layer	**Output Vector Length**
Encoder	Input	n,R $\in [0,255]$
	Dense (ReLU)	128
	Dense (Linear)	M
Channel	Sofmax+one-hot transform	M (one-hot vector)
	Constellation point selection	n (complex number)
	Gaussian Noise	n
Decoder	Dense (ReLU)	128
	Dense (ReLU)	M
	Sofmax	M (probability distribution)

End-to-end learning is introduced to the autoencoder to optimise the occurrence probability distribution of the nD signals for different signal-to-noise ratio



(a) 4D-16-QAM (b) PS-4D-256-QAM by AE

Figure 6.5: Probability distribution among symbols

values, ranging from 5 dB to 20 dB. Specifically, we simulate the proposed scheme for PS-4D-256-QAM with an AWGN channel. The optimised autoencoder-based probabilistic shaping scheme is tested. The system performance is evaluated in terms of BER and mutual information.

6.4.2 Analysis of Learned Constellations

Here, the learned constellations with the optimal probabilistic shaping are analysed to show their advantage relative to the 4D-256-QAM and 2D-16-QAM. Numerical results are demonstrated in terms of the probability distribution of the constellations, the improved BER performance and the enhanced mutual information of the optimised PS-4D-256-QAM relative to the 4D-256-QAM and 2D-16-QAM.

6.4.2.1 Probability Distribution

Fig. 6.5 shows the probability distribution among the 2D projected symbols for the original 4D-256-QAM and the learned PS-4D-256-QAM. As discussed in section 5.2, the 4D-256-QAM constellations are chosen subsets of the infinite nD lattice Λ. One of the key properties of such a symbol constellation is its non-uniform occurrence probability distribution as shown in Fig.6.5 (a). In contrast, as depicted by Fig.6.5 (b), the learned constellation has a more centralised probability distribution.

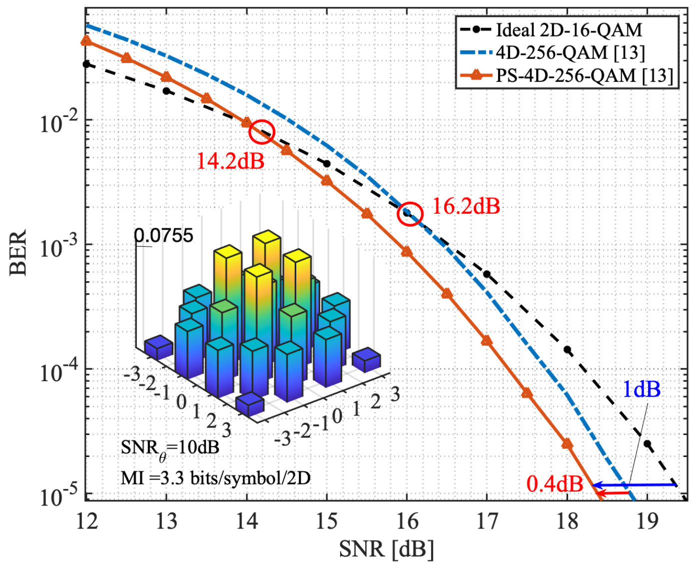


Figure 6.6: BER performance comparison for the 4D signals with/without probabilistic shaping using bit mapping from [159]

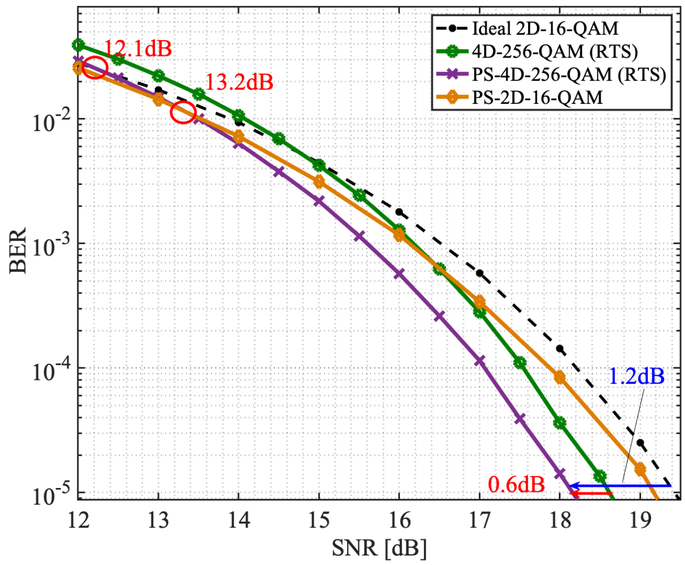


Figure 6.7: BER performance comparison for the 4D and 2D signals with/without probabilistic shaping. The bit labelling of the 4D QAM is obtained by optimisation using the RTS algorithm

6.4.2.2 BER Performance

To verify the designs numerically, the BER performance of 4D-256-QAM is compared with the BER of 2D-16-QAM in the simulations. It is worth noting that the bit-to-symbol mapping from [159] and the optimised mapping, found by the RTS algorithm in this work, are tested for the proposed scheme for the 4D-256-QAM. The probability distribution of the PS-4D-256-QAM constellation was optimised by the end-to-end learning procedure with the autoencoder structure using the original and the optimised bit mappings are given by Fig. 6.6 and Fig. 6.7, respectively. Such optimisation requires predetermined values to be fed into the model for each training. Therefore, varying values of SNR are tested and SNR = 10 dB is chosen for its optimality with the corresponding mutual information of 3.3 bits per 2D symbol.

Fig. 6.6 depicts the comparison of the BER performance of the PS-4D-256-QAM with respect to that of unshaped 2D-16-QAM and 4D-256-QAM. The 4D-256-QAM was proved in [159] to have better BER performance in a high-SNR regime relative to the 2D-16-QAM, which has the same mutual information of 4 bits/symbol/1D. Fig. 6.6 shows that the PS-4D-256-QAM given by autoencoder outperforms the 2D-16-QAM when the SNR exceeds 14.2 dB. An extra shaping gain of 0.4 dB can be achieved for the probabilistically shaped 4D QAM relative to the unshaped 4D QAM, leading to an overall power reduction of 1 dB at BER = 10^{-5}.

Fig. 6.7 shows the power advantage of PS-4D-256-QAM, resulting from the combined effects of optimisation of bit mapping and probabilistic shaping, over 4D and 2D signals, including probabilistically shaped 2D signal similar to that discussed recently in [235] for a fair comparison. The results show that using optimised bit mapping by RTS leads to a better error performance for 4D-256-QAM, such that a 1.2 dB power advantage can be achieved for PS-4D-256-QAM when compared to the 2D-16-QAM. While even compared to PS-2D-16-QAM signal optimised by autoencoder, there is still approximately 1 dB advantage achieved by the PS-4D-256-QAM. Besides, the crossing points of the BER performance show that the optimal the PS-4D-256-QAM with optimised bit mapping and probabilistic shaping outperforms the PS-2D-16-QAM and the unshaped 2D-16-QAM when the SNR exceeds 13.2 dB and 12.1 dB as the corresponding BER is above 10^{-2}.

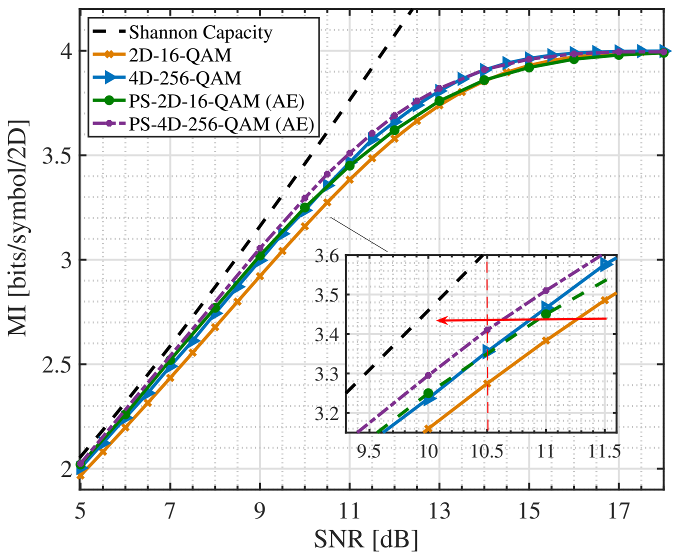
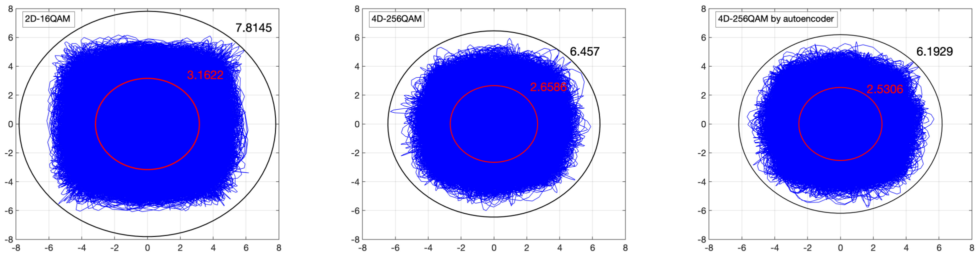


Figure 6.8: Mutual information versus SNR, with/without probabilistic shaping with autoencoder

6.4.2.3 Mutual Information Performance

We assess the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ of the 4D QAM with optimised probabilistic shaping from the proposed scheme regarding other schemes. It is worth noting that the bit-to-symbol mapping does not affect the mutual information of modulation based on its definition in equation (6.12). Therefore the mapping is not taken into account in simulations for the comparative study. The mutual information of probabilistically shaped QAM signals is analysed while the SNR varies for a range from 5 dB to 20 dB. Fig. 6.8 shows significant gain achieved by the probabilistically shaped 4D QAM signal when compared to the 4D-256-QAM and the 2D-16-QAM. The unshaped 4D is shown to achieve higher mutual information relative to the PS-2D-16-QAM generated by the autoencoder in a high SNR regime. Both probabilistic shaping and 4D constellation design achieve enhanced achievable capacity when compared to the regular 2D QAM. It can be observed that the mutual information of the PS-4D-256-QAM approaches Shannon's limit. In particular, the shaped 4D signal is shown to have enhanced capacity in the low SNR regime where the mutual information is below the benchmark of the maximum achievable data rate for the modulation scheme concerned.



(a) (b) (c)

Figure 6.9: PAPR Performance for (a) ideal 2D-16-QAM (b) 4D-256-QAM and (c) PS-4D-256-QAM

6.4.2.4 Power Efficiency

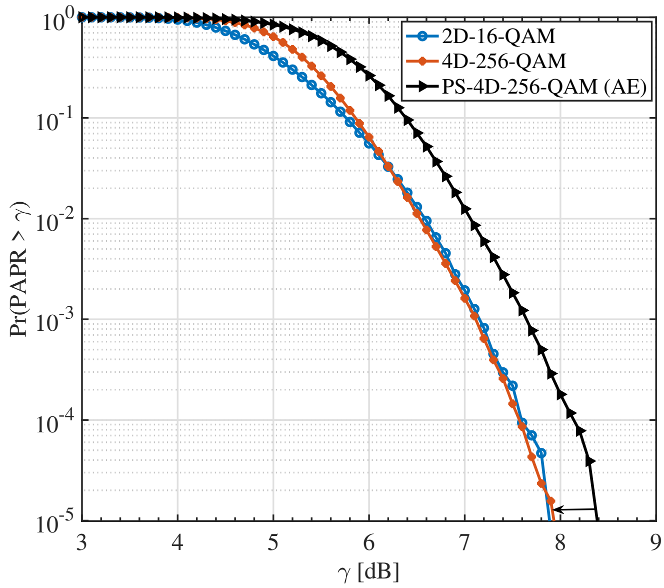


Figure 6.10: CCDF of the PAPR for PS-4D-256-QAM, 4D-256-QAM and 2D-16-QAM symbols with sqaured-root raised cosine (SRRC) pulse shaping applied with roll-off factor $\alpha = 0.2$ and filter length $L = 20$.

The power efficiency performance of the probabilistically shaped 4D signal is investigated by numerical simulations in terms of the PAPR. The calculation of PAPR is detailed in Appendix C. Fig. 6.9 shows the peak and average power of three signals. It can be seen that the 4D signals have smaller peak and average power when compared to the 2D signal. The probabilistic shaping on 4D signal reduces both the peak and average power of the signal. Fig. 6.10 shows the CCDF of the PAPR for the 2D and 4D signals. The results indicate that the unshaped 4D signal has a similar PAPR as the 2D signal. The 4D signal has a slightly higher PAPR when compared to the unshaped signals, as expected. This is due to the nature of probabilistic shaping, such that the average signal power is minimised at a fixed data rate so as to maximise the overall AIR while the peak power remains approximately the same. Nevertheless, the PAPR penalty of the PS-4D-256-QAM is less than 0.5 dB for 99.99% of the signals as shown in Fig. 6.10.

6.5 Conclusion

This chapter explores the application of machine learning algorithms in physical layer communication systems, particularly in signal design. A new optimisation method for multidimensional probabilistic shaping is presented by using the autoencoder model with unsupervised end-to-end learning. Based on the study of constellation shaping in Chapter 5 and multidimensional modulation in Chapter 4, this work aims to amend the research gap of multidimensional probabilistic shaping.

The key challenge of multidimensional probabilistic shaping is the asymmetric feature of the lattice-based multidimensional constellations. In this case, the conventional constellation shaping can not be applied to these multidimensional signals and thus requires new designs. This chapter commences with an overview of machine learning applications in communication systems as well as a brief introduction to the basic concepts of neural networks, deep learning and autoencoder. To provide significant insight into the proposed scheme, it is addressed that the research gap of lacking optimisation designs for multidimensional probabilistic shaping.

The chapter further showcased the autoencoder-based optimisation model, addressing the trainable approximation of the discrete probability distribution using the Gumbel-Softmax trick. Moreover, two main figures of merit used to evaluate the signal performance in communication systems, the AIR mutual information and the categorical cross entropy, are explored for multidimensional signals based on which the loss function is derived.

To verify the efficacy of the autoencoder-based optimisation model and the end-to-end learning method, Simulations were conducted using the 4D-256-QAM signalling model as a prototype. The results revealed learned constellations exhibiting power advantage and enhanced mutual information performance, at the cost of negligible power efficiency in terms of the PAPR when compared to the regular non-shaped 4D or 2D modulation formats that achieve the same spectral efficiency. The overall assessment affirmed the potential of the novel approach of applying autoencoder for the optimisation of multidimensional probabilistic shaping in communication systems.

Chapter 7

Conclusions

7.1 Summary and Discussions

Wireless communication technologies are evolving to achieve higher spectral and energy efficiency. As mobile devices and data-intensive applications continue to proliferate, the demand for faster and more reliable mobile communications continues to increase. However, the limited availability of the RF spectrum and energy constraints pose ongoing challenges for the wireless communication industry. To meet this demand, novel techniques are being developed and employed to provide more efficient use of RF spectrum and lower energy consumption. The main objective of this research work is to examine techniques aimed at enhancing spectral and energy efficiency for wireless communications for next generation networks. In this thesis, new techniques have been proposed and explored to optimise the wireless signal design to meet the different requirements for various use scenarios for next generation communications.

Towards the ultimate goal of providing possible designs for future generations of wireless networks, work in this thesis was conducted in two main parts; firstly, the fundamentals of three basic MCM signals, OFDM, SEFDM and FOFDM, and their associated systems were mathematically modelled and their key feature were studied, with an original coexistence study of orthogonal OFDM and non-orthogonal SEFDM signalling. Second, from Chapter 3 to Chapter 6, each chapter explored a new technique that enhances spectral efficiency and proposed or developed novel methods to improve the overall system performance, which was tested in simulations with numerical investigations. In general, particular attention has been given to MCM waveform formats in the first three chapters, due to good performance in terms of the spectral efficiency and energy advantages, when used in conjunction with FEC coding or advanced detection methods. Specifically, the non-orthogonal SEFDM and its special variant FOFDM have been considered, due to the reported significant spectral efficiency improvement without sacrificing much of the error performance. Chapter 5 and Chapter 6 focus more on optimisation of the higher-dimensional signal designs and therefore SCM system models were adopted to evaluate the signal performance as 'proof-of-concept' implementations. The optimisation problem associated with this implementation was analysed theoretically and modelled mathematically in these two chapters, with new models developed and appropriate cost function derived, which takes into account both the signal Euclidean and Hamming distances. Due to the different characteristics of the optimisation problems, different solutions were chosen; conventional metaheuristic optimisation methods were investigated to optimise the bit-to-symbol mapping for the multidimensional signal; machine learning models were designed and simulated to optimise multidimensional probabilistic shaping.

This thesis commenced with the basic concept and statistical model of spectral efficiency and the theoretical upper bound given by the Shannon limit. Chapter 2 provided a comprehensive review of existing spectrally efficient technologies, starting with the discussion of the development and applications of MCM. Various spectrally efficient MCM schemes were critically reviewed in terms of their performance and applications, identifying the benefits and drawbacks of each and leading to the research carried out through the first part of this thesis, based on the chosen systems. Three key MCM systems were discussed and mathematically modelled, including the most prominent signalling techniques OFDM and two promising spectrally efficient non-orthogonal designs FOFDM and SEFDM. Detailed mathematical studies of the ICI in FOFDM were conducted to prove its maintained orthogonality among only the real subcarriers. This can lead to the conclusion that FOFDM symbols can be perfectly recovered without distortion from ICI when singular dimensional modulation schemes are adopted. A similar statistical model of ICI in SEFDM was investigated, indicating that the deliberately introduced ICI largely depends on the compression level of subcarriers. To test the feasibility of heterogeneous signalling, a coexistence study was carried out by mathematically modelling the transmission of orthogonal OFDM and non-orthogonal SEFDM signals simultaneously in the adjacent BWPs and assuming standard compliant 5G NR frame. Three scenarios were considered firstly without the use of any channel coding; evident performance degradation appeared when two OFDM signals with different SCS were transmitted while signals with larger SCS were shown to be more robust to the inter-numerology interference; acceptable error performance was achieved when mixing two signal formats when a sphere decoder for SEFDM and a matched filter for OFDM were used and AWGN channel was assumed; severe interference was shown and resulted in error degradation in the case of two SEFDM signals with high compression levels. The last set of results led to a further study by employing LDPC coding, where simulation results showed improved error performance due to the amelioration of interference effects by LDPC. This coexistence study provided insight into the potential of non-orthogonal signalling SEFDM to be applied in future communication networks. Notwithstanding, the limitation of further improvement in the spectral efficiency of the aforementioned systems appeared in the studies, for which some advanced signal processing methods are required and such were developed and verified in this thesis.

To enhance the spectral efficiency of the aforementioned MCM signal formats and achieve reasonable error performance, a filtering technique is proposed, to add advantageously another layer of orthogonality, using Hilbert transform filter pair. Two use scenarios were studied in Chapter 3 to verify the efficacy of the proposed signalling method; turbo-coded SEFDM signal was employed and NB-IoT for which FOFDM signal with one-dimensional modulation formats were used. For both cases, the modulation and demodulation schemes of using Hilbert filter pair were designed and modelled. The statistical model of generation of the Hilbert filter pair was studied and provided a theoretical analysis of the orthogonality. The proposed method was evaluated in the presence of channel impairment AWGN through system simulations. For the HT-SEFDM system, compression levels with spectral efficiency gains up to 67% from SEFDM were used while the system performance benefited from the use of a coding scheme. The results show the use of Hilbert filter pair doubles the spectral efficiency of SEFDM without incurring error penalties. The second model was designed in the context of NB-IoT system by combining two orthogonal techniques; the frequency orthogonal FOFDM coupled with the time orthogonal Hilbert filter pair. This work demonstrated the compelling advantages over OFDM scheme of quadrupled effective data rate, within the 5G standard 180 kHz transmission bandwidth, through numerical simulations. Moreover, the layered orthogonality ensures the performance of the HT-FOFDM without adding error penalties and with negligible computational complexity.

Chapter 4 considers non-uniform signalling techniques, more specifically, constellation shaping, which are employed to approach Shannon's limit for MCM systems. This thesis presented studies focused on probabilistic shaping, which exploits a symmetric occurrence distribution among constellations of transmitted symbols and its application to SEFDM to improve further system performance. A new PS-SEFDM system was designed and developed to bring the advantages of probabilistic shaping while maintaining the error performance. At the core of the proposed probabilistic shaping architecture, a distribution matching unit (fixed-to-fixed length matching technique CCDM) was operated jointly with a probabilistic amplitude shaping PAS-unit in conjunction with LDPC coding. This proved to be a successful technique and was verified through the comparative study of PS-SEFDM versus LDPC-coded OFDM and uniformly distributed SEFDM. For fair comparisons, signals parameters were specified mathematically/analytically, to set coding rates and DM rates to reach the same achievable spectral efficiency. The error performance was assessed and results proved that the new PS-SEFDM scheme outperforms OFDM when the same spectral efficiency is achieved. Furthermore, the available variation of the compression level of SEFDM extends the flexibility of rate adaptation of the probabilistic shaping scheme. Until now, the channel impairments are limited to AWGN. For further justification of the performance of the PS scheme in the practical environment, two multipath fading channel models were modelled and simulated mathematically; the static frequency selective channel and the Rummler's two-way channel model. Results proved that the PS-OFDM/SEFDM have better immunity against the multipath fading effect, achieving power advantages relative to regular LDPC-coded signals in fading channels.

Inspired by the ultimate shaping gain of 1.53 dB that can be achieved by probabilistic shaping as the modulation dimension approaches infinity, Chapter 5 delved into high dimensional modulations employing the temporal dimension to increase the dimensionality of signal formats. More specifically, a four-dimensional modulation format that provides non-uniform constellations by processing multiple time symbols jointly was studied. The design of bit-to-symbol mapping for the four-dimensional signal was challenged: conventional optimal labelling via gray mapping cannot be used because of the constellation symbols' neighbourhood structure in multidimensional space. Therefore, this work proposed the first optimisation model for the bit-to-symbol mapping for multidimensional signal format. To formulate the optimisation problem, the cost function was designed for the bit mapping for the four-dimensional signal derived from the analytical upper bound of BER, taking into account both the Euclidean distance between any pair of symbols and the Hamming distance between the corresponding bit vectors. Based on the non-convex characteristic of such problem, metaheuristic algorithms were explored and four metaheuristic algorithms were investigated; BSA, GA, SA and RTS. This work developed algorithms based on the four methods and adjusted the design to the specific problem. Numerical simulations were carried out to compute the optimisations. The performance was evaluated in terms of the computational complexity and convergence speed of each of the four algorithms.

Another study on multidimensional signal optimisation was conducted in Chapter 6, through the application of machine learning techniques. A particular neural network model, *autoencoder*, was proposed to enable multidimensional probabilistic shaping, to improve the signal performance and to optimise the design. Based on the discussion in Chapter 4, the challenge was addressed for multidimensional probabilistic shaping. While conventional probabilistic shaping is well investigated and implemented in 1D, to the best of the author's knowledge, there is no universal design for multidimensional probabilistic shaping so far. Hence, the work in this chapter presented an nD probabilistic shaping design employing an autoencoder-based model. The cost function of the machine learning model was derived by taking two figures of merit into account; the achievable information rates (i.e., the mutual information) and the categorical cross entropy. The same four-dimensional signal described and studied in Chapter 5 was used hereby as an example for the optimisation study. The autoencoder model was designed and trained in an end-to-end manner. The learned probabilistically shaped signals were assessed, showing power advantage and improved mutual information performance at the cost of negligible power efficiency loss in terms of the PAPR.

To summarise, the research presented in this thesis has demonstrated the advantages of different techniques to improve the spectral and energy efficiency of wireless communication systems, including non-orthogonal signalling with pulse shape filtering and non-uniform signalling, specifically, constellation shaping and multidimensional modulation. It proposed new signalling methods that further improve the spectral efficiency of spectrally efficient signal formats without incurring the error performance in ideal Gaussian noise channels and practical multipath propagation models. Nevertheless, it is worth noting that the enhanced spectrum efficiency of the multi-carrier signals is at the cost of slightly increased computational complexity due to the introduced signal processing at both transmitter and receiver ends. Furthermore, the thesis optimised multidimensional modulation signal designs in two aspects: bit-to-symbol mapping and constellation shaping. For these, conventional metaheuristic method based optimisation algorithms were implemented for the former problem, giving optimum bit mapping that provides better power advantages; the optimised multidimensional probabilistic shaping computed by autoencoder-based end-to-end learning has shown substantial improvement in achievable information rate and SNR advantage.

7.2 Future Work

Investigations in this thesis generate a number of new research and engineering questions to be tackled. A non-exhaustive list of potential future investigations, as a continuation of the work reported here, as well as new areas of interest, are briefly described below:

- **Hardware prototyping and testing of an end-to-end system using Hilbert filter pair**: Chapter 3 studied two spectrally efficient systems employing Hilbert filter pair for two specific use scenarios. This places the systems studied on a path for end-to-end implementation. The main challenge for practical hardware implementation is the system complexity, which has been investigated briefly and requires more detailed analytical studies. In addition, only a simple matched filter was used in the reported work for 'proof of concept'. Future investigations suggested above require thorough studies in advanced detection methods and receiver designs, especially for the HT-SEFDM. Applicable detection techniques have been extensively studied in [9, 11–13]. It will also be beneficial to consider the complexity, precision of hardware and performance for the implementation when employing practical detection methods. Such will hopefully open a new avenue of research into the practical implementation of the proposed bit rate expanding techniques.
- **Application in optical-wireless and VLC links**: Given the successful applications and implementations of the FOFDM system in studies for optical communications [68, 70, 238], it is imperative to further examine the potential application of HT-FOFDM system in VLC. Such system demonstrated in [29] and Chapter 3 is of potential advantages for high bit rate optical wireless and VLC systems, with the aligned requirements for signal design as the one in the use scenario of NB-IoT; namely low computational complexity. Hence, HT-FOFDM signal will be expected to show attractive spectral efficiency and

error performance in VLC links. Although the use of Hilbert filter pair only introduces a slight complexity increase due to the filtering procedure, yet, to constrain such increased complexity, the effects of certain system design parameters need to be investigated analytically, such as the filter length and the number of subcarriers. This may well open a new research area with significant practical implications.

- **Enhancing the performance by joint geometric and probabilistic shaping**: The PS-SEFDM system developed in Chapter 3 aims at enhancing spectral efficiency for a certain power consumption or gaining power advantage for a given data rate. Geometric shaping, the other technique briefly mentioned in the chapter, has similar advantages as probabilistic shaping and could be explored to attain further performance improvements. This application of geometric shaping to SEFDM could be studied separately first and then a hybrid shaping scheme could be modelled to examine the efficacy of using both constellation shaping techniques, as similar studies have been reported for optical systems [18]. Having said that, it is worth noting that there is a clear trade-off between complexity and capacity improvement, particularly in practical system implementation. Therefore, optimising the constellations for geometric shaping and hybrid shaping is also a challenge that needs to be tackled. The machine learning technique presented in Chapter 6 would be one useful approach to this issue.

- **Design studies of non-orthogonal signalling system with constellation shaping for mmWave transmission**: The mmWave (30 GHz - 300 GHz) is proposed for use in 6G for its potential to provide high-data-rate and high capacity communication services. The work reported in Chapter 4 and in [30] concerns the application of PS-SEFDM to DVB-S2 system, which might be extended to mmWave transmission. This will be expected to show the advantages of PS-SEFDM proposed format in improving spectral efficiency and power efficiency (better PAPR performance), relative to OFDM. The main challenge of developing such system is to overcome the effects of channel impairment in the mmWave frequency bands. Specific channel models should be considered in such research, such as the model specified in 3GPP Release 14 [239], for mathematical modelling of the mmWave channel using deterministic and stochastic propagation processes. The performance of the SEFDM system with constellation shaping should be compared with a single carrier, OFDM and other non-orthogonal waveform designs, with the aim of optimising the number of sub-carriers and the transmission format.

- **Studies of non-orthogonal and probabilistic shaped signals in MIMO systems**: Whilst the research work presented in Chapters 3, 4 and 5 only considered single transmission links SISO, capacity enhancement using MIMO links may be considered as an extension tom the three types of signals considered. This would necessitate further optimisation and more detailed consideration of the transmission channel and propagation models. In addition, extending the work on signal shaping or dimensionality and assuming recently adopted transmission links optimisation methods, such as adaptive modulation and coding (AMC), will give an added dimension to the AMC as recently proposed for for SEFDM in [240]. These topics will be worth investigating by modelling and experimentation.
- **Studies of optimising higher order and/or higher dimensional modulation formats**: This future research line stems from the research work presented in Chapter 5. The work formulated the optimisation for bit-to-symbol mapping for multidimensional signals and developed general algorithms to tackle the problem. In particular, the algorithms were applied to a four-dimensional signal and their efficacy and performance have been verified. However, to examine the generalisability of the developed algorithms, it will be useful to study the optimisation for higher order and/or higher dimensional signals, such as 8D-16-QAM or 4D-1024-QAM. Intrinsically, the problem formulation remains the same as discussed in the thesis, while the complexity becomes the major issue, requiring a new look at appropriate optimisation methods and techniques

- **Applications of graph neural networks to the optimisation of multidimensional modulation formats**: Chapter 5 presented the developed conventional metaheuristic algorithms to optimise the bit-to-symbol mapping for multidimensional signals. Metaheuristic algorithms are trajectory-based and hence, can lead to high computational complexity. Graph neural networks have been recently considered for the quadratic assignment problem, which has been studied recently as a special type of combinatorial optimisation problem [241]. For instance, instead of learning the large constellation structure directly, the graph tool can observe and extract the feature from a smaller constellation structure in multidimensional space and apply this to the larger constellation. This would have the potential of going beyond traditional optimisation methods and expanding the idea explored in Chapter 5 for further improvement of system performance.

Overall, this thesis and its underlying research work have explored a variety of new engineering design and system techniques, mathematical and analytical methods, and optimisation and machine learning techniques for enhancing capacity and energy efficiency in next generation wireless communication networks. A number of original contributions have been made, with six papers published and two patent applications. It is hoped that the studies in this thesis will pave the way to further investigations and implementation in future communication systems.

Appendix A

Hilbert Transform Preliminaries

Appendix A provides the preliminaries on the mathematical formulations and properties of Hilbert transform used in Chapter 3 of the thesis. This Appendix is used for ease of reference for readers.

Mathematically, the Hilbert Transform $\hat{g}(t)$ of a signal $g(t)$ can be defined as a convolution between the function $\frac{1}{\pi t}$ and the signal $g(t)$ in time domain [115]. The mathematical expression of $\hat{g}(t)$ is given by [115]

$$\hat{g}(t) = \mathscr{H}[g(t)] = g(t) * \frac{1}{\pi t} = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{g(\tau)}{t-\tau}d\tau = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{g(t-\tau)}{\tau}d\tau. \tag{A.1}$$

It is worth noting that both $g(t)$ and $\hat{g}(t)$ are signals defined in time domain. If $g(t)$ is a real function, then $\hat{g}(t)$ is real-valued [115].

For a real function $f(t)$, the portion of the power spectrum in the positive frequency axis contains the complete information about the waveform in time domain [242]. Hence, the spectrum located in negative frequency axis can be removed by using Hilbert transform.

As mentioned in the definition part, the $\hat{g}(t)$ can be regarded as a convolution in time domain. A corresponding expression in frequency domain can be given by [116]

$$\hat{G}(f) = G(f) \cdot \mathscr{F}\{\frac{1}{\pi t}\}, \tag{A.2}$$

where $\hat{G}(f)$ and $G(f)$ are the Fourier Transform of $g(t)$ and $\hat{g}(t)$ respectively. $\mathscr{F}\{\frac{1}{\pi t}\}$ denotes the Fourier Transform of signal $\frac{1}{\pi t}$, which can be expressed by

[115]

$$\mathscr{F}\{\frac{1}{\pi t}\} = -j \cdot sgn(f) = \begin{cases} -j & \text{if } f > 0 \\ 0 & \text{if } f = 0 \\ j & \text{if } f < 0. \end{cases} \tag{A.3}$$

Based on equation A.3, it can be seen that the Hilbert Transform does not change the magnitude of the signal frequency representation $G(f)$. Similar to a filter, it changes the phase of the signal regarding the sign of its frequency. As such, all the signals of negative frequency get a $+90°$ phase shift while all the signals of positive frequency get a $-90°$ phase shift [242]. This is also called quadrature filter. The analytic signal is a complex signal defined as the follows [115]

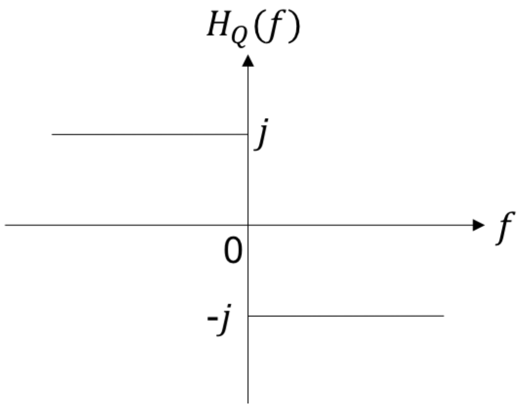


Figure A.1: Quadrature Filter

$$g_+(t) = g(t) + j \cdot \hat{g}(t), \tag{A.4}$$

wherein $g(t)$ is real-valued. As proved in [115], the real function $g(t)$ and its Hilbert transform $\hat{g}(t)$ are orthogonal.

$$\int_{-\infty}^{+\infty} g(t) \cdot \hat{g}(t) = 0. \tag{A.5}$$

The analytic signal is mainly used in two important applications: the SSB amplitude modulation and the mathematical representation of bandpass signals [116].

Some of the basic properties of the Hilbert transform are listed below. For more information about Hilbert transform and its properties, readers are encouraged to refer to [115] and [116], from which the below contents are sourced.

Linearity: The Hilbert transform is linear. For arbitrary scalars c_1 and c_2, and arbitrary functions $g_1(t)$ and $g_2(t)$, if the function $g(t) = c_1 g_1(t) + c_2 g_2(t)$, when the Hilbert transform of $g_1(t)$ and $g_2(t)$ exists, then for all t,

$$\hat{g}(t) = \mathscr{H}[g(t)] = \mathscr{H}[c_1 g_1(t) + c_2 g_2(t)] = c_1 \hat{g_1}(t) + c_2 \hat{g_2}(t). \tag{A.6}$$

Inverse Transform: If $\hat{g}(t)$ is the Hilbert transform of the real function $g(t)$, then the Hilbert transform of function $\hat{g}(t)$ can be given by

$$\mathscr{H}[\hat{g}(t)] = \mathscr{H}[g(t) * \frac{1}{\pi t}] = g(t) * \frac{1}{\pi t} * \frac{1}{\pi t} = g(t) * (\frac{1}{\pi t} * \frac{1}{\pi t}) = -g(t). \tag{A.7}$$

Multiple Transform: As is proved in inverse transform, twice Hilbert transform on a real function derives the original function with altered sign. By the analogue, multiple Hilbert transform used on the same function can be expressed as the follows

$$\mathscr{H}^2[g(t)] = -g(t). \tag{A.8}$$

$$\mathscr{H}^3[g(t)] = \mathscr{H}[\mathscr{H}^2[g(t)]] = -\hat{g}(t) \tag{A.9}$$

$$\mathscr{H}^4[g(t)] = g(t). \tag{A.10}$$

Hence, the Hilbert transform can be used N times, where N is an even integer, to obtain the same real function or its opposite. For a real function $g(t)$, of which the

Hilbert transform $\hat{g}(t)$ exists, then

$$\mathscr{H}^N g(t) = \begin{cases} g(t) & \text{if } N = 4m, m \in \mathbb{Z}_+ \\ \hat{g}(t) & \text{if } N = 4m+1, m \in \mathbb{Z}_+ \\ -g(t) & \text{if } N = 4m+2, m \in \mathbb{Z}_+ \\ -\hat{g}(t) & \text{if } N = 4m+3, m \in \mathbb{Z}_+. \end{cases} \tag{A.11}$$

Convolution: Since the Hilbert transform is defined on the basis of convolution, calculation with respect to convolution can be much simplified when using Hilbert Transform. For example,

$$\mathscr{H}[g_1(t) * g_2(t)] = \hat{g_1}(t) * g_2(t) = g_1(t) * \hat{g_2}(t). \tag{A.12}$$

This can be easily proved due to the associative and commutative properties of convolution operation.

Orthogonality: The real function $g(t)$ and its Hilbert transform $\hat{g}(t)$, or the real part and imaginary part of a strong analytic signal $g_+(t)$, are orthogonal. The proof process is given below.

$$\begin{aligned} \int_{-\infty}^{+\infty} g(t) \cdot \hat{g}^*(t) dt &= \int_{-\infty}^{+\infty} G(f) \cdot \hat{G}^*(f) df \\ &= \int_{-\infty}^{+\infty} G(f) \cdot [-jsgn(f) \hat{G}(f)]^* df \\ &= \int_{-\infty}^{+\infty} j |G(f)|^2 sgn(f) df \\ &= 0, \end{aligned} \tag{A.13}$$

where $|G(f)|^2$ is an even function of f, and $sgn(f)$ is an odd function of f, hence the integration of the product of two functions equals to zero.

Appendix B

AWGN, SNR and E_b/N_0

As there are many methods and norms for mathematically defining these key communication systems metrics, AWGN, SNR and E_b/N_0, the specific mathematical definitions of these metrics in this thesis are briefly present here.

B.1 Gaussian Noise

White noise is the most common form of noise in wireless communication systems[11]. In this thesis, only an AWGN channel is considered in the system design and modelling. The AWGN, by definition, refers to a linear addition of white noise that owns a constant power spectral density and a Gaussian distributed amplitude. Thus, the mean of the AWGN noise can be expressed as [243]

$$E\{w_i\} = 0, \tag{B.1}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. The variance of AWGN noise is given by [243]

$$\sigma^2 = \frac{N_0}{2} \int_{-\infty}^{+\infty} |H(f)|^2 df, \tag{B.2}$$

where N_0 is the defined as the noise power per unit bandwidth, of which the unit is $watts/Hz$, the integral term denotes the power in the infinite frequency band, whereby $H(f)$ is the Fourier transform of the impulse response $h(t)$ of a linear time-invariant system [116] [243].

In the simulation, the AWGN can be represented by a series of Gaussian distributed variables w_i with a zero mean and the variance of σ, where the index i denotes the discrete timestamps. The power spectral density of the AWGN noise sample vector w_i can be given by [243]

$$S_w(f) = \frac{N_0}{2}. \tag{B.3}$$

It is worth noting that equation(B.3) is an approximation in simulation due to the assumption that the frequency range of w_i is wide enough, since $\frac{N_0}{2}$ is termed as the two-sided power spectral density for all f, $-\infty < f < +\infty$.

Another significant parameter of the AWGN noise is auto-correlation function of the process, which is given by

$$R_w(\tau) = \begin{cases} \frac{N_0}{2}\delta(\tau) & \text{if } \tau = 0 \\ 0 & \text{if } \tau \neq 0 \end{cases} \tag{B.4}$$

where $\delta(\tau)$ denotes the unit impulse. Equation(B.4) implies the AWGN variables are independent statistically. In the process of the transmission, the ratio of spectral density between signal and noise is significant, termed as E_b/N_0. Recalling from section 2.3.3, equation(2.14) describes the discrete SEFDM signal. Therefore, the average power per bit can be expressed as $\frac{\|x\|^2}{Nlog_2M}$. According to the conversion between power and energy, E_b can be yielded as [11]

$$E_b = \frac{\|x\|^2 T_b}{Nlog_2M}. \tag{B.5}$$

Subsequently, the noise spectral density N_0 can be calculated based on:

$$E_b/N_0 = 10 \times log_{10}(\frac{E_b}{N_0}). \tag{B.6}$$

B.2 BER versus SNR or E_b/N_0

The error rate is evaluated regarding to either SNR or E_b/N_0 in digital communications. Therefore, the conversion between this two measurements are important.

SNR is signal-to-noise ratio, often expressed in decibels by

$$SNR_{dB} = 10 \times log_{10}(SNR). \tag{B.7}$$

It measures the ratio of signal power and the background noise power at the receiver.

E_b/N_0 is energy per bit to noise power spectral density ratio, also known as SNR per bit. It can be given wherein E_b is the energy per bit and N_0 is the associated noise energy. The relation between SNR and E_b/N_0 in dB can be given by

$$SNR = \frac{P_{sig}}{P_noise}, \tag{B.8}$$

where P_{sig} and P_n are the signal power and noise power respectively. Signal power can be expressed as energy per time unit and the noise power is the product of the power density and the signal bandwidth. Consequently, SNR can be given by,

$$SNR = \frac{E_s/T_s}{N_0 \times B}. \tag{B.9}$$

Thus, SNR as a function of E_b/N_0 is given by

$$\begin{aligned} SNR &= \frac{E_b \times R_b}{N_0 \times B} = E_b/N_0 \times \eta, \\ SNR_{dB} &= \frac{E_b}{N_0}_{dB} + 10 \times log_{10}(\eta), \end{aligned}$$

where $\eta bits/s/Hz$, i.e., data rate per bandwidth unit, denotes the spectral efficiency. In this thesis, taking SEFDM system for an example, the spectral efficiency of SEFDM signal is expressed as

$$\eta_{SEFDM,\alpha,M-QAM} = \frac{log_2(M)}{\alpha} \tag{B.10}$$

wherein M denotes the modulation level of QAM signal and α is the bandwidth compression factor of SEFDM symbol.

Appendix C

Calculation of PAPR

The calculation of PAPR, used first in Section 2.3.3 to compare SEFDM to OFDM, and then in Chapters 3, 5 and 6 as part of the performance evaluation, is based on the standard definition below.

Considering discrete time representation for the k^{th} time sample of a single multi-carrier symbol [119]

$$X[k] = \frac{1}{\sqrt{Q}} \sum_{n=0}^{N-1} s_n \cdot e^{j2\pi nk\alpha/Q}, \tag{C.1}$$

where $\mathbf{s} = [s_0, s_1, ..., s_{N-1}]$ is the complex input symbol for N subcarriers and Q represents the total number of discrete-time samples in one SEFDM symbol. The PAPR of this multi-carrier symbol is generally defined as

$$PAPR = \frac{max|x[k]|^2}{E[|x[k]|^2]}, \tag{C.2}$$

where the numerator denotes the maximum power of the transmitted signal, $|\cdot|$ gives the absolute value of the signal magnitude and $\mathbb{E}[\cdot]$ is the expectation operator that calculates the average power.

List of References

[1] S. Ramseler, M. Arzberger, and A. Hauser, "MV and LV Powerline Communications: New Proposed IEC Standards," in *1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333)*, vol. 1. IEEE, 1999, pp. 235–239.

[2] 3GPP TS 36.201 version 11.1.0, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer; General Description," Rel. 11, Feb. 2013.

[3] 3GPP TS 38.212 version 15.2.0, "5G NR; Multiplexing and Channel Coding," Rel. 15, Jul. 2018.

[4] E. Perahia, "IEEE 802.11n Development: History, Process, and Technology," *IEEE Communications Magazine*, vol. 46, no. 7, pp. 48–55, 2008.

[5] W. Forum, "Mobile WiMAX-A Technical Overview and Performance Evaluation," 2006.

[6] J. Armstrong, "OFDM for Optical Communications," *Journal of lightwave technology*, vol. 27, no. 3, pp. 189–204, 2009.

[7] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[8] I. Darwazeh, H. Ghannam, and T. Xu, "The First 15 Years of SEFDM: A Brief Survey," in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*. IEEE, 2018, pp. 1–7.

[9] I. Kanaras, "Spectrally Efficient Multicarrier Communication Systems: Signal Detection, Mathematical Modelling and Optimisation," Ph.D. dissertation, UCL (University College London), 2010.

[10] S. I. A. Ahmed, "Spectrally Efficient FDM Communication Signals and Transceivers: Design, Mathematical Modelling and System Optimization," Ph.D. dissertation, UCL (University College London), 2011.

[11] R. C. Grammenos, "Spectrum Optimisation in Wireless Communication Systems: Technology Evaluation, System Design and Practical Implementation," Ph.D. dissertation, UCL (University College London), 2013.

[12] T. Xu, "Bandwidth Compressed Waveform and System Design for Wireless and Optical Communications: Theory and Practice," Ph.D. dissertation, UCL (University College London), 2017.

[13] H. Ghannam, "Mathematical Modelling and Signal and System Design for Spectrally Efficient Future Wireless Communications," Ph.D. dissertation, UCL (University College London), 2019.

[14] G. Forney and L.-F. Wei, "Multidimensional Constellations. I. Introduction, Figures of Merit, and Generalized Cross Constellations," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, pp. 877–892, 1989.

[15] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable Signaling on the Gaussian Channel," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 726–740, 1990.

[16] F. Kschischang and S. Pasupathy, "Optimal Nonuniform Signaling for Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 913–929, 1993.

[17] G. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient Modulation for Band-limited Channels," *IEEE journal on selected areas in communications*, vol. 2, no. 5, pp. 632–647, 1984.

[18] Z. Qu and I. B. Djordjevic, "On the Probabilistic Shaping and Geometric Shaping in Optical Communication Systems," *IEEE Access*, vol. 7, pp. 21 454–21 464, 2019.

[19] T. Fehenberger, G. Böcherer, A. Alvarado, and N. Hanik, "LDPC Coded Modulation with Probabilistic Shaping for Optical Fiber Systems," in *Optical Fiber Communication Conference*. Optical Society of America, 2015, pp. 1–3.

[20] A. V. Kisel, "An Extension of Pulse Shaping Filter Theory," *IEEE transactions on communications*, vol. 47, no. 5, pp. 645–647, 1999.

[21] R. Laroia, N. Farvardin, and S. Tretter, "On Optimal Shaping of Multidimensional Constellations," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1044–1056, 1994.

[22] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth Efficient and Rate-matched Low-density Parity-Check Coded Modulation," *IEEE Transactions on communications*, vol. 63, no. 12, pp. 4651–4665, 2015.

[23] B. Farhang-Boroujeny, "OFDM versus Filter Bank Multicarrier," *IEEE signal processing magazine*, vol. 28, no. 3, pp. 92–112, 2011.

[24] A. Sahin, I. Guvenc, and H. Arslan, "A Survey on Multicarrier Communications: Prototype Filters, Lattice Structures, and Implementation Aspects," *IEEE communications surveys & tutorials*, vol. 16, no. 3, pp. 1312–1338, 2013.

[25] 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)*. IEEE, 2015, pp. 66–70.

[26] M. Series, "Future Technology Trends of Terrestrial IMT Systems Towards

2030 and Beyond," *Int. Telecommun. Union, Geneva, Switzerland, Rep. ITU*, 2022.

[27] X. Liu, T. Xu, and I. Darwazeh, "Coexistence of Orthogonal and Non-orthogonal Multicarrier Signals in Beyond 5G Scenarios," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. IEEE, 2020, pp. 1–5.

[28] X. Liu and I. Darwazeh, "Doubling the Rate of Spectrally Efficient FDM Systems Using Hilbert Pulse Pairs," in *2019 26th International Conference on Telecommunications (ICT)*. IEEE, 2019, pp. 192–196.

[29] ——, "Quadrupling the Data Rate for Narrowband Internet of Things without Modulation Upgrade," in *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*. IEEE, 2019, pp. 1–5.

[30] X. Liu, I. Darwazeh, N. Zein, and E. Sasaki, "Spectrally Efficient FDM System with Probabilistic Shaping," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*. IEEE, 2021, pp. 1–6.

[31] ——, "Probabilistic Shaping for Multidimensional Signals with Autoencoder-based End-to-end Learning," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 2619–2624.

[32] "Certain Topics in Telegraph Transmission Theory, author=Nyquist, Harry," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[33] R. V. Hartley, "Transmission of Information 1," *Bell System technical journal*, vol. 7, no. 3, pp. 535–563, 1928.

[34] S. Isam and I. Darwazeh, "Characterizing the Intercarrier Interference of Non-orthogonal Spectrally Efficient FDM System," in *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2012 8th International Symposium on*. IEEE, 2012, pp. 1–5.

[35] M. Doelz, E. Heald, and D. Martin, "Binary Data Transmission Techniques for Linear Systems," *Proceedings of the IRE*, vol. 45, no. 5, pp. 656–661, 1957.

[36] R. W. Chang, "Synthesis of Band-limited Orthogonal Signals for Multichannel Data Transmission," *Bell system technical journal*, vol. 45, no. 10, pp. 1775–1796, 1966.

[37] ——, "Orthogonal Frequency Multiplex Data Transmission System," Jan. 6 1970, US Patent 3,488,445.

[38] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*. Elsevier Ltd., 2011.

[39] ——, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.

[40] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and Its Wireless Applications: A Survey," *IEEE transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1673–1694, 2008.

[41] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1451–1462, 1975.

[42] J. E. Mazo and H. J. Landau, "On The Minimum Distance Problem for Faster-than-Nyquist Signaling," *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1420–1427, 1988.

[43] F. Rusek and J. B. Anderson, "The Two Dimensional Mazo Limit," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.* IEEE, 2005, pp. 970–974.

[44] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist Signaling," *Proceedings of the IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[45] M. Rodrigues and I. Darwazeh, "Fast OFDM: A Proposal for Doubling the Data Rate of OFDM Schemes," 2002.

[46] F. Xiong, "M-ary Amplitude Shift Keying OFDM System," *IEEE Transactions on Communications*, vol. 51, no. 10, pp. 1638–1642, 2003.

[47] D. Karampatsis, *Modelling and Performances Assessment of OFDM and Fast-OFDM Wireless Communication Systems*. University of London, University College London (United Kingdom), 2004.

[48] K. Li, *Fast orthogonal frequency division multiplexing (Fast-OFDM) for wireless communications*. University of London, University College London (United Kingdom), 2008.

[49] T. Xu and I. Darwazeh, "Non-orthogonal Waveform Scheduling for Next Generation Narrowband IoT," in *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2018, pp. 1–6.

[50] P. A. Haigh and I. Darwazeh, "Visible Light Communications: Fast-orthogonal Frequency Division Multiplexing in Highly Bandlimited Conditions," in *2017 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. IEEE, 2017, pp. 1–8.

[51] M. Rodrigues and I. Darwazeh, "A Spectrally Efficient Frequency Division Multiplexing Based Communications System," in *Proc. 8th Int. OFDM Workshop*, 2003, pp. 48–49.

[52] D. Rainnie, Y. Feng, and J. Bajcsy, "On Capacity Merits of Spectrally Efficient FDM," in *MILCOM 2015-2015 IEEE Military Communications Conference*. IEEE, 2015, pp. 581–586.

[53] A. Barbieri, D. Fertonani, and G. Colavolpe, "Time-frequency Packing for Linear Modulations: Spectral Efficiency and Practical Detection Schemes," *IEEE Transactions on Communications*, vol. 57, no. 10, pp. 2951–2959, 2009.

[54] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What Should 6G Be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[55] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal Time Frequency Space Modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.

[56] Z. Wei, W. Yuan, S. Li, J. Yuan, G. Bharatula, R. Hadani, and L. Hanzo, "Orthogonal Time-frequency Space Modulation: A Promising Next-generation Waveform," *IEEE wireless communications*, vol. 28, no. 4, pp. 136–144, 2021.

[57] S. Isam and I. Darwazeh, "Precoded Spectrally Efficient FDM System," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2010, pp. 99–104.

[58] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-advanced for Mobile Broadband*. Academic press, 2013.

[59] 3GPP TS 36.213 version 14.2.0, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Rel. 14, Apr. 2017.

[60] B. Muquet, Z. Wang, G. B. Giannakis, M. De Courville, and P. Duhamel, "Cyclic Prefixing or Zero Padding for Wireless Multicarrier Transmissions?" *IEEE Transactions on communications*, vol. 50, no. 12, pp. 2136–2148, 2002.

[61] J. Oh and M. Lim, "The Bandwidth Efficiency Increasing Method of Multicarrier CDMA and Its Performance Evaluation in Comparison with DS-CDMA with RAKE Receivers," in *1999 IEEE 49th Vehicular Technology Conference (Cat. No. 99CH36363)*, vol. 1. IEEE, 1999, pp. 561–565.

[62] D. Karampatsis, M. Rodrigues, and I. Darwazeh, "Implications of Linear

Phase Dispersion on OFDM and Fast-OFDM systems," in *London Communications Symposium UCL*, 2002, pp. 117–120.

[63] D. Karampatsis and I. Darwazeh, "Performance Comparison of OFDM and FOFDM Communication Systems in Typical GSM Multipath Environments," in *Proc. London Commun. Symp.* Citeseer, 2003, pp. 360–372.

[64] K. Li and I. Darwazeh, "System Performance Comparison of Fast-OFDM System and Overlapping Multi-carrier DS-CDMA Scheme," in *London Communications Symposium*, 2006.

[65] ——, "System Performance Comparison of Fast-OFDM with Overlapping MC-DS-CDMA and MT-CDMA Systems," in *2007 6th International Conference on Information, Communications & Signal Processing*. IEEE, 2007, pp. 1–4.

[66] I. Kanaras, A. Chorti, M. Rodrigues, and I. Darwazeh, "Analysis of Sub-optimum Detection Techniques for a Bandwidth Efficient Multi-carrier Communication System," in *Cranfield Multi-Strand Conference: Creating Wealth Through Research and Innovation*, vol. 6, 2008, p. 505.

[67] T. Xu and I. Darwazeh, "Non-orthogonal Narrowband Internet of Things: A Design for Saving Bandwidth and Doubling the Number of Connected Devices," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2120–2129, 2018.

[68] P. A. Haigh, A. Burton, Z. Ghassemlooy, and I. Darwazeh, "Visible Light Communications: Simplified Co-equalisation of Fast OFDM in a Multiple-input Multiple-output Configuration," in *2019 26th International Conference on Telecommunications (ICT)*. IEEE, 2019, pp. 46–50.

[69] L. Zhu, T. Xu, and I. Darwazeh, "Fast-OFDM Transmission with Duobinary 3-PSK Modulation," in *Proceedings-2019 International Conference on Wireless Networks and Mobile Communications, WINCOM 2019*. IEEE, 2019.

[70] L. C. Vieira, S. Hussein, I. Darwazeh, C.-P. Liu, and J. Mitchell, "A Combined Digital Linearization and Channel Estimation Approach for IM/DD Fast-OFDM Systems," *Optical Fiber Technology*, vol. 67, p. 102725, 2021.

[71] L. C. Vieira, I. Darwazeh, S. Hussein, C.-P. Liu, and J. Mitchell, "Experimental Demonstration of Direct-detection Optical Fast-OFDM Using Memory Polynomials," in *2021 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 2021, pp. 1–5.

[72] L. C. Vieira, W. Ozan, J. Mitchell, and I. Darwazeh, "Modeling and Compensation of Nonlinear Distortion in Direct-detection Optical Fast-OFDM Systems," in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*. IEEE, 2022, pp. 168–173.

[73] J. Zhao and A. D. Ellis, "A Novel Optical Fast OFDM with Reduced Channel Spacing Equal to Half of the Symbol Rate per Carrier," in *Optical Fiber Communication Conference*. Optical Society of America, 2010, p. OMR1.

[74] J. Zhao and A. D. Ellis, "Discrete-Fourier Transform Based Implementation for Optical Fast OFDM," in *36th European Conference and Exhibition on Optical Communication*, 2010, pp. 1–3.

[75] S. K. Ibrahim, J. Zhao, D. Rafique, J. A. O'Dowd, and A. D. Ellis, "Demonstration of World-first Experimental Optical Fast OFDM System at 7.174Gbit/s and 14.348Gbit/s," in *36th European Conference and Exhibition on Optical Communication*, 2010, pp. 1–3.

[76] J. Zhao, S. K. Ibrahim, D. Rafique, P. Gunning, and A. D. Ellis, "Symbol Synchronization Exploiting the Symmetric Property in Optical Fast OFDM," *IEEE Photonics Technology Letters*, vol. 23, no. 9, pp. 594–596, 2011.

[77] ——, "A Novel Method for Precise Symbol Synchronization in Double-side Band Optical Fast OFDM," in *2011 Optical Fiber Communication Confer-*

ence and Exposition and the National Fiber Optic Engineers Conference, 2011, pp. 1–3.

[78] J. Zhao, S. K. Ibrahim, P. Gunning, and A. D. Ellis, "Chromatic Dispersion Compensation Using Symmetric Extension Based Guard Interval in Optical Fast-OFDM," in *2011 37th European Conference and Exhibition on Optical Communication*, 2011, pp. 1–3.

[79] J. Zhao and A. Ellis, "Transmission of 4-ASK Optical Fast OFDM With Chromatic Dispersion Compensation," *IEEE Photonics Technology Letters*, vol. 24, no. 1, pp. 34–36, 2012.

[80] E. Giacoumidis, S. K. Ibrahim, J. Zhao, J. M. Tang, A. D. Ellis, and I. Tomkos, "Experimental and Theoretical Investigations of Intensity-Modulation and Direct-Detection Optical Fast-OFDM Over MMF-Links," *IEEE Photonics Technology Letters*, vol. 24, no. 1, pp. 52–54, 2012.

[81] J. Zhao and A. D. Ellis, "Channel Estimation and Compensation in Chromatic Dispersion Limited Optical Fast OFDM Systems," in *2012 8th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, 2012, pp. 1–4.

[82] J. Zhao and A. Ellis, "Advantage of Optical Fast OFDM Over OFDM in Residual Frequency Offset Compensation," *IEEE Photonics Technology Letters*, vol. 24, no. 24, pp. 2284–2287, 2012.

[83] J. Zhao, "Intensity-modulation Full-field Detection Optical Fast OFDM," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 5, pp. 465–474, 2013.

[84] E. Giacoumidis, S. K. Ibrahim, J. Zhao, J. M. Tang, I. Tomkos, and A. D. Ellis, "Experimental Demonstration of Cost-Effective Intensity-modulation and Direct-detection Optical Fast-OFDM Over 40km SMF Transmission," in *OFC/NFOEC*, 2012, pp. 1–3.

[85] Z. Hu and C. Chan, "A 7-D Hyperchaotic System-Based Encryption Scheme for Secure Fast-OFDM-PON," *Journal of Lightwave Technology*, vol. 36, no. 16, pp. 3373–3381, 2018.

[86] Ming Li, J. Zhao, and Lian-Kuan Chen, "Blind Maximum-likelihood Frequency Offset Estimation for Coherent Fast OFDM Receivers," in *2013 18th OptoElectronics and Communications Conference held jointly with 2013 International Conference on Photonics in Switching (OECC/PS)*, 2013, pp. 1–2.

[87] X. Ouyang and J. Zhao, "Single-Tap Equalization for Fast OFDM Signals Under Generic Linear Channels," *IEEE Communications Letters*, vol. 18, no. 8, pp. 1319–1322, 2014.

[88] D. Wang, C. Lei, R. Luo, H. Chen, M. Chen, S. Yang, and S. Xie, "16QAM Optical Fast-OFDM Transmission System Based on Discrete Fourier Transform and Frequency Multiplexing," in *2014 OptoElectronics and Communication Conference and Australian Conference on Optical Fibre Technology*, 2014, pp. 445–446.

[89] X. Ouyang and J. Zhao, "Single Sideband Modulated Coherent Optical Fast-OFDM With Frequency Oversampling," *Journal of Lightwave Technology*, vol. 35, no. 12, pp. 2436–2443, 2017.

[90] X. Ouyang, W. Jia, P. Gunning, P. D. Townsend, and J. Zhao, "Experimental Demonstration and Field-Trial of an Improved Optical Fast OFDM Scheme Using Intensity-Modulation and Full-Field Detection," *Journal of Lightwave Technology*, vol. 33, no. 20, pp. 4353–4359, 2015.

[91] S. Vappangi and V. V. Mani, "Performance Analysis of Fast Optical OFDM for VLC," in *2017 20th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2017, pp. 206–211.

[92] M. Sung, J. Lee, and J. Jeong, "DCT-Precoding Technique in Optical Fast

OFDM for Mitigating Fiber Nonlinearity," *IEEE Photonics Technology Letters*, vol. 25, no. 22, pp. 2209–2212, 2013.

[93] Z. Hu, J. Zhao, Y. Hong, S. Gao, C. Chan, and L. Chen, "Set-Partitioned QAM Fast-OFDM with Real-Valued Orthogonal Circulant Matrix Transform Pre-Coding," in *2018 Conference on Lasers and Electro-Optics (CLEO)*, 2018, pp. 1–2.

[94] N. H. Nguyen, B. Berscheid, and H. H. Nguyen, "Fast-OFDM with Index Modulation for NB-IoT," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1157–1160, 2019.

[95] X. Ouyang, "Digital Signal Processing for Fiber-optic Communication Systems," Ph.D. dissertation, University College Cork, 2017.

[96] S. Isam and I. Darwazeh, "Design and Performance Assessment of Fixed Complexity Spectrally Efficient FDM Receivers," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*. IEEE, 2011, pp. 1–5.

[97] T. Xu, C. Masouros, and I. Darwazeh, "Waveform and Space Precoding for Next Generation Downlink Narrowband IoT," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5097–5107, 2019.

[98] T. Xu, R. C. Grammenos, F. Marvasti, and I. Darwazeh, "An Improved Fixed Sphere Decoder Employing Soft Decision for the Detection of Non-orthogonal Signals," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1964–1967, 2013.

[99] I. Kanaras, A. Chorti, M. R. Rodrigues, and I. Darwazeh, "Spectrally Efficient FDM Signals: Bandwidth Gain at the Expense of Receiver Complexity," in *2009 IEEE International Conference on Communications*. IEEE, 2009, pp. 1–6.

[100] S. Isam and I. Darwazeh, "Peak to Average Power Ratio Reduction in Spec-

trally Efficient FDM Systems," in *2011 18th International Conference on Telecommunications*. IEEE, 2011, pp. 363–368.

[101] H. Ghannam and I. Darwazeh, "Robust Channel Estimation Methods for Spectrally Efficient FDM Systems," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–6.

[102] S. Osaki, M. Nakao, T. Ishihara, and S. Sugiura, "Differentially Modulated Spectrally Efficient Frequency-Division Multiplexing," *IEEE Signal Processing Letters*, 2019.

[103] B. Yu, H. Zhang, X. Hong, C. Guo, A. P. T. Lau, C. Lu, and X. Dai, "Channel Equalisation and Data Detection for SEFDM over Frequency Selective Fading Channels," *IET Communications*, vol. 12, no. 18, pp. 2315–2323, 2018.

[104] R. G. Gallager, *Low-Density Parity-Check Codes*. MIT Press, 1963.

[105] X. Liu and I. Darwazeh, "Energy and Spectrally Efficient Signalling for Next Generation IoT," in *2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 2022, pp. 202–207.

[106] X. Zhang, L. Chen, J. Qiu, and J. Abdoli, "On The Waveform For 5G," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 74–80, 2016.

[107] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: Principles, Models and Technology Components*. Academic Press, 2018.

[108] X. Zhang, M. Jia, L. Chen, J. Ma, and J. Qiu, "Filtered-OFDM-enabler For Flexible Waveform in The 5th Generation Cellular Networks," in *2015 IEEE global communications conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.

[109] H. Jebbar, S. El Hassani, and A. El Abbassi, "Performance Study of 5G Multicarrier Waveforms," in *Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on*. IEEE, 2017, pp. 1–6.

[110] N. Michailow, M. Lentmaier, P. Rost, and G. Fettweis, "Integration of a GFDM Secondary System in an OFDM Primary System," in *Future Network & Mobile Summit (FutureNetw), 2011*. IEEE, 2011, pp. 1–8.

[111] E. Ozturk, E. Basar, and H. A. Cirpan, "Generalized Frequency Division Multiplexing with Index Modulation," in *Globecom Workshops (GC Wkshps), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[112] R. Datta, N. Michailow, M. Lentmaier, and G. Fettweis, "GFDM Interference Cancellation for Flexible Cognitive Radio PHY Design," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*. IEEE, 2012, pp. 1–5.

[113] M. M. Alavi-Sereshki, "Analytic Signals and Hilbert Tansforms," Ph.D. dissertation, Texas Tech University, 1972.

[114] E. Bedrosian, "The Analytic Signal Representation of Modulated Waveforms," *Proceedings of the IRE*, vol. 50, no. 10, pp. 2071–2076, 1962.

[115] M. Johansson, "The Hilbert Transform," *Mathematics Master's Thesis. Växjö University, Suecia. Disponible en internet: http://w3. msi. vxu. se/exarb/mj_ex. pdf, consultado el*, vol. 19, 1999.

[116] A. B. Carlson, *Communication System*. Tata McGraw-Hill Education, 2010.

[117] P. A. Haigh, P. Chvojka, S. Zvánovec, Z. Ghassemlooy, and I. Darwazeh, "Analysis of Nyquist Pulse Shapes for Carrierless Amplitude and Phase Modulation in Visible Light Communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 5023–5029, 2018.

[118] P. A. Haigh and I. Darwazeh, "Real Time Implementation of CAP Modulation With "Better-Than-Nyquist" Pulse Shaping in Visible Light Communications," *IEEE Communications Letters*, vol. 24, no. 4, pp. 840–843, 2020.

[119] S. Isam and I. Darwazeh, "Simple DSP-IDFT Techniques for Generating Spectrally Efficient FDM Signals," in *2010 7th International Symposium on*

Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010). IEEE, 2010, pp. 20–24.

[120] A. F. Shalash and K. K. Parhi, "Multidimensional Carrierless AM/PM Systems for Digital Subscriber Loops," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1655–1667, 1999.

[121] H. Ghannam and I. Darwazeh, "Signal Coding and Interference Cancellation of Spectrally Efficient FDM Systems for 5G Cellular Networks," in *Telecommunications (ICT), 2017 24th International Conference on.* IEEE, 2017, pp. 1–6.

[122] O. Vermesan, M. EisenHauer, M. Serrano, P. Guillemin, H. Sundmaeker, E. Z. Tragos, J. Valino, B. Copigneaux, M. Presser, A. Aagaard *et al.*, "The Next Generation Internet of Things–Hyperconnectivity and Embedded Intelligence at the Edge," *Next Generation Internet of Things. Distributed Intelligence at the Edge and Human Machine-to-Machine Cooperation*, 2018.

[123] *Ericsson Mobility Report*, Ericsson, Nov.2021.

[124] T. Xu and I. Darwazeh, "Non-orthogonal narrowband internet of things: A design for saving bandwidth and doubling the number of connected devices," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2120–2129, 2018.

[125] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow Band Internet of Things," *IEEE Access*, vol. 5, pp. 20 557–20 577, 2017.

[126] N. C. Beaulieu, C. C. Tan, and M. O. Damen, "A" Better than" Nyquist Pulse," *IEEE Communications Letters*, vol. 5, no. 9, pp. 367–368, 2001.

[127] G. D. Forney, "Trellis Shaping," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 281–300, 1992.

[128] P. Schulte and G. Böcherer, "Constant Composition Distribution Matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, 2015.

[129] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental Demonstration of Capacity Increase and Rate-adaptation by Probabilistically Shaped 64-QAM," in *2015 European Conference on Optical Communication (ECOC)*, 2015, pp. 1–3.

[130] Y. C. Gültekin, W. J. van Houtum, A. G. C. Koppelaar, F. M. J. Willems, and W. J. van Houtum, "Enumerative Sphere Shaping for Wireless Communications With Short Packets," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 1098–1112, 2020.

[131] W. Kang, H. Che, and K. Zhu, "Precoding Optimization for Faster-Than-Nyquist Signaling With Probabilistic Shaping," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 470–11 478, 2020.

[132] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic Shaping and Forward Error Correction for Fiber-optic Communication Systems," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 230–244, 2019.

[133] H. Ghannam and I. Darwazeh, "SEFDM over Satellite Systems with Advanced Interference Cancellation," *IET Communications*, vol. 12, no. 1, pp. 59–66, 2017.

[134] *Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications; Part 1 (DVB-S2)*, ETSI Standard EN 302 307-1 v1.4.1, Nov.2014.

[135] T. Xu, R. C. Grammenos, F. Marvasti, and I. Darwazeh, "An Improved Fixed Sphere Decoder Employing Soft Decision for the Detection of Non-orthogonal Signals," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1964–1967, 2013.

[136] P. N. Whatmough, M. R. Perrett, S. Isam, and I. Darwazeh, "VLSI Architecture for a Reconfigurable Spectrally Efficient FDM Baseband Transmitter," in

2011 IEEE International Symposium of Circuits and Systems (ISCAS), 2011, pp. 1688–1691.

[137] T. S. Rappaport *et al.*, *Wireless Communications: Principles and Practice*. prentice hall PTR New Jersey, 1996, vol. 2.

[138] X. Wang, P. Ho, and Y. Wu, "Robust Channel Estimation and ISI Cancellation for OFDM Systems with Suppressed Features," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 5, pp. 963–972, 2005.

[139] W. Rummler, R. Coutts, and M. Liniger, "Multipath Fading Channel Models for Microwave Digital Radio," *IEEE Communications Magazine*, vol. 24, no. 11, pp. 30–42, 1986.

[140] H. R. Anderson, *Fixed Broadband Wireless System Design*. John Wiley & Sons, 2003.

[141] M. Morelli and U. Mengali, "A Comparison of Pilot-aided Channel Estimation Methods for OFDM Systems," *IEEE Transactions on signal processing*, vol. 49, no. 12, pp. 3065–3073, 2001.

[142] Y. Shen and E. Martinez, "Channel Estimation in OFDM Systems," *Freescale semiconductor application note*, pp. 1–15, 2006.

[143] J.-J. Van De Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On Channel Estimation in OFDM Systems," in *1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, vol. 2. IEEE, 1995, pp. 815–819.

[144] B. Muquet, M. De Courville, and P. Duhamel, "Subspace-based Blind and Semi-blind Channel Estimation for OFDM Systems," *IEEE Transactions on signal processing*, vol. 50, no. 7, pp. 1699–1712, 2002.

[145] S. Zhou and G. B. Giannakis, "Finite-alphabet Based Channel Estimation for OFDM and Related Multicarrier Systems," *IEEE Transactions on Communications*, vol. 49, no. 8, pp. 1402–1414, 2001.

[146] T. Y. Al-Naffouri, A. A. Dahman, M. S. Sohail, W. Xu, and B. Hassibi, "Low-complexity Blind Equalization for OFDM Systems with General Constellations," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6395–6407, 2012.

[147] V. Jones and G. C. Raleigh, "Channel Estimation for Wireless OFDM Systems," in *IEEE GLOBECOM 1998 (Cat. NO. 98CH36250)*, vol. 2. IEEE, 1998, pp. 980–985.

[148] S. Isam and I. Darwazeh, "Robust Channel Estimation for Spectrally Efficient FDM System," in *2012 19th International Conference on Telecommunications (ICT)*. IEEE, 2012, pp. 1–6.

[149] T. Xu, S. Mikroulis, J. E. Mitchell, and I. Darwazeh, "Bandwidth Compressed Waveform for 60-GHz Millimeter-wave Radio over Fiber Experiment," *Journal of Lightwave Technology*, vol. 34, no. 14, pp. 3458–3465, 2016.

[150] Y. Wang, Y. Zhou, T. Gui, K. Zhong, X. Zhou, L. Wang, A. P. T. Lau, C. Lu, and N. Chi, "SEFDM Based Spectrum Compressed VLC System Using RLS Time-domain Channel Estimation and ID-FSD Hybrid Decoder," in *ECOC 2016; 42nd European Conference on Optical Communication*. VDE, 2016, pp. 1–3.

[151] L. Hanzo and T. Keller, "OFDM and MC-CDMA: A Primer," 2006.

[152] R. Negi and J. Cioffi, "Pilot Tone Selection for Channel Estimation in a Mobile OFDM System," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, 1998.

[153] "IEEE Standard for Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High Speed Physical Layer in The 5 GHz Band," *IEEE Std 802.11a-1999*, pp. 1–102, 1999.

[154] C. Kim, Y. H. Yun, K. Kim, and J.-Y. Seol, "Introduction to QAM-FBMC: From Waveform Optimization to System Design," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 66–73, 2016.

[155] D. Nopchinda, T. Xu, R. Maher, B. C. Thomsen, and I. Darwazeh, "Dual Polarization Coherent Optical Spectrally Efficient Frequency Division Multiplexing," *IEEE Photonics Technology Letters*, vol. 28, no. 1, pp. 83–86, 2015.

[156] Yang-Seok Choi, P. J. Voltz, and F. A. Cassara, "On Channel Estimation and Detection for Multicarrier Signals in Fast and Selective Rayleigh Fading Channels," *IEEE Transactions on Communications*, vol. 49, no. 8, pp. 1375–1387, 2001.

[157] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[158] J. Conway and N. Sloane, "A Fast Encoding Method for Lattice Codes and Quantizers," *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 820–824, 1983.

[159] A. Gersho and V. Lawrence, "Multidimensional Signal Constellations for Voiceband Data Transmission," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 687–702, 1984.

[160] E. Agrell and M. Karlsson, "Power-efficient Modulation Formats in Coherent Transmission Systems," *Journal of Lightwave Technology*, vol. 27, no. 22, pp. 5115–5126, 2009.

[161] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Springer Science & Business Media, 2013, vol. 290.

[162] D. Li, *Signal Constellation Design for Gaussian and Non-Gaussian Channels*. Rice University, 1995.

[163] J. Boutros, E. Viterbo, C. Rastello, and J.-C. Belfiore, "Good Lattice Constellations for Both Rayleigh Fading and Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 502–518, 1996.

[164] D. Slepian, "Permutation Modulation," *Proceedings of the IEEE*, vol. 53, no. 3, pp. 228–236, 1965.

[165] T. Ericson and V. Zinoviev, "Spherical Codes Generated by Binary Partitions of Symmetric Pointsets," *IEEE transactions on information theory*, vol. 41, no. 1, pp. 107–129, 1995.

[166] P. Liu and A. Springer, "Space Shift Keying for LOS Communication at mmWave Frequencies," *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 121–124, 2014.

[167] D. A. Basnayaka, M. Di Renzo, and H. Haas, "Massive But Few Active MIMO," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 6861–6877, 2015.

[168] R. Mesleh, R. Mehmood, H. Elgala, and H. Haas, "Indoor MIMO Optical Wireless Communication Using Spatial Modulation," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–5.

[169] "Pindex modulation techniques for 5g wireless networks."

[170] E. Basar, M. Wen, R. Mesleh, M. Di Renzo, Y. Xiao, and H. Haas, "Index Modulation Techniques for Next-generation Wireless Networks," *IEEE access*, vol. 5, pp. 16 693–16 746, 2017.

[171] T. Mao, Z. Wang, Q. Wang, S. Chen, and L. Hanzo, "Dual-mode Index Modulation Aided OFDM," *IEEE Access*, vol. 5, pp. 50–60, 2016.

[172] M. Nakao, T. Ishihara, and S. Sugiura, "Dual-mode Time-domain Index Modulation for Nyquist-criterion and Faster-than-Nyquist Single-carrier Transmissions," *IEEE Access*, vol. 5, pp. 27 659–27 667, 2017.

[173] T. Ishihara and S. Sugiura, "Faster-than-Nyquist Signaling with Index Modulation," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 630–633, 2017.

[174] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design Guidelines for Spatial Modulation," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 6–26, 2014.

[175] C. E. Shannon, "Communication in The Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[176] J. L. Massey, "Coding and Modulation in Digital Communications," *Rec. Int. Zurich Sem. Digital Commun., Zurich, Switzerland. Mar.*, 1974.

[177] G. Ungerboeck, "Trellis-coded Modulation with Redundant Signal Sets Part I: Introduction," *IEEE Communications magazine*, vol. 25, no. 2, pp. 5–11, 1987.

[178] A. J. Viterbi, J. K. Wolf, E. Zehavi, and R. Padovani, "A Pragmatic Approach to Trellis-coded Modulation," *IEEE Communications Magazine*, vol. 27, no. 7, pp. 11–19, 1989.

[179] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error-correcting Codes," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 371–377, 1977.

[180] ——, "Correction to'A New Multilevel Coding Method Using Error-Correcting Codes'," *IEEE Transactions on Information Theory*, vol. 23, no. 6, pp. 784–784, 1977.

[181] E. Zehavi, "8-PSK Trellis Codes for a Rayleigh Channel," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 873–884, 1992.

[182] L. Szczecinski and A. Alvarado, *Bit-interleaved Coded Modulation: Fundamentals, Analysis and Design*. John Wiley & Sons, 2015.

[183] A. G. i Fàbregas, A. Martinez, and G. Caire, *Bit-interleaved Coded Modulation*. Now Publishers Inc, 2008, vol. 5, no. 1-2.

[184] X. Zhou, L. E. Nelson, P. Magill, R. Isaac, B. Zhu, D. W. Peckham, P. I. Borel, and K. Carlson, "High Spectral Efficiency 400 Gb/s Transmission Using PDM Time-Domain Hybrid 32–64 QAM and Training-assisted Carrier Recovery," *Journal of Lightwave technology*, vol. 31, no. 7, pp. 999–1005, 2013.

[185] L. R. Rabiner, "The Design of Finite Impulse Response Digital Filters using Linear Programming Techniques," *Bell System Technical Journal*, vol. 51, no. 6, pp. 1177–1198, 1972.

[186] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[187] M. Beko and R. Dinis, "Designing Good Multi-dimensional Constellations," *IEEE wireless communications letters*, vol. 1, no. 3, pp. 221–224, 2012.

[188] H. Khoshnevis, I. Marsland, and H. Yanikomeroglu, "Design of High-SNR Multidimensional Constellations for Orthogonal Transmission in a Nakagami-$\{m\}$ Fading Channel," *IEEE Access*, vol. 5, pp. 26 623–26 638, 2017.

[189] F. Kayhan and G. Montorsi, "Constellation Design for Memoryless Phase Noise Channels," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2874–2883, 2014.

[190] K. Zeger and A. Gersho, "Pseudo-gray Coding," *IEEE Transactions on communications*, vol. 38, no. 12, pp. 2147–2158, 1990.

[191] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimized Symbol Mappings for Bit-interleaved Coded Modulation with Iterative Decoding," in *GLOBECOM'03. IEEE Global Telecommunications Conference (IEEE Cat. No. 03CH37489)*, vol. 6. IEEE, 2003, pp. 3316–3320.

[192] J. H. Holland, “Genetic Algorithms,” *Scientific american*, vol. 267, no. 1, pp. 66–73, 1992.

[193] S. Katoch, S. S. Chauhan, and V. Kumar, “A Review on Genetic Algorithm: Past, Present, and Future,” *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2021.

[194] A. Misevicius and E. Guogis, “Computational Study of Four Genetic Algorithm Variants for Solving the Quadratic Assignment Problem,” in *International Conference on Information and Software Technologies*. Springer, 2012, pp. 24–37.

[195] Z. H. Ahmed, “An Improved Genetic Algorithm using Adaptive Mutation Operator for the Quadratic Assignment Problem,” in *2015 38th International conference on telecommunications and signal processing (TSP)*. IEEE, 2015, pp. 1–5.

[196] A. Angioi, M. Lixia, and M. Murroni, “Optimized APSK Bit Allocation for Satellite Communication,” in *2010 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop*. IEEE, 2010, pp. 407–412.

[197] A. Meloni and M. Murroni, “On the Genetic Optimization of APSK Constellations for Satellite Broadcasting,” in *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*. IEEE, 2014, pp. 1–6.

[198] M. C. Valenti, R. Doppalapudi, and D. Torrieri, “A Genetic Algorithm for Designing Constellations with Low Error Floors,” in *2008 42nd Annual Conference on Information Sciences and Systems*. IEEE, 2008, pp. 1155–1160.

[199] S. S. Patel, T. Quazi, and H. Xu, “A Genetic Algorithm for Designing Uncoded Space-time Labelling Diversity Mappers,” in *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*. IEEE, 2018, pp. 1–6.

[200] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by Simulated Annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[201] P. M. Pardalos, L. Pitsoulis, T. Mavridou, and M. G. Resende, "Parallel Search for Combinatorial Optimization: Genetic Algorithms, Simulated Annealing, Tabu Search and GRASP," in *International Workshop on Parallel Algorithms for Irregularly Structured Problems*. Springer, 1995, pp. 317–331.

[202] D. Delahaye, S. Chaimatanan, and M. Mongeau, "Simulated Annealing: From Basics to Applications," in *Handbook of metaheuristics*. Springer, 2019, pp. 1–35.

[203] A. Gamal, L. Hemachandra, I. Shperling, and V. Wei, "Using Simulated Annealing to Design Good Codes," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 116–123, 1987.

[204] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 799–809, 1990.

[205] F. Kayhan and G. Montorsi, "Constellation Design for Transmission over Nonlinear Satellite Channels," in *2012 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2012, pp. 3401–3406.

[206] ——, "Joint Signal-labeling Optimization Under Peak Power Constraint," *International Journal of Satellite Communications and Networking*, vol. 30, no. 6, pp. 251–263, 2012.

[207] Y. Huang and J. A. Ritcey, "Optimal Constellation Labeling for Iteratively Decoded Bit-interleaved Space-time Coded Modulation," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1865–1871, 2005.

[208] T. O'shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[209] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep Learning for Physical-layer 5G Wireless Techniques: Opportunities, Challenges and Solutions," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 214–222, 2019.

[210] D. Burth Kurka and D. Gündüz, "Joint Source-channel Coding of Images with (not very) Deep Learning," in *International Zurich Seminar on Information and Communication (IZS 2020). Proceedings*. ETH Zurich, 2020, pp. 90–94.

[211] Y. Jiang, H. Kim, H. Asnani, S. Kannan, S. Oh, and P. Viswanath, "Turbo Autoencoder: Deep Learning based Channel Codes for Point-to-point Communication Channels," *Advances in neural information processing systems*, vol. 32, 2019.

[212] F. A. Aoudia and J. Hoydis, "End-to-end Waveform Learning Through Joint Optimization of Pulse and Constellation Shaping," in *2021 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2021, pp. 1–6.

[213] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep Learning-based Channel Estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, 2019.

[214] H. Ye, G. Y. Li, and B.-H. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[215] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[216] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep Learning for Wireless Physical Layer: Opportunities and Challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, 2017.

[217] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, “Deep Learning in Physical Layer Communications,” *IEEE Wireless Communications*, vol. 26, no. 2, pp. 93–99, 2019.

[218] G. Bonaccorso, *Machine Learning Algorithms*. Packt Publishing Ltd, 2017.

[219] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[220] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The Expressive Power of Neural Networks: A View From the Width,” *Advances in neural information processing systems*, vol. 30, 2017.

[221] H. Robbins and S. Monro, “A Stochastic Approximation Method,” *The annals of mathematical statistics*, pp. 400–407, 1951.

[222] L. Bottou, “Large-scale Machine Learning with Stochastic Gradient Descent,” in *Proceedings of COMPSTAT’2010*. Springer, 2010, pp. 177–186.

[223] B. T. Polyak, “Some Methods of Speeding Up the Convergence of Iteration Methods,” *Ussr computational mathematics and mathematical physics*, vol. 4, no. 5, pp. 1–17, 1964.

[224] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[225] S. Sun, Z. Cao, H. Zhu, and J. Zhao, “A Survey of Optimization Methods from a Machine Learning Perspective,” *IEEE transactions on cybernetics*, vol. 50, no. 8, pp. 3668–3681, 2019.

[226] G. Swirszcz, W. M. Czarnecki, and R. Pascanu, “Local Minima in Training of Deep Networks,” *arXiv preprint arXiv:1611.06310*, 2016.

[227] M. Tschannen, O. Bachem, and M. Lucic, “Recent Advances in Autoencoder-based Representation Learning,” *arXiv preprint arXiv:1812.05069*, 2018.

[228] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du, and S. Marshall, "Novel Segmented Stacked Autoencoder for Effective Dimensionality Reduction and Feature Extraction in Hyperspectral Imaging," *Neurocomputing*, vol. 185, pp. 1–10, 2016.

[229] J. Cho and P. J. Winzer, "Probabilistic Constellation Shaping for Optical Fiber Communications," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019.

[230] Z. Qu, S. Zhang, and I. B. Djordjevic, "Universal Hybrid Probabilistic-geometric Shaping Based on Two-dimensional Distribution Matchers," in *Optical Fiber Communication Conference*. Optical Society of America, 2018, pp. M4E–4.

[231] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On Shaping Gain in the Nonlinear Fiber-optic Channel," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 2794–2798.

[232] P. Schulte, F. Steiner, and G. Bocherer, "Four Dimensional Probabilistic Shaping for Fiber-optic Communication," in *Signal Processing in Photonic Communications*. Optical Society of America, 2017, pp. SpM2F–5.

[233] Q. Guo, W.-R. Peng, Y. Cui, and Y. Bai, "Multi-dimensional Probabilistic Shaping for Higher Fibre Nonlinearity Tolerance," in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–3.

[234] A. Amari and A. Richter, "Multi-dimensional Short Blocklength Probabilistic Shaping for Digital Subcarrier Multiplexing Systems," *arXiv preprint arXiv:2105.04956*, 2021.

[235] M. Stark, F. A. Aoudia, and J. Hoydis, "Joint Learning of Geometric and Probabilistic Constellation Shaping," in *2019 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2019, pp. 1–6.

[236] E. Jang, S. Gu, and B. Poole, "Categorical Reparameterization with Gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[237] "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: https://www.tensorflow.org/

[238] E. Giacoumidis, S. K. Ibrahim, J. Zhao, J. Tang, A. D. Ellis, and I. Tomkos, "Experimental and Theoretical Investigations of Intensity-modulation and Direct-detection Optical Fast-OFDM over MMF-links," *IEEE Photonics Technology Letters*, vol. 24, no. 1, pp. 52–54, 2011.

[239] 3GPP TS 38.901 version 14.3.0, "5G; Study on Channel Model for Frequencies from 0.5 to 100 GHz," Rel. 14, Jan. 2018.

[240] H. Ghannam and I. Darwazeh, "SEFDM over Satellite Systems with Advanced Interference Cancellation," *IET Communications*, vol. 12, no. 1, pp. 59–66, 2018.

[241] Q. Cappart, D. Chételat, E. Khalil, A. Lodi, C. Morris, and P. Veličković, "Combinatorial Optimization and Reasoning with Graph Neural Networks," *arXiv preprint arXiv:2102.09544*, 2021.

[242] F. R Kschischang, "The Hilbert Transform," Ph.D. dissertation, The Edward S. Rogers SR. Department of Electrical and Computer Engineering, Univeristy of Toronto, 2006.

[243] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.