Applications of Non-Orthogonal Waveforms and Artificial Neural Networks in Wireless Vehicular Communications

Scott A. Stainton

School of Engineering
Newcastle University

This thesis is submitted for the degree of
Doctor of Philosophy

May 2022

To my loving parents and family

suos cultores scientia coronat

knowledge crowns those who seek her

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements.

Scott A. Stainton
May 2022

Acknowledgements

First of all, I would like to thank everyone I have met over the years and offer my apologies to anyone that I may have missed. It is all of you who have helped shape my journey and made it what it was.

I would like to acknowledge all of the financial support I have received, including my full PhD scholarship from the university and the EPSRC, and the grants that helped fund the equipment and experiments that made my work possible.

I would like to thank all of my supervisory team for their support, I owe you all a tremendous debt of gratitude and no amount of words here can capture all of the things you have done for me, both personally and professionally. Dr Martin Johnston, who has been generous enough to support me since long before my PhD as an undergraduate, and was always on hand to offer advice and ideas when things didn't always work out. Prof Satnam Dlay, who was always there to offer wisdom and advice on far more than just my PhD, and whose words will stay with me long beyond my time at the university. Dr Paul Haigh, who generously agreed to join my supervisory team not long after joining the research group, and proceeded to go above and beyond what is expected of a supervisor. Without your guidance and feedback at a difficult time, none of this would have been possible. I would also like to acknowledge Dr Rajesh Tiwari, whose knowledge of vehicular communications helped form the early days of my studies.

A huge thanks goes out to all of my friends and colleagues in the research office for making my time at the university so special, it was a pleasure to work alongside you all these years. I would also like to personally thank Shaun Barney. Thank you for all of the distractions, the discussions, and our time outside of the office.

A sincere thank you to my partner Siobhan, and my siblings Anna, and Alex. You have been constants in an ever changing world.

Any finally, a special thanks to my parents who have always believed in me and are the reason I am where I am today. This is for you.

Abstract

We live in an ever increasing world of connectivity. The need for highly robust, highly efficient wireless communication has never been greater. As we seek to squeeze better and better performance from our systems, we must remember; even though our computing devices are increasing in power and efficiency, our wireless spectrum remains limited.

Recently there has been an increasing trend towards the implementation of machine learning based systems in wireless communications. By taking advantage of a neural networks powerful non-linear computational capability, communication systems have been shown to achieve reliable error free transmission over even the most dispersive of channels. Furthermore, in an attempt to make better use of the available spectrum, more spectrally efficient physical layer waveforms are gathering attention that trade increased interference for lower bandwidth requirements. In this thesis, the performance of neural networks that utilise spectrally efficient waveforms within harsh transmission environments are assessed.

Firstly, we investigate and generate a novel neural network for use within a standards compliant vehicular network for vehicle-to-vehicle communication, and assess its performance practically in several of the harshest recorded empirical channel models using a hardware-in-the-loop testing methodology. The results demonstrate the strength of the proposed receiver, achieving a bit-error rate below 10^{-3} at a signal-to-noise ratio (SNR) of 6dB.

Secondly, this is then further extended to utilise spectrally efficient frequency division multiplexing (SEFDM), where we note a break away from the 802.11p vehicular communication standard in exchange for a more efficient use of the available spectrum that can then be utilised to service more users or achieve a higher data throughput. It is demonstrated that the proposed neural network system is able to act as a joint channel equaliser and symbol receiver with bandwidth compression of up to 60% when compared to orthogonal frequency division multiplexing (OFDM). The effect of overfitting to the training environment is also tested, and the proposed system is

shown to generalise well to unseen vehicular environments with no notable impact on the bit-error rate performance.

Thirdly, methods for generating inputs and outputs of neural networks from complex constellation points are investigated, and it is reasoned that creating 'split complex' neural networks should not be preferred over 'contatenated complex' neural networks in most settings. A new and novel loss function, namely error vector magnitude (EVM) loss, is then created for the purposes of training neural networks in a communications setting that tightly couples the objective function of a neural network during training to the performance metrics of transmission when deployed practically. This loss function is used to train neural networks in complex environments and is then compared to popular methods from the literature where it is demonstrated that EVM loss translates better into practical applications. It achieved the lowest EVM error, thus bit-error rate, across all experiments by a margin of 3dB when compared to its closest achieving alternative. The results continue and show how in the experiment EVM loss was able to improve spectral efficiency by 67% over the baseline without affecting performance.

Finally, neural networks combined with the new EVM loss function are further tested in wider communication settings such as visible light communication (VLC) to validate the efficacy and flexibility of the proposed system. The results show that neural networks are capable of overcoming significant challenges in wireless environments, and when paired with efficient physical layer waveforms like SEFDM and an appropriate loss function such as EVM loss are able to make good use of a congested spectrum. The authors demonstrated for the first time in practical experimentation with SEFDM that spectral efficiency gains of up to 50% are achievable, and that previous SEFDM limitations from the literature with regards to number of subcarriers and size of the transmit constellation are alleviated via the use of neural networks.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

5G 5th Generation

ADC Analogue-to-Digital Converter

ANN Artificial Neural Network

AWGN Additive White Gaussian Noise

BER Bit-Error Rate

BSS Basic Service Set

CDF Cumulative Distribution Function

CP Cyclic Prefix

CSI Channel State Information

DAC Digital-to-Analogue Converter

dB Decibel

DFT Discrete Fourier Transform

DSRC Dedicated Short Range Communication

ECDF Empirical Cumulative Distribution Function

EVM Error Vector Magnitude

FEC Forward Error Correction

FFT Fast Fourier Transform

GFDM Generalised Frequency Division Multiplexing

GSD Generalised Sphere Decoding

IBSS Independent Basic Service Set

IBSS User Datagram Protocol

ICI Inter-Carrier Interference

IDFT Inverse Discrete Fourier Transform

IFFT Inverse Fast Fourier Transform

IFI Inter Frame Interference

IP Internet Protocol

ISI Inter Symbol Interference

ITS Intelligent Transport Systems

LED Light Emitting Diode

LOS Line Of Sight

LTI Linear Time Invariant

ML Maximum Likelihood

MMSE Minimum Mean Square Error

MSE Mean Square Error

NAG Nesterov Accelerated Gradient

NLOS Non-Line Of Sight

OFDM Orthogonal Frequency Division Multiplexing

PAM Phase Amplitude Modulation

PCA Principal Component Analysis

PDF Probability Density Function

PDP Power Delay Profile

PSD Power Spectral Density

QAM Quadrature Amplitude Modulation

QPSK Quadrature Phase Shift Keying

RF Radio Frequency

RMS Root Mean Square

RRC Root Raised Cosine

SAE Society for Automotive Engineers

SD Sphere Decoding

SEFDM Spectrally Efficient Frequency Division Multiplexing

SGD Stochastic Gradient Descent

SIC Self Interference Cancellation

SI Self Interference

SNE Stochastic Neighbour Embedding

SNR Signal to Noise Ratio

SSID Service Set ID

SS Service Set

TCP Transmission Control Protocol

t-SNE t-Distributed Stochastic Neighbour Embedding

USRP Universal Software Radio Peripheral

V2I Vehicle-To-Infrastructure

V2V Vehicle-To-Vehicle

V2X Vehicle-To-Anything

VANET Vehicular Ad-Hoc Network

VLC Visible Light Communication

WAVE Wireless Access in Vehicular Environments

WSMP WAVE Short Message Protocol

ZF Zero Forcing

Chapter 1

Introduction

Autonomous vehicles are receiving a great deal of attention, with many automotive companies expending time, money and resources to research and development in this area. However, there are still many challenges that need to be addressed before autonomous cars can be allowed on the public highway. With the anticipation of the 5G network to be in widespread deployment by 2022 [1] we are expecting a new wave of connected devices; data sharing on a massive scale will become common place and many new applications will appear that were not feasible with current technology. Many manufacturers are getting closer to providing autonomous vehicles [2–4] and some of the last hurdles they are facing are from legal and political issues rather than technological. Besides other uses for improved wireless connectivity in vehicles such as infotainment [5, 6] and traffic and congestion management [7, 8], over the years we have seen more and more autonomous functionality being brought into the individual vehicles to aid driver safety [9]. The next step in this evolution is to include information from vehicles in the local vicinity into the autonomous decision making processes. There have been many works that propose partial solutions such as the automated braking proposal of [10] or 'over the horizon awareness' presented in [11] (a review of more can be seen in [12]) however, ultimately what is a required is a joint and concerted effort from academia, industry, and legislators. At the time of writing, the government bodies responsible for allocating the frequency spectrum in the United States, the FCC, and in the United Kingdom, Ofcom, have reserved the frequency ranges 5.850 GHz–5.855 GHz [13] and 5.850 GHz–5.925 GHz [14] respectively for intelligent transport systems (ITS) services. Furthermore, the IEEE working group 1609 have produced the wireless access in vehicular environments (WAVE) standard for dedicated short range communication (DSRC) cooperatively with the society of automotive engineers (SAE) [15] built on top of the 802.11 framework [14]. The key

take-away being that society is being shaped to prepare for the arrival of ITS services, therefore the research we do in this field has the potential for direct societal impact in the immediate future.

1.1 Research Aims and Motivation

The impact we aim to impart through this research relates to the advancement of spectrally efficient modulation schemes in highly mobile vehicular wireless environments which suffer from severe signal degradation, ultimately leading to high error rates. Futhermore, there is a direct need to increase driver safety, therefore in order to reduce the human error either the autonomous controls or the information provided to the driver needs to incorporate more data from the local surroundings and other vehicles in the vicinity. This allows drivers to act accordingly and respond to events in good time, taking a proactive rather than reactive approach to safety. This indicates a need for highly robust, efficient communication in what is known to be a very harsh and unforgiving wireless environment.

From this we can derive a few key aims that the research presented hereafter will seek to explore. The first of which is the need for spectrally efficient communication methodologies, the current United Kingdom radio frequency (RF) spectrum is highly overcrowded, Fig. 1.1 shows the extent to which the available spectrum is subscribed. This is compounded by the fact that the United Kingdom (at the time of writing) has upwards of 38.4 million licensed vehicles on the road [16], thus any communication standard that requires broadcasting of potentially safety critical information needs to be able to handle large volumes of data exchange. We will approach this by investigating, designing, and implementing novel physical layer waveforms. Second is the highlighted need for highly robust communication. As mentioned previously, the environment for vehicle-to-vehicle (V2V) communication can be very harsh and severely degrade the reliability and performance. A secondary, but very serious added consequence of the failure of communication of safety critical information is that there will be higher than normal occurences of poor channel allocation and in some instances network flooding of repeated failed messages [17]. This research will address this issue by investigating, designing, and implementing novel wireless receivers via the use of artificial neural networks.

The latest report (at time of writing) from the United Kingdom Department for Transport [18] provides a lot of context to the motivation for this work. In 2019, there were $1,752$ reported deaths from road traffic accidents. One can see from Fig. 1.2 that

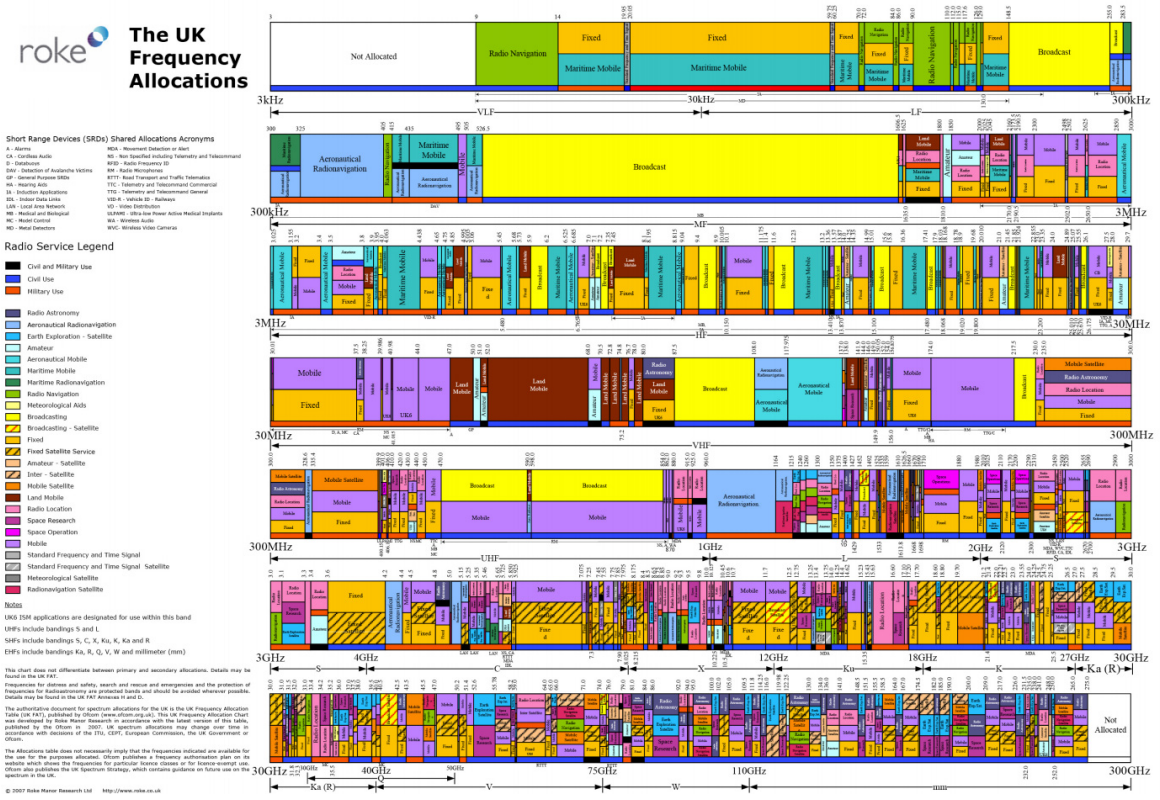

Fig. 1.1 The UK radio frequency spectrum allocation

this number has not changed meaningfully in almost a decade. Whilst vehicle safety improved in the period from 2004 to 2012, it has remained stagnant ever since. To add to this, besides fatalities, in 2019 alone there was a total of 153,158 casualties of road traffic accidents. A study performed in [19] examined the underlying causes of UK road traffic accidents, where they state that the so-called 'unholy trinity' of speed, alcohol, and lack of safety restraint were factors in around two thirds of fatal incidents. In fact, the top 6 factors that account for almost all incidents share one thing in common - human nature. Whilst non-human factors such as weather and road conditions are certainly important, it is the human element that ultimately makes an error that in many cases could have been avoided.

These two factors, namely the requirement for spectrally efficient communication due to RF spectrum overcrowding, and the requirement for robust communication in the face of adverse environments such as those seen in vehicular environments, are the motivation for this work. As a result, the following main themes will occur throughout this thesis:

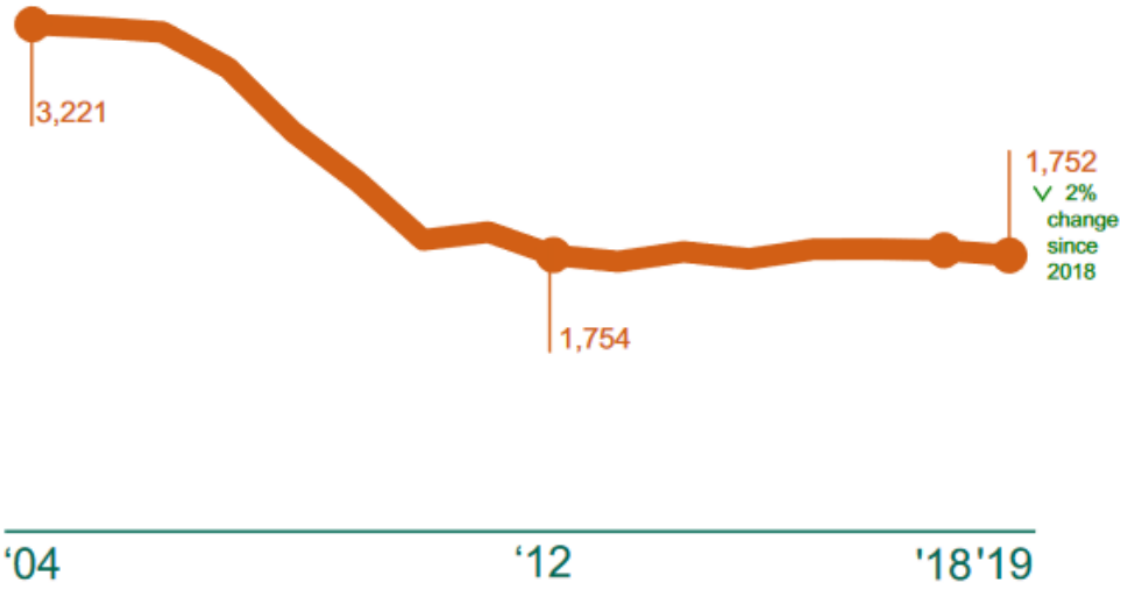

Fig. 1.2 Road traffic accidents in the UK from 2004 to 2019 (from [18])

- Artificial neural networks are proposed and investigated for use as joint channel equalisers and symbol receivers in communications.
- The methodology for generating and training neural networks for communications are investigated, and a new loss function for neural networks is proposed that aligns the training objective with the metrics used in deployment.
- Spectrally efficient waveforms are investigated as a means of both easing the congestion of the limited spectrum resource, and increasing the potential number of serviceable users in densely populated vehicular networks.
- The combination of these methods will be tested in multiple scenarios, to ensure the algorithms and techniques developed scale to other areas of communication, for example, in visible light communications.
- Where possible all experimentation will be undertaken in a practical manner to ensure that conditions are as close to real world deployment environments as possible.

1.2 Contribution

This thesis contributes to three main aspects that overcome significant obstacles and improve the performance of communications in both wired and wireless formats.

- The design and implementation of a novel neural network capable of being deployed in multiple harsh wireless vehicle-to-vehicle communication scenarios. It was proved practically that the neural network could improve communication throughput, achieving an acceptable bit-error rate at a higher spectral efficiency when compared to the baseline IEEE 802.11p implementation. The challenge of a change of environment for the neural network was investigated and it was shown that the proposed neural network was able to perform with no significant detriment to baseline performance.
- The design and implementation of a new loss function for machine learning in a communication setting based on the error vector magnitude. It was reasoned that more tightly coupling the objective function of the training process to the goal of the resulting neural network leads to a significant performance improvement in real world applications. This was demonstrated practically, where the proposed system was able to outperform mean square error variants of the same networks in all experiments by a minimum of 2dB.
- Investigation of the more widespread applicability of novel neural network architectures and cost functions previously derived for vehicle-to-vehicle communication in optical communications. It was shown practically that the benefits of the proposed methods from this work tranferred to other domains of communication, testing in a VLC network demonstrated spectral savings comparable to that of the originally proposed wireless vehicular networks.

1.3 Publications

The following publications have been produced during the tenure of this research. There are related works listed outside of communications included for completeness, such as signal processing and machine learning, as the application of certain topics of this research became applicable in other domains.

1. **Stainton, S.A.** and Haigh, P.A. “Doubling the Spectral Efficiency with EVM as the Objective Function for Training Neural Networks in Non-Orthogonal

Visible Light Communications Systems" *The Optical Fiber and Communication Conference*, 2021.

2. **Stainton, S.A.**, Johnston, M., Dlay, S., and Haigh, P.A., "EVM Loss: A Loss Function for Training Neural Networks in Communication Systems" *MDPI Special Issue on Visible Light Communication, Networking, and Sensing*, 2021, vol. 21, pp. 1094-1103.

3. **Stainton, S.A.**, W. Ozan, M. Johnston, S. Dlay and P. A. Haigh, "Neural Network Equalisation and Symbol Detection for 802.11p V2V Communication at 5.9GHz," *IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1-5.

4. **Stainton, S.A.**, Johnston, M., and Dlay, S. "VANET Group Sizing for GFDM Relaying in Realistic Vehicular Networks" *UK/China Emerging Technologies (UCET)*, 2019, pp. 1-2.

5. **Stainton, S.A.**, Tsimenidis, C., and Murray, A. "Characteristics of Phonocardiography Waveforms that Influence Automatic Feature Recognition" *Computing in Cardiology (CinC)*, 2016, pp. 1173-1176.

6. **Stainton, S.A.**, Barney, S., Catt, M., and Dlay, S. "On the Application of Quantization for Mobile Optimized Convolutional Neural Networks as a Predictor of Realtime Ageing Biomarkers" *11th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 2018, pp. 1-5.

7. Barney, S., **Stainton, S.A.**, Catt, M., and Dlay, S. "Feature Specific Analysis of a Deep Convolutional Neural Network for Ageing Classification" *11th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 2018, pp. 1-6.

8. Gulati, I., Li, H., **Stainton, S.A.**, Johnston, M., and Dlay, S. "Investigation of Ionospheric Phase Scintillation at Middle-Latitude Receiver Station" *International Symposium ELMAR*, 2019, pp. 191-194.

9. Li, H., Gulati, I., **Stainton, S.A.**, Saleh, A.A, Pickert, V, and Dlay, S. "Sliding Mode Control for Vehicular Platoon based on V2V Communication" *32nd International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+)*, 2019, pp. 2078-2089.

10. Sun, Y., Fu, Z., **Stainton, S.A.**, Barney, S., Hogg, J., Innes, W., and Dlay, S. "Automated Identification of Spontaneous Venous Pulsation from Infrared Video of the Optic Nerve Head" *17th International SPECTRALIS® Symposium (ISS)*, 2019 [poster].

11. Sun, Y., Fu, Z., **Stainton, S.A.**, Barney, S., Hogg, J., Innes, W., and Dlay, S. "Automated Retinal Layer Segmentation of OCT Images Using Two-Stage FCN and Decision Mask," *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2019, pp. 1-6.

12. Sun, Y., Fu, Z., Xiangyu Z., **Stainton, S.A.**, Barney, S., Hogg, J., Innes, W., and Dlay, S. "MPG-Net: Multi-Prediction Guided Network For Segmentation Of Retinal Layers In OCT Images" *28th European Signal Processing Conference (EUSIPCO)*, 2020, pp. 1299-1303.

1.4 Thesis Outline

This thesis is organised into six chapters.

Chapter 1 has provided a brief introduction to this research, its aims, objectives and motivation, and the resultant output of publications.

Chapter 2 covers the technical background of wireless communication, visible light communication, and artificial neural networks with relevant reviews of the state-of-the-art in the respective areas.

Chapter 3 covers the design of the novel neural network and its practical implementation in a harsh V2V environment. It then explores the potential of further improving the efficiency of communication by introducing spectrally efficient frequency division multiplexing. An evaluation is then undertook to compare the receiver performance of the proposed implementations against an 802.11p standard conformant baseline implementation.

Chapter 4 covers the exploration of utilising the error vector magnitude as a cost function for training neural networks in a communication scenario. A performance comparison is made between three implementations of neural networks, starting from common neural network architectures found in the literature.

Chapter 5 explores the wider application and advantages stemming from the novel neural network with the proposed error vector magnitude objective loss. Its performance boundaries are explored in a visible light communication setting, demonstrating the widespread applicability of the solution.

Chapter 6 begins the conclusion, where key findings are summarised and recommendations are made on potential future research opportunities.

Chapter 2

Background Theory

2.1 Introduction

This chapter looks at the challenges faced by wireless communication in a vehicular network and explains the background physics of the channel itself. The wireless standards that were put in place for the environment are examined which in turn sets the theme for areas of suggested possible improvement throughout this research. It then delves into some required background theory for later discussions on machine learning, and specifically challenges around artificial neural networks and their training.

The desire for highly robust and spectrally efficient signal processing methods are stated as a driving factor in this research. These two aims work harmoneously with one another, more robust signal processing leads to lower numbers of signal re-transmissions, ergo more efficient use of the available spectrum. More spectrally efficient waveforms mean higher numbers of users can communicate simultaneously over the same available bandwidth. Together, these advantages allow for widescale adoption of vehicular communication in a scalable manner.

Another factor that drives this research relates to the safety critical aspects of vehicular communication. Vehicles today are already able to perform many tasks autonomously, and as vehicles become more and more autonomous, some of this autonomy will rely on information that is shared to and from other vehicles and infrastructure. The potentially safety critical nature of these systems demand that messages are able to be transmitted in even the worst conditions. This will be a critical turning point in our transportation systems, allowing us to take a proactive approach to safety rather than reactive.

2.2 Vehicular Communication Standards

There exists a standard produced by the IEEE working group 1609 that is based on the 802.11p standard named WAVE that sets out the main protocols for vehicle-to-vehicle and vehicle-to-infrastructure communication. Being a part of the 802.11 framework, it shares many similarities with other standards in the same family - the most recognisable of which being Wi-Fi. Table 2.1 below outlines key similarities and differences between them while Table 2.2 gives the channel coding parameters.

Table 2.1 Wi-Fi and WAVE Comparison

Parameters	Wi-Fi	WAVE
Operating Frequency	2.4/5GHz	5.9GHz
Bandwidth	20MHz	10MHz
Supported Data Rates (Mbps)	6,9,12,18,24,36,48,54	3,4.5,6,9,12,18,24,27
Supported Modulation Schemes	BPSK, QPSK, 16QAM, 64QAM	Same as Wi-Fi
Channel coding	Convolutional Code	Same as Wi-Fi
No. Of Data Subcarriers	48	Same as Wi-Fi
No. Of Pilot Subcarriers	4	Same as Wi-Fi
No. Of Virtual Subcarriers	12	Same as Wi-Fi
FFT/IFFT Size	64	Same as Wi-Fi
FFT/IFFT Interval	6.4μs	3.2μs
Subcarrier Spacing	0.3125MHz	0.15625MHz
Cyclic Prefix Interval	0.8μs	1.6μs
OFDM Symbol Interval	4μs	8μs

Table 2.2 WAVE Channel Coding Parameters

Parameter	Value
Generator Polynomial	$(133,171)_8$
Code Rates	$\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$
Puncturing Vectors	$R = \frac{2}{3} \rightarrow 1110$, $R = \frac{3}{4} \rightarrow 110101$

As one can see, through the expectation of a harsh channel, the data rates supported have been halved. Interestingly however, the subcarrier spacing has also been halved

meaning this scheme is more sensitive to Doppler spread and inter-carrier interference (ICI). This will have the effect of decreasing the coherence bandwidth making the entire system more frequency selective, although it has been partially addressed by doubling the symbol and cyclic prefix intervals.

The range of messages that are communicated via this standard can vary, to learn more one can investigate the IEEE 1609.3 WAVE Networking standard [20]:

> "At minimum, a WAVE device implements at least the following features specied in this standard (see Annex D).
>
> - LLC sublayer
> - IPv6 or WSMP, or both
> - Transmit or receive, or both"

The second point from the excerpt above shows the types of packets that the communicating entities have to be able to process. IPv6 is the newest iteration of the Internet Protocol, with an address space of 2^{128} (up from the offering of 2^{32} from IPv4). WSMP stands for WAVE short message protocol, and was designed for the low latency sending of safety messages within a vehicular ad-hoc network (VANET) and for WAVE service advertisement messages. The need for low latency comes from the SAE J2735 protocol [21] which dictates that safety messages must be broadcast every 100-300ms. It does this by reducing the overhead in both the packet design and the network structure. From the network point of view, by being primarily built on the 802.11 framework, the notion of service sets (SS) has been brought forward. An independent basic service set (IBSS) is a set of stations without infrastructure, generally referred to as an ad-hoc network, however in the standard every basic service set (BSS) requires an access point that acts as the controller/master station, which will periodically broadcast a beacon message containing the service set ID (SSID). It is this beacon that allows all stations within the network to synchronise their time and frequency, hence only members of the same service set can communicate with each other. This involves a lot of overhead, and is one of the motivations behind the WSMP as it allows vehicles to skip the steps of forming a BSS. As for reducing the overhead in the packet design, an IPv6 packet has a header size of 40 bytes but must also carry either a UDP or TCP header on top at a minimum, having sizes 8 bytes and 20 bytes respectively, making the entire packet a minimum of 48 to 60 bytes without a payload. Compare this to WSMP which has a header size of 11 bytes. One of the fundamental principles in WAVE is the broadcasting of basic safety messages and road side alerts. As mentioned previously,

basic safety messages must be sent periodically every 100–300ms and contain safety critical information such as other vehicles location, speed, and heading whereas road side alert messages are transmitted when particular events occur, such as poor road conditions or an accident alert. It is therefore imperative that these messages reach surrounding vehicles, however, the harsh channels found in the vehicular environment [22] caused by factors such as severe Doppler effects, multipath fading, and shadowing, can reduce the error free transmission probability significantly.

2.3 Physical Properties of Vehicular Radio Channels

Communication in a VANET poses many unique challenges due to its mobile nature. At typical vehicle speeds of 30–70mph (13.4–31.3 m/s), Doppler effects are large (263.7–616 Hz, assuming stationary transmitter or receiver) and at 5.9 GHz the wavelength of the carrier signal is small. This problem is exacerbated by the environment, where various fading effects such as multipath attenuate this already degraded signal. There has been much research attempting to quantify the effects of vehicular channels [23–25], however the results vary depending on the particular scenario i.e. highways and rural areas have a much stronger ratio of line-of-sight to multipath components than do inner city environments where buildings and other scatterer objects can contribute to stronger non-line-of-sight components. Due to the high variation in the channel depending on the particular context, it is difficult to produce a general model that can be applied, therefore it is not uncommon to see research use varying mixtures of more classic channel fading such as Rayleigh, Ricean and Nakagami distributed models. These distributions can be useful for modelling multipath fading which will be investigated in a later section. Another distinguishing element of vehicular commuication is the fact that both the transmitter and the receiver are at similar, low heights (1m - 2.5m) [26] which results in lower coverage than traditional mobile systems.

One of the key differences between modelling a mobile communication system and modelling vehicle-to-vehicle communication is that while mobile channels are often modelled with stationary channel statistics, vehicle-to-vehicle channels exhibit non-stationary statistics. This means that not only are the impulse responses time varying, but other properties such as the power delay profile (PDP), Doppler spread and small scale fading effects are also time varying. It is because of this that the wide sense stationary with uncorrelated scattering assumption used by many mobile channel descriptors cannot be used, as the region of stationarity for the vehicular channel is small [27]. These characteristics depend upon the distance between the two antennas,

the path(s) taken by the signal, and the environment (buildings and other objects) around the path. To this end, the sections that follow will give some background to the physical phenomena affecting the vehicular wireless channel.

2.3.1 Multipath Propagation And Inter Symbol Interference

The typical antenna radiates energy in an isotropic manner, meaning electromagnetic energy is distributed evenly in all directions, this is demonstrated in Fig. 2.1. Because of this, objects in the environment between the transmitter and the receiver have the opportunity to reflect and scatter the wireless signal, leading to multiple attenuated versions of the signal arriving at the receive antenna. As can be intuitively expected, from the equation $v = {}^{s}/_{t}$ (speed equals distance over time, where the speed $v = c \approx 299,792,458\text{m/s}$ is fixed) one can see that due to the extra distance the signal has travelled at a fixed speed, the time at which the receiver receives the N reflected signals is some $\tau_N > \tau_{\text{LOS}}$.

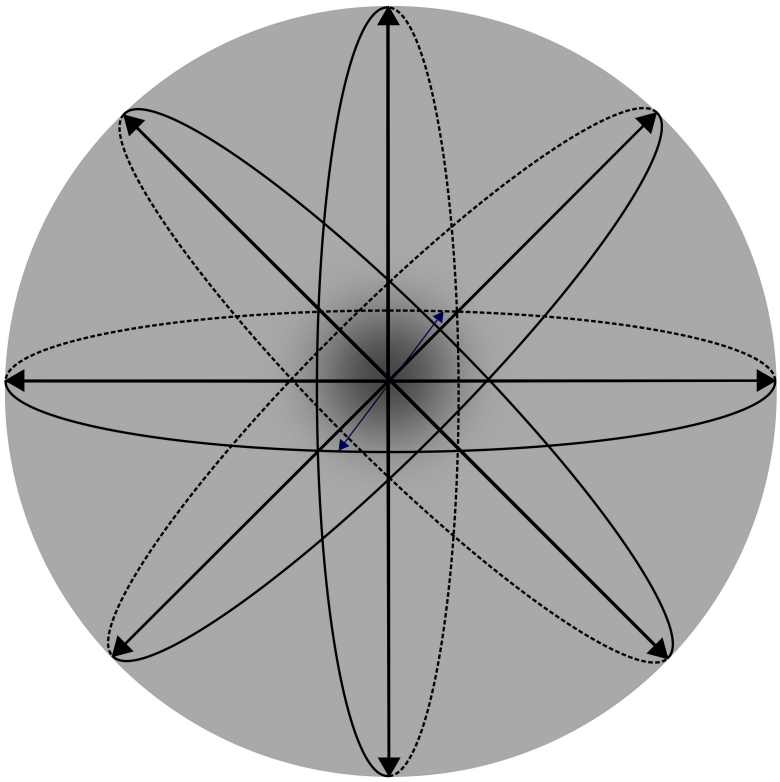

Fig. 2.1 Isotropic radiation pattern example

As mentioned previously, generalised statements about the type of wireless environment can be useful in channel modelling. Knowing whether or not the the wireless environnment is likely to contain a dominant line of sight (LOS) component allows baseline channel models to be created from theoretical parameters. If one observes the reflected signals and their time delays and assumes that they are normally distributed i.e. $\tau \sim \mathrm{N}\left(\mu, \sigma^2\right)$ then the absolute value of the impulse response is given by a Ricean distribution if $\mu \neq 0$ indicating there is a LOS component or a Rayleigh distribution if $\mu = 0$ indicating no dominant LOS component. With a LOS component, one can begin to model the channel using a Ricean distribution, as given by [28]:

$$p\left(r\right) = \frac{r}{\beta^2} e^{-\frac{r^2+A^2}{2\beta^2}} I_0\left(\frac{rA}{\beta^2}\right) \tag{2.1}$$

where $I_0(x)$ is the modified Bessel function of the first kind with zero order [28]:

$$I_0\left(x\right) = \frac{1}{2\pi} \int_0^{2\pi} e^{-x\cos\theta} d\theta \tag{2.2}$$

When there is no dominant line of sight component, also known as a non line of sight (NLOS) channel, the signal degradation can be very severe as reflected signals arrive with phase differences, leading to the phenomena of fast fading. In this instance, the Rayleigh model may be given by [29]:

$$p\left(x|\beta\right) = \frac{x}{\beta^2} e^{\left(\frac{-x^2}{2\beta^2}\right)} \tag{2.3}$$

One such simulation that shows the potential effects of Rayleigh fading on the channel is shown in Fig. 2.2, note the characteristic deep fades.
There exists further generalisations of channels, for example in the case of Weibull. The Weibull probability density function is a general fading model that allows for more control over the extent of the fading (and can reduce to both Ricean and Rayleigh) however at the cost of computational complexity. For completeness, this is given in (2.4) [29].

$$p\left(x|k,l\right) = \frac{l}{k}\left(\frac{x}{k}\right)^{l-1} e^{-\left(\frac{x}{k}\right)^l} \tag{2.4}$$

Due to the scattering and reflection, when these signals arrive at the receiver they are superimposed which results in a combined signal that may have increased or descreased in power depending on whether the superposition of the waves added constructively or destructively. As the wavelength of radio signals in the GHz region are very small, even small path differences can have a dramatic effect.

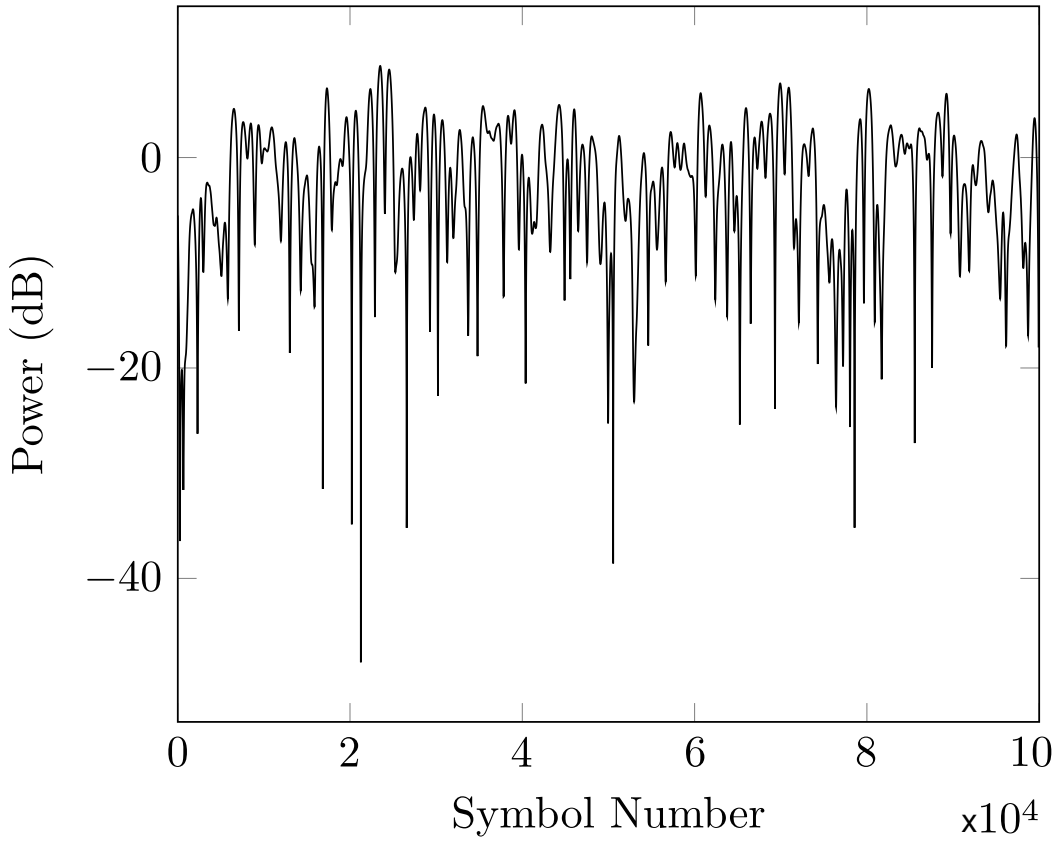

Fig. 2.2 Power spectral density of an example Rayleigh channel

A useful metric for a given channel is the delay spread, typically either the maximum $\tau_{\max}$ or the root mean squared (rms) τ_{rms}. The delay spread gives an indication of how long after a signal is transmitted is the receiver likely to be receiving reflected copies. A large value for τ indicates a highly dispersive channel, and presents a challenge for high throughput, accurate communications. This is demonstrated graphically in Fig 2.3.

To investigate the channel impulse response, without loss of generality let the arbitrary transmitted signal be [30]:

$$\begin{aligned} s(t) &= \text{Re}\left\{u(t)\exp^{j2\pi f_c t}\right\} \\ &= \text{Re}\{u(t)\}\cos(j2\pi f_c t) - \text{Im}\{u(t)\}\sin(j2\pi f_c t) \end{aligned} \tag{2.5}$$

where $u(t)$ is the equivalent lowpass signal for $s(t)$ at a carrier frequency of f_c. The corresponding received signal after transmission over a wireless channel is the sum of the line of sight component and all resolvable multipath components (without a noise term for clarity) is given by [30]:

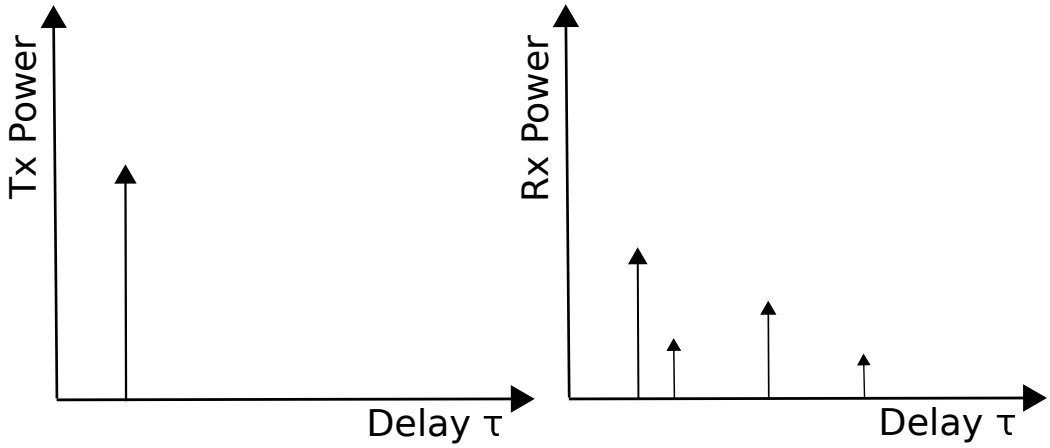

Fig. 2.3 Example power delay profile for a multipath channel

$$r(t) = \mathrm{Re}\left\{ \sum_{n=0}^{N(t)} \alpha_n(t)\, u(t - \tau_n(t)) \exp^{j(2\pi f_c(t-\tau_n(t))+\phi_{D_n})} \right\} \tag{2.6}$$

where $n = 0$ is the line of sight path, $N(t)$ is the number of resolvable multipath components, $\tau_n(t)$ is the delay of n^{th} multipath component, ϕ_{D_n} is the doppler shift of the n^{th} multipath component, and $\alpha_n(t)$ is the amplitude of the n^{th} multipath component. Two multipath components are resolvable if their delay difference significantly exceeds the inverse of the signal bandwidth i.e. $|\tau_1 - \tau_2| >> 1/B$. If this relationship is not satisfied, the two components are binned together at the receiver and appear as a superposition of the pair, and are hence deemed non-resolvable.

This effect can be modelled by a tapped delay line, where each impulse of the channel is characterised by a delay and an impulse coefficient. The channel is then modelled by [30]:

$$h(t,\tau) = \sum_{n=0}^{N(t)} \alpha_n(t) \exp^{j(2\pi f_c(t-\tau_n(t))+\phi_{D_n})} \delta(\tau - \tau_n(t)) \tag{2.7}$$

where $h(t,\tau)$ is the time varying channel impulse response at time t, $\delta[\cdot]$ is the Kronecker delta function.

In reality, when transmitting wirelessly from an isotropic antenna as described, there is no escaping the effects of inter-symbol interference (ISI). To demonstrate this effect, one can model a wireless communication channel as a linear time invariant (LTI) filter [31] with additive noise, typically additive white Gaussian noise (AWGN). This allows for the basic expression of the effect of a wireless communication channel as [31]:

$$x[n] = \sum_{k=-\infty}^{\infty} h[k]s[n-k] + w[n] \tag{2.8}$$

where $h[\cdot]$ is the channel impulse response, $s[\cdot]$ is the transmitted symbol, and $w[\cdot]$ is the additive noise. This equation can be rearranged around $k = 0$ to emphasise the effect for any given symbol $s[n]$ [31]:

$$x[n] = h[0]s[n] + \sum_{k=-\infty, k\neq 0}^{\infty} h[k]s[n-k] + w[n] \tag{2.9}$$

It is clearly outlined by (2.9) that the received symbol that is observed is now a composition of the original transmitted symbol with additive noise, plus an extra term that is a function of the delayed versions of the originally transmitted symbol. This term is the ISI, and it is the job of downstream equalisation to attempt to eliminate or reduce its effect so that the original symbol can be recovered correctly. The design of these depend on the communication methods implemented, therefore, the specific implementations of the receivers used in this work will be discussed in later sections as the modulation schemes themselves are introduced.

2.3.2 Path Loss and Shadowing

Path loss and shadowing are the two main proponents of the wider group of channel degradations classed as 'large scale fading'. The most fundamental of losses that affects a wireless signal is known as free space propagation path loss. This loss is caused by the physical spreading of the RF energy as it moves through space-time. The theoretical received power can be estimated as a function of the transmitted power, the net gain of the antennas, the wavelength of the transmitted signal, and the distance between the transmitter and the receiever. This is shown in Eq 2.10 [30].

$$P_r = P_t \left(\frac{\sqrt{G_l}\lambda}{4\pi d} \right)^2 \tag{2.10}$$

Where P_r is the received power
P_t is the transmitted power
G_l is the net antenna gain
λ is the wavelength of the signal
d is the distance between transmitter and receiever

This is given for an unobstructed LOS signal path, thus in practice observed results are often lower, and a general empirical formula for path loss is often used such as [30]:

$$P_r = P_t K \left(\frac{d_0}{d} \right)^{\gamma} \tag{2.11}$$

Where P_r is the received power

P_t is the transmitted power

$K = \dfrac{P_r(d_0)}{P_t}$ is a constant factor at a reference distance d_0

γ is the path loss exponent

This equation is known as the simplified path loss model, and whilst it theoretically estimates received power it requires empirical validation for the specific channel environment one intends to use it in. Typically, one would use $\gamma = 2$ for estimating a free space model or $\gamma = 4$ for estimating a two-ray model.

In reference to the second proponent of large scale fading, namely shadowing, this is a phenomena that arises due to the presence of obstacles in the wireless path between the transmitter and the receiver causing the attenuation of signal power though reflection, scattering, absorption and diffraction. As one probably can intuit from personal experience, when signal power is completely attenuated in areas it can create what is commonly referred to as signal 'dead spots'. In a mobile environment like the vehicle-to-vehicle one, these obstructions may be present in the environment anywhere on the order of seconds to minutes, which in comparison to the wavelength of the signals in use means these effects occur at a much slower timescale compared to fast fading, which will be discussed in the proceeding section.

It is often not possible to completely characterise a wireless environment and all its obstacles and blockages ahead of time, especially in mobile environments. The location, size, and dielectric properties of the blocking objects, as well as the changes in the reflecting surfaces and scattering objects that cause the random attenuation, are generally unknown. In these instances it is useful to model the obstructions in a more statistically random manner. One common model for this is log-normal shadowing. In log-normal shadowing, the ratio of transmit to receive power $\psi = {}^{P_t}/_{P_r}$ is assumed to be a random variable with a log-normal distribution that is given by [30]:

$$p(\psi) = \frac{\xi}{\sqrt{2\pi}\sigma_{\psi_{\mathrm{dB}}}\psi} \exp\left[-\frac{\left(10\log_{10}\psi - \mu_{\psi_{\mathrm{dB}}}\right)^2}{2\sigma^2_{\psi_{\mathrm{dB}}}}\right], \psi \geq 0 \tag{2.12}$$

where $\xi = {}^{10}/_{\ln 10}$, $\mu_{\psi_{\mathrm{dB}}}$ is the mean of $\psi_{\mathrm{dB}} = 10\log_{10}\psi$ in decibels, and $\sigma_{\psi_{\mathrm{dB}}}$ is the standard deviation of ψ_{dB} in decibels. In this instance, since the mean μ of the log-normal distribution is determined by $\psi = {}^{P_t}/_{P_r}$, it follows that the value for $\mu_{\psi_{\mathrm{dB}}}$ is always greater than or equal to zero. Although the log-normal distribution accepts values $0 \leq \psi \leq \infty$, this implies a region where $\psi < 1$ meaning $P_r > P_t$ which is not possible.

One can independently justify that the normal distribution for the model of the mean received signal power in dB is correct by considering the case where the shadowing is dominated by blocking objects (as opposed to reflective and refractive etc, for simplicity). The attenuation of a signal as it travels through an object of depth d is approximately equal to [30]:

$$s(d) = \exp^{-\eta d} \tag{2.13}$$

where η is an attenuation constant that is dependent on the physical properties of the attenuating object. If one assumes that η is approximately equal for all blocking objects, using a random depth d for object k, the signal attenuation is then given as [30]:

$$\begin{aligned} s(d_t) &= \exp^{-\eta\sum_{k=0}^{K} d_k} \\ &= \exp^{-\eta d_t} \end{aligned} \tag{2.14}$$

where $d_t = \sum_{k=0}^{K} d_k$ is the sum of the depths of all the random blocking objects through which the signal has travelled. If there are many objects between the transmitter and the receiver, say for instance lots of vehicles and infrastructure, then by the central limit theorem d_t can be approximated by a normally distributed random variable. Thus from (2.13) one can easily see $\log s(d_t) = \eta d_t$ will be normally distributed with a mean μ and a standard deviation σ. The value σ will depend on the specific environment.

Models for both the path loss and the shadowing can be superimposed to give a more rounded picture of the communication environment, in this view, one can take a simpler view of the ratio of received to transmitted power which is given by [30]:

$$\frac{P_r}{P_t}\text{dB} = 10\log_{10} K - 1 - \gamma \log_{10}\frac{d}{d_0} - \psi_{\text{dB}} \tag{2.15}$$

where ψ_{dB} is a normally distributed random variable with zero mean and a variance $\sigma^2_{\psi_{\text{dB}}}$. It can be seen from this that the path loss decreases linearly in relation to $\log_{10} d$ with a gradient of 10γdB/decade, where γ is the path loss exponent. The overall combination of path loss and shadowing, or more generally the large scale fading, has large implications on the overall system design. There comes a point where if the received power level drops any lower, the system performance is unacceptable. This leads to the quantification of some power level $P_{\min}$ which sets a target minimum received power level to maintain adequate system performance. Since one has defined the large scale fading effects statistically with a log-normal distribution, there exists a statistical representation for the probability that the received power level will drop below $P_{\min}$, thus causing an outage event. One can define this outage probability $P_{\text{out}}(P_{\min}, d)$ under large scale fading effects to be the probability that at a given distance d, the received power $P_r(d)$ falls below $P_{\min}$, i.e.

$$P_{\text{out}}(P_{\min}, d) = p(P_r(d) < P_{\min}) \tag{2.16}$$

For the combined path loss and shadowing model from (2.15) this is given as [32]:

$$p(P_r(d) \leq P_{\min}) = 1 - Q\left(\frac{P_{\min} - \left(P_t + 10\log_{10} K - 10\gamma \log_{10}\left(\frac{d}{d_0}\right)\right)}{\sigma_{\psi_{\text{dB}}}}\right) \tag{2.17}$$

where the Q function is defined as the probability that some Gaussian random variable X with mean 0 and variance 1 is greater than z [29]:

$$Q(z) \triangleq p(X > z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} \exp^{-\frac{y^2}{2}} dy \tag{2.18}$$

which can be defined in terms of the complimentary error function as [29]:

$$Q(z) = \frac{1}{2}\text{erfc}\left(\frac{z}{\sqrt{2}}\right) \tag{2.19}$$

2.3.3 Doppler Effect

When there is relative motion between a transmitter and a receiver (or any scattering or reflecting surface between the two in the case of multipath), a Doppler shift can

be induced in the transmitted signal. The degree of this shift is directly proportional to the rate of motion between the transmitter and the receiver. Take for instance a signal with a wavelength given by $\lambda = c/f_c$ where c is the speed of light and f_c is the carrier frequency of the signal. With a relative motion given by v, at an angle of ϕ_i to the propoagation direction, one can calculate the phase change experienced by the transmitted signal after some time Δt. In the time Δt, the receiver has travelled a distance of $v\Delta t$, thus meaning it has experienced a change in propagation path length of approximately $\Delta l = v\Delta t \cos(\phi_i)$. The phase change seen caused by this can then be given as [33]:

$$\Delta\Phi = 2\pi \frac{v\cos(\phi_i)}{\lambda}\Delta t \tag{2.20}$$

The rate of this phase variation is characterised by the Doppler frequency, often denoted f_D, and can be given by [33]:

$$f_D = \frac{1}{2\pi}\lim_{\Delta t\to 0}\frac{\Delta\Phi}{\Delta t} = \frac{v\cos(\phi_i)}{\lambda} \tag{2.21}$$

The ultimate effect being that the instantaneous frequency of the receieved signal over the propagation path becomes $f_c \pm f_D$. Due to this effect happening simultaneously on all reflected and scattered propagation paths, the resultant signal will contain mutiple signals of varying energies with varying Doppler shifts. The maximum of these shifts however will occur when $\phi_i = 0$ (when the incident wave and the relative direction of motion are opposite to each other), simplifying (2.21) and giving a single value to characterise the worst case scenario [33]:

$$f_D = \frac{v}{\lambda} \tag{2.22}$$

Related to the Doppler shift is a metric known as the coherence time, T_c. While the Doppler shift can be thought of as characterising the highest speed of phase variation in the propagation channel, the coherence time is a measure of the time domain variation. More specifically, it is a measure of the duration of time in which the impulse response of the channel remains highly correlated. To derive this measure, if one was to begin by investigating the frequency domain correlation of the channels impulse response [33]:

$$\rho(f,\Delta t) = \mathbf{E}[H^*(f,t)H(f,t+\Delta t)] \tag{2.23}$$

where $\mathbf{E}[\cdot]$ denotes the statistical expectation, and $H(f,t)$ is the time varying frequency response of the wireless channel, calculated from Fourier transform of the time domain representation $h(\tau,t)$ with respect to τ. The values of T_c is derived such that it satisfies $\rho(f,\Delta t) \approx 1$ for all f and all $\Delta t < T_c$. The value T_c is inversely proportional to the maximum Doppler shift f_D, and has been shown to hold the approximate relationship of $T_c \approx {}^{0.4}/_{f_D}$.

2.4 Application of VLC in Vehicular Networks

Whilst RF is the most prevalent method of communication in both everyday life and vehicular communication, there is growing attention being paid to the use of visible light communication (VLC). Whilst RF in vehicular communications is often heavily degraded by multipath effects as we have covered, in VLC the lightwaves are highly directional and do not pass through objects thus avoiding much of this multipath problem. Whilst that is an advantage, being highly directional we move away from the isotropic antenna radiation pattern discussed for RF and have to design systems with asymetric radiation patterns in mind. One such example of this exploration can be found in [34] where the authors investigate Lambertian, Gaussian and asymmetrical radiation patterns to model headlights from different vehicle manufacturers. It was found that path loss depends on the radiation patterns as well as the traffic conditions including traffic flow and inter-vehicular spacing, and that the statistical distribution of the path loss is a convolution of the radiation intensity distribution and the inter-vehicle spacing distribution.

In keeping with RF vehicular communication however, it is also well understood that vehicular VLC communication suffers from the same high level problem, namely, the physical environment. There have been many studies undertaken to understand the vehicular VLC channel, one of which is [35] which examines the temporal properties of the channel taking into account dynamic traffic conditions, and the authors state that received multipath components have higher amplitude and lower propagation delay at high density traffic in comparison to the multipath components received at low density traffic. They also state that in both high and low density traffic conditions the vehicular VLC channel is flat and slow fading, making VLC an ideal candidate for vehicular communication systems.

To this end, the following section will explain some brief background theory for implementing VLC communication in practice, which provides the prerequisute knowledge for practical experiments that will be undertaken in later chapters.

2.4.1 Intensity Modulation and Direct Detection

The optical source for the VLC systems used in later experiments is a light emitting diode (LED), with a luminous intensity R_0 that follows Lambert's cosine law [36]:

$$R_0(\theta) = \frac{m+1}{2\pi}\cos(\theta)^m \tag{2.24}$$

Where θ is the angle of emission and m is the Lambertian order, given by [37]:

$$m = -\frac{\ln(2)}{\ln\left[\cos\left(\theta_{1/2}\right)\right]} \tag{2.25}$$

Where $\theta_{1/2}$ is the transmitter semi-angle. The source is known as a Lambertian transmitter when $m = 1$ ($\theta_{1/2} = 60°$). Beyond an optical source, one needs to be able to modulate it in order to transmit data. Whilst it is possible to use modulation methods such as a Mach-Zehnder modulator as seen in [38], it is more common to see VLC communication implemented via intensity modulation and direct detection. As the name suggests, in intensity modulation the signals electrical amplitude is represented via the intensity of the optical power and direct detection performs the opposite, the intensity of the optical power is represented by the output signals electrical amplitude. An example of intensity modulation can be seen in Fig 2.4.

Firstly, a linear region on the LI curve of the optical source is identified. The linearity region is important because any non-linear alterations to the signal need to be accounted for at the receiver otherwise they will act as a source of noise for the system. Having a large linear region for an optical device is desirable as it increases the number of possible applications, making this an active area of research in the literature such as [39]. Secondly, depending on the maximum voltage swing of the intensity modulation signal source, a dc bias current is added such that the peak-to-peak swing can sit optimally inside the linear region.

This optical power in the form of light then hits a photoreceiver after travelling over a channel, with a frequency response $H(f)$ given by [40]:

$$H(0) = \frac{A_r}{d^2} R_0(\theta)\cos(\varphi) \tag{2.26}$$

Where A_r is the photoactive area of the photodetector, d is the distance to the receiver, and φ is the angle of incidence to the photodetector. In an ideal setting, the total received optical power P_r impinging on the photodetector is given by [37]:

$$P_r = H(0)\, P_t \tag{2.27}$$

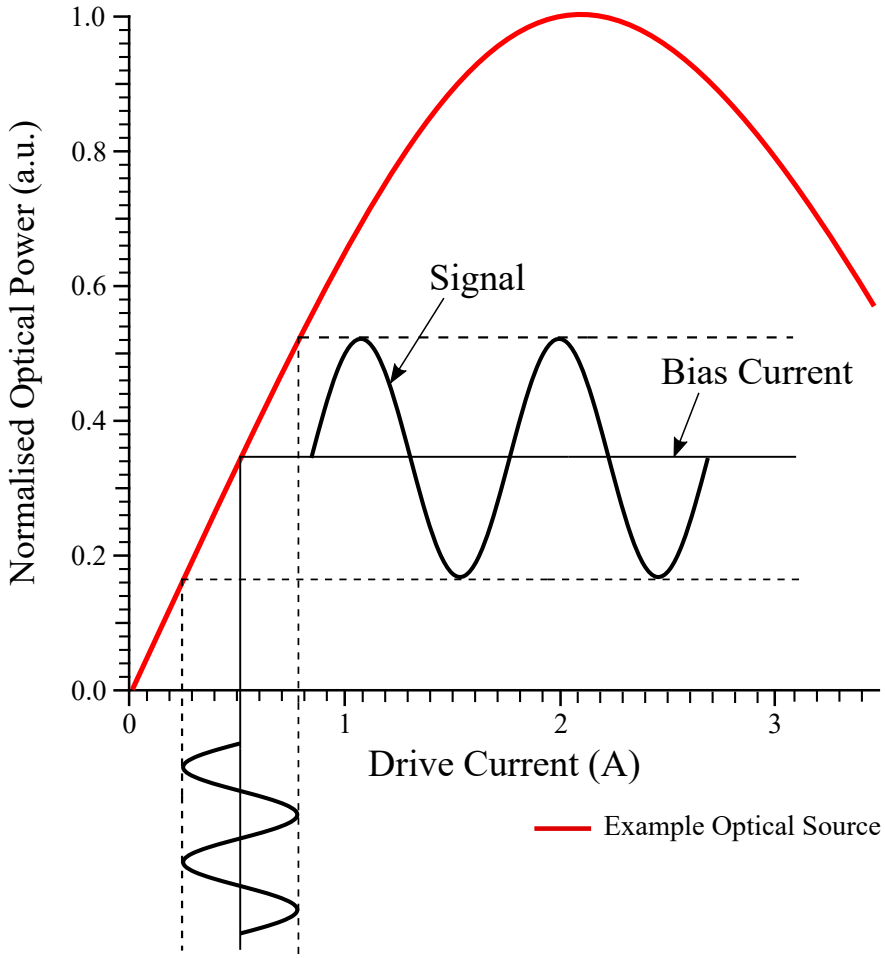

Fig. 2.4 Example of an LI curve demonstrating Intensity Modulation

Where P_r is the received optical power and P_t is the transmitted optical power. Using (2.26) to expand this one gets:

$$P_r = \frac{A_r}{d^2} R_0(\theta) \cos(\varphi) P_t \tag{2.28}$$

This clearly shows that neither the wavelength nor the signal frequency has an impact on the received optical power.

2.5 Physical Layer Waveforms

2.5.1 Orthogonal Frequency Division Multiplexing

Orthogonal frequency division multiplexing (OFDM) is a well known and researched topic, and forms the basis for many modern communications standards. The basic

premise of OFDM is to take a given bandwith of spectrum W and use it to transmit N independent and orthogonal streams, known as subcarriers. This can be seen graphically in Fig 2.5. There are several benefits that arise from splitting the single data stream up into these independent streams. Firstly, the data rate required on each stream is lowered by a factor of N to achieve the same data rate of the original stream, meaning that symbols on each subcarrier are further spaced in time. Secondly, as was seen previously, many real world wideband wireless channels are not uniform over all frequencies, i.e. they display frequency selectivity. Due to this, receiver performance can be poorly detrimented. In the OFDM case, this frequency selective channel is split up into N indepenent narrow band channels, which can be seen as uniform for an appropriately selected value of N.

OFDM transmitter

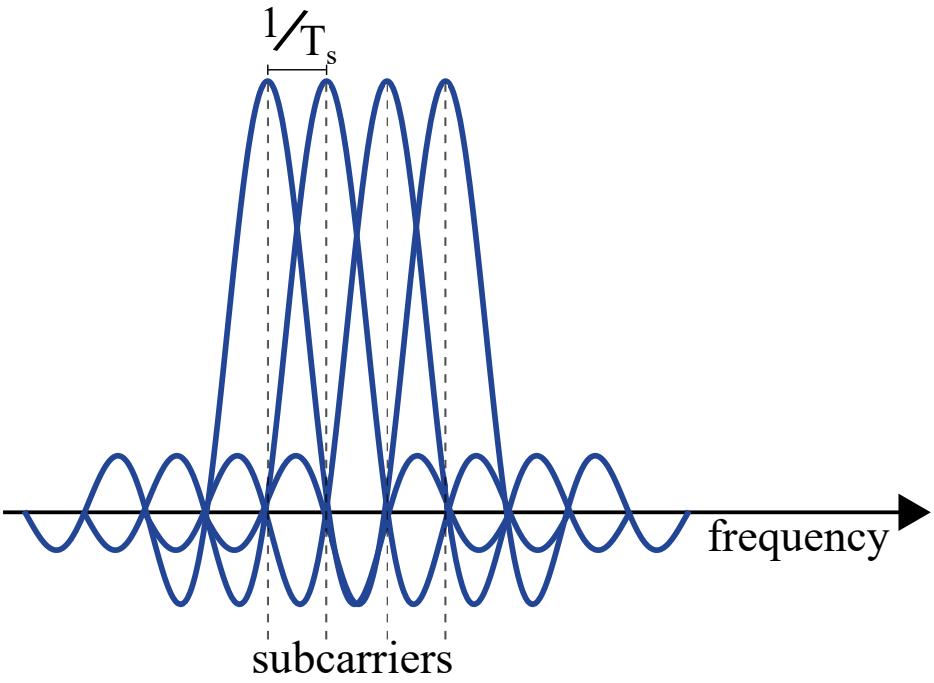

Fig. 2.5 OFDM individual subarriers

For a symbol length of T seconds, which encompasses a time T_{cp} seconds for the length of the cyclic prefix, the generation of an OFDM waveform can be viewed as [41]:

$$\Psi_k(t) = \begin{cases} \frac{1}{\sqrt{T-T_{cp}}} \exp^{j2\pi \frac{W}{N} k(t-T_{cp})}, & \text{if } t \in [0,T] \\ 0, & \text{otherwise} \end{cases} \tag{2.29}$$

where $T = {}^{N}/_{W} + T_{cp}$. These waveforms $\Psi_k(t)$ can then be used to modulate a data stream to create a baseband signal s [41]:

$$s_l(t) = \sum_{k=0}^{N-1} x_{k,l} \Psi_l (t - lT) \tag{2.30}$$

where $x_{k,l}$ complex numbers representing points from a reference constellation.

OFDM receivers

An OFDM receiver is typically made up of a bank of matched filters, matched to the transmitted waveform Ψ from (2.29). These take the form [41]:

$$\Phi_k (t) = \begin{cases} \Psi_k^* (T - t), & \text{if } t \in [0, T - T_{cp}] \\ 0, & \text{otherwise} \end{cases} \tag{2.31}$$

From this, one can see that the cyclic prefix is discarded and not needed for signal recovery. Assuming an adequately long cyclic prefix, all the ISI from the previous signal should be isolated within it, hence the sampled output from the receiver filter bank should contain no ISI, thus increasing performance.

In a modern digital system however, the above is often replaced by its discrete time equivalent, which replaces the continuous time modulation and demodulation with the inverse discrete Fourier transform (IDFT) and the discrete Fourier transform, respectively. Assuming that the cyclic prefix is longer than the channel impulse response, the standard linear convolution operation with the transmission channel is replaced by a circular convolution. This means that the end to end OFDM system can be written as [42]:

$$\begin{aligned} y_l &= \mathrm{DFT}\left(\mathrm{IDFT}\,(x_l) \bigotimes h_l + \hat{n}_l\right) \\ &= \mathrm{DFT}\left(\mathrm{IDFT}\,(x_l) \bigotimes h_l\right) + n_l \end{aligned} \tag{2.32}$$

where

$$\mathrm{DFT}\,(s) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_n \exp^{-\left(\frac{j2\pi nk}{N}\right)} \tag{2.33}$$

and

$$\mathrm{IDFT}\,(s) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_n \exp^{\left(\frac{j2\pi nk}{N}\right)} \tag{2.34}$$

where y_l contains the N received data points, x_l are the N originally transmitted data points from the reference constellation $\mathbb{C}$, h_l is the zero padded impulse response of the channel, and $\hat{n}_l$ is the additive white Gaussian noise of the channel. In (2.32) above, it can be written that $n_l = \mathrm{DFT}\,(\hat{n}_l)$ as the Gaussian noise is assumed to be uncorrelated. The assumption that the cyclic prefix is longer than the channel impulse response needs to be taken into consideration by system designers as the typical channel responses differ between environments. Having a cyclic prefix that is too long results in wasted energy transmitting symbols that will be discarded at the receiver, however, if it is too short then the circular convolution property that the receiver relies on for demodulation is no longer valid, and performance will break down. It can be further noted that if the circular convolution property does hold, one can take advantage of the fact that the DFT of two circularly convolved signals is directly equivalent to the product of their respective individual DFTs, i.e. [42]:

$$\begin{aligned} y_l &= x_l \cdot \mathrm{DFT}\,(h_l) + n_l \\ &= x_l \cdot H_l + n_l \end{aligned} \tag{2.35}$$

where $\cdot$ represents element-wise multiplication, and $H_l = \mathrm{DFT}\,(h_l)$ represents the frequency response of the channel.

OFDM zero forcing receiver

A zero forcing equaliser is a simple form of linear equalisation that attempts to 'undo' the effects of the channel by applying its direct inverse to the received signal. The name comes from the fact that, ignoring other sources of noise besides that of the channel, it reduces the effect of the channel to net-zero.

As OFDM converts a potentially frequency selective, wideband channel into uniform, narrowband subchannels, zero forcing for OFDM takes the same form as that of the AWGN channel i.e. for a channel with frequency response $H(f)$, the zero forcing equaliser $Z(f)$ is given by [42]:

$$Z\,(f) = \frac{1}{H\,(f)} \tag{2.36}$$

Thus producing the relationship $Z(f)H(f) = 1$, which gives a flat frequency response. As we have seen previously however, this has the negative side effect of boosting the noise power within the signal when the frequency response of the channel contains deep nulls that tend toward zero.

OFDM minimum mean square error receiver

A minimum mean square error (MMSE) equaliser exploits a channels correlation in both time and frequency. It typically involves inserting a set of pilot symbols $\mathbb{P}$ into the data stream, these are symbols at known locations in a stream known to both the transmitter and receiver. Each pilot symbol can be uniquely identified by its location in time and frequency. These pilot symbols are used in MMSE to generate an estimated channel transfer function, $\hat{H}_{k,n}$ of the k^{th} subcarrier in the n^{th} OFDM symbol.

One can denote the set of pilots $\mathbb{P}$ as [43]:

$$\mathbb{P} = \{(kP_0, nP_0), (kP_1, nP_1), \ldots, (kP_M, nP_M)\} \tag{2.37}$$

where $kP_i, i = 0, 1 \ldots, M$ represents the pilot index in the frequency direction, and $nP_i, i = 0, 1 \ldots, M$ represents the pilot index in the time direction. For this given pilot set, the least squares estimate of the channel can be given by [43]:

$$\begin{aligned} \hat{H}_{kP,nP} &= \frac{Y_{kP,nP}}{X_{kP,nP}} \\ &= H_{kP,nP} + \frac{N_{kP,nP}}{X_{kP,nP}} \end{aligned} \tag{2.38}$$

where $Y_{kP,nP}$ is the received pilot symbol on the k^{th} subcarrier in the frequency domain, $X_{kP,nP}$ is the transmitted pilot symbol on the k^{th} subcarrier in the frequency domain, and $N_{kP,nP}$ is the AWGN noise in the frequency domain. From this, the minimum mean square error estimate of the channel response $\hat{H}_{k,n}$ can be obtained via Wiener filtering as [43]:

$$\hat{H}_{k,n} = \sum_{kP,nP \in \mathbb{P}} w^{k,n}_{kP,nP} \hat{H}_{kP,nP} \tag{2.39}$$

OFDM summary

OFDM is currently the physical layer waveform of the WAVE protocol for vehicle-to-vehicle communication, and whilst OFDM is a deservingly popular waveform for

wireless communication, it is not without its drawbacks. The use of OFDM places strict requirements on the synchronization process which is needed to keep the subcarriers orthogonal, and with one cyclic prefix per symbol it presents a somewhat problematically low spectral efficiency [44]. It is also known for having a high peak to average power ratio, which effectively means the signals amplitude exhibits a high dynamic range. This high range of values poses issues for RF amplifiers as they require a larger linear region to uniformly amplify signals.

It would be beneficial to utilise a more flexible physical layer waveform that still allows for multiple users and protection from multipath environments, but also allows for better spectral management. For these reasons, non-orthogonal waveforms are investigated in the proceeding sections.

2.5.2 Generalised Frequency Division Multiplexing

Generalised frequency division multiplexing (GFDM) is a (typically) non-orthogonal multicarrier system that incorporates a flexible pulse shaping technique. GFDM was first proposed by Fettweis et. al. in [45], and is a current prospect for the physical layer waveform of 5G [46, 47]. At its core, it operates on blocks of data where each block consists of a number of subcarriers and subsymbols. Firstly, binary data is modulated onto a constellation $\mathbb{C}$ and split into sequences of length KM. Each of these sequences $d[\ell]$, $\ell = 0, 1, \ldots KM-1$ can be decomposed into blocks of K subcarriers and M time slots for transmission. This can be represented as a matrix, i.e. [45]

$$\begin{aligned}\mathbf{D} &= [\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{K-1}]^T \\ &= \begin{bmatrix} d_0[0] & d_0[1] & \cdots & d_0[M-1] \\ d_1[0] & d_1[1] & \cdots & d_1[M-1] \\ \vdots & & & \vdots \\ d_{K-1}[0] & d_{K-1}[1] & \cdots & d_{K-1}[M-1] \end{bmatrix}\end{aligned} \tag{2.40}$$

where $d_k[m] \in \mathbb{C}$ is the data symbol transmitted on the k^{th} subcarrier in the m^{th} timeslot.

GFDM transmitter

The block diagram for the GFDM transmitter is shown in Figure 2.6. Note that the K subcarrier streams have independent J-QAM mappers, therefore allowing different constellation orders on each stream relative to each subcarriers specific channel conditions.

On the k^{th} subcarrier of the transmitter, each data symbol $d_k[m], m = 0, 1, \ldots, M-1$ is upsampled by a factor of N, resulting in the sequence [48]:

$$d_k^N[n] = \sum_{m=0}^{M-1} d_k[m]\,\delta[n - mN], \quad n = 0, 1, \ldots, NM-1 \tag{2.41}$$

where $\delta[\cdot]$ is the Dirac delta function.

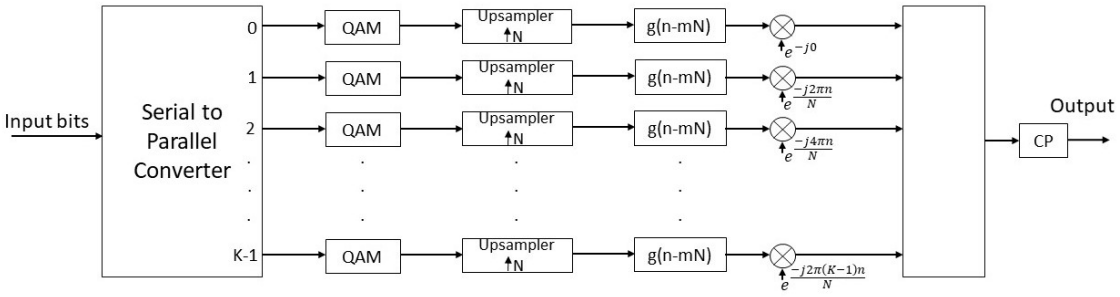

Fig. 2.6 GFDM transmitter block diagram

This sequence is then applied to a transmit filter $\tilde{g}[n]$ of length $L = NM$, with $N \geq K$ to avoid aliasing [45]. $g[n]$ defines the pulse shape that is to be applied to the subcarriers and is related to $\tilde{g}[n]$ as a circularly shifted version i.e [49].

$$\tilde{g}[n] = g[\langle n + MN/2 \rangle_{MN} - MN/2] \tag{2.42}$$

where $\langle \cdot \rangle_{MN}$ denotes a modulo MN operation. The relationship between $g[n]$ and $\tilde{g}[n]$ can be seen in Figure 2.7. The design of the pulse shaping filter has an influence in the overall performance of the system as it introduces ICI. By choosing $g[n]$ to be a root raised cosine (RRC) filter, the ICI introduced by each subcarrier can be limited to its adjacent subcarriers, thus allowing for a relatively low complexity iterative interference cancellation technique to be used, as will be introduced in a later subsection.

After pulse shaping, each subcarrier is up-converted and the resulting K modulated subcarriers are summed to produce the transmit signal. Mathematically this can be seen as [48]:

$$x[n] = \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} d_k^N[m]\tilde{g}[n - mN]\, e^{-j2\pi \frac{kn}{N}} \tag{2.43}$$

This can be re-expressed by gathering the filter and up-converting terms to produce a complex valued modulation matrix $\mathbf{A}$ of order $KM \times NM$, thus allowing the GFDM modulation process to be expressed as [50]:

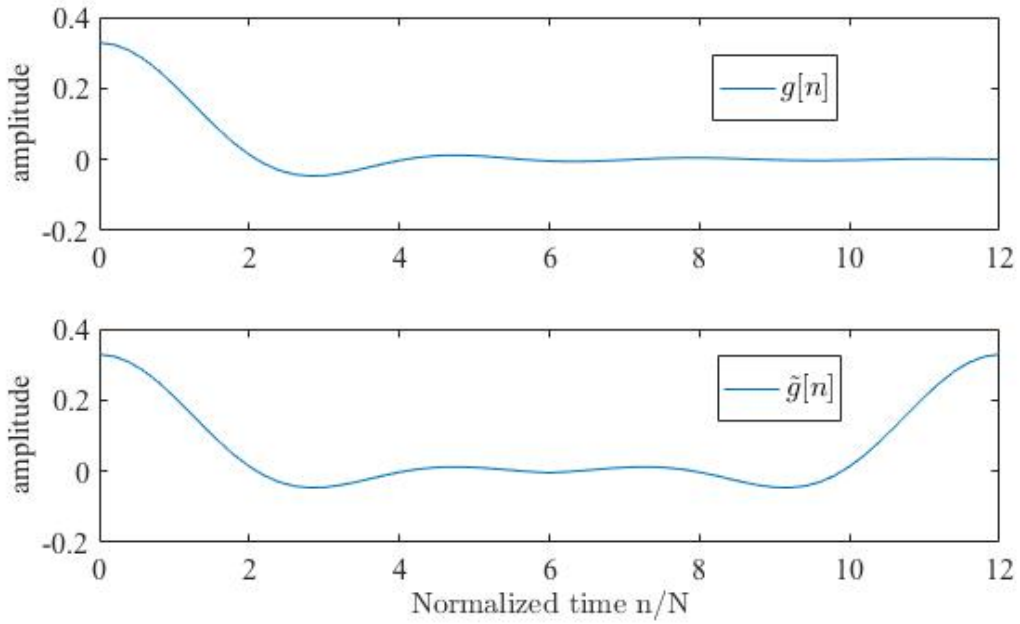

Fig. 2.7 Example GFDM pulse shape showing the relationship between $g[n]$ and $\tilde{g}[n]$ for root raised cosine pulse with $\alpha = 0.5, M = 5,$ and $N = 12$

$$\mathbf{x} = \mathbf{Ad} \tag{2.44}$$

This allows for easier descriptions of the chosen receivers, therefore this is the notation that shall be used in later sections. The composition of this matrix is a circularly shifted version of the pulse shaping filter in both time and frequency, which can be seen in Figure 2.8

Finally, as with OFDM, a cyclic prefix can be added to prevent inter-frame interference (IFI), an important distinction however is that GFDM only requires one CP per block as opposed to OFDM that requires one CP per timeslot. This is shown graphically in Figure 2.9.

GFDM receivers

Given the channel with impulse response $h[n]$, the received signal can be described as [42]:

$$r_{CP}[n] = x_{CP}[n] * h[n] + w[n] \tag{2.45}$$

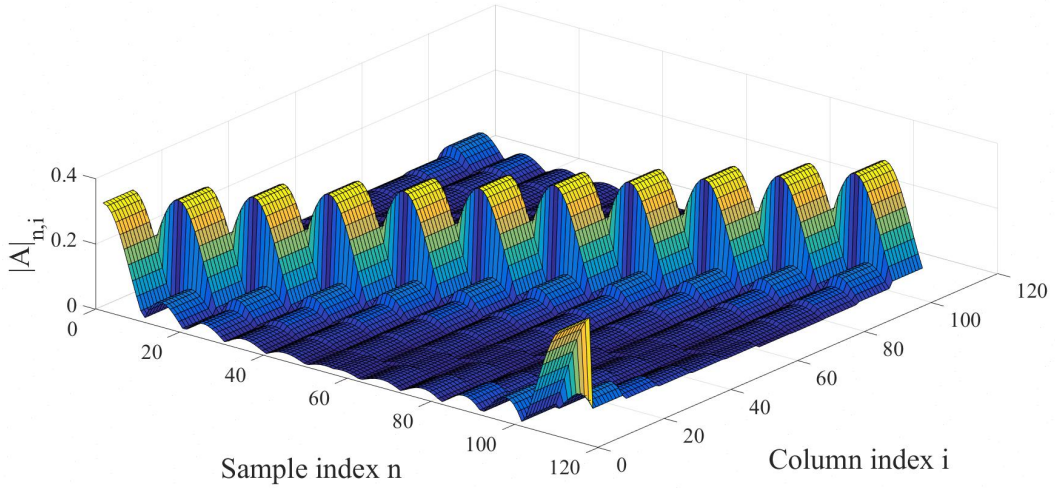

Fig. 2.8 GFDM transmitter matrix with $K = 10, M = 11,$ and $\alpha = 0.1$

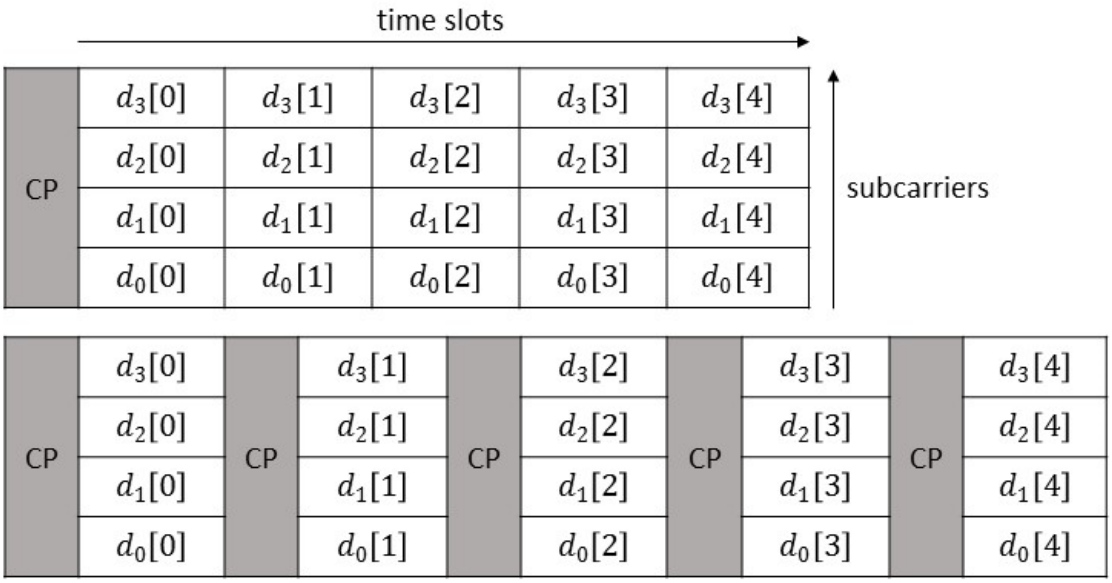

Fig. 2.9 Comparison of GFDM and OFDM block structures for $M = 5$ and $K = 4$

where x_{CP} denotes the transmitted GFDM frame with added CP and $w[n]$ are AWGN samples with distribution $\mathcal{N} \sim (0, \sigma^2)$. We assume that synchronization is handled by the receiver perfectly. The CP is then removed, which as previously mentioned has prevented IFI between two successive GFDM frames assuming a CP of length greater than the maximum channel delay spread. This gives the received signal, r as:

$$\mathbf{r} = \lfloor \mathbf{r}_{CP} \rceil \tag{2.46}$$

where $\lfloor \cdot \rceil$ denotes the slice operator. Assuming the channel can be estimated by the receiver, a zero forcing channel equaliser can be employed to compensate for the frequency response of the channel. Assuming a large enough K, the effect on each subcarrier can be considered frequency flat fading and thus a single tap per subcarrier equaliser can be used resulting in the equalised signal:

$$r_{eq}[n] = \text{IFFT}\left\{ \frac{\text{FFT}\{r[n]\}}{\text{FFT}\{h[n]\}} \right\} \tag{2.47}$$

where $\text{FFT}(\cdot)$ is the fast Fourier transform and $\text{IFFT}(\cdot)$ is the inverse fast Fourier transform respectively. The equalised signal can then be applied to the receiver. The next subsection will introduce some of these receivers, in increasing order of complexity.

GFDM matched filter receiver

The matched filter receiver is given by [44]:

$$\hat{\mathbf{y}}_{MF} = \mathbf{A}^H \mathbf{r}_{eq} \tag{2.48}$$

This operation maximises the signal-to-noise ratio (SNR) on each subcarrier, however, this has the effect of introducing ICI when a non-orthogonal pulse shape is used [44]. The default pulse shape often considered is the root raised cosine pulse, and is thus non-orthogonal, hence the bit-error rate (BER) performance is expected to be negatively effected. In terms of complexity however, this receiver is simple to implement.

GFDM zero forcing receiver

By formulating the GFDM transmitter operation as a matrix in a previous subsection, the zero forcing operation can be seen as the inverse, thus following [42]:

$$\hat{\mathbf{y}}_{ZF} = \mathbf{A}^{-1} \mathbf{r}_{eq} \tag{2.49}$$

However, as pointed out in [51], the transmitter matrix $\mathbf{A}$ may be non-square or ill conditioned meaning the inverse does not exist. In such a case, the Moore-Penrose psuedo-inverse can be used and is given by [52]:

$$\mathbf{A}^{+} = \mathbf{A}^{H}\left(\mathbf{A}\mathbf{A}^{H}\right)^{-1} \tag{2.50}$$

where $\mathbf{A}^{H}$ is the Hermitian matrix of $\mathbf{A}$. The ZF receiver will completely remove any ICI but at the cost of potential noise enhancement which can decrease BER performance.

GFDM minimum mean square error receiver

The linear minimum mean square error receiver is given by [45]:

$$\hat{\mathbf{y}}_{MMSE} = \left(\left(\mathbf{R}_{w}^{2} + \mathbf{A}^{H}\mathbf{H}^{H}\mathbf{H}\mathbf{A}\right)^{-1}\mathbf{A}^{H}\mathbf{H}^{H}\right)\mathbf{r}_{eq} \tag{2.51}$$

where $\mathbf{H}$ is the circulant convolution matrix representation of the channel and $\mathbf{R}_{w}^{2}$ denotes the covariance matrix of the noise. This results in a receiver that reaches a compromise between the matched filter and the zero forcing receivers in terms of self-interference versus noise enhancement. Note in the case of the minimum mean square error receiver the channel is jointly equalised in this process, thus eliminating the need for the zero forcing precursor mentioned previously.

GFDM interference cancellation receiver

As mentioned previously, by choosing the pulse shaping filter to be a RRC filter the ICI can be limited to adjacent subcarriers only as shown in Figure 2.10.

The transmitter matrix $\mathbf{A}$ is readily reproducible at the receiver, and can be used to recreate an estimated interference signal and iteratively remove ICI from the received signal. The process can be described as follows. To reduce the ICI on the k^{th} subcarrier caused by the $(k-1)^{\text{th}}$ and $(k+1)^{\text{th}}$ subcarriers, first apply the matched filter to $\mathbf{r}_{eq}$ and estimate the symbols from the reference constellation being used, $\hat{\mathbf{d}}_{k-1}$ and $\hat{\mathbf{d}}_{k+1}$. To recreate the estimated interference signal, create a dummy data frame $\mathbf{c}$ that contains the estimated $\hat{\mathbf{d}}_{k-1}$ and $\hat{\mathbf{d}}_{k+1}$ and zeros elsewhere, then multiply by the transmitter matrix $\mathbf{A}$ as [53]:

$$\mathbf{v}_{k} = \mathbf{A}\mathbf{c} \tag{2.52}$$

This interference signal is then be subtracted from the original signal $\mathbf{r}_{eq}$ to clean subcarrier k of ICI [53].

$$\mathbf{r}'_{eq} = \mathbf{r}_{eq} - \mathbf{v}_{k} \tag{2.53}$$

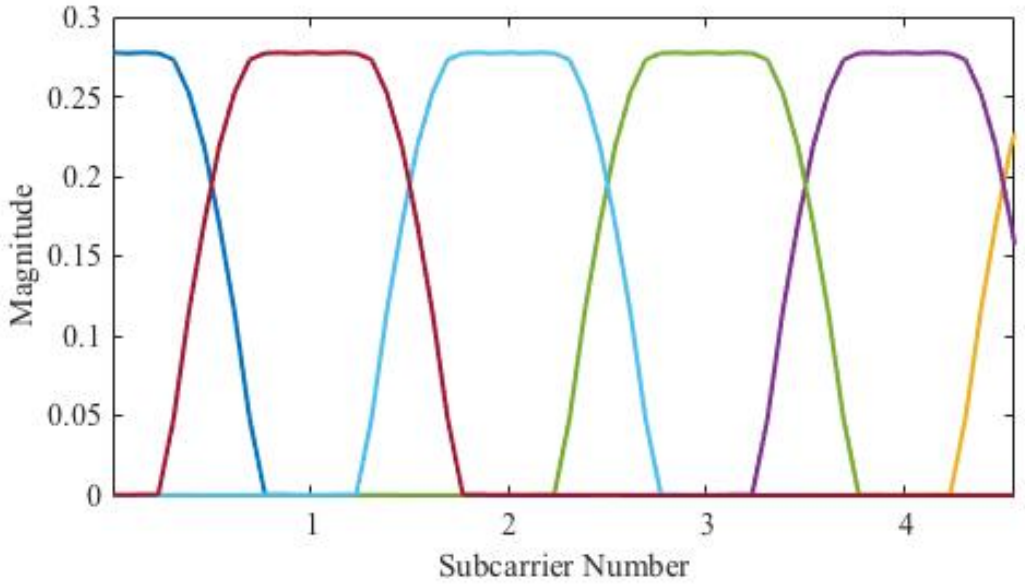

Fig. 2.10 RRC interference pattern - $\alpha = 0.5, M = 13, N = K$

The data symbols on subcarrier k are then re-estimated. This process continues for all K subcarriers, and the most recent estimated symbols are used in the next iteration. The entire process can be repeated I times until the desired level of ICI is reached. The more iterations that are performed, the lower the resultant ICI, however at the clear cost of computational complexity.

GFDM summary

GFDM can be seen as almost a superset of OFDM, but this flexibility often comes at the expense of complexity. GFDM has several attractive qualities when compared to OFDM, through an appropriate selection of pulse shape the peak to average power ratio of GFDM is lower than that of OFDM, and less spectral leakage is achieved which reduces the out of band emissions - ideal for flexible spectrum resource sharing. In GFDM, each subcarrier can also be modulated individually with its own indepent constellation mapper, allowing for a high degree of freedom for multi-user scheduling. This of course comes at the cost of complexity and loss of orthogonality, which detriments the overall bit-error rate performance.

2.5.3 Spectrally Efficient Frequency Division Multiplexing

Spectrally efficient frequency division multiplexing (SEFDM) is another non-orthogonal multicarrier system that was first proposed by Darwazeh et. al. in [54] that is a current prospect for the physical layer waveform of 5G [55]. In SEFDM, the orthogonality constraint of OFDM is purposefully broken for the purposes of utilising more subcarriers per given bandwidth than in OFDM, however, ISI is introduced as a result. It does this by compressing more subcarriers into a given bandwidth, placing them at locations that are below the symbol rate. This compression factor is denoted by α, and is directly linked to the bandwidth saving with respect to an equivalent OFDM signal i.e. a value of $\alpha = 0.8$ means the SEFDM system requires $0.8 \times 100\%$ of the bandwidth, or often it will simply be denoted as $(1-\alpha) \times 100\%$ bandwidth saving. An example of this can be seen in Figure 2.11.

SEFDM transmitter

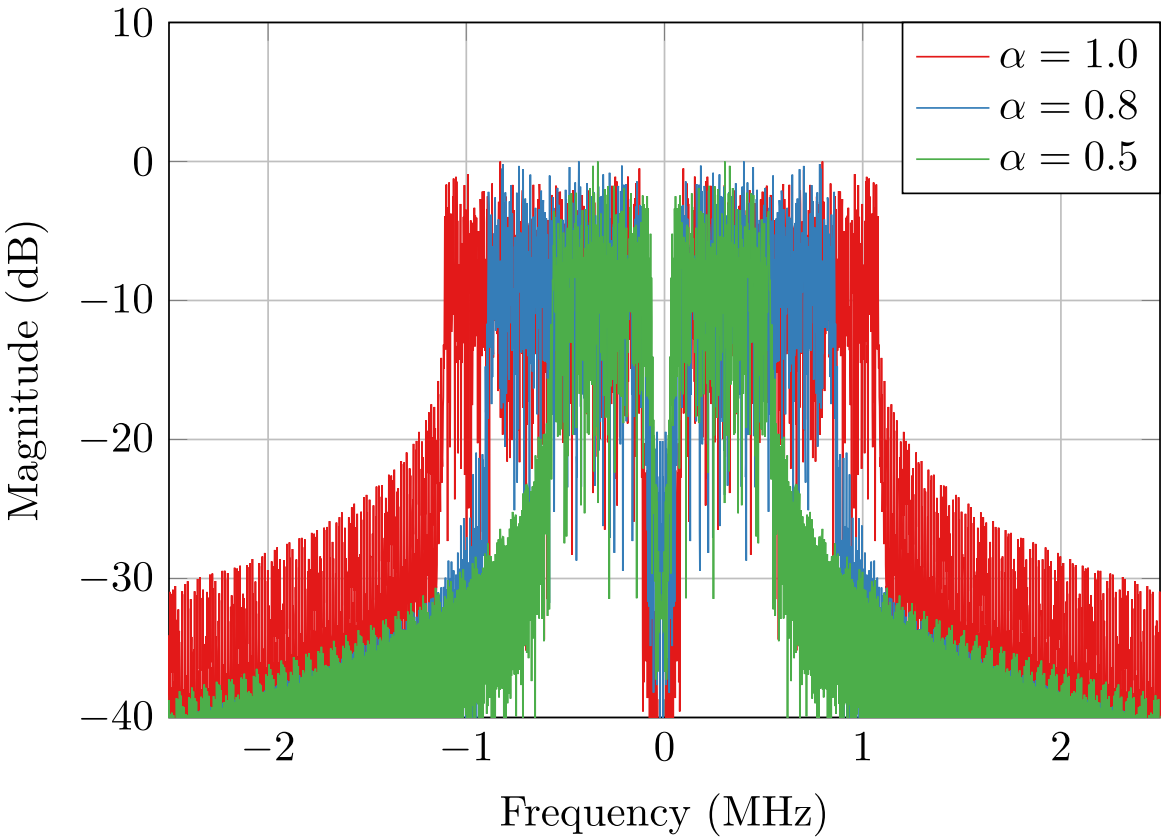

Fig. 2.11 Power spectrum of SEFDM signals with varying value of α and OFDM for reference

An SEFDM signal is created similarly to OFDM by modulating an input data stream onto parallel subcarriers. Unlike in OFDM where subcarriers are spaced evenly at intervals of the symbol period T_s to retain orthogonality, in SEFDM the carriers are spaced in frequency by some fraction of the symbol time according to the value of α for the system via $\Delta f = {}^{\alpha}/_{T_s}$. This results in the following mathematical representation for the creation of a complex baseband SEFDM signal as [56]:

$$x\left(t\right) = \frac{1}{\sqrt{T}} \sum_{l=-\infty}^{\infty} \sum_{n=0}^{N_c-1} s_{l,n} \exp^{\frac{(j2\pi\alpha(t-lT))}{T}} \tag{2.54}$$

where N_c is the total number of subcarriers and $s_{l,n}$ is the symbol that is modulated on the n^{th} subcarrier in the l^{th} SEFDM frame. It is worth noting that for $\alpha = 1$ this reduces to the OFDM case. The representation above in (2.54) however does not lend itself well to implementation using standard hardware for OFDM via the FFT and IFFT as it requires a bank of modulators each set at the subcarriers frequencies. In order to generate SEFDM signals via an IFFT approach, it helps to redefine (2.54) in matrix form.

Considering only the first frame sampled at intervals of ${}^{T}/_{N}$ where $N \geq N_c$, one can express a frame for $k = 0, 1, \ldots, N-1$ as [56]:

$$X\left[k\right] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_n \exp^{\left(\frac{j2\pi nk\alpha}{N}\right)} \tag{2.55}$$

where $X[k]$ is the k^{th} time sample of the first frame of $x(t)$, and s_n is the n^{th} symbol in the frame. Noting the highlighted term α, one can see by comparing this is to (2.34) that this is a minor modification to the FFT/IFFT method utilised by OFDM. Furthermore, this allows for a direct expression in matrix form as [57]:

$$X = \Phi S \tag{2.56}$$

where $X = \left[x_0, \ldots, x_{N-1}\right]^T$ is a vector of the transmitted symbols of $x(t)$ and $S = \left[s_0, \ldots, s_{N_c-1}\right]^T$ is a vector of input symbols, and $[\cdot]^T$ denotes the transpose operation. This results in Φ forming an $N_c \times N$ matrix whose elements contain the sampled SEFDM subcarriers values [56]:

$$\phi_{n,k} = \frac{1}{\sqrt{N}} \exp^{\left(\frac{j2\pi\alpha nk}{N}\right)} \tag{2.57}$$

for $0 \leq n \leq N_c - 1$ and $0 \leq k \leq N-1$. This matrix form of SEFDM allows easy relation to the IFFT implementation of OFDM. Referring back to (2.57), to retain equivalence

after the IDFT operation, the SEFDM symbols within S must be of length $M = {}^{N}/_{\alpha}$. This can be achieved by zero-padding S by appending $M - N$ zeros such that [58]:

$$\hat{S} = \begin{cases} S_i, & 0 \leq i < N \\ 0, & \mathrm{N} \leq i < M \end{cases} \tag{2.58}$$

Assuming that ${}^{N}/_{\alpha}$ is an integer, the resulting SEFDM signal can be expressed as [58]:

$$\hat{X} = \frac{1}{\alpha} Q \hat{S} \tag{2.59}$$

such that $X[k'] = \hat{X}[k']$ for $1 < k' < N$, where Q is an $M \times M$ IFFT matrix. The resulting signal is shown in Fig 2.11 for different values of α.

SEFDM receivers

An SEFDM receiver is comprised of a demodulator and a detector. It is the job of the demodulator to collect statistics R of the received signal by projecting it onto a set of orthonormal bases $b_i(t)$, whilst it is the job of the detector to utilise those statistics by applying algorithms to estimate the originally transmitted symbols. Using an ideal AWGN channel, without loss of generality, a received signal can be given by [42]:

$$y(t) = x(t) + w(t) \tag{2.60}$$

where $w(t)$ is an additive white Gaussian noise term, the i^{th} collected statistics r_i is obtained as [59]:

$$r_i = \int_0^T y(t)\, b_i^*(t)\, dt, \quad i = 0, 1, \ldots, N-1 \tag{2.61}$$

where $b_i(t)$ is the i^{th} base and $[\cdot]^*$ is the Hermitian conjugate operation. Due to the lack of orthogonality introduced by the modulation process and α, the collected statistics contain contributions from non-orthogonal subcarriers. Referring back to (2.57), and noting that the columns of ϕ represent the signals subcarriers, the correlation between two arbitrary subcarriers ϕ_m and ϕ_n is given by [59]:

$$\begin{aligned} c_{m,n} &= \langle \phi_m, \phi_n \rangle \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N-1} \exp^{\frac{j2\pi\alpha km}{N}} \exp^{-\frac{j2\pi\alpha kn}{N}} \\ &= \exp^{j\pi\alpha(m-n)} \exp^{\frac{-j\pi\alpha(m-n)}{N}} \left[\frac{\operatorname{sinc}\left(\alpha\left(m-n\right)\right)}{\operatorname{sinc}\left(\frac{\alpha(m-n)}{N}\right)} \right] \end{aligned} \tag{2.62}$$

The amplitude of which is given by the sinc term as [58]:

$$|c_{m,n}| = \left| \frac{\operatorname{sinc}\left(\alpha\left(m-n\right)\right)}{\operatorname{sinc}\left(\frac{\alpha(m-n)}{N}\right)} \right| \tag{2.63}$$

From (2.62), this allows for the expression of a correlation coefficient matrix C as [59]:

$$C = \begin{cases} 1, & \text{m} = \text{n} \\ \frac{1}{N} \left[\frac{1-\exp^{j2\pi\alpha(m-n)}}{1-\exp^{\frac{j2\pi\alpha(m-n)}{N}}} \right], & \text{m} \neq n \end{cases} \tag{2.64}$$

where C is a Toeplitz and Hermitian $N \times N$ matrix which characterises the energy spillage between the N subcarriers. This is demonstrated graphically in Fig 2.12 for two different values of α to show the increase in energy spillage as the bandwidth is increasingly compressed.

SEFDM zero forcing receiver

Gathering up all the statistics and bases from (2.61) as the vecors $\mathbf{r}$ and $\mathbf{b}$ respectively, one can describe the maximum likelihood detection problem as a least squares minimisation problem as [58]:

$$\text{min.} \left|\left| \mathbf{r} - \mathbf{bs} \right|\right|^2 \tag{2.65}$$

where $\mathbf{s}$ are the originally transmitted symbols such that $\mathbf{s}$ is a symbol from some reference constellation $\mathbb{C}$. The zero forcing estimate, which is also known as the Babai estimate [60], $\hat{\mathbf{s}}$ is given by [58]:

$$\hat{\mathbf{s}} = \lfloor \tilde{\mathbf{s}} \rceil = \left\lfloor \mathbf{b}^{-1}\mathbf{r} \right\rceil \tag{2.66}$$

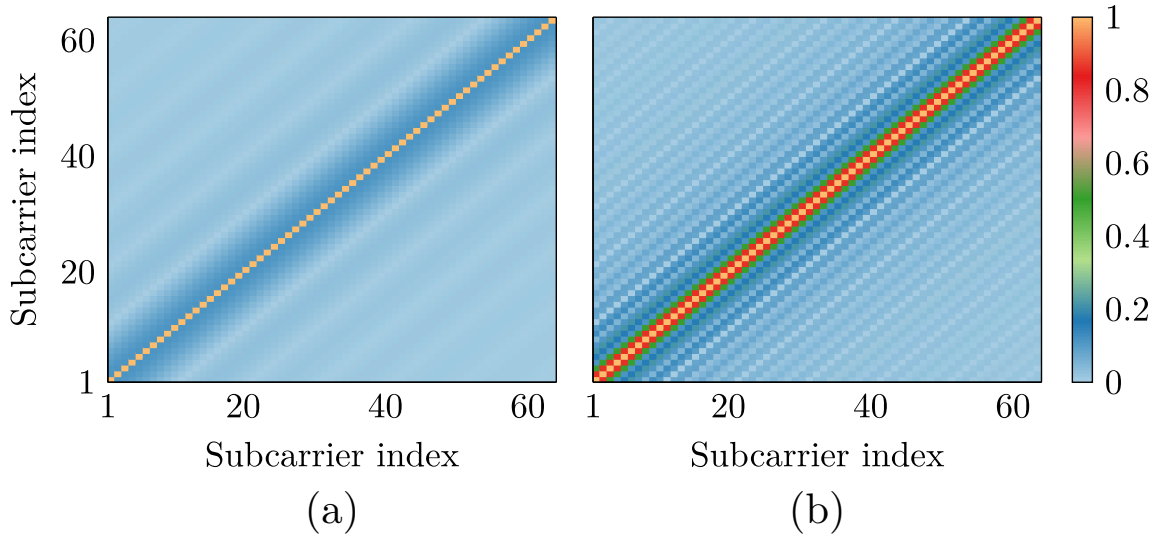

Fig. 2.12 Correlation matrix for $N = 64$ and (a) α=0.9 and (b) α=0.3

where $\lfloor \cdot \rceil$ denotes the slice operator. Much like other zero forcing algorithms investigated, this too has the unwanted side effect of enhancing the noise power when the determinant of $\mathbf{b}$ becomes small, and also has an algorithmic complexity of $O(N^3)$ due to the required inversion of $\mathbf{b}$.

SEFDM minimum mean square error receiver

The full derivation of the minimum mean square error receiver can be found in [58], however, it is reported to be given by:

$$\hat{\mathbf{s}} = \left\lfloor \mathbf{b}^H \left(\mathbf{b}\mathbf{b}^H + \frac{\sigma^2}{\sigma_s^2}\mathbf{I} \right)^{-1} \mathbf{r} \right\rceil \tag{2.67}$$

where $\mathbf{b}^H \left(\mathbf{b}\mathbf{b}^H\right)^{-1}$ is the Moore-Penrose psuedoinverse [52] of the matrix $\mathbf{b}$, $\mathbf{I}$ is the identity matrix, and σ^2/σ_s^2 represents the inverse of the signal to noise ratio following [61].

Logically, one can infer that this is a more effective receiver than the zero forcing receiver. The use of the Moore-Penrose psuedoinverse means that it will never be singular, and therefore will not be limited by the singularity of the projection matrix like in zero forcing. The regularisation term σ^2/σ_s^2 dynamically adapts based on the noise level, therefore experiences less noise enhancement issues, and in a noise-less environment one can see that this solution converges to that of zero forcing.

SEFDM interference cancellation receiver

If one was to write out the analytical expresion for (2.60) as [62]:

$$\begin{bmatrix} r_1 \\ \vdots \\ r_{N-1} \\ r_N \end{bmatrix} = \begin{bmatrix} b_{11} & \dots & b_{1,N-1} & b_{1,N} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & b_{N-1,N-1} & b_{N-1,N} \\ 0 & \dots & 0 & b_{N,N} \end{bmatrix} \begin{bmatrix} s_{11} \\ \vdots \\ s_{N-1} \\ s_N \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_{N-1} \\ w_N \end{bmatrix} \tag{2.68}$$

one notices that $\mathbf{b}$ is triangular. Starting from the N^{th} symbol, the estimate for $\hat{s}_N$ is [63]:

$$\hat{s}_N = \left\lfloor \frac{r_N}{b_{N,N}} \right\rceil \tag{2.69}$$

Estimating the next symbol $\hat{s}_{N-1}$ and subtracting the interference due to $_N$ gives [63]:

$$\hat{s}_{N-1} = \left\lfloor \frac{1}{b_{N-1,N-1}} \left(r_{N-1} - b_{N-1,N} \hat{s}_N \right) \right\rceil \tag{2.70}$$

This process is then repeated until the first symbol is reached. This can be neatly summed up by the following recursive equation for the estimate of the i^{th} symbol as [63]:

$$\hat{s}_i = \left\lfloor \frac{1}{b_{ii}} \left(r_i - \sum_{j=i+1}^{N} b_{i,j} \hat{s}_j \right) \right\rceil \tag{2.71}$$

2.6 Application of Neural Networks

Machine learning has had a profound impact on modern life. It powers much of the digital world we interact with daily, from the relatively innocuous such as shopping recommendations and email filtering, to the grandiose such as autonomous vehicles and space exploration [64]. The aim of machine learning is to provide a system the ability to automatically learn and improve based on experience without being explicitly programmed. It explores the numerical relationships inherent in data and uses that to arrive at conclusions. Broadly speaking, there are three main types of machine learning algorithm; supervised learning, semi-supervised learning, and unsupervised learning. At a basic level, in supervised learning the system is shown examples of inputs and outputs and is tasked with learning the function that is capable of mapping those inputs to the example outputs, i.e. given an input set $\mathbf{x}$ and example output set $\mathbf{y}$

the machine learning algorithm attempts to approximate some function λ such that $\lambda(\mathbf{x}) = \mathbf{y}$. Unsupervised learning on the other hand does not make use of an example set $\mathbf{y}$, instead attempting to derive structure organically present within $\mathbf{x}$ however it can. The final type, semi-supervised learning, as one can imagine from the name, is a hybrid approach. As the volume of data collected worldwide is increasing exponentially year on year [65], creating or maintaining an accurate supervised learning output set $\mathbf{y}$ can become very costly in some instances, therefore the aim of semi-supervised learning is to combine elements of both approaches, depending on the particular problem statement.

Supervised learning is the most common of the three, and will be the main focus of this research. The nature of wireless communication research lends itself well to this problem, since we control both the transmitter and the receiver - i.e. both the producer of $\mathbf{y}$ and consumer of $\mathbf{x}$.

2.6.1 Neural Network Training

A machine learning model, specifically the neural networks that this work will consider, can be seen as a collection of composable functions connected via weights and biases, a simple diagram is shown in Fig. 2.13. Each circle in the diagram is a neuron, which are organised into layers where the data flows from the input(s) on the left to the output(s) on the right. Each arrow is a connection inside the model, which has its own weight and bias matrix, and a (usually non-linear) activation function. This can be seen in Fig. 2.14. These weight and bias matrices are referred to as the models parameters, and it is these parameters that need to be learned during the training process to generate accurate outputs.

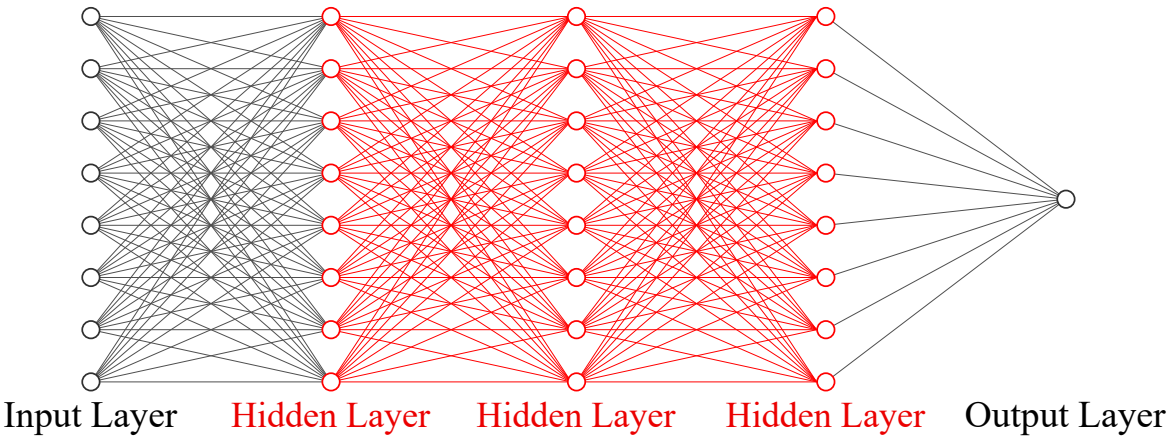

Fig. 2.13 An example neural network layout with three hidden layers; hidden layers are highlighted in red

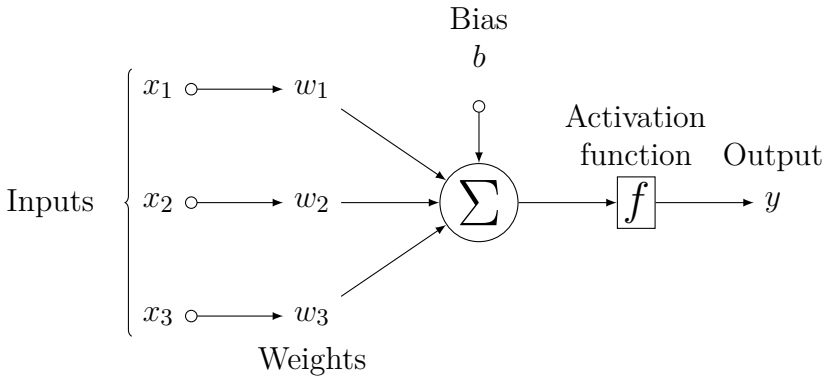

Fig. 2.14 The input to output process of a neuron showing the weight, bias, and activation function

As was discovered by works such as [66, 67], networks configured in this way can be trained via an algorithm called gradient descent and backpropagation. The key finding was that the derivative of the objective function with respect to the input of a neuron can be calculated by working backwards from the gradient with respect to the output of that neuron or the input of the following neuron. Backpropagation can be repeatedly applied in this manner to propagate gradients through all neurons in the model, starting from the output(s) and flowing through to the input(s). Once all the gradients are computed, one can calculate the gradients with respect to the neurons parameters for each neuron. The premise behind training a neural network in this context is as follows; apply $\mathbf{x}$ to the input(s) of the neural network, the neural network applies its internal functions via the weights, biases and activation functions at each successive neuron and layer in the model, and a prediction, $\hat{\mathbf{y}}$, is generated at the output(s). The training algorithm then compares that output via a loss function (sometimes called cost function) to the ground truth label, $\mathbf{y}$, and creates a loss value $L = J(\theta)$ where θ are the models trainable parameters. For a single layer neural network, this process goes as follows [66]:

$$\hat{y}_k = \sum_{i=1}^{M} w_{k,i} x_i + b_k \tag{2.72}$$

where $\hat{y}_k$ is the output from neuron k, M is the total number of examples in the dataset, x_i is the input sample from the dataset in step i, and b_k is the bias term for neuron k. Or in matrix form [66]:

$$\hat{\mathbf{y}} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{2.73}$$

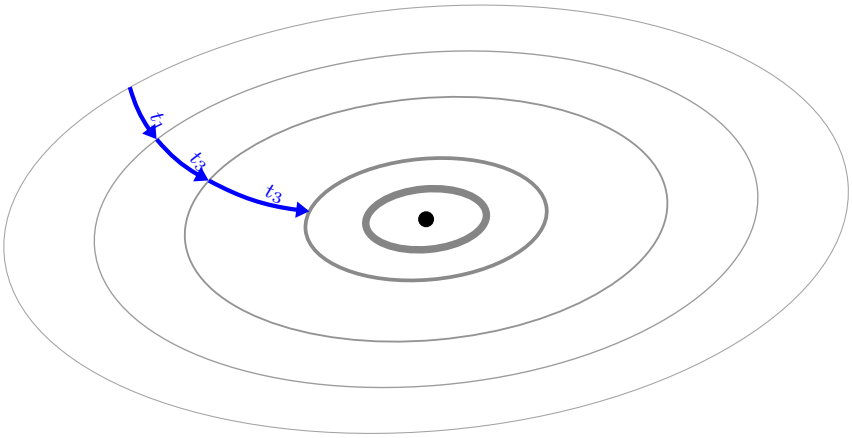

Fig. 2.15 Gradient descent visualised

For example purposes, let us utilise the mean square error as an example loss function [42]:

$$\begin{aligned} E_n &= \frac{1}{2}||\hat{\mathbf{y}} - \mathbf{y}||^2 \\ &= \frac{1}{2}\sum_{m=1}^{M}(\hat{y}_m - y_m)^2 \end{aligned} \tag{2.74}$$

It is this relationship between the learnable parameters of the neural network, θ and the loss, $L = J(\theta)$, that gives rise to a high dimensional space where all possible values of θ map to values of loss. It is the job of gradient descent to iterate over this space to find values of θ that minimise L. It does this by moving iteratively in the opposite direction to the gradient, which happens where L is decreasing most rapidly. We first must decide what type of gradient descent we are using. When training a neural network, there are three general methods of gradient descent that can be opted for, namely batch gradient descent, minibatch gradient descent, and pure stochastic gradient descent. The difference between these methodologies is the subtle difference of how many data points from the dataset are used in each iteration of the training process. Pure stochastic gradient descent takes a random point from the dataset and uses this sample to perform a training update. This data point could take the gradient in a direction that is completely off course from the wider dataset, however, this is not a bad thing as it can aid in escaping from saddle points [68], which are points in the loss space that have zero gradient but are not local minima. In batch gradient

descent, the gradient is calculated over the entire dataset, meaning the resulting loss vector is averaged over many points and those noise-like perturbations from single data points are lost. This is the reason one can end up in a local minima of the loss function or as mentioned previously a saddle point. Depending on the dataset, this can also be very memory intensive too. Comparing these two approaches, one can see there is a balance between computation and optimisation accuracy i.e. performing an expensive parameter update after every single sample as in the pure stochastic gradient descent approach, and potentially stalling the gradient descent process in local-minima and saddle points as in the batch gradient descent case. It is for this reason that the minibatch method is typically used, and will be the method discussed from here on unless explicitly stated. As one can imagine, the minibatch approach is a compromise of these two methods, using more than one sample at a time to perform updates to speed up convergence yet not using too many samples as to retain enough noise-like fluctuations from the dataset to escape local-minima and saddle points. Mathematically, we can view the parameter update process from minibatch gradient descent as [69]:

$$\theta_{k+1} = \theta_k + \eta \Delta\theta_k \tag{2.75}$$

where

$$\Delta\theta_k = -\frac{1}{m}\sum_{i-1}^{m} \nabla_\theta J_i\left(\theta_k\right) \tag{2.76}$$

where η is the learning rate of the system i.e. how large a step to take in the gradient descent process, m is the minibatch size, $m \leq M$ where M is the size of the training dataset.

Whilst (2.75) forms the basis of most computationally feasible neural network training algorithms, there are several optimisation algorithms in prevalent use that look to address specific problems and speed up training convergence. For instance, it was shown in [70] that increasing minibatch size m whilst retaining a constant η allows for greater parallelisation but uses potentially stale gradients, and that a higher η can be used with a smaller minibatch size m as each weight update contains lower variance. The authors in [71] introduced an optimisation method called adagrad, short for adaptive gradient. In adagrad, the learning rate η is not fixed for all parameters θ but instead it adapts η and performs smaller updates on parameters that are associated with frequently occuring features in the dataset and larger updates on parameters that are associated with infrequently occuring features. This helps the neural network when

dealing with sparse data, as the sparse features are usually those of interest and are thus weighted higher when they occur. This means one has to adapt (2.75) to allow for per-parameter updates at each time step t. For brevity, we define the gradient at time step t as [71]:

$$g_{t,i} = \nabla_\theta J(\theta_{t,i}) \tag{2.77}$$

which subsequently allows us to write the update for each parameter θ_i as [71]:

$$\theta_{t+1,i} = \theta_{t,i} - \eta g_{t,i} \tag{2.78}$$

For adagrad to modify the learning rate η at each time step per parameter based on feature occurence, it utilises past historical gradients that have been calculated [71]:

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} g_{t,i} \tag{2.79}$$

where $G_t \in \mathbb{R}^{d \times d}$ is a diagonal matrix where each element on the diagonal i, i is the sum of the squares of the gradients for θ_i for all previous time steps, ϵ is a smoothing term (typically on the order of 10^{-8}) that provides numerical stability by avoiding cases where one might divide by zero.

As one can intuitively see however, whilst this does address some issues it is also not without its drawbacks. The need to record gradient updates for all parameters for all time can be memory intensive, and since the matrix G_t is always updated by addition, the sum increases over the lifetime of the training period until it eventually causes updates to $\to 0$ and no more learning can take place.

2.6.2 Generalisability

A good machine learning model is one that has generalised well to the problem statement one set out to solve. Whilst this statement may seem obvious, generalisation of machine learning models has been a huge issue for a long time and likely always will be e.g. [72–74]. The reason for lack of generalisability of a model could be a multitude of things, however, the main issues normally crossed include; the dataset being used to train the model, the model architecture implemented relative to the problem, and the training process used to teach the model.

When the dataset used whilst training the model has different properties to that of the data seen when the model is deployed, performance degradation is likely. This result can often come as a surprise to many, who do not realise the properties between

training and deployment have changed. Take for instance the work in [75], the authors demonstrated a machine learning system for wireless communications that provided large performance gains in training, however under-performed in reality. The cause was deemed to be the dataset used to train the model, the 'random sequence' used to generate the bit patterns was flawed, resulting in repeating patterns in the dataset. The machine learning model had learned how to map this dataset $\mathbf{x}$ to output $\mathbf{y}$ but the function it learned to do so was unintended, thus the model failed to generalise in real-world deployment. There exists an interesting research topic in this area, namely the so called 'manifold learning' hypothesis [76, 77]. The premise of this hypothesis is that the high dimensionality of many datasets (thus high complexity of learning its consequent mapping function λ) was artificial, and each example could be described with fewer underlying parameters i.e. each example could be seen as a sample from a lower dimension manifold. Worth mentioning are related fields to this in the space of dimensionality reduction, with popular methods such as principle component analysis (PCA) with its seminal paper in [78], and t-distributed stochastic neighbour embedding (t-SNE) first introduced in [79]. The use of t-SNE is prevalent in the exploration of high dimensional data, as it provides an intuitive way of reasoning with complex internal relationships. The introduction of t-SNE was brought about to make improvements to the standard stochastic neighbour embedding (SNE) algorithm, namely the use of a symmetric cost function with simpler gradients and a replacement of the Gaussian distribution with the Student-t distribution to compute point similarity in low-dimensional space. The aim of t-SNE is to minimise the divergence between two distributions, a distribution that measures pairwise similarities of the input data points and a distribution that measures pairwise similarities of the corresponding low dimensional points in the embedding. For instance, given an input set of high dimensional data points $D = \{x_1, x_2, \ldots_N\}$ and some function to compute a distance between any pair of data points in the set $d(x_i, x_j)$ (e.g. Euclidean distance $d(x_i, x_j) = ||x_i - x_j||$), the aim is to learn some s-dimensional embedding in which each data point from D is represented by a point $\varepsilon = \{y_1, y_2, \ldots_N\}$ with $y_i \in \mathbb{R}^s$. To do this, t-SNE defines a set of joint probabilities p_{ij} that measure the pairwise similarity between data points x_i and x_j by creating two conditional probabilities as [80]:

$$p_{j|i} = \frac{\exp^{-d\left(x_i, x_j\right)^2 / 2\sigma_i^2}}{\sum_{k \neq i} \exp^{-d(x_i, x_k)^2 / 2\sigma_i^2}} \tag{2.80}$$

and

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \tag{2.81}$$

where σ_i^2 are Gaussian kernels that are set such that the perplexity of the conditional distribution P_i is equal to a predefined perplexity value u. The optimal value for which can be found via binary search [81]. The second distribution required finds the similarities between two points, y_i and y_k in the s-dimensional embedding ε by computing a normalised Student-t kernel with a single degree of freedom as [80]:

$$q_{ij} = \frac{\left(1 + ||y_i - y_j||^2\right)^{-1}}{\sum_{k \neq l} \left(1 + ||y_k - y_l||^2\right)^{-1}} \tag{2.82}$$

The heavy tails of the normalised Student-t kernal ensure that dissimilar inputs x_i and x_j can be modelled by low dimensional equivalents y_i and y_j that are far apart. Finally, the embedding points y_i are determined by minimising the Kullback-Leibler divergence between the two joint distributions P and Q as [79]:

$$C(\varepsilon) = \sum_{i \neq j} p_{ij} \frac{p_{ij}}{q_{ij}} \tag{2.83}$$

which can be optimised by gradient descent [79]:

$$\frac{\delta C}{\delta y_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij}) q_{ij} \tag{2.84}$$

where $Z(y_i - y_j)$ is a normalisation term equal to [79]:

$$Z = \sum_{k \neq l} \left(1 + ||y_k - y_l||^2\right)^{-1} \tag{2.85}$$

In the context of exploring dimensionality, we can describe the lack of generalisation of a neural network in terms of higher dimensional overfitting i.e. the neural network learned the high dimensional features with high variance that failed to transfer when using new data. In any neural network, we can describe the total error as being a combination of two terms; a constant error term, bias, and a varying error term that depends on the dataset, variance. A neural network with high variance is likely to generalise poorly, as it can mean that the neural network has focussed too much on the particular structure within the training dataset rather than the underlying lower dimensional features. A neural network that has a high bias is symptomatic of not being powerful enough to capture the underlying complexities within the data using the given features. This is often reffered to as the inherent bias-variance tradeoff [82].

2.6.3 Conclusion

This chapter has presented the wireless standards that guide this work, provided a theoretical background into the challenges of wireless communication in a vehicular network, and introduced the part that machine learning has to play in this area of research. The physical properties that govern wireless radio communication were discussed, it was shown how multipath propagation, ISI, path loss, shadowing, and Doppler effects are all natural phenomena that affect a wireless signal as it traverses through an environment. Following the radio background, it was then investigated how a VLC network operates for vehicular networks, which sets the scene for later practical experiments. It was then discussed how the physical layer waveforms used throughout this work are generated in a transmitter and subsequently received with different methods. Finally, the technical background behind neural networks and their training was discussed, highlighting their capability as teachable mathematical models that can generalise relationships between sets of inputs and outputs in preparation for implementation in later experiments.

Chapter 3

Neural Network Equalisation and Symbol Detection

3.1 Introduction

Vehicular communication, specifically vehicle-to-vehicle communication, faces many challenges that differentiate it from more classic communication systems. One of these challenges is the harshness of the communication channel. Both the transmitter and receiver are at a similar height, which causes multiple reflections from surroundings that have slight path length differences, causing fast fading [22, 27]. This is exacerbated by Doppler effects due to the relative motion between the vehicles and the fact that the network topology is changing rapidly [83].

There have been several proposals to overcome these issues in the literature. In [84], the authors propose the use of lasers to achieve high speed communication over several hundred metres. One problem with this however is that line-of-sight is required between devices which cannot be guaranteed in a vehicular setting due to the configuration of roads and intersections and also the intermittent blockages caused by other road users. Another proposal introduced in [85] attempts to predict Doppler changes to allow for easier compensation of the channel effects, the results outperform a benchmark least squares estimation, however only simulation results are presented. Thus, its ability to perform under real world constraints are unknown and require further investigation. Further proposals include the use of multiple input multiple output (MIMO) systems [86], cooperative communication [87], and integration into existing long term evolution (LTE) and future 5G systems [88]. There have also been several proposals to use machine learning in the vehicular network, such as in [89] where they investigate the use of artificial neural networks (ANNs) to increase throughput in a vehicular ad-hoc

network by optimising the medium access control (MAC) layer of the 802.11p protocol. Again, in [90], the use of an online machine learning algorithm is proposed for beam selection in millimetre-wave (mmWave) vehicle-to-anything (V2X) communication to allow network adaptation to traffic and line-of-sight blockages, and the authors in [91] propose machine learning to predict channel state information (CSI). To the best of the authors knowledge at time of writing, there are no attempts at the use of machine learning to equalise and compensate for the vehicular channel.

Therefore in this chapter, the main focus is on the application of artificial neural networks in the context of vehicular channel equalisation and symbol detection. Two physical layer waveforms of interest will be focussed on, namely OFDM and SEFDM.

OFDM has long been the physical layer waveform of choice in the 802.11p vehicular communication standard and is an algorithm commonly employed in many wireless standards due to its robustness in the face of frequency selective fading and fairly simple implementation via the fast Fourier transform. There is an abundance of literature available on OFDM in the decades since its original conception.

SEFDM on the other hand is a more recent creation first described in [54]. SEFDM utilises a modified, discrete fractional Fourier transform in order to compress a signal in the frequency domain beyond the $1/T$ orthogonality limit, where T is the symbol period. This compression in the frequency domain beyond the orthogonality limit causes self-induced deterministic inter-carrier interference as one would expect, but it comes at the benefit of higher spectral efficiency as more subcarriers can be transmitted in a given bandwidth. Following some of its early success, there has been research undertaken that has seen reasonable bit-error rates using 4-level quadrature amplitude modulation (4-QAM) [57] using only 70% of the bandwidth of OFDM for a given data rate. The sometimes crippling effects of the induced ICI means that complex receivers are often required to recover the transmitted symbols, thus far sphere decoders (SDs) are the most popular choice in the literature, and are being used to take advantage of the newly improved spectral efficiency on offer from SEFDM.

When implemented practically, both OFDM and SEFDM have a requirement for available channel state information, used in tandem with known pilot symbols inserted into each transmitted frame to be used in estimating the instantaneous quality of the wireless channel [92]. The resulting channel estimate can then be used to 'undo' the effect of the channel on the wireless transmission before attempting to demodulate the signal. In this chapter, it will be shown that one can move away from the standard SD receiver architecture and how the author successfully designed and implemented a bespoke neural network algorithm, to be used as a joint wireless channel equaliser

and symbol detector. It has been shown previously that artificial neural networks have the ability to outperform some methods of classic receivers such as decision feedback equalisers and other transversal filters [93].

It is for this reason that the author proposes to marry the advantages of SEFDM, namely its lower bandwidth usage i.e. higher spectral efficiency with a neural network receiver to increase the performance of a vehicular network. The higher spectral efficiency can be advantageously used to support more users within the same network or to increase the networks data rates, whilst the neural network receiver should provide robustness to the transmissions.

The proposed neural network receiver model does not require channel state information, as a supervised learning approach to training has been taken. In this approach, the neural network is trained on data that has been transmitted over a realistic vehicular wireless channel so that the relationships and associations between transmitted and received symbols can be done offline and embedded within the knowledge of the neural network. The implementation is tested via hardware co-simulation, using National Instruments universal software radio peripheral re-configurable input/outputs (NI USRP-RIOs), and a Spirent VR5 channel emulator to generate realistic vehicular channels defined in [94]. It is shown that the proposed neural network receiver achieved an RMS error vector magnitude of 4.75 in the OFDM case over 52 data symbols whilst still conforming to the 802.11p standard, and SEFDM under the same conditions achieves a performance of $10.8 \rightarrow 25$ (RMS) error vector magnitude (EVM) as the compression factor decreases from $\alpha = 0.8 \rightarrow \alpha = 0.4$, a significant improvement. It is then demonstrated that the proposed system is highly portable by changing the vehicular environment that the neural network is deployed in without performing any additional training. The results show no significant degradation of performance, thus the proposed system demonstrates a strong generalisability that reinforces its suitability to practical deployment.

3.2 Practical Implementation

3.2.1 Experimental Test Setup

The experimental test set-up that the author used in this work is illustrated in Fig. 3.1, highlighting the end-to-end system implemented and a block diagram is shown in Fig. 3.2. The author opted for a 'hardware-in-the-loop' approach in the generation and transmission of both the OFDM and SEFDM signals under test. The OFDM

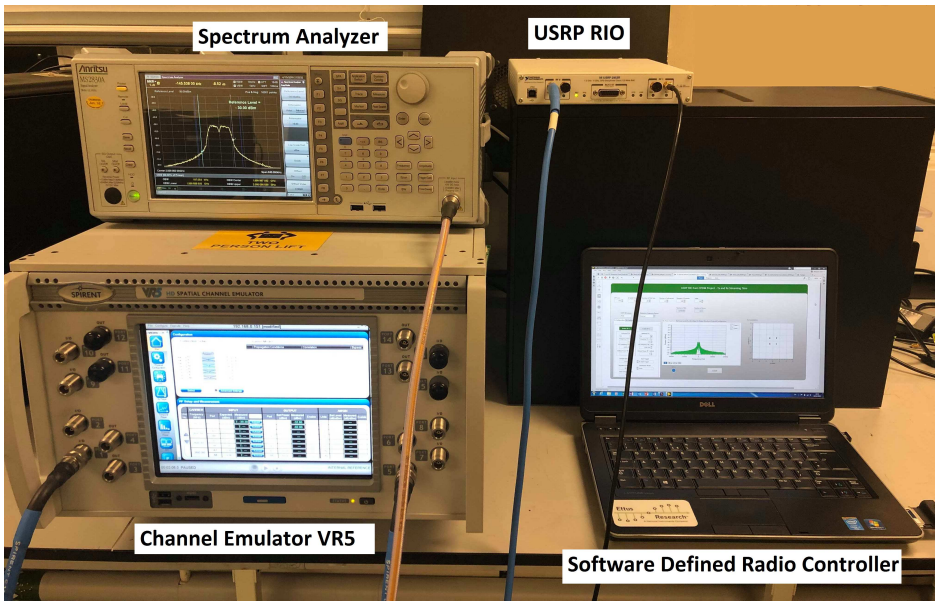

Fig. 3.1 Image capture of the test setup

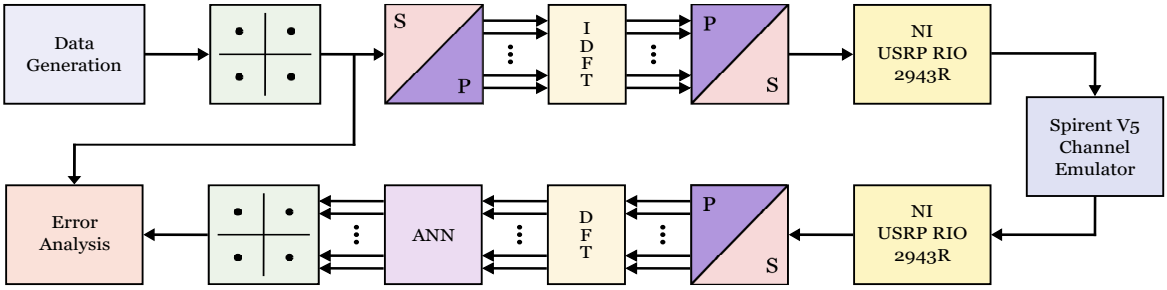

Fig. 3.2 Block diagram of the test setup

signals were generated as described in (2.29), and the SEFDM signals were generated as described in (2.55). For the transmitter, an NI USRP-RIO 2953R was deployed and controlled using LabVIEW software. For the testing, $2^{16}-1$ bits were generated for each physical frame, and these are mapped to a quadrature phase shift keying (QPSK) constellation. In preparation for modulation, these frames are passed through a serial to parallel converter. The signals generated follow the 802.11p standard, i.e. the IDFT size is 64, the number of active subcarriers is $N = 52$ and the signal bandwidth was 10 MHz. The particular specifics are documented further in Table 3.1. Also worth noting is the fact the author test values of α from 1 (OFDM) down to 0.4 (60% bandwidth compression) in steps of 0.2.

Table 3.1 Transmission Parameters

Parameter	Values
Carrier Frequency	5.9 GHz
Sampling Frequency	20 MSs
Signal Bandwidth	10 MHz
Values of α	1 (OFDM), $0.8 : 0.2 : 0.4$
FFT/IFFT Size	64 samples
Cyclic Prefix Size	16 samples
Modulation Scheme	QPSK

3.2.2 Vehicular Channel Model

This work uses two vehicular channel models that were empirically collected in [94], namely the highway NLOS channel model and the urban crossing NLOS model. These models were chosen for this work to test the proposed system in a worst case scenario. The neural network will be built and evaluated on the highway NLOS channel, but then further investigation will look at the effect of channel mismatch on a deployed neural network, where it is used as a symbol receiver in an environment it has never seen during the training process. To extend this and add some further degredation not easily captured in simulation, such as physical wiring and processing imperfections, a Spirent VR5 channel emulator was employed to generate the test signals at the desired 5.9 GHz.

The highway NLOS vehicular channel was chosen as a worst case as its characteristics outline it as the most challenging of the five scenario channels presented by the authors. It has the highest received delay spread, the highest delayed replica receive powers, and the highest spread of non-uniform Doppler shifts. The urban crossing NLOS channel was chosen as the 'unseen' channel for the neural network as it is the second most challenging environment, thus testing the generalisability of the neural network. The main parameters that outline the characteristics of these channels can be found in Table 3.2. All the taps in the delay line have a channel bandwidth of 10 MHz and are defined as Rayleigh distributed, with a mean received power η^2, a receive delay τ, a Doppler frequency f_d, and a given power spectral density (PSD) that the authors had uncovered empirically.

The interesting thing to note about the Doppler profile is that these channel models do not conform to the standard 'bathtub' shape, instead they form a directional

Table 3.2 Channel Parameters

Scenario	Tap	η_i^2 [dB]	τ_i [ns]	$f_{i,d}$ [Hz]	Profile
Highway NLOS	i=1	0	0	0	Static
	i=2	-2	200	689	Half BT
	i=3	-5	433	-492	Half BT
	i=4	-7	700	886	Half BT
Urban Crossing NLOS	i=1	0	0	0	Static
	i=2	-3	267	295	Half BT
	i=3	-4	400	-98	Half BT
	i=4	-10	533	591	Half BT

'half-bathtub' (noted as Half BT in Table 3.2) shape which more accurately reflects the real life application of multiple vehicles communicating in a highway scenario as the dominant contributing scatterers are more densely located directly in front or behind each node [95]. An example PSD from the implemented channel can be seen in Fig. 3.3, along with an empirical cumulative distribution function (ECDF) of the tap realisations in Fig. 3.4, validating their Rayleigh distributions.

The signals are generated as outlined previously, then they are transmitted via the traversal through the Spirent VR5 emulator. On the other side to receive the signal is a USRP-RIO 2953R. At this stage in the process the received signal, considering the application of the vehicular channel H with an added contaminent of the AWGN vector Z may be defined as follows, using (2.56) [57]:

$$R = \mathbf{\Phi}^* H X + \mathbf{\Phi}^* Z = \mathbf{\Phi}^* H \mathbf{\Phi} S + \mathbf{\Phi}^* Z \tag{3.1}$$

where R is the demodulated signal vector estimation of X and $(.)^*$ is the Hermitian transpose operation. Aside from the usual considerations outside the scope of this work and thus not listed mathematically here, such as potential power amplifier non-linearity, it is clear that the main impediment to the goal of successfully demodulating the signals can be identified as the channel, H, and the self induced interference, $\mathbf{C}$.

3.2.3 Neural Network Receiver

The estimation and tracking of both the vehicular channel H and the interference matrix $\mathbf{C}$ becomes highly complex when the channel is both non-static and fading as is the case with the chosen channel model. From empirical use the author has gleaned

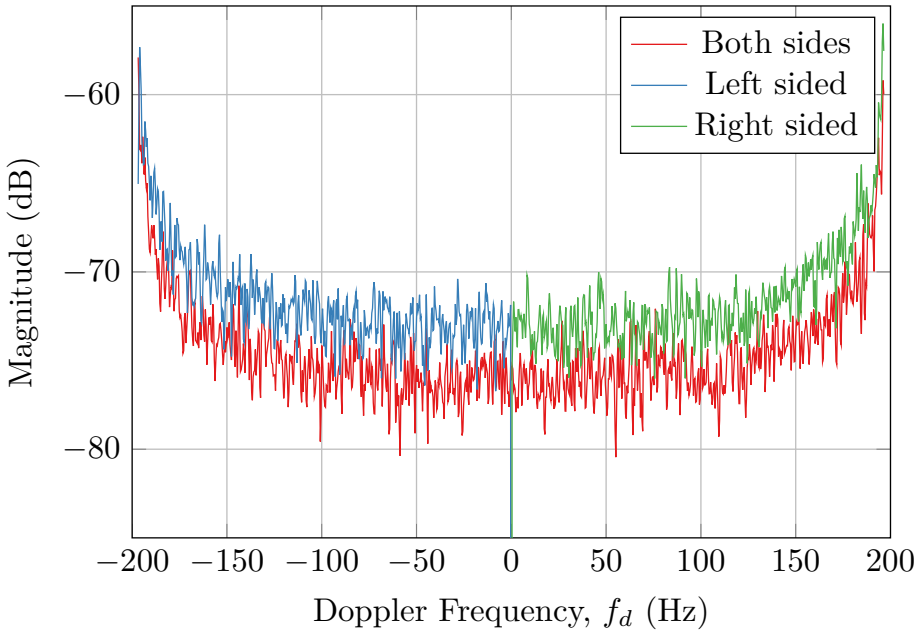

Fig. 3.3 PSD of an original Jakes trace, with the analytic left and right side PSD examples

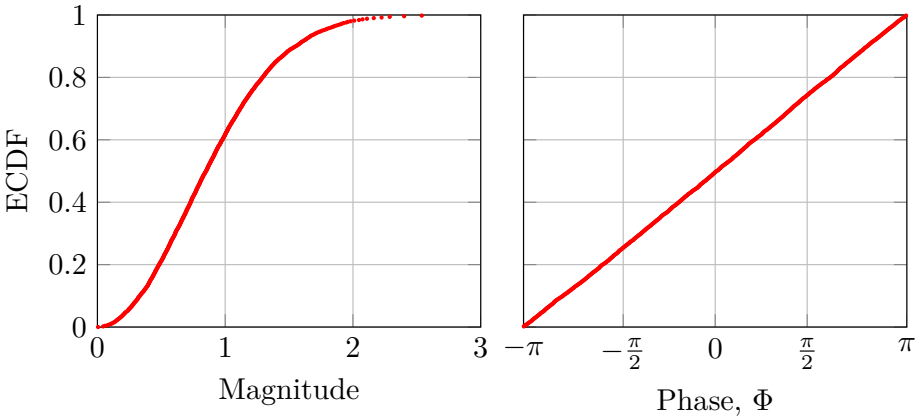

Fig. 3.4 ECDF of the channels magnitude and phase

that conventional information theory based equalisers, for example feedforward or decision feedback equalisers, are inherently limited in their ability to counteract and predict this extremely complex relationship.

For this reason, the author turned their attention to the potential application of artificial neural networks to undertake this task as neural networks posess several advantageous properties that the author believes make them well suited for the problem at hand. One such property is the ability of a neural network with non-linear activation functions, like those that are used herein and will be introduced later, are able to

generalise to any arbitrary input-output relationship given there are sufficient neurons in the hidden layers [96], a property also sometimes referred to as general or universal approximation theorem [97].

The artificial neural network in this work can be seen in Fig. 2.13, and it was designed with the intended use as described in Fig 3.5, acting as a symbol receiver for QPSK symbols after being transmitted with either OFDM multiplexing or SEFDM multiplexing for any of the given bandwidth compression factors. The main parameters for its recreation can be found in Table 3.3. The 802.11p vehicular standard outlines the communication protocol as having 64 subcarriers in total, however 12 of these are reserved as virtual subcarriers which act as a guard band to reduce the out of band radiation [98]. In a typical vehicular system, the communication protocol also defines that 4 of the remaining 52 subcarriers are to be used as pilot symbols to assist with channel estimation and prediction [99].

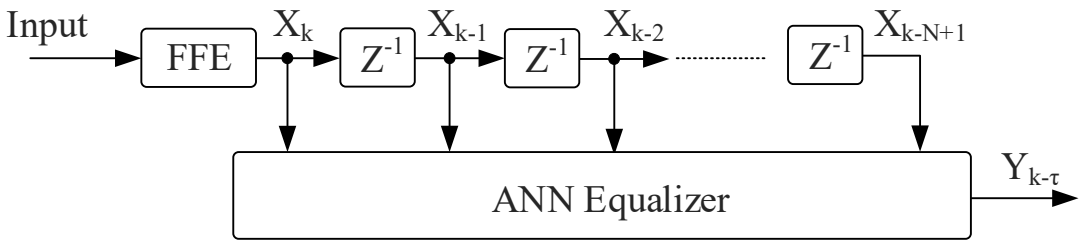

Fig. 3.5 Block diagram of the physical implementation of the neural network equalisation

In the work presented herein, the use of the 4 pilot symbols has been forgone as the variable nature of the channel pertubations have been embedded into the memory of the model during the training stage, thus the neural network implementation can be viewed as being $4/52 \approx 7.7\%$ more spectrally efficient over the standard implementation. In the design of the neural network, the author opts to split the channel observation vector into its real and imaginary constituents. This is an important distinction, we have utilised a 'concatenated-complex' neural network which means that a single neural network is considered as a joint equaliser and symbol receiver. To understand why this is a noteworthy distinction, it is first helpful to know that whilst there is ongoing research in the area of neural networks that fundamentally use complex numbers e.g. [100], this body of work is still in its infancy. In order to overcome this limitation, one has several options.

First, one could use symbol constellations that only use in-phase components on the purely real Cartesian plane, but this would severely limit the number of bits per symbol and hence data throughput. Second, and a common occurence in machine learning

for communications, is to take what the author calls a 'split-complex' approach. The principle of this method is illustrated in Fig 3.6, the real and complex constituents of a signal are separated and a neural network is created for both streams independently - NN_R and NN_I in the figure, respectively. After the neural networks these signals are then recombined, multiplying the output of the 'imaginary' neural network by i and adding it to the output of the 'real' neural network. As one can reason from looking at this approach, during the learning process the real and imaginary components have been treat and optimised independently from one another. It is known in communications that perturbations and degredations to the signal whilst in transmit are not isolated in affecting the purely in-phase or quadrature component at once, but rather any possible combination between the two. This phase relationship is broken and discarded when using the 'split-complex' approach, which leads us to our third option - the aforementioned 'concatenated-complex' neural network. This design of neural network essentially merges the two neural networks of the 'split-complex' approach. Incoming signals are still split into their real and complex components, however now they are both fed into the same neural network as purely real signals. These key differences are visually captured in Figures 3.6, 3.7, and 3.8.

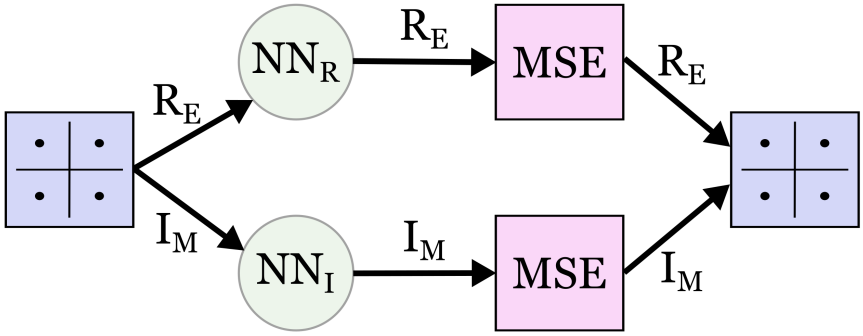

Fig. 3.6 The 'split-complex' neural network approach using MSE as the objective loss function

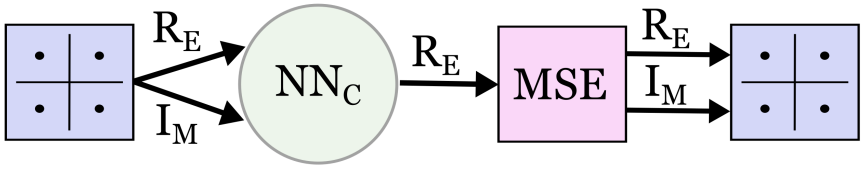

Fig. 3.7 The 'concatenated-complex' neural network approach using MSE as the objective loss function

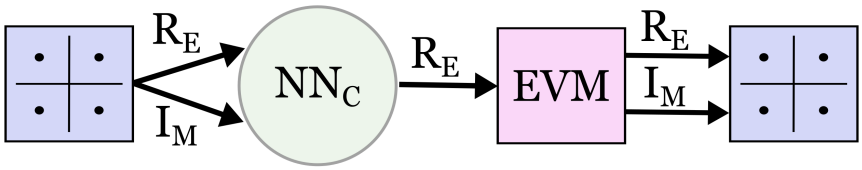

Fig. 3.8 The 'concatenated-complex' neural network approach using EVM as the objective loss function

Due to the design of our neural network, i.e. using full connected layers, meaning each neuron in one layer is connected to every neuron in the immediately proceeding layer, means that the information from the signals that belonged to the 'complex' parts of the original signal are still available to their 'real' counterparts. In addition, as these neurons are fully connected this information is also available to neighbouring neurons. This ultimately means that the neural network, to account for the 52 data subcarriers, has an input size of $2N = 104$ neurons. Beyond this, the 3 hidden layers of the neural network are organised into groups of $4N$, $8N$ and $4N$ neurons respectively. The reason 3 hidden layers were chosen with this neuron configuration were chosen is multifold. Firstly, there is precedent for empirical reasons from the literature [17]. Secondly, the choice for the number of layers and neuron configuration in a neural network has ramifications for the learning process.

Table 3.3 Model Summary

Layer	Type	Output Shape	Notes
Input Layer	Input	104	-
Dense	Dense	208	-
Leaky ReLU	Activation	208	$\beta = 0.3$
Dense	Dense	416	-
Leaky ReLU	Activation	416	$\beta = 0.3$
Dense	Dense	208	-
Leaky ReLU	Activation	208	$\beta = 0.3$
Dense	Dense	104	No Bias
Output Later	Output	104	-

The neural network has to contain enough processing power to capture the complexities of the task, i.e. in this instance both the channel pertubations and the self-induced interference in the case of SEFDM, but one cannot just keep increasing the complexity of the neural network indefinitely and expect better and better performance. As is discussed in [101], there is an optimal neural network size for a problem beyond which

there is a risk of overfitting, that is to say the model will fail to generalise beyond what it has observed in the training data.

Each of the layers in the network used herein are what are referred to as 'dense' layers, meaning that each neuron is directly connected to all neurons from the preceding layer. Mathematically this can be described this as [66]:

$$N_{\text{out}} = \psi\left(\mathbf{N}_{\text{in}} w\right) + b \tag{3.2}$$

where ψ is the activation function of the neuron, N_{in} is the input to the neuron, w are the neurons weights i.e. it's internal kernel, and b is the neurons bias vector. The activation function used in this neural network was the LeakyReLU, which is a divergence from the typical implementations seen in the literature which typically employ the standard ReLU activation function. ReLU is a well founded choice for an activation function and has been popular ever since its introduction in [102], mathematically it is defined as:

$$\psi\left(x\right) = \begin{cases} 0, & x < 0 \\ x, & \text{otherwise} \end{cases} \tag{3.3}$$

It has long been recognised as a useful activation function for learning complex relationships in data as the zero-clamping effect makes the model sparser, therefore only the most important features end up propagating throughout the network during the learning process. As mentioned previously however, the author opted instead in this work to use the LeakyReLU activation function. The author did this for one main reason, ReLU is known to behave poorly when the input data has not been normalised [103]. This can lead to neurons entering a state where its output is perpetually stuck at zero, and becomes a 'dead' neuron which in turns diminishes the computational power of the neural network. Whilst it is true that the transmit data has been normalised, it is important not to forget that whilst in transit the vehicular channel has multiplicative and additive effects. In this case, the signal can no longer be treat as 'truly' normalised in the typical sense as it cannot be mathematically guaranteed that the upper and lower bounds on the received power do not exceed ± 1. LeakyReLU made a minor but significant change to the formulation of ReLU by allowing linear but non-zero outputs for $x < 0$, which the original authors in [104] state can help avoid the death of neurons within the neural network. LeakyReLU can be defined mathematically as:

$$\psi(x) = \begin{cases} \beta x, & x < 0 \\ x, & \text{otherwise} \end{cases} \tag{3.4}$$

Combining both the knowledge of the structure of the neural network and the chosen loss function, it is possible to perform a baseline measure of computational complexity. The number of multiplications and additions required to do a forward pass through the neural network depend on the number of neurons in the preceeding layer and the number of neurons in the current layer. For instance, the first hidden layer of the proposed neural network has $4N = 208$ neurons, and the input layer has $2N = 104$ neurons. Each of the $4N$ neurons in the first hidden layer have to perform one multiplication and one addition for each input from the $2N$ input neurons according to (3.2). The bias is an addition and is applied once per neuron, and to account for the activation function we will take the worst case scenario for LeakyReLU where the output requires one further multiplication to reach the neurons final output value. Following this, we can derive the computational complexity of the proposed neural network in a worst case scenario requiring $(2N \times 4N) + (4N \times 8N) + (8N \times 4N) + (4N \times 2N) = 80N$ multiplications and additions internally, then a further $4N + 8N + 4N = 16N$ bias additions (the output layer contains no biasing), and $4N + 8N + 4N + 2N = 18N$ multiplications for the activation function. This results in a total of $98N$ multiplications and $96N$ additions for the proposed neural network. For completeness, the N used in this work was 52, thus a total of 5096 multiplications and 4992 additions are required end to end.

Once the data has completed its traversal through the layers of the neural network, the first 52 outputs are considered as the real constituent of the output symbol and the following 52 outputs are considered as the imaginary constituent of the output symbol. There is no particular reasoning as to the ordering of the output neurons, they were arranged this way as to mirror the ordering of the input neurons, which again followed no particular reasoning besides conceptually representing the order of the subcarriers. The reason one does not have to worry about the mapping between subcarriers and neurons in this particular instance is due to the use of the dense layers throughout the entirety of the neural network since every neuron is connected to every other neuron from the preceding layer. For the proposed neural network that has 104 neurons in the output layer, we take the first output neuron to represent the real component of the first subcarrier, and the $1 + 52 = 53^{\text{rd}}$ neuron to represent the imaginary component of the first subcarrier. By considering the subcarrier to neuron mapping in this 'split-complex form', the outputs of the neural network are

used to form an estimate of the transmitted complex QPSK symbol, this can be seen in Fig. 3.9.

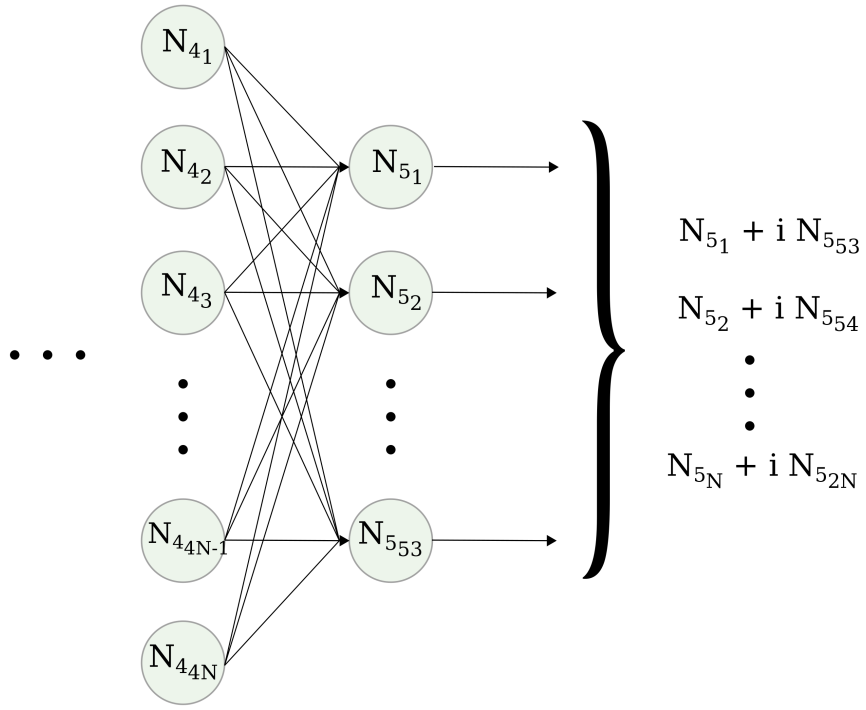

Fig. 3.9 Block diagram of the QPSK symbol estimation from the neural network

During the training process, neural networks learn through a process called gradient descent and backpropagation which was explained previously in Chapter 2. Just using gradient descent however is often far from optimal, whilst one can calculate a loss and propagate its updates back through the neural network, this often leads to highly erratic updates that move in the right direction. However, there is high variance from epoch-to-epoch so it takes a long time to converge. It is at this point that one introduces gradient descent optimisers. Optimisers take the loss values and instead of applying the weight updates to the neurons in the neural network directly from it, perform additional steps.

The additional steps are what distinguish optimisers from each other, however they all share the common goal of updating the weights of the neurons in the neural network to eventually reach the global minima of the loss function with respect to the target. To explain how the author decided on their choice of optimiser, one must first take a step back to the standard stochastic gradient descent (SGD) algorithm.

As mentioned previously, looking at the loss after each training sample to apply the weights can lead to high variance in the backpropagation algorithm updates. It is

known that SGD has performance issues when the surface of the loss landscape curves much more steeply in one dimension than in others e.g. around a local minima, which causes the behavior of the system to be equivalent to a set of coupled and damped harmonic oscillators [105]. In these scenarios, SGD oscillates in this 'ravine' region and converges slowly toward the minima. To alleviate some of these concerns, the concept of momentum was introduced and can be used to amend the gradient descent algorithm to include an extra term to help dampen the oscillations of SGD. The momentum term improves the speed of convergence by bringing some eigen components of the system closer to critical damping and then one can rework (2.75) to include this term as [105]:

$$v_t = \gamma v_{t-1} + \eta \nabla_\phi J_i(\phi_k) \tag{3.5}$$

where, v_t is the resultant weight update vector at time t, ϕ are the parameters i.e. weights and biases of the neural network, $J(\phi)$ is the loss function given the parameters of the neural network, ∇_ϕ is the derivative with respect to the neural network parameters, and η is the current learning rate for the neural network in the training process. The new momentum term, γv_{t-1}, is a function of some numeric factor γ and the weight update vector from the previous time step, v_{t-1}. The γ term here controls the amount of momentum to be used by the optimiser. This momentum term increases for dimensions where the gradient points in the same direction as previously, and reduces the size of weight updates for dimensions where the gradient has changed direction. This ultimately leads to faster convergence and reduced oscillation during the training process as can be seen in Fig. 3.10.

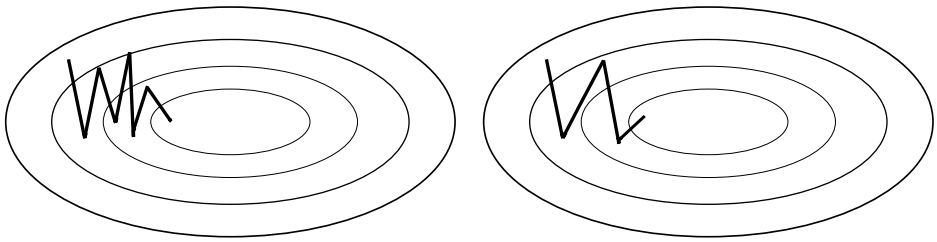

Fig. 3.10 Representation of weight update vectors during the training process without momentum (left) and with momentum (right)

Besides momentum, other alterations to standard gradient descent have been proposed, many of which focus on adaptive learning rates. One such example is Adadelta [106], which adapts learning rates based on a moving window of gradient updates. To demonstrate this with an example, one can denote the gradient at time step t of the loss function with respect to the neural network parameter ϕ_i as [106]:

$$g_{t,i} = \nabla_{\phi} J\left(\phi_{t,i}\right) \tag{3.6}$$

which in turn lets us express the SGD update for each parameter ϕ_i at a time step t as [106]:

$$\phi_{t+1,i} = \phi_{t,i} - \eta g_{t,i} \tag{3.7}$$

where as before η is the learning rate. With adadelta, the sum of gradients is recursively defined as a decaying average of all past squared gradients [106]:

$$E\left[g^2\right]_t = \mu E\left[g^2\right]_{t-1} + (1-\mu)\, g_t^2 \tag{3.8}$$

The running average of $E\left[g^2\right]_t$ at a time t now depends on the previous average and the current gradient with a numeric factor μ which acts similarly to the γ factor in momentum whereby it controls the extent to which the previous previous average has an influence in the current update. This leads us to the basis of the optimiser that was used in this work. The author used the nadam optimiser which was first proposed in [107], but first one needs to be familiar with the optimiser which nadam itself is based on, the adam optimiser (adaptive moment estimation) from [108]. Much like the adadelta optimiser that was just discussed, the adam optimiser computes adaptive learning rates for each parameter of the neural network.

In addition to storing the exponentially decaying average of past squared gradients as was the case for adadelta, adam also retains the exponentially decaying average of the past gradients i.e. the aforementioned momentum addition, causing it to prefer flat minima in the error landscape [109]. Mathematically, one can describe the decaying averages of the past gradients, and the decaying average of the square of the past gradients as [108]:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\, g_t \tag{3.9}$$

and

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\, g_t^2 \tag{3.10}$$

where m_t and v_t denote the average, and square of the average past gradients, respectively. One would note that these can be seen as the estimate of the first and second moment of the gradients respectively, hence why adam received its name. There is one final step in the adam algorithm however, as the original authors note that since m_t

and v_t are initialised as a zero vector they remain biased towards zero, especially with smaller decay rates i.e. when β_1 and β_2 are close to 1. To prevent this zero bias, the author suggests calculating the bias-corrected first and second moment estimates [108]:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{3.11}$$

and

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{3.12}$$

The final leap from this adam optimiser to the nadam optimiser used in the authors work requires a change to how one considers and formulates the momentum factor for weight updates. In the original paper presented in [110], the author proposes altering the momentum term to give it prescience. The method is called Nesterov accelerated gradient (NAG) and uses a 'look ahead' method to more intelligently apply momentum updates. From (3.5) it is apparent that there exists a momentum term γv_{t-1} and that would normally be applied to the neural network parameters, ϕ. Instead of directly applying this momentum term, in NAG one calculates an approximation of the next iteration of the parameters as [110]:

$$\hat{\phi}_{t+1} = \phi_t - \gamma v_{t-1} \tag{3.13}$$

This is only an approximation however, because as one can see from inspecting (3.5), the gradient changes at this point are unknown with respect to the full update for all parameters. It is also worth noting that for a small learning rate η, the approximation is relatively accurate as parameters are updated in small steps. This ultimately means one can now calculate the gradient with respect to the approximate future positions of the neural network parameters ϕ, instead of the actual current network parameters. It can then be said that [110]:

$$v_t = \gamma v_{t-1} + \eta \nabla_\phi J\left(\phi - \gamma v_{t-1}\right) \tag{3.14}$$

The anticipatory nature of this method reduces the ability of the momentum factor to deviate too far from where the weight update and loss function originally intended to travel, and increases the responsiveness of the updates as the momentum is more easily disrupted in the case of a few erroneous steps [111]. It also has the advantage of stronger theoretical convergence guarantees for convex functions [112].

This gives us all the necessary components required to describe the nadam gradient descent optimiser used in this work. The nadam optimiser, at a basic level, is the combination of adam with Nesterov accelerated gradients. To do this, one must return to (3.5) and modify the momentum term to include NAG [107]:

$$g_t = \nabla_{\phi t} J\left(\phi_t - \gamma m_{t-1}\right) \tag{3.15}$$

and

$$m_t = \gamma m_{t-1} + \eta g_t \tag{3.16}$$

The authors also make one more subtle change, this current implementation requires applying the momentum step twice, once for updating the gradient vector g_t, and once for updating the neural network parameters ϕ_{t+1}. The prescient momentum vector must be directly used to update the current parameters [107]:

$$g_t = \nabla_{\phi t} J\left(\phi_t\right) \tag{3.17}$$

and

$$m_t = \gamma m_{t-1} + \eta g_t \tag{3.18}$$

meaning one can arrive at new parameters for ϕ_{t+1} as [107]:

$$\phi_{t+1} = \phi_t - \left(\gamma m_t + \eta g_t\right) \tag{3.19}$$

Note from (3.19) that the author now uses the current momentum vector m_t to 'look-ahead' rather than the previous momentum vector m_{t-1}. Now that the optimiser is defined, the author will set up the rest of the training process. As was found to be empirically best performing in [113], the author set an initial learning rate for the neural network of 10^{-3}. At the end of each training epoch the author examined the validation loss, i.e. the neural networks current prediction capability on a set of examples not included in the training data, as a gauge of whether the learning process had stalled or plateaud. In the case of a plateau, it may be that the learning rate is too high causing oscillations around a minima that it can't escape from, thus after monitoring for several epochs, if a plateau was detected the learning rate was reduced by a factor of 10^{-1}. As is often common in machine learning in communications [114], in this work the author opted for the mean square error criterion to act as the objective loss function [115]:

$$L = \frac{1}{2N} \sum_{i=1}^{2N} (y_i - \hat{y}_i)^2 \tag{3.20}$$

where N represents the number of subcarriers, which in this case is 52, and the factor of $2N$ as mentioned prevously allows for the separation of the real and imaginary components of the signal to be input to a singular neural network.

3.2.4 Practical Results

The results hereafter outline the outcome of the authors practical implementation of both OFDM and SEFDM transmission with increasing levels of bandwidth compression factor, α. They are tested in a realistic vehicular communication environment, namely the fast paced, dynamic, urban NLOS scenario that was previously introduced, in a setting that conforms with the 802.11p vehicular communication standard. The outcome of this work aids in the understanding and comparison of the effects that would need to be understood should the proposal arrive for a more spectrally efficient waveform to be considered as being included in the official 802.11p vehicular communication standard.

The first result to inspect is that of the bit-error rate in Fig. 3.11. The bit error rate is a standard measurement of performance in communications, and is defined as the number of received bits of a data stream that have been received incorrectly, expressed as a rate from the total number of bits in the data stream. Bit-error rate results often require vast numbers of transmitted bits to accurately empirically converge to the theoretical limit, thus the author notes that while statistically significant amounts of data were indeed transmitted and the resulting bit-error rate plot looks to have converged, for feasibility and time reasons inherent in practical demonstration the lesser compressed SEFDM and OFDM cases did not produce any errors beyond 10^{-5}. In order to provide as fair a comparison as possible between all the experiments regardless of convergence restraints, and to inform the reader with the full data, the author has also measured the EVM of the received constellations in all experiments with reference to the originally transmitted symbol constellations. Mathematically, one can calculate the EVM as [116]:

$$EVM = \sqrt{\frac{\frac{1}{N}\sum_{k=1}^{N}\left(I_k - \hat{I}_k\right)^2 + \left(Q_k - \hat{Q}_k\right)^2}{\frac{1}{N}\sum_{k=1}^{N}\left(I_k^2 + Q_k^2\right)}} \tag{3.21}$$

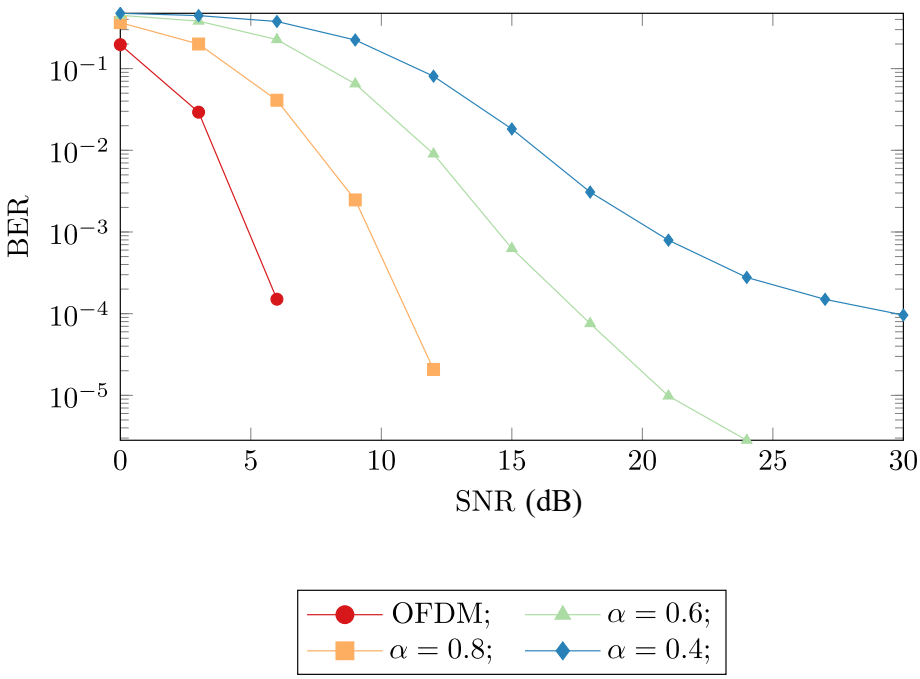

Fig. 3.11 Bit-error rate measurements for all values of α under test

where I_k and Q_k are the k^{th} original transmitted in-phase and quadrature components respectively, $\hat{I}_k$ and $\hat{Q}_k$ are the k^{th} observed in-phase and quadrature components respectively, and N is the number of subcarriers.

The next result demonstrated in Fig. 3.12 illustrates an example EVM plot at a fixed high signal-to-noise ratio (30dB). Note that unlike bit-error rate, which is a summary metric, EVM exists for each individual subcarrier. For a reader new to EVM reporting, and to further aid interpretation, the author has included representative received constellation plots after equalisation via the neural network in Fig. 3.13. As one can see, following from the logic of the bit-error rate and EVM plots, the received constellations are recognisable as to their structure but the inherent self-interference from the increasing levels of bandwidth compression i.e. as α decreases is increasingly apparent.

It is impractical to produce these plots for every value of SNR however, therefore to give a complete picture of the comparisons of performance for this system the author takes the mean value of the EVM per SNR in order to obtain a single 'summary' EVM for that SNR (with a 95% confidence interval of $< 0.1\%$). By doing so, one can observe the trend that is displayed in Fig. 3.14. If one begins by observing OFDM, it can be noted that the value of EVM sharply decreases until around 12dB, at which point it begins to plateau. The neural network was quickly able to equalise the channel until

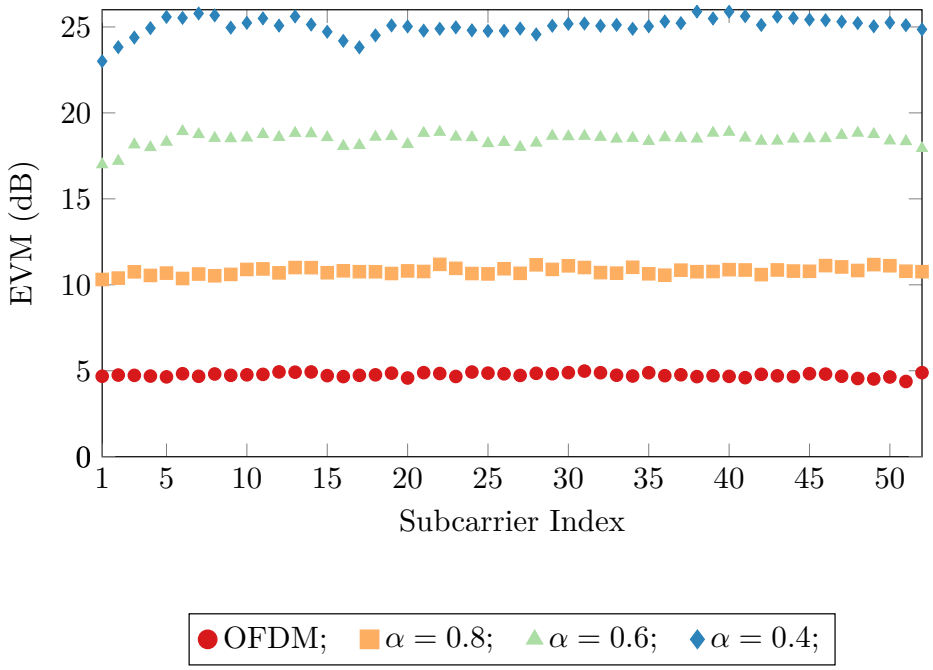

Fig. 3.12 Error vector magnitude for varying values of α at high SNR

the non-deterministic AWGN noise became the limiting factor. If one was to now observe the SEFDM trends, a very similar pattern to OFDM can be observed, which is to logically be expected. In all cases, the neural network is not only able to again equalise the channel, but it has also incorporated and tackled the self-interference caused by the SEFDM process. The limiting factor this time is a combination of the non-deterministic AWGN noise and the ICI. The author notes that with increasing levels of bandwidth compression factor α, there is a performance penalty courtesy of the larger interference contribution from each subcarriers neighbours.

To further investigate the practial feasibility of deploying a neural network in a vehicular environment, the communication channel in the experiment was replaced with the urban crossing NLOS vehicular channel. This is a channel that the neural network has never seen before, as all of the training samples came from the highway NLOS channel. It is discussed previously that generalisability is a desirable feature in neural networks, as it means that the training process has accurately captured true underlying relationships rather than overfit to the dataset. This is an often overlooked aspect of neural network applications in communication, which prohibits practical implementation in real world applications. The vehicle is not restricted to driving in a highway NLOS environment, therefore the neural network has to work for all vehicular environments to be a realistic proposal for practical uptake. Shown in Fig 3.15

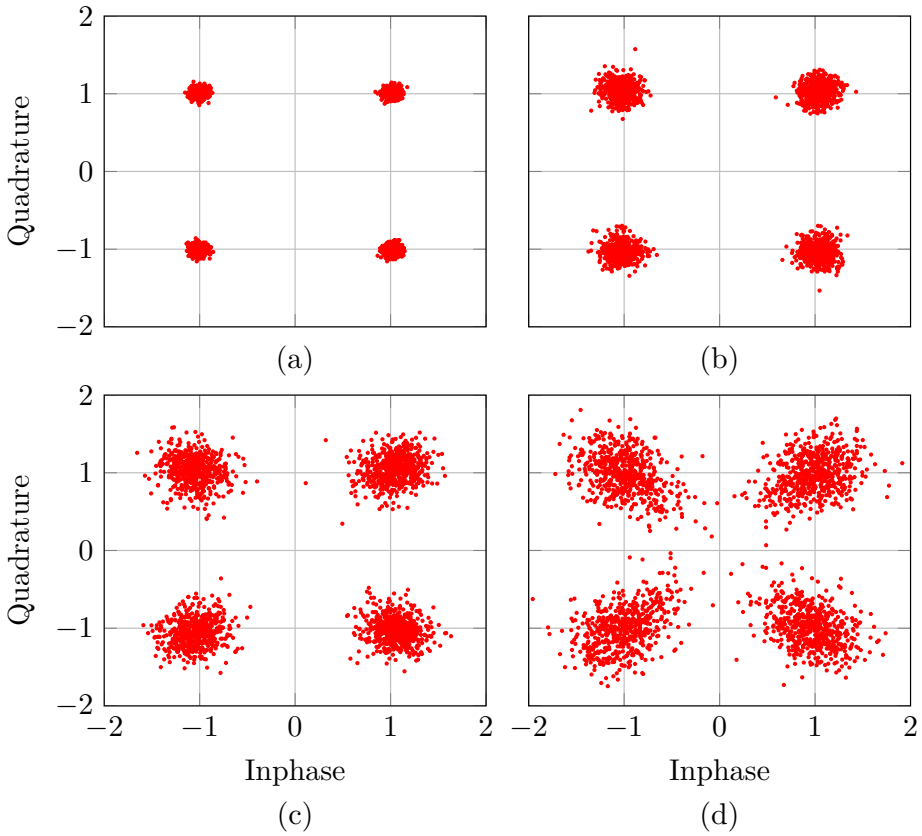

Fig. 3.13 Received constellations for the neural network trained on Highway NLOS data deployed inside a matching Highway NLOS environment for (a) OFDM, (b) SEFDM with $\alpha = 0.8$, (c) SEFDM with $\alpha = 0.6$ and (d) SEFDM with $\alpha = 0.4$

are the received constellations for the neural network trained on the highway NLOS environment but deployed in the urban crossing NLOS environment. One can observe that they are very similar to the original constellations from the correct channel in Fig. 3.13. Following the same process as before, the mean error vector magnitudes over different SNR ranges in the experiment are plotted which are shown in Fig 3.16. One can clearly see a very strong correlation to the results obtained previously for the correct channel, which indicates that the neural network has a strong generalisability to unseen vehicular environments.

3.3 Conclusion

It has been shown that a properly designed and implemented neural network has adequate capability of acting as a joint channel equaliser and symbol detector in a vehicular network to adapt to and overcome the inherent challenges of the rapidly

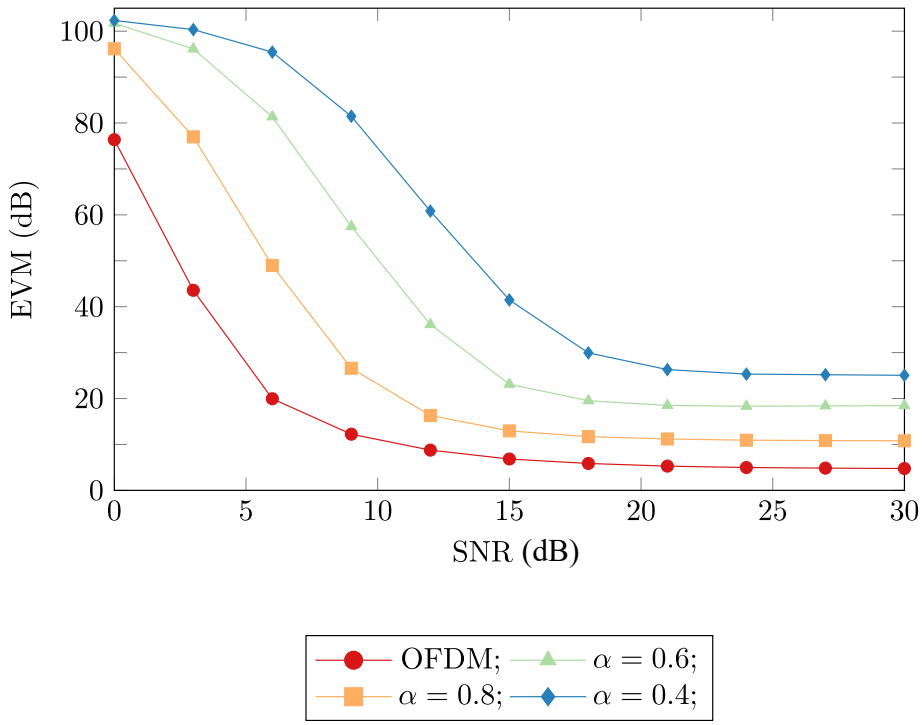

Fig. 3.14 Mean error vector magnitudes for the neural network trained on Highway NLOS data deployed inside a matching Highway NLOS environment for varying values of α at different SNRs

changing communication environment. Practical experiments were undertaken in which the author used a hardware-in-the-loop approach to demonstrate the efficacy of the proposed method. The experiment began by utilising a system that would conform to the communication specification set out by the 802.11p protocol using OFDM. This baseline implementation was then to demonstrate results with a more spectrally efficient waveform, namely, SEFDM. SEFDM was tested with various bandwidth compression factors ranging from 20% up to 60% bandwidth compression. The author would note however that they do not recommend such high bandwidth compressions in practice, but have done so to show that there is a trade-off to be decided upon between bit-error rate performance and spectral efficiency. If one should deem that such a performance trade-off be acceptable for their application, the benefits of the spectral savings could be used to improve data rates or service more users. It was demonstrated that an acceptable BER performance can be achieved even under the harsh conditions of the highway NLOS vehicular channel. Further validation was obtained via measurements of the error vector magnitude to ease the practical constraints of the implementation. The resulting neural network was then tested in a vehicular environment it had never seen before, namely the urban crossing NLOS environment where it was demonstrated

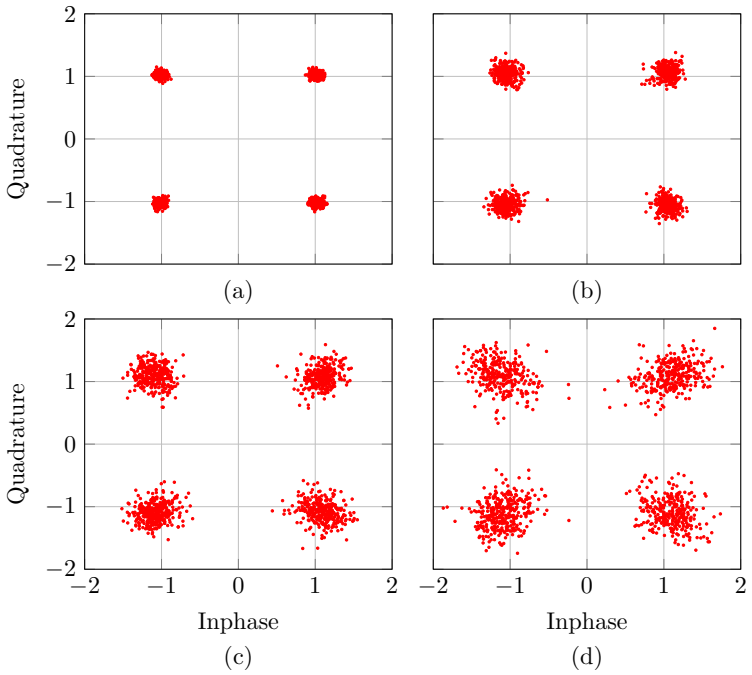

Fig. 3.15 Received constellations using a neural network trained on Highway NLOS data deployed inside the Urban Crossing NLOS environment for (a) SEFDM with $\alpha = 0.8$, (b) SEFDM with $\alpha = 0.6$ and (c) SEFDM with $\alpha = 0.4$

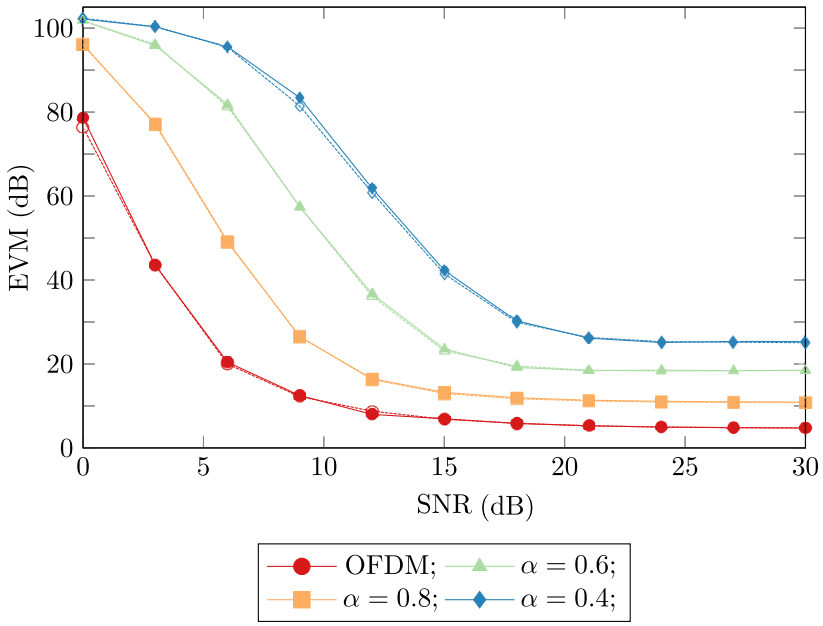

Fig. 3.16 Mean error vector magnitudes using a neural network trained on Highway NLOS data deployed inside the Urban Crossing NLOS environment for varying values of α at different SNRs. The dashed lines with hollow icons are the previous results obtained in Fig. 3.14, showing the minimal reduction in performance.

that the proposed system had the ability to transcend vehicular environments with no significant detriment to performance. The work presented herein has validated the feasibility and efficacy of the application of neural networks in harsh vehicular communications.

Chapter 4

Error Vector Magnitude as a Loss Function

4.1 Introduction

The use of machine learning and neural networks for communication systems is a topic of ever increasing interest. One of the key aspects missing from the current state-of-the-art are domain-related methods of training and evaluating these systems within the communication system context [117]. Communication systems are typically evaluated via their bit-error rate after forward error correction (FEC), however evaluation of bit-error as an objective function for an optimisation target is not feasible, for example when training models in machine learning. Due to this, in the literature for machine learning in communications, authors typically use the mean square error (MSE) or the related l_2 norm as a loss function to train their models [17, 118]. This causes a disconnect between the objective function of the model, and the eventual metrics that will be reported when the model is deployed. In other words, the model is trained to minimise a value that is not representative of the real deployed system.

It is clear that a performance gain would be obtained when the objective function of the training process is aligned with the metrics that the communication system will be evaluated against. To this end, in this paper we propose the exploitation of the error vector magnitude in the training process as the objective loss function and show how it can be effectively utilised to improve the performance of symbol receivers. EVM is a common key performance indicator and has been used in many communication systems since it was first proposed in [119], and is closely related to the bit-error rate [120]. In a machine learning scenario, EVM is useful because it gives an error value in relation to a known ideal value, and is available instantly for any batch size used in the

system, unlike the unit-less bit-error rate metric that requires vast numbers of bits to converge. The EVM requires knowledge of both the in-phase and quadrature parts of a signal, therefore to aid this method, we will also be using the 'concatenated-complex' approach. Whilst not the main contribution of this work, there exists mixed literature between this and the 'split-complex' approach [121, 122], i.e. splitting a complex signal into it's in-phase and quadrature components and training a separate, independent model for each, this work will also demonstrate how including both can aid a models learning ability due to indirect retention of the phase information.

The proposed system demonstrates a spectral efficiency improvement of up to 67% at zero cost within the practical experiment, with a 40% reduced bandwidth shown to be capable of communicating error free at the same data rate of the equivalent OFDM system.

4.2 Practical Implementation

4.2.1 Error Vector Magnitude

EVM provides a convenient way of measuring the SNR for quadrature amplitude modulated signals. It is a method that is used in many communications standards such as; Wireless Local Area Network (WLAN) [123], LTE [124], Universal Mobile Telecommunications System (UMTS) [125] and Enhanced Data rates for GSM Evolution (EDGE) [126]. It encompasses the effects of both magnitude and phase distortions [127] and is evaluated by comparing a measured signal with a reference signal, i.e. a received signal and the transmitted signal. This is illustrated in Fig. 4.1. Mathematically, we define EVM as [128]:

$$\begin{aligned} \text{EVM} &= \sqrt{\frac{\frac{1}{N}\sum_{k=1}^{N}\left(I_k - \hat{I}_k\right)^2 + \left(Q_k - \hat{Q}_k\right)^2}{\frac{1}{N}\sum_{k=1}^{N}\left(I_k^2 + Q_k^2\right)}} \\ &\approx \sqrt{\frac{N_0}{E_s}} \\ &= \sqrt{\frac{1}{\text{SNR}}} \end{aligned} \tag{4.1}$$

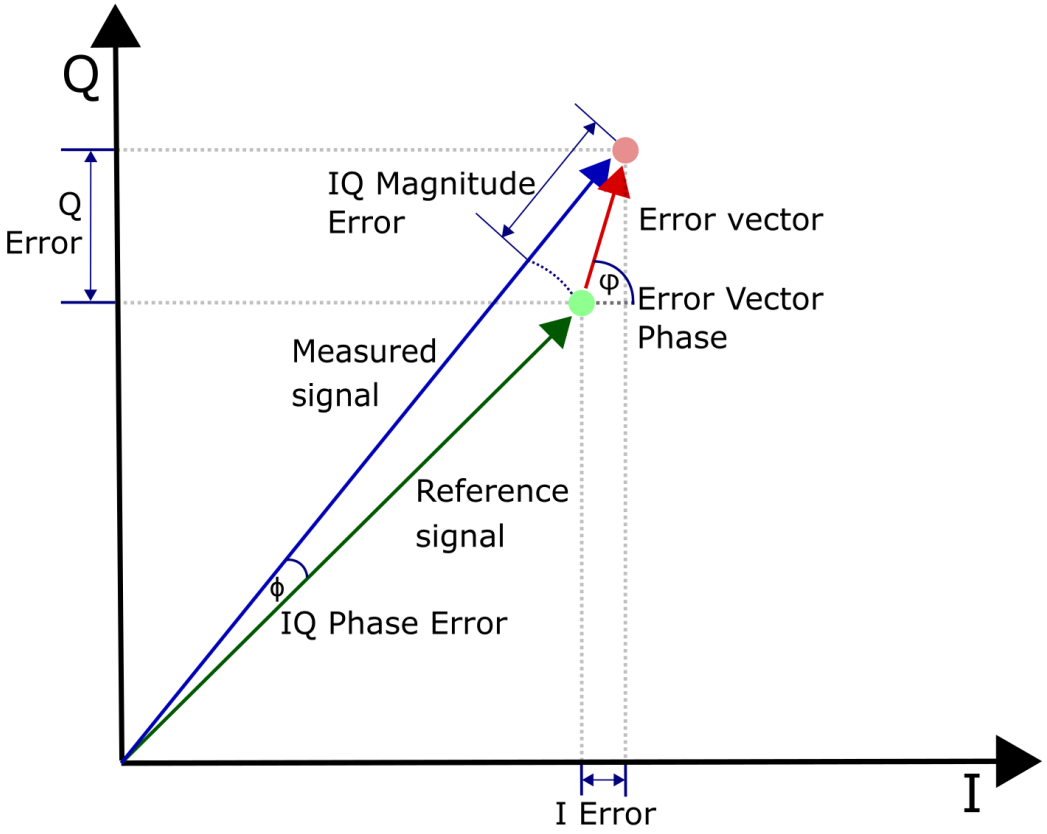

Fig. 4.1 Illustration of EVM

where I_k and Q_k are the real and complex transmitted symbol on the k^{th} subcarrier, respectively, $\hat{I}_k$ and $\hat{Q}_k$ are the real and complex received symbol on the k^{th} subcarrier, respectively, N_0 is the noise spectral density, and E_s is the symbol energy. EVM provides a more rounded picture about the quality of the communication link than both the bit-error rate and the symbol error rate as it is not merely a unitless aggregate statistic, but contains actual magnitude and phase information. The authors in [128] note that the error vector magnitude is available instantly, as opposed to the bit-error rate which is normally calculated via error counting using a Monte-Carlo methodology, and requires time to converge. Therein lies the key differentiator from the application in the machine learning space. Having an error metric that is available, observable and provides reliable feedback in real time means that the iterative process used during gradient descent to train the neural network is fully in tune with the ultimate objective. In other words, the training process of the neural network is more tightly coupled to the real world scenario in which the neural network will be deployed.

The probability of a bit-error occuring is an unescapable phenomena in communications that include at a minimum additive white Gaussian noise channel perturbations. Furthermore, it is known that the bit-error performance for fading channels is always worse and can be directly related to that of the additive white Gaussian noise channel [129]. Considering M-ary modulation in an additive white Gaussian noise channel under otherwise ideal conditions, it can be shown that [130]:

$$P_b = \frac{2\left(1-\frac{1}{L}\right)}{\log_2 L} Q\left[\sqrt{\left[\frac{3\log_2 L}{L^2-1}\right]\frac{2E_b}{N_0}}\right] \tag{4.2}$$

where P_b is the probability of a bit-error, L is the number of levels in the of the M-ary modulation scheme, E_b is the energy of each individual bit, and $N_0/2$ is the noise power spectral density. $Q[\cdot]$ is the Gaussian error function and is given by [131]:

$$Q(x) = \int_x^\alpha \frac{1}{2\pi} \exp^{\frac{-y^2}{2}} dy \tag{4.3}$$

One can also rewrite 4.2 in terms of signal to noise ratio for the symbol, which is often more useful than describing at a bit level [130]:

$$P_b = \frac{2\left(1-\frac{1}{L}\right)}{\log_2 L} Q\left[\sqrt{\left[\frac{3\log_2 L}{L^2-1}\right]\frac{2E_s}{N_0\log_2 M}}\right] \tag{4.4}$$

where E_s is the energy of a symbol and $\log_2 M$ is the number of bits per symbol. This is a simple modification that assumes each bit in the symbol receives equal share of power, an assumption that is generally true for regular constellations unless using a system that explicitly modifies otherwise.

With this well defined, closed form theoretical bit-error rate for M-ary modulation, one can combine this with (4.1) to produce an approximate relationship between the probability of a bit error P_b and the error vector magnitude. This is given as [128]:

$$P_b \approx \frac{2\left(1-\frac{1}{L}\right)}{\log_2 L} Q\left[\sqrt{\left[\frac{3\log_2 L}{L^2-1}\right]\frac{2}{\mathrm{EVM}^2\log_2 M}}\right] \tag{4.5}$$

which when analysed via Monte-Carlo simulation correctly converges to the theoretical bit-error rate curves for M-ary modulation as seen in Fig. 4.2.

4.2.2 Experimental Test Setup

The test setup can be found in Fig. 4.3. The decision for placing the neural network after the FFT owes to the property of neural networks known as the ‘universal approximation

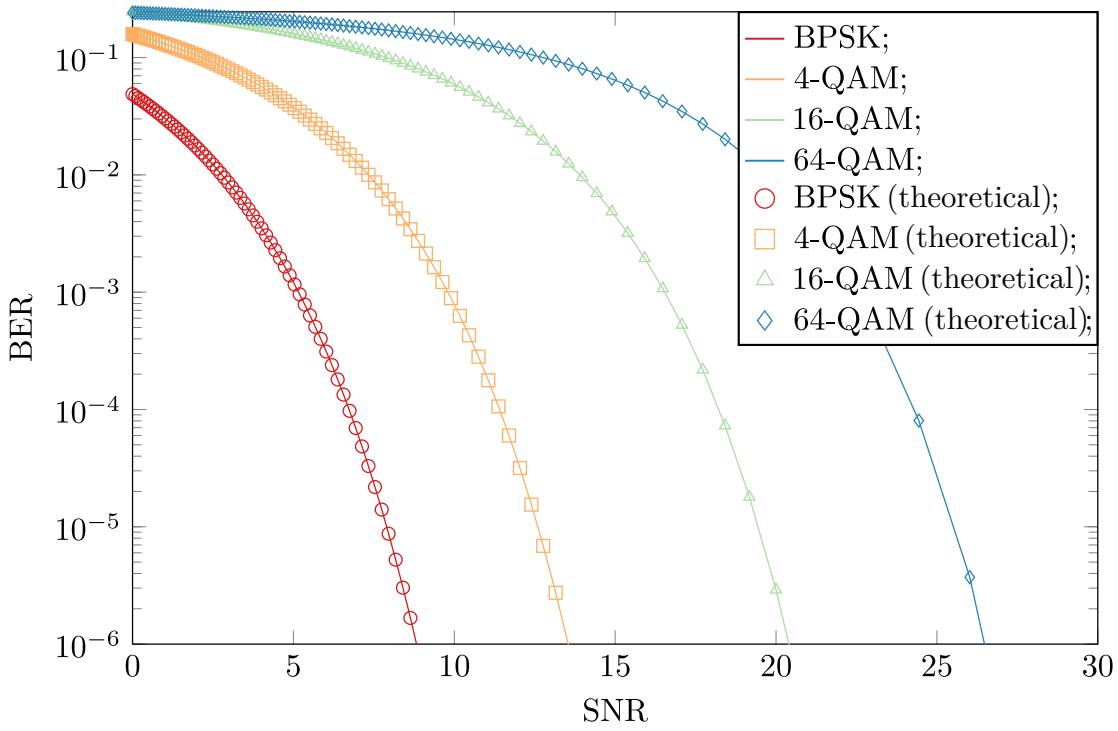

Fig. 4.2 BER vs SNR using the EVM approximation

theorem' [132]. The theorem states that a neural network can learn to approximate any function, however since the FFT/IFFT pairing is already exactly known and computationally efficient, we do not incorporate this into the neural networks task. We again opted for a hardware-in-the-loop approach to generate and transmit the SEFDM and OFDM signals that are under test. The OFDM signals were generated as described in (2.29), and the SEFDM signals were generated as described in (2.55). The transmitter was an NI USRP-RIO 2953R that was operated via the accompanying LabVIEW software. For each independent value of alpha tested, we generated an independent pseudo-random binary sequence of length $2^{16}-1$ which has been mapped to a quadrature phase shift keying constellation scheme. After passing through a serial to parallel converter, the Hermitian symmetry property of the DFT is utilised so that the OFDM and SEFDM symbols that are generated by the inverse FrFT (IFrFT) are purely real valued. The generated signals are then passed through a parallel to serial converter before being digital to analog converted and passed to the transmitter driving circuit to perform intensity modulation as described in Fig. 2.4. The LED used was a Vishay Semiconductors VLMS1500-GS08, which features approximately 50 MHz of bandwidth [133]. The signal was modulated with 90 mA dc and 2 Vpp data swing using a Mini-Circuits ZFBT-4R2GW+ bias-T, maximally occupying the linear

region of the LED. The addition of the bias voltage results in dc biased optical OFDM (DCO-OFDM), a form of OFDM suitable for transmission over an LED by biasing the input signal such that it is contained within the linear region of the LED's LIV response. As SEFDM is generated via the same process as OFDM but with smaller bandwidths however, the process of dc biasing SEFDM to produce an optical SEFDM signal works the same way. The transmission distance was 0.5m and the receiver used was a Thorlabs PDA10A2, featuring 150 MHz bandwidth and a receiver responsivity of 0.38 A/W at a wavelength of 630nm output from the LED. The signal bandwidth was set to be 1 MHz, centred around 10 MHz, much smaller than the link bandwidth, to ensure flat-band response and that the experiment was explicitly testing the EVM loss performance and avoiding any attenuation via out-of-band transmission. The aim of this work was to evaluate the effectiveness of the proposed loss function for receiving symbols after transmission, not to evaluate the transmission medium or transmission channel. In this spirit, the experiment was performed inside an isolated anechoic chamber and all experiment equipment was operated from a control station externally. This results in a stable, repeatable configuration to isolate any external factors on the systems' performance.

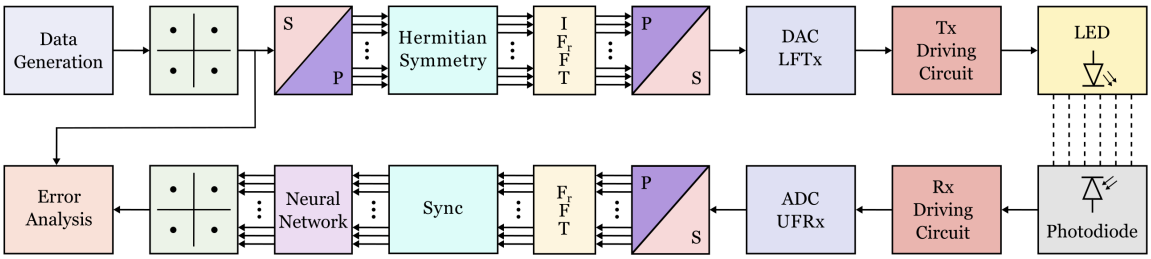

Fig. 4.3 Block diagram of the physical system implemented used to test the proposed methods

4.2.3 EVM as a Cost Function

The EVM that was defined above needs to be reworked to be directly implemented in a practical machine learning setting. As was defined, there is no relation to the neural network or the machine learning process, thus it needs to be in a form that the gradient descent algorithm can aim to optimise the learning process as to minimise the EVM at each subcarrier. We first introduce an error vector magnitude vector, E_k, as [116]:

$$E_k = \left(I_k - \hat{I}_k\right)^2 + \left(Q_k - \hat{Q}_k\right)^2 \tag{4.6}$$

where I_k and Q_k are the k^{th} ideal in-phase and quadrature component respectively, and $\hat{I}_k$ and $\hat{Q}_k$ are the measured versions. This allows for the calculation of the rms EVM as [116]:

$$\mathbf{E}_{\text{rms}} = \sqrt{\frac{\frac{1}{N}\sum_{k=1}^{N} E_k}{\frac{1}{N}\sum_{k=1}^{N}\left(I_k^2 + Q_k^2\right)}} \tag{4.7}$$

We incorporate this into the loss function when training the neural network receiver, instead of the mean square error. As is clear from the above, this requires the loss function to have knowledge of both the real and imaginary constituents of the input signal, therefore the design of the neural network has to take this into consideration. This is a simple modification that requires increasing the number of neurons of the input layer by a factor of 2. Whilst this does increase the complexity of the individual network, overall complexity is maintained by the fact only a single network is required, rather than each one for in-phase and quadrature signal components. A second advantage of doing this, is by enabling the neural network to observe both signal components, it is better able to learn how to recover the signal. This is because there is additional co-information embedded in the Cartesian form of the constellation without splitting, such as combined phase information that can lead to more accurate joint recovery of symbols. If two neural networks are trained separately for each signal component, this information would remain unknown to the other and hence performance deteriorates comparatively. This will be further discussed in the proceeding section.

4.2.4 Neural Network EVM Receiver

The estimation and tracking of both the transmission channel H and the interference matrix $\mathbf{C}$ becomes highly complex when the channel is both non-static and fading as is the case with vehicular channel models. From empirical use we have gleaned that conventional information theory based equalisers, for example feedforward or decision feedback equalisers, are inherently limited in their ability to counteract and predict this extremely complex relationship. For this reason, we turned our attention to the potential application of artificial neural networks to undertake this task as neural networks posess several advantageous properties that we believe make them well suited

for the problem at hand. One such property is the ability of a neural network with non-linear activation functions, like those that are used herein and will be introduced later, are able to generalise to any arbitrary input-output relationship given there are sufficient neurons in the hidden layers [96], a property also sometimes referred to as general or universal approximation theorem [97].

Neural networks are able to generalise any input-output sequence given sufficient neurons in the hidden layers [96]. In principle, this means that given a sufficient signal-to-noise ratio, the neural network should be able to estimate $\mathbf{C}$, for any value of α. In this work, we test $\alpha \geq 0.6$, meaning a bandwidth saving up to 40%. The neural network is designed to be trained on sequences of received QPSK symbols after transmission over a visible light link using SEFDM with a given bandwidth compression factor, or OFDM where $\alpha = 1$. The VLC link is characterised in [134] with a bandwidth of > 50 MHz. We set our signal bandwidth to be 1 MHz to ensure flat-band response since this work is not about improving data rates. The signal was dc biased at 90 mA with an ac signal voltage of 2 Vpp, i.e. keeping within the linear operating region of the LED. A generalised diagram of the model used in this work can be seen in Fig. 4.4.

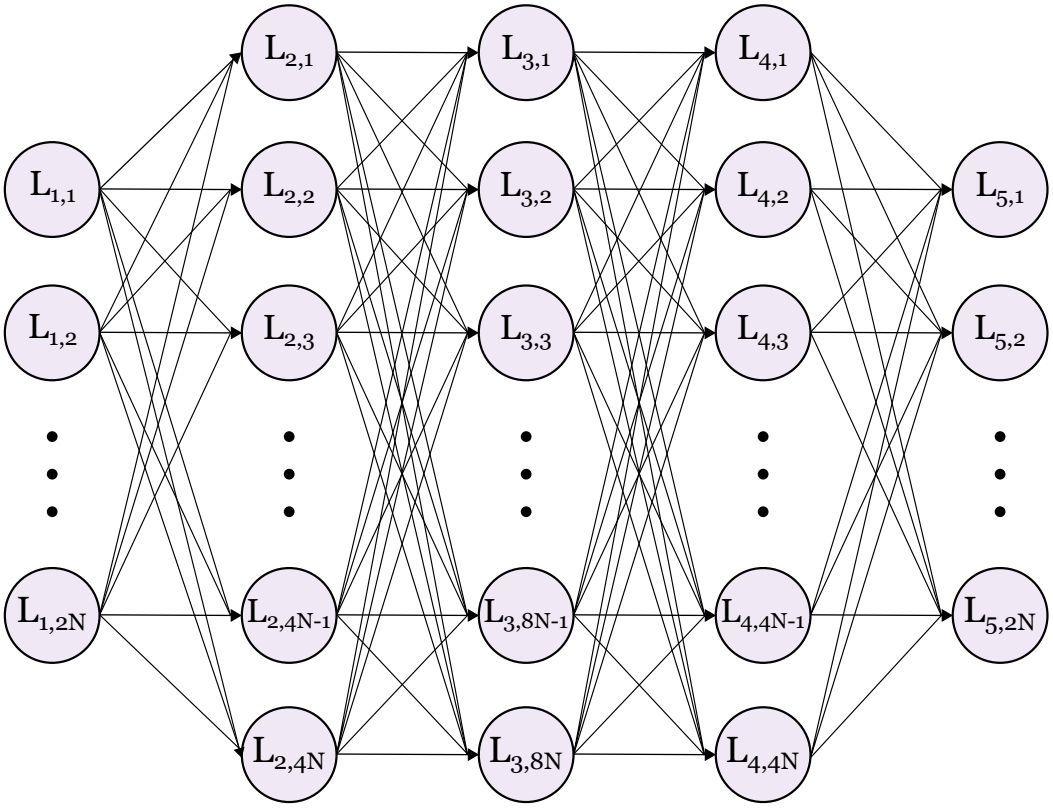

Fig. 4.4 Block diagram of the 'concatenated-complex' neural network

The neural network input layer contains $2N$ neurons, which allows for separation and concatenation of the real and imaginary components of N subcarriers, the reasoning for which was given in the previous chapter. The network requires sufficient neurons in its hidden layers to learn the complexities of the ICI and transmission channel, thus, inspired from [17], we opt for 3 hidden layers with $4N$, $8N$ and $4N$ neurons respectively. Each hidden layer uses the LeakyReLU activation function with the final output layer using a linear activation, and the model was trained with the Nadam optimiser. The received symbols are fed through the network until they reach the output layer, which also contains $2N$ neurons, mirroring the input. The estimate of the originally transmitted QPSK symbols are then obtained by combining the real and imaginary outputs. The typical mean square error which we are comparing against as a baseline is given as [115]:

$$L = \frac{1}{2N} \sum_{i=1}^{2N} (y_i - \hat{y}_i)^2 \tag{4.8}$$

where N in is the number of subcarriers, y_i is the ground truth value and $\hat{y}_i$ is the predicted value.

In the case of SEFDM there is known intercarrier interference, which is compounded by frequency selective and frequency relativistic perturbations. These perturbations also affect OFDM signals, however without the obvious preset intercarrier interference. If one views the neurons of the neural network in the context of being the real and complex data from neighbouring subcarriers, it can be seen that having this information available within the network means that both the correlated interference between subcarriers and the correlated frequency perturbations can be reasoned with internally. This in turn should increase the performance of the neural network within a communication setting, which will be demonstrated in the proceeding section.

4.3 Practical Results

We test the proposed method with three neural network architectures across several different values of $\alpha = \{1, 0.9, 0.8, 0.7, 0.6\}$, starting with an MSE loss on a split-complex network as a baseline. We then extend this to show the gain from utilising concatenated-complex networks, and then the further gain of implementing the proposed 'EVM loss' system. Fig 4.5 to Fig 4.8 show the received constellations for these three tested approaches for the different values of α. The results show notable reductions in variance as one first converts from a split-complex network to a concatenated-complex network,

owing to the extra indirect phase information being retained within the network. A further notable gain is then seen when converting from MSE loss in the concatenated-complex network to the proposed EVM loss, demonstrating the effectiveness of the approach.

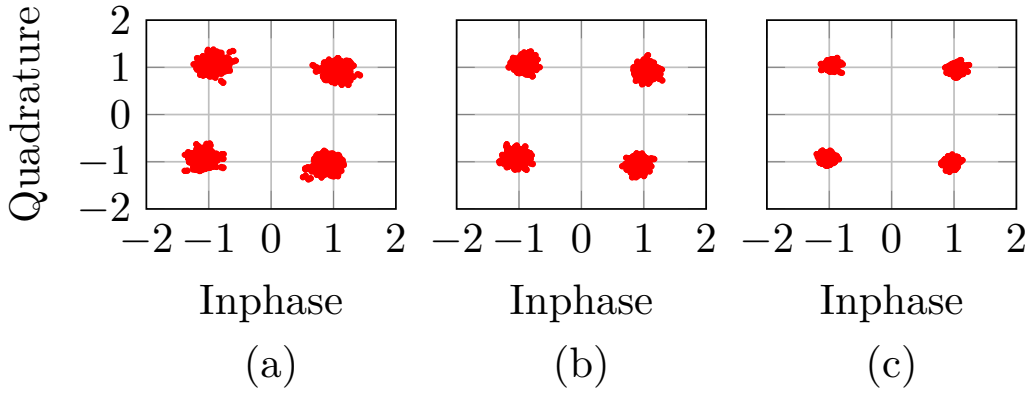

Fig. 4.5 Received constellations for (a) MSE loss w/Split Complex, (b) MSE loss w/Concatenated Complex, (c) EVM loss w/Concatenated Complex for SEFDM with $\alpha = 0.9$

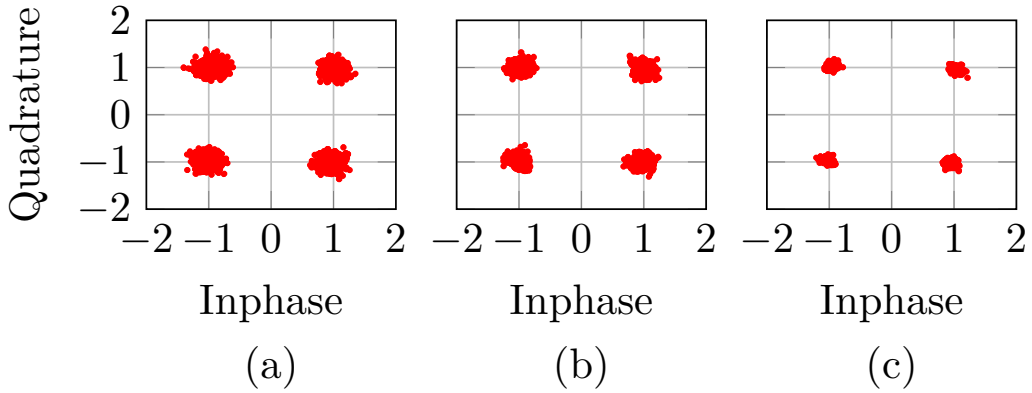

Fig. 4.6 Received constellations for (a) MSE loss w/Split Complex, (b) MSE loss w/Concatenated Complex, (c) EVM loss w/Concatenated Complex for SEFDM with $\alpha = 0.8$

The first test was performed for the OFDM data transmitted over the VLC link. It should be noted that for all experiments, the subcarriers at the extremes display a worse performance. This happens because the leftmost and rightmost subcarriers experience less interference from neighbouring subcarriers as there are less of them, this leads to a lower ratio of interference to noise at the extremes. As interference is not stochastic, a neural network can attempt to learn to undo the effects. On the other hand, when there

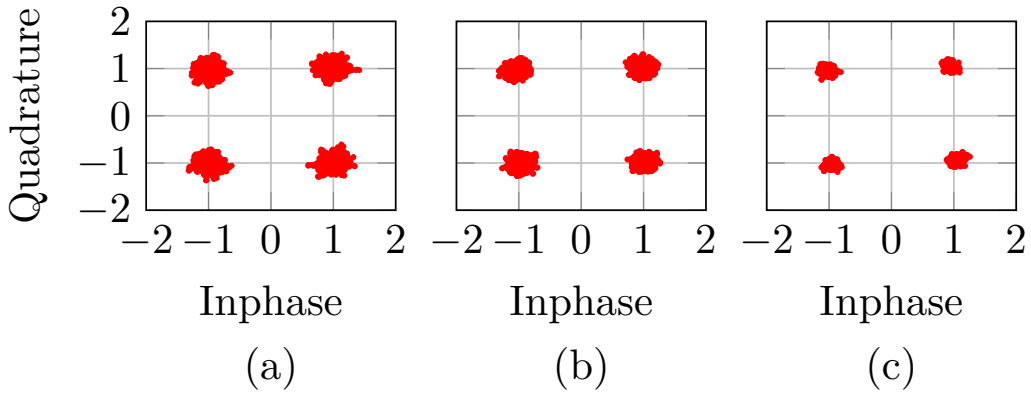

Fig. 4.7 Received constellations for (a) MSE loss w/Split Complex, (b) MSE loss w/Concatenated Complex, (c) EVM loss w/Concatenated Complex for SEFDM with $\alpha = 0.7$

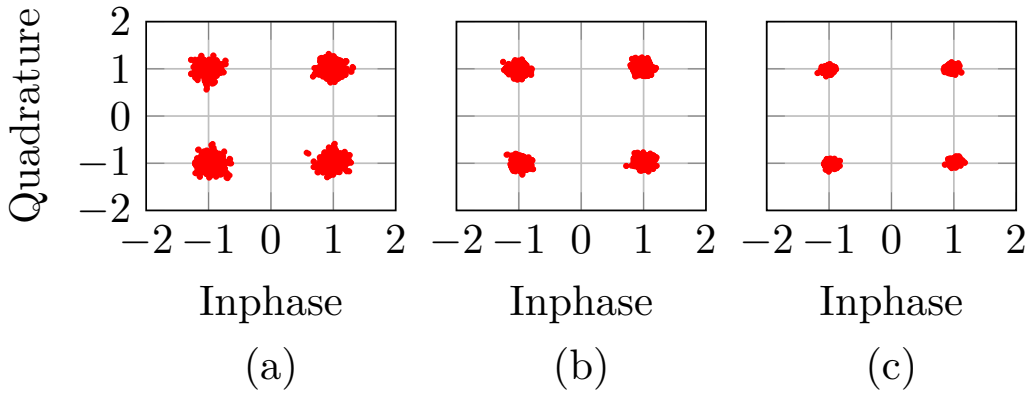

Fig. 4.8 Received constellations for (a) MSE loss w/Split Complex, (b) MSE loss w/Concatenated Complex, (c) EVM loss w/Concatenated Complex for SEFDM with $\alpha = 0.6$

is less interference the only pattern for the neural network to learn is the stochastic noise element, which reduces the performance of these subcarriers as a stochastic process cannot be learned. This also explains why this effect is most prelevent in the OFDM experiment, being orthogonal OFDM has no other noise except interference from neighbours which only arise due to channel effects, whereas with all SEFDM variants there is an inherent built in interference so as we see more interference noise present we see less differences in the performance at the extremes of the subcarriers. As can be observed from Fig. 4.9, the highest EVM was reported by the split-complex approach with the MSE error. The reason for this was eluded to previously, by creating a neural network for both the real and imaginary constituents of the input signal individually, one deprives the system of information that is embedded in the relationship between

the two components of the input signal. Communication system channels that affect the phase of the signal being transmitted alter this relationship, which when divided into real and imaginary signals and trained separately is not taken into consideration. This is rectified by modifying a single neural network to allow for both the real and imaginary constituents as input. The result of this allows the neural network to exploit the phase relationship during training, meaning a lower overall error vector magnitude.

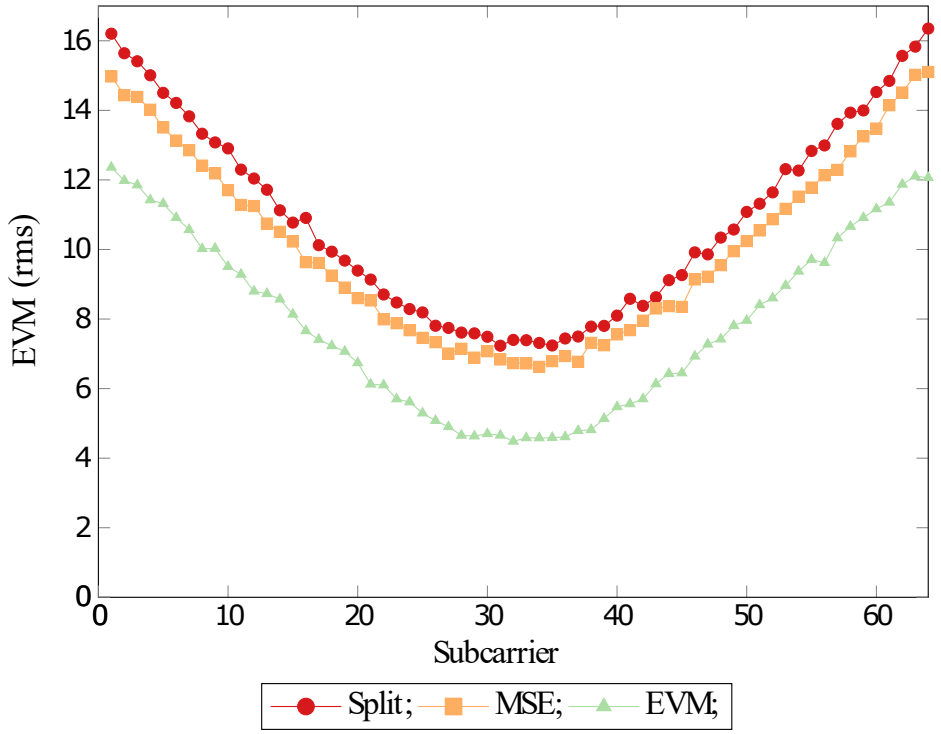

Fig. 4.9 Error vector magnitudes across the subcarriers for the 3 neural network receiver approaches for OFDM

Finally, the lowest error vector magnitude calculated across all subcarriers is achieved when the neural network uses the proposed EVM loss method as the objective function rather than the typical mean square error. This demonstrates that the mean square error is not an optimal objective function when training neural networks in a communication systems context. Whilst the two other neural networks have minimised their mean square error objective functions as optimally as possible, the fact still remains that when being used in a communication system the training metric and the evaluation metric no longer align, leading to a degradation of overall system performance. This can be seen when considering the implication of the implementation of the EVM loss function. The MSE loss function calculates a per neuron MSE loss, where the neuron has implicit access to the available information from its real and imaginary counterpart

via the hidden layer neurons, thus the real and imaginary neurons are loosely coupled. That is to say that the neurons have been made aware of the relationship at their inputs, but have then had their outputs evaluated independently from one another. The EVM loss function also calculates a per neuron loss, however, as well as the implicit information from the network being available at their inputs, the loss function itself explicitly examines the available information from its real and imaginary counterpart at the output, forcing a tight coupling all the way through the training process.

These same trends can again be seen in Fig. 4.10 to Fig. 4.12, where SEFDM was evaluated with $\alpha = \{0.8, 0.7, 0.6\}$ being used as the transmitted physical layer waveform respectively. With the system parameters set to transmit QPSK over a 1MHz VLC link, the data rate can be calculated at 2Mbps, or a spectral efficiency of 2b/s/Hz. At a maximum bandwidth compression factor of $\alpha = 0.6$ in these experiments, we have been able to transmit error free the same 2Mbps over 600kHz. This results in a spectral efficiency of $3.3\dot{3}$, a 67% improvement over the OFDM baseline.

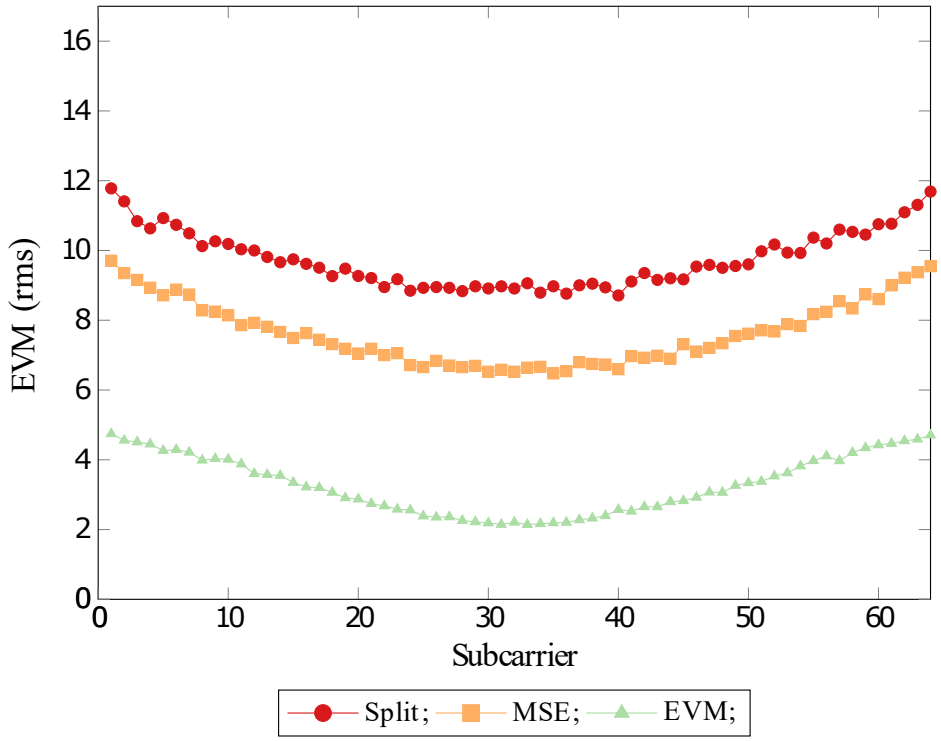

Fig. 4.10 Error vector magnitudes across the subcarriers for the 3 neural network receiver approaches for SEFDM with $\alpha = 0.8$

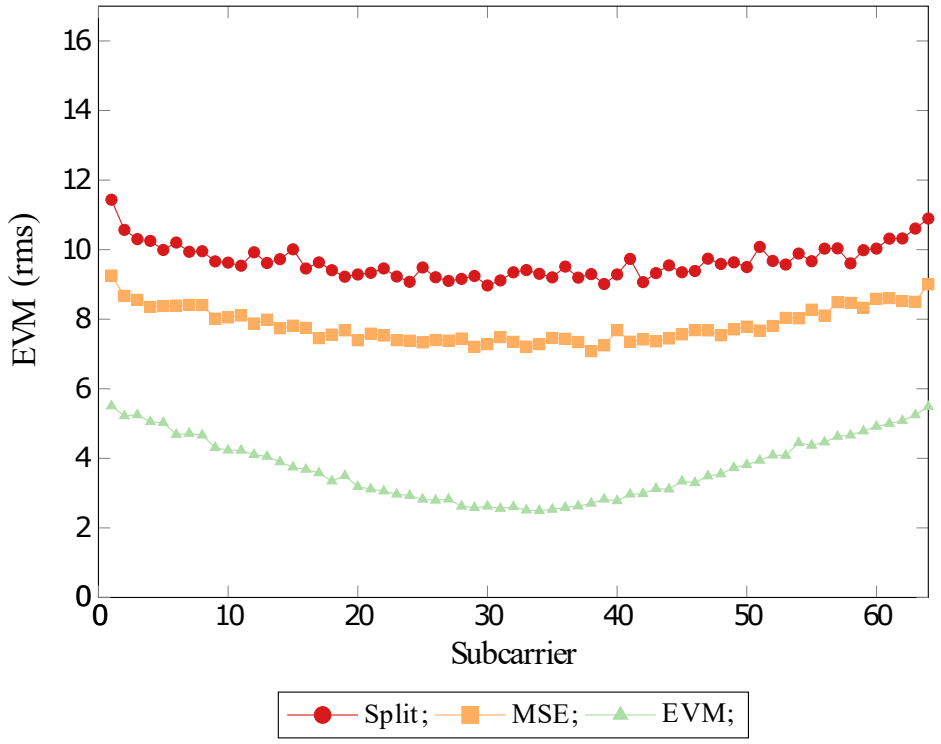

Fig. 4.11 Error vector magnitudes across the subcarriers for the 3 neural network receiver approaches for SEFDM with $\alpha = 0.7$

4.4 Conclusion

This work has demonstrated the shortcomings of two of the typical approaches used in the implementation of neural networks in communication systems, namely; the split complex approach, whereby transmitted signals are split into their real and imaginary constituents and used to train two separate neural network models, and the use of mean square error as an objective function.

It was shown that by splitting the signals into their real and imaginary components, the system was being deprived of valuable information that can be exploited and used to more accurately recover the originally transmitted symbols. Extending this, it was also shown that there is a disconnect introduced between the training of a neural network using the mean square error objective function when ultimately the communication system will be deployed and evaluated via its bit-error rate performance. This was rectified via the implementation of the proposed EVM loss method which seeks to reconnect the disparity between the evaluation performed when training a neural network and the evaluation of the overall end-to-end communication system. The results validated the capabilities of the method, which consistently obtained the

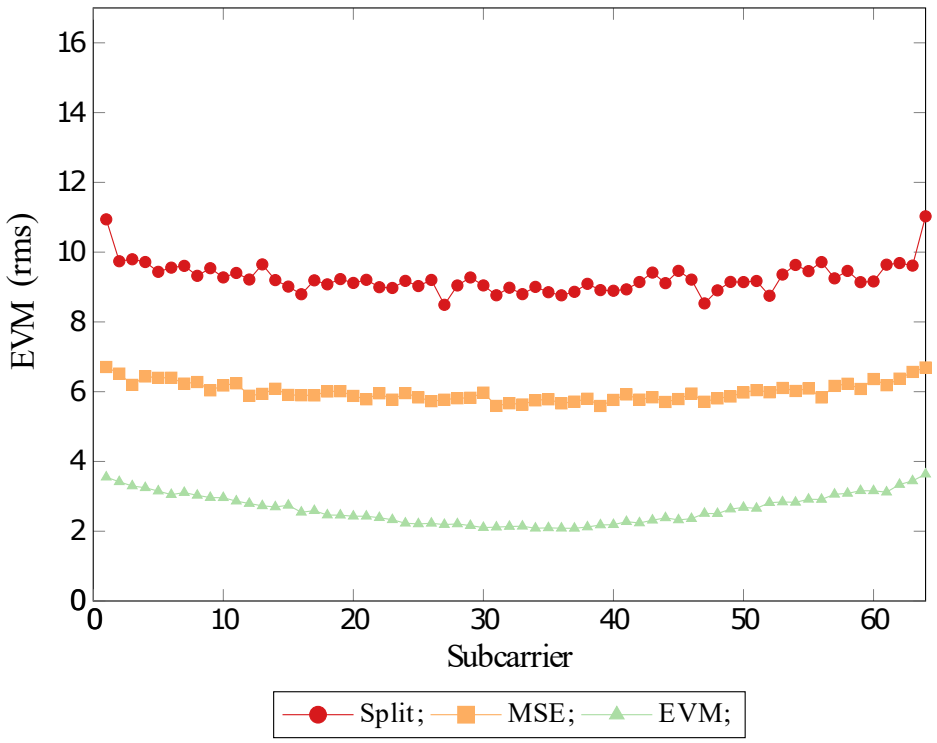

Fig. 4.12 Error vector magnitudes across the subcarriers for the 3 neural network receiver approaches for SEFDM with $\alpha = 0.6$

lowest overall EVM when deployed in a real experimental setup using OFDM and SEFDM with varying bandwidth compression factors. It was also shown that a spectral efficiency improvement of 67% was obtained at zero cost within the experiment, with a 40% reduced bandwidth being capable of communicating error free at the given data rate.

Chapter 5

Doubling Spectral Efficiency with SEFDM and EVM Loss

5.1 Introduction

A long-standing problem in visible light communication has been the low modulation bandwidths exhibited by the light-emitting diodes used as transmitters [135]. Spectrally efficient advanced modulation formats attempt to optimise bandwidth usage and as such, non-orthogonal schemes such as faster-than-Nyquist and spectrally efficient frequency division multiplexing have been gaining increasing attention over the last decade.

The first SEFDM proposal was reported in 2006 by Rodrigues and Darwazeh [54] as a variant of OFDM where subcarrier orthogonality is broken by reducing the space between subcarriers beyond the $1/T$ orthogonality limit. This results in a reduced signal bandwidth without compromising data rates, at the cost of self-induced inter-carrier interference [54]. Obviously, ICI comes with a bit-error rate penalty but there have been propositions to overcome this, including sphere decoding, iterative detection and artificial neural networks [136], enabling compression rates up to 20% with low SNR penalties.

Simultaneously, the use of artificial neural networks for compensation of perturbations such as ICI in communication systems is an ever-increasing topic of interest. One of the key aspects missing from the current state-of-the-art are domain-related methods of training and evaluating these systems, as communication systems are typically evaluated via bit-error rates. However, as demonstrated previously, evaluation of bit-error rate as an objective function for an optimisation target is not feasible when training neural network models, for this reason the mean square error is typically used

as the objective loss function of the training optimiser [137]. This causes a disconnect between the objective function of the model, and the eventual metrics that will be reported when the model is deployed. Thus, a performance gain would be expected to be obtained when the objective function of the training process is aligned with the metrics that the communication system will be evaluated against. Therefore, it was stipulated and shown that the deployment of the error vector magnitude, due to its close relationship with bit-error rate [120], can be used as the objective loss function. It will next be shown how one can utilise this method effectively to improve the performance of symbol receivers outside of the vehicle-to-vehicle communication scenario.

In a general neural network for communications setting, EVM is useful because it gives an error value in relation to a known ideal value and is available instantly for any batch size used in the system, unlike the unit-less bit-error rate metric that requires vast numbers of bits to converge. Experimental investigation examines the performance of the EVM objective loss function against that of a traditional MSE-based system in split-complex and concatenated neural network architectures. SEFDM signals are created in a software defined radio to generate increasingly severe ICI by introducing bandwidth compressions of up to 70% in comparison to the nominal OFDM bandwidth. The SEFDM signals are then transmitted via LED over free-space and captured via a photodetector. The signals are then taken offline to test the proposed neural network architectures via a Python environment. It is observed that optimal performance is achieved when using EVM as the objective loss function in the neural network and up to 50% bandwidth compression can be supported whilst maintaining the same data rate as the OFDM case.

5.2 Practical Implementation

5.2.1 SEFDM Receivers

In this work SEFDM will be further examined with data that was obtained experimentally using a wireless visible light communication link; what practical difficulties has it faced, what is the real world impact on its spectral design, and why these make SEFDM a good candidate for the target of a neural network symbol receiver.

It is stipulated that this technique is medium independent and was tested on a VLC link due to availability, but also works for radio links.

The reader is referred to the generation of an SEFDM signal as described in (2.55), where upon generating the signal, it is noted that the resulting SEFDM signal contains deterministic ICI, which can be modelled as a correlation matrix, Λ, between the subcarriers as shown in (5.1) [138]:

$$\Lambda_{m,n} = \exp\left(j\pi\alpha\left(m-n\right)\right)\exp\left(\frac{-j\pi\alpha\left(m-n\right)}{Q}\right)\left[\frac{\operatorname{sinc}\left(\alpha\left(m-n\right)\right)}{\operatorname{sinc}\left(\frac{\alpha(m-n)}{Q}\right)}\right] \tag{5.1}$$

which reduces to an identity matrix, $\mathbf{I}$, for OFDM. A visual example of the change in the correlation matrix can be seen in Fig 2.12.

As can be seen from the figure, the correlation i.e. the ICI for higher values of α is more contained and isolated. This means data subcarriers are not leaking over their boundary and contaminating neighbouring samples. On the other hand, as the value of α decreases it can be noted that the correlation is not very well isolated, meaning a subcarriers data stream is going to be impeding on subcarriers not localised to its immediate vicinity. In the OFDM case, as by their orthogonal design, this correlation matrix is perfectly localised hence the resulting identity matrix $\mathbf{I}$.

In [139] it was proved that the ICI on its own can be modelled as a Gaussian process for high N, characterised by its first and second moments i.e. the mean and variance, respectively. Using a symmetric QAM constellation as we are in this work, results in a mean $\mu = 0$, thus the ICI can be completely modelled by its variance. We can then infer that the ICI variance σ_n^2 on the n^{th} subcarrier caused by all other subcarriers is [139]:

$$\sigma_{I,n}^2 = \sum_{m=0, m\neq n}^{N-1} \sigma_{m,s}^2 \left|\Lambda_{m,n}\right|^2 \tag{5.2}$$

where $\sigma_{m,s}^2$ is the variance of the data symbols, s. This leads to an expression for total variance due to ICI within a single SEFDM symbol as [139]:

$$\sigma_I^2 = \sum_{n=0}^{N-1} \sigma_{I,n}^2 \tag{5.3}$$

which in a system with normalised transmit power P, the total interference can be upper bounded by $P\sigma_I^2$.

The effect of this variance is ultimately to reduce the bit-error rate performance by increasing the probability of a bit error P_e. From [140] we know the mathematical probability of an error for a standard QPSK OFDM system is given by:

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{5.4}$$

where E_b and N_0 are the bit energy and the noise power spectral density, respectively. Q is the tail distribution function of the normal distribution. This equation however no longer holds for SEFDM with an $\alpha < 1$ due to the aforementioned extra variance induced by the ICI. To account for this we modify the error probability equation as [139]:

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_0 + 2E_b\sigma_I^2}}\right) \tag{5.5}$$

From the above, it can easily be reasoned that this simplifies to (5.4) when $\alpha = 1$ as the variance parameter $\sigma_I^2 \to 0$. If one now turns their attention to the practicalities of utilising SEFDM in an application, it will now be explored some of the difficulties faced. The generation of SEFDM signals based on a simple fractional Fourier transform is well covered herein, thus will not be discussed further. For more information on practical generation the reader is referred to [141]. Of interest to this work are the practical difficulties faced with the symbol receiver in an SEFDM system. One could discuss multiple types of symbol receivers, but this work shall focus on two; the optimal maximum likelihood detector and the state-of-the-art sphere decoder. For a more full review of a broader range of receivers, the reader is directed to [142].

The maximum likelihood receiver is common throughout a lot of communications systems, however is rarely used in practice. It states that given some received signal y, determine the transmitted vector x with the highest a posteriori probability. This results in the following inequality [131]:

$$\arg\min ||y - Hx||^2 \tag{5.6}$$

where y are the received symbols, H is the channel matrix, x are the originally transmitted symbols, and $||\cdot||$ denotes the Euclidean norm.

One thing of importance to note is the reliance on H, the channel matrix. Whilst a common requirement, it is worth pointing out that this is a large assumption. The transitter nor the receiver can know about the channel conditions ahead of time, therefore methods must be used to estimate the channel matrix such as pilot symbol insertion. This however reduces the spectral efficiency of the communication, as they utilise bandwidth that otherwise would contain data symbols. One way of solving the maximum likelihood inequality as it is written is to perform an exhaustive

search over the possible combinations and outcomes. As one can imagine this is very computationally inefficient, infact, it is known to be of exponential complexity [143] with the cardinality of the communication constellation and the number of subcarriers.

In an attempt to achieve near optimal symbol reception but with a more realistic implementation from a practical standing, methods such as sphere decoding, a branch of tree search detectors, have arisen and are currently considered state-of-the-art in SEFDM symbol detection at time of writing. Sphere decoding is based on the premise of the maximum likelihood operator, but uses techniques to reduce its complexity. It performs a search over received signals that lie within a hypersphere of radius R. At its heart, has a relatively simple objective; only search for points in a lattice that are contained within some sphere of radius R around the given vector. By placing the radius, one has restricted the search space and therefore the number of computations that need to be performed. Via the transitative property, it is clear that the closest lattice point within the sphere will also be the closest lattice point for the whole lattice, therefore the point chosen is guaranteed to be correct. Whilst this explanation is simple, it still poses some difficult questions. How does one choose a suitable value for the search radius R? And how can one determine which lattice points lie within the sphere? As for the question pertaining to selecting a suitable value for the radius R; a value too small and one runs the risk of electing a radius that contains zero points inside, a value too large and one is performing more exponentially more computation than is required. The ideal candidate is known as the covering radius, which is the smallest radius of spheres centered within the lattice that covers the entire space. This guarantees that a point exists for any input vector, however determining the covering radius for a given lattice is itself known to be an NP hard problem [144]. The second question regarding the derivation of the lattice points that are inside the sphere is an important question, as if one needs to test every single point to determine whether it lies within the radius, then one might as well not have begun the process as an exhaustive search of all points is already being carried out. It is important to note that sphere decoding addresses this second question, it poses no solution to choosing the value of radius R.

Mathematically, one can describe the inequality that sphere decoding is attempting to solve as [145]:

$$\arg\min ||y - Hx||^2 \leq R^2 \tag{5.7}$$

where y are the received symbols, H is the channel matrix, x are the originally transmitted symbols, and $||\cdot||$ denotes the Euclidean norm. Note the similarity to the

maximum likelihood inequality, and again the reliance on obtaining an accurate value for the channel matrix H. Under this constraint, one only needs to examine lattice points that exist within the hypersphere, calculate their distance metric under the maximum likelihood assumption, and prune any paths that cannot give a minimum solution. Following this principal, the distance metric can be viewed as the sum of non-negative functions that take an increasing number of arguments the further from the root node one gets, i.e.

$$d = f(x_1, \ldots, x_k) = h(x_k) + h(x_k, x_{k-1}) + \cdots + h(x_k, \ldots, x_1) \tag{5.8}$$

This can be considered as a tree as illustrated in Fig 5.1. Each branch in the tree corresponds to an intermediate distance value.

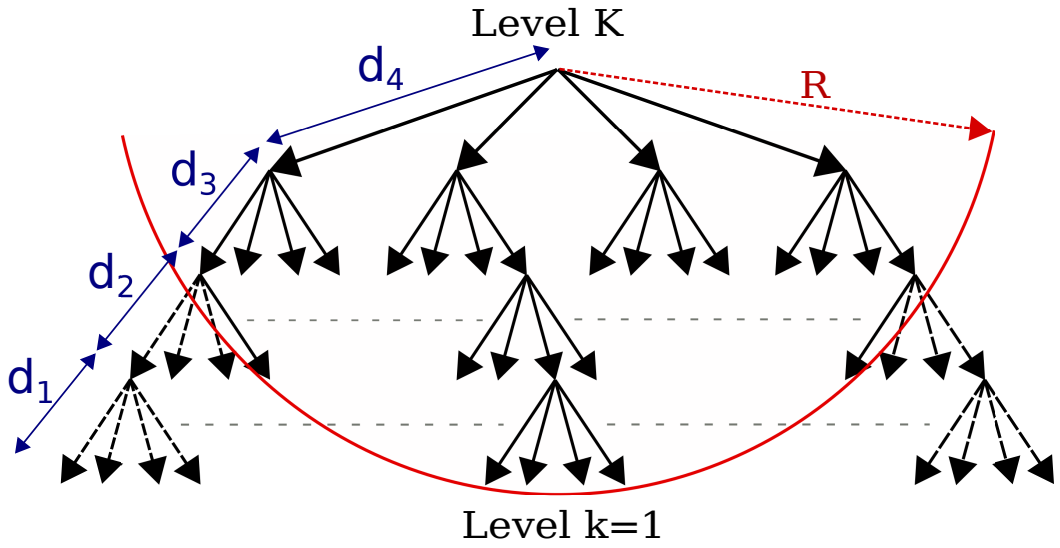

Fig. 5.1 Graphical illustration of the sphere decoder process

The aim is to fully explore as few nodes in the tree as possible in order to keep the computational cost low. Once a node is explored all the way until $k = 1$, the lowest distance path is retained and all other candidate paths are pruned. If this candidate path is lower than the previous lowest candidate path, this path becomes the new solution. This continues until all paths have been pruned, and the remaining solution is known to be the absolute minima.

Ultimately, it is well established that maximum likelihood is the optimal solution to symbol detection, but as its computational complexity grows exponentially with the number of subcarriers and the size of the communication constellation, it is not a suitable candidate for real world applications. To combat this, sphere decoding was presented as a more computationally efficient alternative, but it still has several serious limitations.

Firstly, it still depends on the invertability of the ill-conditioned correlation matrix which limits its practical efectiveness to systems with a small number of subcarriers and a value of $\alpha \geq 0.8$ [142]. Secondly, because complexity still increases significantly with the number of subcarriers, and also increases significantly with lower α (due to the fact that it introduces extra noise which causes uncertainty, so the search radius needs to be altered to reflect this), this places a bound on the realistic performance achievable in a real setting. For instance, in [62] the authors attempted to address this limitation with a method known as generalised sphere decoding (GSD) which performs regularisation on the correlation matrix based on Cholesky decomposition to improve its ill-conditioning, and still concluded its applicability was restricted to $N \leq 32$ and $\alpha \geq 0.75$.

5.2.2 Experimental Test Setup

To investigate the effectiveness of the proposed EVM loss approach, it is experimentally verified using OFDM and SEFDM over a VLC link. The data was generated and tested over two real time National Instruments software defined radios (NI 2943 and NI 2953). In the first device, an independent data stream of $2^{16} - 1$ bits were generated and mapped to a 4-ary quadrature amplitude modulation (4-QAM) constellation.

After serial-to-parallel conversion, an inverse FFT is used to generate the OFDM and SEFDM frames. The number of subcarriers used are $N = \{4, 8, 16, 32, 64\}$, with 64 being at least four times more than any similar SEFDM-NN work [136] to the best of the authors knowledge. The induced ICI is the basis for the requirement of the neural network at the receiver, since the ICI generated is deterministic and correlated, it is possible to remove it with advanced equalisation. The size of the IFFT(/FFT) is dependent on the compression factor α, following [146].

The symbols are then serialised and converted to the analogue domain via an Ettus Research LFTX (30 MHz bandwidth) before transmission over the VLC link. It was decided not to use a cyclic prefix in this work and instead used zero-padding between data frames, following [147]. The experimental setup for the VLC circuit is as described previously and shown in Fig. 4.3

At the receiver, the signal was post-amplified by a factor of ten using a Texas Instruments THS3202 amplifier before being fed to the second NI software defined radio via analogue-to-digital conversion by an Ettus Research LFRX. Then, the data is synchronised and converted to parallel and passed through the forward FFT before being committed to storage.

The data was then loaded into a Python environment (offline) where the neural networks under test are deployed and the signals are applied to the neural network as described in Fig 3.5. The calculation of EVM follows the form given in (4.1). As mentioned, an advanced equalisation format is required to remove the ICI introduced by carrier compression of the SEFDM system. It is well known that neural networks can approximate any input-output sequence, given sufficient neurons in the hidden layer(s) [148]. In principle, this means that given a sufficient SNR, number of neurons and hidden layers, the neural network can estimate the ICI for any value of α. In this work, we evaluate $\alpha = \{1, 0.9, 0.8, 0.7, 0.6, 0.5\}$, saving up to 50% of the nominal 1 MHz signal bandwidth, doubling the spectral efficiency.

Two different neural network architectures are tested; the split-complex approach, and the concatenated approach. In both architectures, there is an input layer, three hidden layers and an output layer. Three hidden layers were selected as to enable sufficient neurons and hidden layers to provide sufficient processing power to estimate the complex ICI and channel characteristics, inspired by [137]. In the split-complex approach, two neural networks are used, one each for the in-phase and quadrature symbols. The number of neurons was set in the (a) input layer to N $(4-64)$, (b) hidden layer to $2N, 4N, 2N$ and (c) output layer to N. For the concatenated approach, the neural network is widened by a factor of two, doubling the number of neurons in each layer in order to allow the input of both the in-phase and quadrature symbols into the same network. This maintains an equal number of neurons per input across the two networks.

Each hidden layer uses the leaky rectified linear unit as its activation function, while the final output layer uses a linear activation, and the model was trained with the Nadam optimiser. The initial learning rate was set to 10^{-3} with a reduction of factor 10^{-1} when the validation loss did not decrease over a period of five epochs. The received symbols are fed through each neural network under test until they reach the output layer, where the 4-QAM constellation is reconstructed. While training, the optimiser requires a function to evaluate the training performance, and this is an objective function which is normally assumed to be MSE (defined in [149]). In general, using MSE loss as the objective function for the optimiser works well for one-dimensional constellations such as pulse amplitude modulation, but the performance of MSE loss in both split complex and concatenated networks suffers when considering complex data. This is because MSE loss focuses explicitly on sample amplitudes and does not consider any phase offset from the ideal constellation points. This forms the basis of our motivation to replace it with the EVM which does, and allows the

network to gain knowledge of the relationship between the in-phase and quadrature components, unlike split-complex/MSE-activated networks. This is because there is additional co-information embedded in the Cartesian form of the constellation without splitting, such as combined phase information that can lead to more accurate joint recovery of symbols. If two artificial neural networks are trained separately for each signal component, this information would remain unknown to the other and hence performance deteriorates comparatively. The proposal of using EVM as the objective function in a concatenated network is thoroughly tested and compared with (a) MSE in a split-complex network and (b) MSE in a concatenated network. The extent to which the choice of subcarrier number N affects the output is also investigated.

5.3 Practical Results

Following on from the previous chapters, it raises the question - if one was to keep compressing further, will this same pattern of spectral efficiency saving be observed? The answer is yes to a point, for example, using the lowest compression ratios tested, i.e. $\alpha = \{0.5, 0.6, 0.7\}$, in Fig. 5.2 the measured EVM at the output of the neural network is shown for the three architectures under test. It is apparent immediately that the concatenated-complex neural network with EVM as the objective loss function achieves the most optimal performance for all compression bandwidth ratios tested. The overall highest performing scenario of the entire system was obtained when $\alpha = 0.6$, where the neural network reduced the EVM to approx. 3 dB, i.e. approximately error free. For values of $\alpha < 0.5$, the ICI becomes overwhelming and performance decreases, so these will be discussed later in this section.

Evaluating the other neural network architectures, for the concatenated network using MSE as the objective loss function the EVM value was measured to be of the order of 6 dB, and for the split complex case again using MSE as the objective loss function the measured EVM was approximately 9 dB, showing the performance penalty of adding 3 dB of EVM between successive architectures under test. A similar relationship is found when for $\alpha = 0.5$, with slightly greater difference ratios between the different architectures. This can be attributed to the hugely increased ICI, as at $\alpha = 0.5$ one is describing a scenario where subcarriers are 50% overlapped on both sides with their neighbours. At this level, split-complex network has reached its performance limit, shown by the further reduction in performance in comparison to the concatenated network.

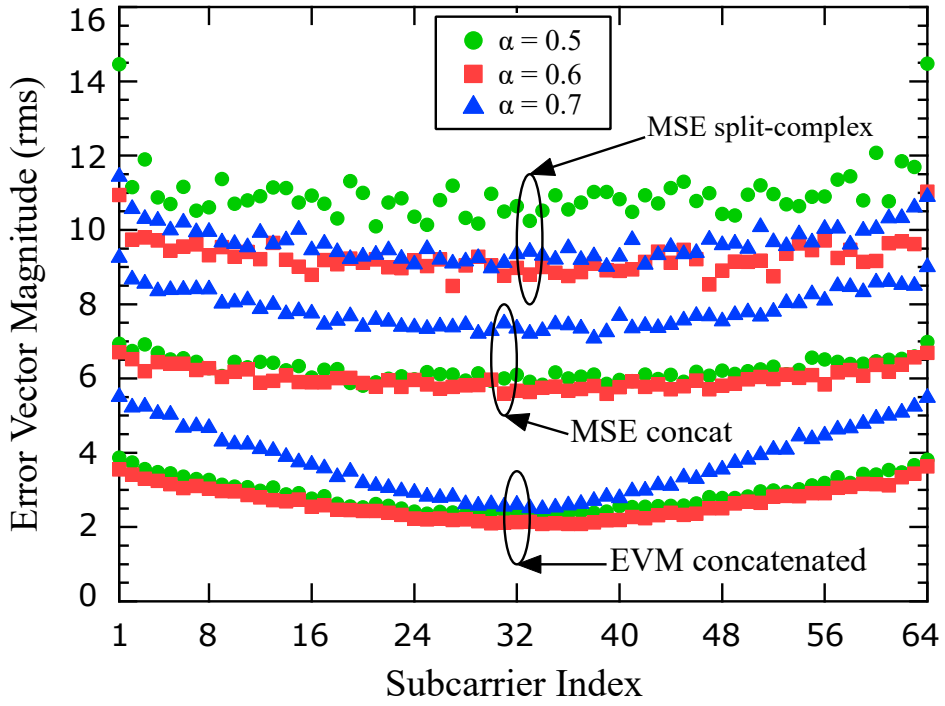

Fig. 5.2 Error vector magnitude results for each of the networks under test for $N = 64$

It is also worth noting that, to the best of the authors knowledge at time of writing, this is the highest value of α ever obtained in an SEFDM system which one does not attribute solely to the proposed EVM architecture, but also attribute partially to the high signal to noise ratio in our VLC link along with the fact that the linear, flat-band response of the visible light system response was exploited.

In Fig. 5.3, the measured EVM values for the concatenated system with EVM as the objective loss function for all compression ratios are shown. One unforeseen aspect in this work was the predelection of the neural network to over-fitting when faced with negligible amounts of ICI (i.e. high values of α). Whilst neural networks are powerful tools for approximating complex input-output relationships, in the case of OFDM over a high SNR stationary channel as was used here, the optimal receiver is simply the inverse of the transmit side transform. In this scenario where the ICI is much less (referring back to (5.1), $C = I + \Lambda \rightarrow I + 0$ for OFDM) than the noise of the channel (AWGN in this case), the input-output relationship that is attempted to be modelled by the neural network is that of AWGN, which is known to be uncorrelated and stochastic, as opposed to learning meaningful relationships from ICI which has been shown to be correlated. As a result, the worst performance shown in Fig. 5.3 is surprisingly the OFDM system itself. The reason for this is as just preluded to, there is no ICI and no channel perturbations or non-linearities except for power attenuation and small levels

of noise that lead to the neural network estimating input-output patterns that do not exist due to the purely stochastic nature of the channel perturbations it was observing.

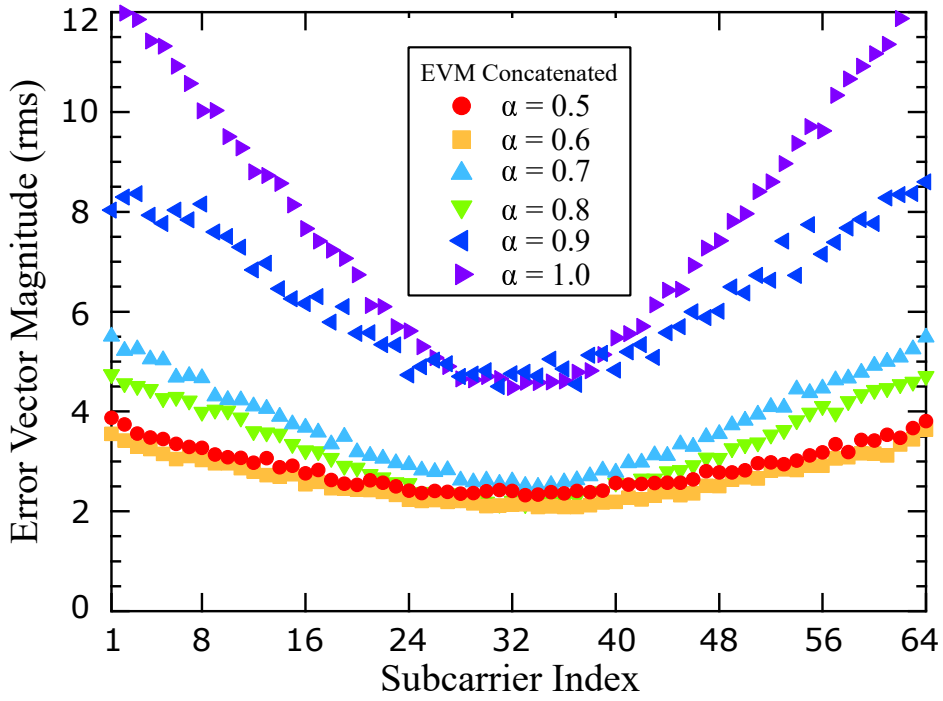

Fig. 5.3 The overall performance of the proposed system for all values of α tested for $N = 64$

Finally, the effect of varying N was investigated due to its close relationship to the ICI as was described in (5.2) and (5.3). Fig. 5.4 shows the measured EVM values for the neural network receiver including the α values < 0.5. For all cases, it was found that for $\alpha < 0.5$, our neural network performance degraded and the EVM values were increased, as is to be expected. Therefore, a maximum compression ratio of 50% was obtained while maintaining the same data rate as in OFDM hence its dismissal from previous results. It can be seen that the value of N has an impact on the neural networks performance, with higher values of N being preferred for lower values of α. This is likely due to the construction of the neural network, as more neurons are created for a higher N, which provides additional computational power that can take advantage of the increased levels of correlated self interference between subcarriers. This can be demonstrated numerically by referring to the complexity analysis discussion in Chapter 3. With this form of neural network, we have $18N$ neurons, therefore using 4 subarriers we have a total of 72 neurons versus the 288 neurons when using 16 subarriers and versus the 1152 neurons when using 64 subcarriers. This is a vast difference in learning

power between the networks which explains the disparity of the performance between the systems.

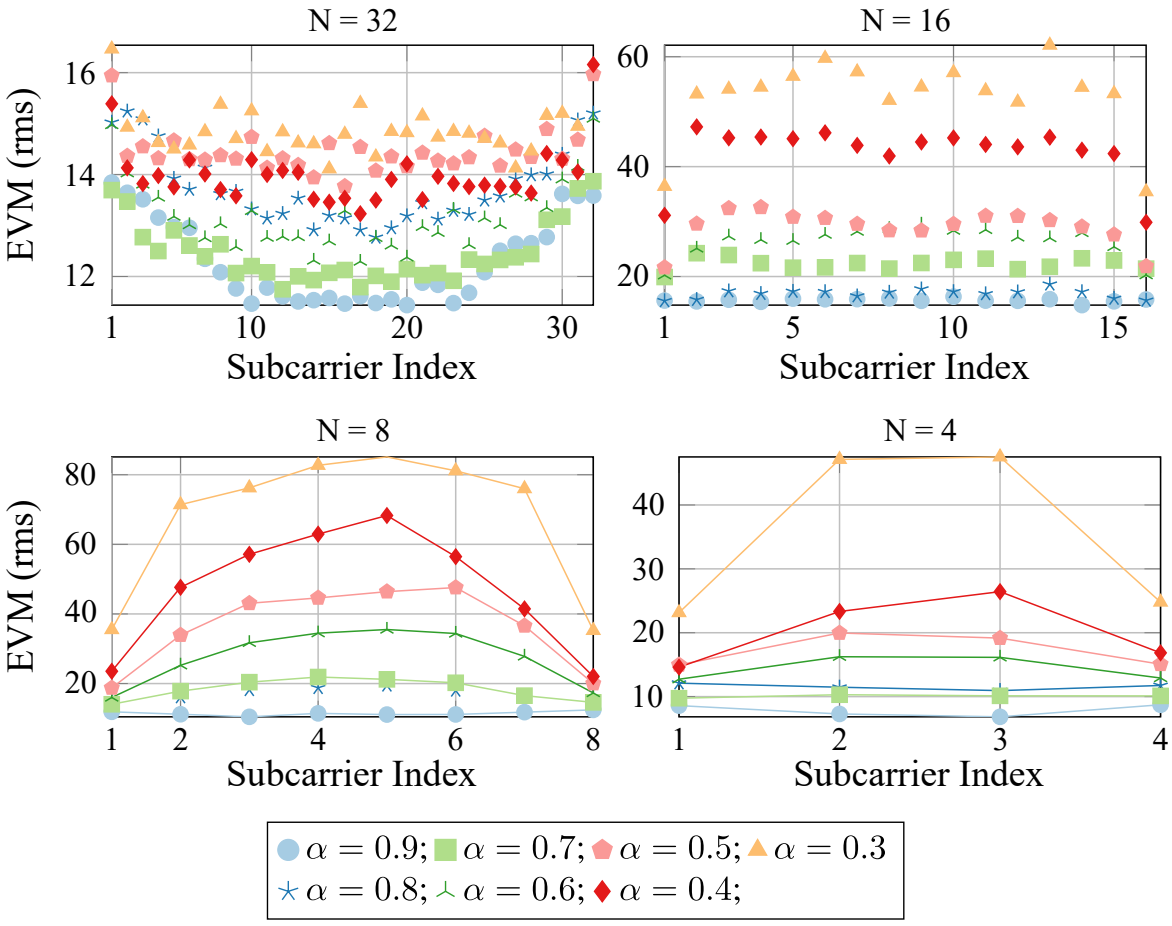

Fig. 5.4 Error vector magnitudes for varying N

5.4 Conclusion

This chapter has extended upon the previous works to demonstrate the efficacy of the use of neural networks in general communications outside of strictly vehicle-to-vehicle communication. By performing analysis in VLC networks, one can clearly see the spectral efficiency gains that extend beyond the initial scope. To the best of the authors knowledge at time of writing, the 50% spectral saving at a value of $\alpha = 0.5$ is the highest ever presented in the literature with any degree of signal recoverability.

The use of EVM loss has proved to outperform MSE as an objective loss function for the use of neural networks in communications, and further looked at the foundational reasoning for its successful implementation for SEFDM. This is not to say that neural

networks do not have their downside, it was also discussed and reasoned that under certain conditions the neural network was prone to overfitting and thus adapted poorly to the given task on new data. On the other hand, the difficulties SEFDM face in a practical implementation have been reviewed, it was demonstrated that the optimal maximum likelihood receiver is far too complex to be considered for practical use due to the fact it grows exponentially in complexity with the number of subcarriers and the constellation size. It was then shown that the current state-of-the-art, namely sphere decoding, whilst attempting to address the complexity issue of the maximum likelihood receiver only partially manages to do so and suffers limitations on its practial implementation as a result, with literature showing a restriction of $N \leq 32$ and $\alpha \geq 0.75$.

It was clearly shown that our novel neural network approach provides a viable alternative for practical implementations of SEFDM symbol receiving, whilst also placing a tighter upper bound on the complexity as the neural network presented herein does not grow in complexity with the size of the output constellation. The performance analysis has shown that neural networks provide the flexibility and the ability to reduce the impact of the combined channel and ICI in practial implementations, and that the novel EVM loss function with a concatenated complex network achieved the highest performance at the lowest measured EVM in all experiments undertaken, achieving ≈ 3 dB lower EVM than the common MSE loss.

Chapter 6

Conclusion and Future Work

Neural networks and non-orthogonal spectrally efficient waveforms have been regarded as promising technologies for future wireless networks such as vehicular communications. In this thesis, a thorough investigation has been undertaken into the practical efficacy of the joint application of these areas, with practical implementations being conducted in all experiments. In the following sections, a brief summary of this research and its key findings are outlined, followed by a brief discussion on potential topics aligned to areas related to the work of this thesis are presented as possible future research directions.

6.1 Conclusion

In chapter three, a novel neural network acting as a joint channel equaliser and symbol receiver was designed and implemented for use in a vehicular network for vehicle-to-vehicle communication. The harshest empirical channel model from the literature was emulated using a hardware-in-the-loop practical implementation. Experiments started from the point of an 802.11p compliant transmitter using OFDM as the physical layer waveform, then after setting a baseline, proceeded to break protocol and use SEFDM as the physical layer waveform to investigate the potential spectral efficiency benefits from the use of a non-orthogonal waveform. The proposed neural network receiver achieved an RMS error vector magnitude of 4.75 in the 802.11p compliant OFDM case over 52 data symbols, and SEFDM under the same conditions achieved a performance of $10.8 \rightarrow 25$ (RMS) EVM as the compression factor decreased from $\alpha = 0.8 \rightarrow \alpha = 0.4$, demonstrating a significant improvement to spectral efficiency which could in turn be used to service more users or achieve a higher total throughput.

In chapter four, the shortcomings of two approaches of implementing neural networks for communications were demonstrated, and proposals were made for both that improved the bit-error rate performance of the resulting network. Firstly, the common methods of forming the input and output layers of neural networks were investigated and the 'concatenated complex' neural network model was formalised which predicated that creating 'split complex' neural networks should not be preferred over 'contatenated complex' neural networks in most settings as to allow the neural network to gain valuable information from the relationship that exists between the real and complex components of a signal. Secondly, a new loss function was created for the specific purpose of training neural networks in communication settings. The error vector magnitude of a received signal when compared to the reference constellation was used to tightly couple the objective function of the neural network whilst training to the metrics of the wider communication system that the neural network was to be deployed in. The results showed that both methods resulted in increases in performance over the baseline, and the proposed EVM loss system was able to demonstrate a 40% spectral efficiency improvement over the OFDM baseline with error free transmission.

In chapter five, the methodologies from the previous chapter were extended to investigate the applicability of neural networks and EVM loss in a wider communications setting, namely VLC networks, with a particular focus on SEFDM as a spectrally efficient physical layer waveform. It was shown for the first time that a spectral efficiency saving of 50% could be obtained, effectively halving the bandwidth requirement for communication. It was also reasoned that compared to the state-of-the-art sphere decoder for SEFDM which in literature has stated restrictions of $N \leq 32$ and $\alpha \geq 0.75$, neural networks bound the complexity more reasonably as the size of the output constellation has no effect on the neural network complexity, unlike the sphere decoder.

6.2 Future Work

The application of artificial neural networks, and more broadly machine learning, in communication systems is a vast topic with many potential areas for research. In this section, a brief proposal of some possible future research topics are presented.

1. Extending on the work of chapter 3, an investigation into the possible role of including the transmitter in the learning process could be undertaken. The transmitter could alter and transform the transmitted signal in ways it has jointly learned with its counterpart neural network on the receiver to combat the link between them. This is not dissimilar to existing works such as [150, 151], where

autoencoders are deployed end-to-end and the wireless channel represents the hidden internal states of the network. The difference being autoencoders are a form of unsupervised learning, whereas the topic proposed herein would take advantage of supervised learning methods.

2. To the best of the authors knowledge, chapter 4 is the first work that challenges the foundations of the machine learning process for the specific context of neural networks in communications. This opens a plethora of future research directions that could be investigated to further improve the latency, performance, and complexity constraints that currently prohibit widespread adoption. One such example would be the investigation of non fully connected neural networks. By only connecting k neighbouring neurons, one could still provide subcarriers with inter-carrier information from those relevant within the vacinity, but reduce the complexity of the neural network by not sharing information from distant subcarriers with little impact on ICI. What value of k would be dependant on the communication method and the channel that would be used for communication. For example, high values of α in the SEFDM case would mean more neighbours should be included, as would channels that have high frequency selectivity or phase effects. Another potential example of future research would be into signal carriers outside of frequency division transmission. For example in [152] the authors use a neural network to schedule transmission slots at the MAC level based on observations made of the wireless spectrum, which they claim reduced the number of collisions by a factor of 15.

3. Extending on the work of chapter 5, future work could look to tackle the growing complexity due to the number of subcarriers, while this work has lowered the complexity resulting from increasing the cardinality of the output constellation it has not tackled the growing complexity that arises from increasing the number of subcarriers and matches the state-of-the-art sphere decoder in that regard. It is known from the correlation of the ICI that the major contributors are from the immediate neighbours of each subcarrier, therefore this limits the information gain distant neurons within the neural network will have via their shared connections due to the fully connected nature of the network. It is postulated that pruning the neural network of connections beyond a tunable, integer valued region n of each subcarrier would convert the exponential growth from increased subcarriers into linear growth.

References

[1] "Ensuring the connected 5g experience." [Online]. Available: https://www.anritsu.com/en_GB/testmeasurement/technologies/5g_everything_connected

[2] "Press Kit | Tesla UK." [Online]. Available: https://www.tesla.com/en_GB/presskit/autopilot

[3] "Google self-driving car project." [Online]. Available: https://www.google.com/selfdrivingcar

[4] "Pittsburgh, your self-driving uber is arriving now." [Online]. Available: https://newsroom.uber.com/pittsburgh_selfdriving-uber

[5] H. Jiang and G. Gui, *Channel Modeling in 5G Wireless Communication Systems*. Springer International Publishing, 2020.

[6] N. Zhang, S. Zhang, J. Zheng, X. Fang, J. W. Mark, and X. Shen, "QoE Driven Decentralized Spectrum Sharing in 5G Networks: Potential Game Approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7797–7808, Sep 2017.

[7] L. C. Bento, R. Parafita, and U. Nunes, "Intelligent traffic management at intersections supported by V2V and V2I communications," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2012, pp. 1495–1502.

[8] S. A. Ahmad, A. Hajisami, H. Krishnan, F. Ahmed-Zaid, and E. Moradi-Pari, "V2V System Congestion Control Validation and Performance," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2102–2110, Mar 2019.

[9] J. C. De Winter, R. Happee, M. H. Martens, and N. A. Stanton, "Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 196–217, Nov 2014.

[10] V. Milanés, E. Onieva, J. Pérez, J. Simó, C. González, and T. De Pedro, "Making transport safer: V2V-based automated emergency braking system," *Transport*, vol. 26, no. 3, pp. 290–302, Sep 2011.

[11] R. S. Schwartz, M. Van Eenennaam, G. Karagiannis, G. Heijenk, W. K. Wolterink, and H. Scholten, "Using V2V communication to create over-the-horizon awareness in multiple-lane highway scenarios," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2010, pp. 998–1005.

[12] F. Jiménez, J. Naranjo, S. Sánchez, F. Serradilla, E. Pérez, M. Hernández, and T. Ruiz, "Communications and Driver Monitoring Aids for Fostering SAE Level-4 Road Vehicles Automation," *Electronics*, vol. 7, no. 10, p. 228, Oct 2018.

[13] IEEE Vehicular Technology Society and Intelligent Transportation Systems Committee, "1609.4 - IEEE Standard for Wireless Access in Vehicular Environments (WAVE) — multi-channel operation, Networking Services IEEE Vehicular Technology Society," Tech. Rep. March, 2010.

[14] ETSI, "ES 202 663 - V1.1.0 - Intelligent Transport Systems (ITS); European profile standard for the physical and medium access control layer of Intelligent Transport Systems operating in the 5 GHz frequency band," *ETSI Standard*, pp. 1–27, 2009.

[15] IEEE Vehicular Technology Society and Intelligent Transportation Systems Committee, "1609.0 - IEEE Guide for Wireless Access in Vehicular Environments (WAVE) — architecture, Networking Services IEEE Vehicular Technology Society," Tech. Rep. March, 2014.

[16] Department for Transport, "Vehicle Licensing Statistics: 2020 Q2," Tech. Rep.

[17] L. E. Ye, D. Chen, Q. Han, L. Zeng, Q. He, F. Liu, and J. Lei, "V2V channel allocation method based on SWM communication demand matrix," *Ad-Hoc and Sensor Wireless Networks*, vol. 41, no. 3-4, pp. 225–246, 2018.

[18] D. Lloyd, "Reported road casualties in Great Britan: annual report 2012," *Statistical Release*, pp. 1–9, 2013. [Online]. Available: https://www.gov.uk/government/

[19] D. D. Clarke, P. Ward, C. Bartle, and W. Truman, "Killer crashes: Fatal road traffic accidents in the UK," *Accident Analysis and Prevention*, vol. 42, no. 2, pp. 764–770, Mar 2010.

[20] IEEE Vehicular Technology Society and Intelligent Transportation Systems Committee, "1609.3 - IEEE Standard for Wireless Access in Vehicular Environments (WAVE)— Networking Services IEEE Vehicular Technology Society," Tech. Rep. December, 2010.

[21] Society of Automotive Engineers, "Dedicated short range communications (DSRC) message set dictionary superseding j2735," Tech. Rep. January, 2016.

[22] H. Boeglen, B. Hilt, P. Lorenz, J. Ledy, A. M. Poussard, and R. Vauzelle, "A survey of V2V channel modeling for VANET simulations," in *2011 8th International Conference on Wireless On-Demand Network Systems and Services, WONS 2011*. IEEE, Jan 2011, pp. 117–123.

[23] Y. J. Li, "An Overview of the DSRC/WAVE Technology," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, vol. 74, pp. 544–558, 2012.

[24] L. Cheng, B. E. Henty, F. Bai, and D. D. Stancil, "Highway and rural propagation channel modeling for vehicle-to-vehicle communications at 5.9 GHz," *2008 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting, APSURSI*, vol. 15213, no. 1, 2008.

[25] A. Paier, J. Karedal, N. Czink, H. Hofstetter, C. Dumard, T. Zemen, F. Tufvesson, A. F. Molisch, and C. F. Mecklenbräuker, "Car-to-car radio channel measurements at 5 GHz: Pathloss, power-delay profile, and delay-Doppler spectrum," *Proceedings of 4th IEEE Internatilonal Symposium on Wireless Communication Systems 2007, ISWCS*, pp. 224–228, 2007.

[26] M. Shafi, J. Zhang, H. Tataria, A. F. Molisch, S. Sun, T. S. Rappaport, F. Tufvesson, S. Wu, and K. Kitao, "Microwave vs. Millimeter-Wave Propagation Channels: Key Differences and Impact on 5G Cellular Systems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 14–20, Dec 2018.

[27] A. F. Molisch, F. Tufvesson, J. Karedal, and C. F. Mecklenbräuker, "A survey on vehicle-to-vehicle propagation channels," *IEEE Wireless Communications*, vol. 16, no. 6, pp. 12–22, 2009.

[28] K. L. Du and M. N. S. Swamy, *Wireless communication systems: From RF subsystems to 4G enabling technologies*, 2010.

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

[30] A. Goldsmith, *Wireless communications.* Cambridge: Cambridge University Press, 2005.

[31] A. Rojas, J. H. Braslavsky, and R. H. Middleton, "Control over a bandwidth limited signal to noise ratio constrained communication channel," in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, 2005, pp. 197–202.

[32] Y.-C. Ko, M. Alouini, and M. K. Simon, "Outage probability of diversity systems over generalized fading channels," *IEEE Transactions on Communications*, vol. 48, no. 11, pp. 1783–1787, 2000.

[33] Y. Hong-Chuan, *Introduction to Digital Wireless Communications.* The Institution of Engineering and Technology, 2017.

[34] F. M. Alsalami, N. Aigoro, A. A. Mahmoud, Z. Ahmad, P. A. Haigh, O. C. Haas, and S. Rajbhandari, "Impact of vehicle headlights radiation pattern on dynamic vehicular vlc channel," *Journal of Lightwave Technology*, vol. 39, pp. 3162–3168, May 2021.

[35] F. M. Alsalami, Z. Ahmad, P. A. Haigh, O. C. Haas, and S. Rajbhandari, "The statistical temporal properties of vehicular visible light communication channel," *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing*, Jul 2020.

[36] T.-H. Do and M. Yoo, "Potentialities and challenges of vlc based outdoor positioning," in *2015 International Conference on Information Networking (ICOIN)*, 2015, pp. 474–477.

[37] J. Kahn and J. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, 1997.

[38] P. Devi and R. K. Maddila, "Modeling of lithium niobate based mach-zehnder modulator for visible light communication system with ber analysis," *Optical and Quantum Electronics*, vol. 53, pp. 1–15, Jun 2021.

[39] P. Haigh, I. Papakonstantinou, F. Arca, S. Tedde, O. Hayden, and E. Leitgeb, "A 1-mb/s visible light communications link with low bandwidth organic components," *IEEE Photonics Technology Letters*, vol. 26, pp. 1–1, Apr 2014.

[40] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using led lights," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 100–107, 2004.

[41] O. Edfors, M. Sandell, J.-J. van de Beek, D. Landstrom, and F. Sjoberg, "An introduction to orthogonal frequency-division multiplexing," 1996.

[42] S. Haykin, *Communication Systems*, 5th ed. Wiley Publishing, 2009.

[43] H. Wei, Y. Huang, T. Zhang, and L. Li, "An universal mmse channel estimator for ofdm receiver," *Wireless Personal Communications*, vol. 94, 2016.

[44] N. Michailow, M. Matthe, I. S. Gaspar, A. N. Caldevilla, L. L. Mendes, A. Festag, and G. Fettweis, "Generalized frequency division multiplexing for 5th generation cellular networks," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3045–3061, 2014.

[45] G. Fettweis, M. Krondorf, and S. Bittner, "GFDM - Generalized Frequency Division Multiplexing," *IEEE 69th Vehicular Technology Conference, Spring 2009.*, pp. 1–4, 2009.

[46] I. Gaspar, L. Mendes, M. Matthé, N. Michailow, D. Zhang, A. Albertiy, and G. Fettweis, "GFDM - A Framework for Virtual PHY Services in 5G Networks," 2015.

[47] G. R. Al-Juboori, A. Doufexi, and A. R. Nix, "System level 5G evaluation of GFDM waveforms in an LTE-A platform," in *Proceedings of the International Symposium on Wireless Communication Systems*. IEEE, Sep 2016, pp. 335–340.

[48] M. Matthe, L. L. Mendes, and G. Fettweis, "Generalized frequency division multiplexing in a gabor transform setting," *IEEE Communications Letters*, vol. 18, pp. 1379–1382, Aug 2014.

[49] M. Sameen, A. A. Khan, I. U. Khan, N. Azim, and I. Shafi, "Comparative analysis of OFDM and GFDM," vol. 4, pp. 267–274, 2016.

[50] N. Michailow, S. Krone, M. Lentmaier, and G. Fettweis, "Bit Error Rate Performance of Generalized Frequency Division Multiplexing," *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sep 2012.

[51] B. Alves, L. Mendes, D. Guimaraes, and I. Gaspar, "Performance of GFDM over Frequency-Selective Channels," *International Workshop on Telecommunications*, vol. 15, 2013.

[52] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[53] R. Datta, N. Michailow, M. Lentmaier, and G. Fettweis, "GFDM interference cancellation for flexible cognitive radio phy design," in *IEEE Vehicular Technology Conference*. IEEE, Sep 2012, pp. 1–5.

[54] M. R. D. Rodrigues and I. Darwazeh, "A Spectrally Efficient Frequency Division Multiplexing Based Communication System," *Proceedings of the 8th International OFDM-Workshop (InOWo'03)*, pp. 70–74, 2003.

[55] F. Luo and C. Zhang, *Signal processing for 5G: Algorithms and implementations*, Jan 2017.

[56] A. B. Kislitsyn, A. V. Rashich, and N. N. Tan, "Generation of SEFDM-Signals Using FFT/IFFT," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer International Publishing, 2014, pp. 488–501.

[57] W. Ozan, P. A. Haigh, B. Tan, and I. Darwazeh, "Experimental SEFDM Pipelined Iterative Detection Architecture with Improved Throughput," in *VTC 2018*. IEEE, Jul 2018, pp. 1–5.

[58] I. D. Kanaras, *Spectrally Efficient Multicarrier Communication Systems: Signal Detection, Mathematical Modelling and Optimisation*. UCL Discovery, 2010.

[59] S. Isam and I. Darwazeh, "Characterizing the intercarrier interference of non-orthogonal Spectrally Efficient FDM system," in *2012 8th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, 2012, pp. 1–5.

[60] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm i. expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2806–2818, 2005.

[61] T. Wang, J. Proakis, and J. Zeidler, "Techniques for suppression of intercarrier interference in OFDM systems," in *IEEE Wireless Communications and Networking Conference, 2005*, vol. 1, 2005, pp. 39–44.

[62] 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.

[63] T. Xu and I. Darwazeh, "Experimental Validations on Self Interference Cancelled Non-Orthogonal SEFDM Signals." IEEE, Jul 2018, pp. 1–5.

[64] Z. Chen and B. Liu, "Lifelong Machine Learning," pp. 1–27, Nov 2016.

[65] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa, and I. Yaqoob, "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.

[66] Y. Le Cun, "Une procedure d'apprentissage pour reseau a seuil assymetrique (A learning procedure for assymetric threshold networks)," in *Proceedings of Cognitiva 85*, 1985, pp. 599–604.

[67] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[68] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points-online stochastic gradient for tensor decomposition," vol. 40. JMLR: Workshop and Conference Proceedings, 2015, pp. 1–46.

[69] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," in *Proceedings of COMPSTAT*. Physica-Verlag HD, 2010, pp. 177–186.

[70] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," 2018.

[71] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," in *COLT 2010 - The 23rd Conference on Learning Theory*, vol. 12, no. 61, 2010, pp. 257–269.

[72] N. Cesa-Bianchi, A. Conconi, and C. Gentile, "On the generalization ability of on-line learning algorithms," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2050–2057, Sep 2004.

[73] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, Nov 2016.

[74] J. Gama and P. Brazdil, "Cascade Generalization," *Machine Learning*, vol. 41, no. 3, pp. 315–343, 2000.

[75] T. A. Eriksson, H. Bulow, and A. Leven, "Applying Neural Networks in Optical Communication Systems: Possible Pitfalls," *IEEE Photonics Technology Letters*, vol. 29, no. 23, pp. 2091–2094, Dec 2017.

[76] L. Cayton, "Algorithms for manifold learning," Tech. Rep., 2005.

[77] H. Narayanan and S. Mitter, "Sample complexity of testing the manifold hypothesis," in *Advances in Neural Information Processing Systems 23*. Curran Associates, Inc., 2010, pp. 1786–1794.

[78] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, Aug 1987.

[79] L. Van Der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2625, 2008.

[80] L. V. D. Maaten, "Learning a parametric embedding by preserving local structure," 2009.

[81] G. Hinton and S. Roweis, "Stochastic neighbor embedding," 2002.

[82] G. Wahba, X. Lin, F. Gao, D. Xiang, R. Klein, and B. Klein, "The bias-variance tradeoff and the randomized GACV," in *Advances in Neural Information Processing Systems*, 1999, pp. 620–626.

[83] C. F. Mecklenbraüker, A. F. Molisch, J. Karedal, F. Tufvesson, A. Paier, L. Bernadó, T. Zemen, O. Klemp, and N. Czink, "Vehicular channel characterization and its implications for wireless system design and performance," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1189–1212, 2011.

[84] Y. Kaymak, S. Fathi-Kazerooni, and R. Rojas-Cessa, "Indirect Diffused Light Free-Space Optical Communications for Vehicular Networks," *IEEE Communications Letters*, vol. 23, no. 5, pp. 814–817, May 2019.

[85] R. Alieiev, T. Hehn, A. Kwoczek, and T. Kurner, "Predictive Communication and Its Application to Vehicular Environments: Doppler-Shift Compensation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7380–7393, Aug 2018.

[86] T. Izydorczyk, F. M. Tavares, G. Berardinelli, M. Bucur, and P. Mogensen, "Performance evaluation of multi-antenna receivers for vehicular communications in live LTE networks," in *IEEE Vehicular Technology Conference*. Institute of Electrical and Electronics Engineers Inc., Apr 2019.

[87] E. Ahmed and H. Gharavi, "Cooperative Vehicular Networking: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 996–1014, Mar 2018.

[88] S. A. A. Shah, E. Ahmed, M. Imran, and S. Zeadally, "5G for Vehicular Communications," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 111–117, Jan 2018.

[89] M. A. Karabulut, A. S. Shah, and H. Ilhan, "Performance Optimization by Using Artificial Neural Network Algorithms in VANETs," in *42nd TSP 2019*. IEEE, 2019, pp. 633–636.

[90] G. H. Sim, S. Klos, A. Asadi, A. Klein, and M. Hollick, "An online context-aware machine learning algorithm for 5g mmwave vehicular communications," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2487–2500, Dec 2018.

[91] J. Joo, M. C. Park, D. S. Han, and V. Pejovic, "Deep Learning-Based Channel Prediction in Realistic Vehicular Communications," *IEEE Access*, vol. 7, pp. 27 846–27 858, 2019.

[92] W. Ozan, H. Ghannam, T. Xu, P. A. Haigh, and I. Darwazeh, "Experimental Evaluation of Channel Estimation and Equalisation in Non-Orthogonal FDM Systems," in *11th CSNDSP*, July 2018, pp. 1–6.

[93] K. Burse, R. N. Yadav, and S. Shrivastava, "Channel equalization using neural networks: A review," *IEEE transactions on systems, man, and cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 3, pp. 352–357, 2010.

[94] T. Blazek, M. Ashury, C. F. Mecklenbrauker, D. Smely, and G. Ghiaasi, "Vehicular channel models: A system level performance analysis of tapped delay line models," in *15th ITST*. IEEE, May 2017, pp. 1–8.

[95] A. Paier, L. Bernadó, J. Karedal, O. Klemp, and A. Kwoczek, "Overview of vehicle-to-vehicle radio channel measurements for collision avoidance applications," *VTC 2010*, pp. 4–8, 2010.

[96] A. Zerguine, A. Shafi, and M. Bettayeb, "Multilayer perceptron-based dfe with lattice structure," *IEEE transactions on neural networks*, vol. 12, no. 3, pp. 532–545, 2001.

[97] A. Gorban and D. Wunsch, "The general approximation theorem." Institute of Electrical and Electronics Engineers (IEEE), Nov 1998, pp. 1271–1274.

[98] *Introduction to OFDM*. John Wiley Sons, Ltd, 2010, ch. 4, pp. 111–151.

[99] M. Sternad and D. Aronsson, "Channel estimation and prediction for adaptive OFDM downlinks," in *IEEE Vehicular Technology Conference*, vol. 58, no. 2. IEEE, 2003, pp. 1283–1287.

[100] C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep Complex Networks," Tech. Rep., 2017.

[101] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural Network Studies. 1. Comparison of Overfitting and Overtraining," *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 5, pp. 826–833, 1995.

[102] V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 2010, pp. 807–814.

[103] I. Arnekvist, J. F. Carvalho, D. Kragic, and J. A. Stork, "The effect of Target Normalization and Momentum on Dying ReLU," May 2020.

[104] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[105] R. Sutton, "Two problems with backpropagation and other steepest-descent learning procedures for networks," 1986.

[106] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," Dec 2012.

[107] T. Dozat, "Incorporating Nesterov Momentum into Adam," *ICLR Workshop*, no. 1, pp. 2013–2016, 2016.

[108] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, Dec 2015.

[109] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2017, p. 6629–6640.

[110] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$,"1983.

[111] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu, "Advances in optimizing recurrent networks," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Oct 2013, pp. 8624–8628.

[112] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *30th ICML 2013*, no. PART 3, 2013, pp. 2176–2184.

[113] A. Hernandez, "Model Calibration with Neural Networks," *SSRN Electronic Journal*, 2018.

[114] E. Balevi and J. G. Andrews, "One-Bit OFDM Receivers via Deep Learning," in *IEEE Transactions on Communications*, vol. 67, no. 6. Institute of Electrical and Electronics Engineers Inc., Jun 2019, pp. 4326–4336.

[115] W. Schmidt, S. Raudys, M. Kraaijveld, M. Skurikhina, and R. Duin, "Initializations, back-propagation and generalization of feed-forward classifiers," in *IEEE International Conference on Neural Networks*. IEEE, pp. 598–604.

[116] S. T. Le, K. J. Blow, V. K. Menzentsev, and S. K. Turitsyn, "Comparison of numerical bit error rate estimation methods in 112Gbs QPSK CO-OFDM transmission," in *IET Conference Publications*, vol. 2013, no. 622 CP, 2013, pp. 1083–1085.

[117] S. Rajbhandari, J. Faith, Z. Ghassemlooy, and M. Angelova, "Comparative study of classifiers to mitigate intersymbol interference in diffuse indoor optical wireless communication links," *Optik*, vol. 124, no. 20, pp. 4192–4196, 2013.

[118] Z. Zhou, L. Liu, V. Chandrasekhar, J. Zhang, and Y. Yi, "Deep Reservoir Computing Meets 5G MIMO-OFDM Systems in Symbol Detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 1266–1273, Apr 2020.

[119] R. Hassun, M. Flaherty, R. Matreci, and M. Taylor, "Effective evaluation of link quality using error vector magnitude techniques," in *Annual Wireless Communications Conference, Proceedings*. IEEE, 1997, pp. 89–94.

[120] R. Schmogrow, B. Nebendahl, M. Winter, A. Josten, D. Hillerkuss, S. Koenig, J. Meyer, M. Dreschmann, M. Huebner, C. Koos, J. Becker, W. Freude, and J. Leuthold, "Error vector magnitude as a performance measure for advanced modulation formats," *IEEE Photonics Technology Letters*, vol. 24, no. 1, pp. 61–63, Jan 2012.

[121] S.-S. Yang, C.-L. Ho, and S. Siu, "Sensitivity analysis of the split-complex valued multilayer perceptron due to the errors of the iid inputs and weights," *IEEE transactions on neural networks*, vol. 18, no. 5, pp. 1280–1293, 2007.

[122] T. Xu and I. Darwazeh, "Design and Prototyping of Neural Network Compression for Non-Orthogonal IoT Signals," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[123] IEEE Standardization, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications; High-speed Physical Layer in the 5 GHz Band," 1999.

[124] 3GPP Standardization, "3GPP Technical Specification 36.101 V10.2.1 (2011–04). Document: 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 10)," 2011.

[125] 3GPP Standardization, "ETSI TS 125.101 V10.5.0 (2012–03). Document: Universal Mobile Telecommunications Systems(UMTS); User Equipment (UE) radio transmission and reception (FDD) (Release 10)," 2012.

[126] 3GPP Standardization, "TS 145 005 - V6.9.0 - Digital cellular telecommunications system (Phase 2+); Radio transmission and reception (3GPP TS 45.005 version 6.9.0 Release 6)," pp. 0–130, 2005.

[127] J. L. Pinto and I. Darwazeh, "Error vector magnitude relation to magnitude and phase distortion in 8-PSK systems," *Electronics Letters*, vol. 37, no. 7, pp. 437–438, Mar 2001.

[128] R. A. Shafik, M. S. Rahman, and A. H. Islam, "On the extended relationships among EVM, BER and SNR as performance metrics," in *Proceedings of 4th International Conference on Electrical and Computer Engineering*. IEEE Computer Society, 2006, pp. 408–411.

[129] L. Hanzo, W. T. Webb, and T. Keller, "Single-and Multi-carrier Quadrature Amplitude Modulation Principles and Applications for Personal Communications, WLANs and Broadcasting," pp. 401–423, 2010.

[130] K. Cho and D. Yoon, "On the general ber expression of one- and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. 50, no. 7, pp. 1074–1080, 2002.

[131] J. G. Proakis, *Digital communications*, 5th ed. New York: McGraw-Hill Higher Education, 2008.

[132] T. Chen and H. Chen, "Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical Systems," *IEEE Transactions on Neural Networks*, vol. 6, no. 4, pp. 911–917, 1995.

[133] P. Chvojka, A. Burton, P. Pesek, X. Li, Z. Ghassemlooy, S. Zvanovec, and P. Anthony Haigh, "Visible light communications: increasing data rates with polarization division multiplexing," *Optics Letters*, vol. 45, no. 11, p. 2977, Jun 2020.

[134] P. Chvojka, A. Burton, P. Pesek, X. Li, Z. Ghassemlooy, S. Zvanovec, and P. A. Haigh, "Visible light communications: increasing data rates with polarization division multiplexing," *Optics Letters*, vol. 45, no. 11, pp. 2977–2980, 2020.

[135] P. A. Haigh, *Visible Light*. IOP Publishing Ltd, 2020.

[136] T. Xu, T. Xu, and I. Darwazeh, "Deep Learning for Interference Cancellation in Non-Orthogonal Signal Based Optical Communication Systems," in *Progress in Electromagnetics Research Symposium*, 2018, pp. 241–248.

[137] H. He, C. K. Wen, S. Jin, and G. Y. Li, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, Oct 2018.

[138] W. Ozan, P. A. Haigh, B. Tan, and I. Darwazeh, "Time precoding enabled Non-Orthogonal frequency division multiplexing," in *30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep 2019.

[139] H. Ghannam and I. Darwazeh, "SEFDM: Spectral Efficiency Upper Bound and Interference Distribution," in *2018 11th International Symposium on Communication Systems, Networks and Digital Signal Processing*, 2018.

[140] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *IEEE Vehicular Technology Conference*, vol. 3, 1997, pp. 1639–1643.

[141] S. Isam and I. Darwazeh, "Simple DSP-IDFT techniques for generating spectrally efficient FDM signals," in *2010 7th International Symposium on Communication Systems, Networks and Digital Signal Processing*. IEEE Computer Society, 2010, pp. 20–24.

[142] 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 and Digital Signal Processing*. IEEE, Jul 2018.

[143] R. Jothikumar, "Reduced complexity analysis for ML MIMO systems," in *2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology*, 2013, pp. 390–393.

[144] H. S. M. Coxeter, J. H. Conway, and N. J. A. Sloane, "Sphere Packings, Lattices and Groups." *The American Mathematical Monthly*, vol. 96, no. 6, p. 538, 1989.

[145] C. Y. Hung and T. H. Sang, "A sphere decoding algorithm for MIMO channels," in *Sixth IEEE International Symposium on Signal Processing and Information Technology, ISSPIT*, 2006, pp. 502–506.

[146] P. N. Whatmough, M. R. Perrett, S. Isam, and I. Darwazeh, "VLSI architecture for a reconfigurable spectrally efficient FDM baseband transmitter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 5, pp. 1107–1118, 2012.

[147] W. Ozan, R. Grammenos, and I. Darwazeh, "Zero Padding or Cyclic Prefix: Evaluation for Non-Orthogonal Signals," *IEEE Communications Letters*, vol. 24, no. 3, pp. 690–694, 2020.

[148] A. Zerguine, A. Shafi, and M. Bettayeb, "Multilayer perception-based DFE with lattice structure," *IEEE Transactions on Neural Networks*, vol. 12, no. 3, pp. 532–545, May 2001.

[149] W. F. Schmidt, S. Raudys, M. A. Kraaijveld, M. Skurikhina, and R. P. Duin, "Initializations, back-propagation and generalization of feed-forward classifiers," in *1993 IEEE International Conference on Neural Networks*. Publ by IEEE, 1993, pp. 598–604.

[150] 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," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019, pp. 2758–2768.

[151] P. G. Pachpande, M. H. Khadr, H. Hussien, H. Elgala, and D. Saha, "Autoencoder Model for OFDM-based Optical Wireless Communication," in *OSA Advanced Photonics Congress (AP) 2019 (IPR, Networks, NOMA, SPPCom, PVLED)*. Optical Society of America, 2019.

[152] R. Mennes, M. Camelo, M. Claeys, and S. Latre, "A neural-network-based MF-TDMA MAC scheduler for collaborative wireless networks," in *IEEE Wireless Communications and Networking Conference*. Institute of Electrical and Electronics Engineers Inc., Jun 2018, pp. 1–6.