THE UNIVERSITY
of EDINBURGH

Channel Prediction in Wireless Communications

Alan John Anderson

A thesis submitted for the degree of Doctor of Philosophy

Institute for Digital Communications
University of Edinburgh

January 2015

Abstract

Knowledge of the channel over which signals are sent is of prime importance in modern wireless communications. Inaccurate or incomplete channel information leads to high error rates and wasted bandwidth and energy. Although active channel measurement is commonly used to gain channel knowledge, it can only accurately represent the channel at the time the measurement was taken, makes energy and bandwidth demands, and adds significant complexity to the radio system. Due to the highly time variant nature of wireless channels, active measurements become invalid almost as soon as they are taken, making alternative approaches to predicting future behaviour highly attractive. Such systems would allow maximum advantage to be taken of the limited bandwidth available and make significant power savings. This thesis investigates a number of complementary technologies, leading towards a channel prediction scheme suitable for mobile devices.

As a first step towards channel prediction, anomaly detection is investigated within periodic wireless signals to establish when radical changes in the channel occur. In previous experiments, long monotonic sequences had been observed to coincide with certain anomalies but not others when using Kullback-Leibler Divergence (KLD) analysis, possibly allowing the characterisation of anomaly types. An investigation is described to explain the origin of these features in a rigorous mathematical sense. A proof is given for the causes of the monotonic sequences, followed by a discussion of the types of signal anomaly which would underly such a feature and the value of this information.

The second part describes a novel channel characterisation method which uses a class of Recurrent Neural Network (RNN) called an Echo State Network (ESN). Using this tool, a channel characterisation system can be constructed without an explicit statistical or mathematical model of the wireless environment, relying instead on observed data. This approach is much more convenient than existing models which require detailed information about the wireless system's parameters and also allows for new channel classifications to be added easily. It is able to achieve double the correct classification rate of a conventional statistical classifier, and is computationally simple to implement, making it ideal for inclusion on low-power mobile devices.

Following their successful use in characterisation, ESNs are used in the final part in an investigation into channel prediction in a number of different scenarios. They were however found to be unable to produce useful predictions for all but the most trivial channel models. An alternative method is described for indoor environments using an approach inspired by ray tracing. It is simple and computationally lightweight to implement, again making it suitable for mobile devices. Simulation results show that it can outperform pilot-assisted methods by a significant margin, while not wasting bandwidth on channel measurement.

Lay Summary

Over the last two decades it has become an expectation that wireless communications will have ever increasing speeds, and be available in all places at all times. This consumer demand for mobile internet connected devices has put a great deal of pressure to develop more efficient ways of using the limited amount of radio spectrum for such devices. In order to best make use of the available bandwidth, information must be known about the channel over which the wireless communication travels. Existing methods to measure the channel do produce reasonably useful results, but waste both time and energy in performing these measurements.

This thesis examines a variety of different methods to characterise and predict channel conditions, in order to allow the best possible use to be made of the channel. Information theoretic methods are used to characterise channels, and a type of neural network known as an echo state network is used to both characterise and predict some channels. Finally, a new prediction method is developed which is related to the existing method of ray-tracing.

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself were composed and originated entirely by myself in the Institute for Digital Communications of the School of Engineering at The University of Edinburgh.

Alan Anderson

Edinburgh, UK

January 2015

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Harald Haas for being a constant inspiration to me, and for his guidance and attention to detail throughout my studies. His enthusiasm and constant stream of suggestions have made this work so much richer. His close involvement from the very early stages of this work has been invaluable, and from him I have learned priceless lessons. I am also grateful to Dave Laurenson for his assistance to me in his role as my second supervisor.

I would also like to acknowledge the generous financial and practical support provided by Agilent Technologies. In addition, Lance Tatman has supported this work by providing equipment and taking an active interest at every step. His faith in my research has been unwavering.

Thanks to all my colleagues at the Institute for Digital Communications who have provided many lively debates and discussions, as well as technical advice and input.

Last but not least, I would like to offer my heartfelt thanks to my family. Thanks go to my parents and parents-in-law, whose love and support have meant so much over the years. To my wonderful wife Esmé, thank you for putting up with my late nights and early mornings, for your practical help, for your steady encouragement and for helping me always to see the bigger picture. Your patience, understanding and unfailing kindness are worth far, far more to me than I can could ever express.

Contents

List of Figures

List of Tables

Acronyms

AR Autoregessive.

AWGN Additive White Gaussian Noise.

BER Bit Error Rate.

BPSK Binary Phase-Shift Keying.

BPTT Backpropagation Through Time.

CoMP Coordinated multipoint.

CR Cognitive Radio.

DECT Digital Enhanced Cordless Telecommunications.

ESN Echo State Network.

ESPRIT Estimation of Signal Parameters via Rotational Invariant Techniques.

FCC Federal Communication Commission.

FPGA Field Programmable Gate Array.

FSA Fixed Spectrum Access.

FSPL Free Space Path Loss.

GNSS Global Navigation Satellite Systems.

GPS Global Positioning System.

GSM Global System for Mobile Communications.

GSMA GSM Alliance.

ISM Industrial, Scientific and Medical.

KLD Kullback-Leibler Divergence.

LOS Line of Sight.

LTE Long Term Evolution.

MATLAB Matrix Laboratory.

MI Mutual Information.

MIMO Multiple Input, Multiple Output.

MLP Multilayer Perceptron.

MSE Mean-Squared Error.

MUSIC Multiple Signal Classification.

NLOS Non-Line of Sight.

NN Neural Network.

Ofcom Office of Communications.

OFDM Orthogonal Frequency Division Multiplexing.

OFDMA Orthogonal Frequency Division Multiple Access.

PDF Probability Density Function.

PMF Probability Mass Function.

QAM Quadrature Amplitude Modulation.

RADAR Radio Distance and Ranging.

RF Radio Frequency.

RNN Recurrent Neural Network.

SC-FDMA Single-Carrier Frequency Division Multiple Access.

SNR Signal to Noise Ratio.

SOS Sum of Sinusoids.

UMTS Universal Mobile Telecommunications System.

WiMAX Worldwide Interoperability for Microwave Access.

WINNER Wireless World Initiative New Radio.

WSN Wireless Sensor Network.

Chapter 1

Introduction

This chapter begins by outlining the real-world demands for improved wireless communication technology to make the best possible use of the available bandwidth. A broad discussion of the issues related to obtaining improved channel information is provided, followed by the aims, objectives and methodologies of this research. Next, the unique research contributions contained in this thesis are laid out. This is followed by an overview of each of the chapters in the thesis.

1.1 Motivation

The rapid and global adoption of wireless systems for digital communication, from Bluetooth, WiFi and Digital Enhanced Cordless Telecommunications (DECT) for short range, to a range of cellular technologies (e.g. Global System for Mobile Communications (GSM), Universal Mobile Telecommunications System (UMTS) and Long Term Evolution (LTE), etc.) for medium to long range, is clear evidence that users find these systems highly satisfactory, even indispensable. From an engineering perspective, however, much remains to be done to increase the speed and reliability of signal transmission and to reduce the required energy. Current approaches to handling wireless communications have ample scope for improvement, often falling a long way short of the theoretical upper performance limits imposed by finite bandwidth and unavoidable noise. The consumer-led exponential rise in data traffic over such networks is creating a seemingly insatiable demand for ever higher speed links. According to the GSM Alliance (GSMA), between 2012 and 2017 in Europe alone, the monthly volume of data transmitted over cellular networks will increase from 0.18EB (Exabyte) to over 1.3EB

per month [48] - an astonishing sevenfold increase.

In order to meet this demand for increasing data usage, new approaches must be developed to optimally exploit the limited radio spectrum. Historically, the radio spectrum has been centrally controlled (in the UK by the Office of Communications (Ofcom) and the Federal Communication Commission (FCC) in the United States) using a Fixed Spectrum Access (FSA) policy and divided up into frequency bands dedicated to one particular purpose such as television broadcasting and marine or military radio. While this is an effective approach at guaranteeing a particular level of service for the licensed users, it means that when the main user is not transmitting, the bandwidth allocated to them is wasted. A range of new methods proposed in 1998 by Mitola [78] known as Cognitive Radio (CR) are being developed to make radios 'smarter', allowing them to use the spare capacity in existing radio bands in their locality. These radios could dynamically sense or predict, and 'jump in' and use bands currently sitting idle [68]. This adaptive approach has the potential to vastly expand the bandwidth available to the increasing number of interested users, however it is of prime importance that the new systems do not in any way interfere detrimentally with the existing ones. To that end, novel technologies are required in the areas of anomaly detection, interference management, spectrum sensing and channel prediction in order to coexist with legacy radio systems.

As well as using idle spectrum to boost wireless system throughput, the existing spectrum must be used as efficiently as possible. Paramount to achieving this is channel knowledge: if a system has incomplete or inaccurate channel knowledge, its ability to compensate for the distortion effects imposed on the transmitted signal will be limited. The more accurate the information about the channel is, the more faithfully the originally transmitted signal can be reconstructed. This leads to fewer errors, and the possibility of using high-order modulation techniques (such as 256-Quadrature Amplitude Modulation (QAM)) which can carry larger amounts of data then lower-order techniques (such as 16-QAM or Binary Phase-Shift Keying (BPSK)). Channel knowledge is also of vital importance to Multiple Input, Multiple Output (MIMO) systems, as their ability to accurately recover the transmitted data depends on being able to distinguish between distinct channels.

Although any given channel may be considered to be unique to a particular time, location and system, there are certain features which will be similar among channels in similar situations. For example, channels between fixed base stations and fast-moving

users will have higher variability than channels between fixed points. While this is a long way short of complete channel knowledge, it can be useful nonetheless. This task of characterising channels according to their behaviour is vital to adaptive radio systems, and CR in particular. If a radio can determine some information about its environment, it can adjust its communication parameters (such as data rate, power and time scheduling algorithms) to best exploit the channel's characteristics, and this is often done using active channel measurement. By sending a signal (referred to as a pilot signal) with parameters (phase, magnitude and duration) which are known *a-priori* to the receiver, an accurate measurement of the channel can be arrived at by comparing the received signal which has been passed through the channel with the ideal signal which was known to have been transmitted.

The use of active channel measurement in radio systems has become an accepted inconvenience. By embedding them in transmissions, pilot signals allow system designers to obtain reasonably accurate channel estimation data at the expense of a percentage of their system's time and energy budget, limiting the maximum possible throughput since these pilot signals consume both time and bandwidth, yet convey no user data. However, the cost associated with channel measurements may not stop here; to make optimal use of these measurements for adaptive modulation and coding, they must be communicated back to the transmitter in order to influence future transmissions. This additional signalling further increases power and time usage, without transmitting any user data.

This setup is further complicated by the time-variant nature of the channel. If the receiver delays beyond a certain length of time in sending data back to the transmitter in response to the pilot signals, the channel information may expire before it can be used. As customer demand pushes systems towards higher and higher bandwidths, this problem is further exacerbated, since higher bandwidth channels require more channel information.

As is clear from the preceding description, this system of measurement and feedback incurs significant cost to the radio system, in time and energy wasted, in bandwidth expended on repeated measurements and in added complexity. In some conditions, these overheads can even outweigh any potential improvement in throughput that can be derived from the added channel knowledge. A clear example of this effect has been seen in recent years with the push to include Coordinated multipoint (CoMP) into the LTE standard. CoMP aims to increase throughput by having multiple base stations

transmit the same data to a given user, allowing users near cell-boundaries to benefit from a stronger signal from several cooperating base stations. In order to avoid the base stations' transmissions simply cancelling each other out, very accurate channel information needs to be communicated to each of the associated transmitters. Doing so requires a large amount of signalling, and due to the great complexity of the system, its inclusion into the LTE standard has been delayed by 5 years [101, 37].

If a perfect channel prediction algorithm were ever developed (even if it could only predict a fraction of a second into the future) the benefits would be huge. The need for active channel measurements would be reduced or eliminated, the channel's bandwidth could be completely available for data, and the power requirements for transmitters would be reduced significantly. Although current state of the art transmission systems fall far short of their maximum possible throughput (if full channel knowledge were available), even modest improvements in channel prediction methods have the potential to provide large benefits to users.

By reducing dependence on active channel measurement (which wastes a small percentage of a radio's link budget), such a system would not only save energy on pilot signals and regain some of its lost bandwidth, but the significant overheads introduced by feedback signalling could be much reduced. Other benefits could include power saving from lower processing demands, leading to longer battery life in a mobile device, and a more efficient, environmentally friendly solution at a base station.

1.2 Research Aims and Objectives

The central aims of this study are to explore a variety of techniques designed to improve throughput in wireless systems by gaining a better understanding of the channel characteristics and reducing the need for wasteful active channel measurement. The most promising approaches will be extended and developed, and their potential for improved wireless communication will be assessed.

The separate research objectives identified to achieve the aims are to investigate a number of approaches in channel characterisation and prediction, and for each approach:

- An in-depth discussion will be provided to clarify in what ways it may contribute to increasing knowledge of the channel.
- Simulation-based experiments will be undertaken to understand its strengths,

weaknesses and potential value.

- Based on insights gained, the most promising approach(es) will be identified, and methods to evaluate and enhance its performance will be developed.

Previous work on wireless channel prediction has met with varying degrees of success. However, no one approach stands out as the best, faced as they are with the major challenges posed by a highly dynamic environment, rapidly varying signal quality and noise, and the need for realtime computational performance. Building on the work of others, the focus of this thesis is to improve throughput by reducing reliance on active channel measurement and to enhance wireless channel prediction.

1.3 Research Contributions

Building on a framework laid by Afgani *et al.*, an explanation is presented in Chapter 3 for long monotonically increasing sequences in the Kullback-Leibler Divergence (KLD) when it is used as an anomaly detection algorithm in periodic wireless signals. A mathematical proof is shown which demonstrates the kind of anomalous events which cause these sequences. The finding is that such sequences can be caused by signals whose information content experiences a significant drop (i.e. the signal becomes more deterministic). While this is clearly atypical in a communications system (where the purpose is to communicate information) and indicative of an event outside normal expectations, this is not a sufficiently narrow criterion to enable it to be used to characterise the underlying cause of the anomaly. This contribution has further validated the KLD as an efficient anomaly detection algorithm for use in periodic wireless systems.

The Echo State Network (ESN) approach is used in Chapter 4 to characterise an unknown wireless channel according to its similarity to previously observed channels. The system's primary advantage is that it is entirely data-driven, meaning that it can be trained entirely on real, observed data, and does not have any need for an explicit system model. This makes it a very attractive approach as defining model parameters is complex and error-prone, whereas gathering training data is often very easy. The system is also capable of including new channel classifications simply by adding appropriate training data. Experiments show that the Neural Network (NN)is able to achieve double the accuracy of a statistical, data-driven classification method, illustrating the power of the ESN to infer relationships within the data, a task which other systems can not

accomplish. As ESNs are computationally much simpler (especially during the training phase) than other NNs, this method is ideal for inclusion in low-power electronics such as mobile phones.

This ESN-based method has the potential to improve everyday communications, as it provides a degree of channel information (i.e. a characterisation of the type of channel in use) with zero signalling overhead. As an example, the characterisation system could be run on an LTE device to help make decisions about modulation, coding, and scheduling. However, the method developed here is not tied to any particular radio system (be it LTE, WIFi or a new standard), as it deals only with the channel, not the higher level protocols, potentially allowing it to be implemented in a variety of radio systems. Although the ESN approach has many positive aspects, a number of promising research directions were also identified to refine the setup (primarily improving the ability to generalise channel scenarios). With some further research, the ESN method has potential to make a meaningful impact on practical radio systems.

Following the successful use of ESNs in the wireless characterisation setting, they are subsequently investigated in Chapter 5 with respect to their ability to predict future channel conditions. Experiments showed that although unsuitable for complex channel models (such as those used to generate data for the characterisation system), ESNs are able to predict the future channel magnitude gain values of some simple indoor channel models. Beyond a certain level of complexity, the network however becomes unable to capture the complexity inherent in the signal, and is unable to generate a useful prediction.

Lastly a channel prediction scheme based on ray-tracing concepts is presented. When given information about a number of rays used to represent a channel, the scheme can make accurate predictions along a straight-line trajectory, designed to model a moving device's channel. Simulations show that it can out-perform conventional Orthogonal Frequency Division Multiplexing (OFDM) transmission strategies in indoor scenarios, while having no overheads from active channel measurement. These savings translate directly into either increased throughput or reduced energy consumption.

By performing channel prediction using ray interpolation, the way is opened to much higher quality channel information which is a vital enabler of a host of technologies. The method developed here is noteworthy because of its low computational demands, making it ideal for low power and mobile devices. By reducing reliance on active channel measurement, some energy and bandwidth can be saved. However with accurate channel

predictions, the need for a feedback loop (and all the associated signalling it entails) is either reduced or eliminated, allowing these now redundant transmission slots to be better employed. With the added channel information, radio systems are able to make more intelligent decisions about resource allocation, coding, scheduling, power usage and throughput. The radio link can then be optimised for maximum data speeds, fairness amongst users or energy efficiency, depending on the particular scenario. Although there are outstanding implementation challenges (ray estimation and noise immunity) these are challenging but not impossible.

1.4 Thesis Outline

The rest of this thesis is organised as follows. Chapter 2 provides relevant background information to prepare the reader for the technical chapters. A review of anomaly detection methodologies is provided along with their applications in the wireless communications domain. Channel modelling, characterisation and prediction are covered, followed by neural networks and finishing with a brief overview of LTE and how accurate channel information is vital to its success.

Due to evolving insights gained as this research progressed, this thesis reports on three strongly related lines of research, presented in Chapters 3, 4 and 5 respectively, all of which are related to wireless signal analysis and modelling, leading towards channel prediction. Chapter 3 discusses the use of information theoretic methods for anomaly detection in periodic wireless signals. It begins with an overview of work by Afgani *et al.* which forms the starting point of the investigation [4, 5, 7, 8]. It then focuses on monotonic increases in the KLD output which have been observed in real-world test signals. A mathematical proof for this behaviour of the KLD algorithm is derived, and a discussion of the possible causes and implications of these sequences is presented. While this work does yield some insights and mathematical proofs, it is not evident that it can reliably characterise anomalies with different causes.

In light of the apparent limitations of using information theoretic methods to characterise wireless signals, Chapter 4 addresses the problem of characterising wireless channels using a recently developed class of neural networks known as ESNs. This approach is attractive as it is data-driven, not requiring the explicit construction of a mathematical model, using only observed data to make its characterisations. Experiments showing the effectiveness of this approach are contrasted with a more conventional statistical tool

and the advantages and drawbacks of these alternative methods are covered.

After some success using ESNs in characterisation work, Chapter 5 begins by investigating their suitability as a channel prediction tool. They are found to be capable of performing some channel prediction on simple channel models, but are unable to cope with more complex scenarios. With this result in mind, a new approach is presented in the area of channel prediction. It attempts to predict the channel characteristics based on modelling a known environmental context. A novel linear framework for indoor channel prediction is first presented and is then enhanced with higher order polynomial interpolation which further increases accuracy. Simulation techniques are used to demonstrate that this method has considerable potential to predict channel behaviour and reduce the bit error rate without the need for repeated channel measurement overheads.

Finally, Chapter 6 concludes the thesis with a review of the aims and objectives outlined in Section 1.2, summarising the main findings of this research. Limitations are discussed, and suggestions for future research are presented.

Copies of the academic papers published during this research are included in the appendices, as well as a selection of source code listings and details of the custom software packages used during experiments.

Chapter 2

Background

This chapter provides supporting background material for the following technical chapters, and is broken into four sections. In Section 2.1 the area of anomaly detection is presented, laying out some historical and current approaches and the wide range of application areas is discussed. Section 2.2 discusses neural networks, their uses and the specific type later employed. The third section introduces wireless channels and the related topics of modelling, characterisation and prediction (relevant to Chapters 4 and 5). Finally Section 2.4 presents an overview of some of the challenges a practical radio system faces and describes selected parts of the Long Term Evolution (LTE) standard.

2.1 Anomaly Detection

Anomaly detection (also sometimes referred to as outlier or novelty detection) is a broad field with a huge range of applications, including computer network intrusion detection [87, 67], banking fraud discovery [65], identification of cancerous tumours from medical scans [23] and even radio astronomy [52]. Its basic assumption is that a system has data or a state which can be considered as 'normal', and that anything not conforming to this definition is in some way of interest. Sometimes the detection method is able to rely on having a well-defined description of normal data. However in many complex scenarios such a description many not be available. Nonetheless, it may be possible to classify data or performance parameters which are outside the normal range by analysing past behaviour.

Since the field is so broad (as evidenced by the large number of 'survey' style papers [23, 54, 76, 75, 67, 14, 123]) with a large number of approaches taken (many of

which are specific to one particular application), making comparisons among different algorithms is difficult. Although the particular method used in a given system will be specifically tailored to the type of data expected, it is possible to group anomaly detection approaches into three broad categories [87], as described below.

Data mining

The range of approaches grouped into the 'data mining' field (sometimes also referred to as Knowledge Discovery) contains a wide variety of technical approaches which all have one common feature - they analyse large amounts of historical data to identify patterns which may not otherwise become apparent, even to a trained human observer. Once such patterns have been discovered, they can be used to characterise what normal or anomalous behaviour looks like for a particular system [14]. Given the complex and extensive amounts of data typically involved, there may be quite a large 'grey area' separating what can clearly be classified as anomalous or normal. Therefore in some cases data mining may be used to create a system model which can be combined with a statistical approach. This allows a detection and classification scheme to be defined that can be used in assessing the probability that a particular anomaly has been detected.

Machine learning

Machine learning approaches use self-learning methods to generate a profile of observed activity [75]. Once sufficient learning has been achieved, the system will attempt to identify anomalies in the data it processes. This uses the basic assumption underlying all anomaly detection, that outliers will be rare. In addition, some machine learning systems will attempt to improve the accuracy and reliability of anomaly discovery as they analyse more data. For example, a newly implemented system might not correctly recognise a particular anomaly, however given more time and observed data, the same system might later be able to correctly identify the data as anomalous. Another benefit that machine learning systems have over other types is that they can change their definition of what counts as anomalous behaviour as patterns change. While at one time a particular type of event may have appeared very infrequently (and perhaps have been correctly classified as anomalous), if such events started occurring more frequently, a machine learning system could adjust its detection criteria and stop classifying these events as anomalous, flagging other types as of interest.

Statistical

Purely statistical approaches assume some model of the expected data, and then compare the observed statistics with the model [76], presuming it to be accurate. Such a system might assume its data should be normally distributed with a certain variance, and observed values not conforming to this model are assumed to be anomalous [39]. An obvious advantage to such a scheme is its simplicity - any data point (or set of points) not included in the set defined by the statistical model is automatically classified as being of interest. For some settings, particularly where a well-defined model of the system's normal behaviour can be accurately observed and characterised by statistical processes, this method can be highly effective. However in cases where an accurate system model cannot be formulated, or when a very imperfect or incorrect model is assumed, the results are likely to of little or no value.

Since the work presented in Chapter 3 takes a statistical approach, this third area is further expanded in the following subsection.

2.1.1 Statistical Anomaly Detection

Due to the wide range of application areas, direct comparison among different methods and algorithms is often not possible. Patcha and Park note that there is no standard metric for evaluating anomaly detection systems, making comparison among different systems rather difficult [87].

With this in mind, a critical review of a number of leading statistical anomaly detection approaches follows.

A straightforward approach with wide applicability was outlined by Basu *et al.* [15]. This method assumes the data stream is to some degree locally smooth, and that data points closer in time tend to have a higher correlation than those more widely separated. A pair of similar approaches are proposed which use a time window (referred to as a neighbourhood) and compare each value to the median value within the window. The first approach compares data with past and future results (within its neighbourhood) while the second examines only its immediate past. A threshold is then used to determine whether it should be classified as an outlier data point. Depending on the values selected for the neighbourhood size and the threshold, the algorithm can be tuned to accept more or less noisy data at the expense of accuracy. The low computational complexity and windowed nature of the algorithm make it suitable for real-time

applications. The authors remark however that the values used for window width and thresholds were determined somewhat arbitrarily, even differing for different parts of the same signal.

Although some suggestions are made about automating the choice of these parameters, there is no clear analytical solution offered. While this approach has the advantage of real time computability, the lack of a theoretically underpinned basis for the selection of the window width and threshold values is a major weakness. Without such a development, this method has no assurance of giving optimal performance without extensive testing.

Desforges *et al.* [31] proposed a statistical method for anomaly detection which can be applied to arbitrary types of data. Rather than being designed to identify single anomalous data samples, this approach attempts to identify when a system as a whole starts to deviate from its normal behaviour.

The method creates its Probability Mass Function (PMF) estimates using kernel density techniques. By combining many individual kernel functions centred on sampled data points, a PMF estimate can be constructed. If the constituent kernel functions used are simple rectangular functions, then the result is identical to a histogram whose bin spacing is equal to that of the spacing between the kernel functions. If the underlying data is known to belong to a particular distribution model, then this is the ideal choice of kernel function for creating that data's PMF estimate. The authors note that a Gaussian distribution was chosen as a starting point for their investigations into choice of kernel function. It is noted that these methods take large amounts of computational resources when the datasets are of high dimensionality, so a wavelet transform may be employed to speed run time, though no comment is made on the effects this has on accuracy.

Once two PMF estimates have been calculated they can be compared in order to identify differences. The suggestion is that a human operator would have a PMF made using data from a correctly functioning system to compare with unknown data. Although it provides a usable visual output for a trained operator to read (one of the examples given is of Radio Distance and Ranging (RADAR) target data which is often interpreted visually), the authors suggest no algorithmic method to detect when anomalous conditions have occurred. Without such a method, this anomaly detection scheme cannot be implemented in an automated system.

Zhang *et al.* [129] designed a statistical anomaly detection method to identify either

malfunctioning or maliciously transmitting nodes within a Wireless Sensor Network (WSN) (i.e. an insider-attack problem). It uses the mean and standard deviation of each sensor node's packet transmission rate to profile what normal behaviour looks like, allowing abnormal transmissions to be detected based purely on the frequency with which packets are sent. This method is unique among the others reviewed here in that it is designed explicitly for a cooperating, distributed system of nodes, allowing this method to scale up easily to scenarios where there are a large number of nodes. While this approach processes time series data, its applicability to anomaly detection in radio communications is limited because of the 'bursty' nature of many transmission technologies, in contrast to the regular and predictable patterns typical in a WSN.

Kun *et al.* [80] proposed a novel statistical outlier detection method which uses clustering to spot anomalous data points. This method has the advantage that it requires no parameter tuning (a common inconvenience in many anomaly detection schemes), making it an entirely data-driven system. It is specifically tailored to defeat the so-called 'curse of dimensionality' [17, 32] which causes an exponential increase in search space with increasing dimensionality. It avoids the problems commonly associated with highly-dimensional data by projecting each data point from the high-dimensionality feature space it originated from into a 'distribution difference' space of much lower dimensionality. 'Distance spheres' are defined by computing the distances between all data points in the original feature space and creating spheres with radii equal to these distances and centred at each point. After computing the density of these spheres (by counting how many other points are contained within each sphere), a space of much lower dimensionality can be defined using this density definition. Clustering of data points is then performed in this new space, and small clusters are defined as being outliers. This final step is common to many classification systems, and can be achieved using a variety of means, such as k-Nearest Neighbour [26]. Although this method provides a good detection ratio when applied to real-world data in off-line systems, its computational performance makes it unsuitable for the real-time requirements of communication systems, despite its run time increasing only linearly with number of dimensions. Also there is no direct comparison of running times with other, similar methods to allow a comprehensive comparison.

An alternative approach with an explicit on-line/real-time design focus was proposed by Filippone *et al.* [39]. It targets applications where the amount of training data is small, and uses information theoretic methods to detect anomalies [107, 27]. It functions

by estimating the information content of a new data point given a training dataset. It reduces the false positive probability by explicitly taking the size of the training set into account. This is because a smaller training data set provides less reliable sampling of a system's behaviour. The system is somewhat limited in its usefulness as it assumes the data distribution to be Gaussian-shaped (or a mixture of Gaussians). While this assumption will be true in some situations, it does not cover the general case, making it a rather niche solution. An ideal solution would make no such parametric assumptions about a system's input characteristics, but would be entirely data-driven.

Paper	Method	Comment
Basu [15]	Time windowed comparison	Low complexity, but requires hand-tuning of several parameters
Desforges [31]	PMF comparison using kernel density method	Requires knowledge of expected PMF shape, and results must be human-interpreted, preventing its use in automated systems
Zhang [129]	Standard deviation and mean-based profiling	A completely data-driven method, good for time series data with constant mean and standard deviation under normal circumstances
Kun [80]	Clustering with dimensionality reduction techniques	Decreases run-time when compared to classic approaches such as k-Nearest Neighbour when applied to highly-dimensional data. Good detection ratio on real-world test data.
Filippone [39]	Information theoretic analysis	Small amount of training data required, and 'learns' while running. Assumes that input dataset is Gaussian-shaped, which limits applicability.

Table 2.1: Outline table summarising the previously described statistical anomaly detection methods.

The following subsection will now introduce how statistical anomaly detection can be applied to the field of wireless communication systems.

2.1.2 Statistical Anomaly Detection in Wireless Systems

Abnormal radio signals have a wide range of causes such as interference from other wireless devices, a poor transmission channel or malfunctioning transmitters, and the effect of such anomalous behaviour will almost always be a corruption in the data being transferred. It is immediately obvious that detection of these anomalies is paramount

to making sure user data is transferred in an error-free manner, but further analysis of anomalies will have additional benefits beyond simply detecting when data corruption has occurred. Such of these anomalies are the result of channel effects, which if understood sufficiently could be used to infer information about the channel, perhaps leading to improved prediction algorithms.

Although error coding is widely used to detect and correct small transmission errors [50, 18, 42], such codes do not provide insight into what could have caused the error in the first place. Anomaly detection techniques, as well as performing their primary function (of detecting such transmission errors), may offer some insight into the cause of such transmission errors, potentially allowing the radio system to change appropriate transmission parameters (e.g. bit-rate, modulation order, scheduling algorithms, etc.) to prevent the problem re-occurring. The idea of modifying transmission parameters to maximise throughput is key to the emerging field of Cognitive Radio (CR), in which radio systems have some degree of adaptive intelligence, allowing them to coexist with existing legacy radio systems (often referred to as 'Primary Users') without causing detrimental interference. By using anomaly detection techniques a CR could effectively detect if it were causing interference with the primary user, and then modify its scheduling appropriately. This is the backdrop to the next selection of literature examined.

Afgani *et al.* developed a pair of complementary statistical anomaly detection methods based on related information theoretic principles. Firstly, they proposed a method to detect anomalies by analysing the information content of a signal's envelope in aperiodic signals [2, 8]. By making the assumption that the more common an event is, the less information it conveys (and conversely uncommon or anomalous events convey a lot of information) a highly effective anomaly detection scheme can be created. It constructs an estimated histogram of the signal's power distribution over a particular time window, meaning that this approach does not require any training data and is very flexible and potentially applicable to a wide range of wireless systems. It does however require some careful choice of parameters (such as histogram resolution, cluster size and alarm threshold) in order to function correctly, though these can be chosen at the design stage and typically remain fixed for a particular implementation. Since it functions on the signal envelope, it has the added benefit that demodulation of the signal is unnecessary for anomaly detection.

The second approach is tailored specifically to periodic signals [4, 5, 7, 8]. Similarly

to the first method, it uses histogram estimation based on the signal envelope to perform anomaly detection, but as the signal is assumed to be periodic, an additional useful assumption can be made. In a correctly functioning periodic system, consecutive transmission bursts (usually corresponding to radio packets) will have a high degree of statistical similarity, meaning that two PMFs from successive transmissions ought to have similar shapes. Thus if the signal does not conform to this assumption, it is likely to be anomalous. Calculating the dissimilarity between two estimated histograms (separated by the signal's period) allows the system to detect when the signal is in this anomalous state. A range of algorithms was investigated to perform this comparison between the two histograms, but the most promising (from both an effectiveness and efficiency standpoint) was the Mutual Information (MI), also known as the Kullback-Leibler Divergence (KLD). It measures the difference in information content between two distributions, and a very computationally efficient algorithm is available for its implementation. As with the first approach, demodulation of the signal is not required, further simplifying the design and making it a very promising choice for real-time anomaly detection.

In experiments on real-world data captures, the KLD method was able to detect a range of errors (frame timing errors, interference on a WiFi network from a Bluetooth device [4, 5, 7]). These anomalies would have produced a variety of errors in the data stream, so detecting them at an early stage (i.e. before demodulation and error code checking) is highly advantageous. The KLD approach can also be used for boundary detection, which can be used to determine the signal's period, and then set the anomaly detection system to the correct time base. The approach's suitability for real-time Radio Frequency (RF) operation was demonstrated by the construction of an Field Programmable Gate Array (FPGA) based hardware prototype which could effectively detect a range of anomalies in a signal with a bandwidth of 10MHz [5].

2.2 Neural Networks

The Neural Network (NN) was proposed in the 1940s and 1950s as a computational method for solving arbitrary computational problems, inspired by the design of animal central nervous systems [77, 38, 96, 99]. It is for this reason that the basic unit which comprises a NN is still referred to as a neuron. Although its popularity as a method has waxed and waned in the decades since they emerged, the advent of fast and affordable computers has allowed them to be used in a huge array of applications. Since it has been

claimed that standard, 3-layer feedforward networks can approximate any function to an arbitrarily high precision, there is no function for which a NN based approach could not work, although in practice there are certain constraints such as collecting sufficient training data and affording sufficient processing time [28]. They have frequently found uses in the fields of function approximation, data processing, robotics, control and classification [113].

One of the greatest strengths (and also perhaps also the greatest weakness) of NNs is that they are entirely dependent on training data to mimic the system's behaviour (referred to as a data-driven approach, as opposed to a model-driven one). As first noted by Cybenko [28], given the correct training data, a neural network can approximate any measurable function. Training data consists of real input and output data observed from the system being modelled, which in many circumstances is much more readily available than a well-defined system model [91]. For example, in a weather forecasting application, training data could be comprised of historical weather patterns, which is much easier to collect than to create an accurate statistical model of weather patterns.

While this data-driven approach is very convenient in a large variety of settings, it does have a weakness that unless a properly representative set of training data of sufficient length is provided, the network will not be able to approximate the system's function accurately, leading to erroneous results. Good training data must be representative of expected data [91], and must adequately represent statistical variation. Additionally, though care must be taken not to 'over-train' a network, where the NN learns the training data perfectly, but is unable to cope with inputs which are different. Ensuring that these criteria are met are key to a successful NN.

NNs are generally split into two broad categories: 'Feedforward' or 'Recurrent'. The majority of NNs employed today in real-world applications are feed-forward networks [56] (meaning that connections between neurons never form a cycle), and they have the advantage that training is simple because there are no feedback loops (which are much harder to train optimally). See Fig. 2.1 for a diagram of a typical NN structure.

The less commonly used Recurrent Neural Network (RNN) type form loops in their structure. While this can cope with a more complex system because it has no restraints on which neurons may connect to one another, training is often extremely resource intensive and complex. There is some experimentally based evidence that RNNs are better suited to non-linear system modelling because they can exploit the feedback loops inherent in their structure [49, 62]. However, they were were not heavily researched

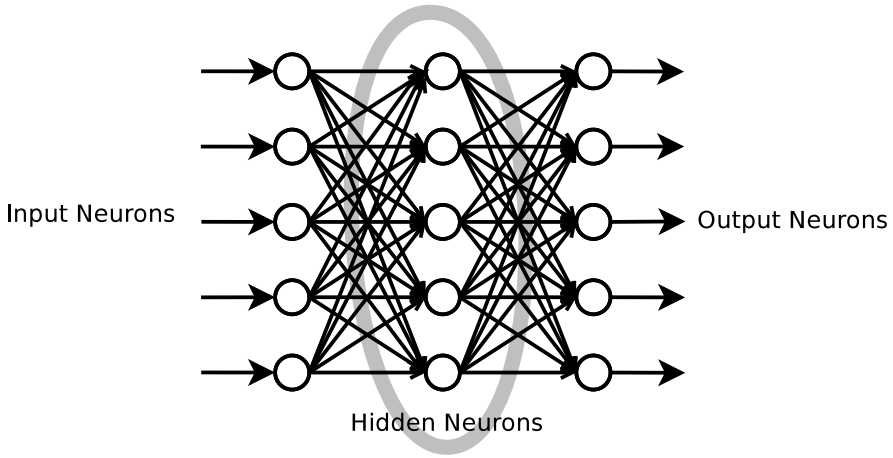


Figure 2.1: A typical 3-layer feedforward neural network. As is often the case in such networks, every neuron in one layer is connected to every other neuron in the next layer (referred to as full connectivity). Usually all the connective weights between neurons are varied during the training stage. Neurons with no direct connections to inputs or outputs are known as hidden neurons, and form the middle layer in this example.

for several decades, until refinements were made to overcome the training complexity challenge.

Since Chapters 4 and 5 use a sub-class of RNNs, they will be further discussed in Section 2.2.3. Firstly however, the use of NNs as classifiers is presented.

2.2.1 Neural Networks as Classifiers

Like anomaly detection, the problem of classification is a well-studied one with with many application areas. An archetypical neural network classification problem is one of speech recognition: from a set of pre-defined words, the NN must identify the words(s) in a spoken input based on earlier recordings of speech (with clearly identified words) used to train the NN [104, 69, 20]. Other areas which have used NNs as classification tools are intrusion detection [130, 131], Nuclear Magnetic Resonance Imaging [84] and even analysis of satellite photography to determine land use [46].

In contrast to traditional statistical classification methods (in which an underlying probability model must be assumed), NNs take a data-driven approach to classification, needing only appropriate training data to make a classification.

As evidenced by the weight of literature, in the vast majority of cases where NNs are employed for classification, feedforward networks rather than RNNs are used [128, 44, 11, 115, 63, 93]. There is however a large amount of common ground between the

two approaches, and several relevant contributions to the literature are briefly examined below.

In response to an earlier study by Tam and Kiang [110] which had produced very promising results, Wang examined the unpredictable nature of NNs in classification problems [11, 115]. He demonstrates that within a classification system driven by training data (as is the case with all NNs) there is an infinite number of solutions which correctly classify the training data, but make different classifications of new data. This observation clearly highlights two points: firstly, the training data must be carefully chosen to represent the system under investigation, and secondly that there is no unique correct solution for classification which can be arrived at without additional data. Wang proposes a smoothing method for the data being classified, but recognises that this is only relevant in certain fields, and requires domain specific knowledge.

Rogova proposed an approach to improve the accuracy of NN classifiers by using multiple different types of NNs to classify the same data points [98]. As in [115], he implemented a system which outperforms any individual classifier by taking a weighted combination of the individual networks' probabilistic classifications. While this method typically halved the number of data points mis-classified, it does not address the fundamental problem of the unpredictability typically exhibited by NNs.

Kittler *et al.* compared the performance of NNs with a range of probabilistic classification approaches in a number of problems to observe the merits of each method [63]. Like Rogova [98], they found that by incorporating output from several different systems, results could be improved. They found that while the NN method worked well, the error rate could be improved by having a 'voting' system, where each data point was classified in parallel by a number of competing systems, and the most popular classification would be chosen. This suggests that NNs will make misclassifications which more conventional systems (e.g. Hidden Markov Models or Gaussian Classifiers) will not make, and vice versa. By having a range of classifiers, a consensus can be arrived at which outperforms any single classification method.

2.2.2 Neural Networks as Predictors

From their inception, NNs have been studied for the task of prediction (also referred to as 'forecasting' in some of the literature). Machine learning and function approximation are building blocks which very naturally support time series prediction. Although it was not until the late 1980s that computers were available which were capable of running

large NNs for practical applications, the foundations had been laid in the 1950s and 1960s. For the last 30 years they have found applications throughout computer science, engineering, biological sciences and the world of finance. Due to the extensive range of application areas where NNs have been used in recent decades, any overview of the topic can only scratch the surface. With this in mind, a few of the more prominent developments are outlined in the following section.

An area where they have attracted much interest is in weather forecasting [127]. The problem has much in common with problems faced by wireless engineers: the system to be predicted is extremely complex, and although many previous models have been made, none are clearly optimal. Also there are numerous relationships within the data which are difficult to capture, but sufficient historical data is available to allow training of neural networks.

French *et al.* [40] used a NN to perform rainfall predictions. As in the vast majority of cases, they used a 3-layer feedforward network, and used the backpropagation through time training algorithm. Their results were comparable or marginally better than the accepted domain-specific prediction methods. Although the NN approach was significantly more computationally intense, improved results could be achieved by either increasing training time or the number of neurons used. An additional benefit noted is that the network could be dynamically retrained as it runs, so it could continue to improve its predicitons by learning. This can be a key feature of NN systems.

Another area NNs have found a variety of applications is in the financial sector. Clearly large sums of money depend on correct predictions of the future, whether in stock market trading, risk analysis, investment choices and a variety of similar areas. Odom and Sharda [81] used a simple 3-layer network with the standard back-propagation though time training algorithm to perform bankruptcy predictions on companies. When compared to conventional techniques used in the financial arena (known as linear discriminant analysis) the neural network was able to produce a higher ratio of correct predictions, as well as giving more robust results when operating on small sample sizes.

White [116] examined the daily stock price of a particular company, using this data to train a feed-forward network. As in the previous case, the standard back-propagation thorough time training method was used. The network was not able to give a prediction which would give a potential investor an advantage, suggesting that perhaps the behaviour of the stock price is a random process, rather than one which can be predicted with only historical data. This highlights that although NNs are very

versatile, they are no more capable of any method of predicting a random process than any other method.

A popular test case for testing a NN's predictive ability is sunspot behaviour data. It is well established that the sun has a fairly regular cycle lasting approximately 11 years, however the time series is quite chaotic. Good predictive ability on this data is often used as evidence that a particular network is capable of performing predictions on chaotic time series data. Koons and Gorney [64] used a 3-layer feedforward network with 33 months worth of historical data to predict the number of sunspots present for the then upcoming solar cycle. They claim their predictions were more accurate than other methods which did not use neural networks, and the results would prove useful in compensating for interference with satellites which is a common result of increased sunspot activity.

2.2.3 Echo State Networks

First proposed in 2001 by Jaeger [57, 59] the Echo State Network (ESN) is a type of RNN which has been shown to have applications in modelling highly non-linear dynamic systems. They have the unique property that training the system only modifies the weights assigned to the output neurons, rather than additionally modifying all internal connecting weights (as happens in classical NNs), making training a computationally simple task. The internal connecting weights in an ESN are initialised randomly, and remain fixed for the life of the network. ESNs are also sparsely connected internally (often having only around 1% interconnectivity), which allows the internal neurons (referred to as reservoir neurons) to produce a number of loosely interconnected subsystems which cooperate to give the desired output. Since training an ESN is a relatively computationally inexpensive task, it is quite feasible to have a network with several thousands of neurons, allowing complex systems to be modelled efficiently. An almost identical approach known as Liquid State Machines was developed independently from and simultaneously with ESNs by Maass [72], and the combined area has become known as Reservoir Computing.

They derive their name from the "Echo State Property" which any ESN must demonstrate. Informally it can be summarised as

> a condition of asymptotic state convergence of the reservoir network, under the influence of driving input. [57]

More informally, it means that the effects of initial conditions will vanish as time passes.

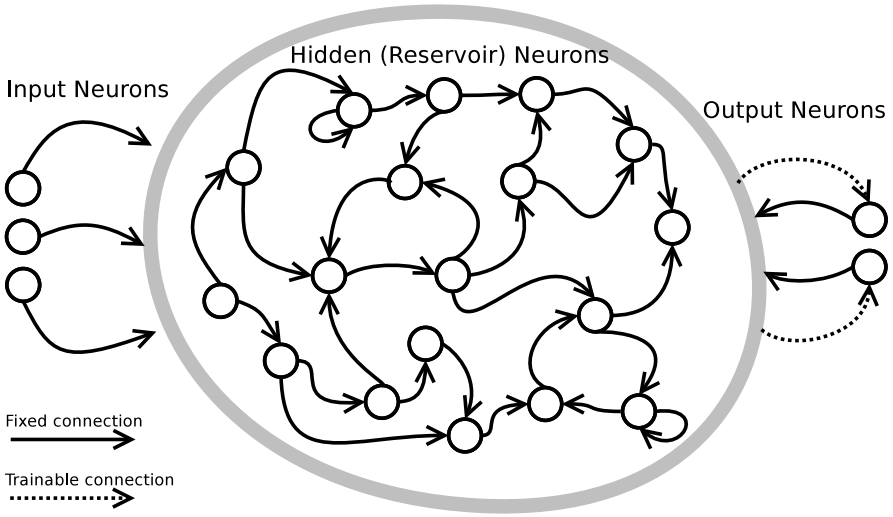


Figure 2.2: An example echo state network, illustrating that the only trainable connections are those from the reservoir to the output neurons. Also visible is the recurrent nature of the network, and the randomly chosen, fixed connections between internal neurons. The number of input and output neurons is determined by the dimensionality of the input and output. In contrast to the NN shown in Fig. 2.1, ESNs typically have quite sparse internal connections.

Echo-state networks have been used to model and classify a wide range of complex systems - they have found applications from communication channel equalisation, [60] to future wind speed forecasting [117] and prediction of hydroelectric dam water inflow [29]. Their ability to characterise a system based purely on empirically obtained training data and not on a mathematical model, is extremely attractive in many fields, as mathematical models may be complicated, inaccurate or non-existent, while historical records of past performance are often much more readily available. ESNs allow the user to treat the complex system as a 'black box', whose input and output relationships are unknown at the outset, and for the neural network to infer what function the system performs using only the training data. This will result in a model based on the real system, rather than on imperfect assumptions. This does, however, require careful selection of the training data and system parameters to ensure that it is representative of normal behaviour and that the system does not suffer from a symptom common to all NN types, known as "over-training".

Because ESNs are a subset of NNs, they suffer from some of the same problems. One of the most serious is that of over-training. This occurs when the network reproduces the training data 'from memory', rather than by successfully modelling the underlying

system model. If over-training has occurred, then the network is unable to process any data other than that on which it was trained. When using ESNs, it has become common practise to add a low level of noise to the system's internal neurons during training. This tends to reduce how closely the ESN fits the training data, preventing over-training and often resulting in a more generally applicable training solution than if no noise is added. Adding too much noise will however obscure some of the desired signal features and prevent a useful solution from being reached. Deciding how much noise to add is currently a matter of trial and error specific to each setting in which ESNs are used, as no widely applicable rigorous research has been done to quantify the effect of noise on training fitting and system generality [97].

While ESNs are usually described as being sparsely connected, this is not a formal requirement. Although early reports regarding ESNs cited low connectivity as being responsible for creating multiple, nearly independent systems, there is now a growing body of evidence showing that fully connected reservoirs perform as well as sparsely connected ones [59].

There are however practical reasons for wanting a sparsely connected network: conventional RNNs with full connectivity containing n neurons exhibit computational complexity of $O(n^2)$. An ESN in which each neuron is constrained to be connected to a fixed number of other neurons only has a run-time complexity which grows with $O(n)$. Jaeger notes [59] that as there is little evidence of improved results from a fully connected network over a sparsely connected one, so naturally the network with a lower run-time is preferred.

Another problem common to NNs in general is that of neuron saturation. This typically happens when a neuron with a non-linear compression function (such as sigmoid) gives an almost constant output for most input values. This is a more significant problem in RNNs than in non-recurrent networks, as the feedback generated can cause the saturation to spread throughout the network, and significantly affect the output results. Ensuring training values and inputs are scaled to an appropriate range is usually sufficient to protect against over saturation, however when using ESNs it is not possible to eliminate the possibility completely due to the randomised nature of the inter-neuron weights. If an output neuron saturates in an ESN, it can be a sign that the network is unable to adequately process the current input signal, given its training experience.

A feature of ESNs is that, unlike feedforward NNs, often they cannot make good predictions for a number of time steps after they begin processing due to uninitialised

data within the reservoir. When data starts to appear at the output of an ESN, it is initially based on the uninitialised states of the neurons, rather than on any real input data. For a large ESN, it can take several hundred steps for this junk data to be replaced with data from the valid input. If the network does possess the echo state property, then it can be assumed that the initial junk data will contaminate the output for only a limited number of steps.

There now follows a critical review of some of the more prominent publications concerning echo state networks. Work which models non-linear systems will be particularly highlighted, along with developments to the echo state method itself.

Jaeger *et al.* demonstrated the ability of ESNs to predict highly chaotic time series [60]. It was shown that such networks are capable of learning the Mackey-Glass system [73] (a standard benchmark technique used for time series prediction) by treating it as a 'black box', where no data other than inputs and outputs are used to deduce its operation. The ability of ESNs to learn time series such as this one made it apparent that other applications were possible, including wireless communication channel equalisation. Under high Signal to Noise Ratio (SNR) conditions, the conventional channel equalisation schemes performed two orders of magnitude worse than the proposed method, highlighting the versatility of the Echo State approach.

Venayagamoorthy found ESNs useful in predicting load in a distributed power generation grid [114]. In a direct comparison between an ESN and a time-delayed NN, the ESN reacted more quickly and with a higher accuracy. It has the additional advantage that it is capable of on-line learning. While such networks always undergo a training phase at the start of their operation, ESNs can also adaptively train themselves during operation, so that as the environment changes they can adjust their parameters to yield better results. This feature means that as a system being modelled evolves over time (in this case a power transmission network) it can be accurately represented in its dynamic state.

This work was further extended by Dia *et al.* to include Multilayer Perceptron (MLP) networks, in addition to ESNs and RNNs to solve a related problem in power transmission of 'harmonic pollution' due to load nonlinearities [30]. The comparison is of note because it highlights the differences between the three different neural network types, as well some of the peculiarities of the ESN. The MLP network (a classical feed-forward type) performed the least well of the three systems, requiring a larger number of neurons than the two recurrent networks for its task. The authors comment that

the smaller network was sufficient to learn the complexity of the system, and it was noted that the performance of a 30 neuron ESN was inferior than that of a smaller, 15 neuron ESN. It is also mentioned that the ESN's performance depended more heavily on the initial weights than either of the other neural network types used, as some of the weights were not changed during training.

Additionally they remark that in their experiments increasing the number of hidden neurons did not imply better performance, as additional neurons tend to change the characteristics of the dynamic reservoir.

This result is not unique to ESNs as it is a recognised phenomenon within other NN types, although the untrainable randomly initialised weights in the reservoir may accentuate the issue in the case of ESNs. It is also noted that the computational time to train the ESN is lower than the MLP and RNN.

Schrauwen *et al.* demonstrated the use of ESNs in the classic NN field of speech recognition [104]. They showed that not only were they a suitable choice, but that they could also be implemented efficiently on FPGA hardware. This is a significant contribution, as previously all implementations had been purely in software. Moving the processing to dedicated hardware further reduces the power requirements, a necessary step in being able to implement ESNs in portable wireless devices.

Yang *et al.* found ESNs useful in predicting the temperature of transmission power lines [125]. Since temperature and electrical resistance are so interlinked, it can be a complex task to estimate the temperature of a conductor, especially when ambient weather conditions are considered. Using only historical training data, they were able to find a close match between the ESN results and the widely accepted IEEE standard for the applicable conductor type. This then allowed accurate prediction of how much current could be safely passed along a given conductor line. They did not use any on-line training, but mentioned that it could be of use in further research.

Seth *et al.* modified the ESN architecture to process complex-valued numbers, with the goal of using them in the signal processing domain [106]. This was achieved by replacing the connecting weights between neurons with complex numbers and the activation functions with complex nonlinearities. The complex variant of the standard backpropagation algorithm for training was used. In the task of symbol decoding, these modified ESNs outperformed other complex-valued NNs, though no comparison was made with conventional Quadrature Amplitude Modulation (QAM) decoders. This work was extended by Xia *et al.* who placed an additional non-linear output neuron

layer between the reservoir and the output [121]. This modification to the ESN appears to improve the degree of function approximation that can be achieved.

Boccato *et al.* propose a similar modification to the ESN architecture, again adding a non-linear output layer in order to increase the processing abilities of the network [21]. They chose a Volterra filtering layer as it captures a certain amount of temporal information, providing better additional feedback to the system. They were able to demonstrate that their modified ESN outperformed a conventional ESN, sometimes by a wide margin. The authors do not however comment on the applicability of their approach to other problems commonly addressed by RNNs.

Recently there has been significant interest in understanding the reasons for the efficacy of ESNs for solving particular problems. Both the recurrent nature of the network, and the randomly assigned connections and weights make a rigorous understanding of ESNs' behaviour a complicated problem. While there exists much experimental evidence of their worth, the literature dealing with how they function is limited, especially when compared with the much more mature field of RNNs.

Manjunath and Jaeger introduced new research from the field of nonautonomous dynamical systems theory which brings a previously absent degree of mathematical rigour to the ESN method [74]. Specifically, they examined the common case of input-driven neural systems and addressed questions of system stability. They found that for an ergodic input, it can be determined whether a network does or does not exhibit echo state properties by examining its response and without analysing the network itself. This is a good first step to providing an analytical grounding to the ESN method.

A useful study was undertaken by Rodan *et al.* to formalise the complexity requirements of ESNs [97]. They note that:

1. There are properties of the reservoir that are poorly understood.
2. Specification of the reservoir and input connections requires numerous trials and even luck.
3. Strategies to select different reservoirs for different applications have not been devised.
4. Imposing a constraint on spectral radius of the reservoir matrix is a weak tool to properly set the reservoir parameters.
5. The random connectivity and weight structure of the reservoir is unlikely to be optimal and does not give a clear insight into the reservoir

dynamics organization. [97]

As there are some in the scientific community [90] who have reservations about using ESNs in real-world applications due to the lack of mathematical rigour, they address a number of these concerns, and find that although rigorous analytical solutions are difficult, it is possible to reach some useful conclusions. Firstly they give experimental evidence that a fully-connected reservoir is often sufficient to give equivalent performance to an ESN. Secondly they show that reservoirs can be constructed in a deterministic manner and that they can perform as well as randomly connected ones. Thirdly they have the advantage of being much easier to analyse, opening the way to a fully rigorous understanding of ESN behaviour.

The field of reservoir computing and ESN research is currently very active, with new applications and architectural variants being proposed. Perhaps the most active area presently is the search for a better understanding of why ESNs are so effective in certain areas. As has been mentioned, there is some hesitancy in the scientific community to use such a tool in the absence of a firm understanding of its behaviour or a high level of confidence that it will function reliably. A rigorous mathematical framework would be extremely beneficial because it would underpin system designers' ability to assess the reliability and accuracy of an ESN system's functionality and output.

2.3 Wireless Channels

This section introduces some relevant background to wireless channels, how they are modelled, and work which has been carried out to both characterise and predict their behaviour.

A channel is a mathematical model used to describe the various effects (such as attenuation, fading, Doppler shifts, multipath reflections and scattering) on a transmitted signal between its transmitter and receiver [109]. Without accurate channel information, decoding transmitted signals correctly becomes an intractable problem, since the originally transmitted signal cannot be reconstructed to a high level of accuracy without channel knowledge.

As a radio wave propagates through an environment, it is subjected to a variety of effects which (from a communications standpoint) can corrupt the signal being transmitted. A clear understanding of these effects is crucial if these (often detrimental) effects are to be compensated for.

2.3.1 Fading Channels

As a signal passes over a transmission channel, its phase and magnitude are changed. If coherent modulation techniques are being used, a change in the phase will affect performance if it is not corrected for. Similarly, if the modulation is sensitive to the signal's envelope then this must be accounted for also. There are several effects which combine to change the phase and magnitude of a signal which are outlined below.

Fading is the name of the variation in attenuation which a channel will exhibit when a signal is transmitted over it. Fading may vary over time, frequency and geographical location. When evaluating a fading channel, a distinction is made between slow and fast fading. A channel is referred to as exhibiting slow fading if the symbol time is less than the channel's coherence time, T_c. This is the period over which the channel impulse response is considered to be non-varying. This is a critical value for a radio system to know, as it is the interval during which the validity of an active channel measurement can be assumed. Beyond the coherence time, either a new measurement must be taken, or an estimate must be made.

If the symbol time is greater than T_c, then the channel is said to exhibit fast fading. In slow fading, a particular fade will affect multiple symbols, whereas in fast fading the channel is not correlated from one symbol to the next. Fast fading channels therefore require consideration to be taken about the variation of the fading from one symbol to the next.

The maximum Doppler spread, f_d, is related to the coherence time via 2.1 [109]:

$$T_c \simeq \frac{1}{f_d} \tag{2.1}$$

Frequency Selectivity of Fading Channels

If a channel exhibits fading across its bandwidth (i.e. the magnitude response is not the same at all frequencies) it is said to exhibit frequency selective fading, and is referred to as a wideband channel. If, on the other hand, the channel has a uniform magnitude response over its entire bandwidth, the fading is referred to as being flat, and the system is referred to as being narrowband.

The crucial measurement for quantifying frequency selective fading is the coherence bandwidth, f_c, and is the frequency range over which very similar amplitude attenuation and phase rotation effects are exhibited. In a wideband signal, coherence bandwidth

influences the minimum frequency spacing of pilot signals for channel estimation. The maximum delay spread, $\tau_{\max}$, is related to the coherence bandwidth via 2.2 [109]:

$$f_c \simeq \frac{1}{\tau_{\max}} \tag{2.2}$$

Multipath Fading

Multipath fading occurs due to the constructive and destructive superposition of a large number of reflected, refracted and time-delayed signal components. Depending on the nature of the channel environment, different statistical models can be employed to describe the fading envelope. A wide range of models (several of which are outlined below) can be used to calculate the average channel statistics depending on the nature of the propagation environment.

In the following distributions, the received carrier amplitude is modulated by α, the fading amplitude. α is a random variable with mean-square value of $\Omega = \overline{\alpha^2}$ and a Probability Density Function (PDF) of $p_\alpha(\alpha)$.

Rayleigh Model The Rayleigh distribution is used to describe multipath channels without a strong Line of Sight (LOS) component. The fading amplitude α has a distribution described by 2.3:

$$p_\alpha(\alpha) = \frac{2\alpha}{\Omega}\exp\left(-\frac{\alpha^2}{\Omega}\right), \qquad \alpha \geqslant 0 \tag{2.3}$$

Such environments are typical in indoor and urban scenarios.Similar environments are also found in the troposphere and ionosphere where there is a high degree of reflection and refraction.

Nakagami-n (Rice) Model In environments with a strong LOS component and a number of weaker scattered paths the Nakagami-n (also known as the Rice) model is often used. In the following equation, I_0 is the zeroth order modified Bessel function of the first kind and n is the Nakagami-n fading parameter, ranging from 0 to ∞. The PDF is described by:

$$p_\alpha(\alpha) = \frac{2(1+n^2)e^{-n^2}\alpha}{\Omega}\exp\left[-\frac{(1+n^2)\alpha^2}{\Omega}\right] I_0\left(2n\alpha\sqrt{\frac{1+n^2}{\Omega}}\right), \qquad \alpha \geqslant 0 \tag{2.4}$$

Similar channel scenarios can be observed in small-cell deployments, satellite communications and some ship-to-ship radio links.

Nakagami-q (Hoyt) Model In the Nakagami-q (or Hoyt) model, q represents the Nakagami-q fading parameter. The PDF is given by:

$$p_\alpha(\alpha) = \frac{2(1+q^2)\alpha}{q\Omega}\exp\left[-\frac{(1+q^2)^2\alpha^2}{4q^2\Omega}\right] I_0\left(\frac{(1-q^4)\alpha^2}{4q^2\Omega}\right), \qquad \alpha \geqslant 0 \tag{2.5}$$

This model is typically applied to satellite communication channels which are subjected to atmospheric scintillation (sometimes referred to as 'twinkling').

Nakagami-m Model The Nakagami-m can be used to represent a wide variety of different channel types, since the one-sided Gaussian distribution and the Rayleigh distribution are simply special cases, and the Hoyt distribution can also be very closely approximated. m represents the Nakagami-m fading parameter.

$$p_\alpha(\alpha) = \frac{2m^m\alpha^{2m-1}}{\Omega^m\Gamma(m)}\exp\left(-\frac{m\alpha^2}{\Omega}\right), \qquad \alpha \geqslant 0 \tag{2.6}$$

This model often gives the best fit for land-mobile and indoor-mobile scenarios, as well as ionospheric communications subjected to scintillation.

Shadowing

Much as the name would suggest, shadowing effects are caused by large objects (buildings, trees, terrain etc.) obscuring certain links from the transmitter, and are modelled in a different fashion to fast fading in the multipath context, due to the slower nature of shadowing. It has been shown by extensive experimental evidence [24, 47, 19] that the large-scale variations caused by shadowing follow a log-normal distribution over distance.

Pilot Signal Channel Measurement

While there are a large number of refinements to this approach, all pilot-assisted channel measurement schemes work in essentially the same way. A signal of known amplitude and phase (i.e. a pilot signal) $X(t)$ is transmitted over the channel being measured. The receiver then detects this signal $Y(t)$ after it has been subjected to the channel's complex

gain $H(t)$ and noise (often modelled as Additive White Gaussian Noise (AWGN)) $w(t)$ [55].

The transmitted signal can be expressed as:

$$Y(t) = X(t)H(t) + w(t) \tag{2.7}$$

By trivial rearrangement, an estimate of $H(k)$ can be arrived at:

$$H(t) = \frac{Y(t) - w(t)}{X(t)} \tag{2.8}$$

allowing $H(t)$ to be estimated, although the uncertainty in the noise term $w(t)$ will inevitably introduce errors into the calculation. Unless a channel's response can be considered to be completely invariant over time, the measurement of the channel will quickly become obsolete. There are two options available: either a new measurement must be made, or an estimate can be computed based on an assumed model. A new measurement will have the advantage of being more accurate, while an estimate will not waste bandwidth.

For channels subject to frequency-selective fading, the response must be known across the channel's bandwidth. Although it would be possible to measure the response of every subchannel, doing so would be unnecessary. Since there is often correlation among subchannels, the complete channel's response can be estimated using a number of pilot signals spaced by f_c across the entire bandwidth and using interpolation techniques. See Section 2.4 for details of how pilot signals are used in an LTE system.

2.3.2 Computer Modelling

There is a variety of approaches which can be taken to computer modelling of a wireless channel. Statistical models (such as the Okumura model [83]) can provide an estimate of mean expected path loss conditions based on a small number of parameters (such as frequency, distance, height of base station and mobile stations) for a particular scenario (such as a deployment in a dense city or in a sparse suburban environment). Other statistical models can provide estimates of channel statistics for a given scenario, as mentioned in the preceding subsection. Such tools are useful in infrastructure planning and radio design, but far more precise tools exist (although they require many more parameters to be known about the environment) which allow much finer-grained data

to be produced.

The most flexible models use a geometry based approach. In such simulations, the 3D physical layout of transmitters, receivers, buildings, terrain and other scatterers and reflectors is represented. The simulation then incorporates a large number of paths (referred to as rays) leaving the transmitter(s) and interacting with the model of the physical environment. This style of modelling is known as ray-tracing.

A frequency-selective fading channel for a wideband signal can be simulated using the following equation [109]:

$$h(t) = \sum_{l=1}^{L} \alpha_l e^{-j\theta_l} \delta(t - \tau_l) \tag{2.9}$$

with L being the total number of resolvable paths, l the channel index and α_l, θ_l and τ_l corresponding respectively to the channel amplitudes, phases and path delays. Each resolvable path in the model corresponds to a physical path through the propagation environment, with $l = 1$ being the first component which arrives, and its τ_l is zero by convention, with later parts arriving with larger time delays. It is often the case that the first signal to arrive contains the LOS component, and commonly has the largest α_l. The parts which arrive later tend to have travelled further through the physical environment, and exhibit deduced amplitude due to reflection or refraction. For a wideband channel, this process must be repeated over its entire bandwidth, as different frequencies will have various interactions with the environment. Not only will fades occur at sundry points due to constructive and destructive interference, but also materials have diverse permittivity across frequency.

By applying path loss functions, modelling how waves pass through different materials, are reflected or refracted by them and by measuring path lengths, very accurate values for α_l, θ_l and τ_l can be arrived at for each individual path. Doppler shifts can be taken into account, and noise can be introduced into the model to more closely resemble real-world conditions.

As may be evident, the challenge of mimicking the behaviour of such a dynamic real-world system is a difficult one, and these simulations will often tax the limits of the available computational power. Additionally, the amount of information that must be known about the surrounding environment (even down to knowing the permittivities of building construction materials) is huge, and gathering it would be a major undertaking.

Such simulation models are frequently used in research environments, allowing the testing of new algorithms and transmission schemes, and their flexibility in being able to simulate an arbitrary propagation environment is one of their main strengths.

However, due to the computational costs, these accurate models are invariably used in offline situations, sometimes taking hours or days to produce useful results. If however this same computation were to be achieved in a matter of milliseconds (and using only the small amount of energy permitted by a mobile device's battery) then ray-tracing models could be of use in real-time radio systems. However this is unlikely to become possible in the foreseeable future.

2.3.3 Channel Characterisation and Prediction

Sometimes general knowledge about a channel can be of use to an RF system (e.g. how often on average the received signal's magnitude becomes nearly zero (known as a "deep fade") or a general description of its likely surroundings). Such approximate data can be used to choose a modulation and coding scheme to match the limitations of the channel. More specific channel data allows greater exploitation of the channel's characteristics, but generally comes at the expense of increased channel measurement. The area of research concerned with obtaining general channel data is referred to as channel characterisation. Sometimes this will involve comparing a channel with a set of previously observed channels, and pairing it with the most similar subset. This then allows adaptive modulation and coding logic to use the optimal transmission scheme for that channel type. The advantages of channel characterisation is that often it can be achieved without active channel measurement.

Knowledge of a wireless channel's characteristics is vital to enhancing modern wireless systems, where a fast-fading channel's gain can vary by several orders of magnitude over a millisecond timescale.

Poor channel knowledge leads to high error rates, wasted power and sub-optimal throughput of data. Modern adaptive transmission technologies [12, 70, 22, 61] take advantage of channel information to choose a modulation scheme, power level and scheduling algorithm to maximise efficiency in transmission. A channel's behaviour is often treated as being so complex as to be indistinguishable from a truly random process. As such, channel characteristics are directly measured using pilot signals - typically short signals of known phase, magnitude and duration which can be used to measure a channel's distortion effects, but carry no end-user data. The information

supplied by the pilot signal is then treated as being valid over the coherence time of the channel. Once that time period has expired, further measurement is performed or estimates are computed. This infinitely repeated cycle of pilot signalling and measurement provides reasonable channel knowledge at the expense of wasted bandwidth and power, and also increased system complexity. A scheme which could reduce the number of pilot signals (or even eliminate them entirely) while maintaining the ability to supply accurate channel information is highly desirable in systems looking to improve their spectral efficiency.

In modern radio systems such as LTE, pilots and associated power control signalling can use upwards of 30% [102] of a system's link budget. Eliminating this overhead would make this bandwidth available to the end user. Any method which gets closer to this ideal than existing methods is highly desirable.

Channel measurement is relatively easy and widely deployed, typically using pilot signals, however it can only be relied upon for the exact time during which the pilot was sent. Due to the time-variant nature of typical channels, measurements are often out of date before the information can be used. Therefore channel prediction has seen a great deal of recent research [34, 124, 86, 53, 33, 85, 111, 92], with a variety of different approaches being taken. The goal is the same in each case: to be able to predict how the future channel will behave. Some systems aims for a general predictions, such as the average channel gain for the next time slot, while other systems aim for a more specific prediction of exact magnitude and phase predictions. The list of benefits to such systems is long: optimised power usage, maximised throughput, reduced reliance on expensive channel measurements, improved modulation and coding systems and eliminated power control signalling to name but a few. Although wireless channels are difficult to model, and sometimes they behave in a truly random fashion, there is often some information about future behaviour which can be deduced by examining past behaviour. A few of the more promising approaches are now outlined.

The most widespread channel prediction schemes use Autoregessive (AR) models [13], which presume that a channel's future behaviour can be predicted by a weighted linear combination of its previous values. AR models have been used successfully to predict long-term fading effects quite effectively [34], however AR algorithms can be very sensitive to noise, and so prove less successful when used to predict short-term fades [41]. The AR approach has the advantage that the model itself can be computationally simple to implement. Good prediction of the AR coefficients relies on accurate channel

measurements in order to estimate the correlation function of the channel coefficients, which in practice can be complex.

Dong *et al.* [25] proposed a simple AR channel prediction model specifically with CR in mind. When using a simple radio model (ALOHA [1]) to simulate both licensed and unlicensed users accessing the same band, they were able to achieve a lower Bit Error Rate (BER) at all SNRs when compared to using a Kalman filter under the same conditions. The performance gap between the two systems is most notable at under high SNR conditions, (with a BER difference of 7dB at the highest simulated SNR of 20dB) suggesting this system is best suited to favourable channel conditions.

It performs its predictions using a combination of AR modelling and Bayesean filtering (referred to in the paper as 'particle filtering'). While the computational simplicity of the system is noted, making it suitable for inclusion in a mobile device, the system falls far short of the upper bound for performance, that is when perfect channel knowledge is available.

Aguiar *et al.* [9] suggested a simple approach to channel prediction in WiFi networks: rather than using more complex predictors as in [13, 10, 35, 36], they employed heuristics. By using either a moving average or linear prediction scheme based on the received signal strength of the WiFi packets, decisions are made to change the modulation or coding scheme to maximise throughput. They argued that their 15% loss in capacity (compared to perfect channel knowledge) is a worthwhile tradeoff, given the simplicity of the predictors they implemented, especially considering that existing channel prediction systems for this WiFi scenario typically fall short of maximum capacity by 40%. Although this result does show some promise, even the authors remark that they did not understand how prediction errors affect channel-adaptive mechanism. Without a rigorous explanation of how the method works in relation to wireless signal theory, the method does not seem to show promise for use outside of the very specific WiFi setup described by the authors.

Wong *et al.* [118]used a sinusoidal modelling technique for outdoor moving receivers using Orthogonal Frequency Division Multiplexing (OFDM). They assume that the receiver is moving at a constant velocity in a straight line through a time and frequency selective channel. They assume that the channel model parameters vary slowly, which would be typical in an outdoor pedestrian environment. The authors have used the classical Estimation of Signal Parameters via Rotational Invariant Techniques (ESPRIT) [89] algorithm to extract channel parameters, and then use Bayesian-type model ex-

trapolation to produce a prediction scheme which can predict channel conditions ahead several wavelengths of where the receiver is currently placed. In simulation results, this scheme is shown to outperform a number of comparable schemes, however the authors note that their model is not suitable for indoor or dense urban scenarios because of the large number of scatterers often present, and comment that perhaps a hybrid system with an AR component would be more suitable in such scenarios.

Øien *et al.* [82] discussed the impact of inaccurate or delayed information on performing predictions of what the channel's future conditions are likely to be. They noted that both noise and delay have a marked impact on channel prediction algorithms, and proposed an algorithm to predict the channel's fading envelope within delay and noise constraints. They present closed-form expressions for BER and spectral efficiency for different channel models. Also, a linear envelope predictor was derived which used frequent pilot signals, although since the authors do not compare their design with other predictors, it makes direct comparison impossible. This work is nonetheless a significant development as it recognises the realities of round-trip time in feedback-based systems, in contrast to the many others which assume an error-free instantaneous feedback link.

The Sum of Sinusoids (SOS) approach [94, 120, 122, 88] refers to a range of methods, all of which make the common assumption that since the real channel is composed of sinusoids of differing gains, phases and Doppler shifts, this can be modelled given enough information on the transmitters, receivers, scatterers, reflectors and other physical channel parameters. Taken to the extreme, this is treated as a ray-tracing problem, aiming to replicate exactly the entire channel environment. If the model is close enough to reality, then the sinusoids can be extrapolated into the future, and a very close prediction of the channel at that point can be derived. The assumption that information on a device's surroundings is available is rarely realistic, and the computational requirements for such modelling can be significant with current approaches, further adding to the difficulty of using such models in real-world systems. As such, these systems are commonly found in simulations (where knowledge of the complete channel environment can be assumed), but are rarely employed to model real scenarios.

Some other work has also been carried out using so-called basis expansion techniques [126, 71]. These incorporate band-limited process model-based prediction algorithms that predict fading coefficients. This is a comparatively new area and results on real world channels have been mixed, seeming to be heavily dependent on channel type.

2.4 LTE

LTE has become the world's most widely deployed '4G' cellular radio system, and now accounts for a huge amount of data transmitted. To achieve the high data rates customers expect requires timely, accurate channel information in order to overcome the challenges of narrowband fast fading. This section briefly outlines some of the key technologies employed within LTE to meet consumer demands and highlights where additional channel information would improve the system's performance. Here LTE is used as an example of the specific challenges faced by many modern digital ratio systems.

OFDM

For the downlink channel, LTE uses OFDM, a digital multi-carrier transmission scheme which uses a large number of closely-spaced subcarriers to encode its data. The data in each subcarrier is encoded using a conventional modulation format (such as QAM) with a low symbol rate. The subcarriers are spaced in frequency at the reciprocal of the symbol rate, such that the nulls in the spectrum of one line up with the peaks of the adjacent subcarriers, meaning that there is effectively zero inter-symbol interference. It is from this property of subcarrier orthogonality that OFDM derives its name.

Because OFDM systems are composed of many narrowband signals, each subcarrier is subjected to frequency-selective fading. In order to accurately recover the data from each separate stream, channel information (whether estimated or real) must be used. This means that the amount of channel information increases with the number of subcarriers that comprise the whole bandwidth. Unless channel information is provided, the signal cannot be reliably decoded.

It should be noted that although the uplink uses a different transmission scheme (known as Single-Carrier Frequency Division Multiple Access (SC-FDMA)), it also stands to benefit in many of the same ways as the downlink.

OFDMA

An extension of OFDM used in LTE is known as Orthogonal Frequency Division Multiple Access (OFDMA). This scheme allows multiple users to dynamically share time and frequency resources according to each user's needs, meaning that the entire bandwidth

could be allocated to one intensive user, or shared among multiple lower-demand users. This is ideal in the cellular setting, as the bandwidth requirements for each user can change very rapidly.

Additionally, if intelligent scheduling can be combined with sufficient channel information, frequency-selective fading can be worked around for a particular user by allocating that user subcarriers which are not at that point in time being subjected to destructive interference. In practice, ensuring that the transmitter has sufficiently accurate channel information to allow it to make optimal decisions often requires a large degree of active channel measurement and extra signalling.

Transmission structure

With the possibility of dynamic resource allocation afforded by OFDMA, much more complicated radio frames can be constructed. The smallest unit of resources that can be allocated to one user in an LTE system is known as a resource block (see Fig. 2.3). Many of these resource blocks are then dynamically combined across time and frequency to encode more complex message types (those for user data, signalling, etc.).

The precise number of subcarriers and OFDM symbols in a resource block are determined by the size chosen for the cyclic prefix, however 2.3 shows a commonly used configuration. Each resource element represents a single OFDM symbol of modulated data. If all the resource elements in Fig. 2.3 used 64-QAM, then the resource block would represent 504 bits of un-coded data (the product of 12 subcarriers, 7 OFDM symbols and 6 bits per resource element).

Adaptive Modulation

As previously mentioned, one of the advantages of OFDM as a transmission scheme is that each of its subcarriers can be considered orthogonally, and this extends to channel effects. This means that each subcarrier can be thought of as being transmitted over its own narrowband channel with a flat fading profile. Appropriate channel information about each subcarrier allows the transmitter to use a high-order modulation technique (e.g. 64-QAM) over subchannels with a high SNR, while modulating data with lower-order techniques (e.g. BPSK) for sending over lower quality subchannels.

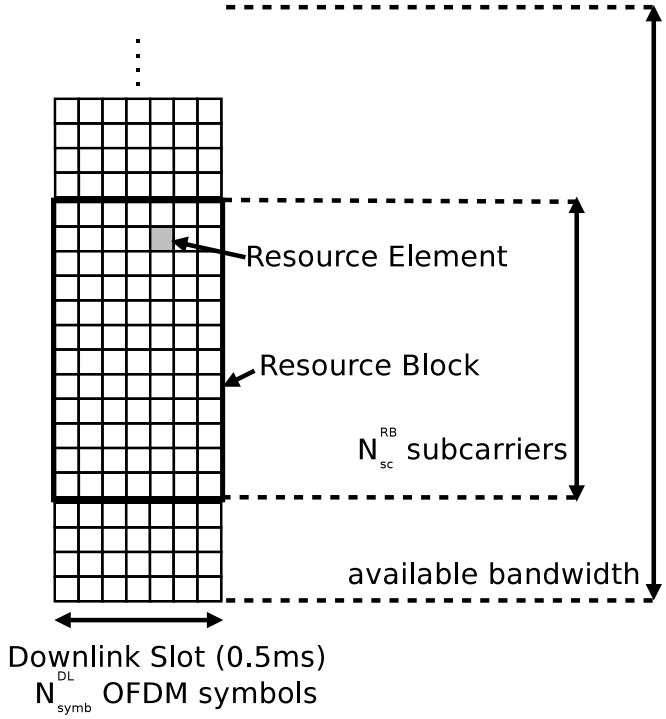


Figure 2.3: Diagram showing a downlink timeslot, with frequency along the vertical axis and time along the horizontal axis. The smallest unit which can be scheduled for transmission in an LTE system is known as a resource block. It is composed of N_{sc}^{RB} subcarriers and N_{symb}^{DL} OFDM symbols. Each of the smallest squares represents a single complex modulation datum, known as a resource element. Since resource elements can be modulated using different schemes (e.g. Binary Phase-Shift Keying (BPSK) or QAM, etc) transmissions can be matched to channel conditions.

Channel Measurement

To achieve optimal use of the shared bandwidth available to an LTE system, adaptive modulation and scheduling techniques have been developed which enable a larger number of users to concurrently share a very limited frequency resource. The nature of OFDM as a transmission scheme requires channel information to be provided in order to decode the data. Without estimates of phase shift and channel gain for each subcarrier, correct demodulation is impossible.

In order to obtain channel information, the LTE standard defines a regular pattern of pilot signals (referred to as reference signals) which are spaced in both time and frequency throughout the resource blocks. A typical scenario is shown in Fig. 2.4. Reference signals are also sent with a power 6 dB greater than the surrounding resource elements in order to make then easy to detect and demodulate.

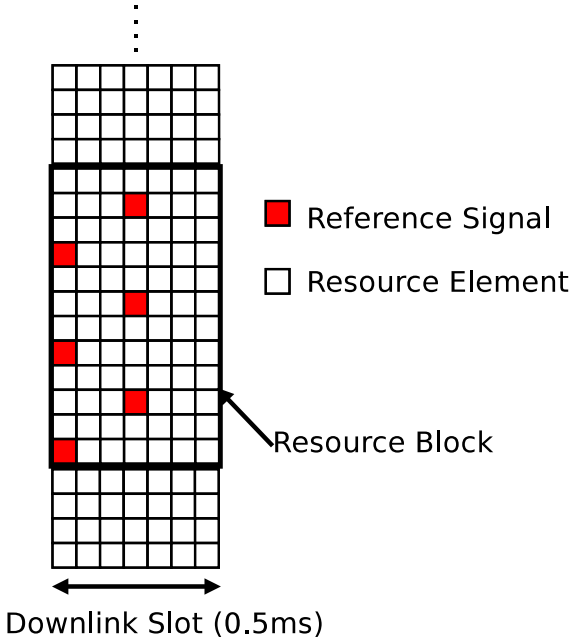


Figure 2.4: A diagram showing one possible arrangement of reference signals within a resource block. Using them for active channel measurement provides reasonably accurate information, however in this example 6 of the possible 84 resource elements within the block are unavailable for data. An additional disadvantage is that the reference signals are transmitted at a significantly higher power than the surrounding resource blocks which inevitably increases power usage.

These reference signals provide channel measurements only at the time and frequency when they are sent. In order to use the information they provide effectively, this data must be used to estimate the channel for other subcarriers and time instants to enable coherent decoding.

The LTE specification does not prescribe any particular algorithm or method for employing the data from these reference signals, but leaves it as an open problem to those building radios. This allows competing innovative approaches to be developed. The two most common implementation methods are linear and lowpass filter interpolation [51].

Linear Interpolation When using a linear interpolation scheme, the channel's response is estimated by assuming that phase and magnitude between adjacet pilot symbols changes linearly. The following equation is used to produce channel extimates:

$$\hat{H}(n) = \hat{H}(p_i) + \frac{\hat{H}(p_{i+1}) - \hat{H}(p_i)}{p_{i+1} - p(i)}.(n - p_i) \qquad \text{for} \qquad p_i \leq n \leq p_{i+1} \tag{2.10}$$

where $\hat{H}(n)$ represents an estimate of the channel's transfer function, and $\hat{H}(p_i)$ is a channel measurement derived from a pilot signal. Clearly this is a simple to implement method, though a large proportion of the symbols may need to be devoted to channel measurement in order for it to produce accurate enough results.

Lowpass Filter Interpolation A family of interpolators which use lowpass filters has been developed which aim to outperform the simple linear method previously outlined. Although they more closely model the nature of the channel, due to the practical implementation issues surrounding lowpass filters, only a limited set of sets are vaild. This can lead to poorer performance than the linear method in practise.

While such methods provide usable estimates, they still rely on the presence of regular chunks of the transmission bandwidth dedicated to channel measurement. If reliable channel prediction algorithms were developed however, many of these reference signals could be made redundant, freeing up the bandwidth and power they occupied for increased throughput or reduced power consumption.

2.5 Summary

This chapter has presented the historical and technical background relevant to the research presented in the following chapters. Firstly, anomaly detection was introduced, and the data mining, machine learning and statistical fields were outlined. More detail was then given about statistical approaches, and how this applies to RF signals. Section 2.2 then outlined neural networks, how they can be used as classifiers and predictors, and finished with a description and discussion of ESNs. Section 2.3 discussed the linked fields of channel modelling, characterisation and prediction, and outlined the most recent contributions in the literature. The chapter concluded with an overview of a subset of the LTE standard in order to illustrate the challenges faced by real-world digital radio systems.

The research described in this chapter is quite varied, but two common observations can be made. First, not all approaches are suited to real-time operation. Only those

approaches with reasonable prospect of real-time performance will be pursued further in this thesis. Second, it is repeatedly evident that many algorithms need to be adjusted on a rather *ad hoc* and undocumented basis by the researchers to suit a particular wireless context. While this is sometimes unavoidable, automatic adjustment of system parameters seems to be a highly desirable goal.

The following three chapters present original research in the areas introduced, beginning with anomaly detection in periodic wireless systems, followed by channel characterisation and lastly channel prediction.

Chapter 3

KLD-based Anomaly Detection in Periodic Wireless Signals

The research background for detecting anomalies in periodic wireless channels has been presented in Section 2.1, along with the relevant underlying theory and a description of some of the most recent and promising techniques that may contribute to improved system performance. Among the classes of anomaly detection methods, the Kullback-Leibler Divergence (KLD) has been identified as one of the most promising measures that may be used with periodic radio signals. This chapter extends existing work by investigating if additional information can be extracted from a received signal, and particularly whether anomalies can be characterised to provide more information about the type of anomaly detected, perhaps also providing additional information about the wireless channel, with potential to aid channel prediction.

The first two sections serve as background, detailing previous research on which the subsequent new work depends. In Section 3.1, the contributions by Afgani *et al.* [8, 7, 5] on anomaly detection in periodic wireless signals will be described. The way in which estimates of a system's current Probability Mass Function (PMF) are generated, and how this can be used within an anomaly detection scheme (see Chapter 2) will be discussed in Section 3.2. This section will also detail how the approach in [8, 7, 5] is implemented.

New work will the be presented in the following sections. Section 3.3 will present the observations of monotonic sequences, develop a new hypothesis about what their presence may infer, and discuss the possibility and potential advantages of anomaly characterisation. Section 3.4 will consider the results in detail and analyse monotonic

sequences by decomposing the KLD equation. Based on this, a proof of the mathematical causes of sequences that indicate high anomaly likelihood will be discussed. Section 3.5 will show how simulations can be used to both verify the mathematical model and to accurately model real-world observations. An hypothesis for how such monotonic sequences can occur is also developed and tested in computer simulation. Finally, section 3.6 will discuss the key findings and limitations of the presented work. It will reflect on the value of the KLD-based anomaly detection observations and simulations with respect to coping with signal anomalies in real-world and real-time mobile communication systems, and comment on the suitability of the KLD as a characterisation tool for anomalies.

3.1 Introduction

Radio systems such as Bluetooth, WiFi and a range of cellular technologies (Global System for Mobile Communications (GSM), Universal Mobile Telecommunications System (UMTS), Long Term Evolution (LTE) etc.) are ubiquitous, enabling a host of useful communication services. Many of these technologies send and receive radio packets on a fixed, regular time base. Such signals have a frame structure which repeats on a defined time period, t_p (referred to as the signal period), although each frame carried unique data. This is illustrated in Fig. 3.1.

Afgani *et al.* [4, 5] exploited this property of periodicity within the radio signal to allow anomaly detection without the need for demodulation, working only on the baseband signal's envelope. It is a particularly attractive method as it requires no knowledge of the higher-level protocols used, with the only information required in advance being the signal's period. It was found that of a number of possible statistical similarity functions, the KLD gave the best performance and lowest false-positive probability. The following two sections describe previous work carried out by Afgani *et al.* in order to provide context for the new research which follows it.

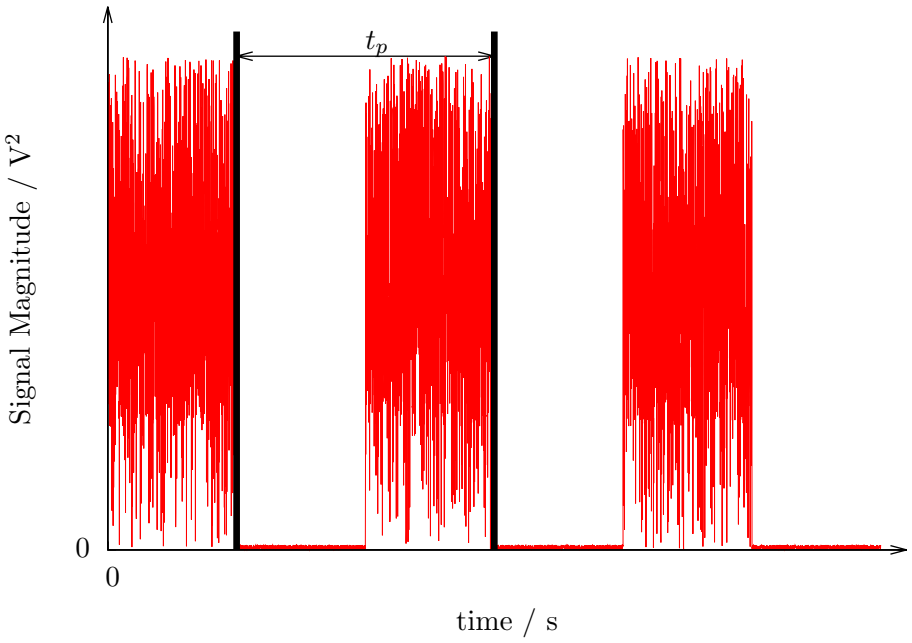


Figure 3.1: A graph of signal magnitude showing three distinct data packets, each carrying unique data. The signal period t_p is labelled.

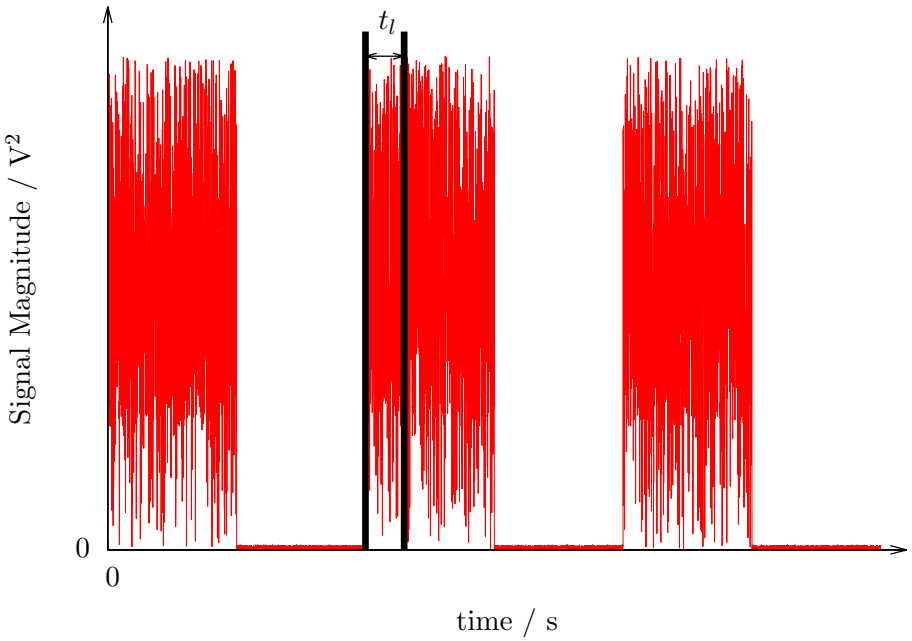
3.2 General Experimental Setup

By analysing a received radio signal over a short time period (often referred to as a window), Afgani *et al.* were able to construct a histogram to represent the distribution of the signal's power. Afgani chose the PMF as the parameter by which to measure a system for a number of reasons. In a periodic system, the overall shape of the PMF should remain consistent from one frame to the next, so a measurement of the shape of the PMF would allow a system to detect when there had been some major change to the probability distribution of the underlying signal. It has the added advantage that it is both quick and simple to construct, leading to a less complex system which does not require a large amount of processing power.

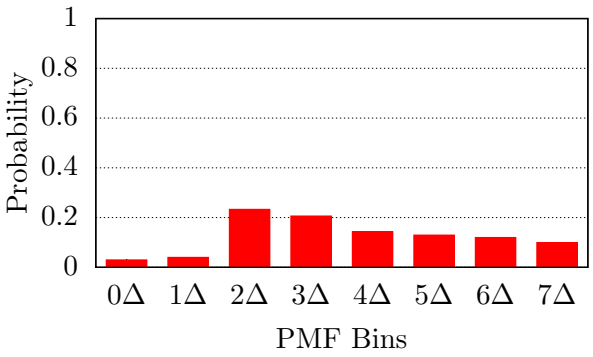
Fig. 3.2 shows such a histogram, which sorts the power into a discrete number (in this case, eight) of 'bins'. Since a PMF corresponds to a direct measurement of the signal's PMF over this time period, the histogram is normalised to ensure the sum of the areas of the bins is equal to one.

The number of frequency bins chosen affects the system's performance: if few bins are chosen, the PMF becomes insensitive to small but possibly relevant changes in the signal's power spread, while having more bins increases the system's susceptibility to

transient noise. The optimal number of bins for a minimum error criterion has been studied [108] and should provide a balance between resolution and sensitivity.



(a) Time series showing the sampling window

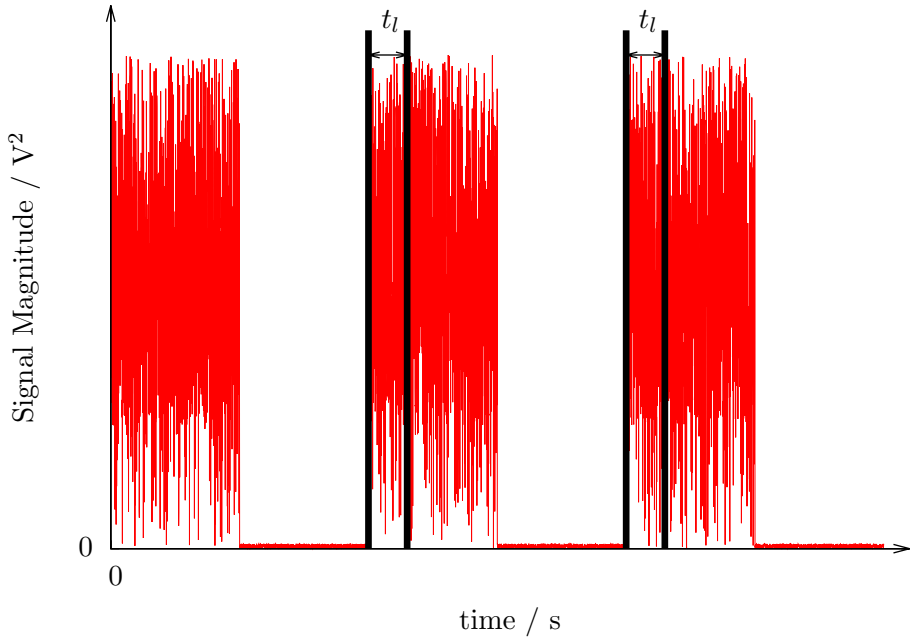


(b) Normalised histogram showing the spread of signal power in the window interval, t_l

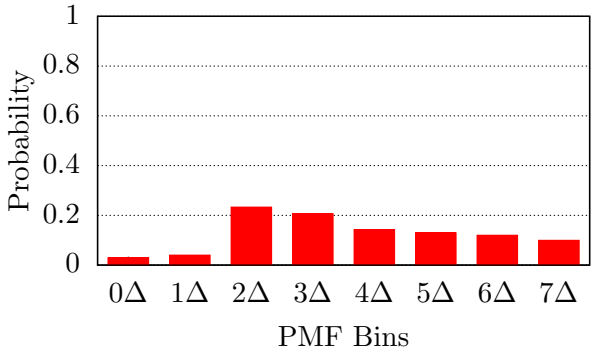
Figure 3.2: By analysing the signal over a time period t_l, a histogram of signal power values can be generated. Note that t_l can be much smaller than the signal period, t_p

By using two sampling windows separated in time by one signal period, t_p, corresponding points in the periodic signal can be compared. This is illustrated in Fig. 3.3. If a signal is periodic, then its statistical properties will remain relatively static over time, meaning that the PMF for a particular interval in one packet should look very similar to the PMF for the same interval in a subsequent packet. Thus the two PMFs

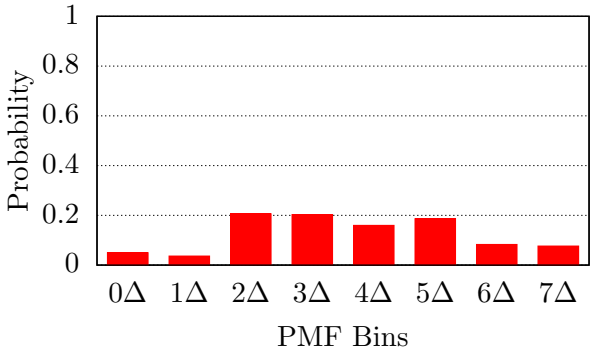
would be expected to look similar under normal signal conditions.



(a) A typical periodic signal showing the two sampling windows



(b) Power histogram for first sampling window



(c) Power histogram for second sampling window

Figure 3.3: By having two windows of width t_l, two histograms can be created, and then compared. If they are separated by the signal period t_p (as in this example) corresponding parts of the repetitive signal are being compared to each other.

It is then possible to quantify the difference between the two PMFs by using some statistical distance or similarity measure. Five candidate methods are outlined in the following section.

3.2.1 Statistical PMF Comparison Methods

By assuming that one packet is by some definition similar to another under normal circumstances, anything else can be classified as anomalous. By comparing two PMFs, their similarities can be characterised and used to detect anomalous signal conditions. There is a range of different statistical measures of similarity which can be employed for this purpose. In each case, two different PMFs, p and q are referred to, with X bins each.

L1 Distance

The L1 distance (variously also known as the Manhattan distance, rectilinear distance or city block distance) is defined as:

$$\sum_{x \in X} |p(x) - q(x)| \tag{3.1}$$

In the case of comparing power PMFs, it corresponds to the sum of the absolute differences between corresponding histogram bins. It is very simple to compute, and was investigated by Afgani *et al.* [8] as an alternative to the KLD algorithm. It was passed over as a choice however because the L1 distance for very dissimilar PMFs is much smaller than the KLD between the same PMFs.

L2 Distance

The L2 distance (also known as the Euclidean distance) is defined as:

$$\sqrt{\sum_{x \in X} (p(x) - q(x))^2} \tag{3.2}$$

This is a commonly used general metric, and is what is usually meant informally by 'distance'. Its performance is much like the L1 distance, where very dissimilar PMFs produce a smaller value of L2 distance than with the KLD. It has slightly higher runtime complexity than the L1 distance.

Mean Squared Difference

The mean squared difference metric compares corresponding histogram bins:

$$\frac{1}{n}\sum_{x\in X}(p(x)-q(x))^2 \tag{3.3}$$

It is the sum of the squares of the differences between corresponding bins in the two PMFs. Very different PMFs will produce a large difference.

Variance

Variance is a measure of deviation from the signal's mean value, quantifying the spread of the signal's power. This means that vastly dissimilar PMFs could easily give the same values for variance. For example, a PMF with a full bin 0 will have the same variance as one with any other full bin. This also requires the separate computation of μ_p and μ_q, the mean of the PMFs p and q.

$$\left|\frac{1}{n}\sum_{x\in X}(p(x)-\mu_p)^2-\frac{1}{n}\sum_{x\in X}(q(x)-\mu_q)^2\right| \tag{3.4}$$

The Kullback-Leibler Divergence

The KLD (also known as relative entropy or mutual information) is defined [66] as:

$$D(p||q)=\sum_{x\in X}p(x)\log_n\frac{p(x)}{q(x)} \tag{3.5}$$

where p and q are probability distributions composed of samples from the domain X, using a base-n logarithm. It is formally described by Cover *at al.* below:

> The *relative entropy* $D(p||q)$ is a measure of the inefficiency of assuming the the distribution is q when the true distribution is p. For example, if we knew the true distribution p of the random variable, we could construct a code with average description length $H(p)$. If, instead we used the code for a distribution q, we would need $H(p)+D(p||q)$ bits on the average to describe the random variable. [27]

Unlike several of the previous algorithms outlined, the KLD is not a distance metric, since generally $D(p||q) \neq D(q||p)$. Any logarithm base n can be used, however using $n = 2$ results in a divergence measured in bits. In this setting, the KLD can be used as a function for comparing the difference in information content between two datasets. Two similar sets will have a small KLD, while very different sets would have a larger KLD. It is this comparison operation which makes it well suited to anomaly detection in periodic wireless signals, since correctly formed frames should have a similar information content.

3.2.2 The KLD Method

The work done by Afgani *et al.* [8, 7, 5] has shown the KLD to be both a convenient and effective anomaly detection measure for use with periodic signals. It quantises the difference between two probability distributions to a single, positive real value which can be used to make decisions. It was chosen in preference over the previously outlined measures (notably the L1 distance, (3.1), belonging to the same class of distance measures as the KLD) for sensitivity to anomalies and greater immunity from normal signal noise. A diagrammatic overview of their experimental flow is shown in Fig. 3.4.

In their work, the KLD is used in conjunction with a threshold detector to find anomalies in periodic wireless signals. In this case, the system designer uses personal judgement to define a value of KLD which is considered to be greater than values typically expected from normal signals, and when the KLD exceeds this value it is deemed that an anomaly has occurred. It should be noted that the KLD is almost never exactly zero, but fluctuates very slightly since the two PMFs are unlikely to be identical (due to both noise and differences in the data content of differing frames).

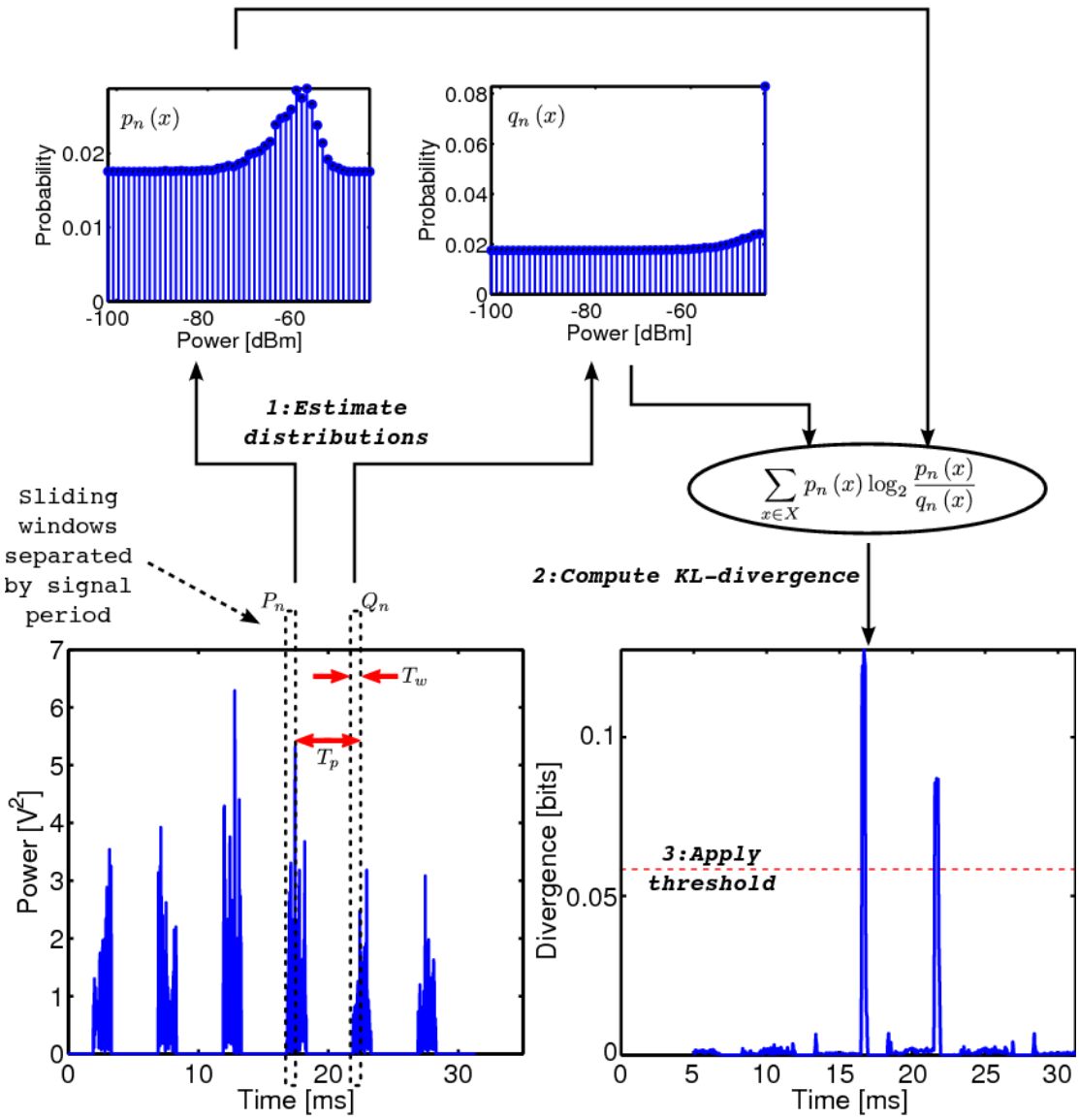


Figure 3.4: An overview of the KLD anomaly detection method developed by Afgani *et al.* It shows time window sizing and separation, PMF estimation, KLD calculation and the threshold method for detecting anomalies in the signal. (Source: [8], used by permission)

3.3 Observations of Monotonic Sequences

During the research by Afgani *et al.* [4, 5, 8], a number of cases were observed where the KLD output would change in a monotonic manner over a large number of samples, in stark contrast to its usual fluctuating behaviour. Fig. 3.5 shows both normal fluctuating behaviour and several monotonic sequences. Every time such a sequence occurred in tests, it coincided with a known anomaly, though not all anomalies coincided with such a feature. It was suggested that this distinction might indicate the presence of a certain

class of anomaly, and that this information could be exploited for the purpose of channel prediction, however this investigation was not further examined by Afgani *et al.*

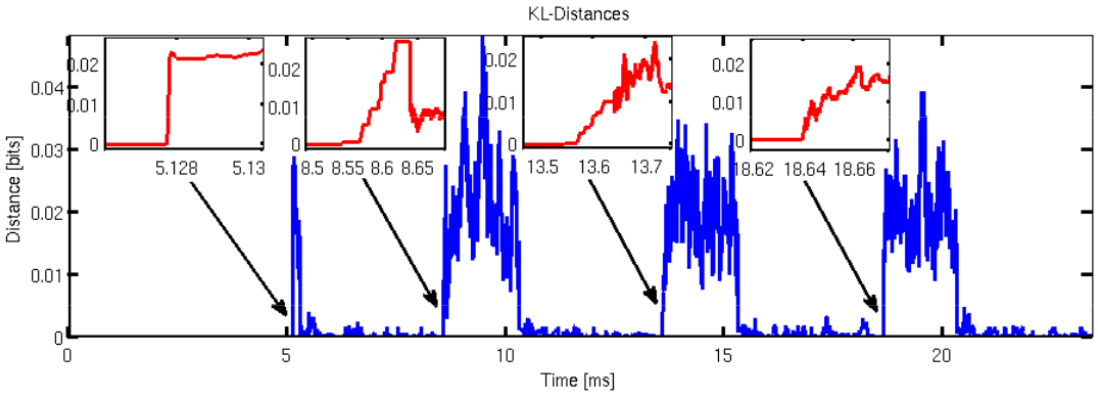


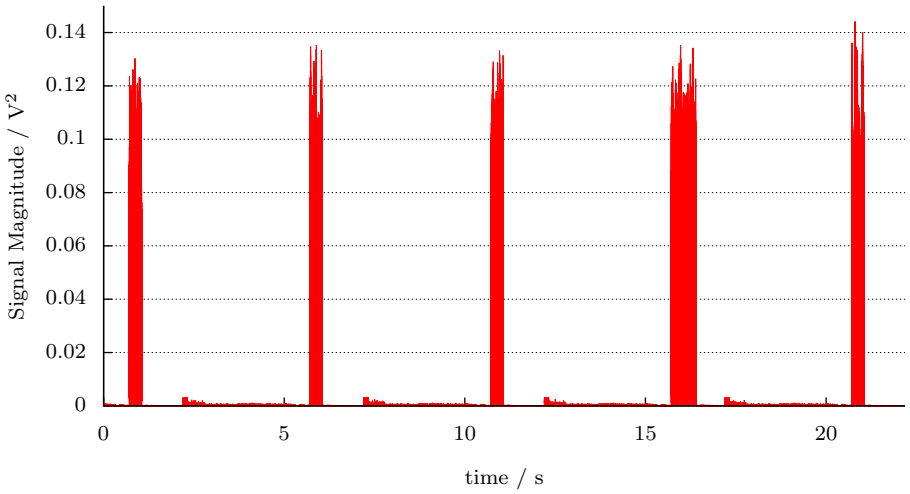
Figure 3.5: A plot of KLD versus time, showing four anomalies within a signal, with only the second showing strictly monotonic increase. It was hypothesised that this different behaviour might indicate something about the nature of the detected anomaly. (Source: Unpublished presentation by Mostafa Afgani, October 2010)

If the KLD method were capable of distinguishing among different specific types of anomalies with different causes (a task known as anomaly characterisation) this would open up a range of benefits. In such a system, further information about the wireless channel could be gleaned from the received signal, and corrupted transmissions could be of value in diagnosing problems. Thus adaptive modulation and coding methods could be employed to intelligently make use of the improved channel information coming from the anomaly characterisations. If the system could distinguish between anomalies caused by fading and those caused by interference from a competing transmission, it would easily find broad applicability in the field of Cognitive Radio (CR) where detrimental interference between radios signals on the same frequency must be avoided. The more specifically such a system could characterise a signal feature, the more useful it would be.

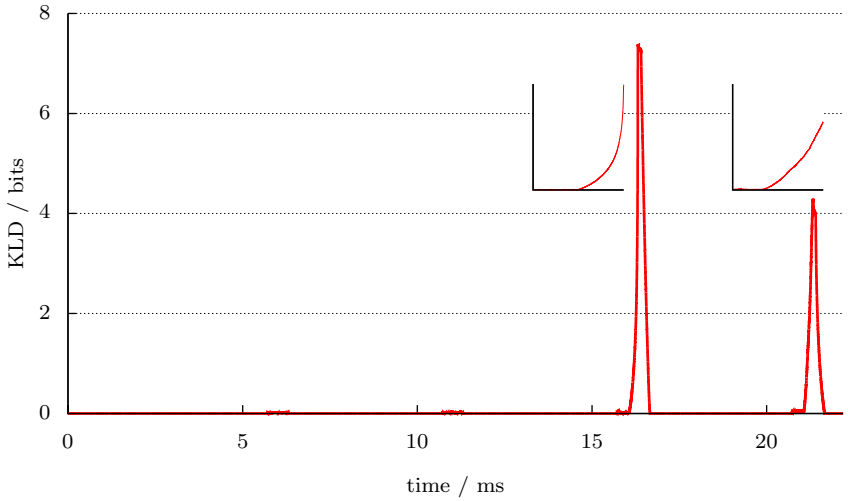
During examination of a mobile Worldwide Interoperability for Microwave Access (WiMAX) signal which included a known anomaly, several instances where the KLD increased monotonically for a large number of samples were observed [5]. The anomaly can be seen on the time series (Fig. 3.6a), where the fourth frame is abnormally long, and was diagnosed as being caused by a malfunctioning transmitter.

Each time these long sequences were observed, they coincided with an anomaly in the wireless signal. Fig. 3.6 shows the Radio Frequency (RF) data, and how the KLD increases over time. It should be noted that in a KLD system, an anomaly will

always create two peaks in the KLD: one as the anomalous frame is compared with the preceding correct frame, and a second as the next (presumably correct) frame is compared with the anomalous frame.



(a) Time series data of observed signal. The anomaly can be spotted pictorially, as the fourth frame is abnormally long, and was caused by a malfunctioning transmitter.



(b) Plot of KLD with inset graphs highlighting its monotonic growth.

Figure 3.6: Time series of signal magnitude and KLD, taken from a real-world WiMAX recording by Afgani from a malfunctioning transmitter. The KLD is shown to be very small during the period of normal frames, spiking dramatically when an anomalous frame is detected. The strictly monotonic increase in KLD is interesting, as it is only present for some anomalies and not others. This distinction is later investigated. (Source: RF data provided by Afgani, used by permission)

The following section presents new research which investigates the causes of the

monotonically increasing KLD, and demonstrates the suitability of the KLD method for anomaly characterisation.

3.4 Monotonic Sequence Analysis

By observing how the two PMFs p and q change over time, it became clear that one PMF's mass would shift, concentrating all the data points into one bin, while the other PMF would retain its previous evenly distributed characteristic. Since these two PMFs are radically different, clearly a large KLD would be expected. However, it was not immediately clear why the KLD would exhibit this monotonically increasing characteristic over time.

In order to investigate this behaviour, the KLD calculation itself is examined.

3.4.1 The General Case

At each time step, the time windows move forward one data sample, meaning both windows admit one new reading to their front edge and remove the oldest reading from their back edge. The updated histograms p and q are calculated by adding the new samples to the histograms and removing the old ones. This results in 4 different bins being modified:

- a is the bin in histogram p which is increased
- b is the bin in histogram p which is decreased
- c is the bin in histogram q which is increased
- d is the bin in histogram q which is decreased

Since only these 4 bins will change, by simply expanding the KLD equation (3.5), the difference between two successive KLD calculations can be calculated. $p_{\mathrm{n}-1}(x)$ and $q_{\mathrm{n}-1}(x)$ refer to the old PMFs in the preceding time step, while $p_{\mathrm{n}}(x)$ and $q_{\mathrm{n}}(x)$ refer to the new PMFs in the current time step.

$$
\begin{aligned}
&\mathrm{D}_{\mathrm{n}-1} - \mathrm{D}_{\mathrm{n}} = \\
&p_{\mathrm{n}-1}(a) \log \frac{p_{\mathrm{n}-1}(a)}{q_{\mathrm{n}-1}(a)} - p_{\mathrm{n}}(a) \log \frac{p_{\mathrm{n}}(a)}{q_{\mathrm{n}}(a)} + \\
&p_{\mathrm{n}-1}(b) \log \frac{p_{\mathrm{n}-1}(b)}{q_{\mathrm{n}-1}(b)} - p_{\mathrm{n}}(b) \log \frac{p_{\mathrm{n}}(b)}{q_{\mathrm{n}}(b)} + \\
&p_{\mathrm{n}-1}(c) \log \frac{p_{\mathrm{n}-1}(c)}{q_{\mathrm{n}-1}(c)} - p_{\mathrm{n}}(c) \log \frac{p_{\mathrm{n}}(c)}{q_{\mathrm{n}}(c)} + \\
&p_{\mathrm{n}-1}(d) \log \frac{p_{\mathrm{n}-1}(d)}{q_{\mathrm{n}-1}(d)} - p_{\mathrm{n}}(d) \log \frac{p_{\mathrm{n}}(d)}{q_{\mathrm{n}}(d)}
\end{aligned} \tag{3.6}
$$

Clearly this difference must be positive over a number of successive readings to produce a monotonic sequence. However, each of the terms in (3.6) can be either positive or negative, depending on the state of the individual bins in p and q. Thus there is no simple criterion for ensuring that this difference is kept positive without previous knowledge of the states of the PMF bins.

3.4.2 Simplified General Case

If the assumption is made that a, b, c and d are all different from each other, then a simplified version of (3.6) can be derived. In this case, $p_{\mathrm{n}-1}, q_{\mathrm{n}-1}$ and $p_{\mathrm{n}}, q_{\mathrm{n}}$ can be replaced using the following substitutions, simply using the definitions of a, b, c and d from Section 3.4.1. If the four affected bins are not unique, then this simplification cannot be made. In experiments it is rare for this property of uniqueness to occur over a long time series, and making this assumption greatly simplifies the analysis, and exposes a more elegant expression for KLD calculation which yields extremely similar experimental results to the unsimplified version (3.6).

$$
\begin{aligned}
p_{\mathrm{n}}(a) &= p_{\mathrm{n}-1}(a) + 1 & \qquad q_{\mathrm{n}}(a) &= q_{\mathrm{n}-1}(a) \\
p_{\mathrm{n}}(b) &= p_{\mathrm{n}-1}(b) - 1 & \qquad q_{\mathrm{n}}(b) &= q_{\mathrm{n}-1}(b) \\
p_{\mathrm{n}}(c) &= p_{\mathrm{n}-1}(c) & \qquad q_{\mathrm{n}}(c) &= q_{\mathrm{n}-1}(c) + 1 \\
p_{\mathrm{n}}(d) &= p_{\mathrm{n}-1}(d) & \qquad q_{\mathrm{n}}(d) &= q_{\mathrm{n}-1}(d) - 1
\end{aligned}
$$

By substituting the above terms into (3.6), much simplified expression for the change

in KLD is developed.

$$
\begin{aligned}
&\mathrm{D_n} - \mathrm{D_{n-1}} = \\
&p_{\mathrm{n-1}}(a) \log \frac{(p_{\mathrm{n-1}}(a)+1)}{p_{\mathrm{n-1}}(a)} + \log \frac{(p_{\mathrm{n-1}}(a)+1)}{q_{\mathrm{n-1}}(a)} + \\
&p_{\mathrm{n-1}}(b) \log \frac{(p_{\mathrm{n-1}}(b)-1)}{p_{\mathrm{n-1}}(b)} - \log \frac{(p_{\mathrm{n-1}}(b)-1)}{q_{\mathrm{n-1}}(b)} + \\
&p_{\mathrm{n-1}}(c) \left(\log \frac{q_{\mathrm{n-1}}(c)}{(q_{\mathrm{n-1}}(c)+1)} \right) + p_{\mathrm{n-1}}(d) \left(\log \frac{q_{\mathrm{n-1}}(d)}{(q_{\mathrm{n-1}}(d)-1)} \right)
\end{aligned} \tag{3.7}
$$

From this form of the expression it is clear that there are two terms which are always positive, and two which are always negative. This means the state of PMF p (and particularly the state of the four modified bins within it) will have the dominant impact on the change in KLD. From this, it is clear that the fuller a modified bin is, the larger its contribution to the change.

3.4.3 Further Simplification

The general proof in Section 3.4.1 offers little insight into how the monotonic behaviour occurs, while Section 3.4.2 simplifies the equation.

Below is a further simplification of (3.7) which should make the cause of monotonic behaviour clearer. The further assumption made in this instance (as well as c and d being unique) is that the PMF p remains unchanged from one time step to the next. This assumption has been observed to hold true for long time periods in a number of test cases (such as in the recording shown in Fig. 3.6). It occurs most often in the inter-frame gap in a periodic signal, or when there is no active transmission happening on the channel.

$$
\begin{aligned}
&\mathrm{D_n} - \mathrm{D_{n-1}} = \\
&p_{\mathrm{n-1}}(c) \left(\log \frac{q_{\mathrm{n-1}}(c)}{(q_{\mathrm{n-1}}(c)+1)} \right) + p_{\mathrm{n-1}}(d) \left(\log \frac{q_{\mathrm{n-1}}(d)}{(q_{\mathrm{n-1}}(d)-1)} \right)
\end{aligned} \tag{3.8}
$$

This can be re-arranged to the following form. In order for monotonic behaviour to occur, the following inequality must hold over a number of successive calculations:

$$\left| p_{\mathrm{n}-1}(c) \left(\log \frac{q_{\mathrm{n}-1}(c)}{(q_{\mathrm{n}-1}(c)+1)} \right) \right| \leq \left| p_{\mathrm{n}-1}(d) \left(\log \frac{q_{\mathrm{n}-1}(d)}{(q_{\mathrm{n}-1}(d)-1)} \right) \right| \tag{3.9}$$

In the above equation, since the differences between the numerator and denominator in the log terms will usually be small, the value of the log terms will be close to 0. The value of the log terms will grow larger as the bin it refers to becomes emptier (smaller). Clearly the dominant terms in the inequality are the scaling factors of $p(c)$ and $p(d)$.

3.4.4 Worked Example using Theory

By examining the PMFs p and q at different points in time during the monotonic sequences generated from real-world observations, it was observed that a single bin would become very full in one PMF, while the same bin in the other PMF remained at a lower level. The following example shows how this observed behaviour could provide the monotonic sequences observed.

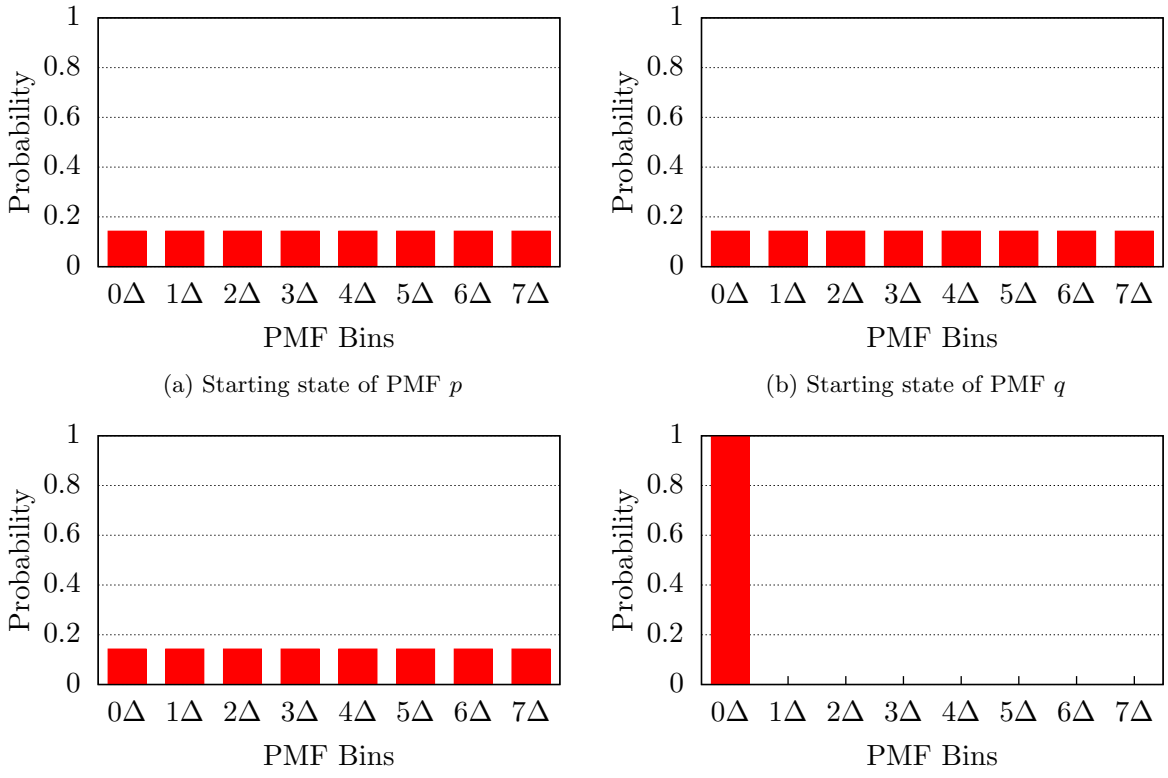
The system's two PMFs, $p(x)$ and $q(x)$, each have n bins, numbered 0 to $n-1$. For this artificial example, both $p(x)$ and $q(x)$ begin with each bin filled equally (i.e. a uniform distribution). PMF $q(x)$ will tend towards a full bin number 0, and empty bins 1 to $n-1$, while $p(x)$ will not change with time. This is shown graphically in Fig. 3.7. Bins 1 to $n-1$ in $q(x)$ are depleted evenly, such that every $n-1$ time steps bin 0 will have grown by $n-1$ samples, and bins 1 to $n-1$ have each one sample less.

Since it is assumed that PMF p is time-invariant and each bin in p is always equal to $\frac{1}{n}$, the KLD calculation can be simplified to:

$$\begin{aligned} \mathrm{D} =& p(x) \sum_{x=0}^{n} \log p(x) - \log q(x) \\ =& \frac{1}{n} \sum_{x=0}^{n} \log \left(\frac{1}{n}\right) - \log q(x) \end{aligned} \tag{3.10}$$

By introducing a time step variable, t, the value of each bin in the changing PMF q can be calculated using:

$$q(x) = \begin{cases} \frac{1}{n} + (n-1)t & : 0 \\ \frac{1}{n} - t & : 1, \ldots, n-1 \end{cases} \tag{3.11}$$



(a) Starting state of PMF p

(b) Starting state of PMF q

(c) Finishing state of PMF p

(d) Finishing state of PMF q

Figure 3.7: It is clear that PMF q, although initially identical to p, becomes radically different by the end time.

For this system, t ranges from 0 to $\frac{1}{n}$.

When this PMF equation is expanded and substituted into (3.10), the value of the KLD at any time between 0 and $\frac{1}{n}$ can be calculated:

$$\mathrm{D} = \frac{1}{n}\Big(\underbrace{n \log \frac{1}{n}}_{p(0),\ldots,p(n-1)} \underbrace{- \log\left(\frac{1}{n} + (n-1)t\right)}_{q(0)} \underbrace{- (n-1) \log\left(\frac{1}{n} - t\right)}_{q(1),\ldots,q(n-1)} \Big) \tag{3.12}$$

(3.12) does indeed show monotonic behaviour over the time interval $0 \leq t < \frac{1}{n}$, and is shown in Fig. 3.8.

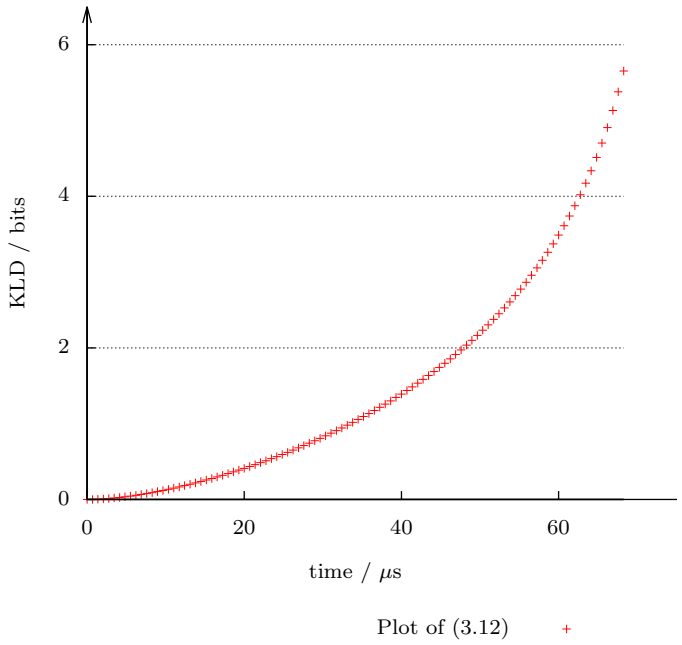


Figure 3.8: A plot of 3.12 demonstrating clear monotonic behaviour.

3.5 Simulation

Although the example in Section 3.4.4 demonstrates monotonicity, the idea that a real signal would deplete the PMF bins of q perfectly evenly is a convenient, though unrealistic assumption. In this further example a system similar to that in the previous example is simulated, but the requirement that the bins be depleted evenly is removed, making it a much more realistic example.

In this simulation, the two PMFs p and q begin identically, (just as in Section 3.4.4) with each of the 8 bins filled equally. As time progresses, p remains unchanged, while q tends towards a full first bin. At each timestep one of the other seven bins is chosen at random and its count value is decremented, while the first bin's count value is incremented until all the other bins are empty. The start and end states of both PMFs will be the same as shown in Fig. 3.7. The plot of the KLD between the two PMFs is shown in Fig. 3.9.

This first simulation (see Fig. 3.9) agrees very closely with that predicted by the theory in (3.12) with only a small deviation due to the discrete randomised nature of the bin emptying.

In order to make the simulation resemble real world conditions as closely as possible,

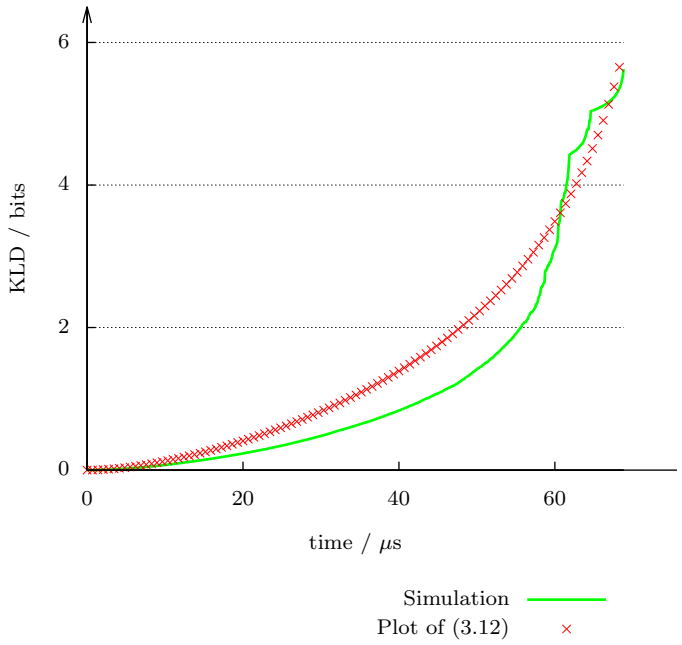


Figure 3.9: A plot of the KLD over time from the simulation, assuming bins in q are randomly depleted, (as would be likely to be found in a real signal capture) while keeping p unchanged. Also shown is the predicted data from (3.12), showing that the results from the simulation agree with those obtained from the theoretical approach take in Section 3.4.4.

the assumption that the PMF p remained entirely unchanged for the second simulation is relaxed. A low-level white noise signal is applied to p, causing its bins to change very slightly at each time step. This is exactly what can be seen when the distribution of power is even in the signal, (which is the same model used for p, as is evident in its histogram in Fig. 3.7). Thus the PMF will not change its overall shape significantly, but will change slightly at each time step, maintaining its overall shape with time. The PMF q is treated exactly the same as in the previous simulation.

This final simulation reveals that when the assumptions made in the example in Section 3.4.4 are replaced with real-world conditions, the behaviour of the system remains largely unchanged. This test introduces a more realistic set of circumstances under which monotonic increase can be achieved.

Although these simulations represent a particular scenario, since they describe the scenario which was theorised to produce monotonic time series, they are good validation of the equations developed.

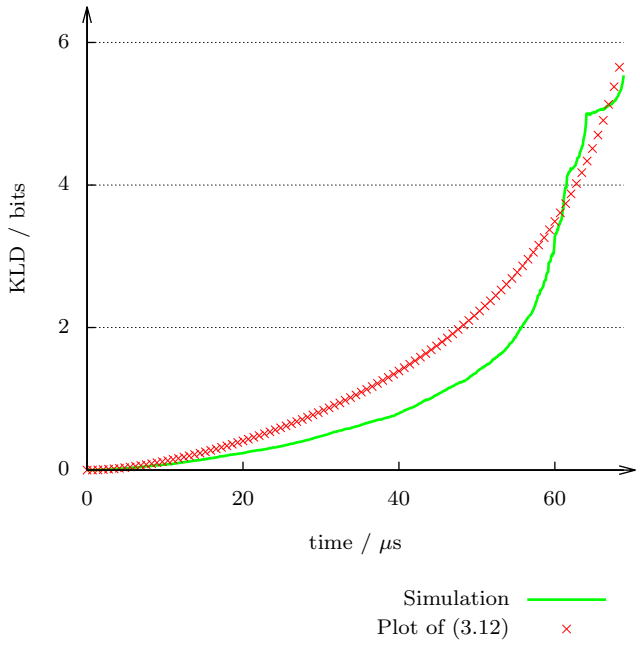


Figure 3.10: A simulation of a system with a randomly varying p, as well as a changing q. This result is almost identical to both the previous example, and the outcome predicted in 3.4.4

3.6 Results

The preceding derivations show that monotonic sequences are caused by a shifting of mass of one PMF with respect to the other, as the example in Section 3.4.4 illustrates.

As is highlighted in (3.9) and (3.7), the change in KLD from one time step to the next is heavily dependent on the absolute value of only 4 PMF bins, and there will always be two pairs of terms of opposing sign. Since there is no way to eliminate the negative terms entirely in the calculation, there is no trivial way of ensuring an always positive result, producing monotonic behaviour. Without *a priori* knowledge of the PMFs, the sign of the change in KLD cannot be determined. It has however been possible to produce a model for monotonic sequences which agrees very closely with observations from a variety of real-world signals.

Since the PMFs which cause such behaviour have a wide range of causes, it is difficult to attribute any particular class of anomalous events to such a shift of mass. However it can be said that the radio signal which generates such a PMF has become more deterministic and has a lower information content, which may offer some limited insight into the problem. This result agrees with the observations in Section 3.4.3, which

identifies no simple way of guaranteeing monotonic behaviour.

While the KLD method has previously been demonstrated as being very effective at detecting anomalies in periodic wireless signals, this work has demonstrated that it is not capable of characterising anomalies into narrow types.

3.7 Summary

The KLD method had been shown by Afgani *et al.* [8, 7, 5] to be highly effective at anomaly detection in periodic signals, and the current investigation has explained the occurrence of previously observed monotonic sequences.

A discussion of the causes of monotonic sequence events in a KLD-based anomaly detection system has been presented in this chapter. A new proof of the mathematical causes of these sequences has been demonstrated, along with simulation examples verifying the proof. An hypothesis for how such monotonic sequences can occur has been presented and it has been established that this agrees very closely with results in observed signals.

This chapter's contribution demonstrates that such signal features are not indicative of any narrow classification of anomalous events (as had been hypothesised), but of a broader class of anomalies characterised by a drop in the information content of the signal. This feature renders the KLD unsuitable for the additional task of anomaly characterisation (or determining what caused the anomaly) - a task which would be very valuable in real-world applications, especially in emerging technologies such as CR. It also fails to provide any additional information about the wireless channel over which the transmission was sent.

As a continuation of the search for improved channel knowledge, the following chapter will examine the problem of characterisation of wireless channels using Echo State Network (ESN)s.

Chapter 4

Radio Channel Characterisation using Echo State Networks

The previous chapter discussed anomaly detection within radio systems, the particular case of monotonic sequences when using the Kullback-Leibler Divergence (KLD) as a detection algorithm, and the possibility of anomaly characterisation. As the KLD method was found to be unsuitable for the job of characterisation, the research in this chapter is to characterise wireless channels using Echo State Network (ESN)s as a stepping stone towards channel prediction. In an earlier chapter (see Sections 2.2 and 2.3) the relevant background has been presented for the ESN approach and channel characterisation. Section 4.1 will provide the context and motivation for the research in this chapter. The reasons for choosing the echo state method are discussed in Section 4.2, and a detailed description of the experimental setup is then given in Sections 4.3. The results are discussed, along with their strengths and limitations in Section 4.4, where they are also compared with a conventional statistical characterisation approach. Finally, the chapter is summarised in Section 4.5 with an overview of the main findings.

4.1 Introduction

The ability to detect channel characteristics and use this information to select the best transmission scheme is core to maximising throughput in modern digital wireless systems. One cause of apparent anomalies is a signal which has been incorrectly decoded due to inaccurate or out of date channel information. If more information is available about the channel, receivers can more accurately decode a transmission, resulting in

fewer apparent anomalies. As a prelude to predicting a channel's future behaviour, this chapter addresses the challenge of channel characterisation. This is however a very complicated issue, as high-bandwidth channels (such as those used in today's and tomorrow's cellular systems) are subject to frequency selective fading, and extremely rapid variations, especially when devices are mobile.

To tackle this problem, a number of approaches have been devised to measure and characterise these channels, so that the most effective transmission schemes can be used to best exploit their characteristics. Often these approaches involve active measurement using pilot signals, which although providing good channel information, can be very resource intensive. The channel estimation and related signalling overhead in Long Term Evolution (LTE) deployments can use 30% [102] of all the available time and frequency resources which could otherwise be used to increase throughput.

Another range of classifiers (more fully discussed in Section 2.3) examine the channel's history and try to determine the type of wireless environment in which the device is operating. These local wireless characteristics can then be used to apply a pre-computed best-fit solution for the transmission parameters and determine optimal resource allocation for those conditions [43]. This data could also be used at the receiver to improve decoding accuracy.

This second approach makes fewer assumptions about the channel than the first, and is the subject of this investigation. The ultimate goal of this research is to predict a channel's future behaviour by examining its past. Doing so would potentially allow for vastly improved resource allocation, a reduction in channel estimation overhead, (often a limiting factor in more sophisticated transmission techniques such as Multiple Input, Multiple Output (MIMO) and Coordinated multipoint (CoMP)) and maximised spectral and energy efficiency. It is hoped that by investigating the more straightforward area of channel characterisation, insight may be gained which would enable new methods for channel prediction to be developed.

For this task, the aim is to correctly identify which previously observed scenario an unknown channel response most closely resembles. For example, it would be advantageous to be able to detect that the current observed channel response closely resembles previously observed data taken in an urban environment, allowing the device to infer it is in a built-up setting. This information could then be used by other parts of the radio chain to make decisions about power control and resource allocation. One of the large benefits of this type of system is that it operates without needing explicit signalling or

a feedback channel.

4.2 Echo State Network Suitability

Neural Network (NN)s are known for their ability to process arbitrary input sequences, and they were chosen for this research in the hope that they could extract information from a channel's magnitude response which would otherwise not be detected. They have the advantage that they do not require a detailed model of how a system works, but attempt to determine the relationships between inputs and outputs automatically. In the case of channel modelling, this is a hugely useful property, as developing an exact system model is almost impossibly complex.

NNs in general have the property of being able to detect patterns which are hidden, whether by noise, errors or simply the sheer volume of data to process, and there are many variations with differing conceptual perspectives and assumptions to choose from. Recurrent Neural Network (RNN)s are generally more useful than feed-forward NNs when the amount of historical data to be examined is non-trivial, since feed-forward networks have close to zero memory. However, this may not hold true when NNs are augmented by pre-computation or are given a vector of recorded data (often via a tapped delay line).

Having narrowed down the choice to RNNs, there are still many variations to choose from. ESNs however have two particularly attractive properties for the tasks of channel characterisation or prediction: firstly their ability to characterise non-linear systems at secondly to do so at low computational cost.

The suitability of ESNs for processing datasets with highly non-linear or chaotic internal relationships has been recognised by a number of researchers, who note that these complex associations and patterns are otherwise difficult to detect [117, 29]. The seminal paper introducing them [57] used the example of predicting the Mackey-Glass attractor system (a set of complex non-linear time delay differential equations) as unmistakable evidence of its ability to work with non-linear systems. Additionally, they have already proven effective in some applications in the wireless communication space [60]. While ESNs' ability to work with highly non-linear systems is well documented, the underlying reasons are only poorly understood. It is suggested in the literature that their abilities in this areas are partly due to the neuron activation function chosen, and partly because of the randomised nature of inter-neuron connectivity. Further research

into the theoretical underpinning of ESNs is necessary before these characteristics are as well comprehended as more established types of NNs.

Wireless channels are affected by a huge number of factors, some of which may have a detectable and characterisable impact on the overall performance of the channel, while others will appear more like noise, making the whole system very chaotic. If a NN could discriminate among features with different causes, it would be of great value in channel characterisation. For example, if an ESN could anticipate long-term fading, (while not being influenced by more transient effects such as noise or specular reflections from moving vehicles), the output could go beyond characterisation and provide prediction about the channel's future behaviour.

As discussed in Section 2.2, the computational cost of training an ESN is much lower than the cost of training more traditional networks (whether recurrent or feed-forward) because only a small number of connections are modified (i.e. connections to output neurons) the great majority will remain unaffected. This would be a particularly important consideration if such a system were to be integrated into a portable device where power and computational resources are much more constrained than in powerful desktop computers. Since such a wide variety of modern radio systems are targeted at the mobile area, this energy conservation feature of ESNs is vitally important when compared with the wider range of NNs available.

For the task of system characterisation, ESNs can be trained to give an output which indicates how closely the observed system matches its internal representations of each of the individual scenarios on which it was trained. An output which perfectly corresponded to one of the training systems would indicate that it was an exact match. All the similarity ratings can form a useful data set when the observed measurements display characteristics similar to those in several different training sets. This can be used to construct a matrix showing how closely each channel matches each of the training scenarios. Since such a system does not make definitive classifications, but rather indicates degrees of closeness to the range of available classifications, it can be more flexible, providing richer data to the higher layers of the radio chain, in turn allowing better resource allocation decisions to be made.

In summary, ESNs show considerable promise for use as classifiers. They share many of the advantages common to all NNs, and have some unique properties making them particularly well suited to the task of channel characterisation.

4.3 Experimental Setup

This section describes the experimental setup and design decisions taken. Conceptually the system is intended to operate in the following manner: channel data is obtained from known sources, and used to train an ESN. Next, channel data is obtained from an unknown source, and then supplied to the network. After computation, the system should be able to identify which of the channel scenarios present in the training data the unknown channel most closely resembles. This result could then be used by the rest of the radio chain to make decisions about modulation, scheduling and coding.

The following subsections describe each component of the system in finer detail.

4.3.1 Wireless Channel Modelling

Modelling a wireless channel is a complex task, with a huge number of variables affecting a simulation. Ray-tracing can be a highly accurate approach, however (as mentioned in Section 2.3) it requires a large amount of data about the environment to be available. Since this task is in the realm of channel characterisation, it is reasonable to assume that such a detailed level of information would not be available. As a result, a statistical modelling approach was chosen instead.

The Wireless World Initiative New Radio (WINNER) project [16] created a parametrisable channel modelling software package, allowing for a wide variety of wireless scenarios, but particularly targeted at those likely to be encountered in third and fourth generation cellular networks. Accordingly, this suite was selected to model the channels to be used, as it is widely seen to be the most comprehensive channel modelling suite for this purpose [79]. As well as being able to generate arbitrary channels (given appropriate parameters), the WINNER model has a set of thirteen channel scenarios commonly seen in cellular systems, encompassing a range of indoor, outdoor, urban and rural channel scenarios, as outlined in Table 4.1.

Indoor office (A1) This scenario assumes both transmitter and receiver are located inside an office building, with rooms of dimensions 10 x 10 x 3 m spaced regularly along corridors of 5 x 100 x 3m. This allows for the model to cover the Line of Sight (LOS) case (along corridors) and also the Non-Line of Sight (NLOS) cases (within the rooms). The number of floors in the building is also parametrisable, or can be automatically chose at random by the channel simulator.

Code	Channel Description
A1	Indoor office
A2	Indoor to outdoor
B1	Urban micro-cell
B2	Bad Urban micro-cell
B3	Indoor hotspot
B4	Outdoor to indoor
B5	Stationary Feeder
C1	Suburban macro-cell
C2	Urban macro-cell
C3	Bad urban macro-cell
C4	Urban macro outdoor to indoor
D1	Rural macro-cell
D2	Moving networks

Table 4.1: A description of each of the thirteen channel scenarios used in the characterisation problem. These were designed to cover the most common cellular scenarios, and to give a diverse set of channel types for system training and testing. These channel types are outlined below, but are explained in full detail in [112]. Some channel types are radically different from one another (e.g. B1 and D1), while others are more similar (e.g. A1 and B3).

Indoor to outdoor (A2) This setup has the transmitter located inside an office building (modelled as in the A1 case) and the receiver located outside the building. The receiver is at a height of 1-2m, and the transmitter is between 2-2.5m above floor height (another parametrisable option) and is designed to model the pedestrian use case. The outdoor environment is modelled using the B1 case.

Urban micro-cell (B1) In a typical urban setup, both transmitter and receiver are located below the tops of surrounding buildings, and are outdoors. The B1 scenario creates a geometry of streets and buildings in a Manhattan-style grid. This provides both LOS and NLOS environments.

Bad Urban micro-cell (B2) Bad urban micro-cells share the same geometry as the B1 case, but include certain locations at which multipath energy can be received with significant power and may have long excess delay. This models clear radio paths which occur in urban environments with large open spaces (such as open squares or bodies of water).

Indoor hotspot (B3) The B3 scenario represents a typical conditions for an indoor hotspot with wide coverage of large indoor space (such as might be found in a train

station, conference hall airport). Dimensions of the space can be in excess of 100m in length, and allows for both LOS and NLOS areas.

Outdoor to indoor (B4) In a similar manner to the A2 case, the B4 scenario uses the B1 and A1 models. The transmitter is located outside the building, 5-15m below surrounding buildings, and the receiver is located 1-2m above floor height.

Stationary Feeder (B5) In order to model links in which both transmitter and receiver are in fixed locations, scenario B5 can be used. Such links are common in cellular backhaul when a hard connection is not possible due to geography or cost of installation. Since the two radio elements do not move, the calculation of Doppler shifts depends entirely on the motion of scatterers.

Suburban macro-cell (C1) The transmitter is assumed to be located above the height of residential buildings (allowing for wide coverage), which are not laid out in a regular grid as in the urban setting. Vegetation is assumed to be modest, and a number of open areas (such as parks or playgrounds) are added into the environment. The receiver is located near street level.

Urban macro-cell (C2) To simulate a cell with large coverage, the transmitter is assumed to be well above surrounding building height, and the predominant path the the receiver will be via a single diffraction from a rooftop. The receiver is presumed to be at pedestrian height. Both LOS and NLOS cases will be found depending on obstructions from buildings. Building density and type is approximately homogeneous.

Bad urban macro-cell (C3) The bad urban micro-cell is characterised by hetrogeneous building height and layout, and also includes some open spaces in the geography, which induce high excess delay into the propagation environment. The transmitter is also located below the height of some buildings to model the effects of very tall buildings.

Urban macro outdoor to indoor (C4) The outdoor environment is the same as in C2, and the indoor environment the same as in A1. Overall this is a similar setup to the urban micro-cell setup (B1) but the signals will have longer delay spread.

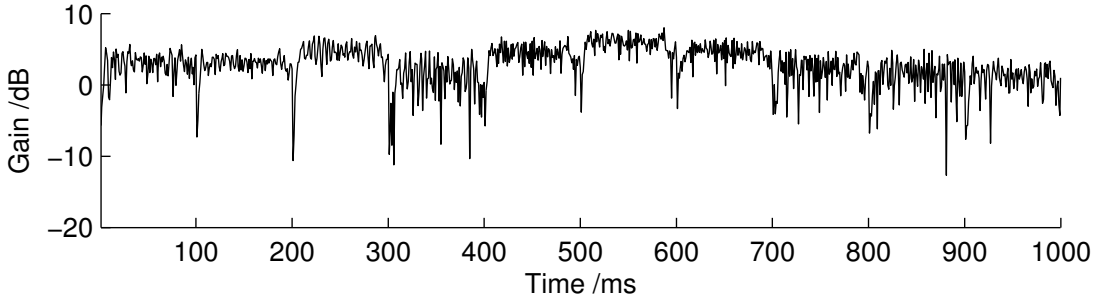
Rural macro-cell (D1) Used for modelling links of up to 10km, the rural macro-cell assumes that the transmitter is located much higher than any buildings (between 20-70m) and so LOS conditions will be common.

Moving networks (D2) This scenario is used to model environments where both the transmitter and receiver are moving at speed (e.g. in a train).

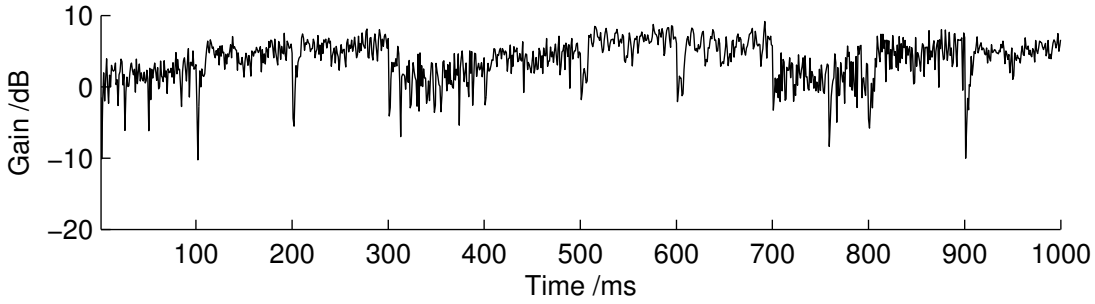
Each time the WINNER software package generates a channel, it creates a simplified, randomised model of the effects of buildings and the many surfaces that act as wireless scatterers and reflectors which it uses to create a mathematical model for that particular setup. This has the effect that using the software to generate two channels with identical scenario designations (e.g. two indoor office (A1) scenarios) will generate different results. These two channels will have similar statistical properties (such as delay spread, fade depth and frequency) but produce distinct time series. This property is useful as it allows for easy generation of different but statistically similar channel models.

In order to eliminate effects due to antennas, ideal point isotropic antennas were used in simulations. This allowed the channels to be fairly compared with one another without needing to compensate for antennas pointing in different directions in the various scenarios.

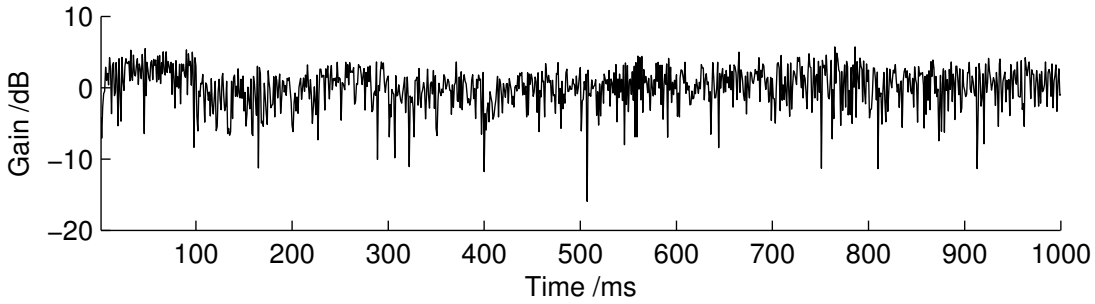
For the characterisation system, a set of the thirteen standard wireless channel scenarios (as described in Table 4.1) is generated, and a randomised signal is transmitted over each, and the channel magnitude data is subsequently extracted. To generate each channel (used for both training and testing) a geographical model is created of where each transmitter and receiver is located for each of the thirteen channel types. A semi-random layout is chosen as provided by the WINNER software in order to generate a model representative of the real world. Each channel's magnitude response is then extracted (seen in the top graph of Fig. 4.3). These thirteen channels are kept as training and testing data for the NN.



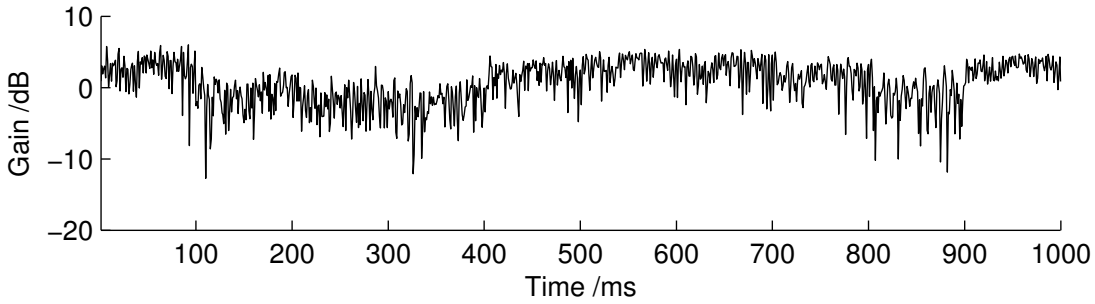
(a) Indoor office scenario (type A1)



(b) Urban micro-cell (type B1)



(c) Suburban macro-cell (type C1)



(d) Moving Network, as found on a high-speed train (type D2)

Figure 4.1: Normalised magnitude response time series for a selection of different channel types, taken from Table 4.1, each showing different characteristics (such as frequency and depth of deep fades, slow and fast fading, etc.). These, along with other similar channels are used in the characterisation task.

4.3.2 Channel Data Pre-processing

In early exploratory experiments the channel magnitude response data was used directly as input to the ESN. Although it was successful in characterisation in a small number of cases, the noisy nature of the signal caused unpredictable behaviour within the ESN. The network would often classify two datasets from the same channel as different types, and was highly inconsistent in its output. Additionally, in a significant number of experiments the ESN output completely saturated, giving no useful information about the input signal. Due to the randomised nature of the ESN's internal structure, these test runs were not repeatable, even when using the same training data. It was then decided that some pre-computation should be applied to the raw time series before it was fed into the ESN (either as an input or as training data) in order to achieve more consistent results. A range of approaches was tested:

Windowed Mean of Signal Envelope

The first approach attempted involved computing the mean of the channel's magnitude response over a sliding time window. This approach did not provide useful data to the ESN. Evidently it over-simplified the input, hiding useful characteristics. For example, once a highly variable channel (such as a high-mobility channel like D2 in table 4.1) had been pre-processed, it could appear very similar to a pre-processed version of a much less rapidly varying channel. Changing the size of the time window over which the mean was calculated was investigated, however no value was found which would eliminate unwanted variations while still preserving enough signal information to produce a usable result. The time windowing method was however carried over and used in the final approach as it is useful for identifying local signal traits.

Windowed Dynamic Range

A method using dynamic range was investigated to supply the ESN with the difference between the largest and smallest values observed within a sliding time window. Although this a system did have limited success at characterisation, it often mischaracterised one channel type as another similar one. Effectively it could only discriminate between channels with a large or small dynamic range, and so this approach was not used.

Local Minima and Maxima Detection

Also considered was the possibility of using the number of local maxima or minima within a sliding time window. The justification for investigating this method was that it could highlight areas of greater signal variability. It was found to be unsuitable however as it could not distinguish between a signal with a small or large variability, and additionally does not take signal strength into account. Also, adding noise (even a small amount) to a 'clean' signal completely changed the nature of the output, and so this approach too was rejected.

Frequency of Deep Fades

Another approach examined involved extracting the frequency with which very deep fades occurred in the channel. This approach had limited success, as channels without frequent deep fades produced very sparse data for characterisation, though it did allow for effective identification of those with frequent fading. Additionally, the decision of what constitutes a 'deep' fade is a somewhat subjective one, and is highly dependent on the channel being investigated. A feature considered a deep fade in one scenario (perhaps because it is particularly rare in that context) could be much more common in other channel types, making comparison of one channel to another subject to interpretation. However, the idea of calculating the frequency of notable events (deep fades in this case) inspired the final approach which was used to pre-process channel data and achieved broadly successful characterisation results (see Section 4.4).

Zero-Crossing Detection

After the disappointing performance of the local minima and maxima method and of the deep fade detection method, attention turned to the most visually evident differences between one channels' response (seen clearly in graphs such as Fig. 4.1) and another that is the amount of variation in the channel over time. With this in mind, zero crossing data was computed for each channel. Although this greatly reduced the amount of data provided to the ESN, the data points were too sparsely spaced for the network to hold sufficient historical data to relate one point to the next.

Windowed Standard Deviation of Zero-Crossing Spread

The method that was finally selected to pre-process the channel data for the ESN involved computing the standard deviation of the distances between adjacent zero crossings within a moving time window. The number of zero crossing points was used to quantify the variability of a channel (in a similar way to the deep fade detection method and the local minima and maxima method), and computing the standard deviation of distances between consecutive points provides a concise statistical description of how closely spaced they are. It also introduces an element of additional 'memory' into the system which the ESN did not adequately capture.

For a window of n points, individual samples x_i and mean $\bar{X}$, each point in the training data S is generated from the windowed zero crossing times T using:

$$X = \{t_i - t_{i-1} | t_i, t_{i-1} \in T\} \tag{4.1}$$

$$s = \sqrt{\frac{\sum_{i=0}^{n}(x_i - \bar{X})^2}{n}} \tag{4.2}$$

The Matrix Laboratory (MATLAB) code to perform this computation is shown in Fig. 4.2, and it is illustrated graphically in Fig. 4.3.

```
for(k=1:time_increment:number_of_readings-window_size)
%count the zero crossings in a given window
    a=crossing(10*log10(input(k:k+window_size)));
    [x b]=size(a);
    %compute the distances between adjacent zero crossings
    dif=a(2:b)-a(1:b-1);
    %compute the s.d. of distances between adjacent zero crossings
    std_diff(c) = std(dif);
    c=c+1;
end
```

Figure 4.2: This is the MATLAB code for the approach selected to pre-process the channel magnitude data before being input to the ESN. It reduces the variability of the the data to a degree that allows the neural network to use it in channel characterisation without the problems encountered when directly feeding the channel magnitude data to the NN.

The algorithm combines a number of features from previously investigated pre-processing methods in order to extract useful data from the simulated channel responses,

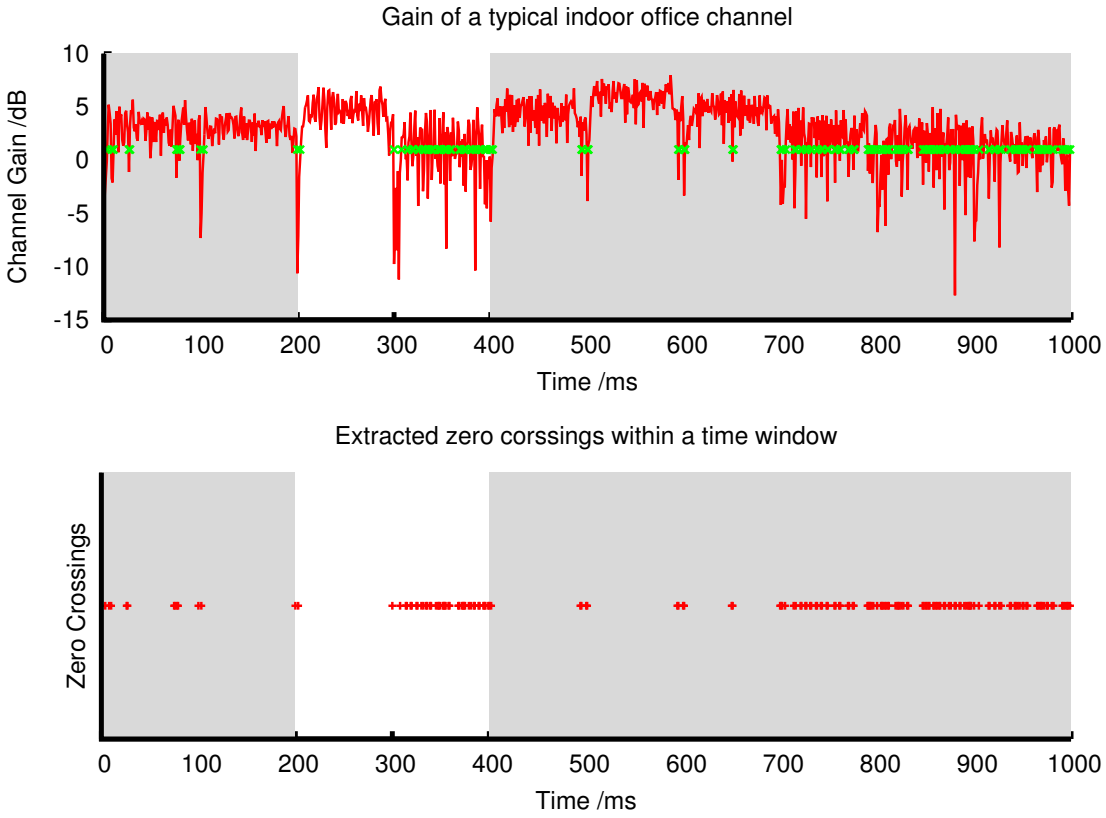


Figure 4.3: The top graph shows the magnitude response of a fast-fading wireless channel for a typical indoor office scenario, with the 0dB crossing points marked. The lower graph shows the same crossing points, and the highlighted section shows the time window over which the standard deviation of distances between adjacent points are calculated. This window is moved along the time series, performing the same calculation at each time step, and the resulting output is then used to train the ESN.

while smoothing out the noisy nature of the signal which cause the network to saturate. This was selected as the data pre-processing algorithm as it was the only algorithm which produced output data which allowed the ESN to classify the data without ever causing the NN to saturate.

This computation is used in both the training and characterisation stages.

4.3.3 ESN Configuration

Generally, when using NNs for general time series prediction, a tapped delay line can be employed. In this setup, the input signal is loaded into shift registers, and fed into multiple input neurons with different time delays (see Fig. 4.4 for a diagram). This has the effect of increasing the dimensionality of the signal input to the network, and giving

access to a small number of historical values, often dramatically enhancing performance. The practice is however much more common in feed-forward networks. In the case of RNNs and especially ESNs, such an arrangement is not only unnecessary, but can be detrimental to performance [58].

Recurrent networks have a far greater historical memory capacity than those which are feed-forward only. Adding a number of extra dimensions simply serves to increase training time and run time. A delay line is also only capable of storing as much of the signal's history as there are shift registers. In some systems, storage of a small number of immediate historical values is sufficient to allow useful characterisation or prediction work, however a huge number of taps would be necessary to capture the complexity in the signal provided by the WINNER channel models. As a result, the signal was pre-processed before being fed to the ESN for characterisation as discussed in Section 4.3.2.

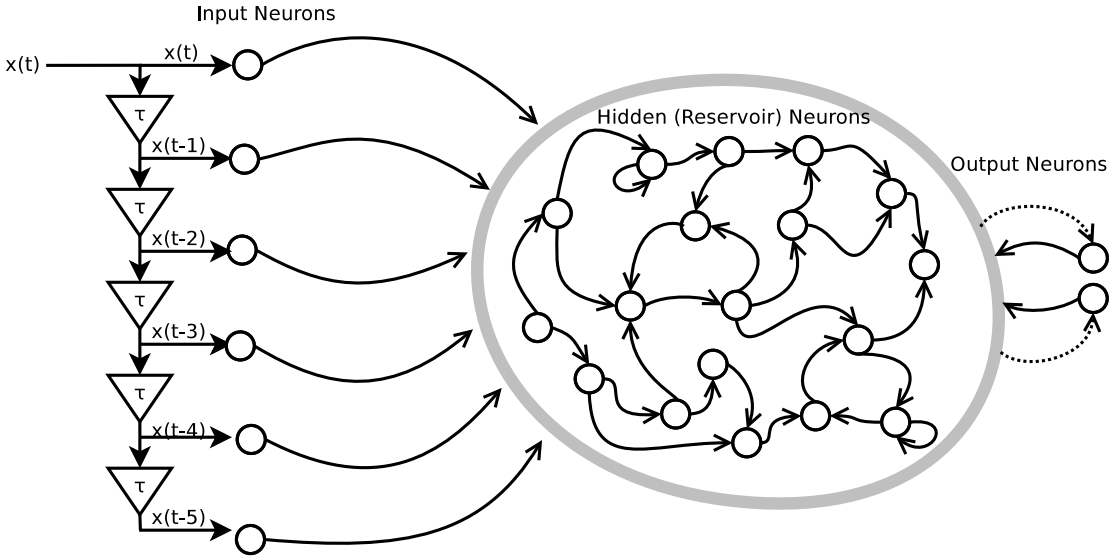


Figure 4.4: Diagram showing the possible use of a tapped delay line from the input signal $x(t)$. This example expands a 1-dimensional input to a 6-dimensional one simply by supplying time delayed versions of $x(t)$ to separate input neurons. Although common and often effective in feed-forward networks, such an arrangement is rarely useful in RNNs due to their inherent historical memory.

In the following experiments, networks of 1,000 internal neurons were used. Experiments were performed (see Section 4.4.5) to choose the optimal tradeoff between accuracy (which increases in a non-linear fashion with neuron count) and running time (which increases exponentially with neuron count). A connectivity of 1% was selected between all neurons. This is a typical value for an ESN suggested by Jeager [57]. Modi-

fying the connectivity did not have a significant impact on accuracy, though a very high connectivity increased running time. For each channel, 100s of Radio Frequency (RF) data was generated to give the ESN a large amount of training data.

ESN Training

Before data from a simulated channel can be characterised, the ESN must be trained to recognise each scenario that defines a particular classification. Since there are thirteen different channel types available in the WINNER simulation suite, an unknown set of channel data must be assessed for its similarity to each of these scenarios. This constraint dictates the number of input and output neurons in the ESN; a single input neuron for the pre-processed time series data, and thirteen output neurons, each corresponding to a particular scenario from Table 4.1.

Training data must first be generated, and then provided to the ESN at both the input and the output in order for the training algorithm to operate. The input is simply a representative signal likely to be seen when the ESN is in characterisation mode. The training data supplied to the output neurons is what the system designer intends the output to be under the conditions imposed by the input signal. It can be though of as the 'right answer' being given to the network for that particular input signal. In an ideally trained network, if the training input signal were provided when the network was running, it would perfectly reproduce at its outputs the same data provided to them during training. Unfortunately in practical systems, if a network is able to perfectly replicate the training data it can be an indicator of over-training, meaning it has not deduced the underlying model for the data, but has simply reproduced it verbatim from memory.

Fig. 4.5 is a diagram of the structure of the training data used, illustrating how a network could be trained for four channel types. It can be trivially extended to cover the thirteen WINNER classifications.

The ESN is supplied with the training data and undergoes the training operation. During this process the ESN adjusts its trainable neuron weights with the goal of replicating the training output signal if it were subsequently supplied the same training data.

In all experiments the ESN-specific variant of the Backpropagation Through Time (BPTT) training algorithm [58, 95, 100] is used. A third-party Matlab library was used

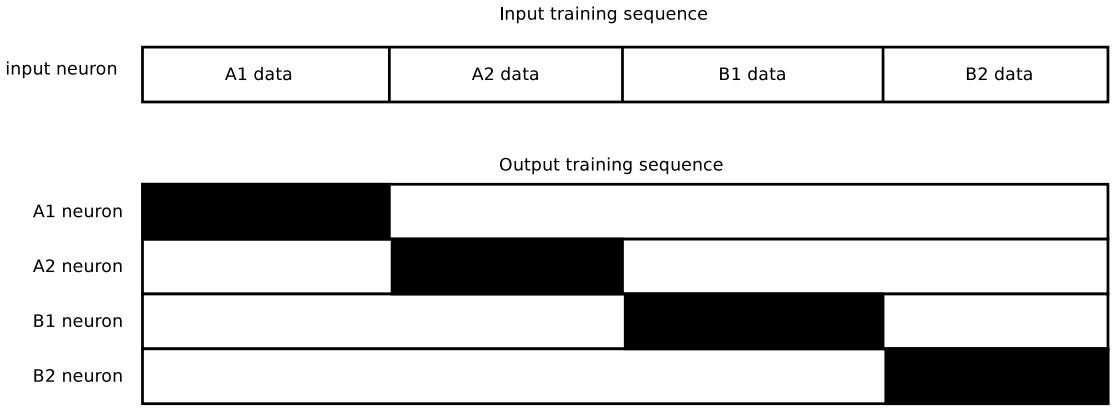


Figure 4.5: Diagram of how training data for the ESN is constructed. Sequences of zero-crossing data from each of the channel types are concatenated to form the 1-dimensional input data. There is a 1-dimensional time series for each of the output neurons (four are illustrated here), corresponding to one particular channel scenario. The value of the time series for each output neuron is zero except when the input sequence is composed of data generated by that particular channel model, when it is set to one.

for ESN simulation, and is detailed in Appendix C. An example of how the library can be used in a larger body of code is included in Appendix B.

The BPTT algorithm was originally developed for feedforward NNs, however it has subsequently been extended to cover RNNs, and ESNs in particular [58]. Training an ESN is computationally much simpler than most other networks of a similar size, simply because the number of neurons weights which can be adjusted is smaller. As explained in Section 2.2.3, only the connections leading to the output neurons can be trained, as all internal weights are fixed to random values when the network is first generated.

The BPTT algorithm iteratively alters the weights of the trainable connections, aiming to minimise the error between the training output and the observed output when the training input is supplied to the ESN's input neuron. It does this by 'unfolding' the network (i.e. making copies of it) for each time step it would go through. At each step it updates the trainable weights in order to arrive at an output which is closer to the training output data than was achieved at the previous step. It can execute an arbitrary number of times, each time aiming to reduce the overall error.

Once the training step has completed, the channel to be characterised goes through the same procedure as the training data. When the time series of standard deviation of zero crossings has been calculated, this data is fed into the ESN to be characterised. The ESN has a one-dimensional time series output (coming from an output neuron) for each of the thirteen channels on which it was trained. The input time series is passed into

the ESN, where the network computes how closely this unknown data matches each of the thirteen channel types it was trained to recognise. The output time series with the highest mean value denotes the closest match. It should be noted that the classification is probabilistic. In the case where a particular mean value is markedly larger than the other 12 means, then the classification is highly likely to be correct. However, where two or more mean values are very similar, then the ESN has not been particularly successful at determining the best classification. The system setup is illustrated in Fig. 4.6.

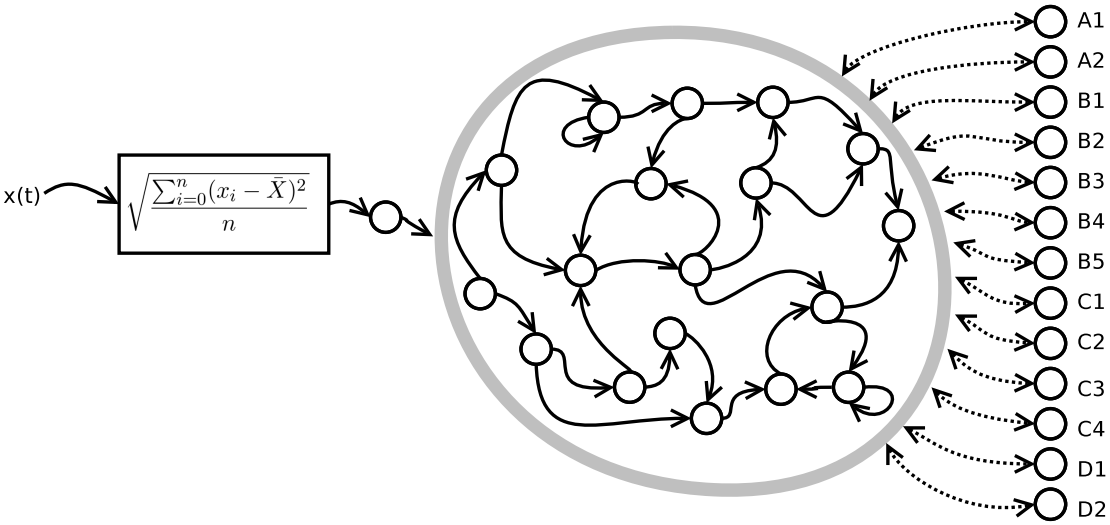


Figure 4.6: Diagram of how the classification of channel data $x(t)$ from the WINNER modelling software is used. After the raw channel data has been created, it is pre-processed, transforming it into a time series of standard deviations of zero-crossing separation data which is then fed to the trained ESN. The 13 output neurons each rate how closely the current input matches the channel type they were trained to recognise. The outputs from these neurons are then compared with each other. The output with the largest mean (i.e. most closely matched the scenario it was trained to) is judged to be the most likely classification, while less similar channel types will be rated as matching less closely.

In simulation the first several hundred output values from the ESN (often referred to as 'dummy steps') are ignored as they will not be based on initialised data within the NN. In each simulation, 100 seconds of training data for each channel type is generated in order to create a large, representative sample of each scenario on which the network can be trained. The ESN is then trained on each of the thirteen channel types in turn, before any unknown signals are applied.

For the experiments described in this chapter, a reservoir of 1,000 neurons was used. See Section 4.4.5 for a discussion of the impact of neuron count on characterisation

results.

4.3.4 Adding Basic Location Data

In the interests of increasing the likelihood of the characterisation scheme giving the correct classification for an unknown channel, additional methods were examined. As is clear from the different designations given to the thirteen WINNER scenarios, they are heavily based on location - whether in a city, suburbia or within an office. This is entirely understandable as these different environments are very different from an RF propagation standpoint.

In a cellular radio system, every device can determine its location to some degree of accuracy, whether it is only which macrocell it is connected to, or if it is connected to a femtocell whose coverage extends a few tens of metres. Even if this data were not available, a Global Navigation Satellite Systems (GNSS) such as Global Positioning System (GPS) could be used to determine location to a very high degree of accuracy. A limited amount of location-specifc information could even be sent by the transmitter if it has knowledge of its surroundings.

If each channel is marked as 'urban' or 'rural' and 'indoor' or 'outdoor' when the system is trained, (and it can be assumed that the mobile device is able to determine this same data) this information could be used to narrow the search space of channels considerably, leading to more accurate characterisations.

Two different approaches to integrating this data into the characterisation scheme were investigated and these are discussed below.

Additional ESN input

An obvious way to integrate this additional information into the characterisation scheme is to provide it as an extra input to the ESN. NNs are ideally suited to taking multi-dimensional data as an input. Experiments were carried out which converted the urban/rural and indoor/outdoor data into discrete values, which were then supplied to the network during training and characterisation. This new input changed the structure of the system to have two input neurons, and this revised layout is illustrated in Fig. 4.7

Using this setup exploits one of the most attractive features of NNs - their ability to infer relationships within data automatically. Providing this data as an additional

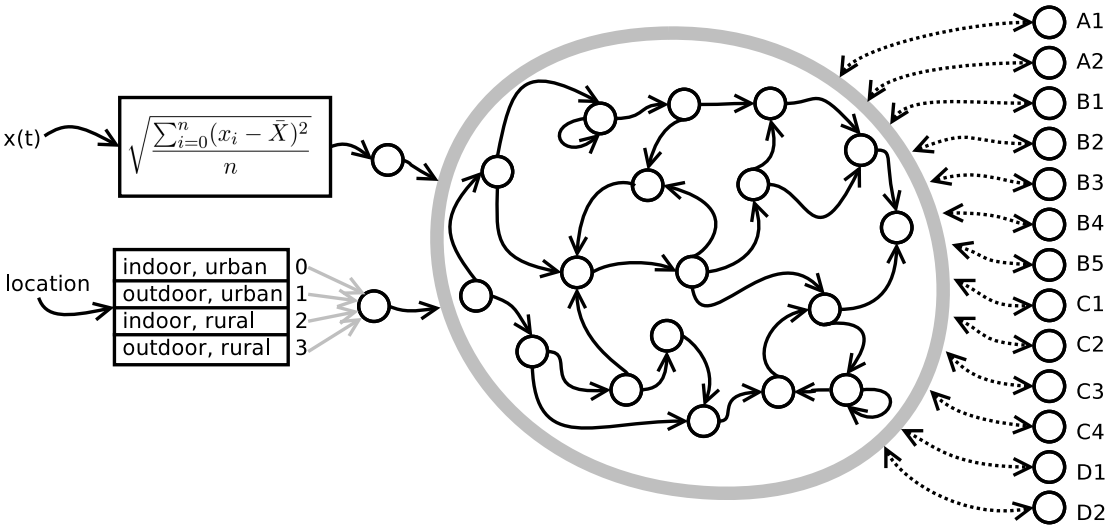


Figure 4.7: Diagram of the characterisation system with the inclusion of basic location data provided as an additional input to the ESN. This extra information is converted into a discrete numerical value, corresponding to one of the possible combinations of urban/rural and indoor/outdoor.

input does require more processing power at both the training and characterisation phase. Furthermore, this setup is not tolerant of datasets where either the availability or reliability of location data is in question. It was also open to the potential conflict of identifying a channel type which was expressly excluded by the location data provided. Ultimately a different approach was taken, as outlined below.

Combinatorial Logic

Since providing location data directly to the ESN entailed potential conflicts, a simpler and more robust system was devised. By using Boolean combinatorial logic, a very effective system can be added to filter the results after the ESN has performed its characterisation work. At its most basic such a system could work by excluding channel classifications which were incompatible with the location data available. This setup could for example exclude all urban classifications if the location data indicated the receiver was in a rural setting.

There are three advantages over using the ESN directly to perform these calculations. Firstly, no additional processing is incurred within the ESN, and the logic required to implement the operation can be trivial. Secondly, it also becomes simpler to make use of the classification result produced by the ESN if no location data is available, as it could simply be bypassed (giving the same result as having no combinatorial logic at

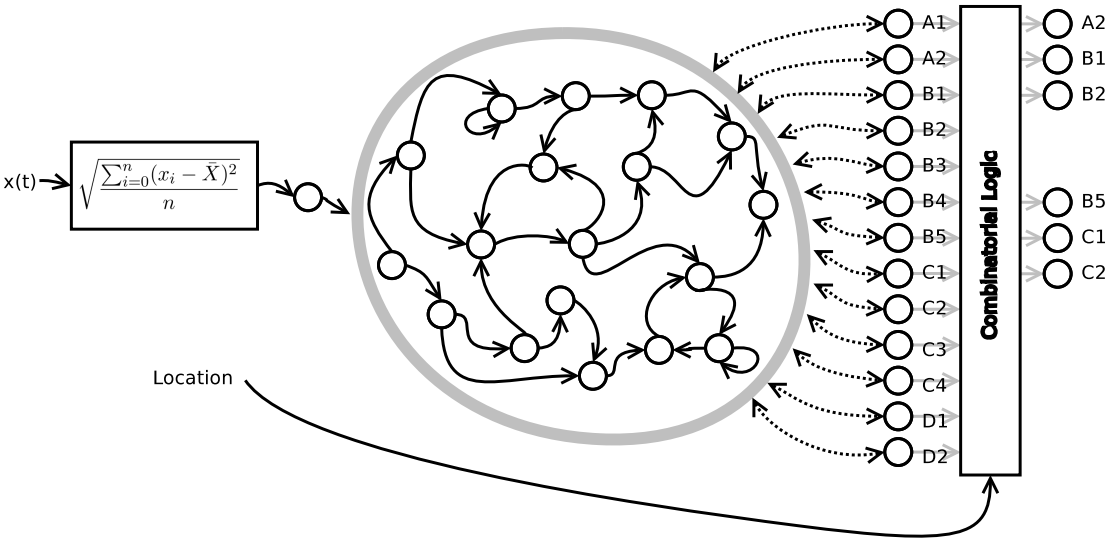


Figure 4.8: Diagram of how a combinatorial logic system could be used to exclude any channel classifications which were inconsistent with the location data available. This example shows the results of the logic system being fed with 'outdoor' and 'urban', thus excluding any channel classifications being labelled as rural or indoor.

all). This would not be possible if the location data were supplied as an input to the ESN. Thirdly, such a system is quite flexible, as it need not be restricted to simple Boolean logic. If the location data came with a 'confidence rating' (i.e. how likely it is to be correct) it could be used as a weighting factor when computing the probability that a channel belongs to a particular classification.

Ultimately, a Boolean logic system was adopted for adding location data both for its simplicity and flexibility. If location data is available it is given precedence because it will normally be reliable, and the system excludes those classifications which disagree with the spatial information available.

4.4 Results

This section presents the results of using the various ESN-based characterisation methods discussed in the previous section. Firstly however, a purely statistical characterisation system is described in order to compare a conventional approach with the newer ESN method.

4.4.1 Statistical Classification System

In order to assess the ESN's effectiveness, a statistical characterisation system was constructed, which like the ESN system did not use a pre-defined system model. Instead of using a neural network to characterise signals, it created Probability Mass Function (PMF)s of the signal power, and compared the power distributions between known and unknown data using the KLD and the L1D equation, which are respectively computed using (4.3) and (4.4). The statistics are generated from two PMFs p and q, each with X bins.

$$\mathrm{KLD}(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{4.3}$$

$$\mathrm{L1D}(p||q) = \sum_{x \in X} |p(x) - q(x)| \tag{4.4}$$

The KLD metric was chosen to compare PMFs because it measures the dissimilarity between two distributions and has previously shown promise in comparing power histograms [3, 6]. Further detail about the KLD algorithm is provided in Chapter 2. The KLD is calculated between the PMF of a known signal, p, and the PMF of the unknown signal, q. If the two distributions are identical, the KLD will be equal to zero, probably indicating that the two distributions were generated from the same channel modelling scenario. The more dissimilar the PMFs are, the larger the KLD between them will be.

The L1D was also chosen to compare the two PMFs, as it has seen use in comparing power histogram estimates [5, 7]. It is conceptually similar to the KLD metric, although the values generated by L1D can be smaller then those from the KLD for the same dissimilar PMFs. It was included to assess whether in this setting it might give different (and perhaps more accurate results) than the KLD.

The system chooses the channel which has the smallest difference (using the KLD and L1D) between the known and an unknown signal as the most likely to be correct. The system is supplied with the same channel magnitude information as used in the ESN system, although it is not passed though the pre-processing phase, since PMFs by their nature are more suited to rapidly varying time series typically generated by the channel modelling software. As the system is only capable of comparing PMFs it is not suitable for processing the location data. While this could be implemented using combinatorial

logic after the system provides output in the same way as it can be with the ESN, the classification system and the location filtering can be considered two independent operations, and this statistical system is concerned only with a comparative evaluation of the echo state approach.

4.4.2 Simulation Results

The ESN system (using only the smoothed zero crossing data, and no additional location data) was able to correctly identify which of the thirteen channels was used in 68% of cases over a run of 10,000 randomly generated channels. This is a noteworthy result, since this characterisation work is done without an explicit physical model of the wireless system. Any model exists solely within the echo state network, and was generated entirely autonomously from the training data. This is a major strength of this method - that there is no need for a system model to be programmed for the system to be able to discriminate between different channel types. Since it is entirely data-driven (rather than model-driven) the effort required to construct such a system is much less than an equivalent system using more conventional methods, and is also more agile - adding a new or updated classification simply involves gathering suitable training data and re-training the system with it. The use of an ESN allowed the system training phase to be much less computationally demanding than with other types of NNs.

This result compares favourably with the statistical comparison systems: using the same channel data, the KLD system makes a correct characterisation in 35% of cases, while the L1D system is marginally inferior, making a correct classification in 33.5% of cases. While this is a much better result than a completely random guess would produce (approximately 7.7% in the case of 13 channels), it is clear that the NN system is a much superior classifier in this setting.

By comparing the results generated using the ESNs to those from the statistical classifier, it is clear that the echo state approach is able to extract much more useful channel information. The fact that it is able to do so without any domain-specific knowledge of how the system works (such as what kind of radio system is being employed) is testament to the convenience of using ESNs in the task of channel characterisation.

Error Clustering

Fig. 4.9 is a plot of mis-classified channels from the statistical system, using the KLD algorithm. It shows quite a uniform spread of errors, with no obvious patterns visible. However, on closer inspection of the incorrectly classified channels from the ESN system, it was found that in a substantial number of cases the correct channel had a prediction score extremely close to the mis-predicted channel. This was often the case with similar channel scenarios, meaning that the ESN perceived these two channels as very alike each other. This led to the clustering of errors, where one channel would often be mis-classified as another similar type. This is immediately evident in the spread of errors in Fig. 4.10 - channel type A1 (indoor office) is a closely comparable scenario to both A2 (indoor to outdoor) and B3 (indoor hotspot). These mis-classifications have the following probabilities: $P(A2|A1) = 0.18$ and $P(B3|A1) = 0.23$.

The likely reason for these errors is the limited amount of information available to the ESN (i.e. the standard deviation of zero-crossings). This pre-processing step removes detail from the signal, meaning that the differences in other channel parameters cannot be extracted. However, as reported in Section 4.3.2, providing the network with the un-processed magnitude response data does not yield useful results.

Although these results are errors (in the sense that they have not correctly identified the channel scenario from the 13 possible models) misclassification of this kind may not adversely affect a real system. In real-world scenarios, observations may not always fit neatly into the set of classifications outlined in the WINNER models. If the channel parameters are genuinely similar, then the transmission scheme used for those sets of channel conditions are also likely to be similar.

The remainder of the mis-classifications from the ESN system (those not identified as a similar channel) seem to occur randomly, being given the highest probabilistic prediction score of any channel, sometimes by a large margin. The reason for this behaviour is not clear. It is a possibility that the neural network had reached its limit for the complexity of the modelled system, however this seems unlikely because increasing the numbers of neurons above 1,000 (as used in the main experiments) showed no change in the rate of incorrect classifications (see Fig. 4.13). No pattern could be found to this type of mis-predication, and it was not repeatable except by using exactly the same ESN parameters, neuron weights, channels and RF data.

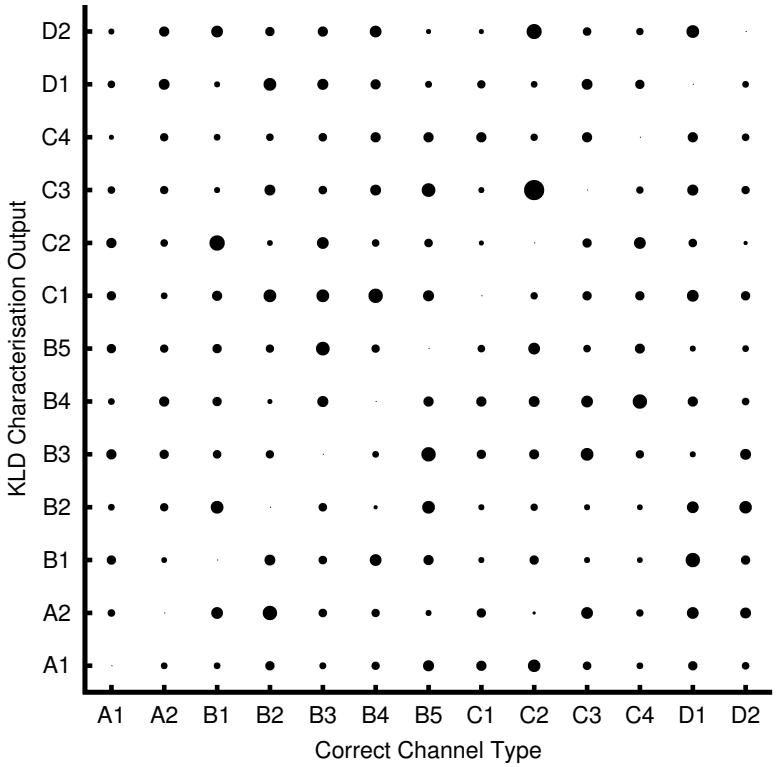


Figure 4.9: Graph showing errors in the predicted vs. actual channel types from a randomised simulation using the KLD metric to measure the similarity of training data to each unknown channel. All correctly classified channels have been omitted for clarity (hence the empty $x = y$ diagonal on the grid). This shows a relatively uniform spread of errors, as would be expected from such a system which uses a purely statistical classification method.

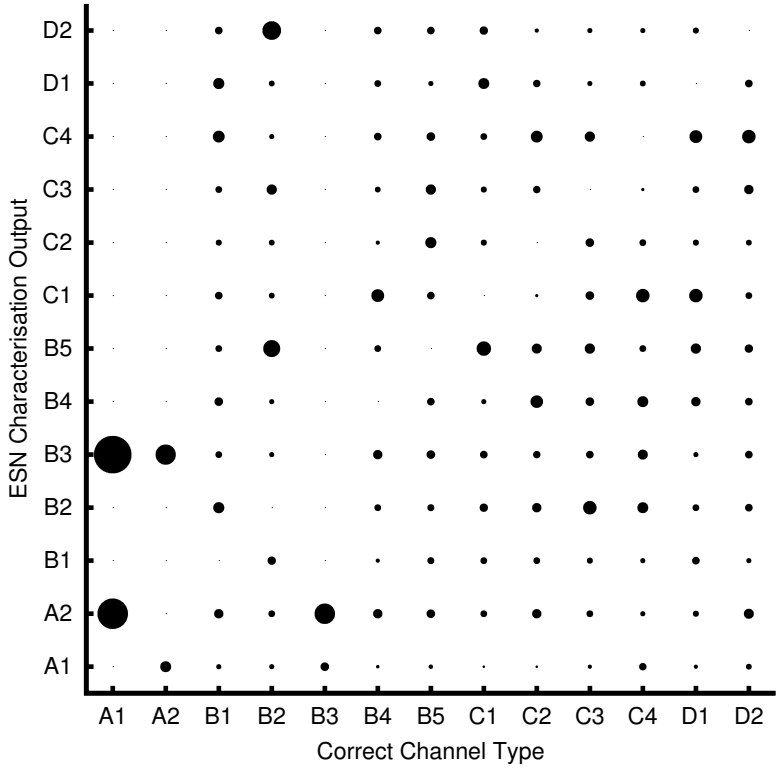


Figure 4.10: Graph showing errors in the predicted vs. actual channel types from a randomised simulation using the trained ESN system. As in Fig. 4.9, all correctly classified results have been omitted. Circle size denotes the probability of each type of mis-classification. Heavy clustering around similar channel types is evident in a number of cases, most visibly in the case of type A1 channels.

4.4.3 Using Basic Location Data

Further tests were carried out on the same 10,000 randomly generated channels used in previous experiments, but with the addition of the combinatorial logic system discussed in Section 4.3.4. Adding the coarse location data (by tagging each piece of RF data as urban/rural and indoor/outdoor) improves the ESN's correct characterisation rate to 72%. This is a helpful improvement over the baseline system and if this information is available, it is a worthwhile addition, especially given the simplicity of the filtering. It works by excluding some of the channel types, allowing only the results from channels which are tagged with the same location data to be made available for comparison at the output.

That the location data does not give a larger improvement to the characterisation performance of the system is a good indicator that this information is already being inferred automatically by the ESN in the majority of cases. If the ESN were not already inferring this information but making a randomised guess, the error rate would drop by a much larger amount than it does with the trained system used here.

System Robustness

To generate testing and training data, rather than generate completely new channel models for each individual test point, one channel model was created for each of the thirteen WINNER scenarios, and then its output was subdivided for use in multiple tests. This was done for practical reasons: the WINNER modelling suite is very resource intensive, and setting up a new channel simulation can take several seconds, even on a modern workstation-class computer. When performing a simulation with a large number of data points (of the order of 10,000 channels) this can easily amount to an additional half day of CPU time on top of the run time associated with the channel data pre-processing and the ESN training and simulation. If the channel modelling scenarios are set up only once each, a longer time series can be generated and the total processing time can easily be an order of magnitude faster than producing many short time series from individual channel setups, while the same amount of time series data can be generated.

In experiments, the first 100 seconds of RF data from each channel model was used as training data. The RF data subsequently generated from the same model was then stored and later used as input to the ESN in smaller sections. See Fig. 4.11 for an

illustration of this setup.

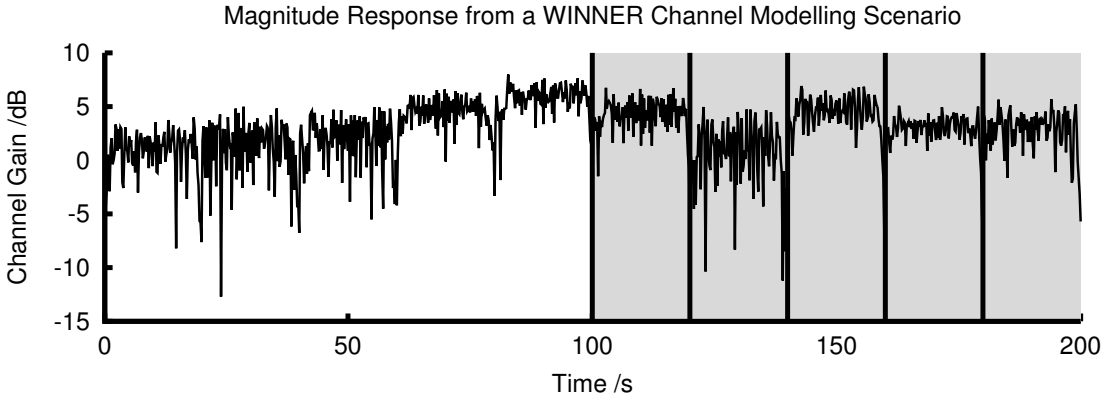


Figure 4.11: Diagram showing the method used to create training and testing data. For each channel scenario, the first 100 seconds of channel data is used as training data for the ESN. The rest of the time series is divided into shorter chunks of 20 seconds (the diagram shows 5 such periods) to be supplied individually to the NN to test how effectively it can identify from which channel type each chunk originated.

After the encouraging results presented earlier in this section, further tests were performed which used a separate instance of each channel type for each individual characterisation test, rather than using different parts of one channel model's time series. From this further investigation, it became clear that in the earlier tests the ESN had learned the individual channel setup, rather than the more general channel type. Since the WINNER models produce a randomised environment layout each time they are run, two instances of an indoor office channel will have different placement of transmitter and receiver, walls and corridors, etc. After investigation it was found that the network was becoming accustomed to the very particular channel conditions present in the training data, so when another channel of the same designation (but different physical layout) was provided, the ESN was unable to recognise it correctly, resulting in an almost perfectly random distribution of results. From this, it is clear that the ESN made a representative model of the particular environmental layout, rather than the whole channel scenario type.

After this discovery was made, further experiments were carried out to determine if this issue could be overcome. To achieve this, the training output signal was modified so that the training for each channel scenario was composed by concatenating of many different individual channel models, all of the same scenario (i.e. many different A1 channel models were made, followed by many different A2 models, etc.). This was in the hope that a more diverse training set might give both a more representative

sample of the characteristics essential to the different scenarios, and also prevent the ESN becoming over-reliant on any one channel modelling instance.

The results from these experiments were no better than when the training and testing datasets came from different channel models. From this we can deduce that while the ESN approach is capable of 'learning' the characteristics of a particular channel, its ability to generalise to a wider set of similar channels seems limited. While this feature is an unwanted one for this characterisation task, the ability to extract information about a particular channel could be a very useful one in the field of channel prediction which will be explored in Chapter 5.

A further investigation was carried out to assess the system's robustness to other channel models. For this task, a simple Rayleigh channel was chosen with a maximum Doppler shift of 100Hz (similar to the values in the WINNER models). The same pre-processing algorithm was applied to the time series data, and a 1,000 neuron ESN was used with a connectivity of 1%. The system was not able to characterise data from this channel model. This is likely because a pure Rayleigh channel model (as implemented by the standard Matlab library) is modelled as a random process. A more accurate channel model (such as the WINNER suite) will not be a purely random process, as it incorporates a geometric model of its surroundings, and it is this non-random nature which allows the ESN to characterise a signal which has historially been treated as being purely random.

4.4.4 Adding Further Pre-processed RF Data

As previously discussed, ESNs are not suited to the direct processing of radio channel data, but require some domain specific pre-computation in order to produce useful results. After settling on the zero-crossing approach for pre-computation, the range of algorithms was re-examined in the hope that by using more than one pre-computation method, the correct characterisation rate might be further improved.

Since the method which supplied the ESN with the dynamic range over a sliding window had shown some limited promise in exploratory experiments (see Section 4.3.2), it was selected as a candidate function to complement the zero-crossing method used elsewhere. Fig. 4.12 shows the modified characterisation system, where the time series undergoes two parallel pre-processing steps before being fed to two input neurons.

The necessary system modifications significantly increased both the training and

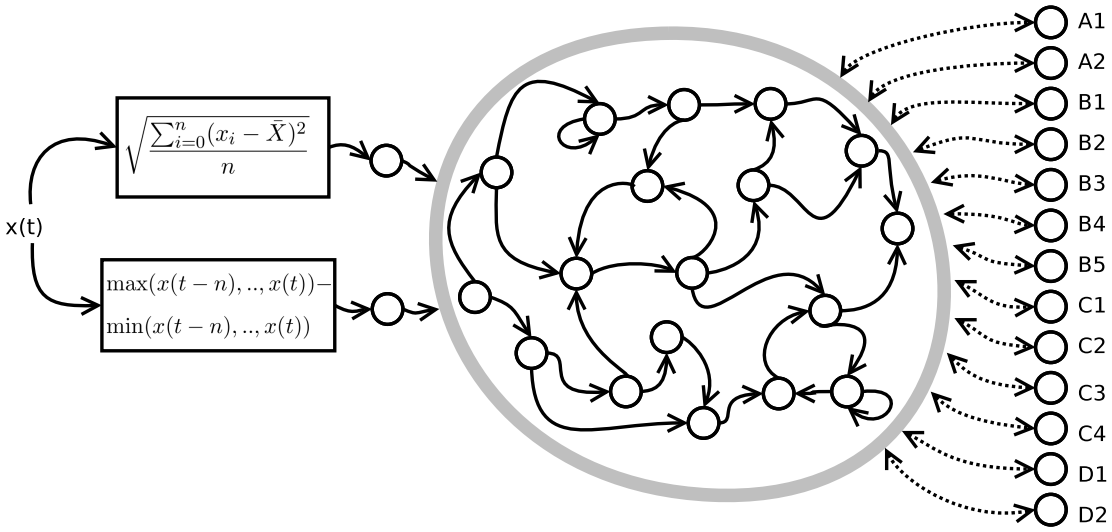


Figure 4.12: System model using two parallel methods of channel data pre-computation; the previously selected zero-crossing data and the windowed dynamic range data. Training data is generated in the same way as in previous experiments, with the addition of dynamic range data as an input.

simulation time required for the ESN, without yielding a measurable increase in correct classification rate. Although the dynamic range method had previously shown some promise, it seems clear that it fails to add any information not already present in the zero-crossing data.

While it is possible that additional RF domain-specific pre-computation could give further improvements, using the windowed zero-crossing method has achieved very encouraging results.

If data such as the channel's delay spread, pilot signals or Doppler characteristics were made available to a similar system, these could perhaps provide further information cannot be easily estimated from the channel's magnitude response data. Whether or not this data could be gathered in a real-world setting however would depend very heavily on the precise type of radio system being used, and could incur significant extra processing overheads.

4.4.5 Neuron Count vs Accuracy

In a tutorial written by Jaeger on choosing ESN parameters [58], he remarks that having too large a reservoir can result in a model which is too powerful for the dataset, picking up on irrelevant statistical fluctuations, and extrapolating these into a model.

The suggestion given is to increase the size of the network until performance on test datasets deteriorates.

Following this advice, the graph in Fig. 4.13 was produced to show the impact of the number of reservoir neurons on classification performance.

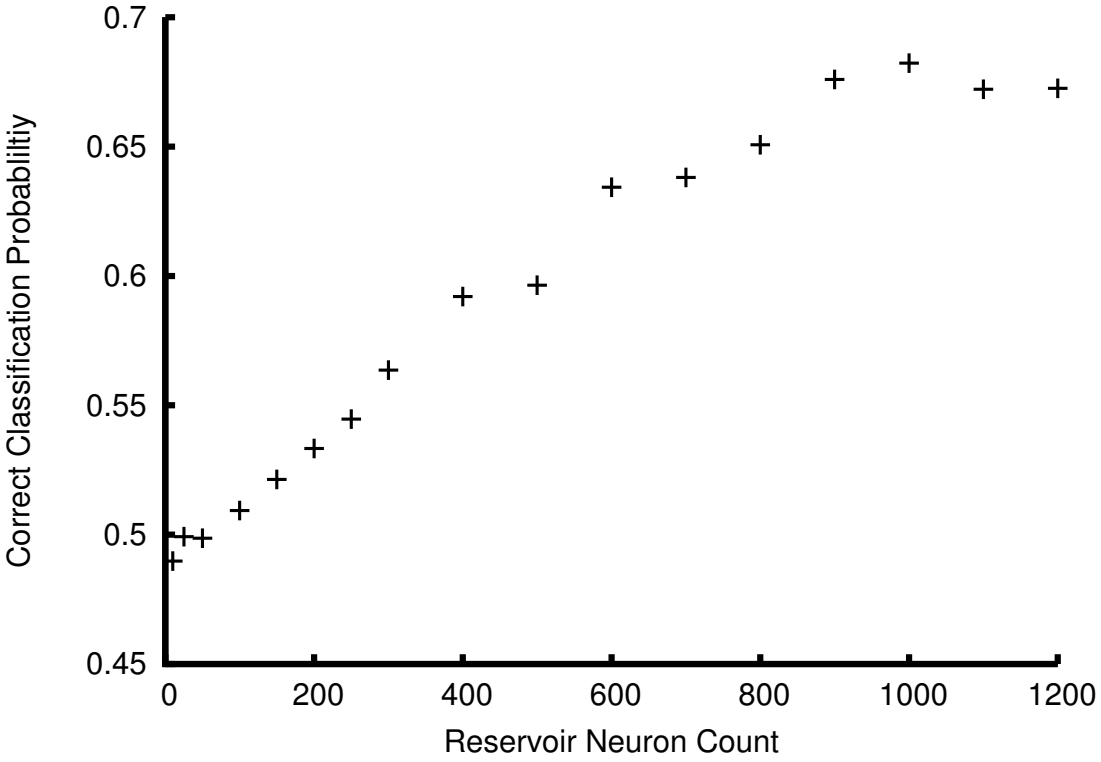


Figure 4.13: Graph showing how reservoir neuron count influences system performance. As can be seen from the final two data points, the probability of correct classification peaks at 68% for this setup. Larger neural networks (of over 1,000 neurons) use significant additional system resources but do not produce any increase in the correct characterisation rate. Very much larger networks (beyond 10,000 neurons) have a tendency to become over-sensitive, saturating frequently.

Each data point in Fig. 4.13 was generated by training and testing 100 ESNs. Since each time a new ESN is generated it has a unique random arrangement of internal neurons, each network must undergo training. This makes each iteration distinct from the next, even though the same training data was supplied to each ESN of each size. The mean value of each set of 100 was plotted.

As a result of these experiments, a reservoir size of 1,000 neurons was chosen to perform the characterisation work previously outlined. The probability of correct classification of an unknown signal increases approximately linearly with the number of neurons used (see Fig. 4.13), reaching a peak at 1,000 neurons, beyond which there

is no further improvement. Also as mentioned in the tutorial, very much larger networks became extremely sensitive, often having their outputs saturate, effectively giving random classification scores to unknown channel input. Although varying the spectral radius of the network is known to make saturation somewhat less likely, these large networks never achieved as high a classification rate as the networks of 1,000 neurons.

Another factor which often influences the number of neurons chosen is running time. In a fully connected NN each neuron is connected to every other neuron. This means that for an n-neuron network, running time increases with n^2. ESNs are typically sparsely connected - often around 1% [58]. While this sparse connectivity does reduce the processing time when compared to a fully connected NN of the same size, the n^2 running time bound cannot be avoided. For the data used to produce Fig. 4.13, simulations using the smallest networks ran in a matter of milliseconds, while the 1,000 neuron simulation took in excess of 30 minutes to process the same data. Even longer run times were encountered, with the 10,000 size network taking may days of CPU time to simulate.

Interestingly, even a network with 10 neurons is able to achieve an almost 50% correct classification rate. This could be useful if the application called for extremely low power draw or lacked sufficient memory to represent larger networks.

4.5 Summary

The echo state network approach to channel characterisation has shown good potential, and most notably does not require a system model. Since the system is trained entirely on measured data and makes no assumptions about the physical system it is trying to characterise, it is highly adaptable to a wide variety of problems. As it outperforms other statistical methods (such as PMF comparison methods) it clearly demonstrates that the learning nature of the echo state approach gives insight other systems are not capable to detect. Provided the information can be supplied to the receiver, the addition of simple location data is an effective method for marginally but usefully increasing the likelihood of a correct classification. This could conceivably be extended to cover a more diverse or specific set of location types, depending on the types of channels expected. The system is somewhat constrained in its ability to generalise channels, as it has a tendency to learn characteristics of a specific channel instance, rather than a more comprehensive scenario.

To optimally solve the problem of channel characterisation, a hybrid system is a possible solution, incorporating an echo state network for the classification work, and tailoring the information made available to it closely to the specific RF system. Supplying information from pilot signals from the wireless link to the ESN (in addition to the windowed zero-crossing rate employed here) might allow such a system to distinguish among a wider range of channel types than this system, with an even higher rate of correct characterisation. Further investigation of this possibility is however necessary to verify these suggestions.

This chapter has presented a novel method to characterise wireless channels using an entirely data-driven approach. The ESN is both simple to implement and computationally cheap to train (especially when compared to other classes of RNNs). It was highlighted that some pre-processing was necessary on the time series data to allow the ESN to process it effectively. The ESN approach can achieve double the accuracy of statistical approaches, and adding new channel classifications, only requires suitable training data. With the addition of some simple location data, results can be marginally improved.

The following chapter builds on the work on channel characterisation developed in this chapter, but moves on to the even stronger challenge of channel prediction. It will discuss the suitability of ESNs and move on to present a novel method based on some concepts from the field of ray tracing.

Chapter 5

Channel Prediction

The previous chapter addressed the area of channel characterisation, and showed that by using an entirely data-driven approach, the type of channel environment could be inferred with reasonable reliability using an Echo State Network (ESN).

As a further development of the theme of increasing channel knowledge to improve radio systems, an investigation into channel prediction methods is presented in this chapter. The relevant background to the area has been outlined in detail in Section 2.3, while Section 5.1 introduces the ideas, concepts and motivation behind channel prediction. In light of the positive results achieved when using ESNs in channel characterisation (see Chapter 4), Section 5.2 investigates their suitability as channel predictors. An alternative method for channel prediction is introduced in Section 5.3 and further refined in Section 5.4. The impact of real-world effects on this prediction method is discussed in Section 5.5, and Section 5.6 summarises the main findings of this chapter.

5.1 Introduction

As has been established in Chapter 2, schemes which depend on active channel measurement devote a significant proportion of their resources to pilot signals, power control signalling and calculations, none of which communicate data across the channel, and so limit the system's maximum throughput, all of which increases power usage. This range of negative issues could potentially be solved with increased channel knowledge.

If a method for perfectly predicting a channel's future response were discovered, it would allow huge gains in energy efficiency to be made. If the transmitter had knowledge of the channel conditions, it could use adaptive transmission techniques to best match

the downlink channel conditions and use scheduling and modulation schemes to optimise throughput. Deep fades could be worked around by re-scheduling the transmission in time or frequency, and power efficiency could be maximised by exploiting the perfect knowledge of the channel's gain and phase shift. The need for pilot signals to actively measure the current channel would also be reduced.

Additionally, if the receiver had accurate channel knowledge, it could decode the transmitted data with increased accuracy, needing less computational power to compensate for corrupted messages. Moreover, if both transmitter and receiver have channel knowledge, it removes the need for a large amount of the signalling overhead.

If all these techniques could come together, it would bring great benefits to both the transmitter and receiver, translating into cheaper, more efficient radio systems.

5.2 Using ESNs for Channel Prediction

Following the encouraging results presented in the previous chapter, an investigation was started into using the same Neural Network (NN) tools in channel prediction (i.e. predicting future channel characteristics). As was discussed in Section 2.2.3, ESNs have demonstrated their ability to predict chaotic time series, sometimes with surprisingly high accuracy. The assumption made in this work was that the channel's future behaviour can be deduced to some extent by examining its history. The first part of this chapter will discuss the merits of using such ESNs to predict a channel's future state.

5.2.1 Preliminary Experiments

Initial experiments began by using an ESN to produce a rolling prediction. In such a setup, the NN provides a prediction for a given number of points in the future at its output, while simultaneously being given current system conditions at its input. The output can be fed back as input for further predictions, however this is a very error-prone method and is only used when current data is not easily available. The more usual approach is to compare the prediction with the system's actual behaviour. This setup is illustrated in Fig. 5.1.

As in the earlier channel characterisation work (see Chapter 4), the Wireless World Initiative New Radio (WINNER) suite, noted for providing excellent channel modelling, was used in experiments to generate the magnitude response data for selected channel scenarios, such as an indoor office or a moving train.

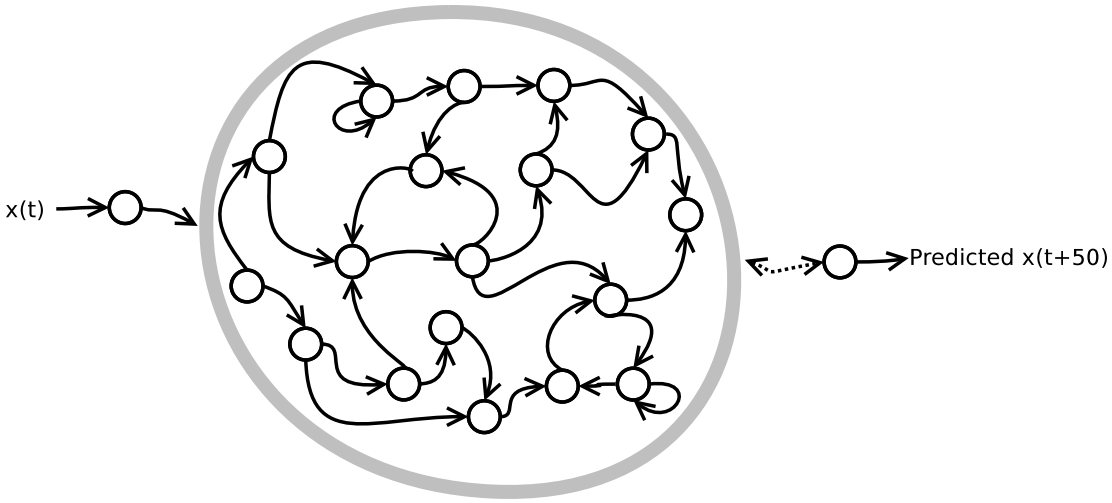


Figure 5.1: Diagram showing the ESN layout used in prediction experiments. The current channel magnitude values, $x(t)$ are provided at the input neuron, while the network attempts to produce a prediction for $x(t + 50)$ at its output.

In order to carry out these experiments, two sets of channel data of the same WINNER scenario (see Table 4.1) were generated. The first set of channel data is used as training data for the ESN while the second set of channel data is used to represent the channel to be predicted. The ESN is then tasked with predicting a number of time instants ahead (corresponding to distance for a moving receiver) on the second channel.

In these experiments, channel measurements are taken every millimetre along the trajectory. The software is configured to provide a prediction 50 data points into the future, corresponding to 50mm (i.e. 5cm). Such values were chosen based on a pedestrian model - with a walking speed of a few metres per second, 5cm in distance corresponds to a few milliseconds in time. Having channel predictions, even for this short time ahead would be most advantageous.

With respect to the NN being used, exactly the same setup can be used as in the channel characterisation work - only the number of output neurons and the training data need to be changed. The ESN's single output neuron gives its prediction for future conditions, while the input neuron is supplied with the current channel conditions. As in the characterisation task, a reservoir of 1,000 neurons was used, with a connectivity of 1%. The standard Backpropagation Through Time (BPTT) training algorithm was used.

A single run of such an experiment is shown in Fig. 5.2.

As is immediately evident from the graph, the simulated channel varies with great

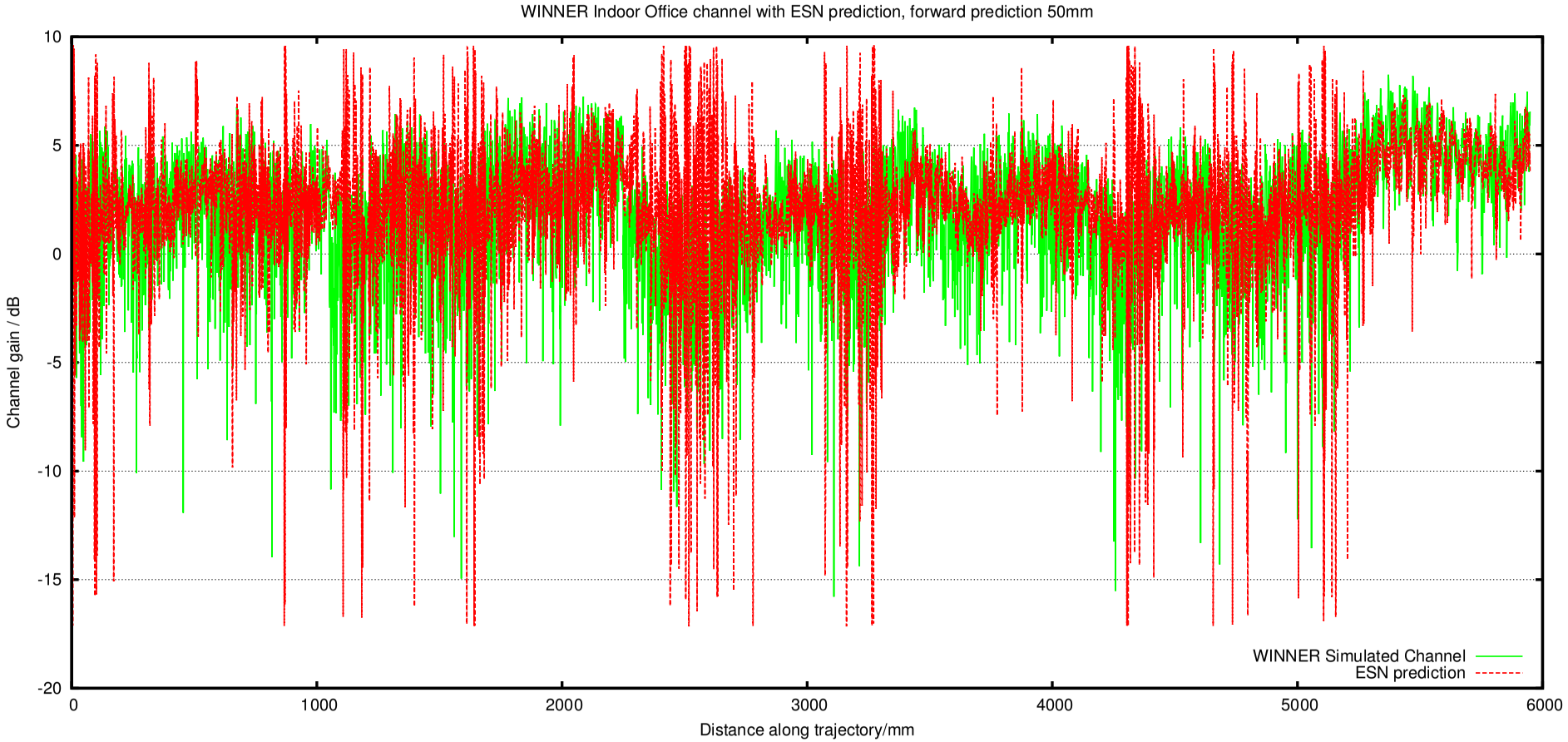


Figure 5.2: Results of an experiment to predict 50mm ahead for a type A1 (Indoor Office) WINNER channel. It employs the ESN in rolling prediction mode, where current conditions are provided as an input as each future prediction is produced. The highly variable nature of this channel is clearly visible, as are a number of very large prediction errors.

rapidity. This proved to be a problem for the ESNs used in the characterisation task (see Section 4.3.2), and the same was found to be true for channel prediction. In the particular instance shown in the figure, the Mean-Squared Error (MSE) over the entire distance was 0.85 dB, which seems quite a strong result at first impression. While the ESN's prediction output does track the approximate magnitude of the channel, it frequently mis-predicts the value, often by a large margin (errors greater than 10 dB are clearly visible).

In order to better understand the nature of these problems, the channel magnitude data was first smoothed to give a less variable signal. The mean value of the channel gain was computed over a window of 100 samples in length, eliminating some of the large variations of the channel, while still preserving the longer-term variations. This allowed the function of the ESN to become more evident. The results of re-running a prediction test under these conditions are shown in Fig. 5.3.

This second graph (Fig. 5.3) reveals more clearly an effect also found in the un-smoothed channel. The simulated channel is followed closely by the prediction which is seldom accurate. By plotting a graph showing the MSE between the simulated channel and horizontally translated versions of the predicted channel data, it is clear that the ESN is not making an effective prediction. Were that the case, it would be expected that the smallest error between the datasets would be found in the unmodified data (i.e. at zero shift). As is clearly visible, the minimum error occurs when the prediction is shifted by approximately 50 samples. These 50 points correspond exactly to the number of samples into the future that the ESN is trying to predict.

From this result, it seems that the trained NN is not producing useful predictions, but is doing little more than replicating the current input value at its output. This is visually evident, as the predicted trace is almost identical to the correct channel, but shifted laterally. Examination of the un-filtered data used to create Fig. 5.2 shows an identical pattern: that the ESN does not produce a useful prediction much beyond reproducing the input value at its output.

Fig. 5.4 makes this more easy to see. In a good predictor, the minimum error would occur at the point it is trying to predict for (i.e. 0 offset). In this case, the minimum error in the prediction occurs after 50 samples, with a near linear increase in error before and after this point. This graph makes it clear that not only is the ESN failing to provide a useful prediction, but that it simply uses the value from the input as its output.

Figure 5.3: A re-run of the experiment from Fig. 5.2 with the channel data being smoothed before being fed to the ESN. Although the two time series track one another quite closely, having this less noisy channel makes it clear that the ESN is not performing its task of predicting. As is evident, the predicted values lag behind the actual system's simulated measurements by a small degree, rather than matching it, as would be expected.

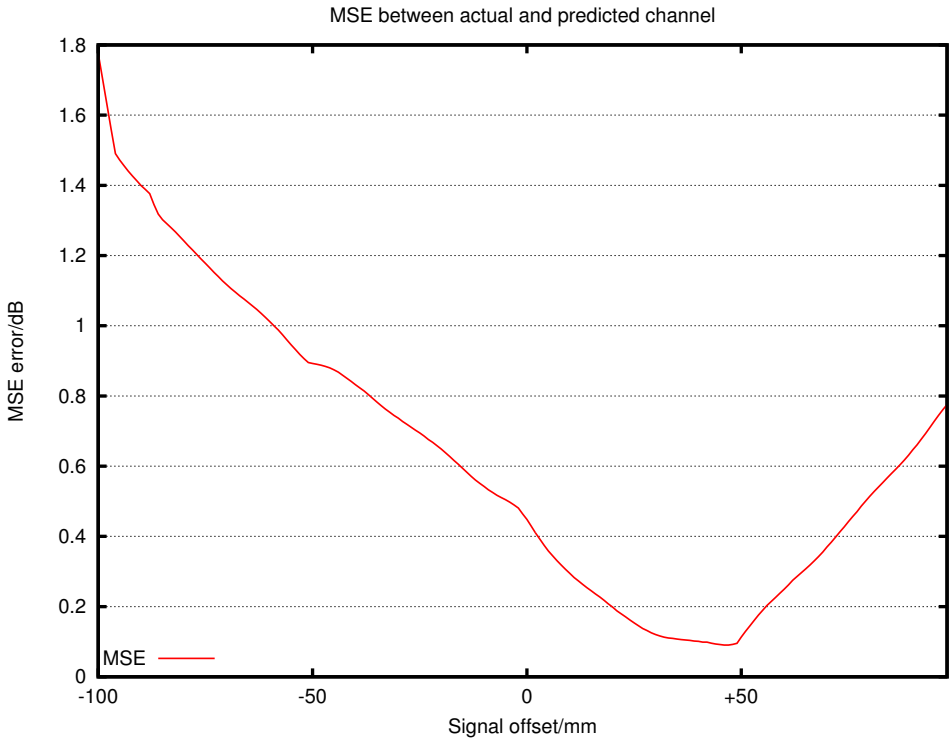


Figure 5.4: This graph shows the MSE between the smoothed channel's actual response and the ESN's prediction when shifted by varying amounts. In a good prediction, it would be expected that the smallest error value would occur when the signal is not offset. However in this particular setup, the minimum error value is found when the prediction is shifted by 50mm, the exact amount the network is supposedly trying to predict.

The preceding experiments were all performed with a reservoir neuron count of 1,000 nodes. To test whether the complexity evident in the channel data generated by the WINNER modelling suite was too great to be adequately captured by such a size of reservoir, larger reservoir sizes were investigated. The experiments were repeated with 2,000, 5,000, 10,000 and 20,000 reservoir neurons, but increasing the pool size (even to the huge numbers used here) showed no noticeable changes in the results, and only served to vastly increase runtime. In each case, the ESN did little more than pass the current input on as an output, although in a small percentage of tests the network's output would saturate completely.

Following these experiments, it became clear that using an ESN to predict the channels generated by the WINNER suite was not feasible. Since varying the ESN's parameters had no effect (e.g. number of neurons) it seems that they are simply not a tool suited to this task. It has been well established [28] that NNs are capable of producing correct predictions for any deterministic system, given enough representative

training data. The WINNER modelling suite produces channel model data which, although based on an geographical model of the scenario, there is a random elements to the results (i.e. is non-deterministic). In the previous task of channel characterisation, the ESN was able to average out the randomised portion of the data, and could classify a channel based on the geographical model. In the case of channel prediction, the non-deterministic nature of the signal means that no matter how much training data, or how many neurons the network was given, the task of performing channel prediction on channel generated by the WINNER framework is fundamentally not suited to computation using a NN. The same can be said of any statistical channel model which incorporates a random element, as many of those listed in Section 2.3 do.

Statistical channel models are a useful tool, and although they do match closely with measurements, they are based on the underlying principle that a channel environment is too complex to be accurately and correctly modelled, and so a parametrised random approximation will suffice. The following work removes this assumption, resulting in channel models which are entirely deterministic.

Since ESNs do however have a history of being able to predict non-linear systems, the decision was taken to change the channel modelling software in order to assess the capabilities of the ESN approach in the area of channel prediction, and perhaps better understand why the WINNER channel models could not be correctly predicted.

5.2.2 2-ray Model

Given the failure of the ESN to predict a realistic channel model, the following experiments were designed to begin with extremely simple channels, and evaluate how effectively an ESN system could predict such a channel's future behaviour. To this end, one of the simplest channel models was chosen, specifically a 2-ray model.

A 2-ray model determines the channel between a transmitter and receiver by considering it as the sum of a Line of Sight (LOS) ray and a second ray which is reflected once off the ground before it arrives at the receiver. Fig. 5.5 presents a visualisation of the setup of such a model.

As in the previous ESN prediction experiments, training data different to the testing data is used. These two data sets are produced by generating channel data from the 2-ray model twice, using two parallel trajectories through space. This is intended to give the ESN a representation of the system it is to emulate and predict, without simply

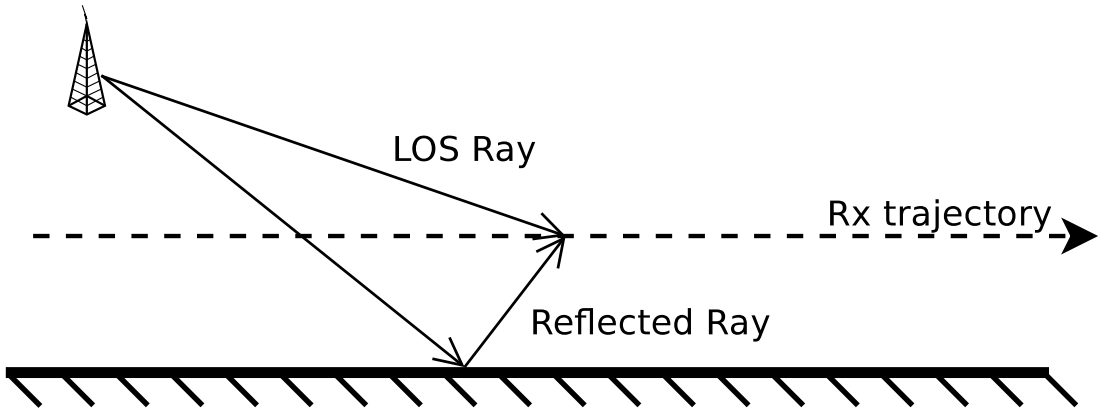


Figure 5.5: A diagram showing the simple 2-ray channel model used in Section 5.2.2. Using such a simple channel was necessary to assess how suitable ESNs might be for the task of channel prediction.

replicating the training data.

Predictive results for this simple channel model (as shown in Fig. 5.6) were promising, although initially the ESN time series does not produce correct predictions. This is likely due to a combination of uninitialised 'junk data' being left over in the ESN from training and a slow rate of convergence towards the new channel data. The graph shows a very close agreement between the model and the prediction.

By plotting the MSE between the two datasets (see Fig. 5.7) with the prediction shifted by varying amounts, it can be seen that the ESN is indeed doing a much better job in this instance than in Fig. 5.4.

The graph (Fig. 5.7) shows two features which would be expected in a good predictor. Firstly the minimum error value occurs very close to the zero offset position. Secondly, the erorrs become larger as the offset from the zero offset point becomes larger. Interestingly, a local maximum value occurs at approximately 70 samples. This can be explained by referring to the channel data in Fig. 5.6. The channel's magnitude response itself displays some periodicity, so shifting the prediction by this amount will cause errors to start to drop.

This encouraging result demonstrates that ESNs are a suitable tool for at least some channel modelling tasks. In order to assess how much complexity they can capture in a channel, a modelling tool with a number of variable parameters was needed. Specifically, the following set of requirements was drawn up.

The channel modelling software must be able to:

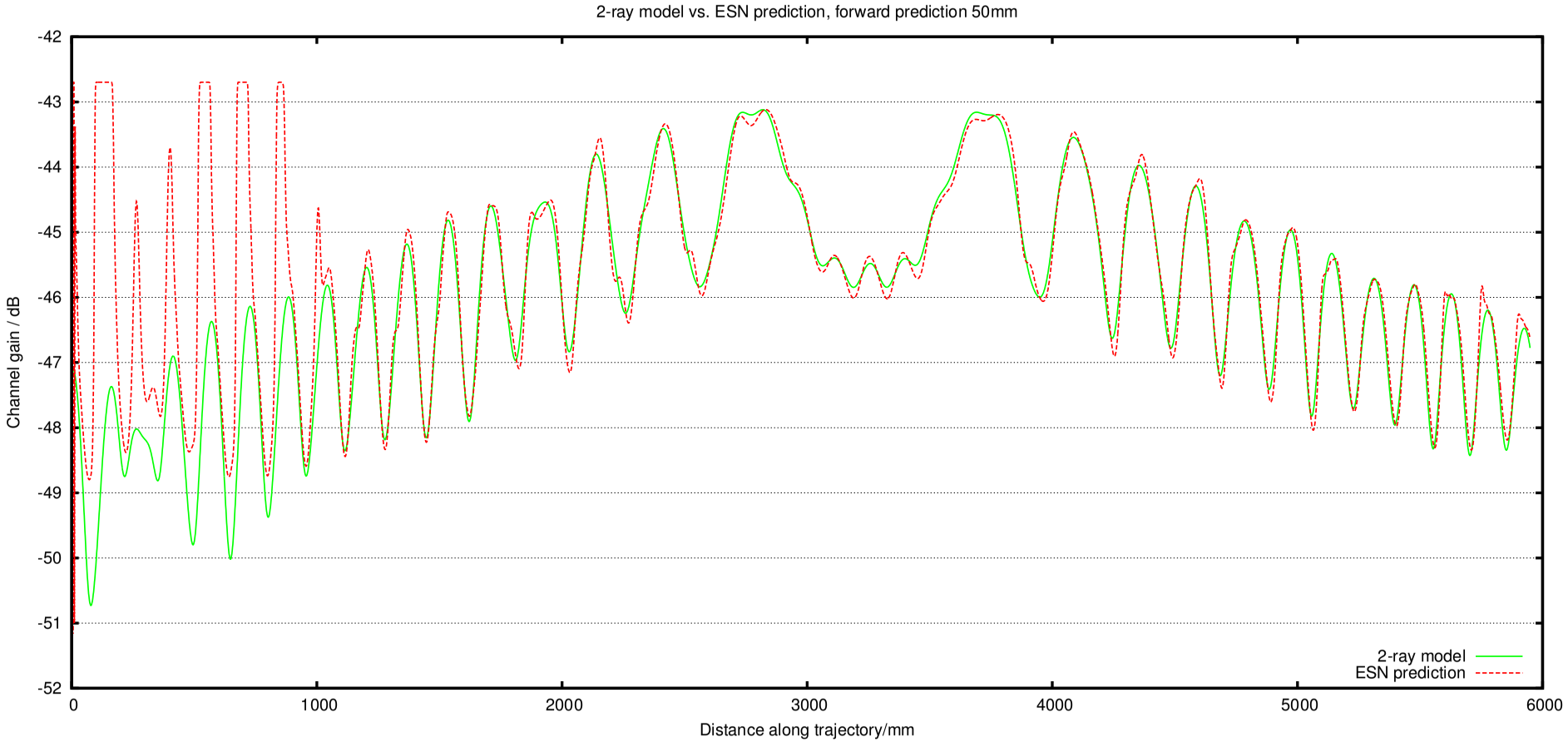


Figure 5.6: Graph showing a channel generated by the 2-ray model, and the output from a trained 1,000 neuron ESN attempting to predict 50 samples (50mm) ahead. After an initial period of poor predictions (including some evidence of saturation) the network is able to very accurately predict future channel conditions.

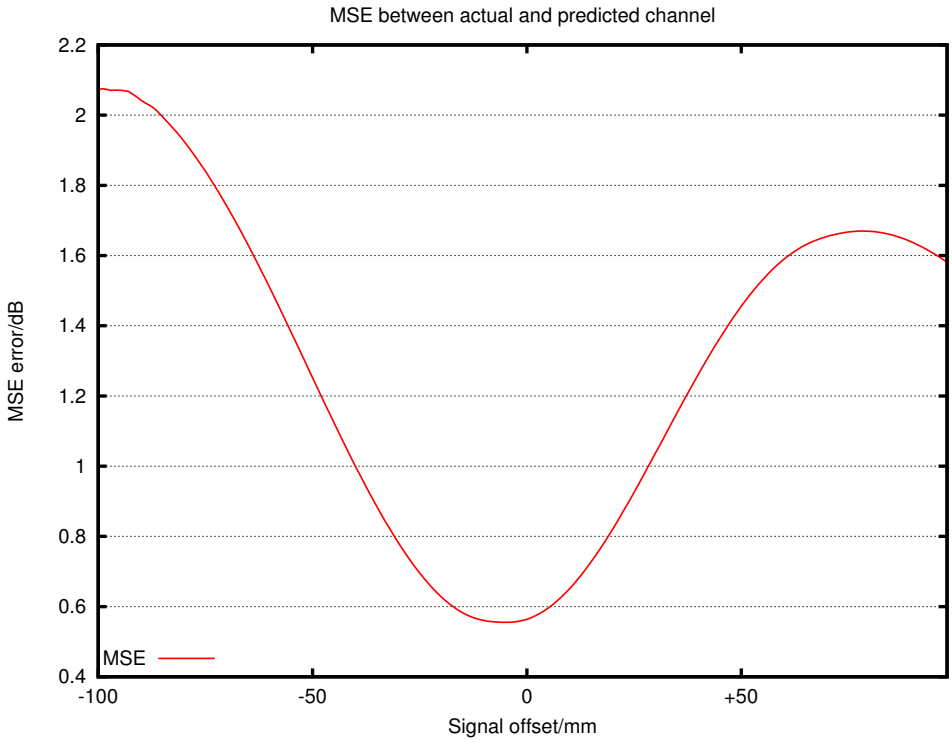


Figure 5.7: Graph showing the MSE between the 2-ray model and the ESN's prediction at a range of offsets. As would be expected from a functioning predictor, the minimum error value is found when the ESN prediction is not shifted. This is good evidence that the NN was able to capture the model's functionality during the training phase and then apply it to the slightly different scenario encountered during the prediction phase.

- Model an indoor narrowband channel
- Calculate LOS and first-order reflections
- Incorporate path-loss and reflection functions
- Add varying amounts of noise to the channel
- Produce channel models deterministically
- Provide data about each ray used in the calculation at any spatial point

After a search of available commercial and open-source Radio Frequency (RF) channel modelling tools, none could be found which provided the necessary features, so a new ray tracer was written to facilitate the generation of channels to train the ESN for the task of prediction. The following section describes this tool, which is used in subsequent experiments.

5.2.3 Ray Tracer Design

Since the WINNER channel models produce such a dynamic, rapidly-varying channel which ESNs were not capable of modelling accurately, a new method was needed which would provide channels whose complexity could be increased or decreased according to need. This would allow the predictive power of ESNs to be better understood in the area of wireless channels.

The ray tracer to generate channels of varying complexities was written in Matrix Laboratory (MATLAB). It uses a 3 dimensional model of a 6-sided room (4 walls, a floor and ceiling) to compute an adjustable number of rays from transmitter to receiver. Both the transmitter and receiver are modelled as being inside the room, and can be placed at arbitrary points as needed by simulations. Additionally the room size and reflectivity of each surface can be configured according to the scenario to be modelled. This setup was chosen because it closely resembles a typical deployment for a femtocell or WiFi access point serving a nearby mobile device. The model was designed to emulate transmissions within the Industrial, Scientific and Medical (ISM) band around 2.4 GHz, which is typical for a range of common radio technologies.

A crucial feature is that results from one simulation will be completely identical to a second simulation when run with the same input options, allowing accurate reconstruction of experiments. Unlike the WINNER suite, if any noise is added to the model, it will be deterministically generated from a pseudo-random number generator with a fixed seed (rather than being based on a random entropy source).

5.2.4 ESN Prediction using Ray Tracer Channels

By replacing the WINNER channel modelling software with the customisable ray tracer, a range of more finely controlled channels could be used to assess how much potential ESNs have for channel prediction.

Firstly, a model composed of 7 rays was used to calculate how a channel varies along a 6m long room. The rays represent a LOS component, and one ray with a single reflection from each of the room's six flat surfaces. The channel data was computed for two parallel trajectories through the room, one being used for training the ESN, with the other as the path to be predicted. The transmitter is located in the centre of the ceiling, and the receiver follows one of the parallel paths through the room. Results from a single run of this setup are shown in Fig. 5.8.

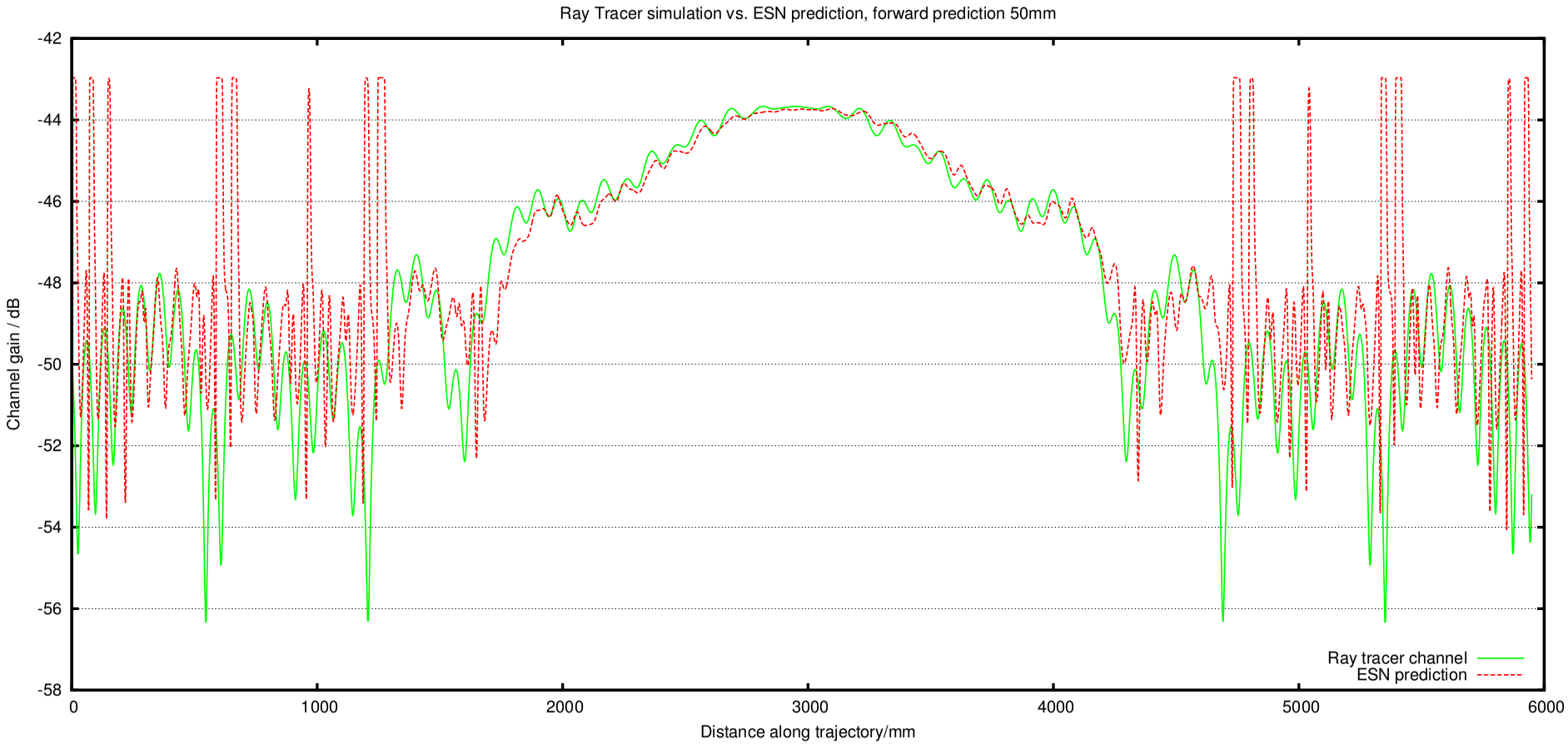


Figure 5.8: Results of using an ESN to predict 50mm ahead along a 6 metre trajectory though a room. The ESN had previously been trained on a different path through the same room. It is clear that in areas where the channel's gain is least variable (in this case the centre of the room) the prediction matches most closely with the ray tracer's results. In areas of higher variability the ESN is less able to accurately predict future conditions. The extreme high values show points at which the NN's output neuron reached saturation.

Visually, the results are quite promising, as the generated channel and the predicted channel are generally quite closely matched. There are however some very obvious instances of saturation. It is clear that they occur in areas where the channel is varying more rapidly over distance due to patterns of constructive and destructive interference. The rapidly varying nature of the channel seems to be what caused ESNs to fail to predict the WINNER models accurately, and a similar effect is evident here.

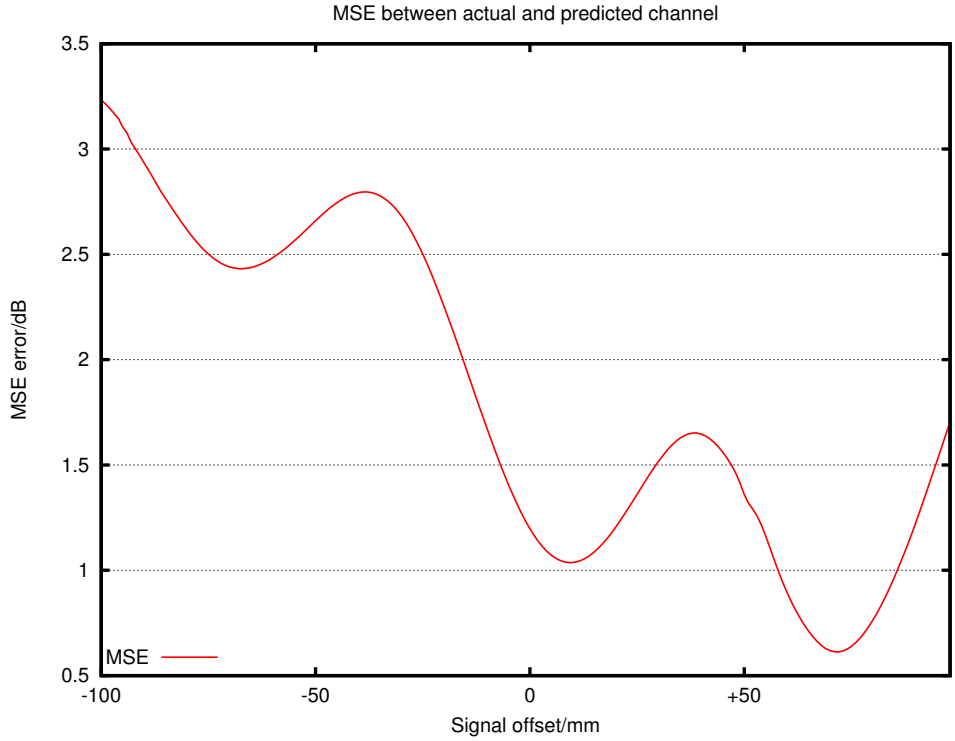


Figure 5.9: Graph showing the MSE between the ray-tracer calculated channel and shifted versions of the ESN's output from the experiment shown in Fig. 5.8. As was the case when the ESN was trained on a WINNER channel model, the smallest error value occurs when the prediction has been shifted. However, there are two noteworthy features. Firstly, that the minimum error is not at a delay of exactly 50 samples, indicating that the ESN is doing some form of computation beyond replicating its input value. Secondly, and more importantly, there is a noticeable local minimum around the point where the signal is not shifted, a good indicator that a reasonably accurate prediction is being made.

Fig. 5.9 shows a plot of the MSE between the calculated channel and the predicted values when shifted by varying amounts. This graph illustrates features different to previous error graphs (Figs. 5.4 and 5.7). While Fig. 5.4 showed a minimum error value at 50mm (i.e. the exact amount the ESN was supposed to predict ahead), this is not the case here. It is in fact even further ahead than would be expected, illustrating that some sort of processing is occurring within the ESN, rather than simply directly

propagating the input value to the output.

The second point of interest is that there is a significant local minimum around the zero shift point. While Fig. 5.4 showed an almost linear decrease in error until the minimum value, the local minimum seen here is a clear indicator that some degree of prediction is being made by the ESN, giving a better result than simply replicating the value at the network's input (as happened in the experiments using the WINNER models).

To better understand the observed effects, simulations were performed which used between 2 and 7 rays to simulate the channel along the length of the room. Since the ESN performed well with a 2-ray model, but achieved only limited success with a 7-ray model, these experiments were carried out to determine the level of channel complexity the ESN could model accurately.

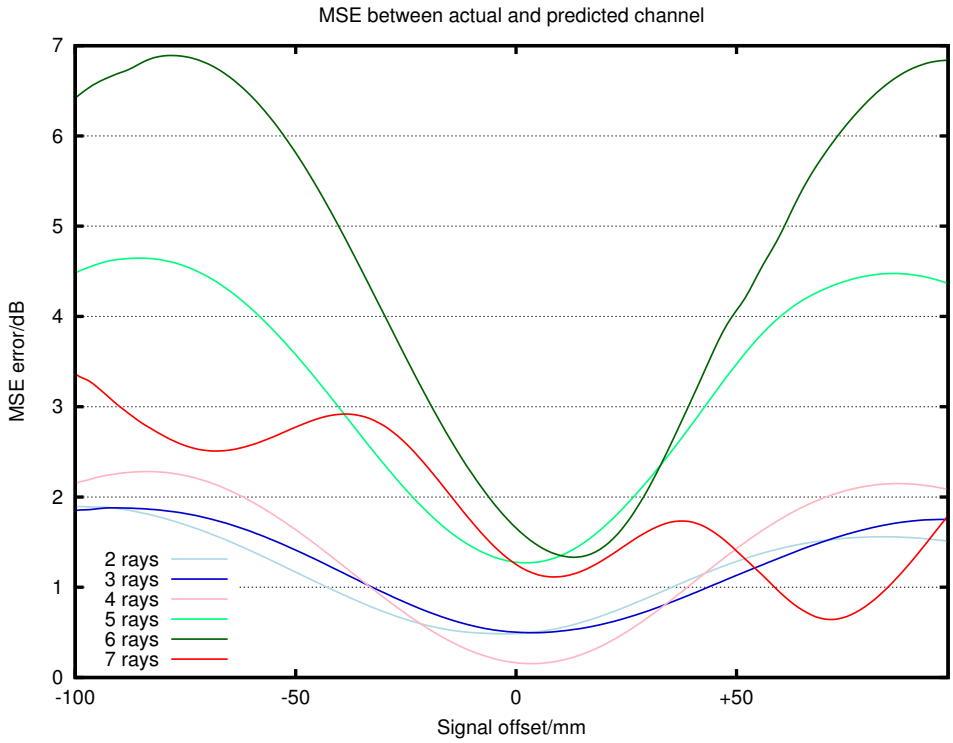


Figure 5.10: Graph showing MSE data at a range of offsets from prediction experiments with a varying number of rays used to simulate the channel. It is clear that channel models with a smaller number of rays (i.e. simpler models) are more easily predicted by the ESN than those with a larger number of rays.

Fig. 5.10 illustrates that simpler channel models (i.e. those computed using fewer rays) are more accurately predicted by the ESN system, while the higher MSE values associated with more complex models indicate greater inaccuracy. Interestingly, all the

channel models except for the 7-ray model have their minimum MSE values at (or close to) zero shift. This suggests that the limit for channel modelling complexity with this ESN setup has been reached.

While the advantages of using NNs (and ESNs in particular) have already been highlighted (the data-driven approach, their simplicity, their ability to approximate arbitrary systems and the ESN's efficient training operation), it is clear that they are subject to limitations as well. Although the ray-tracer used in these experiments is by design an entirely deterministic system, ESNs are not able to reliably predict its behaviour when even seven rays are used to model a channel. This effect is seen much more dramatically when using the WINNER modelling software. In those experiments, the best solution the ESN's training algorithm (which aims to minimise the error between the training data and the network's output) could find to the prediction problem was to use the current input value as its prediction. Admittedly, the WINNER models take a very large number of parameters into consideration when generating the channel data, and it is not a deterministic system as random noise is introduced, however there is still an underlying model to the data, which the ESN was unable to infer.

The question of what the minimum acceptable error between reality and prediction depends on a number of factors, all of which are dependent on the particular radio system being implemented. The modulation type for example will have a large impact on the minimum acceptable error, as clearly a system using QPSK to moulate its data could accommodate a much larger error in its channel prediction algorithm than one using 64QAM or 256QAM. The higher level binary coding scheme used would also influence this decision, as employing a coding scheme with good error correcting abilities could compensate for a channel prediction scheme with an otherwise unacceptably large error.

This set of experiments provides strong evidence that, although the echo state approach has merit (both in channel characterisation and the prediction of very simple channel models), it does not hold much promise for the prediction of channels in real-world radio environments. Even though larger networks are able to model systems with more inherent complexity, the computational demands of such large networks make them prohibitively slow when simulated on current workstation-grade computers. Integrating an ESN of this scale into a power-constrained mobile device to improve channel prediction performance is currently not a realistic proposition.

For these reasons, a different approach was taken to channel prediction in the remainder of this chapter.

5.3 Linear Ray-based Prediction

In light of the negative results produced by the ESN prediction scheme, an alternative approach was taken to channel modelling. In this section, the impact the physical environment has on the channel is examined, and a new ray-based prediction technique is presented. Although factors such as background noise and interference from other systems do contribute to channel effects, physical factors such as the positions of walls, floors, reflectors and scatterers often play the dominant role. Since these are usually static (or change very slowly on the channel's timescale) their effects can be modelled as being static.

This assumption of a static environment is taken by ray tracing tools. They model a channel's physical environment by simulating a large number of rays and computing how they interact. This method requires a huge amount of detailed information about the environment's layout, its composition and a wide variety of other variables often making it infeasible to model with sufficient accuracy. The following approach (although it shares some characteristics with ray tracing) requires less information about the physical environment, making it a better suited computational approach than ray-tracing, especially when faced with a real-time processing constraint.

In order to make predictions about a channel's behaviour in the spatial domain, a number of assumptions are made for this model. It is assumed that:

- the channel can be approximated as being composed of a discrete number of rays
- the phase and magnitude information for each individual ray at the receive antenna is available
- for the duration of the prediction simulation, the number of rays is constant, with no existing rays being cut off, or new rays being received

Our system model incorporates a fixed transmitter and a mobile receiver within the same room. This was chosen to model a typical indoor femtocell deployment or WiFi router and client device.

The prediction scheme first measures the phase and magnitude of each ray at two spatial points, A and B separated by a known displacement vector (i.e. distance and direction) as shown in Fig. 5.11. The difference in phase and magnitude between the two points for each component ray is used to create a linear extrapolation in the spatial

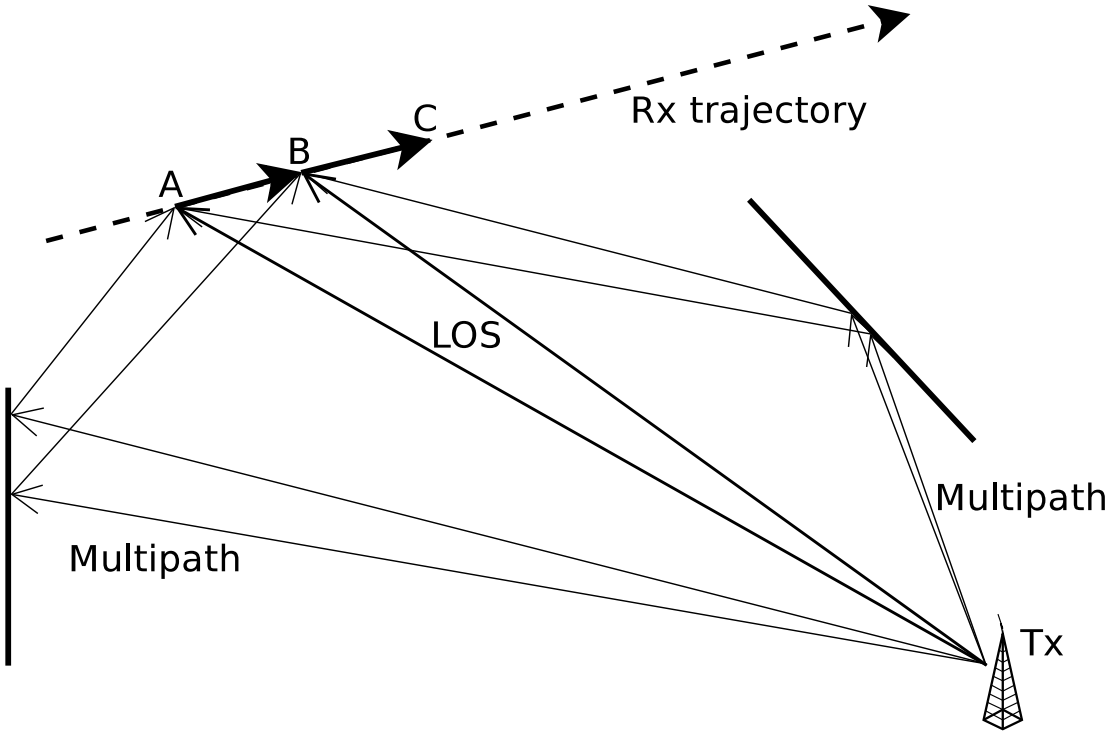


Figure 5.11: Simplified system diagram showing the receiver's trajectory and three rays coming from the transmitter, including one LOS and two multipath components. A and B are the two spatial points at which each of the individual rays are known, while point C is the point for which a prediction will be made. The channel from the transmitter (Tx) to the receiver (Rx) is modelled as being composed of the sum of these three rays.

direction AB for each of the individual rays. By using the approximation that the change in path length (and hence phase) for each ray will increase linearly with distance along the trajectory, the individual ray's phase can be predicted at any given point along its current track. In a similar fashion, path loss (due to changing distance) is modelled as a linear change in magnitude over distance.

The rays at the two points, A and B on which the prediction is based can be written as:

$$Ae^{j\theta^A} = \sum_{i=1}^{I} a_i e^{j\theta_i^a} \tag{5.1}$$

$$Be^{j\theta^B} = \sum_{i=1}^{I} b_i e^{j\theta_i^b} \tag{5.2}$$

with I discrete rays being used in the calculation. a_i, b_i correspond to the gains, and θ_i^a, θ_i^b the phases of each ray. Hence, their difference can be calculated by:

$$\Delta e^{j\theta^{\Delta}} = \sum_{i=1}^{I} (b_i e^{j\theta_i^b} - a_i e^{j\theta_i^a}) \tag{5.3}$$

This Δ calculation is then used to predict the next data point (corresponding to point C in Fig. 5.11). This Δ can then be used iteratively to form a prediction for any point along the receiver's trajectory. Similarly, this calculation could also be used to predict in the opposite direction, to calculate previous channel conditions if required, of if using a multi-antenna array.

Data: Two complex valued channel measurements (A and B)
Result: Complex valued prediction (C)

phase_delta ← angle(A)-angle(B);
magnitude_delta ← abs(A)-abs(B);
C = PolarToComplex(angle(B)-phase_delta, abs(B)-magnitude_delta);

Figure 5.12: Psuedocode to generate the prediction for the point C for a single ray, when given In-phase and Quadrature (IQ) data about that ray at points A and B

Fig. 5.13 shows the magnitude response of four particular rays, and the complete channel along with the predictions. In some cases, the predictions for a particular ray are very good while in other cases a pronounced divergence is visible as distance increases. The bottom graph displays the magnitude of the summation of all the rays (of which only 4 are shown), demonstrating that the predicted channel is able to achieve a very close prediction indeed, although as distance increases and the prediction errors in the individual rays sum together, the predictive success of the whole channel suffers.

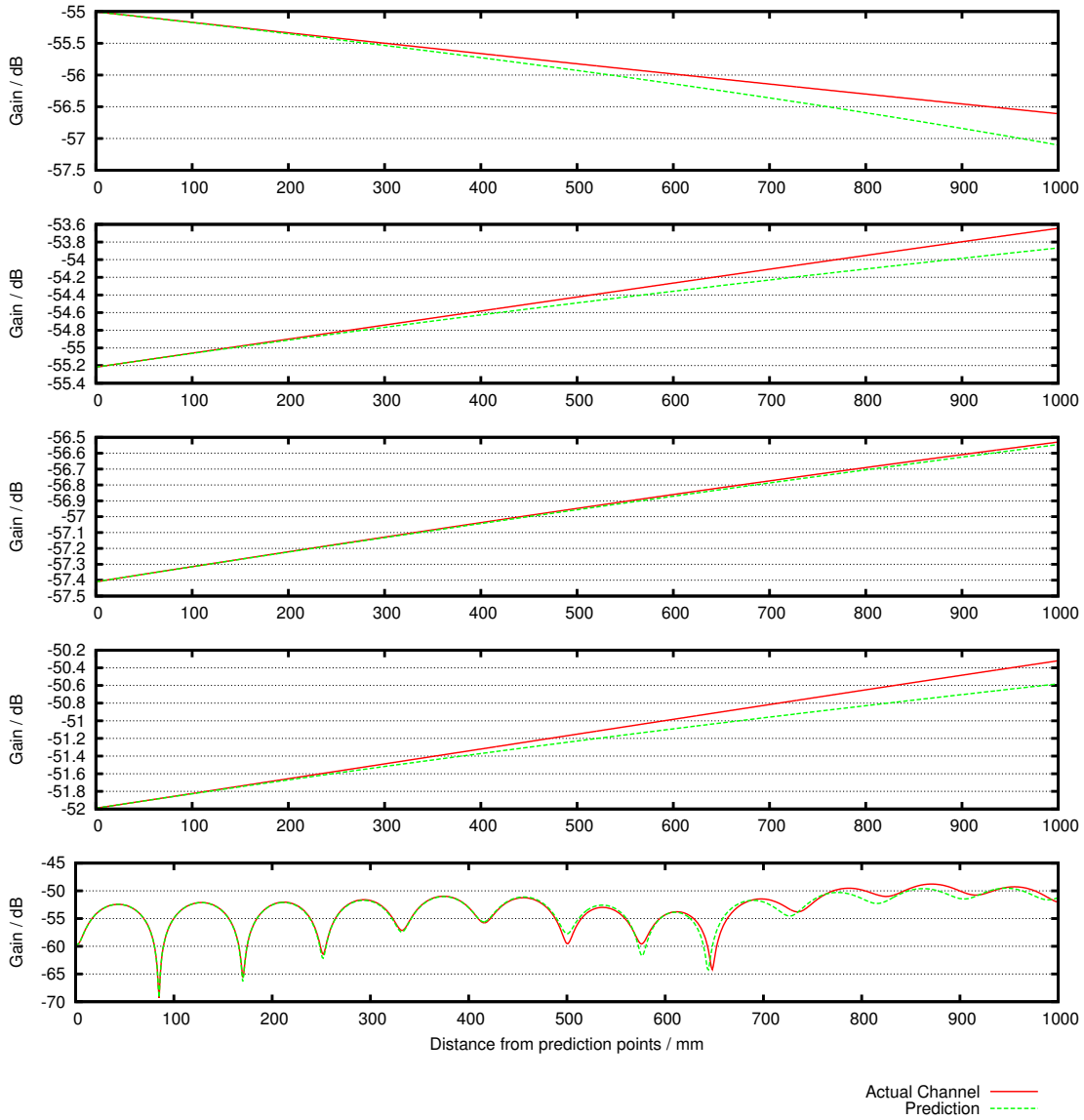


Figure 5.13: Each of the top four graphs show how an individual ray's magnitude response changes over distance - both its correct value (in red) and the predicted value (in green). The difference between the purely linear model from the prediction scheme and the non-linear real-world response becomes increasingly visible as distance increases. The bottom graph shows the overall channel's magnitude response calculated from the sum of the all the component rays. The prediction scheme performs the same interpolation operation on the phase data to produce that prediction. (omitted from this diagram for clarity).

5.3.1 Prediction Results

Since this research depends on knowing the individual component rays comprising a channel, a ray-tracing computer model seems well suited, as it allows trivial extraction of data on each individual ray without the need for advanced processing. The ray-tracer

developed previously for making simplified channels was found to be a good solution, and was re-used.

The scenario has a fixed transmitter inside the room and a mobile receiver which moves on a randomly assigned straight line path through the room. Two consecutive readings of the rays are taken along the receiver's path and the are used as points A and B in the prediction algorithm. The prediction algorithm's performance is then compared with the ray-tracing result (which is treated as error free). Noise is additionally introduced into the simulation system. By specifying an Additive White Gaussian Noise (AWGN) power, it aims to simulate the real world more closely. The AWGN is used to model the many small reflectors a real room would have, representing the multiple household or office items which cause more complex and unpredictable reflections than those from the walls or other large flat surfaces.

Fig. 5.14 shows the output from a single run of the simulator, comparing the simulated channel to the prediction of that same channel based on just two prediction points, A and B.

In order to assess the comparative performance of this prediction scheme, it is compared to a computer-simulated Orthogonal Frequency Division Multiplexing (OFDM) system which can use pilots in order to actively measure the current channel conditions and correct the received signal accordingly. For additional comparison, also included are the optimal solution (where the OFDM receiver has access to perfect channel knowledge) and the case where no pilots or predictions are used in decoding the data.

First a random stream of binary data is generated, which is convolutionally encoded, interleaved and converted into an OFDM frame (see Fig. 5.15). This frame is then transmitted over the channel model provided to the simulator by the ray tracer with the addition of a specified noise power (modelled as AWGN). Next the frame undergoes demodulation; if pilots are being used, they are used to correct the subcarrier data, however if channel prediction is used this alternative data is employed instead. Next, it is deinterleaved and passed through a Viterbi decoder, resulting in a binary data stream. This is finally compared with the binary data originally encoded into the frame before transmission, and the number of bits which differ are counted.

This process is repeated a large number of times, all the while keeping track of the total number of bits transmitted and the number of erroneous bits for each Signal to Noise Ratio (SNR) and scenario. When a sufficiently large number of repetitions have been completed (Fig. 5.16 and 5.21 were created using 10,000 packets per data point),

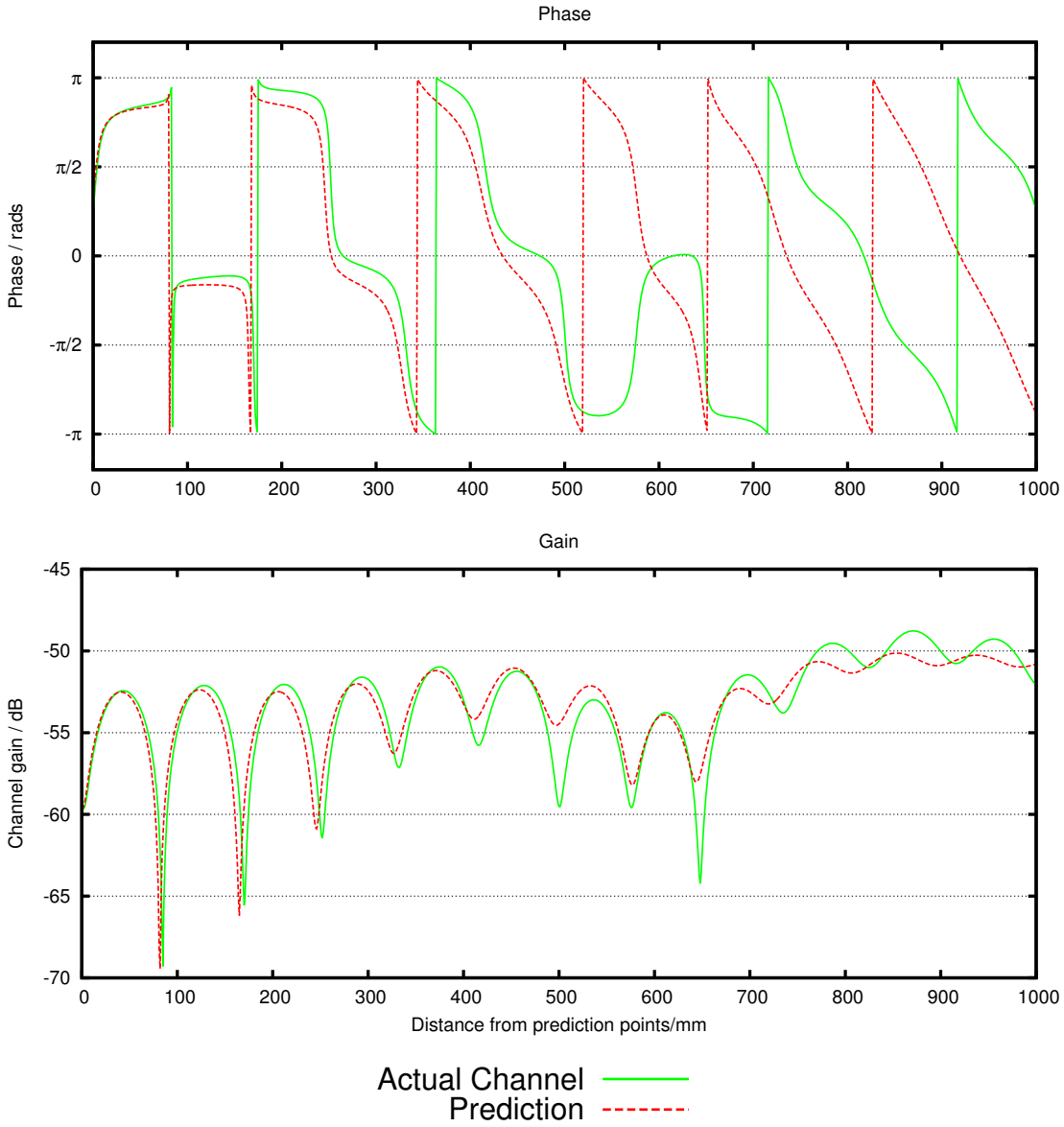


Figure 5.14: A plot of phase and magnitude response of one channel taken from the dataset used to build Fig. 5.17. It shows the actual channel data (from the ray-tracing software) and the prediction. It can be seen how prediction accuracy decreases as distance increases from the two initial measured data points. Note particularly how the accuracy of deep fade depth prediction reduces as distance increases.

the results are collated, and the Bit Error Rate (BER) for each SNR and scenario is calculated. The equation used for calculating the error rate is:

$$\text{BER} = \frac{\text{bits in error}}{\text{total number of bits transmitted}}$$

For each scenario in this simulation, the OFDM frame uses 16-Quadrature Am-

plitude Modulation (QAM) modulation on each subcarrier and the cyclic prefix is set to 25% of the frame length. The modulation order was chosen as 16-QAM as it a commonly used format in wireless OFDM systems, while a 25% cyclic prefix is also a common value. Each frame includes 4 pilot signals, equally spaced in frequency among the data subcarriers, as illustrated in Fig. 5.15. For the prediction scheme, the two known points are assumed to have been calculated with ideal channel knowledge. The simulation results are presented in Fig. 5.16.

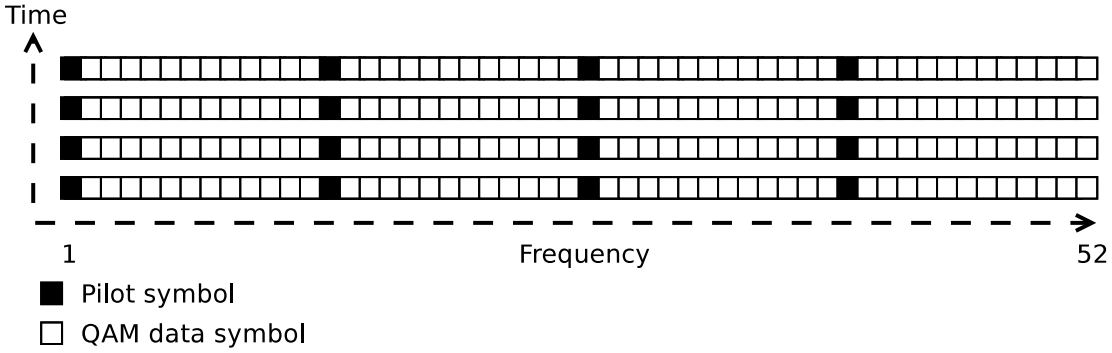


Figure 5.15: Structure showing OFDM frames, and how pilots signals are spaced among the 52 subcarriers.

As would be expected, the perfect channel knowledge scenario achieves the lowest error rate in all SNRs, and the zero channel knowledge scenario always has the highest error rate. Also, in line with the results in Fig. 5.17, since prediction accuracy degrades as distance from the prediction points increase, the error rate increases.

The first prediction data set (at 0cm from the known points) produces near-ideal performance, since it is based on two points with ideal channel knowledge. As can be been seen clearly from the graph, its BER is almost an order of magnitude lower than pilot-assisted OFDM, which represents a typical best-in-class practical system. Achieving such gains in a real-world scenario would be highly advantageous.

At a distance of 10cm from the known points, the results are comparable to those from the pilot-assisted OFDM simulation, showing that the system can achieve similar BER performance to an active channel sampling system without the need for repeated pilot signals. In this particular implementation, the 4 pilot signals could be replaced with more data carrying QAM symbols, increasing the data carried by the frame by almost 17%. This could be achieved with no increase in power usage or decrease in error performance. The power usage would in fact be fractionally lower, as pilots are transmitted with maximum power.

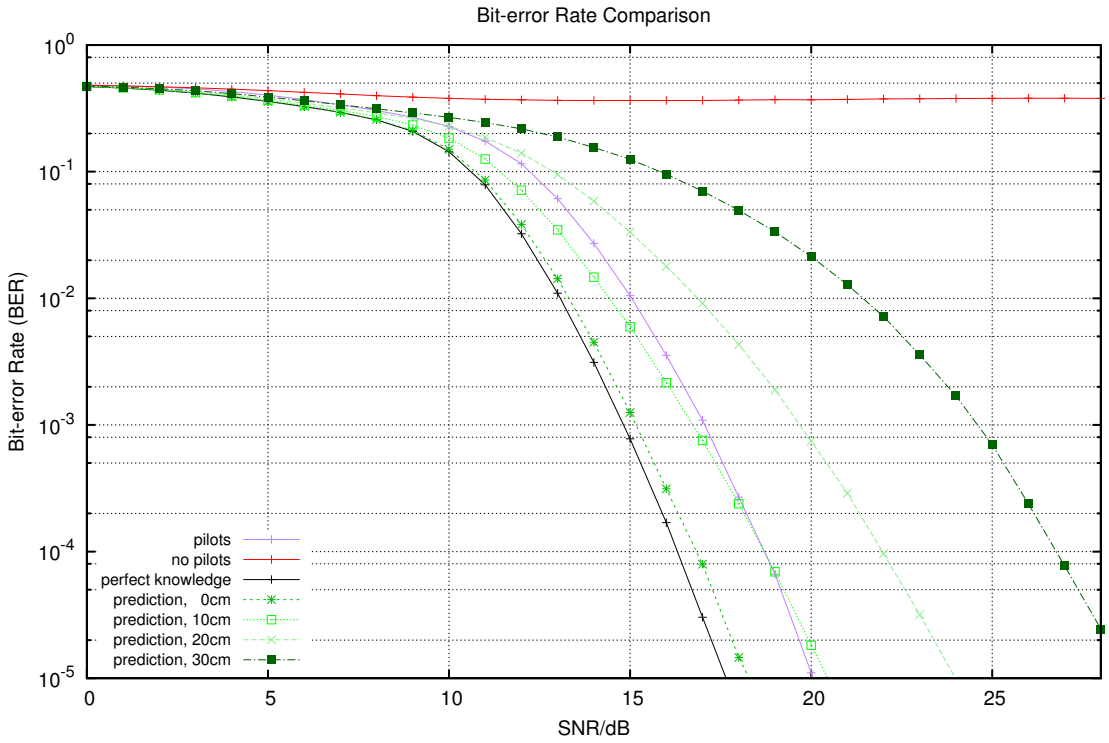


Figure 5.16: A plot showing the BER at a range of SNRs for four different scenarios: imperfect channel knowledge from pilot signals, no channel knowledge from pilot signals, perfect channel knowledge, and predicted channel knowledge (at four distances). It is interesting to note that the performance of the prediction system is marginally better than even the pilot-assisted system at 10cm. The MATLAB code used in calculations is included in Appendix B.

The predictions at 20 cm and 30 cm from the known points do not perform as well as the reference pilot-OFDM system, so a real-world system would have to choose the optimal distance over which to trust the predictions before errors become too large. This could be done by starting a new set of predictions after a certain distance had been travelled, or when the bit error rate starts becoming larger than a pre-determined threshold.

Fig. 5.17 has been generated by Monte Carlo simulation of 10,000 iterations per data point, with SNR from 10 dB to 60 dB. It shows how prediction accuracy decreases as distance increases from the two known points, A and B, under a wide range of noise conditions, showing that even at extremely favourable SNRs, prediction accuracy displays the same features as at more realistic SNRs. These results show an increase in errors at particular spatial points. These correspond to where deep fades occur in the channel. Fig. 5.14 shows that although the prediction scheme does identify these deep fades, the accuracy of fade depth prediction deteriorates as distance from the two

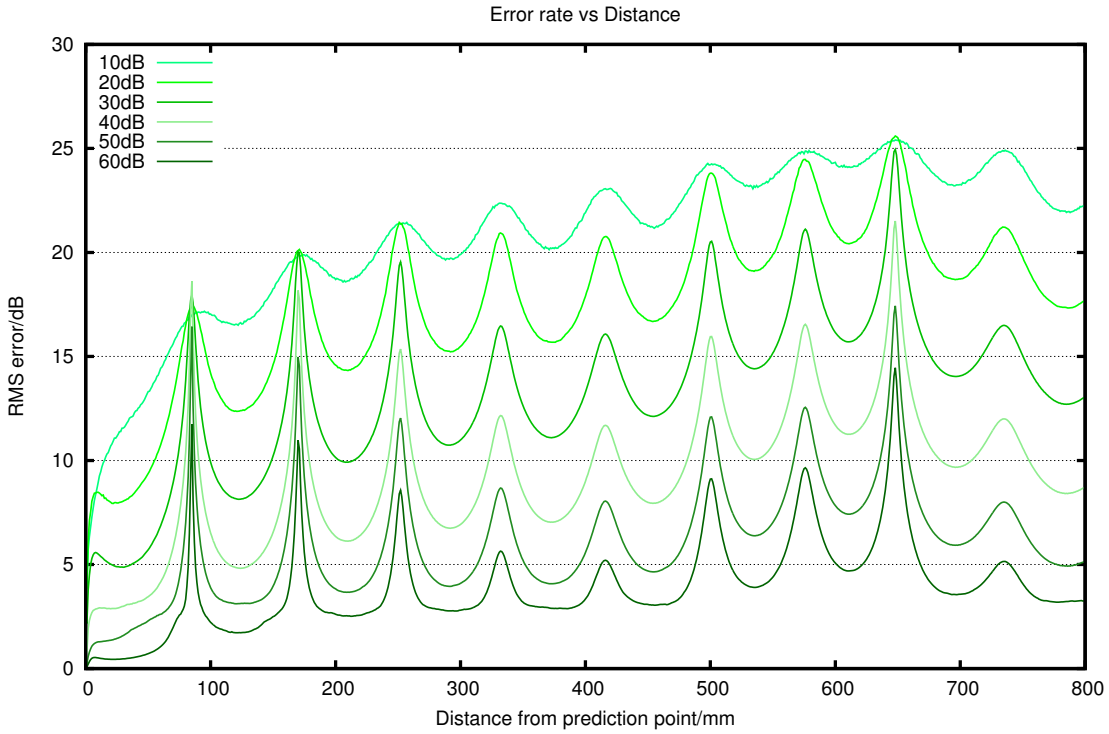


Figure 5.17: A diagram showing error rates as distance increases for a particular simulation scenario with varying SNR. It clearly shows the relationship between SNR and prediction error rate. Also evident are the areas of deep fading when the prediction of the fade depth is not as accurate as in other areas.

known points increases. This feature can be seen in the the gain plot in Fig. 5.14 at a distance of 650mm.

5.4 Polynomial Ray-based Prediction

The following section extends the previous research with the aim of further improving prediction accuracy. As is visible from Fig. 5.13, the earlier assumption, that ray magnitude and phase change linearly with distance, introduces errors. One error is introduced by using the approximation that the length of a ray path increases linearly with trajectory distance. Although this is true for LOS rays, the distance travelled by a ray reflected off a wall is non-linear with respect to the receiver's trajectory. Treating it as linear is analogous to using the approximation $\sin(x) = x$ which is only useful for small values of x, but otherwise breaks down. Path loss functions are also non-linear functions over distance; for example the Free Space Path Loss (FSPL) is related to the inverse of the distance squared [45].

With these factors in mind, the earlier prediction method is here extended using

higher-order polynomials. The two-point interpolation method described so far is identical to using n=1 order polynomial interpolation, so extending the method to higher order polynomial interpolation is a logical next step.

Using the standard method for Vandermonde polynomial interpolation, it is possible to produce a characteristic polynomial which passes through each of the observed points for a given ray. This method is able to model the behaviour of each ray much more accurately than the linear method. Although this method does require more computational resources than a simple linear (two-point) interpolation, the complexity of the newer method is still trivial for today's processors. While accurate predictions can be achieved with this new method, this comes at the expense of measuring the rays at n+1 known spatial points.

Data: Array of complex valued channel measurements (known_points), the polynomial degree (n) and the number of points to predict ahead (n_points)
Result: Array of complex valued channel predictions (prediction_IQ)

poly_m ← VandermondePolynomialGenerator(abs(known_points) , n);
poly_p ← VandermondePolynomialGenerator(angle(known_points) , n);
predictions_p ← EvaluatePolynomial(poly_m , n_points);
predictions_m ← EvaluatePolynomial(poly_p , n_points);
prediction_IQ ← = PolarToComplex(phase, mag);

Figure 5.18: Psuedocode to generate predictions for a single ray when supplied with an array of measurements (`known_points`). The degree of the polynomial can be easily specified by setting the variable `n` appropriately. Setting `n` to 1 gives the same result as linear prediction (as outlined in Fig. 5.12) while higher orders of polynomials can improve prediction performance.

5.4.1 Prediction Results

The same simulation scenario is used here as was employed in Fig. 5.16, however this time the polynomial prediction algorithm (with $n = 3$) was used, rather than the simple linear version.

As can be seen in Fig. 5.19, much improved results can be achieved with the polynomial method. The most important difference between it and the linear method is that it requires a larger number of known data points to produce the prediction. For a given degree of polynomial, n, measurements are required at $n + 1$ spatial points. In experiments, it was found that the most dramatic improvements are made when using $n = 3$, while further increasing n gives only marginal improvement. In the case of $n = 3$, ray

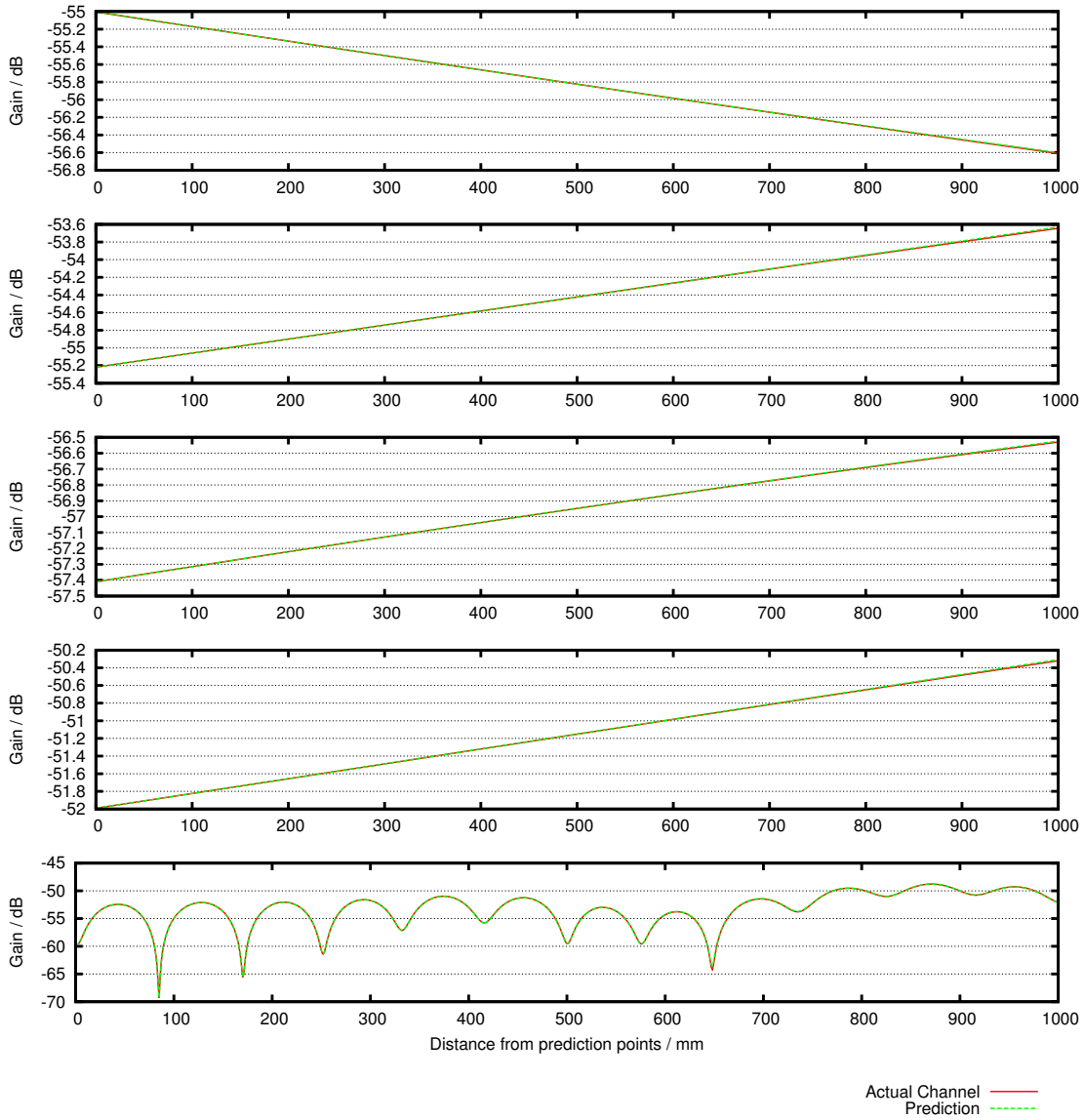


Figure 5.19: This plot shows the same channel scenario and rays as Fig. 5.13, but uses cubic (n=3) polynomial interpolation. Rather than using two known points to make the prediction, (previously referred to as A and B) four points must be sampled. However, it is immediately visually obvious that this predicts more closely than the linear (n=1) scheme. It is of special interest that the deep fades are much more accurately predicted when using polynomial interpolation, albeit at the expense of requiring more sampled data points.

data is needed at 4 spatial points, rather than the 2 required by the linear (equivalent to $n = 1$ polynomial) method.

Fig. 5.21 illustrates the dramatic improvement achieved by using 4 prediction points (i.e. $n = 3$ polynomial) over the linear method presented earlier. In order to demonstrate that the prediction scheme operates independently of the modulation type chosen,

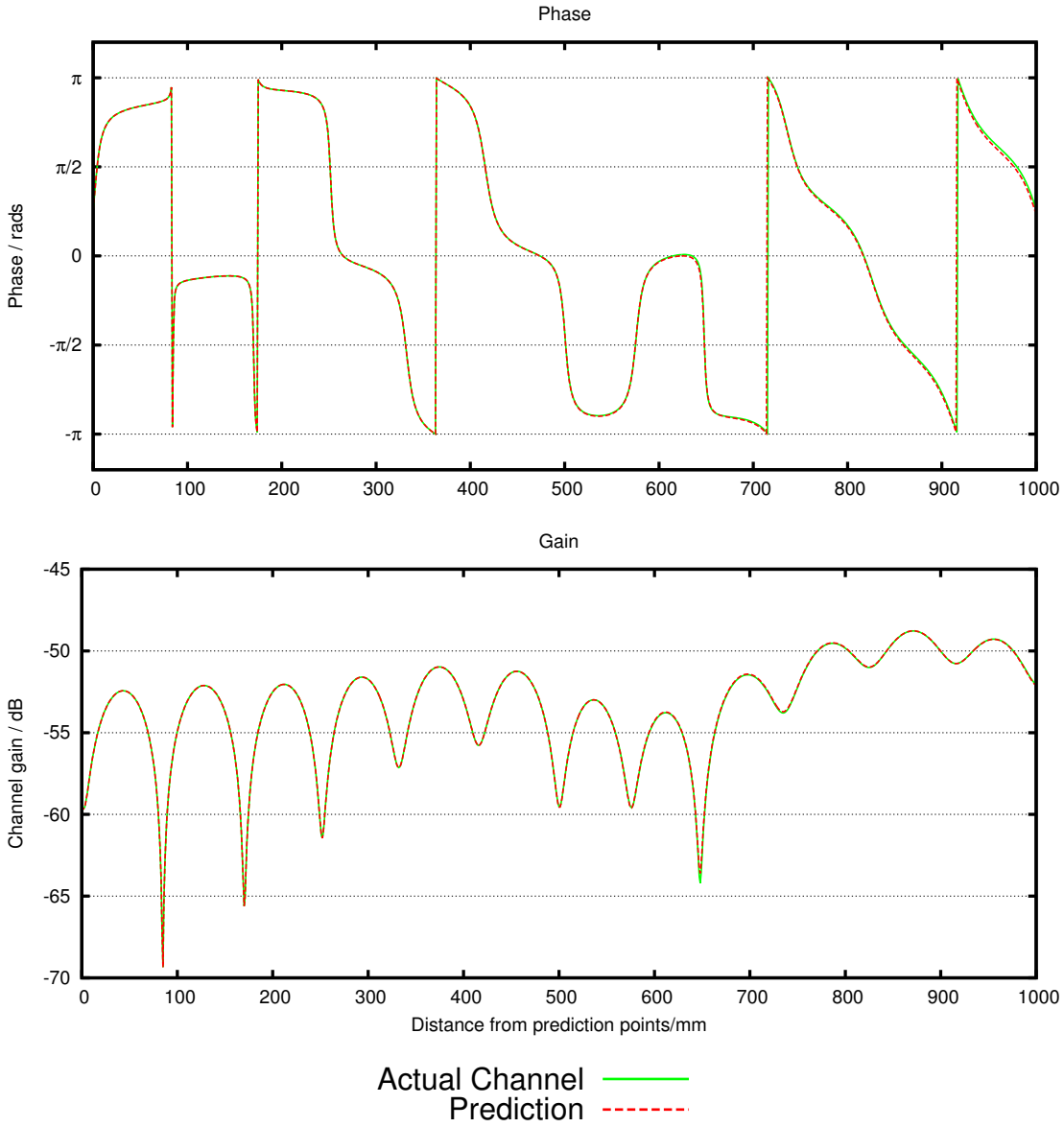


Figure 5.20: Prediction of the same channel shown in Fig. 5.14, but using n=3 order polynomial interpolation, with ray data collected at 4 spatial points. Over the distance shown the largest error in magnitude prediction is 0.5dB, marking a significant improvement over the previous result, at the cost of gathering more ray data.

the same simulation was re-run, but each symbol in the OFDM frame was modulated using 64-QAM, rather than 16-QAM. The results are shown in 5.22.

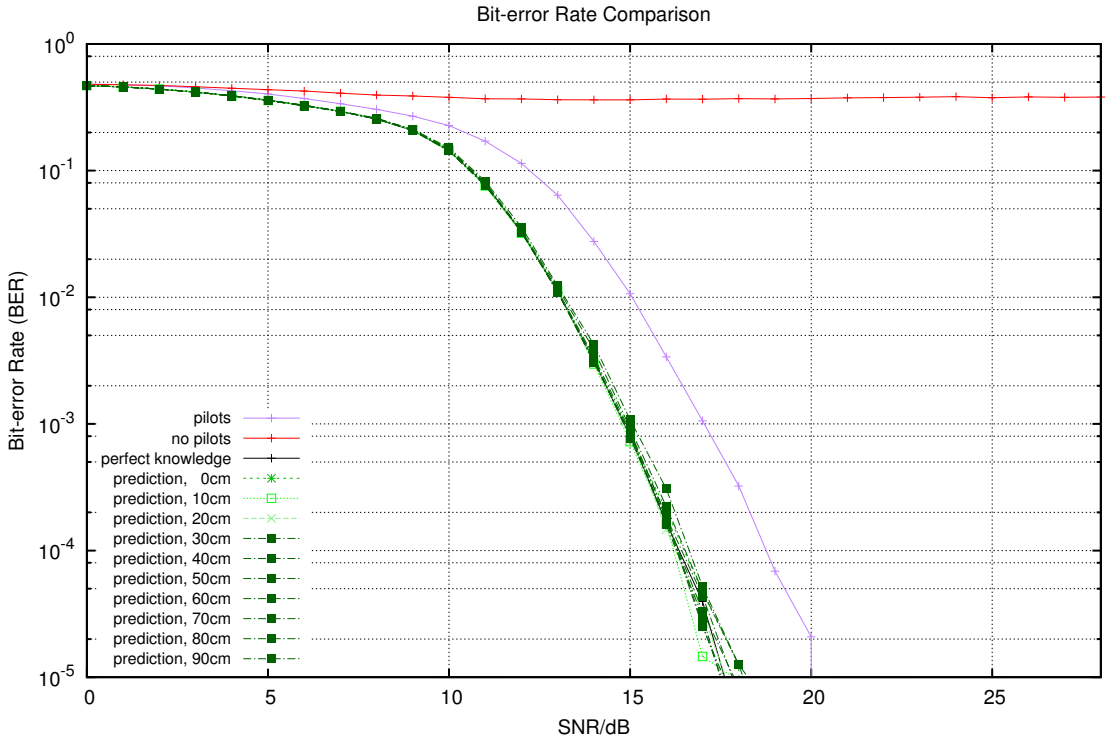


Figure 5.21: A plot showing BER over the same range of SNRs and scenarios as Fig. 5.16, but this time using polynomial $n = 3$ predictions. It is immediately clear that the polynomial prediction method is far superior to the linear method initially developed. It far outperforms the pilot-assisted system, even at 90cm from the prediction points.

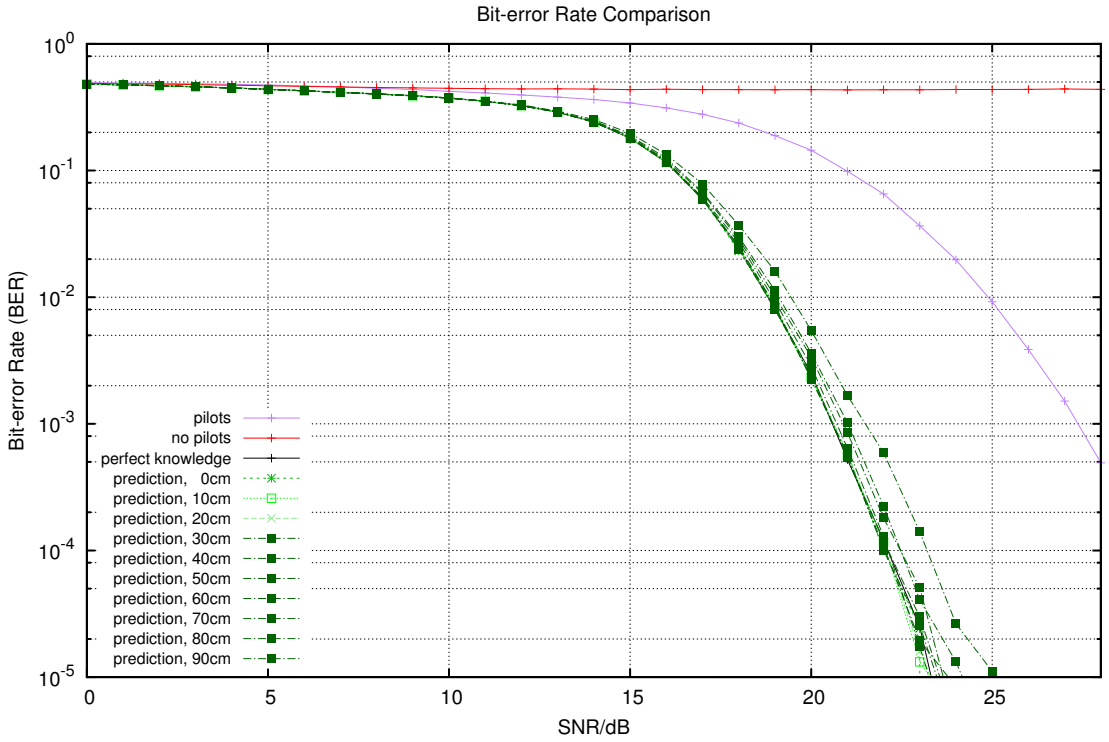


Figure 5.22: A plot showing BER vs SNR, using $n = 3$ polynomial prediction and 64-QAM modulation. It clearly shows all the same trends as Fig. 5.21, but has a higher bit error rate for the same SNRs than the previous 16-QAM case. This is precisely what would be expected from using a higher order modulation technique, and clearly demonstrates that the channel prediction algorithm is modulation order independent.

5.5 Real-World Considerations

As this method requires knowledge of the component rays, applying it to a real scenario requires the additional step of estimating them from their summation. Algorithms such as Multiple Signal Classification (MUSIC) [103] and Estimation of Signal Parameters via Rotational Invariant Techniques (ESPRIT) [89] exist to estimate parameters of a composite signal's component rays, and have already been used to aid channel prediction [119, 105]. Integrating such methods into this system would enable it to make predictions on real-world channels.

Having a method to determine when the set of rays being observed by the receiver changes significantly would also be necessary. The system would need to either correct its existing set of predictions, or take a new set of measurements to create new predictions to take account of the new set of rays.

Finally, knowledge of the receiver's position is key to this prediction scheme. Having accurate location data is required to achieve a useful prediction. This information could come from a range of sources - many modern smartphones are location-aware, and this information could be applied to the prediction scheme.

Given sufficiently accurate ray information at the two prediction points, this system is able to predict a large number of samples into the future, which in the modelled system corresponds to distances in the range of tens of centimetres. In a typical indoor channel scenario with a moving receiver, this could represent hundreds of milliseconds of prediction into the future, depending on the mobile device's velocity. This makes it ideally suited to adaptive transmission systems. Such accurate predictions would completely eliminate the need for any pilot signals during this period, allowing those resources to be used for data transmission, increasing spectral efficiency, which can be converted into power savings or increased throughput. However, as is illustrated in Fig. 5.17, noise at the prediction points has a large impact on the system's accuracy, and ensuring that accurate measurements are taken is absolutely vital to achieve correct operation. Additionally, this algorithm requires very few computational resources to provide an accurate output, making it suited to real-time applications.

While this method is capable of good future predictions, there are a number of areas which if improved upon could further increase the potential for more accurate, even longer-term predictions.

1. As the entire prediction dataset is based upon having accurate readings at a

limited number of points, naturally the algorithm is sensitive to noise at these points. The results clearly show how noise at the prediction points causes errors to grow as the initially small discrepancy is multiplied over distance. Having a larger number of data points to base the prediction from would allow de-noising algorithms. An algorithm such as ESPRIT [89] to estimate the rays' strengths could provide this benefit, as well as calculating the individual ray components. A trade-off would have to be reached here, as a larger number of measurements would mean sacrificing some of the system's gains in spectral efficiency.

2. This research has dealt with the single-antenna case, however extending it to cover Multiple Input, Multiple Output (MIMO) arrays could lead to greater gains in prediction accuracy, as well as more reliable channel measurements to base future predictions on.

3. The algorithm assumes that the same set of rays will be present for the duration of the prediction period. It is however possible for new rays to appear or existing paths be blocked out by room geometry, such as windows or corners. In this case, the system would have to sample a new set of new prediction points, and start a new set of calculations.

5.6 Summary

This chapter has described an investigation into channel prediction, beginning with ESNs and assessing their capabilities in this application domain. After they were unable to reliably predict the complicated channel characteristics generated by the WINNER modelling suite, simpler models were investigated to determine whether ESNs could be of use. Finding they were only able to make accurate predictions about simple channel types, a different approach was adopted.

A method for channel prediction using an Sum of Sinusoids (SOS) approach has been developed. The basis for a linear, followed by a polynomial model is given, and simulation results are presented. The approach shows promise as an efficient channel predictor, particularly well suited to indoor channels. It is very computationally efficient, making it ideal for inclusion into mobile receivers as well as transmitters, and has the potential to significantly reduce the overhead inherent in systems dependent on pilot signals and explicit power control signalling. Because it can provide comparatively

long-term predictions, it has the additional benefit that it enables the transmitter to use adaptive transmission technologies to make optimal choices when scheduling time, power and frequency resources.

Increased resilience to noise is of prime interest when applying this work to real-world conditions, and is discussed. Also incorporating a method for detecting when errors have increased to a point of rendering the method less effective would be useful, to allow the prediction scheme to take action, either to make minor corrections to its existing predictions, or to start a new set of predictions entirely.

The following final chapter will summarise the thesis, present and discuss its main findings and suggest possible future work.

Chapter 6

Conclusions

Digital wireless systems are undergoing a sustained period of exponential growth in traffic, with technologies like WiFi and Long Term Evolution (LTE) becoming ubiquitous. This fast paced growth is straining the limits of what these radio systems can deliver, and with customer expectations of ever increasing data rates, providers are under immense pressured to deliver. As these high bandwidth data services have such a large economic impact, there are huge profits to be made for anyone who can improve the situation.

While these services are performing adequately today, constant innovation is demanded by the fast paced growth of customer demands for data services. There is great scope for improvement, not only in raw throughput but in improved customer experience, especially in challenging signal conditions.

The research in this thesis on anomaly detection could help to address some of the problems encountered by detecting when a signal's quality is too poor to sustain a useful level of service. The channel characterisation work would also be useful in identifying the nature of the signal propagation environment without the need for additional signalling. This information could be used to make intelligent decisions about power allocation, modulation, coding and scheduling of transmissions from the mobile device.

Perhaps the biggest gains stand to come from the channel prediction work. It has potential to eliminate the wasted energy and bandwidth currently allocated to pilot signals, allowing a more energy efficient system, leading to greater battery life for users and lower power bills for cellular operators. Eliminating active measurements would also reduce related overheads; if the transmitter and receiver are both able to predict the future characteristics of a channel, then power control signalling could also be signif-

icantly reduced, which can free up an even larger amount of system resources than pilot signals. These benefits would combine to give a more environmentally friendly, energy efficient system with lower complexity and higher capacity due to lowered signalling overheads.

6.1 Review of Aims and Summary of Thesis

In Section 1.2, the aims of the thesis were stated:

> The central aims of this study are to explore a variety of techniques designed to improve throughput in wireless systems by gaining a better understanding of the channel characteristics and reducing the need for wasteful active channel measurement. The most promising approaches will be extended and developed, and their potential for improved wireless communication will be assessed.

Each of the three major sections of work (the Kullback-Leibler Divergence (KLD) anomaly investigation, channel characterisation and channel prediction) were chosen in the hopes of improving channel knowledge.

After replicating the KLD results of Afgani *et al.*, the system's output would exhibit a monotonic increase when examining certain anomalies as had been previously noted. This distinction was hypothesised to indicate something about the nature of the anomaly, perhaps reflecting some effect introduced by the channel over which the signal had been transmitted. As no cause was immediately obvious, the required signal properties for such monotonic changes were derived, showing that the sequences were caused by a signal whose information content had dropped suddenly. This proof was verified by simulations demonstrating the predicted effect by reproducing Probability Mass Function (PMF)s in a variety of shapes and examining the KLD algorithm's behaviour. Although this new KLD work ultimately did not provide any additional channel information, an explanation was found for the observed monotonically increasing sequences.

In order to investigate methods for increasing channel knowledge, characterisation was chosen because it can provide some additional knowledge without more signalling. Echo State Network (ESN)s were a good candidate tool due to their history of being able to work well with very non-linear systems. Despite initial failure using the unfiltered magnitude response, after a number of candidate pre-processing methods were

investigated, one was chosen which allowed the ESN to work well and provide reasonably accurate characterisations of channels. By generating large amounts of magnitude response data using the Wireless World Initiative New Radio (WINNER) package and training an ESN, an extensive set of Monte Carlo experiments was carried out to quantify the method's effectiveness. In runs of 10,000 iterations the method was able to correctly identify which scenario a section of channel data was generated from in 68% of cases. The incorporation of environmental data (e.g. indoor/outdoor and urban/rural) was investigated, however only marginal gains were seen when adding this or some other forms of pre-processed channel data.

Since successful channel prediction offers such great rewards and ESNs had shown promise in characterisation, the research naturally progressed to unite the two. The channel models generated by the advanced modelling software which had been used previously could not be accurately predicted using ESNs. By closely examining the system, it was found that the ESN was not able to sufficiently learn the complexity of the channel models (as it simply replicated the input at its output without making a prediction), so less complex models were needed in order to assess whether ESNs were viable channel predictors. After developing a parametrisable ray-tracer to model indoor environments and using this to train an ESN, it was found that only channels modelled with a handful of rays could be learned by the Neural Network (NN), while more realistic channels were beyond the ESN's predictive scope.

The time spent developing the ray-tracer for use in previous experiments led channel prediction to be considered in terms of a ray-tracing problem, allowing the final prediction method to be developed. After developing the mathematical basis for linear ray extrapolation, it was natural to extend this framework to higher order polynomial interpolation. A simulator of an Orthogonal Frequency Division Multiplexing (OFDM) transmission system was written in Matrix Laboratory (MATLAB) to compare the performance of the prediction algorithm with a realistic pilot-assisted system, as well as a system which had somehow obtained flawless knowledge of the channel. Results showed that the new methods could enable improved throughput when compared to the pilot-assisted method, without the need for repeated active measurement. The difference in performance between linear and polynomial methods was investigated, and another large Monte Carlo experiment also examined the noise resilience of the system.

6.2 Limitations and Scope for Further Research

As mentioned in Chapter 4, the ESN classification scheme, (although noteworthy for learning classifications entirely from training data) suffered from an inability to infer the underlying essential differences between alternative channel scenarios, rather becoming too reliant on one particular instance of each scenario. Further work on enabling the ESN to learn more general channel characteristics would open this approach up to much broader appeal, especially on real-world channel datasets. Also comparing the ESN approach with other NN architectures would be worthwhile to find out the relative advantages of using different computational models.

The characterisation work could benefit from some of the recently proposed modifications to the basic ESN architecture (see Section 2.2.3) in order to improve results while minimising computational load. As noted in Chapter 4, running time became a practical constraining factor in some of the experiments, and any improvements to this situation would be welcome, especially in low-power mobile devices.

Adding higher-level channel information (such as estimates of the coherence time or Doppler spectrum) in addition to magnitude response data might help the ESN distinguish between different channel scenarios, however this increased dimensionality will come with higher computational demands. An investigation into this trade-off has potential for further improving the accuracy of the characterisation results.

Some of the more interesting results from the ESN work carried out in this thesis are the instances when one channel type is misclassified as another quite similar scenario. This may suggest that the widely recognised channel designations chosen in this work could usefully be reconsidered, as some seem quite similar (see Section 4.3.1). It follows that the same transmission parameters might be optimal for several of the channel scenarios used in this research. It could prove more useful to classify channels according to a different set of scenarios, or to reduce the number of classifications to those which are sufficiently distinguishable that significantly altered communication parameters need to be established.

The work on channel prediction in Chapter 5 offers a practical starting point for the algorithms to be tested in real-world settings, however there are the challenges of ray estimation and noise immunity to explore. Using multi-antenna algorithms such as Multiple Signal Classification (MUSIC) and Estimation of Signal Parameters via Rotational Invariant Techniques (ESPRIT) to perform ray estimation in a real-world

environment would allow the ray-based modelling method proposed to be verified or expanded upon. Since the method relies on modelling a channel as being composed of a number of rays, an investigation would be worthwhile into how many rays are required for good prediction of real-world channels, as well as balancing this with the increasing computational workload. This work assumes a 7-ray model (Line of Sight (LOS) and 6 reflected rays from walls and floors), however this is only going to be the case when inside a building or similar environment. The ability to generalise the algorithm and then allow it to cover a much wider range of channel scenarios is absolutely vital to its success.

A related area worth exploring is how the prediction scheme could benefit multi-antenna radio systems. Since many modern digital wireless systems use Multiple Input, Multiple Output (MIMO) technologies, it is more important then ever to have accurate channel information (as coherent decoding is extremely difficult without it in the MIMO case). Given that there is often correlation between channels of closely spaced antennas, milti-antenna systems may achieve even higher efficiency gains than the single-antenna case outlined in this work.

In a similar fashion to extending into the spatial dimension with MIMO, an extension into the frequency domain with broadband channels would further extend the applicability of this method. The most straightforward way to apply this method is to treat a broadband channel as a number of narrowband subchannels, and use the existing prediction scheme for each. However, given the correlation among such closely spaced channels, results from a subset of the subchannels can be used to interpolate values for the other subchannels (as is commonly done when using pilot signals in broadband OFDM systems such as LTE and WiFi).

Appendix A

Publications

[1] Anderson, A., and Haas, H. Kullback-Leibler Divergence (KLD) Based Anomaly Detection and Monotonic Sequence Analysis In *Vehicular Technology Conference (VTC Fall), 2011 IEEE* (2011), pp. 1–5.

[2] Anderson, A., and Haas, H. Using Echo State Networks to Characterise Wireless Channels. In *Vehicular Technology Conference (VTC Spring), 2013 IEEE* (2013), pp. 1–5.

[3] Anderson, A., and Haas, H. Indoor Channel Prediction Using an Efficient Sum of Sinusoids Linear Prediction Scheme. In *Vehicular Technology Conference (VTC Fall), 2013 IEEE* (2013), pp. 1–5.

Kullback-Leibler Divergence (KLD) Based Anomaly Detection and Monotonic Sequence Analysis

Alan Anderson and Harald Haas
Institute for Digital Communications
The University of Edinburgh
King's Buildings, Mayfield Road, Edinburgh, EH9 3JL, UK
Email: {a.anderson, h.haas} @ed.ac.uk

***Abstract*—Cognitive Radio systems require detailed feedback about their environment, and detecting anomalies is core to this task. The KLD metric can be used to detect a variety of anomalies in radio signals, and has been previously demonstrated to be both effective, and efficient enough to run in real-time. In tests, it was observed that some anomalous signals caused the KLD to increase monotonically for long time periods, while others did not. After analysing the KLD equation and comparing the findings with the results from the tests, we present a hypothesis for how such monotonic sequences could occur and demonstrate that this agrees very closely with results in observed signals.**

***Index Terms*—anomaly detection, cognitive radio, Kullback-Leibler Divergence**

I. Introduction

Cognitive Radios (CR) are an emerging class of radio devices which are rapidly reconfigurable using software. Since they are software defined, they can be much more flexible than traditional radios, allowing the use of measurements to maximise spectral efficiency. Such systems adjust themselves to the current wireless environment, taking account of radio link conditions and other factors to make a more efficient system.

CR systems must be capable of rapidly responding to changing signal parameters. They rely on being able to acquire accurate information about the current wireless environment, allowing them to adapt, however this information can be costly to gather and process, and this is an area of much interest to CR designers.

Abnormal signals have a wide range of causes, from interference from other wireless devices, to malfunctioning transmitters. Being able to identify such signals as abnormal is clearly vital to the operation of CR systems. Once identified as abnormal, the CR can decide whether to change its parameters in response to the anomalous signal.

This problem of detecting anomalies in a system is not unique to the field of CR: this research began as a test and measurement triggering problem. It is also widely applied in network intrusion detection [1] and credit card fraud detection systems [2], as well as in the field of medical imaging and sensor networks. There are three broad classes of anomaly detection systems: those which are taught using both known normal and anomalous data (supervised), those which are 'taught' using only a known normal data set (semi-supervised), and those for which training data is not labelled as normal or anomalous (unsupervised). Unsupervised anomaly detection systems work by assuming that since anomalies are generally considered rare, a rare event is likely to be anomalous.

In the field of radio communications, there are a number of approaches that can be taken to detecting anomalous behaviour [3]. A simple brute-force approach with an exhaustive set of known good waveforms, although theoretically ideal, is hugely impractical: creating the set of good signals would take significant effort, and searching the space is far beyond any current technology.

The interference temperature metric has been suggested [4] as a method for quantifying how a radio system interacts with interference, however it has a number of drawbacks: it must have multiple antennae to account for spatial variations, and its computational complexity is very high, preventing it from being used in real-time systems (which a CR must be by definition). Additionally it is restricted to detecting anomalies associated with interference events.

The rest of this paper is organised as follows: Section II introduces our methodology for detecting anomalies, while Section III outlines our observations of monotonic sequences in test datasets. Sections IV, V and VI take the KLD equation and simplify it, while Section VII shows an example of how monotonic behaviour could be achieved which is verified by simulation in Section VIII. Section IX concludes the paper.

II. Kullback-Leibler Divergence

The KLD is calculated [5] by:

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{1}$$

Any logarithm base can be used, however using a base-2 logarithm results in a quantity measured in bits.

The KLD can be used as a function for comparing the statistical similarity of two datasets. Two similar sets will have a small KLD, while very different sets would have a larger KLD. It is this comparison operation which makes it well suited to anomaly detection in wireless signals.

Fig. 1 shows the general set-up used in a KLD anomaly detection system. PMFs with a discrete number of 'bins', (each of width Δ) are computed over two equal-sized time windows, separated by some fixed amount. In the case of periodic signals, by separating the two windows by the signal's

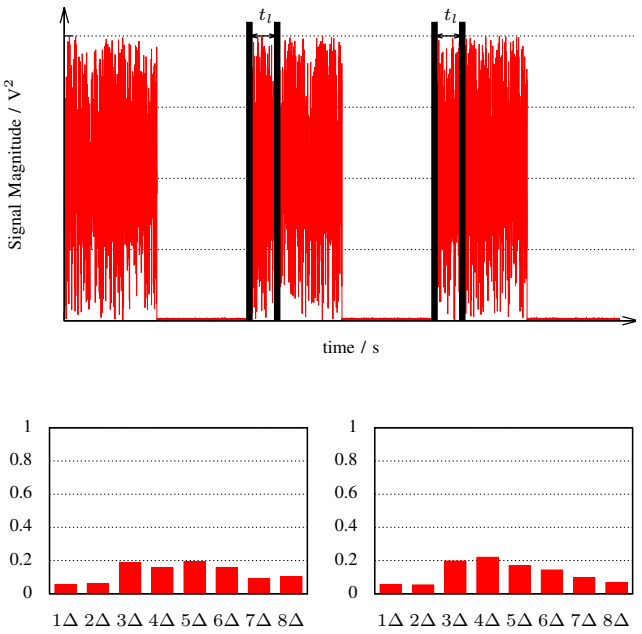


Fig. 1: By having two windows of width t_l, two histograms, $q(x)$ and $q(x)$ can be created, and then compared. If they are separated by the signal period t_p (as in this example) corresponding parts of the repetitive signal are being compared to each other.

period (as in the example), we can compare corresponding parts of the repeating signal structure. Each time a new signal sample is taken, the PMFs are re-calculated, leading to a new KLD. By making the assumption that each frame in a periodic wireless signal is largely similar to the next, we can define normal conditions as being those when the divergence between adjacent frames is small. When this value becomes large, we can assume that this is anomalous, and raise an alert to the wider system. In previous work ([6]) a simple threshold was used so an anomaly would be flagged if the value of the KLD ever exceeded a certain pre-determined value, and this proved very effective in identifying a variety of anomalies [7], [8].

It should be noted that the KLD is almost never exactly zero, but fluctuates very slightly since the two PMFs are rarely absolutely identical due to noise. However, we observed a number of cases where the KLD would increase monotonically over a long time period, in stark contrast to its usual behaviour. Every time such a sequence occurred in tests, it coincided with a known anomaly, though not all anomalies coincided with such a feature. Fig. 2 shows both the normal fluctuating behaviour and several monotonic sequences.

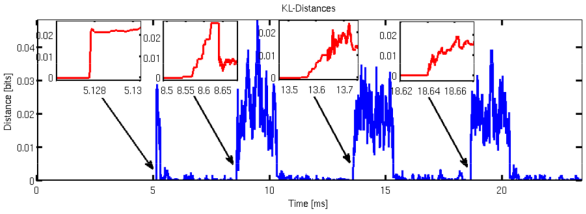


Fig. 2: A plot of KLD showing four anomalies, only some of which show strict monotonic increase.

This sudden change from a fluctuating KLD to this monotonically increasing behaviour stands out as a point of interest, especially since it coincides with known anomalies occurring. This paper is concerned with this behaviour, and it investigates if the behaviour can be related to a particular class of events or anomalies in the input signal. Since this monotonic behaviour does not always occur when an anomaly is detected, but only in some cases, it could potentially be related to a particular type of anomaly.

III. Observed Monotonic Behaviour

During examination of a mobile WiMAX signal which included a known anomaly, we observed a number of instances where the KLD increased monotonically for a large number of samples. The anomaly can be seen on the time series (Fig. 3a), as the fourth frame is abnormally long, and was caused by a malfunctioning transmitter.

Each time these long sequences were seen, they coincided with an anomaly in the wireless signal. Fig. 3 shows the RF data, and how the KLD increases over time. It should be noted that in a KLD system, an anomaly will always create two peaks in the KLD: one as the anomalous frame is compared with the preceding correct frame, and a second as the next correct frame is compared with the anomalous frame.

By observing how the two PMFs change over time, it became clear that one PMF's mass would shift, concentrating all the data points into one bin, while the other PMF would retain its previous evenly distributed characteristic. Since these two PMFs are radically different, we can clearly expect a large KLD. However, it was not immediately clear why the KLD would exhibit this monotonically increasing characteristic over time.

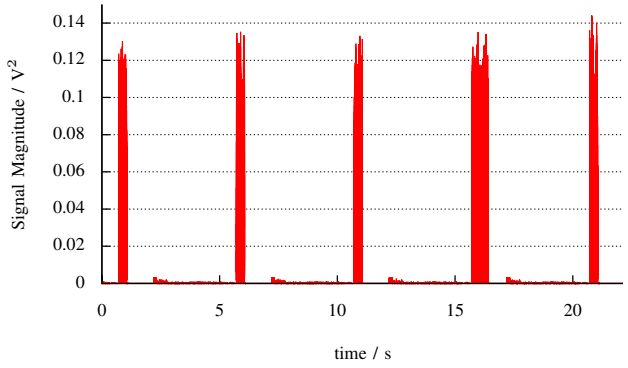
In order to investigate this behaviour, we begin by examining the KLD calculation itself.

IV. The General Case

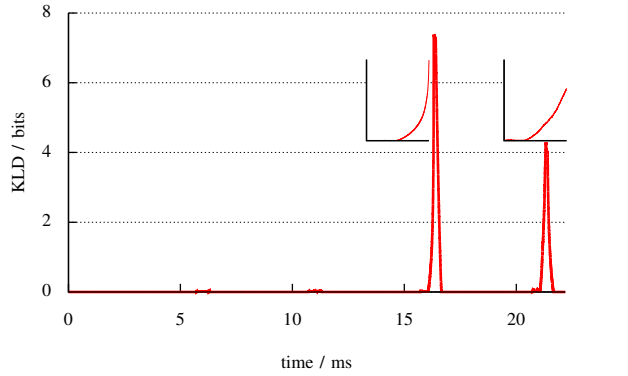
At each time step, the time windows move forward one data point, meaning both windows admit one new reading to their front edge and remove the oldest reading from their back edge. The updated histograms are calculated by adding the new readings to the histograms, and removing the old ones. This results in 4 different bins being modified:

- a is the bin in p which is increased
- b is the bin in p which is decreased
- c is the bin in q which is increased
- d is the bin in q which is decreased

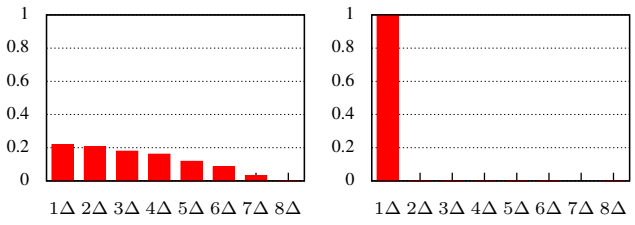
Since only these 4 bins will change, by simply expanding (1), we can work out the difference between two successive KLD calculations. $p_{\mathrm{n}-1}(x)$ and $q_{\mathrm{n}-1}(x)$ refer to the old PMFs in the preceding time step, while $p_{\mathrm{n}}(x)$ and $q_{\mathrm{n}}(x)$ refer to the new PMFs in the current time step.



(a) Time series data of observed signal

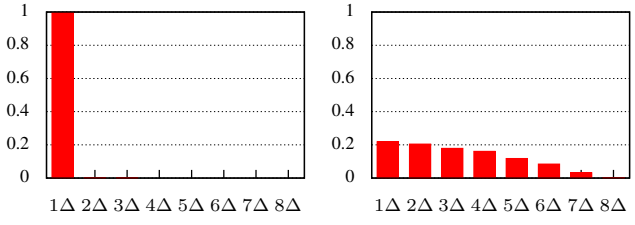


(b) Plot of KLD with inset graphs highlighting its monotonic growth.



(c) PMF p for first spike

(d) PMF q for first spike



(e) PMF p for second spike

(f) PMF q for second spike

bin width $\Delta = 0.0173796$ for each PMF

Fig. 3: This figure shows a mobile WiMAX RF trace and its accompanying KLD plot. There are two long monotonic KLD sequences in this time period, and the PMFs are shown for each at the instant of maximum KLD. In both cases, the p and q PMFs are radically different, with one being almost uniformly distributed and the other being very concentrated.

$$
\begin{aligned}
&\mathrm{D}_{\mathrm{n}-1} - \mathrm{D}_{\mathrm{n}} = \\
&p_{\mathrm{n}-1}(a)\log\frac{p_{\mathrm{n}-1}(a)}{q_{\mathrm{n}-1}(a)} - p_{\mathrm{n}}(a)\log\frac{p_{\mathrm{n}}(a)}{q_{\mathrm{n}}(a)} + \\
&p_{\mathrm{n}-1}(b)\log\frac{p_{\mathrm{n}-1}(b)}{q_{\mathrm{n}-1}(b)} - p_{\mathrm{n}}(b)\log\frac{p_{\mathrm{n}}(b)}{q_{\mathrm{n}}(b)} + \\
&p_{\mathrm{n}-1}(c)\log\frac{p_{\mathrm{n}-1}(c)}{q_{\mathrm{n}-1}(c)} - p_{\mathrm{n}}(c)\log\frac{p_{\mathrm{n}}(c)}{q_{\mathrm{n}}(c)} + \\
&p_{\mathrm{n}-1}(d)\log\frac{p_{\mathrm{n}-1}(d)}{q_{\mathrm{n}-1}(d)} - p_{\mathrm{n}}(d)\log\frac{p_{\mathrm{n}}(d)}{q_{\mathrm{n}}(d)}
\end{aligned}
\tag{2}
$$

Clearly this difference must be positive over a number of successive readings to produce a monotonic sequence. However, each of the terms in (2) can be either positive or negative, depending on the state of the individual bins in p and q. Thus there is no simple criterion for ensuring that this difference is kept positive without knowledge of the states of the PMF bins.

V. Simplified General Case

If we make the assumption that a, b, c and d are all different from each other, then we can replace $p_{\mathrm{n}-1}, q_{\mathrm{n}-1}$ and $p_{\mathrm{n}}, q_{\mathrm{n}}$ using the following substitutions, simply using the definitions of a, b, c and d from Section IV. If the four affected bins are not unique, then this simplification cannot be made. In experiments it is rare for this property of uniqueness to occur over a long time series, however making this assumption greatly simplifies the analysis, and exposes a more elegant expression for KLD calculation which yields extremely similar results to the unsimplified version (2).

$$
\begin{aligned}
p_{\mathrm{n}}(a) &= p_{\mathrm{n}-1}(a) + 1 & q_{\mathrm{n}}(a) &= q_{\mathrm{n}-1}(a) \\
p_{\mathrm{n}}(b) &= p_{\mathrm{n}-1}(b) - 1 & q_{\mathrm{n}}(b) &= q_{\mathrm{n}-1}(b) \\
p_{\mathrm{n}}(c) &= p_{\mathrm{n}-1}(c) & q_{\mathrm{n}}(c) &= q_{\mathrm{n}-1}(c) + 1 \\
p_{\mathrm{n}}(d) &= p_{\mathrm{n}-1}(d) & q_{\mathrm{n}}(d) &= q_{\mathrm{n}-1}(d) - 1
\end{aligned}
$$

By substituting the above terms into (2), we can have a much simplified expression for the change in KLD.

$$
\begin{aligned}
&\mathrm{D}_{\mathrm{n}} - \mathrm{D}_{\mathrm{n}-1} = \\
&p_{\mathrm{n}-1}(a)\log\frac{(p_{\mathrm{n}-1}(a)+1)}{p_{\mathrm{n}-1}(a)} + \log\frac{(p_{\mathrm{n}-1}(a)+1)}{q_{\mathrm{n}-1}(a)} + \\
&p_{\mathrm{n}-1}(b)\log\frac{(p_{\mathrm{n}-1}(b)-1)}{p_{\mathrm{n}-1}(b)} - \log\frac{(p_{\mathrm{n}-1}(b)-1)}{q_{\mathrm{n}-1}(b)} + \\
&p_{\mathrm{n}-1}(c)\left(\log\frac{q_{\mathrm{n}-1}(c)}{(q_{\mathrm{n}-1}(c)+1)}\right) + \\
&p_{\mathrm{n}-1}(d)\left(\log\frac{q_{\mathrm{n}-1}(d)}{(q_{\mathrm{n}-1}(d)-1)}\right)
\end{aligned}
\tag{3}
$$

From this form of the expression it is clear that there are two terms which are always positive, and two which are always negative. This means the state of PMF p (and particularly the state of the four modified bins within it) will have the dominant impact on the change in KLD. From this, we can see that

the fuller a modified bin is, the larger its contribution to the change.

VI. Further Simplification

The general proof in Section IV offers little insight into how the monotonic behaviour occurs, while Section V simplifies the equation.

Below is a further simplification of (3). The further assumption made in this instance (as well as c and d being unique) is that the PMF p remains unchanged from one time step to the next. This assumption has been found to be held true for long time periods in a number of test cases. It occurs most often in the inter-frame gap in a periodic signal, or when there is no active transmission happening on the channel.

$$\begin{aligned}\mathrm{D_n} - \mathrm{D_{n-1}} =& p_{\mathrm{n-1}}(c)\left(\log\frac{q_{\mathrm{n-1}}(c)}{(q_{\mathrm{n-1}}(c)+1)}\right)+\\ & p_{\mathrm{n-1}}(d)\left(\log\frac{q_{\mathrm{n-1}}(d)}{(q_{\mathrm{n-1}}(d)-1)}\right)\end{aligned} \tag{4}$$

In order for monotonic behaviour to occur, the following inequality must hold over a number of successive calculations:

$$\begin{aligned}&\left|p_{\mathrm{n-1}}(c)\left(\log\frac{q_{\mathrm{n-1}}(c)}{(q_{\mathrm{n-1}}(c)+1)}\right)\right| \leq\\ &\left|p_{\mathrm{n-1}}(d)\left(\log\frac{q_{\mathrm{n-1}}(d)}{(q_{\mathrm{n-1}}(d)-1)}\right)\right|\end{aligned}$$

In the above equation, since the differences between the numerator and denominator in the log terms will usually be small, the value of the log terms will be close to 0. The value of the log terms will grow larger as the bin it refers to becomes emptier (smaller). Clearly the dominant terms in the inequality are the scaling factors of $p(c)$ and $p(d)$.

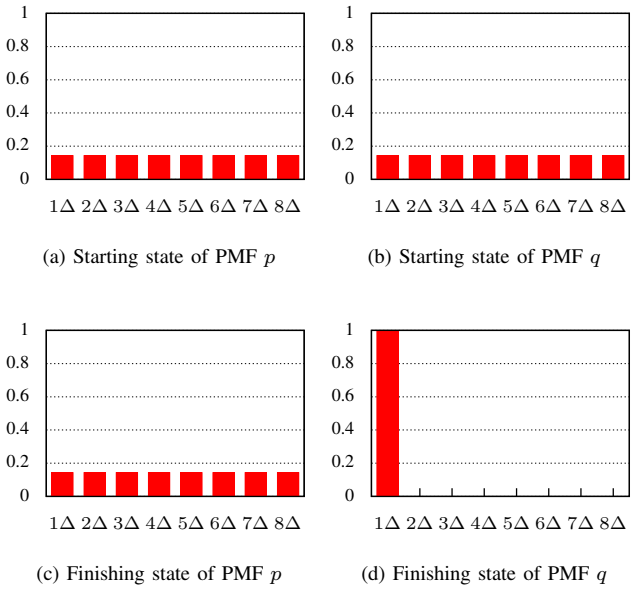
VII. Worked Example using Theory

By examining the PMFs p and q at different points in time during the monotonic sequences, it was observed that a single bin would become very full in one PMF, while the same bin in the other PMF remained at a lower level. The following example shows how this observed behaviour could provide the monotonic behaviour observed:

The system's two PMFs, $p(x)$ and $q(x)$, each have n bins, numbered 0 to $n-1$. For this artificial example, both $p(x)$ and $q(x)$ begin with each bin filled equally (i.e. a uniform distribution). PMF $q(x)$ will tend towards a full bin number 0, and empty bins 1 to $n-1$, while $p(x)$ will not change with time. This is shown graphically in Fig. 4. Bins 1 to $n-1$ in $q(x)$ are depleted evenly, such that every $n-1$ time steps bin 0 will have grown by $n-1$ samples, and bins 1 to $n-1$ have each one sample less.

Since we know PMF p is time-invariant and each bin in p is always equal to $\frac{1}{n}$, we can simplify the KLD calculation to:

$$\begin{aligned}\mathrm{D} =& p(x)\sum_{x=0}^{n}\log p(x) - \log q(x)\\ =& \frac{1}{n}\sum_{x=0}^{n}\log\left(\frac{1}{n}\right) - \log q(x)\end{aligned} \tag{5}$$



(a) Starting state of PMF p (b) Starting state of PMF q

(c) Finishing state of PMF p (d) Finishing state of PMF q

Fig. 4: It is clear that PMF q although initially identical to p becomes radically different by the end time

By introducing a time step variable, t, we can calculate the value of each bin in the changing PMF q using:

$$q(x) = \begin{cases}\frac{1}{n}+(n-1)t & :0\\ \frac{1}{n}-t & :1,\ldots,n-1\end{cases} \tag{6}$$

For this system, t ranges from 0 to $\frac{1}{n}$.

When this PMF equation is expanded and substituted into (5), the value of the KLD at any time between 0 and $\frac{1}{n}$ can be calculated:

$$\begin{aligned}\mathrm{D} = \frac{1}{n}\Big(& \underbrace{n\log\frac{1}{n}}_{p(0),\ldots,p(n-1)} - \underbrace{\log\left(\frac{1}{n}+(n-1)t\right)}_{q(0)} -\\ & \underbrace{(n-1)\log\left(\frac{1}{n}-t\right)}_{q(1),\ldots,q(n-1)}\Big)\end{aligned} \tag{7}$$

(7) does indeed show monotonic behaviour over the valid time interval $0 \leq t < \frac{1}{n}$, and is shown in Fig. 5.

The following section will verify the results from this example by performing simulations of a similar system and comparing them to the outcome predicted here.

VIII. Example using Simulation

Although the example in Section VII demonstrates monotonicity, the idea that a real signal would deplete the PMF bins of q perfectly evenly is a convenient, though unrealistic assumption. In this further example we will simulate a system similar to that in the previous example, but remove the requirement that the bins be depleted evenly, making it a much more realistic example.

In this simulation, the two PMFs p and q begin identically, with each of the 8 bins filled equally. p remains unchanged,

while q tends towards a full first bin, depleting all other bins randomly. The start and end states of both PMFs will be the same as shown in Fig. 4.

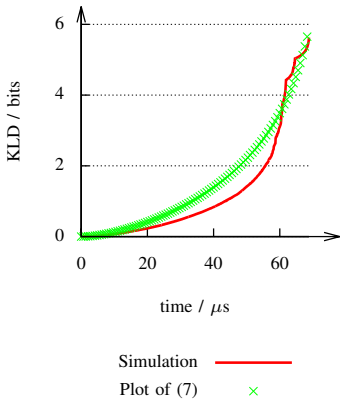


Fig. 5: A plot of the KLD over time from the simulation, assuming bins in q are randomly depleted, as would be likely to be found in a real signal capture. Also shown is the predicted data from (7), showing that the results from the simulation agree with those obtained from the theoretical approach take in Section VII.

This first simulation agrees very closely with that predicted by the theory in (7).

In order to make the simulation resemble real world conditions as closely as possible, we relaxed the assumption that the PMF p remained entirely unchanged for our final simulation. We applied a low-level white noise signal to p, causing its bins to change very slightly at each time step. This is exactly what can be seen when the distribution of power is even in the signal, (which is exactly the model we use for p, as can be seen in its histogram in Fig. 4). Thus the PMF will not change its overall shape by a significant amount, but will change slightly at each time step, maintaining its overall shape with time.

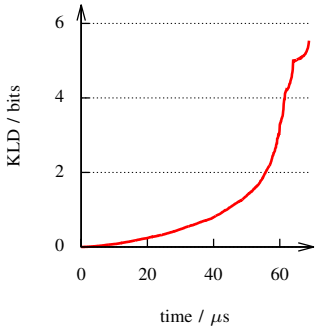


Fig. 6: A simulation of a similar system with a randomly varying p, as well as a changing q. This result is very similar to both the previous example, and the outcome predicted in VII

This final simulation shows that when the assumptions made in the example in Section VII are replaced with real-world conditions, the behaviour of the system remains largely unchanged, and showing a realistic set of circumstances under which monotonic increase can be achieved.

IX. Conclusions

In this paper we aimed to find if there was a link between the observed monotonic sequences and the class of anomaly they accompany. The results suggest that long monotonic sequences cannot be specifically related back to any particular class of anomalies or signal events, as it seems to be caused by a simple shifting of mass of one PMF with respect to the other, as the example in Section VII illustrates. Such a PMF has a very wide range of causes, so we find no direct link to a particular anomalous behaviour. However it can be said that the radio signal which generates such a PMF has become more deterministic, which may offer some insight into the problem. This result agrees with the observations in Section VI, since it identifies no simple way of guaranteeing monotonic behaviour.

As is highlighted in (4) and (3), the change in KLD from one time step to the next is heavily dependent on the absolute value of only 4 PMF bins, and there will always be two pairs of terms of opposing sign. Since there is no way to eliminate the negative terms in the calculation, there is no trivial way of ensuring an always positive result, producing monotonic behaviour. Without *a-priori* knowledge of the PMFs, the sign of the change in KLD cannot be determined.

We were however able to produce a model for monotonic sequences which is very close to our observations from a variety of real-world signals.

Acknowledgements

The authors would like to thank Agilent Technologies UK and USA for sponsoring this research and for providing the large volume of data sets for testing.

References

[1] A. Patcha and J.-M. Park, "An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, Aug. 2007.

[2] Y. Kou, C. Lu, S. Sirwongwattana, and Y. Huang, "Survey of fraud detection techniques," in *Networking, Sensing and Control, 2004 IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 749–754.

[3] M. Filippone and G. Sanguinetti, "Information theoretic novelty detection," *Pattern Recognition*, vol. 43, no. 3, pp. 805–814, 2010.

[4] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[5] T. Cover, J. Thomas, and J. Wiley, *Elements of information theory*, 2nd ed. Wiley-Interscience, 2006, vol. 1.

[6] M. Afgani, S. Sinanović, and H. Haas, "Anomaly Detection Using the Kullback-Leibler Divergence Metric," in *Proc. of the First International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*. Aalborg, Denmark: IEEE, Oct. 25–28 2008, p. 5.

[7] ——, "Hardware Implementation of a Kullback-Leibler Divergence Based Signal Anomaly Detector," in *Proc. of the Second International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL) (Invited Paper)*. Bratislava, Slovak Republic: IEEE, Nov. 24 – 27 2009, p. 6.

[8] M. Afgani, S. Sinanović, and H. Haas, "Information Theoretic Approach to Signal Feature Detection for Cognitive Radio," in *Proc. of the 2008 IEEE Global Telecommunications Conference (GLOBECOM)*. New Orleans, Louisiana, USA: IEEE, Nov. 30–Dec. 4 2008, p. 5.

Using Echo State Networks to Characterise Wireless Channels

Alan Anderson and Harald Haas
Institute for Digital Communications
The University of Edinburgh
King's Buildings, Mayfield Road, Edinburgh, EH9 3JL, UK
Email: {a.anderson, h.haas} @ed.ac.uk

***Abstract*—We propose the use of echo state networks for the task of wireless channel characterisation to select the most similar channel to the current observed channel from a pre-defined set, based entirely on received signal information. This allows the system to select the optimal resource allocation scheme and transmission parameters from pre-computed solutions. Using suitable training data, the neural network was able to learn to characterise a signal correctly 68% of the time, which can be further improved to 72% by adding some simple location data to the signals being examined. Our system out-performs a comparable statistical method by a factor of two, demonstrating echo state networks' ability to infer information from their training data which other systems can not.**

***Index Terms*—Echo-state networks, reservoir computing, channel characterisation, Kullback Leibler divergence**

I. Introduction

The ability to detect channel characteristics and use this information to select the best transmission scheme is core to maximising throughput in modern digital wireless systems. The problem is however a very complicated one, as high-bandwidth channels (such as those used in today and tomorrow's cellular systems) are subject to frequency selective fading, and vary extremely rapidly, especially when devices are mobile. To tackle this problem, a number of schemes have been devised to measure and characterise them, so that the most effective transmission scheme can be used to best exploit the channel's characteristics. Often these involve active measurement using pilot signals, which although provide good channel information, can be very resource intensive. The channel estimation overhead in LTE deployments can use 30% [1] of time and frequency resources which could otherwise be used to increase throughput in an increasingly pressurised network. Another range of classifiers examine the channel's history and try to determine the type of wireless environment the device is in. This can then be used to apply a pre-computed best-fit solution for the transmission parameters and resource allocation for those conditions [2]. This second approach assumes less about the channel than the first, and is the subject of our investigation. The ultimate goal of this research is to predict a channel's future behaviour by examining its past with as little additional information as possible. Doing so would allow for vastly improved resource allocation, a reduction in channel estimation overhead, which is a limiting factor in more sophisticated transmission techniques such as MIMO and CoMP (cooperative multipoint) and maximised spectral and energy efficiency.

Echo State Networks

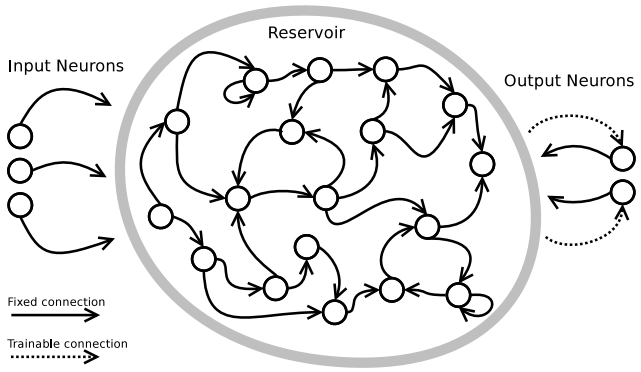


Fig. 1: An example echo state network, illustrating that the only trainable connections are those from the reservoir to the output neurons. Also visible is the recurrent nature of the network, and the random, fixed connections between internal neurons. The number of input and output neurons are simply determined by the dimensionality of the input and output.

First proposed in 2001 [3] echo state networks (ESNs) are a class of recurrent neural networks which have been shown to have applications in modelling highly non-linear systems. They have the unique property that training the system only modifies the weights assigned to the output neurons, rather than additionally modifying all internal connecting weights (as happen in classical neural networks), making training a computationally simple task. The internal connecting weights in an ESN are initialised randomly, and remain fixed for the life of the network. ESNs are also sparsely connected internally (often having only around 1% interconnectivity), which allows the internal neurons (referred to as reservoir neurons) to produce a number of loosely interconnected subsystems which cooperate to give the desired output. Since training an ESN is a relatively computationally inexpensive task, it is quite feasible to have a network with several thousands of neurons, allowing complex systems to be modelled efficiently.

Echo-state networks have been used to model and classify a wide range of complex systems - they have found applications from communication channel equalisation, [4] to future wind speed forecasting [5] and predicting hydroelectric dam water inflow[6]. Their ability to characterise a system based purely on measured training data (and not on a mathematical

model) is extremely attractive in many fields, as mathematical models may be complicated, inaccurate, or non-existent, while historical records of past performance are often much more easily available. ESNs allow the user to treat the complex system as a 'black box', whose input and output relationships are unknown at the outset, and for the neural network to infer what function the system performs using only the training data. This will result in a model based on the real system, rather than on imperfect assumptions. This does, however require careful selection of training data to ensure it is representative of normal behaviour.

For the task of system characterisation, ESNs can be trained to give an output which indicates how similar the measured system is to each of the systems it was trained on. One which perfectly matched one of the training systems would indicate it was an ideal match, but the ESN would also rate how closely the system matched each of the others it was trained on. This can be a useful feature when the system has characteristics similar to those seen in several different training systems.

Wireless Channel Modelling

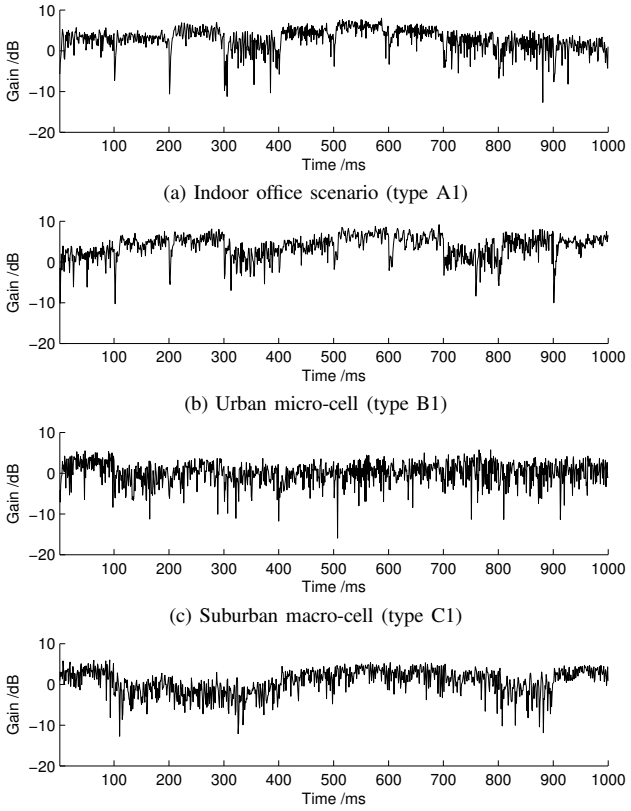
The WINNER project [7] created a parameterisable channel modelling software package, allowing for a wide variety of wireless scenarios, but particularly targeted those likely to be encountered in third and fourth generation cellular networks. It is for this reason we selected this suite to model the channels to be used, as it is seen by some [8] to be the most comprehensive channel modelling suite for this purpose. As well as being able to generate arbitrary channels (given appropriate parameters) the WINNER model has a set of thirteen channel types commonly seen in cellular systems, encompassing a range of indoor, outdoor, urban and rural channel scenarios, as outlined in Table I.

Code	Channel Description
A1	Indoor office
A2	Indoor to outdoor
B1	Urban micro-cell
B2	Bad Urban micro-cell
B3	Indoor hotspot
B4	Outdoor to indoor
B5	Stationary Feeder
C1	Suburban macro-cell
C2	Urban macro-cell
C3	Bad urban macro-cell
C4	Urban macro outdoor to indoor
D1	Rural macro-cell
D2	Moving networks

TABLE I: A description of each of the thirteen channel scenarios used in the characterisation problem. These were chosen to cover the most common cellular scenarios, and to give a diverse set of channel types for system training and testing. These channel types are explained in full detail in [9].

II. Experimental Setup

We begin by generating a set of thirteen standard wireless channel scenarios (as described in Table I), and transmit a randomised signal over each. To generate each channel (used for both training and testing) we begin by creating the geographical model of where each transmitter and receiver is located for each of the thirteen channel types, along with



(a) Indoor office scenario (type A1)

(b) Urban micro-cell (type B1)

(c) Suburban macro-cell (type C1)

(d) Moving Network, as found on a high-speed train (type D2)

Fig. 2: Time series for a selection of different channel types, taken from Table I, each showing different characteristics. These, along with other similar channels are used as training data for the echo state network.

positions of interferers and scatterers (such as buildings). We choose semi-random layouts as provided by the WINNER project in order to generate a real-world representative model. For each channel we extract the channel's magnitude response (seen in the top graph of Fig. 3).

In our early experiments we used this magnitude response data directly as input to the ESN and found that although it was successful in characterisation in some cases, the noisy nature of the signal caused unpredictable behaviour within the ESN. The ESN would sometimes classify two datasets from the same channel as different types, and was not consistent in its output. We decided to perform some pre-computation on the raw time series before feeding it to the ESN, in order to de-noise the signal, and provide more consistent results. After experimenting with a number of methods (including taking moments of the signal, measuring power spread and derivatives of the signal) we chose to measure the spread of zero crossings as it significantly improved the classification results over the original system.

We take the channel's magnitude response, select the zero crossing points (shown on both graphs in Fig. 3), and apply a time window to the data. We then compute the standard deviation of the distances between adjacent zero crossings within the time window. This standard deviation time series is then supplied to the ESN as training data, along with

the correct channel classifications (fed into the system as a teacher signal). This teacher signal is composed of thirteen 1-dimensional time series, each corresponding to one of the training channels.

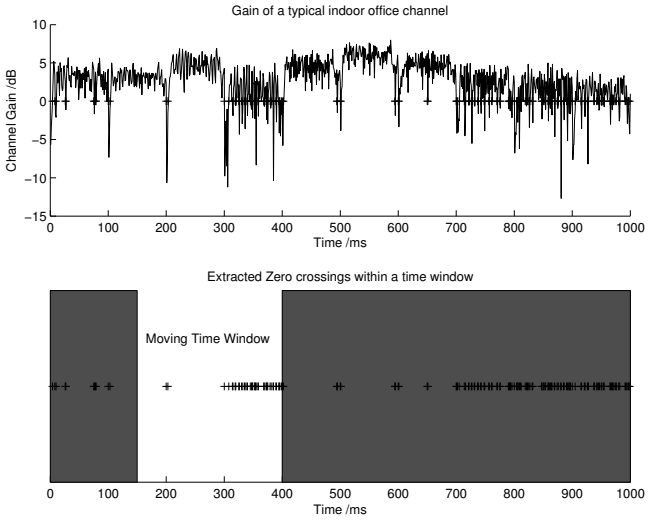


Fig. 3: The top graph shows the magnitude response of a fast-fading wireless channel for a typical indoor office scenario, with the 0dB crossing points marked. The lower graph shows the same crossing points, and the highlighted section shows the time window over which the standard deviation of distances between adjacent points are calculated.

We supply the ESN with the training input and the teacher signal and train the network. Once training has completed, the channel to be characterised goes through the same procedure that the training data do. When the time series of standard deviation of zero crossings has been calculated, this data is fed into the ESN to be characterised. The ESN has a one-dimensional time series output (coming from an output neuron) for each of the thirteen channels it was trained with. The whole input time series is passed into the ESN, where it computes how closely this unknown data matches each of the thirteen channel types it was trained to recognise. The output time series with the closest mean to that of one of the teacher signals denotes the closest match to that teacher signal.

We use a reservoir of 1000 neurons to perform the characterisation work. The probability of correct classification of an unknown signal increases approximately linearly with the number of neurons used (see Fig. 4), up to around 1000 neurons, beyond which there are is no further improvement. In simulation we ignore the first several hundred output values from the ESN as they will not be based on initialised data (often referred to as 'dummy steps'). In each simulation we generate 100 seconds of training data for each channel type. We train the network on each of the thirteen channel types in turn, before supplying the unknown signal.

Adding Simple Location Data

If we mark each channel as 'urban' or 'rural' and 'indoor' or 'outdoor' when we train the system, and then assume

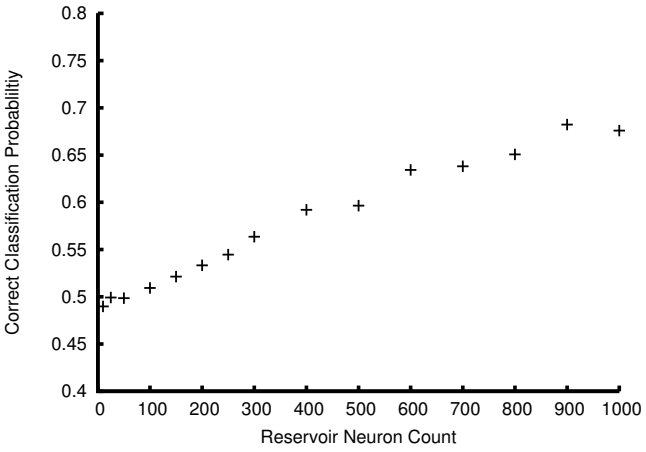


Fig. 4: Graph showing how reservoir neuron count influences system performance. As can be seen from the final two data points, the probability of correct classification peaks at 68% for this setup. Larger neural networks (of over 1000 neurons) use significant additional system resources but do not produce any increase in the correct prediction rate.

the mobile device is able to determine this same data, it narrows the search space of channels considerably, leading to more accurate characterisations. Such information could be gathered using the mobile device's camera to take pictures of the environment and analysing the images. It could also be inferred if the mobile device could work out its latitude and longitude from an embedded GPS receiver, or from other separate means. We assume we have deduced this information, and supply it to the ESN as inputs, both at the training and characterisation stages. The teacher signal remains unchanged, as it already supplies the network with the correct characterisation result.

Adding Further Data

In other tests we also supplied the ESN with further extracted radio data. Using the same time window as used in the zero crossing detection system, we calculated the dynamic range of the channel's magnitude response, and added this to the training data for the ESN. Doing so significantly increased both training and simulation time of the ESN. Again, there was no need to alter the teacher signal.

Comparison System

In order to asses the echo state network's effectiveness, we also constructed a statistical characterisation system, which like the ESN system did not use a pre-defined system model. Instead of using a neural network to characterise signals, it created probability mass functions (PMFs) of signal power, and compared the power distributions between teacher and unknown data using the Kullback-Leibler divergence (KLD). The KLD is computed using (1).

$$\mathrm{KLD}(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \qquad (1)$$

The KLD metric was chosen to compare PMFs as it measures the dissimilarity between two distributions and has previously shown promise in comparing power histograms [10], [11]. Any base of logarithm can be used, however using base 2 logarithms gives a divergence in bits between the two systems. We calculate the KLD between the PMF of the teacher signal, p, and the PMF of the unknown signal, q. If the two distributions are identical, the KLD will be equal to zero. The more dissimilar the PMFs are, the larger the KLD between them will be. Our system chooses the channel which has the smallest KLD between the teacher and unknown signals as the most likely to be the correct result. We supply this system with the same channel magnitude information supplied to the ESN system for a fair comparison. It is not suitable for processing simple location data, although this could be implemented using combinatorial logic after the system's output. By narrowing the search by excluding certain channels, a similar effect as in the ESN system could be reached.

III. Results

In our baseline ESN system using only the zero crossing data, we were able to correctly characterise which of the thirteen channels was used in 68% of cases over a large run of tests. We consider this a success, since this characterisation work is done without an explicit physical model of the wireless system. Any model exists solely within the echo state network, and was generated entirely autonomously from training data. We see this as the primary strength of this method - that there is no need for a system model to be programmed for the system to be able to discriminate between different channel types. Since it is entirely data-driven (rather than model-driven) the effort required to construct such a system is much less than an equivalent system using more traditional methods, and is also more agile - adding a new or updated classification simply involves gathering suitable training data and re-training the system with it.

This result compares favourably with the traditional statistical method: using the same channel data, our KLD system makes a correct characterisation in only 35% of cases (see Fig. 6). Additionally, the incorrect classifications are more uniformly spread when using the KLD approach than with the ESN. Rather than mis-classifying one channel as another similar one, the KLD metric produces a randomised spread of errors. In the case of the ESN, there are a number of cases (especially with the A1, A2 and B3 channel types) where the errors are clustered heavily (see Fig. 5). In these cases, the ESN has recognised that the two channels have similar characteristics, while the KLD method makes no such inference.

Analysis of Errors

After an investigation into the incorrectly classified channels, we found that in a significant number of cases, the correct channel had a prediction score extremely close to the mis-predicted channel. This was often the case with similar channel types, meaning that the ESN perceived these two channels as having very similar characteristics. This is immediately evident in the spread of errors in Fig. 5 - channel type A1

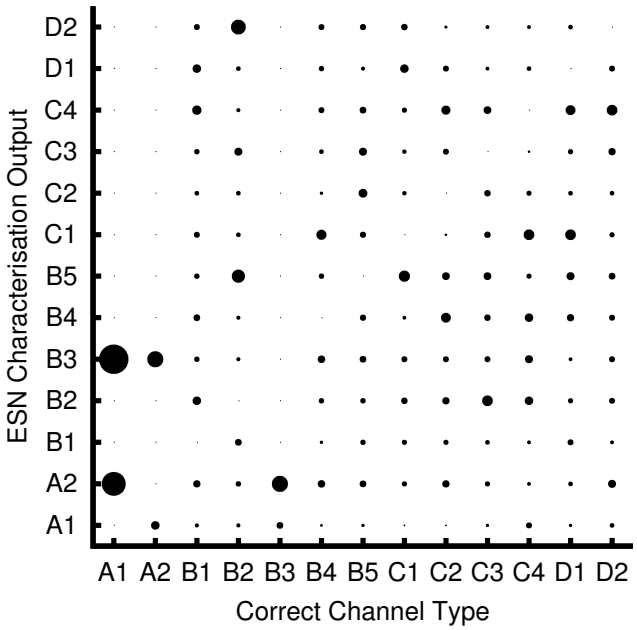


Fig. 5: Graph showing predicted vs. actual channel types from a randomised simulation using the trained echo state network system. All correctly classified channels have been omitted. Heavy clustering around similar channel types is evident in a number of cases, most visibly in the case of type A1 channels.

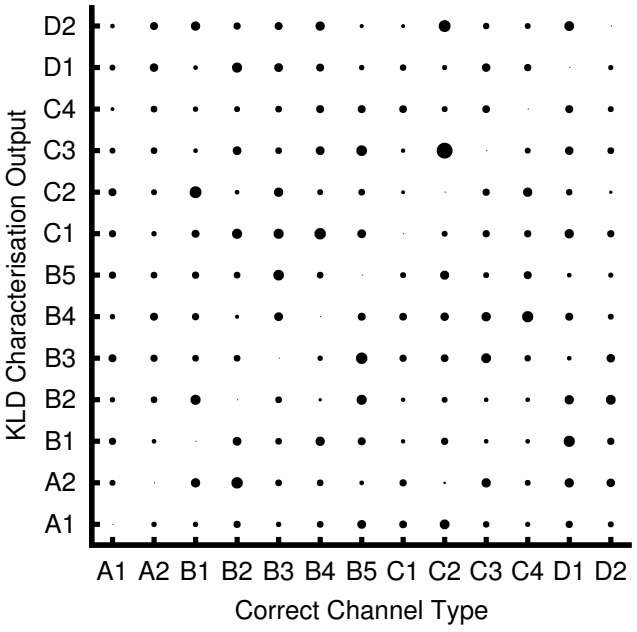


Fig. 6: Graph showing predicted vs. actual channel types from a randomised simulation using the KLD metric to measure the similarity of training data to each unknown channel. As in Fig. 5, all correctly classified results have been omitted. This shows a relatively uniform spread of errors, as would be expected from such a system which uses a purely statistical classification method, with only a small amount of clustering around certain similar channel types.

(indoor office) is a very similar scenario to both A2 (indoor to outdoor) and B3 (indoor hotspot). These mis-classifications have the following probabilities: $P(A2|A1) = 0.18$ and $P(B3|A1) = 0.23$. This is likely because from the limited amount of information available to the ESN (the standard deviation of zero-crossings) the differences in other channel parameters (e.g. delay spread, phase information) cannot be extracted. This class of mis-prediction could probably be eliminated by supplying the ESN with more data from the RF domain. Additionally, an error of this kind may not adversely affect a real running system. If the channel parameters are genuinely similar, then the transmission scheme used for those sets of channel conditions are also likely to be similar.

The remainder of the mis-classified channels from the ESN system (those not identified as a similar channel) seem to be classified entirely randomly, having the highest prediction score of any channel, sometimes by a large margin. The reason for this behaviour is not clear. It may be because the signal contained some feature not present in the training data, however using a longer training sequence than we have for these results did not affect the rate of these random mis-classifications. It is also a possibility that the Neural Network had reached its limit for the complexity of the modelled system. This also seems unlikely, as (in Fig. 4) increasing the numbers of neurons above 1000 (as used in the main experiments) showed no change in the rate of incorrect classifications. We could find no pattern to this type of mis-predication, and it was not repeatable except by using the exact same ESN parameters, neuron weights, channels and RF data.

This comparison with another 'blind' system indicates that ESNs have the ability to extract information from their training data that other traditional statistical methods do not, allowing for significantly more accurate characterisation results.

Using simple location data

Adding the coarse location data improved the ESN's correct predication rate to 72%. This is a very useful improvement over the baseline system and if this information can be gathered we feel it is a worthwhile addition. It works by excluding some of the channel types, allowing the ESN to focus only on the possible set of channels, which may have the additional benefit of reducing the computational workload.

Interestingly, adding the magnitude response dynamic range data did not further improve the prediction rate above 72%. This may be because the limited amount of information contained within this data gives no further knowledge to the system over that which can be extracted from the zero-crossing spread data.

IV. Conclusions

The echo state network approach to channel characterisation clearly has potential, most notably due to the lack of need of a system model. Since the system is trained entirely on measured data and makes no assumptions about the physical system it is trying to characterise, it is highly adaptable to a wide variety of problems. As it outperforms other statistical methods, it clearly demonstrates the learning nature of the echo state approach gives insight other systems are incapable of detecting. Provided the information can be supplied to the receiver, the addition of simple location data is a very effective method for increasing the likelihood of a correct classification. This could conceivably be extended to cover a more diverse or specific set of location types, dependent on the types of channels expected.

To optimally solve the problem of channel characterisation, we would suggest a hybrid system, incorporating an echo state network for the classification work, and to tailor the information made available to it closely to the specific RF system. Supplying information from pilot signals appropriate to the wireless link to the ESN (in addition to the windowed zero-crossing rate we have employed here) is likely to allow the system to distinguish between a wider range of channel types than this system can, with an even higher rate of correct prediction.

Acknowledgements

The authors would like to thank Agilent Technologies UK and USA for sponsoring this research.

References

[1] J. Salo, M. Nur-Alam, and K. Chang, "Practical introduction to lte radio planning," *A white paper on basics of radio planning for 3GPP LTE in interference limited and coverage limited scenarios, European Communications Engineering (ECE) Ltd, Espoo, Finland*, 2010.

[2] R. Ganesh and K. Pahlavan, "Statistical characterization of a partitioned indoor radio channel," in *Communications, 1992. ICC '92, Conference record, SUPERCOMM/ICC '92, Discovering a New World of Communications., IEEE International Conference on*, jun 1992, pp. 1252 –1256 vol.3.

[3] H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks-with an erratum note," Technical Report GMD Report 148, German National Research Center for Information Technology, Tech. Rep., 2001.

[4] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, p. 78, 2004.

[5] M. Win and R. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor channels: a communication-theoretic view," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 9, pp. 1613 – 1627, dec 2002.

[6] J. Dai, G. Venayagamoorthy, and R. Harley, "An introduction to the echo state network and its applications in power system," in *Intelligent System Applications to Power Systems, 2009. ISAP'09. 15th International Conference on*. IEEE, 2009, pp. 1–7.

[7] D. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3g systems: extending the 3gpp spatial channel model (scm)," in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, vol. 5. IEEE, 2005, pp. 3132–3136.

[8] M. Narandzic, C. Schneider, R. Thoma, T. Jamsa, P. Kyosti, and X. Zhao, "Comparison of SCM, SCME, and WINNER channel models," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th.* Ieee, 2007, pp. 413–417.

[9] Various, "WINNER II Channel Models," Winner, Tech. Rep., 2007, also available at http://www.ist-winner.org/deliverables.html.

[10] M. Afgani, S. Sinanović, and H. Haas, "Anomaly Detection Using the Kullback-Leibler Divergence Metric," in *Proc. of the First International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*. Aalborg, Denmark: IEEE, Oct. 25–28 2008, p. 5.

[11] ——, "Hardware Implementation of a Kullback-Leibler Divergence Based Signal Anomaly Detector," in *Proc. of the Second International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL) (Invited Paper)*. Bratislava, Slovak Republic: IEEE, Nov. 24 – 27 2009, p. 6.

Indoor Channel Prediction Using an Efficient Sum of Sinusoids Linear Prediction Scheme

Alan Anderson and Harald Haas
Institute for Digital Communications
The University of Edinburgh
King's Buildings, Mayfield Road, Edinburgh, EH9 3JL, UK
Email: {a.anderson, h.haas} @ed.ac.uk

***Abstract*—In this paper we present a method for predicting a wireless channel using an efficient sum of sinusiods (SOS) method. Our method models a channel as the sum of a discrete number of rays, and extrapolates these rays forward along a trajectory when given accurate measurements of the component rays at distinct spatial points. In simulations, we have been able to eliminate the need for regular pilot signals in an OFDM system over a distance of 10 cm, increasing system data throughput by 17% in the case of a moving receiver.**

***Index Terms*—Channel prediction, ray tracing, sum of sinusoids**

I. Introduction

A. Importance of Channel Knowledge

Knowledge of a wireless channel's characteristics is vital to modern wireless systems, where a fast-fading channel's gain can vary by several orders of magnitude over a millisecond timescale. Poor information leads to high error rates, wasted power, and sub-optimal throughput of data. Adaptive transmission technologies take advantage of channel information to choose a modulation scheme, power level and scheduling algorithm to maximise efficiency in transmission. A channel's behaviour is often treated as being so complex as to be indistinguishable from a truly random process. As such, channel characteristics are directly measured using pilot signals - typically short signals of known phase, magnitude and duration which can be used to measure a channel's distortion effects, but carry no end-user data. The information supplied by the pilot signal is then treated as being valid over the coherence time of the channel. After that time period has expired, further measurement is performed. This endless cycle of pilot signalling allows for reasonable channel knowledge at the expense of wasted bandwidth and power transmitting pilot signals, as well as increased system complexity. A scheme which could reduce the number of pilot signals while maintaining the ability to supply accurate channel information could be very useful in increasing spectral efficiency.

If a method for perfectly predicting a channel's future response were discovered, it would allow huge gains in energy efficiency to be made. Deep fades could be worked around by re-scheduling transmission in time or frequency, power efficiency could be maximised by exploiting the perfect knowledge of the channel's gain and phase shift, and the need for pilot signals to actively measure the current channel would be reduced. In modern multiple-input, multiple-output (MIMO) radio systems (such as LTE) power control signalling can use upwards of 30% [1] of a system's link budget. Eliminating this would make this bandwidth available to the end user.

B. Existing Work in Channel Prediction

Channel prediction, because of its potential benefits, has seen much research [2], with the bulk of work fitting into two broad classifications: autoregressive methods and sum of sinusiods modelling.

1) Autoregressive Models: The most widespread channel prediction schemes use autoregressive (AR) models [3], which presume that the channel's future behaviour can be predicted by a weighted linear combination of its previous values. AR models have successfully been used to predict long-term fading effects quite effectively [4], however AR algorithms can be very sensitive to noise, and so prove more challenging when being used to predict short-term fades [5]. The AR approach has the advantage that the model itself can be computationally simple to implement, however good prediction of the AR coefficients relies on accurate channel measurements needed to estimate the correlation function of the channel coefficients, which in practise can be complex.

2) Sum of Sinusoids: The sum of sinusoid (SOS) method [6] assumes that since the real channel is composed of sinusoids of differing gains, phases and Doppler shifts, this can be modelled given enough information on the transmitters, receivers, scatterers, reflectors and other physical channel parameters. Taken to the extreme, this is treated as a ray-tracing problem, aiming to exactly replicate the entire channel environment. If the model is close enough to reality, then the sinusiods can be extrapolated into the future, and an exact prediction of the channel at that point can be arrived at. The assumption that information on a device's surroundings is available is rarely realistic, and the computational requirements for such modelling are significant, further adding to the difficulty of using such models in real-world systems. As such, these systems are commonly used in simulations (where knowledge of the complete channel environment can be assumed), but are rarely used to model real scenarios.

Some other work has also been carried using so-called basis expansion techniques [7], [8] using band-limited process model-based prediction algorithms, which predict fading coefficients. This is a comparatively new area and results on real world channels have been mixed, seeming to be heavily dependent on channel type.

C. Motivation

As has been established, schemes which depend on active channel measurement devote a significant proportion of their resources to pilot signals, power control signalling and calculations which do not communicate data across the channel, increasing power usage, and decreasing system throughput. If the transmitter has knowledge of the channel conditions, it can use adaptive transmission techniques to best match channel conditions, and use scheduling and modulation schemes to optimise throughput. If the receiver has accurate channel knowledge, it can decode the transmitted data with increased accuracy. Additionally, if both transmitter and receiver have channel knowledge, it removes the need for a large amount of signalling overhead, which can use a significant proportion of the link budget to communicate power control data.

In this paper we present a method to predict a channel's future behaviour based on measurements at a number of spatial points, allowing for much less frequent channel measurement, leading to a system with a higher spectral efficiency. It will allow higher data throughput, better scheduling decisions, fewer errors and overall lower power usage.

II. Methodology

In order to make predictions about a channel's behaviour in the spatial domain, we make a number of assumptions for our model. We assume that:

- the channel can be viewed as being composed of a discrete number of rays
- we are able to obtain phase and magnitude information for each individual ray at our receive antenna
- for the duration of the prediction simulation, the number of rays is consistent, with no existing rays being cut off, or new rays being received
- the receiver will continue moving in a straight line along its current trajectory at constant speed

As the distances over which we predict the channel are of the order of several tens of centimetres, these assumptions are reasonable - it models a mobile pedestrian user in an indoor environment very well. The number of reflections from walls (corresponding to rays) is likely to be consistent, as is the receiver's trajectory and speed. Simulating higher mobility scenarios may become possible by extending the model.

We begin by measuring the phase and magnitude of each ray at two spatial points, A and B separated by a known displacement vector. We take the difference in phase and magnitude between the two points for each component ray, and use this to create a linear extrapolation in the spatial direction AB for each of the individual rays. By using the approximation that the change in path length (and hence phase) for each ray will increase approximately linearly as distance along the trajectory increases, we can predict the individual ray's phase at a given point along its current trajectory. Similarly, we model path loss as a linear change in magnitude over distance.

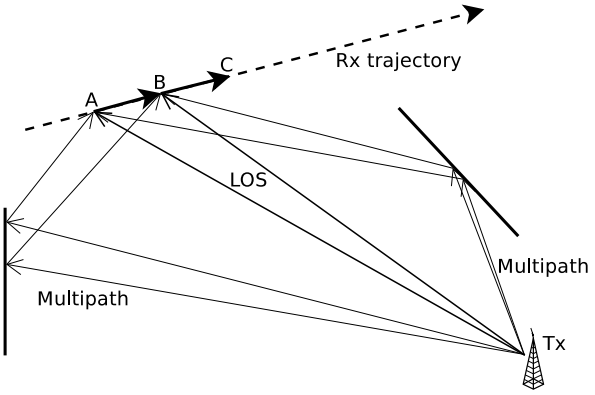


Fig. 1: A diagram showing the receiver's trajectory and the multiple rays coming from the transmitter, including both line-of-sight (LOS) and multipath components. A and B are the two spatial points at which all the separate rays are known, while point C is the point at which a prediction will be made. We model the channel from the transmitter to the receiver as being composed of the sum of these rays.

The the rays at the two points, A and B which the prediction is based on can be written as:

$$Ae^{j\theta^A} = \sum_{i=1}^{I} a_i e^{j\theta_i^a} \tag{1}$$

$$Be^{j\theta^B} = \sum_{i=1}^{I} b_i e^{j\theta_i^b} \tag{2}$$

Where we use I discrete rays in the calculation. a_i, b_i correspond to the gains, and θ_i^a, θ_i^b the phases of each ray. Hence their difference can be calculated by:

$$\Delta e^{j\theta^\Delta} = \sum_{i=1}^{I} (b_i e^{j\theta_i^b} - a_i e^{j\theta_i^a}) \tag{3}$$

This Δ calculation is then used to predict the next data point (corresponding to point C in Fig. 1). We can then project this prediction forwards along the receiver's trajectory and make a prediction about the channel at any arbitrary point along it. Similarly, this calculation could also be used to predict in the opposite direction, to calculate previous channel conditions if required.

III. Simulation Setup

Since this work depends on knowing the individual component rays comprising a channel, a ray-tracking computer model is ideally suited, as it allows trivial extraction of data on each individual ray without the need for advanced processing. Our simulator models a discrete, parameterisable number of rays within a room. We take account of line of sight rays and 1st order reflections. Room size, number of rays and reflectivity of the internal walls are all changeable. We chose a centre frequency of 2.4GHz for these simulations.

In our scenario we have a fixed transmitter inside the room and a mobile receiver which moves on a randomly assigned straight line path through the room. We take two consecutive

readings of the rays along the receiver's path and use them as points A and B in our prediction algorithm. We then compare our prediction algorithm's performance with the ray-tracing result (which we treat as being error free). We also additionally introduce noise into the simulation system. By specifying an additive white Gaussian noise (AWGN) power, we attempt to simulate the real world more closely. The AWGN is used to represent the many small reflextors a real room would have, representing the many small items which would cause more complex and unpredictable reflections that those from the walls or other large, flat surfaces.

Fig. 2 shows the output from a single run of the simulator, comparing the real channel to our prediction of that same channel.

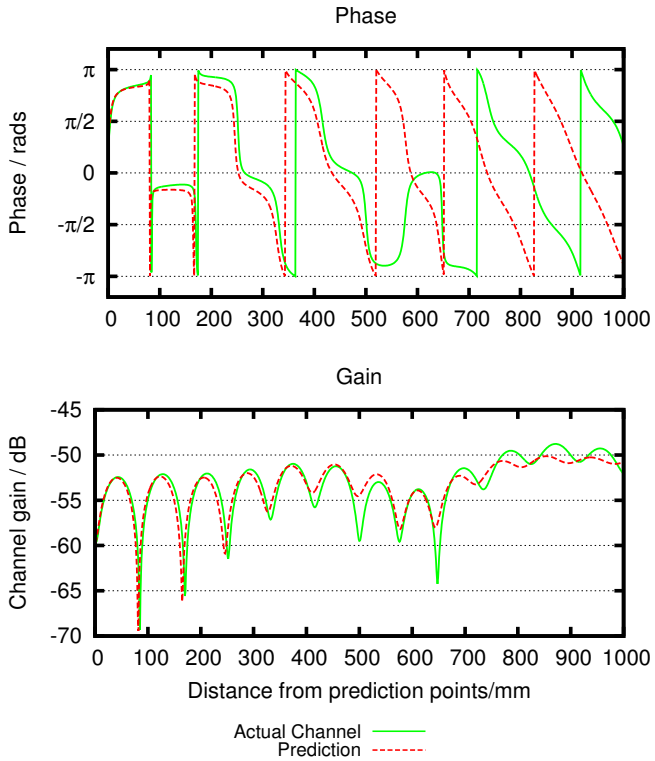


Fig. 2: A plot of phase and magnitude response of one channel taken from the dataset used to build Fig. 3. It shows the actual channel data and the prediction, and it can be seen how prediction accuracy decreases as distance increases from the two initial measured data points. Note particularly how the accuracy of deep fade depth prediction changes a distance increases.

IV. Results

Fig. 3 has been generated by Monte Carlo simulation of 10,000 iterations per data point, with signal to nose ratios (SNR) from 10 dB to 60 dB. It shows how prediction accuracy decreases as distance increases from the two known points, A and B, under a wide range of noise conditions, showing that even at extremely favourable SNRs, prediction accuracy displays the same features as at more realistic SNRs. These results show an increase in errors at particular spatial points. These correspond to where deep fades occur in the channel.

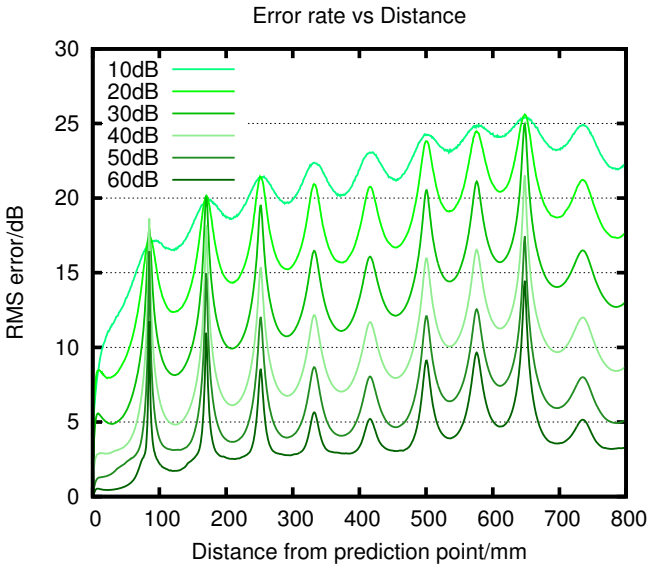


Fig. 3: A diagram showing error rates as distance increases for a particular simulation scenario with varying SNR. It clearly shows the relationship between SNR and prediction error rate. Also evident are the areas of deep fading when the prediction of the fade depth is not as accurate as in other areas.

Fig. 2 shows that although the prediction scheme does identify these deep fades, the accuracy of fade depth prediction deteriorates as distance from the two known points increases. This feature can be seen in the the gain plot in Fig. 2 at a distance of 650mm.

A. Bit-error rate performance

In order to assess the comparative performace of this prediction scheme, we compare it to a standard OFDM system which can use pilots in order to actively measure the current channel conditions and correct the received signal accordingly. For additional comparison we have included the optimal solution (where the OFDM receiver has complete channel knowledge) and the case where no pilots or predictions are used in decoding the data. For each scenario in this simulation, the ODFM frame encodes 96 bits, uses 16-QAM modulation on each subcarrier and the cyclic prefix is set to 25% of the frame's length. It includes 4 pilot signals, equally spaced in frequency among the data subcarries, as illustrated in Fig. 4. For the prediction scheme, the two known points are assumed to have been calculated with ideal channel knowledge. The simulation results are shown in Fig. 5.

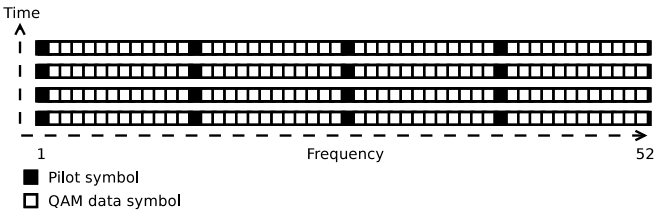


Fig. 4: Structure showing OFDM frames, and how pilots signals are spaced among the 52 subcarriers.

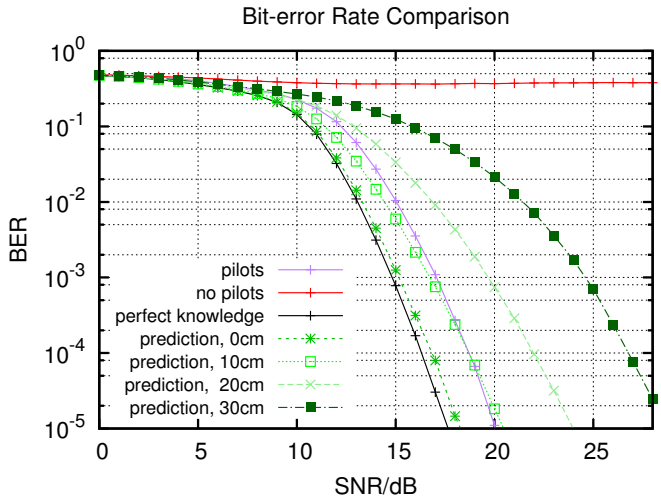


Fig. 5: A plot showing BER at a range of SNRs for four different scenarios: imperfect channel knowledge from pilot signals, no channel knowledge from pilot signals, perfect channel knowledge,and predicted channel knowledge (at four distances). It is interesting to note that the performance of the prediction system is marginally better than even the pilot-assisted system at 10cm.

As would be expected, the perfect channel knowledge scenario achieves the lowest error rate in all SNRs, and the zero channel knowledge scenario always has the highest error rate. Also, in line with the results in Fig. 3, since prediction accuracy degrades as distance from the prediction points increase, the error rate increases.

The first prediction data set (at 0 cm from the known points) produces near-ideal performance, since it is based on two points with ideal channel knowledge. As can be been seen clearly from the graph, its BER is almost an order of magnitude lower than pilot-assisted OFDM, which represents a typical best-in-class practical system. Achieving such gains in a real-world scenario would be highly advantageous.

At a distance of 10 cm from the known points (corresponding to 100 prediction samples), the results are comparable to those from the pilot-assisted OFDM simulation, showing that our system can achieve similar BER performance to an active channel sampling system without the need for repeated pilot signals. In this particular implementation, we could replace the 4 pilot signals with more data carrying QAM symbols, increasing the data carried by the frame by a further 16 bits. This represents an increase of almost 17% in data throughput, with no increase in power usage, or decrease in error performace.

The predictions at 20 cm and 30 cm from the known points do not perform as well as the reference pilot-OFDM system, so a real-world system would have to choose the optimal distance over which to trust the predictions before errors become too large.

B. Advantages of this System

Given sufficiently accurate ray information at the two prediction points, our system is able to predict a large number of samples into the future, which in the modelled system corresponds to distances in the range of tens of centimetres. In a typical indoor channel scenario with a moving receiver, this could represent hundreds of milliseconds of prediction into the future, depending on the mobile device's velocity. This makes it ideally suited to adaptive transmission systems. Such accurate predictions would completely eliminate the need for any pilot signals during this period, allowing those resources to be used for data transmission, increasing spectral efficiency, which can be converted into power savings or increased throughput. However, as is illustrated in Fig. 3, noise at the prediction points has a large impact on the system's accuracy, and ensuring that accurate measurements are taken is absolutely vital.

Additionally, our algorithm requires very few computational resources to provide an accurate output, making it suited to real-time applications.

V. Generalising to Polynomial Interpolation

The two-point interpolation method described so far is identical to using n=1 order polynomial interpolation, so we can extend the method to higher order polynomial interpolation. More accurate predictions can be arrived at, however it comes at the expense of measuring the rays at n+1 known spatial points.

Using the standard method for Vandermonde polynomial interpolation, we are able to produce a set of characteristic polynomials which pass through each of the measured points for each ray. This method is able to much more accurately model behaviour of each ray, as signal strength does not change linearly with respect to distance, but a polynomial fit is able to provide a very close approximation. Although this method does require more computational resources than a simple linear (two-point) interpoltion, the increase is minimal.

A. Sources of Errors

Although this method is capable of good future predictions, there are a number of areas which if improved upon could further increase the potential for more accurate, even longer-term predictions.

1) As the entire prediction dataset is based upon having accurate readings at a limited number of points, naturally the algorithm is sensitive to noise at these points. The results clearly show how noise at the prediction points cause errors to grow as the initially small discrepancy is multiplied over distance. Having a larger number of data points to base the prediction from would allow denoising algorithms. An algorithm such as ESPRIT [9] to estimate the rays' strengths could provide this benefit, as well as calculating the individual ray components. A trade-off would have to be reached here, as a lager number of measurements would mean sacrificing some of the system's gains in spectral efficiency.
2) The algorithm assumes that the same set of rays will be present for the duration of the prediction period. It is however possible for new rays to appear or existing rays be blocked out by room geometry such as windows or corners. In this case, the system would have to sample a new set of new prediction points, and start a new set of calculations.

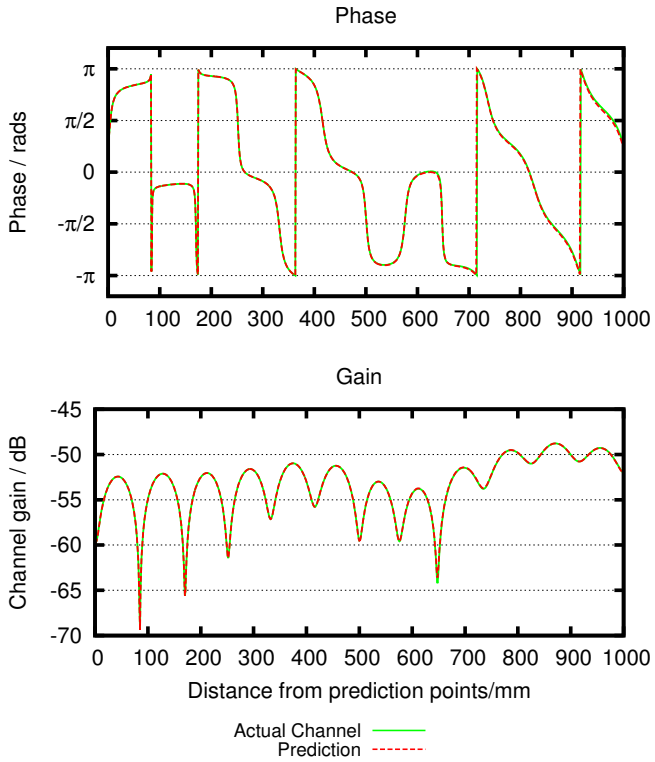


Fig. 6: Prediction of the same channel shown in Fig. 2, but using n=3 order polynomial interpolation, using ray data collected at 4 spatial points. Over the distance shown the largest error in magnitude prediction is 0.5dB, marking a significant improvement over the previous result, at the cost of gathering more ray data.

VI. Challenges in a Real-World Setting

As this method requires knowledge of the component rays, applying it to a real scenario requires the additional step of estimating them from their summation. Algorithms such as MUSIC [10] and ESPRIT [9] exist to estimate parameters of a composite signal's component rays, and have already been used to aid channel prediction [11], [12]. Integrating such methods into this system would enable it to make predictions on real-world channels.

Having a method to determine when the set of rays being observed by the receiver changes significantly would also be necessary. The system would need to either correct its existing set of predictions, or take a new set of measurements to create new predictions to take account of the new set of rays.

Finally, knowledge of the receiver's position is key to this prediction scheme. Having accurate location data is required to achieve a useful prediction. This information could come from a range of sources - many modern smartphones are location-aware, and this information could be applied to the prediction scheme.

VII. Conclusions and Future Work

The method of channel prediction described in the paper shows promise as an efficient channel predictor, particularly well suited to indoor channels. It is very computationally efficient, making it ideal for inclusion into mobile receivers as well as transmitters, and has the potential to significantly reduce the overhead inherent in systems dependent on pilot signals and explicit power control signalling. Because it can provide comparatively long-term predictions, it has the additional benefit that it enables the transmitter to use adaptive transmission technologies to make optimal choices when scheduling time, power and frequency resources.

Future work will investigate the potential to have semi-regular points at which the channel is measured, allowing for corrections to be made to an existing prediction, rather than having to produce an entirely new set of predictions, and finding the optimal rate at which such measurements should be made. It will also investigate the optimal degree of polynomial to provide the best possible predictions.

Increased resilience to noise is of prime interest, especially when applying this work to real-world conditions. Also incorporating a method for detecting when errors become very large would be useful, to allow the prediction scheme to take action, either to make minor corrections to its existing predictions, or to start a new set of predictions entirely.

Acknowledgements

The authors would like to thank Agilent Technologies UK and USA for sponsoring this research.

References

[1] J. Salo, M. Nur-Alam, and K. Chang, "Practical introduction to lte radio planning," *A white paper on basics of radio planning for 3GPP LTE in interference limited and coverage limited scenarios, European Communications Engineering (ECE) Ltd, Espoo, Finland*, 2010.

[2] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2299 –2313, dec. 2007.

[3] K. E. Baddour and N. C. Beaulieu, "Autoregressive models for fading channel simulation," in *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, vol. 2. IEEE, 2001, pp. 1187–1192.

[4] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2299–2313, 2007.

[5] C. Fung and S. Chan, "Estimation of fast fading channel in impulse noise environment," in *IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002.*, vol. 4, 2002, pp. IV–497 – IV–500 vol.4.

[6] M. Pätzold and B. O. Hogstad, "Classes of sum-of-sinusoids rayleigh fading channel simulators and their stationary and ergodic properties-part i," *WSEAS Transactions on Mathematics*, vol. 5, no. 2, p. 222, 2006.

[7] T. Zemen, C. Mecklenbräuker, and B. Fleury, "Time-variant channel prediction using time-concentrated and band-limited sequences," in *IEEE International Conference on Communications, 2006. ICC '06.*, vol. 12, june 2006, pp. 5660 –5665.

[8] R. Lyman and A. Sikora, "Prediction of fading envelopes with diffuse spectra," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 3, march 2005, pp. iii/753 – iii/756 Vol. 3.

[9] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques- esprit," in *Circuits, Systems and Computers, 1985. Nineteeth Asilomar Conference on*, nov. 1985, pp. 83 –89.

[10] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276 – 280, mar 1986.

[11] I. C. Wong and B. L. Evans, "Wlc43-5: Low-complexity adaptive high-resolution channel prediction for ofdm systems," in *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE.* IEEE, 2006, pp. 1–5.

[12] S. Semmelrodt and R. Kattenbach, "Investigation of different fading forecast schemes for flat fading radio channels," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 1, oct. 2003, pp. 149 – 153 Vol.1.

Appendix B

Sample Source Code Listings

The following section contains a listing of some of the important files of code written for this thesis. Each of the following files are written in MATLAB.

ESN Channel Characterisation Simulation

The listing given here shows an experiment to generate a channel of each of the standard WINNER scenarios, train an ESN with 1,000 neurons, and then evaluate whether the trained network can correctly identify which scenario some further Radio Frequency (RF) data comes from.

```
function []=fixed_scenario_sim()
%   Create one instance of each channel, and generate n+x seconds of random samples.
%   Use the first n seconds from each network to train an ESN (1000 input units)
%   Then play the remaining x seconds of RF from one of the channels, and
%   try to identify it

%add the winner directory to the path
path(path,'../../winner/winner2/');
setpath;

SCENARIO_SET={'A1', 'A2', 'B1', 'B2', 'B3', 'B4', 'B5', 'C1', 'C2', 'C3',
'C4', 'D1', 'D2'};

[i num_channels] = size(SCENARIO_SET);

time=150;       %150 * 100 samples
channels=zeros(num_channels,100*time);

for(scenario = SCENARIO_SET)
```

```
    channels(i,:) = generate_channel(scenario,time);
    i=i+1;
end

training_output = zeros(num_channels,10000*num_channels);
for(i=1:num_channels)
    training_output(i,:) = ones_and_zeros(num_channels,i);
end

training_input=reshape(channels(:,1:10000).',1,[]);

% initialize ESN weights
% 1% of recurrent weights set from [-0.40, 0.40]
% 100% of input weights (bias) set from [-0.20, 0.20]
% 100% backward weights set randomly from [-1.00, 1.00]
probInp  = [  1.00 ];
rngInp   = [  1.00 ];
probRec  = [  0.01 ];
rngRec   = [  0.40 ];
probBack = [  0.00 ];
rngBack  = [  0.00 ];

% unit counts (input, hidden, output)
IUC = 1;         %one input time series
HUC = 1000;
OUC = num_channels;         %corresponds to the number of different channels

% create esn network
Lambda = 0.0;
UnitAct = 11;
[net] = rnn_esn_new(IUC, HUC, OUC, probInp, rngInp, probRec, rngRec, probBack, rngBack, Lambda,

%train the network
[net, MSE] = rnn_esn_train(net, training_input,training_output,1000, 1e-10);

means=zeros(num_channels);         %creates a 13x13 matrix

for( i = 1:num_channels)

IP=channels(i,10001:15000);
TP = zeros(net.numOutputUnits,size(IP,2));            %needs an output of the right size,
[AO, ACT] = rnn_esn_sim(net, IP, TP, 0, 0.0);

means(:,i) =     mean(AO,2) ;            %calculate the mean along dimension 2
end

print "done";
par
function x=ones_and_zeros(tot,n)
```

```
x=0.99*horzcat(zeros(1,(n-1)*10000),ones(1,1*10000),zeros(1,(tot-n)*10000));
```

OFDM BER Simulation Code

This code was used to produce the data plotted in Figs. 5.16 and 5.21. A second script was used to call this code repeatedly, providing the varying arguments passed into the function, for collating the large number of results and writing them to a file to allow them to be plotted.

```
% 64 carriers, 96 bits per frame 16-QAM modulation
% 4 pilots, cylic extension 25% (16)
function [retval]=OFDM_sim(packets,pilot,channelIQ,graph,snr)

% Generating and coding data
t_data=randi([0 1],packets*96,1)';
for p=1:packets;
data=t_data(((p-1)+1):((p-1)+96));
k=3;
n=6;
s1=size(data,2);  % Size of input matrix

% Convolutionally encode data
constlen=7;
codegen = [171 133];     % Polynomial
trellis = poly2trellis(constlen, codegen);
codedata = convenc(data, trellis);

%Interleaving coded data
s2=size(codedata,2);
matrix=reshape(codedata,s2/4,4);
intlvddata = matintrlv(matrix',2,2)'; % Interleave.
intlvddata=intlvddata';

% Binary to decimal conversion
dec=bi2de(intlvddata','left-msb');

%16-QAM Modulation
M=16;
QAM_data = qammod(dec,M);

% Pilot insertion
pilotIQ=3+3j;

k=1;
for a=(1:13:52)
    unpadded_pilot_data(a)=pilotIQ;
    unpadded_pilot_data(a+1:a+12)=QAM_data((k:k+11));
    k=k+12;
```

```
end

% upsizing to 64 pad the top and bottom to avoid ISI
pilot_data(1:6)   = unpadded_pilot_data(1:6)';
pilot_data(7:58)  = unpadded_pilot_data(1:52)';
pilot_data(59:64) = unpadded_pilot_data(47:52)';

% IFFT
ifft_sig=ifft(pilot_data',64);

% Adding Cyclic Extension
cext_data=zeros(80,1);
cext_data(1:16)=ifft_sig(49:64);
cext_data(17:end)=ifft_sig;

% SNR
SNR_range=snr;
for snr_count=1:length(SNR_range)

ofdm_sig=zeros(size(cext_data));
for a=1:7
    ofdm_sig = ofdm_sig+ abs(channelIQ(a))*abs(cext_data).*exp(j*(angle(cext_data)+angle(channel
end
ofdm_sig=awgn(ofdm_sig,SNR_range(snr_count),'measured');% Adding white Gaussian Noise

%Removing Cyclic Extension
rxed_sig=ofdm_sig((1:64)+16).';

% FFT
ff_sig=fft(rxed_sig,64);

% Pilot Synch
synched_sig1=ff_sig(7:58);

%decode using pilots as reference
synched_sig=[];
for a=(1:13:52)
        if(pilot=='auto')       %uses the pilots in the OFDM signal
            synched_sig = [synched_sig synched_sig1(a+1:a+12)*(pilotIQ/synched_sig1(a))];
        elseif(pilot=='none')     %ignores pilots and assumes everything is fine
            synched_sig = [synched_sig synched_sig1(a+1:a+12)];
        else                    %uses a single correction value
            synched_sig = [synched_sig synched_sig1(a+1:a+12)*(1/pilot)];
        end
end

% Demodulation
dem_data= qamdemod(synched_sig,16);
```

```
% Decimal to binary conversion
bin=de2bi(dem_data','left-msb');
bin=bin';

% De-Interleaving
deintlvddata = matdeintrlv(bin,2,2); % De-Interleave
deintlvddata=deintlvddata';
deintlvddata=deintlvddata(:)';

n=6;    %Decoding data
k=3;
decodedata =vitdec(deintlvddata,trellis,5,'trunc','hard');
rxed_data=decodedata;

% Calculating BER
rxed_data=rxed_data(:)';

c=xor(data,rxed_data);
errors=nnz(c);

BER(p,snr_count)=errors/length(data);

 end % SNR loop ends here
end % main data loop

% Time averaging for optimum results
[~, snrs]=size(BER);
for col=1:snrs;
    ber(1,col)=0;
    for row=1:packets;
        ber(1,col)=ber(1,col)+BER(row,col);
    end
end
ber=ber./packets;

if(graph==1)
 figure
 i=SNR_range;
 semilogy(i,ber);
 title('BER vs SNR');
 ylabel('BER');
 xlabel('SNR (dB)');
 grid on
end

retval = [SNR_range;ber]';
filename=sprintf('OFDMresults/noise%g.txt',50.0);
dlmwrite(filename,retval);
end
```

Polynomial Ray Interpolation

This code was used during the development of the polynomial ray interpolation work. It compares the prediction result to the channel created by the ray tracer, and optionally plots graphs of a particular ray (specified by interesting_ray) for debugging purposes.

```
function polynomial_interpolation(poly_degree,interesting_ray)

number_of_rays=7;
trace=zeros(1000,8);
for count=1:1000
   trace(count,1:number_of_rays)=rays_at_point(0.255,1,count*0.001,ones(1,number_of_rays));
end
trace(:,8)=sum(trace(:,1:7),2);

polyprediction = zeros(1000,8);
for count=1:7
    mags   = polyval(polyfit([1:poly_degree+1],...
          abs(trace(1:poly_degree+1,count))',poly_degree),[1:1000]);  %mag response
    phases = polyval(polyfit([1:poly_degree+1],...
        angle(trace(1:poly_degree+1,count))',poly_degree),[1:1000]);    %phase response

    [real complex] = pol2cart(phases,mags);
    polyprediction(:,count) = real+(j*complex);
end
polyprediction(:,8)=sum(polyprediction(:,1:7),2);

prediction_mags = 10*log10(abs(polyprediction));
actual_mags = 10*log10(abs(trace));

for i=1:8
        subplot(8,1,i)
        plot(1:1000,prediction_mags(:,i),'g',1:1000,actual_mags(:,i),'r');

        fname = 'mags.txt';
        mags = [prediction_mags(:,i) actual_mags(:,i)];
        dlmwrite(fname,mags,'-append');
        fid = fopen(fname, 'a');  fprintf(fid, '\n\n'); fclose(fid);
end

IQ = [trace(1:1000,interesting_ray) polyprediction(1:1000,interesting_ray)];
plot_phase_and_mag_prettily_from_IQ(IQ)

end
```

Appendix C

Software Packages Used

ESN Library

The library functions used to create, train and simulate ESNs are from a publicly available suite written by Michal Cernansky, a research fellow at the Faculty of Informatics and Information Technologies at the Slovak University of Technology in Bratislava. They are available to download from

`http://www2.fiit.stuba.sk/ cernans/main/download.html`

The software has been validated with a number of datasets, and comes with demonstrations illustrating its ability to predict the Mackey-Glass system

WINNER Channel Models

The WINNER project was a collaboration between 29 partner companies and research institutions to research new technologies to drive forward research into cellular systems. The part of their work used in this thesis is limited to their comprehensive channel modelling software, written in MATLAB. It is publicly available, along with full documentation at

`http://projects.celtic-initiative.org/winner+/deliverables.html`

Graph and Diagram Tools

All graphs in this thesis were created using the free gnuplot graphing software. It allows highly customisable plotting functions from plain text data files.

The majority of the diagrams were created using the Dia editor, with the remainder being created using the Inkscape vector graphics editor.

Bibliography

[1] Abramson, N. The aloha system: another alternative for computer communications. In *Proceedings of the November 17-19, 1970, fall joint computer conference* (1970), ACM, pp. 281–285.

[2] Afgani, M., and Haas, H. Information content analysis and clustering for signal anomaly detection. In *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th* (2009), IEEE, pp. 1–5.

[3] Afgani, M., Sinanović, S., and Haas, H. Anomaly Detection Using the Kullback-Leibler Divergence Metric. In *Proc. of the First International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)* (Aalborg, Denmark, Oct. 25–28 2008), IEEE, p. 5.

[4] Afgani, M., Sinanovic, S., and Haas, H. Information theoretic approach to signal feature detection for cognitive radio. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE* (2008), IEEE, pp. 1–5.

[5] Afgani, M., Sinanović, S., and Haas, H. Hardware Implementation of a Kullback-Leibler Divergence Based Signal Anomaly Detector. In *Proc. of the Second International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL) (Invited Paper)* (Bratislava, Slovak Republic, Nov. 24 – 27 2009), IEEE, p. 6.

[6] Afgani, M., Sinanović, S., and Haas, H. Hardware Implementation of a Kullback-Leibler Divergence Based Signal Anomaly Detector. In *Proc. of the Second International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL) (Invited Paper)* (Bratislava, Slovak Republic, Nov. 24 – 27 2009), IEEE, p. 6.

[7] Afgani, M., Sinanović, S., and Haas, H. The Information Theoretic Approach to Signal Anomaly Detection for Cognitive Radio. *International Journal of Digital Multimedia Broadcasting (Invited Paper) 2010* (June 2010), 18.

[8] Afgani, M. Z. *Exploitation of signal information for mobile speed estimation and anomaly detection.* PhD thesis, The University of Edinburgh, 2011.

[9] Aguiar, A., and Wolisz, A. Channel prediction heuristics for adaptive modulation in wlan. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th* (2007), pp. 1091–1095.

[10] Andersen, J. B., Jensen, J., Jensen, S. H., and Frederiksen, F. Prediction of future fading based on past measurements. In *Vehicular Technology Conference, 1999. VTC 1999-Fall. IEEE VTS 50th* (1999), vol. 1, IEEE, pp. 151–155.

[11] Archer, N. P., and Wang, S. Fuzzy set representation of neural network classification boundaries. *Systems, Man and Cybernetics, IEEE Transactions on 21*, 4 (1991), 735–742.

[12] Awad, Y. A., and Nakamura, M. Adaptive modulation and coding, 2006. EP Patent 1,387,517.

[13] Baddour, K. E., and Beaulieu, N. C. Autoregressive models for fading channel simulation. In *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE* (2001), vol. 2, IEEE, pp. 1187–1192.

[14] Bakar, Z., Mohemad, R., Ahmad, A., and Deris, M. A comparative study for outlier detection techniques in data mining. In *Cybernetics and Intelligent Systems, 2006 IEEE Conference on* (2006), pp. 1–6.

[15] Basu, S., and Meckesheimer, M. Automatic outlier detection for time series: an application to sensor data. *Knowledge and Information Systems 11*, 2 (2007), 137–154.

[16] Baum, D., Hansen, J., and Salo, J. An interim channel model for beyond-3g systems: extending the 3gpp spatial channel model (scm). In *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st* (2005), vol. 5, IEEE, pp. 3132–3136.

[17] Bellman, R. E., and Dreyfus, S. E. *Applied dynamic programming*, vol. 7962. Princeton University Press, 1966.

[18] Berrou, C., and Glavieux, A. Near optimum error correcting coding and decoding: Turbo-codes. *Communications, IEEE Transactions on 44*, 10 (1996), 1261–1271.

[19] Bertoni, H. L. *Radio propagation for modern wireless systems*. Pearson Education, 1999.

[20] Bishop, C. M. *Neural networks for pattern recognition*. Oxford university press, 1995.

[21] Boccato, L., Lopes, A., Attux, R., and Von Zuben, F. An echo state network architecture based on volterra filtering and pca with application to the channel equalization problem. In *Neural Networks (IJCNN), The 2011 International Joint Conference on* (2011), pp. 580–587.

[22] Catreux, S., Erceg, V., Gesbert, D., and Heath Jr, R. W. Adaptive modulation and mimo coding for broadband wireless data networks. *Communications Magazine, IEEE 40*, 6 (2002), 108–115.

[23] Chandola, V., Banerjee, A., and Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. 41*, 3 (2009), 1–58.

[24] Chrysanthou, C., and Bertoni, H. L. Variability of sector averaged signals for uhf propagation in cities. *Vehicular Technology, IEEE Transactions on 39*, 4 (1990), 352–358.

[25] Chunli, D., Yuning, D., and Li, W. Autoregressive channel prediction model for cognitive radio. In *Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09. 5th International Conference on* (2009), pp. 1–4.

[26] Cover, T., and Hart, P. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on 13*, 1 (1967), 21–27.

[27] Cover, T., Thomas, J., and Wiley, J. *Elements of information theory*, 2 ed., vol. 1. Wiley-Interscience, 2006.

[28] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems 2*, 4 (1989), 303–314.

[29] Dai, J., Venayagamoorthy, G., and Harley, R. An introduction to the echo state network and its applications in power system. In *Intelligent System*

Applications to Power Systems, 2009. ISAP'09. 15th International Conference on (2009), IEEE, pp. 1–7.

[30] Dai, J., Zhang, P., Mazumdar, J., Harley, R. G., and Venayagamoorthy, G. A comparison of mlp, rnn and esn in determining harmonic contributions from nonlinear loads. In *Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE* (2008), IEEE, pp. 3025–3032.

[31] Desforges, M., Jacob, P., and Cooper, J. Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 212*, 8 (1998), 687–703.

[32] Devijver, P. A., and Kittler, J. *Pattern recognition: A statistical approach*. Prentice/Hall International Englewood Cliffs, NJ, 1982.

[33] Dong, L. Turbo equalization with channel prediction and iterative channel estimation. In *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE* (2009), pp. 1–6.

[34] Duel-Hallen, A. Fading channel prediction for mobile radio adaptive transmission systems. *Proceedings of the IEEE 95*, 12 (2007), 2299–2313.

[35] Duel-Hallen, A., Hu, S., and Hallen, H. Long-range prediction of fading signals. *Signal Processing Magazine, IEEE 17*, 3 (2000), 62–75.

[36] Ekman, T., Sternad, M., and Ahlén, A. Unbiased power prediction on broadband channels. In *IEEE VTC 2002-Fall* (2002), Citeseer, pp. 24–28.

[37] Ericsson, L. Release 12–taking another step towards the networked society, 2013.

[38] Farley, B. G., and Clark, W. Simulation of self-organizing systems by digital computer. *Information Theory, Transactions of the IRE Professional Group on 4*, 4 (1954), 76–84.

[39] Filippone, M., and Sanguinetti, G. Information theoretic novelty detection. *Pattern Recogn. 43*, 3 (2010), 805–814.

[40] French, M. N., Krajewski, W. F., and Cuykendall, R. R. Rainfall forecasting in space and time using a neural network. *Journal of hydrology 137*, 1 (1992), 1–31.

[41] Fung, C., and Chan, S. Estimation of fast fading channel in impulse noise environment. In *IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002.* (2002), vol. 4, pp. IV–497 – IV–500 vol.4.

[42] Gallager, R. Low-density parity-check codes. *Information Theory, IRE Transactions on 8*, 1 (1962), 21–28.

[43] Ganesh, R., and Pahlavan, K. Statistical characterization of a partitioned indoor radio channel. In *Communications, 1992. ICC '92, Conference record, SUPERCOMM/ICC '92, Discovering a New World of Communications., IEEE International Conference on* (jun 1992), pp. 1252 –1256 vol.3.

[44] Geman, S., Bienenstock, E., and Doursat, R. Neural networks and the bias/variance dilemma. *Neural computation 4*, 1 (1992), 1–58.

[45] Glover, I., and Grant, P. *Digital Communications*, 2 ed. Pearson Prentice Hall, 2004.

[46] Gopal, S., Woodcock, C. E., and Strahler, A. H. Fuzzy neural network classification of global land cover from a 1 avhrr data set. *Remote sensing of Environment 67*, 2 (1999), 230–243.

[47] Greenstein, L. J., Erceg, V., Yeh, Y. S., and Clark, M. V. A new path-gain/delay-spread propagation model for digital cellular channels. *Vehicular Technology, IEEE Transactions on 46*, 2 (1997), 477–485.

[48] "GSM Alliance". Mobile economy europe 2013. Tech. rep., "GSM Alliance", 2013.

[49] Hallas, M., and Dorffner, G. A comparative study on feedforward and recurrent neural networks in time series prediction using gradient descent learning, 1998.

[50] Hamming, R. W. Error detecting and error correcting codes. *Bell System technical journal 29*, 2 (1950), 147–160.

[51] Hanzo, L., Münster, M., Choi, B., and Keller, T. *OFDM and MC-CDMA*. Wiley, March, 2004.

[52] Henrion, M., Mortlock, D., Hand, D., and Gandy, A. Classification and anomaly detection for astronomical survey data. In *Astrostatistical Challenges for the New Astronomy*, J. M. Hilbe, Ed., vol. 1 of *Springer Series in Astrostatistics*. Springer New York, 2013, pp. 149–184.

[53] Heo, J., Wang, Y., and Chang, K. A novel two-step channel-prediction technique for supporting adaptive transmission in ofdm/fdd system. *Vehicular Technology, IEEE Transactions on 57*, 1 (2008), 188–193.

[54] Hodge, V., and Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev. 22*, 2 (2004), 85–126.

[55] Inc., M. *Mathworks LTE Systems Toolbox Channel Estimation*.

[56] Inc., W. *Wolfram Mathematica Neural Network Documentation*.

[57] Jaeger, H. The 'echo state' approach to analysing and training recurrent neural networks-with an erratum note. Tech. rep., Technical Report GMD Report 148, German National Research Center for Information Technology, 2001.

[58] Jaeger, H. Tutorial on training recurrent neural networks, covering bppt, rtrl, ekf and the" echo state network" approach, 2002.

[59] Jaeger, H. Echo state network. *Scholarpedia 2*, 9 (2007), 2330.

[60] Jaeger, H., and Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science 304*, 5667 (2004), 78.

[61] Jia, M., Ma, J., Zhu, P., Yu, D.-S., and Tong, W. Adaptive modulation and coding, Sept. 5 2006. US Patent 7,103,325.

[62] Jim, K.-C., Giles, C. L., and Horne, B. G. An analysis of noise in recurrent neural networks: Convergence and generalization. *Neural Networks, IEEE Transactions on 7*, 6 (1996), 1424–1438.

[63] Kittler, J., Hatef, M., Duin, R. P., and Matas, J. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 20*, 3 (1998), 226–239.

[64] Koons, H. C., and Gorney, D. J. A sunspot maximum prediction using a neural network. *Eos, Transactions American Geophysical Union 71*, 18 (1990), 677–688.

[65] Kou, Y., Lu, C., Sirwongwattana, S., and Huang, Y. Survey of fraud detection techniques. In *Networking, Sensing and Control, 2004 IEEE International Conference on* (2004), vol. 2, IEEE, pp. 749–754.

[66] Kullback, S., and Leibler, R. A. On information and sufficiency. *The Annals of Mathematical Statistics 22*, 1 (1951), 79–86.

[67] Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., and Srivastava, J. A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the 2003 SIAM International Conference on Data Mining* (2003), pp. 25–36.

[68] Liang, Y.-C., Chen, K.-C., Li, G. Y., and Mahonen, P. Cognitive radio networking and communications: An overview. *Vehicular Technology, IEEE Transactions on 60*, 7 (2011), 3386–3407.

[69] Lippmann, R. P. Review of neural networks for speech recognition. *Neural computation 1*, 1 (1989), 1–38.

[70] Liu, Q., Zhou, S., and Giannakis, G. B. Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links. *Wireless Communications, IEEE Transactions on 3*, 5 (2004), 1746–1755.

[71] Lyman, R., and Sikora, A. Prediction of fading envelopes with diffuse spectra. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on* (march 2005), vol. 3, pp. iii/753 – iii/756 Vol. 3.

[72] Maass, W., Natschläger, T., and Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation 14*, 11 (2002), 2531–2560.

[73] Mackey, M. C., and Glass, L. Oscillation and chaos in physiological control systems. *Science 197*, 4300 (1977), 287–289.

[74] Manjunath, G., and Jaeger, H. Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural computation 25*, 3 (2013), 671–696.

[75] Markou, M., and Singh, S. Novelty detection: a review - part 2:: neural network based approaches. *Signal Processing 83*, 12 (2003), 2499 – 2521.

[76] Markou, M., and Singh, S. Novelty detection: a review–part 1: statistical approaches. *Signal Processing 83*, 12 (2003), 2481 – 2497.

[77] McCulloch, W., and Pitts, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics 5*, 4 (1943), 115–133.

[78] Mitola III, J., and Maguire Jr, G. Q. Cognitive radio: making software radios more personal. *Personal Communications, IEEE 6*, 4 (1999), 13–18.

[79] Narandzic, M., Schneider, C., Thoma, R., Jamsa, T., Kyosti, P., and Zhao, X. Comparison of SCM, SCME, and WINNER channel models. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th* (2007), Ieee, pp. 413–417.

[80] Niu, K., Huang, C., Zhang, S., and Chen, J. Oddc: Outlier detection using distance distribution clustering. In *Emerging Technologies in Knowledge Discovery and Data Mining*, T. Washio, Z.-H. Zhou, J. Huang, X. Hu, J. Li, C. Xie, J. He, D. Zou, K.-C. Li, and M. Freire, Eds., vol. 4819 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009, pp. 332–343.

[81] Odom, M., and Sharda, R. A neural network model for bankruptcy prediction. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on* (June 1990), pp. 163–168 vol.2.

[82] Oien, G., Holm, H., and Hole, K. J. Impact of channel prediction on adaptive coded modulation performance in rayleigh fading. *Vehicular Technology, IEEE Transactions on 53*, 3 (2004), 758–769.

[83] Okumura, Y., Ohmori, E., Kawano, T., and Fukuda, K. Field strength and its variability in vhf and uhf land-mobile radio service. *Rev. Elec. Commun. Lab 16*, 9 (1968), 825–73.

[84] Ott, K.-H., Aranibar, N., Singh, B., and Stockton, G. W. Metabonomics classifies pathways affected by bioactive compounds. artificial neural network classification of nmr spectra of plant extracts. *Phytochemistry 62*, 6 (2003), 971–985.

[85] Palleit, N., and Weber, T. Prediction of frequency selective simo channels. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on* (2011), pp. 1428–1432.

[86] Palleit, N., and Weber, T. Channel prediction in multiple antenna systems. In *Smart Antennas (WSA), 2012 International ITG Workshop on* (2012), pp. 1–7.

[87] Patcha, A., and Park, J.-M. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks 51*, 12 (2007), 3448 – 3470.

[88] Pätzold, M., and Hogstad, B. O. Classes of sum-of-sinusoids rayleigh fading channel simulators and their stationary and ergodic properties-part i. *WSEAS Transactions on Mathematics 5*, 2 (2006), 222.

[89] Paulraj, A., Roy, R., and Kailath, T. Estimation of signal parameters via rotational invariance techniques- esprit. In *Circuits, Systems and Computers, 1985. Nineteeth Asilomar Conference on* (nov. 1985), pp. 83 –89.

[90] Prokhorov, D. Echo state networks: Appeal and challenges. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on* (2005), vol. 3, IEEE, pp. 1463–1466.

[91] Rajkumar, T., and Bardina, J. Training data requirement for a neural network to predict aerodynamic coefficients. In *Proceedings of 17 th Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls. Orlando, EUA* (2003), Citeseer.

[92] Ramya, T. R., and Bhashyam, S. On using channel prediction in adaptive beamforming systems. In *Communication Systems Software and Middleware, 2007. COMSWARE 2007. 2nd International Conference on* (2007), pp. 1–6.

[93] Reed, R. Pruning algorithms-a survey. *Neural Networks, IEEE Transactions on 4*, 5 (1993), 740–747.

[94] Rice, S. O. Mathematical analysis of random noise. *Bell Systems Tech. J., Volume 23, p. 282-332 23* (1944), 282–332.

[95] Robinson, A., and Fallside, F. *The utility driven dynamic error propagation network*. University of Cambridge Department of Engineering, 1987.

[96] Rochester, N., Holland, J., Haibt, L., and Duda, W. Tests on a cell assembly theory of the action of the brain, using a large digital computer. *Information Theory, IRE Transactions on 2*, 3 (1956), 80–93.

[97] Rodan, A., and Tino, P. Minimum complexity echo state network. *IEEE Transactions on Neural Networks 22*, 1 (2011), 131–144.

[98] Rogova, G. Combining the results of several neural network classifiers. *Neural networks 7*, 5 (1994), 777–781.

[99] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review 65*, 6 (1958), 386.

[100] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. Tech. rep., DTIC Document, 1985.

[101] Rumney, M., et al. *LTE and the evolution to 4G wireless: Design and measurement challenges*. John Wiley & Sons, 2013.

[102] Salo, J., Nur-Alam, M., and Chang, K. Practical introduction to lte radio planning. *A white paper on basics of radio planning for 3GPP LTE in interference limited and coverage limited scenarios, European Communications Engineering (ECE) Ltd, Espoo, Finland* (2010).

[103] Schmidt, R. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on 34*, 3 (mar 1986), 276 – 280.

[104] Schrauwen, B., D'Haene, M., Verstraeten, D., and Campenhout, J. Compact hardware liquid state machines on FPGA for real-time speech recognition. *Neural Networks 21*, 2-3 (2008), 511–523.

[105] Semmelrodt, S., and Kattenbach, R. Investigation of different fading forecast schemes for flat fading radio channels. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th* (oct. 2003), vol. 1, pp. 149 – 153 Vol.1.

[106] Seth, S., Ozturk, M. C., and Principe, J. C. Signal processing with echo state networks in the complex domain. In *Machine Learning for Signal Processing, 2007 IEEE Workshop on* (2007), IEEE, pp. 408–412.

[107] Shannon, C. E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review 5*, 1 (2001), 3–55.

[108] Shimazaki, H., and Shinomoto, S. A method for selecting the bin size of a time histogram. *Neural Computation 19*, 6 (2007), 1503–1527.

[109] Simon, M. K., and Alouini, M.-S. *Digital communication over fading channels*, vol. 95. Wiley. com, 2005.

[110] Tam, K. Y., and Kiang, M. Y. Managerial applications of neural networks: the case of bank failure predictions. *Management science 38*, 7 (1992), 926–947.

[111] Tseng, S.-M., Zheng, Y., Hsu, Y.-T., and chou Chang, M. Fuzzy adaptive parallel interference cancellation and vector channel prediction for cdma in fading channels. In *Communications, 2002. ICC 2002. IEEE International Conference on* (2002), vol. 1, pp. 252–256.

[112] Various. WINNER II Channel Models. Tech. rep., Winner, 2007. Also available at http://www.ist-winner.org/deliverables.html.

[113] Various. Artificial neural network, 2014.

[114] Venayagamoorthy, G. K. Online design of an echo state network based wide area monitor for a multimachine power system. *Neural Networks 20*, 3 (2007), 404–413.

[115] Wang, S. The unpredictability of standard back propagation neural networks in classification applications. *Management Science 41*, 3 (1995), 555–559.

[116] White, H. Economic prediction using neural networks: The case of ibm daily stock returns. In *Neural Networks, 1988., IEEE International Conference on* (1988), IEEE, pp. 451–458.

[117] Win, M., and Scholtz, R. Characterization of ultra-wide bandwidth wireless indoor channels: a communication-theoretic view. *Selected Areas in Communications, IEEE Journal on 20*, 9 (dec 2002), 1613 – 1627.

[118] Wong, I., and Evans, B. Sinusoidal modeling and adaptive channel prediction in mobile ofdm systems. *Signal Processing, IEEE Transactions on 56*, 4 (2008), 1601–1615.

[119] Wong, I. C., and Evans, B. L. Wlc43-5: Low-complexity adaptive high-resolution channel prediction for ofdm systems. In *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE* (2006), IEEE, pp. 1–5.

[120] Wu, T.-M., and Tzeng, S.-Y. Sum-of-sinusoids-based simulator for nakagami-m fading channels. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th* (2003), vol. 1, IEEE, pp. 158–162.

[121] Xia, Y., Jelfs, B., Van Hulle, M., Principe, J., and Mandic, D. An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals. *Neural Networks, IEEE Transactions on 22*, 1 (2011), 74–83.

[122] Xiao, C., Zheng, Y. R., and Beaulieu, N. C. Novel sum-of-sinusoids simulation models for rayleigh and rician fading channels. *Wireless Communications, IEEE Transactions on 5*, 12 (2006), 3667–3679.

[123] Xie, M., Han, S., Tian, B., and Parvin, S. Anomaly detection in wireless sensor networks: A survey. *Journal of Network and Computer Applications 34*, 4 (2011), 1302 – 1325. ¡ce:title¿Advanced Topics in Cloud Computing¡/ce:title¿.

[124] Xing, X., Jing, T., Cheng, W., Huo, Y., and Cheng, X. Spectrum prediction in cognitive radio networks. *Wireless Communications, IEEE 20*, 2 (2013), 90–96.

[125] Yang, Y., Harley, R. G., Divan, D., and Habetler, T. G. Overhead conductor thermal dynamics identification by using echo state networks. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on* (2009), IEEE, pp. 3436–3443.

[126] Zemen, T., Mecklenbräuker, C., and Fleury, B. Time-variant channel prediction using time-concentrated and band-limited sequences. In *IEEE International Conference on Communications, 2006. ICC '06.* (june 2006), vol. 12, pp. 5660 –5665.

[127] Zhang, G., Eddy Patuwo, B., and Y Hu, M. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting 14*, 1 (1998), 35–62.

[128] Zhang, G. P. Neural networks for classification: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 30*, 4 (2000), 451–462.

[129] Zhang, Y.-Y., Yang, W.-C., Kim, K.-B., and Park, M.-S. Inside attacker detection in hierarchical wireless sensor network. In *Innovative Computing Information and Control, 2008. ICICIC '08. 3rd International Conference on* (2008), pp. 594–594.

[130] Zhang, Z., Li, J., Manikopoulos, C., Jorgenson, J., and Ucles, J. Hide: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification. In *Proc. IEEE Workshop on Information Assurance and Security* (2001), pp. 85–90.

[131] Zhang, Z., and Manikopoulos, C. Neural networks in statistical anomaly intrusion detection. *Neural Network World 11*, 3 (2001), 305–316.