Application of Non-linear Optimization Techniques in Wireless Telecommunication Systems

by

Farzaneh Kohandani

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Electrical & Computer Engineering

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Non-linear programming has been extensively used in wireless telecommunication systems design. An important criterion in optimization is the minimization of mean square error. This thesis examines two applications: peak to average power ratio (PAPR) reduction in orthogonal frequency division multiplexing (OFDM) systems and wireless airtime traffic estimation. These two applications are both of interests to wireless service providers. PAPR reduction is implemented in the handheld devices and low complexity is a major objective. On the other hand, exact traffic prediction can save a huge cost for wireless service providers by better resource management through off-line operations.

High PAPR is one of the major disadvantages of OFDM system which is resulted from large envelope fluctuation of the signal. Our proposed technique to reduce the PAPR is based on constellation shaping that starts with a larger constellation of points, and then the points with higher energy are removed. The constellation shaping algorithm is combined with peak reduction, with extra flexibilities defined to reduce the signal peak. This method, called MMSE-Threshold, has a significant improvement in PAPR reduction with low computational complexity.

The peak reduction formulated into a quadratic minimization problem

is subsequently optimized by the semidefinite programming algorithm, and the simulation results show that the PAPR of semidefinite programming algorithm (SDPA) has noticeable improvement over MMSE-Threshold while SDPA has higher complexity. Results are also presented for the PAPR minimization by applying optimization techniques such as hill climbing and simulated annealing. The simulation results indicate that for a small number of sub-carriers, both hill climbing and simulated annealing result in a significant improvement in PAPR reduction, while their degree of complexity can be very large.

The second application of non-linear optimization is in airtime data traffic estimation. This is a crucial problem in many organizations and plays a significant role in resource management of the company. Even a small improvement in the data prediction can save a huge cost for the organization. Our proposed method is based on the definition of extra parameters for the basic structural model. In the proposed technique, a novel search method that combines the maximum likelihood estimation with mean absolute percentage error of the estimated data is presented. Simulated results indicate a substantial improvement in the proposed technique over that of the basic structural model and seasonal autoregressive integrated moving average (SARIMA) package. In addition, this model is capable of updating the parameters when new data become available.

Acknowledgment

I would like to thank my supervisor, Professor Amir K. Khandani, for all his support and guidance throughout the course of my graduate studies. I would also like to express my gratitude to my committee members for their reviews of this thesis.

I express my appreciation to the organizations that funded this research: Communications and Information Technology of Ontario (CITO), National Sciences and Engineering Research Council of Canada (NSERC), Nortel Networks, Ontario Center of Excellence (OCE), Ontario Graduate Scholarship in Science and Technology (OGSST), Canadian Wireless Telecommunication Association (CWTA), Datatel Foundation, and University of Waterloo.

I am thankful to my friends at the CST lab and outside the CST lab for their friendship, support, and help.

I would like to thank my family, my beloved parents, and beloved sisters for their endless love and support, without which I would not have succeeded.

Last but not least, I would like to thank my best friend and the love of my life, Jan Bakus, who has been my rock the past few years. I dedicate my thesis to him and my family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Non-linear and quadratic optimization techniques have always been important research problems in wireless communications. One criterion in optimization of wireless communication systems is minimum mean square error (MMSE). This thesis deals with two applications of non-linear optimization techniques with minimum mean square error criterion. We examine two applications: peak to average power ratio (PAPR) reduction in orthogonal frequency division multiplexing (OFDM) systems and wireless airtime traffic estimation. These two applications are both of interests to wireless service providers. PAPR reduction is implemented in the handheld devices and low complexity is a major objective. On the other hand, exact traffic prediction can save a huge cost for wireless service providers by better resource man-

agement through off-line operations. First, we look at peak to average power ratio (PAPR) reduction of an orthogonal frequency division multiplexing (OFDM) system.

OFDM used in fourth-generation wireless technology is a multi-carrier multiplexing technique. Recent advances in digital signal processing (DSP) and very large scale integrated circuit (VLSI) technologies have allowed OFDM to be implemented in many standards [94]. OFDM technology allows many users to transmit in an allocated band by sub-dividing the available bandwidth into many narrow bandwidth carriers [54]. The narrow bandwidth carriers result in the signal having a high tolerance to multi-path delay spread, because the delay spread must be very long to cause significant inter-symbol interference. The major disadvantage of OFDM technology is when all the signal peaks happen at the same time which results in a signal with a large peak. This problem is called peak to average power ratio (PAPR). A lot of research is devoted to PAPR reduction techniques categorized as clipping [83], scrambling [9], coding [79], unused spectrum cancellation [17], nonbijective constellation [88], tone reservation and tone injection [90]. An overview of most important PAPR reduction techniques is given in [23].

The PAPR proposed techniques are based on constellation shaping algorithm with lower average energy where sequences with higher average energy are pruned. The consequence of the pruning is that fewer sequences are

available for the encoding, resulting in a reduced data rate for the system and a $1dB$ shaping gain [44]. The peak reduction is performed by the definition of extra bits in different sub-spaces to reduce the signal peak. A novel low complexity minimization technique called MMSE-Threshold is proposed for PAPR reduction in Chapter 2. In addition, the peak reduction formulated into a quadratic minimization problem is subsequently optimized by the semidefinite programming algorithm [74].

The second application of non-linear optimization with MMSE criterion is the airtime data traffic estimation, a crucial problem for wireless telecommunication services companies. By the exact estimation of airtime traffic, these companies can manage their resources more effectively and reduce costs. Later, these companies have the opportunity to sell their extra bandwidths to other companies in daily basis. To generate predictions, information about past events, called "time series data", is collected. A lot of research is performed in data traffic prediction categorized as naive [24], moving average [18], exponential smoothing [87], autoregressive integrated moving average [100] and state space models [57].

The proposed estimation technique called extended structural model (ESM) is based on basic structural model presented in Chapter 3. ESM includes a new optimization algorithm where maximum likelihood estimation is combined with minimum mean absolute percentage error (MMAPE), to estimate

the model parameter over the validation period of the data. This combination prevents a large degree of complexity of the optimization algorithm in maximum likelihood estimation and provides the opportunity to update the model parameters whenever new data become available.

1.1 Summary of the Dissertation

The organization of this thesis is as follows. After the introduction, PAPR reduction in OFDM systems is defined in Chapter 2. The OFDM system, used in fourth-generation standards, deals with the major problem of large envelope fluctuations that can be described with PAPR [90]. Our proposed method is based on the constellation shaping algorithm, which starts with a larger constellation of points and the points with higher average energy are removed [44]. We have combined this selection of constellation points with the peak reduction algorithm by definition of dummy bits to lower the peak energy as well as the average energy. The peak reduction technique leads to binary quadratic optimization and proposal of a low complexity method called MMSE-Threshold.

We have also applied the semidefinite programming algorithm (SDPA) to the defined quadratic problem where its optimization package is available in [15]. These techniques are compared with hill climbing [62] and simulated

annealing algorithms [1], which are directly applied to minimize the PAPR reduction. The simulation results represent that the MMSE-Threshold technique offers a considerable improvement in PAPR reduction, which in comparison with SDPA has much less complexity. Also compared with state-of-the-art techniques, MMSE-Threshold has better or comparable PAPR with low complexity and it has $1dB$ reduction in average energy.

Chapter 3 defines another non-linear optimization problem with maximum likelihood estimation criterion. Airtime traffic estimation is a very important issue for telecommunication service providers and works with maximum likelihood estimation [3]. Over-estimation of future traffic can result in wasteful usage of resources while under-estimation negatively impacts service and causes customer dissatisfaction. This chapter presents a novel model: an extension of the basic structural model (BSM) [35]. The new model along with a novel search technique that combines maximum likelihood estimation with mean absolute percentage error (MAPE) is presented. A simple optimization technique (steepest descent) is used for optimization.

For comparison, a seasonal autoregressive integrated moving average package [95] is used to estimate airtime traffic. The simulation results for traffic estimation represent that our proposed extended structural model (ESM) offers improved performance over BSM and seasonal autoregressive integrated moving average (SARIMA) techniques while it is capable of updating the pa-

rameters, unlike the SARIMA method. Chapter 4 presents the contribution of the thesis along with future work.

Chapter 2

PAPR Reduction in OFDM Systems

2.1 Introduction

High data rate services are becoming popular. Physical media capable of supporting broadband data transmission—including both wired (digital subscriber lines, cable modems, power lines) and wireless media—will be used for transmission. Usually, high data rate services require very reliable data transmission, but most of these transmission systems experience many degradations, such as large attenuation, noise, multi-path, interference, time variation and non-linearities. However, these systems must meet many con-

straints, such as limited transmit power and, most importantly, low cost. One popular physical-layer technique that can deal with these deficiencies is multi-carrier modulation or orthogonal frequency division multiplexing (OFDM).

A major problem with multi-carrier signals is large envelope fluctuation. This is usually measured by a parameter called peak to average power ratio (PAPR). Since power amplifiers are peak-power limited to work in linear regions, we need to operate the system at power levels well below the maximum power available. In practice, occasional saturation of the power amplifiers or clipping in the digital-to-analog-converters is allowed. This non-linear distortion creates inter-modulation distortion that increases the bit error rate in standard linear receivers, and it also causes spectral widening of the transmitted signal, which increases adjacent channel interference to other users [90].

High PAPR is also a major disadvantage of code division multiple access (CDMA) systems. Since both OFDM and CDMA deal with high PAPR, we compare OFDM technology with the CDMA technology, to emphasize the advantages of OFDM over CDMA and why OFDM technology is suggested for fourth-generation systems. That is why in this thesis we have just focused on PAPR reduction in OFDM systems.

2.2 OFDM and CDMA Technologies and Their Comparison

Code division multiple access (CDMA) is a spread spectrum technique that uses orthogonal or non-orthogonal spreading codes to multiplex users. For example, Walsh codes, which are orthogonal and based on the Walsh matrix, are one of the spreading codes in CDMA. With a wide bandwidth signal or spread spectrum, any dips in the spectrum result only in a small loss of signal power. CDMA is used in both second-generation (2G) and third- generation (3G) mobile networks. 2G CDMA standards are branded CDMA-one, including interim standard 95, version A (IS-95A) and interim standard 95, version B (IS-95B). CDMA is the foundation for 3G services as well. The two dominant international mobile telecommunications-2000 (IMT-2000) standards are CDMA2000 and wide-band CDMA (WCDMA) [21].

In OFDM, with the spectrum split into many small bandwidth carriers, fading will affect only a small number of such carriers. OFDM has been implemented in many standards, such as digital audio broadcasting (DAB), digital video broadcasting (DVB), IEEE 802.11a wireless local area network (WLAN), IEEE 802.16a wireless metropolitan area network (WMAN), and asymmetric digital subscriber line (ADSL).

References [28, 54, 63] compare the performance of CDMA and OFDM technologies. [28] studies the performance of direct sequence CDMA (DS-CDMA), multi-carrier CDMA (MC-CDMA), and OFDM. De-spreading the received signal at the receiver is done through the minimum mean square error (MMSE) approach. [54] studies the suitability of OFDM in comparison with CDMA as a multiplexing technique for a fixed wireless phone system for rural areas of Australia. Several of the main factors affecting the performance of an OFDM system are considered, including multi-path delay spread, channel noise, distortion (clipping), and timing requirements. [63] studies the performance comparison between CDMA and OFDM systems in selective fading channels. In addition, two existing standards, universal mobile telecommunications system (UMTS) terrestrial radio access frequency division duplex (UTRA-FDD) and DVB, of the two systems are compared with the comparison focused on the downlink.

These studies show that in a sharp selective fading channel, OFDM outperforms DS-CDMA and MC-CDMA, and the capacity of an OFDM system for a single cell is more than that of a CDMA system. In a multiple cell system, OFDM has almost the same or more capacity than CDMA. It is found that OFDM performs better than CDMA, providing a very high tolerance to multi-path delay spread, peak power clipping, and channel noise. In addition, it provides high spectral efficiency. The CDMA system with a

RAKE-receiver exhibits an inferior bit error rate (BER) performance compared to an OFDM system. However, it has not been proven that OFDM systems are superior to CDMA systems in every aspect. The implementation of orthogonal frequency division multiple access (OFDMA) can lead to a reduction of the bandwidth efficiency of OFDM (especially in up-link). In addition, ignored are the problem of frequency offset, which is the weakness of OFDM, and the multiple access interference (MAI) problem, caused by the RAKE-receiver.

There has been some interest in the combination of these two technologies, named OFDM-CDMA technology. OFDM-CDMA is a method in which a single data symbol is replicated into N parallel copies. Each branch of the parallel stream is multiplied by a code of length N and then OFDM modulation is applied. To overcome the effect that deep fades significantly degrade system performance, OFDM can be combined with the direct sequence spread spectrum technique, through which the signal is spread over various sub-carriers to achieve frequency diversity.

References [5, 42, 68, 67, 58, 10, 40] study this technology. [5] studies the applicability of OFDM combined with CDMA in an indoor environment, with particular emphasis on channel equalization techniques. [42] studies the performance evaluation of an OFDM-CDMA system in the frequency domain using autoregressive (AR) modeling. [68] studies the performance of

an OFDM-CDMA system in a slow Nakagami fading frequency non-selective channel. [67] studies the effect of non-perfect power amplification on OFDM-CDMA systems. [58] studies a scheme that combines the advantages of OFDMA and CDMA to cope simultaneously with both frequency spread and time spread. [10] studies the analysis of the performance of a convolutionally coded CDMA system combined with OFDM in a frequency/time selective fading channel, taking into account the Near-Far problem. [40] studies the performance of OFDM-CDMA systems combined with convolutional and Turbo channel coding.

The literature shows that OFDM-DS-CDMA performs well in the downlink, where each subscriber receives its own useful signal and the interference signals from all the other users through the same channel. The equalization based on channel inversion preserves the orthogonality characteristics of the spreading codes, thus permitting the detection of the useful signal by means of a single branch CDMA receiver. In general, OFDM-CDMA systems perform better than OFDM systems with non-ideal power amplification, by using a simple code word assignment scheme.

2.3 OFDM Systems

Orthogonal frequency division multiplexing (OFDM) is a multi-carrier multiplexing technique. Recent advances in digital signal processing (DSP) and very large scale integrated circuit (VLSI) technologies have allowed OFDM to be implemented in many standards [94]. One recent successful implementation of OFDM is in digital audio broadcasting (DAB), which was developed in Europe for terrestrial and satellite broadcasting of multiple digital audio programs to mobile receivers. Another recent implementation is in asymmetric digital subscriber line (ADSL) technology, which has been selected by American national standard institute (ANSI) for transmission of digitally compressed video signals over telephone lines. A pan-European project digital video broadcasting (DVB) was launched in 1993 and developed the digital video broadcasting-terrestrial (DVB-T) system, a terrestrial distribution system with data throughput up to 24 Mbps over an 8 MHz channel using OFDM.

OFDM technology allows many users to transmit in an allocated band by sub-dividing the available bandwidth into many narrow bandwidth carriers [54]. Each user is allocated several carriers in which to transmit their data. The transmission is generated so that the carriers used are orthogonal to one another, thus allowing them to be packed together much more closely than in

standard frequency division multiplexing (FDM). The result is high spectral efficiency. Each carrier in an OFDM signal has a very narrow bandwidth, so the resulting symbol rate is low, resulting in the signal having a high tolerance to multi-path delay spread, because the delay spread must be very long to cause significant inter-symbol interference.

2.3.1 Basic Principles of OFDM

The concept of OFDM is to transmit data in parallel quadrature amplitude modulation (QAM) modulated sub-carriers using frequency division multiplexing. The carrier spacing is selected so that each sub-carrier is located on all the other sub-carriers' spectra zero crossing points. Although there are spectral overlaps among sub-carriers, they do not interfere with each other if they are sampled at the sub-carrier frequencies. In other words, they have orthogonality in the frequency domain. The mathematical expression of an OFDM signal is

$$x(t) = \sum_{n=0}^{N-1} (a_n \cos \omega_n t + b_n \sin \omega_n t), \tag{2.1}$$

where a_n and b_n are the in-phase and quadrature terms of the QAM signal, and ω_n is the sub-carrier frequency. By observing (2.1), one notices that the OFDM signal is actually the real part of the inverse discrete Fourier transform (IDFT) of the original data $d_n = a_n + jb_n$, i.e.,

$$\begin{aligned} x(t) &= Re\sum_{n=0}^{N-1}(a_n + jb_n)exp(-j\omega_n t) \\ &= Re\{\sum_{n=0}^{N-1}(a_n + jb_n)(\cos\omega_n t - j\sin\omega_n t)\} \\ &= \sum_{n=0}^{N-1}(a_n\cos\omega_n t + b_n\sin\omega_n t), \end{aligned} \tag{2.2}$$

where $\omega_n = 2\pi n/(N\Delta t), t = l\Delta t$, l is the number of OFDM symbol in t and Δt is the symbol duration of the input serial data d_n.

2.3.2 Guard Interval and Its Implementation

Linear distortions such as multi-path delay and micro-reflection cause inter symbol interference (ISI) among OFDM symbols, resulting in loss of orthogonality and an effect similar to co-channel interference.

A simple solution to multi-path delay is to increase the OFDM symbol duration so that it is much larger than that of the delay spread. However, this solution may be difficult to implement because a large delay spread requires a large number of sub-carriers and a large FFT.

Another method to combat multi-path distortion is to have a cyclically extended guard interval, with each OFDM symbol preceded by a periodic extension of the signal itself. The total symbol duration is $T_{total} = T_g + T_s$, where T_g is the guard interval and T_s is the useful symbol duration. Since

the guard insertion will reduce data throughput, T_g is usually less than $T_s/4$. When the guard interval is longer than the channel impulse response or the multi-path delay, ISI can be eliminated.

2.4 PAPR Definition

The peak of a signal $x(t)$ is given by the maximum of its envelope $|x(t)|$. However, for a continuous random process, $\max|x(t)|$ can reach infinity provided that the observation interval is long enough. Even in a discrete random process where $\max|x(t)|$ is bounded, the maximum may occur at a very low probability. Therefore, a more useful definition of peak is in probability terms given by

Definition 1. A signal $x(t)$ is said to have a peak x_p at cut-off probability P_c if

$$\Pr[|x(t)| < x_p] = P_c. \tag{2.3}$$

Therefore, the peak to average ratio (PAR) of a random process $x(t)$ can be specified by its histogram. The PAR definition in (2.4) refers to the probability density function (PDF) generated from its time samples (i.e. collect R samples of $x(t)$, plot its histogram, and the PDF is obtained as R

tends to infinity).

PAR usually refers to a discrete-time measurement using $\{x_g\}$ which is equal to $x(gT_{total})$. Given an N-dimensional sample signal x of an OFDM system, where N is the number of sub-carriers, the PAR of OFDM is defined as,

$$PAR(x(t)) = \frac{\max_g |x_g(t)|^2}{E_x\left[\frac{1}{N}|x(t)|^2\right]} \ . \tag{2.4}$$

Similarly, the peak to average power ratio (PAPR) is usually defined on the continuous time signal $x(t)$ as,

$$PAPR(x(t)) = \frac{\max_t |x(t)|^2}{E_x\left[\frac{1}{NT_{total}}\int_0^{NT_{total}} |x(t)|^2 dt\right]} \ . \tag{2.5}$$

In general, $PAPR(x(t)) \geq PAR(x(t))$ and therefore evaluating performance in the discrete-time domain may lead to optimistic values. Peak-power limitations are usually placed at the power amplifier, which limits the continuous time signals; therefore, PAPR is generally the more relevant metric in practice [53].

2.5 Related Work

Much research has been devoted to the analysis of methods for reducing PAPR. The following sections briefly explain the various techniques and their

potential disadvantages. [23] is an overview of most important PAPR reduction techniques. All existing solutions involve some form of compromise. The most obvious one is the trade-off between bandwidth and peak power. Obviously, if we reduce the data rate, we may reduce the system peak power even without explicit peak reduction mechanisms. Most coding and some phase optimization schemes have these problems. Unused spectrum cancellation is also an implicit form of the trade-off between bandwidth and peak power. If usable bandwidth is sacrificed for peak reduction, we are effectively trading the data rate for lower system peak power.

Clipping

The simplest way to deal with PAPR is to clip the output signal. Many researchers have analyzed this method [83], [29]. With this method, the clipped signal will take higher-order harmonics that can end up in the out-of-band spectrum. There is usually a strict limit on the out-of-band emission. Therefore, these schemes may cause more problems than they solve. Peak windowing schemes [75], [64] try to limit the out-of-band distrotion through narrow-band windows. For example, a Gaussian window may be used to attenuate a large peak. It offers reasonably good peak reduction at the cost of increasing BER and smaller out-of-band distortion. Repeated clipping and frequency domain filtering [2], [92] can reduce the PAPR without increasing

the out-of-band power by clipping the over-sampled time domain signal followed by filtering using an FFT-based, frequency domain filter designed to reject out-of-band discrete frequency components. Filtering results in peak regrowth. The distortion of the in-band signal results in shrinking of the overall signal constellation and an added noise-like effect.

Scrambling

Scrambling is a technique to reduce the PAPR that can be applied on the pattern of the input bits or sub-carriers. Correlated bit patterns generate constellation points with high magnitude in time. Therefore, scrambling the input bit streams may reduce the probability of large peaks generated by those bit patterns. Some designers [9] selectively use the output of one of four scramblers. The modulator selects the scrambled sequence that produces the smallest peak. However, an error in the bits that encode the choice of scrambling sequence may lead to long propagation of decoding errors. Scrambling can also be done on the pattern of the sub-carriers to reduce the PAPR [7]. Introducing a disturbance in the correlation among the sub-carriers reduces PAPR. PAPR can be further reduced with the fixed scrambling pattern that is separately optimized for each user code and the scrambling pattern held while the same code combination is maintained.

Coding

Coding schemes [79], [36], [80], [56] and [73] use well-known block codes to reduce the peak power of OFDM systems with signal constellations, such as binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), and m-array phase shift keying (MPSK). With block code, these schemes can remove some constellation combinations with large peaks in the time domain, a trade-off between data rate and peak power.

Phase Optimization

Some researchers, [13], [71], [14] and [30], have observed that by rotating the channel constellations properly, the peaks can be reduced. Muller and Huber [71] have introduced the partial transmit sequence (PTS) scheme, in which the N-dimensional constellation (where N is the number of sub-carriers) is divided into several blocks and the phases between blocks are optimized. Friese [14] uses differential modulation across channels. Ho [30] presents a dual-layered phase sequencing approach in partial transmit sequences (PTS) to reduce the PAPR with lower complexity. The layers are classified as micro and macro optimization layers. In the micro optimization layer, the C_1 sub-blocks are grouped into D_1 divisions. On the second layer, each individually optimized division is considered as a block for further PAPR reduction. Most

researchers assume that the phase-rotation information is transmitted separately via a reliable channel. For OFDM, a separate channel or additional error-correcting codes are needed for that purpose. Either option reduces the overall data rate of the system. Thus, it is once again a trade-off of bandwidth vs. peak power.

Unused Spectrum Cancellation

In discrete multi-tone (DMT), a few high-frequency channels with SNRs usually are too low to transmit any significant number of bits. This unused spectrum may be used to transmit signals that can cancel out the signal peaks. Gatherer and Polley [17] introduced this technique, which was later improved in [89]. They are reasonably effective for DMT systems in the discrete-time domain. Furthermore, for OFDM, the channels that carry the data will be sacrificed for cancellation. Thus, for OFDM, these schemes trade bandwidth for peak power.

Nonbijective Constellation

In these techniques, instead of each distinct bit pattern being mapped to an N-dimensional constellation point, it is mapped to a set of constellation points. Then, the constellation point with the lowest peak is chosen from

the selected class; these methods are called selected mapping. Tellado and Cioffi [88] extend the QAM constellation in each channel periodically and partition the expanded constellation into cosets. The input bit pattern selects a coset for each channel. Then, optimization is done to select the point within each coset that minimizes the peak. Jones [37] extends the corner point of the QAM constellation into a continuous region. [55] presents a new selected mapping technique for an OFDM system with lower computational complexity compared to a conventional selected mapping technique. It transforms an input symbol sequence into a set of OFDM signals by multiplying the phase sequences to the signal (not at the output of IFFT but after a certain stage of IFFT), and then the OFDM signal with the lowest PAPR is selected. These algorithms do not sacrifice bandwidth for data rate; however, they may violate the power constraint on the channel.

Tone Reservation and Tone Injection

In tone reservation methods, the basic idea is to reserve a small set of tones for PAPR reduction [90], [52] and [81]. The problem of minimizing the PAPR can be formulated as a convex problem and can be solved exactly. The amount of PAPR reduction depends on the number of reserved tones, their locations within the frequency band, and the amount of complexity. This method describes an additive method for reducing PAPR in multi-carrier

transmission and shows that reserving a small fraction of tones leads to large reductions in PAPR even with simple $O(N)$ algorithms at the transmitter, and with no additional complexity at the receiver, where N is the number of sub-carriers. When N is small, the set of tones reserved for PAPR reduction may represent a non-negligible fraction of the available bandwidth and can result in a reduction of data rate, thus motivating the use of PAPR reduction methods with no rate loss.

In tone injection methods, the constellation size is increased so that each point in the original constellation can be mapped to several equivalent points. These extra degrees of freedom can be used towards PAPR reduction in [90] and [31].

2.6 Proposed PAPR Reduction Method

In conventional shaping, one tries to minimize the average energy of the constellation for a given number of points from a given packing. The price to be paid for shaping involves: (i) an increase in the constellation-expansion-ratio (CER), (ii) an increase in the peak to average power ratio (PAPR), and (iii) an increase in the addressing complexity, which is the assignment of the data bits to the constellation points.

In shaping, one starts with a number of points greater than what is re-

quired for a specific bit rate, with the objective of providing some degree of flexibility in the selection of the final constellation. In traditional shaping, this flexibility is used to select the points of the least average energy, but it could also be used to select the points with a low average energy and at the same time result in a small value for the peak power along the time dimensions. This result can be achieved with a signal constellation that has several choices of points (in a multi-dimensional space) available for a given input bit label. In this case, the transmitter side will select the constellation point that results in a small average energy and at the same time has a small peak power (among the possible choices corresponding to the given binary input) as shown in Fig. 2.1. We mix the objectives of reducing the peak and the average energy in the selection of the constellation. This requires finding a proper cost function to incorporate the combined effects of the peak and the average energy, while allowing for an efficient search procedure [44].

2.6.1 Definition of the System for the Average Energy Reduction

Shaping is a method for reducing the average energy required to transmit data relative to the average energy required for an unshaped (cube) constellation while maintaining the minimum distance between constellation points.

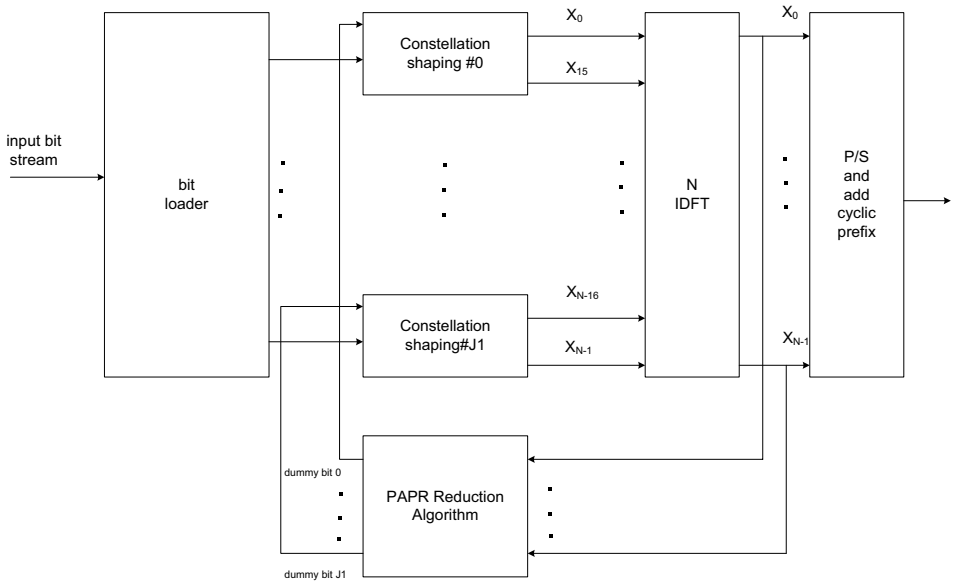

Figure 2.1: OFDM transmitter structure with constellation shaping and PAPR reduction algorithm

This reduction in energy is measured by shaping gain, which is achieved with a larger constellation size compared to an unshaped constellation where the increase in constellation size is given by the constellation expansion ratio (CER) as shown in Fig. 2.2. CER is defined as the ratio of the number of points per 2-D of a shaped constellation to the minimum required number of points per 2-D to achieve the same overall rate in an unshaped constellation [16].

In an OFDM system, the input bit stream is used to generate an address,

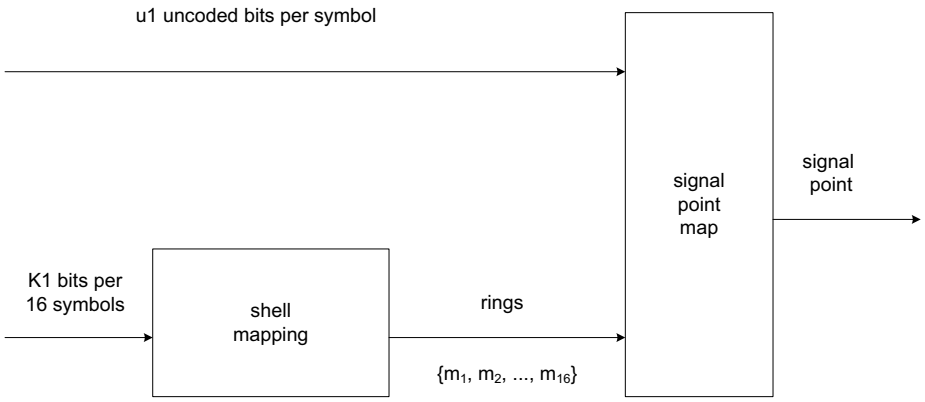

Figure 2.2: Constellation shaping

which is then mapped by the quadrature amplitude modulation to N complex frequency points. Each of these points corresponds to a sub-carrier. With redundancy in the addressing scheme, the constellations of the modulated points can be shaped to reduce the overall signal energy. The modulated points are passed through the OFDM IFFT modulator, whose outputs, after parallel-to-serial conversion, represent N Nyquist rate time domain complex samples of the baseband OFDM waveform.

Using the combination of addressing and QAM modulation allows us to reduce the energy of the signal. Performing the constellation shaping for all N sub-carriers simultaneously is difficult, and therefore we divide the N sub-carriers into m sub-spaces each with k sub-carriers ($N = mk$), and

the constellation shaping is performed individually for each sub-space. We choose k equal to 32 because it has been shown that one can achieve most of the available shaping gain using 32 sub-carriers [43]. To handle systems with more sub-carriers, several of the sub-spaces are concatenated together.

Each sub-space is composed of k QAM points, and the objective of the constellation shaping is to remove the sequences with the highest energy. This removal is achieved as the sequences are grouped by energy, such that sequences in the same group have similar energy compared to sequences in other groups. Then, the groups with high energy sequences are pruned out until the desired number of sequences is reached [44]. The consequence of the pruning is that fewer sequences are available for the encoding, resulting in a reduced data rate for the system and a $1dB$ shaping gain [44]. The block diagram of the system is presented in Fig. 2.3.

2.6.2 Peak Reduction Algorithm with Low Complexity

One of the problems of the OFDM system is the disproportionate peaks compared to the signal average. We propose a technique that works in conjunction with constellation shaping to reduce the amplitude of these peaks. The N transmitted sub-carriers are composed of $m = N/k$ sub-spaces. In each sub-space, we designate one bit as a dummy bit that does not carry any

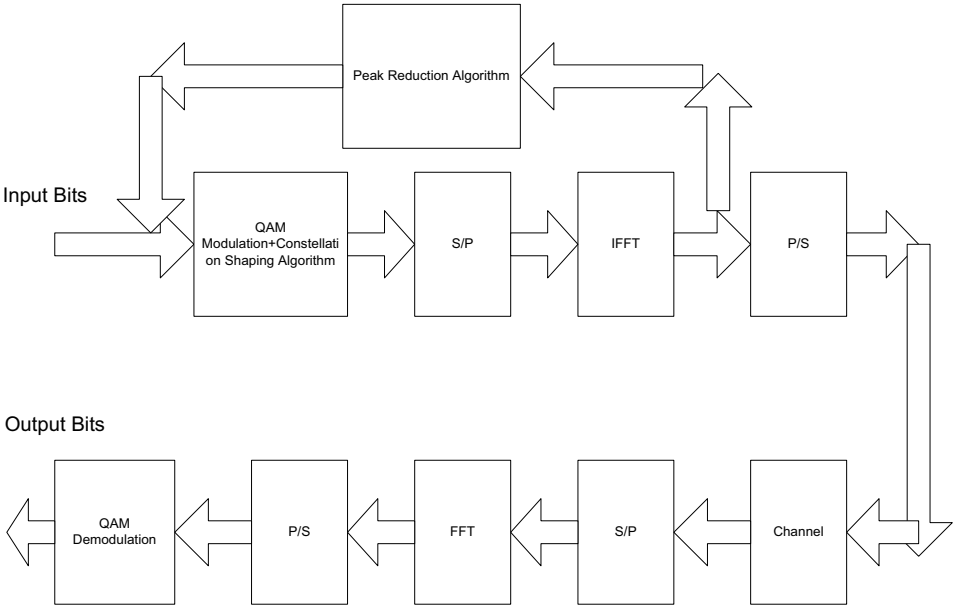

Figure 2.3: System block diagram

data. This gives us flexibility to select values of the dummy bits to reduce the signal peak in the time domain. Since these dummy bits are used in the shaping algorithm, they can have a significant effect on peak reduction. Some preliminary results are given in [48].

To generate the time signal, we first set all the dummy bits in all the sub-spaces to 0 and perform the addressing, modulation, and the IFFT for all N points. The result is a time domain base band vector of N complex points: $\mathbf{z} = z_1, \ldots, z_N$.

We define a complex clipping function as $\mathbf{y} = \text{clip}(\mathbf{z}, T)$, such that a point z_i in vector $\mathbf{z}$ will be mapped to point y_i in vector $\mathbf{y}$ according to

$$y_i = \begin{cases} T e^{j \arg(z_i)} & |z_i| \geq T \\ z_i & \text{otherwise} \end{cases} \tag{2.6}$$

We define the error vector $\mathbf{e}$ for a given threshold T as

$$\mathbf{e} = \mathbf{z} - \text{clip}(\mathbf{z}, T), \tag{2.7}$$

where $\mathbf{z}$ is the time domain signal vector for all the dummy bits set to 0.

The addressing operates on the individual sub-spaces, and therefore changing a dummy bit in one sub-space changes the frequency domain points in that sub-space with no effect on the points in the other sub-spaces. We can therefore construct a set of orthogonal vectors, with each vector correspond-

ing to one sub-space.

Since the IFFT is an orthonormal transform, the orthogonal vectors in the frequency domain result in new orthogonal vectors in the time domain. We can construct a vector $\boldsymbol{\epsilon}_i$ for each sub-space i. To create this vector, we set the dummy bit to 0 in the sub-space i and perform the addressing and QAM modulation for that sub-space. Then, we set the points in other sub-spaces to zero, resulting in a frequency domain vector $\bar{\mathbf{s}}_i$. Afterward, we set the dummy bit in sub-space i to 1 and zero in other sub-spaces and perform the same operation to obtain the frequency domain vector $\mathbf{s}_i$. The corresponding time domain vector $\boldsymbol{\epsilon}_i$ is defined as

$$\boldsymbol{\epsilon}_i = \text{IFFT}(\bar{\mathbf{s}}_i - \mathbf{s}_i). \tag{2.8}$$

Each vector $\boldsymbol{\epsilon}_i$ corresponds to the change of the time domain vector by changing the dummy bit in sub-space i. Using the vectors $\boldsymbol{\epsilon}_i$, we can define the transmitted time domain signal $\mathbf{w}$ for given dummy bit values $\alpha_1, \ldots, \alpha_m$ as

$$\mathbf{w} = \mathbf{z} - \sum_{i=1}^{m} \alpha_i \boldsymbol{\epsilon}_i, \tag{2.9}$$

and based on definition of $\mathbf{z}$, we have

$$\mathbf{z} = \text{IFFT}(\bar{\mathbf{s}}_i). \tag{2.10}$$

From (2.8) and (2.10), we can rewrite (2.9) in the following

$$\mathbf{w} = \text{IFFT}(\bar{\mathbf{s}}_i) - \sum_{i=1}^{m} \alpha_i \text{IFFT}(\bar{\mathbf{s}}_i - \mathbf{s}_i). \tag{2.11}$$

To show that $\mathbf{w}$ is the transmitted time domain signal for selected value of α_i, we transfer $\mathbf{w}$ to frequency domain called $\mathbf{w}_f$ as

$$\mathbf{w}_f = \bar{\mathbf{s}}_i - \sum_{i=1}^{m} \alpha_i (\bar{\mathbf{s}}_i - \mathbf{s}_i), \tag{2.12}$$

where

$$\begin{cases} \text{if} \;\; \alpha_i = 0, & \mathbf{w}_f(i) = \bar{\mathbf{s}}_i(i) \\ \text{if} \;\; \alpha_i = 1, & \mathbf{w}_f(i) = \mathbf{s}_i(i). \end{cases} \tag{2.13}$$

From equation (2.7) and (2.9), we can define the actual clipped error $\mathbf{e_{act}}$ as

$$\mathbf{e_{act}} = \mathbf{z} - \sum_{i=1}^{m} \alpha_i \boldsymbol{\epsilon}_i - \text{clip}\left(\mathbf{z} - \sum_{i=1}^{m} \alpha_i \boldsymbol{\epsilon}_i, T\right). \tag{2.14}$$

Minimizing the actual error defined in (2.14) is a complicated problem because of non-linear clip function, therefore we define a new minimization problem in (2.15). The performance of this minimization problem in terms of PAPR reduction and symbol error rate are evaluated by simulation and the results are presented in section simulation results.

$$\text{Minimize} \left|\left| \mathbf{e} - \sum_{i=1}^{m} \alpha_i \boldsymbol{\epsilon}_i \right|\right|^2. \tag{2.15}$$

Consider a system with only one sub-space i and the corresponding dummy bit value α_i in (2.15). The first case has $\alpha_i = 0$, where the objective function in (2.15) is $||\mathbf{e}||^2$. In the second case, the dummy bit is set to one, i.e. $\alpha_i = 1$, the objective function in (2.15) becomes $||\mathbf{e}||^2 + ||\boldsymbol{\epsilon}_i||^2 - 2\text{Real}(\boldsymbol{\epsilon}_i^H \cdot \mathbf{e})$. We can therefore define a decision function $\gamma(i)$ for each dummy bit α_i as:

$$\gamma(i) = ||\boldsymbol{\epsilon}_i||^2 - 2\text{Real}(\boldsymbol{\epsilon}_i^H \cdot \mathbf{e}), \tag{2.16}$$

where $\boldsymbol{\epsilon}_i^H$ is the hermitian of $\boldsymbol{\epsilon}_i$. We select $\alpha_i = 0$ for $\gamma(i) \geq 0$ and $\alpha_i = 1$, otherwise. We iterate over all the dummy bits, and use (2.16) to select the value of each dummy bit. Then, new error vector is calculated from (2.9) and (2.7), and new values of α_i are obtained. This is done iteratively until no further reduction is achieved in (2.15). This technique is called MMSE-Threshold and the peak reduction algorithm is summarized in Fig. 2.4.

The complexity of MMSE-Threshold algorithm is $O(N)$ scalar multiplications from (2.16) and the total complexity is the complexity of MMSE-Threshold algorithm plus the complexity of IFFT. Therefore the total complexity of MMSE-Threshold is $O(N \log N)$.

To evaluate the performance of MMSE-threshold which is based on (2.15), this equation is also minimized with one of the classical methods of quadratic programming called semidefinite programming [93]. Then to evaluate how good (2.15) estimates the PAPR reduction, the peak to average energy of $\mathbf{w}$

{*initialization*}

Select random values for all dummy bits and perform addressing and IFFT to get $\mathbf{z}$

Calculate $\mathbf{e}$ using (2.7)

Calculate all $\boldsymbol{\epsilon}_i$ using (2.8)

{*main loop*}

while

  for i from 1 to m **do**

    if $\gamma(i)$ in (2.16) ≥ 0 **then**

      $\alpha_i \leftarrow 0$

    else

      $\alpha_i \leftarrow 1$

    end if

  end do

  if not reached minimum PAPR

    Select another random set for all dummy bits to calculate $\mathbf{e}$

  else

    break

  end if

end while

{*end main loop*}

return $\{\alpha_1, \ldots, \alpha_m\}$

Figure 2.4: Peak error minimization algorithm for MMSE-Threshold

given in (2.9) is minimized in terms of α_i by two heuristic algorithms called hill climbing and simulated annealing [66, 72].

The minimization of the quadratic problem given in (2.15) is known as an unconstrained quadratic program (UQP) for binary variables. Many combinatorial optimization problems pertaining to graphs such as determining maximum cliques, maximum cuts, maximum vertex packing, minimum coverings, maximum independent sets, and maximum independent weighted sets are also formulated as the UQP problem [46].

There are many non-linear optimization techniques that can be utilized to solve the quadratic problem defined in (2.15). From the classical methods, we select semidefinite programming [93] for comparison with MMSE-Threshold [93] to minimize (2.15).

To describe semidefinite programming, first we define a standard quadratic problem as [59]

$$\begin{aligned} &\max \ \ \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ &\mathbf{x} \in \{-1,+1\}^{n_1} \end{aligned} \tag{2.17}$$

where $\mathbf{Q}$ can be any symmetric matrix. Since $\mathbf{x}^T\mathbf{Q}\mathbf{x} = \text{Trace}(\mathbf{x}^T\mathbf{Q}\mathbf{x}) = \text{Trace}(\mathbf{x}\mathbf{x}^T\mathbf{Q})$, (2.17) is equivalent to the following problem,

$$\begin{aligned} &\max \ \ \text{Trace}(\mathbf{XQ}) \\ &\text{s.t. } \mathbf{X} = \mathbf{x}\mathbf{x}^T, \quad \mathbf{x} \in R^{n_1} \\ &\mathbf{X}_{l_1 l_1} = 1, \ \ l_1 = 1, \ldots, n_1. \end{aligned} \tag{2.18}$$

The constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^T$ implies that $\mathbf{X}$ is symmetric, positive semidefinite and rank-1. Due to the constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^T$, (2.18) is a non-convex optimization problem. If we remove the rank-1 constraint, we obtain the following relaxed problem known as semidefinite programming (SDP) problem,

$$\begin{aligned} &\max \ \ \text{Trace}(\mathbf{XQ}) \\ &\text{s.t. } \mathbf{X} \geq 0 \\ &\mathbf{X}_{l_1 l_1} = 1, \ \ l_1 = 1, \ldots, n_1, \end{aligned} \tag{2.19}$$

where $\mathbf{X} \geq 0$ means that $\mathbf{X}$ is symmetric and positive semidefinite.

The quadratic problem defined in (2.15) is solved by semidefinite programming package given in [15]. First, the binary variables α_is are changed to $\mathbf{x}$, $(-1, 1)$ of semidefinite programming according to [70], then the solution of semidefinite programming is mapped to the solution of binary quadratic problem by randomization technique defined in [59]. The complexity of semidefinite programming is estimated in polynomial time according to [59].

To evaluate how good (2.15) estimates the PAPR reduction, the peak to average energy of $\mathbf{w}$ given in (2.9) is minimized in terms of α_i by two heuris-

tic algorithms called hill climbing and simulated annealing. Hill climbing (discrete form of steepest descent algorithm) [66] is a search technique which starts with a known solution and at each step examines all possible changes to the input parameters, dummy bits in our case, and selects the change that results in the best improvement. In this application, we set all the dummy bits to 0, perform the entire encoding process and calculate the PAPR. Then, we iterate over all the dummy bits, and for each dummy bit we flip the value and perform the encoding and calculate the PAPR. For each flip, we must undo all previous flips, i.e. each new solution is different from the current one by one bit. We find the bit that results in the largest decrease, flip it and continue until no further improvement is observed. The complexity of hill climbing can be as large as that of the exhaustive search.

The other optimization technique examined is the simulated annealing [66]. Simulated annealing is a generalization of a Monte Carlo method for examining the equations of state and frozen states of a system [65]. To apply the simulated annealing technique to our problem, we start with all dummy bits set to zero and the solution with temperature $Temp$. In this application, the PAPR is interpreted as the energy E of the system. Similar to the hill climbing method, all the possible bits are flipped, and for each possible dummy bit flip, ΔE is calculated as the difference in the PAPR in the two cases. If the change is negative, the new configuration is accepted and if it is

positive, it is accepted with the probability of $\exp(-\Delta E/(K_B Temp)$, where K_B is the Boltzman's constant. Over time, the temperature $Temp$ will be slowly decreased until it reaches 0, where the simulated annealing becomes a hill climbing search. Once no further improvement is made, the algorithm terminates. The complexity of simulated annealing can be as large as that of the exhaustive search.

2.7 Simulation Results

The simulation is done for QAM-256 in two different cases for MMSE-Threshold.

1- When the clipping threshold is set to zero, and starting point for the minimum PAPR search is set to all dummy bits equal to zero. In this case, the simulations are done for two different numbers of sub-carriers, $N = 128$ and $N = 1024$.

2- When the optimum value of threshold for minimum PAPR is determined by simulation and few random starting points are tried to select the minimum PAPR among them. In this case, the simulations are performed for two different numbers of sub-carriers, $N = 256$ and $N = 1024$.

In both of these cases, as the number of sub-carriers changes, PAPR changes considerably. In addition, simulation is done for different numbers of

bits per dimension (QAM-4, QAM-16, QAM-64); however, there is a negligible change in PAPR. Therefore, the corresponding graphs are not presented.

In MMSE-Threshold, we have considered one dummy bit to have the choice of selecting the set with lower PAPR. Simulation results show that increasing the number of dummy bits has a negligible effect on PAPR but a noticeable effect on the clipped energy above the threshold.

Since MMSE-Threshold is based on a clipping algorithm and clipping is a non-linear operation, MMSE-Threshold does not guarantee that we can get the minimum PAPR, but the simulation results show a large improvement in PAPR for MMSE-Threshold compared to original OFDM. We minimize the error by (2.15) at an appropriate threshold level and 20 random starting points for case 2 of simulations for MMSE-Threshold.

In Fig. 2.6 and Fig. 2.5, the time domain signal is obtained by assuming zero value for dummy bits in different sub-spaces and the threshold is set to zero. In these figures, cumulative distribution of PAPR in terms of PAPR values are depicted. From Fig. 2.6 and Fig. 2.5, it is clear that by the proposed MMSE-Threshold, we have gained an improvement of approximately $5dB$ and $4dB$ at 10^{-3} in PAPR reduction for $N = 128$ and $N = 1024$, respectively. Because the average energy in the proposed method has reduced $1dB$, the PAPR graphs of the proposed method are shifted to the left with

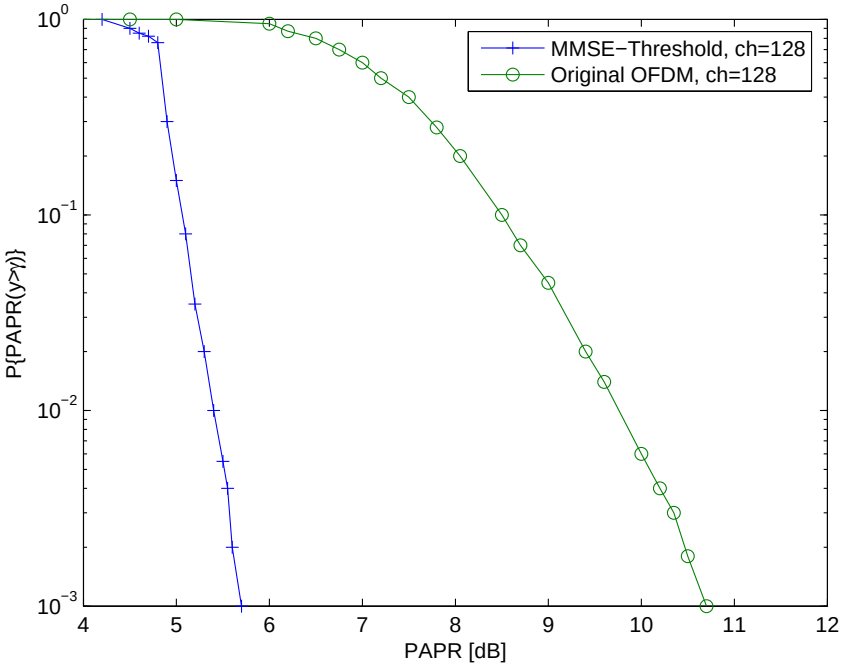

Figure 2.5: Probability that PAPR is greater than PAPR value on the x axis (N=128)

the amount of $1dB$ for a fair comparison. Fig. 2.7, shows the energy clipped above threshold for different threshold levels. As it is clear from the graph, by increasing the number of dummy bits, total clipped energy decreases but it has a negligible effect on PAPR, which is shown in Fig. 2.8. Therefore, we assume only one dummy bit in every sub-space.

Fig. 2.9 shows the probability that the PAPR is larger than some defined value using a system with $N = 256$ for the following optimization techniques:

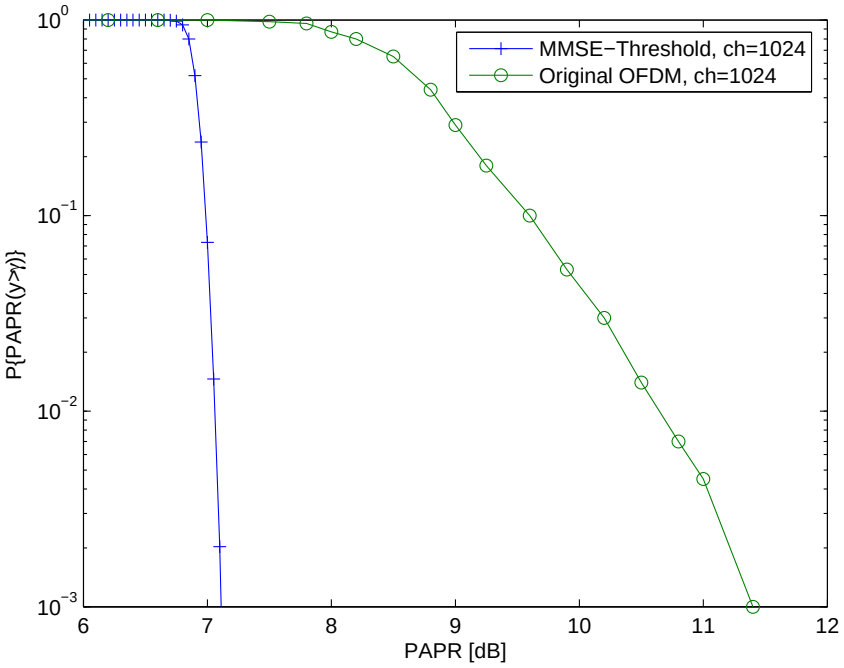

Figure 2.6: Probability that PAPR is greater than PAPR value on the x axis (N=1024)

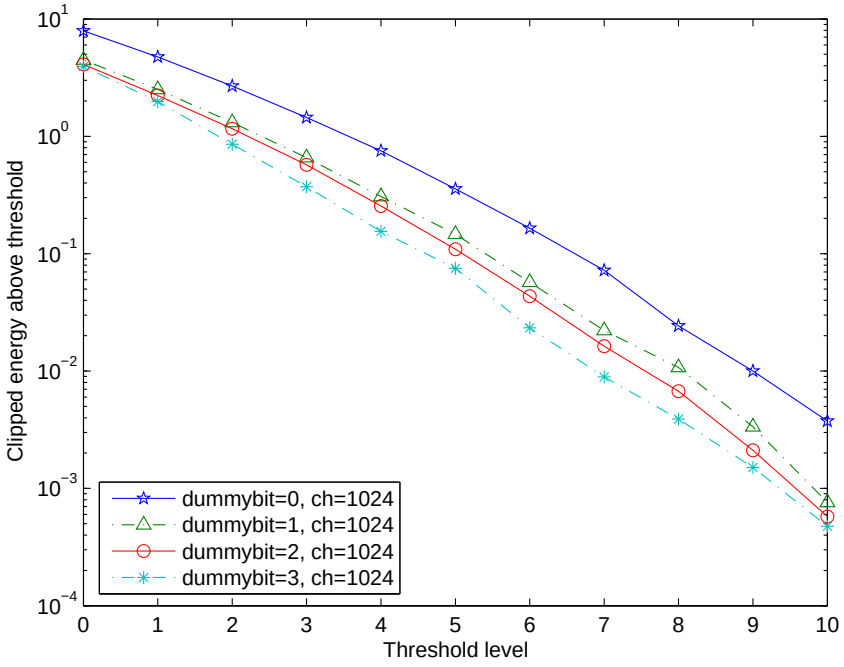

Figure 2.7: Clipped energy above threshold for different numbers of dummy bits (N=1024)

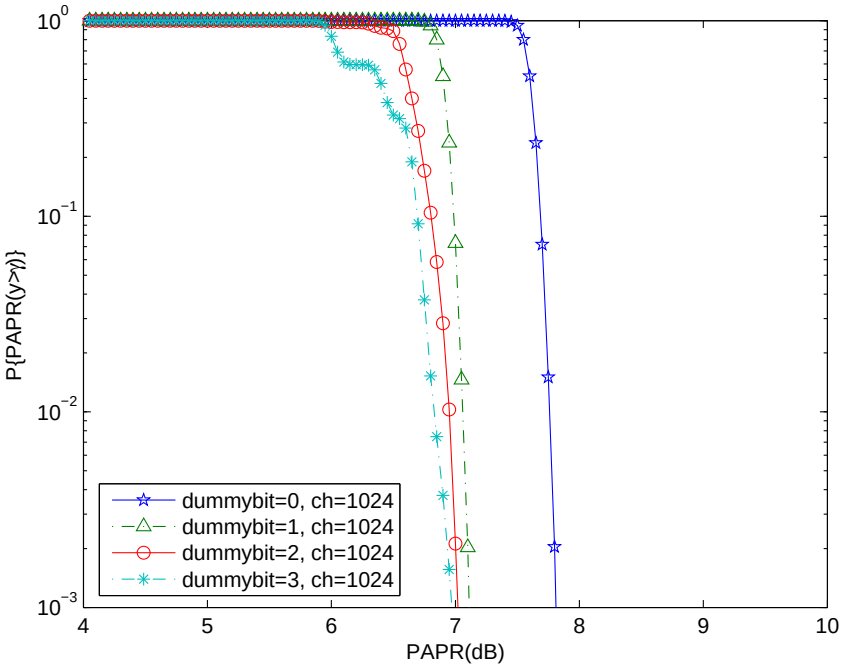

Figure 2.8: PAPR reduction for different numbers of dummy bits (N=1024)

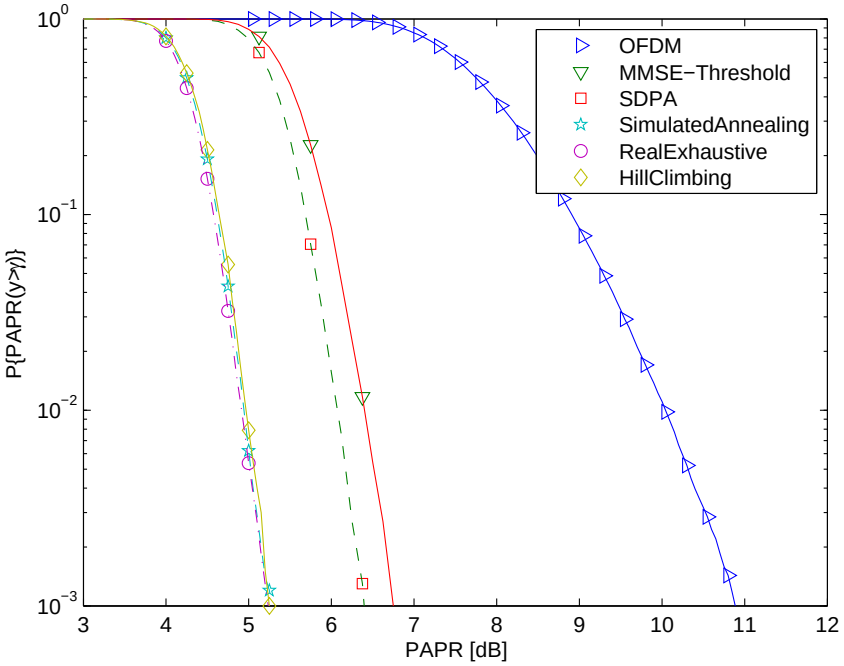

Figure 2.9: Probability that PAPR is greater than PAPR value on the x axis (N=256)

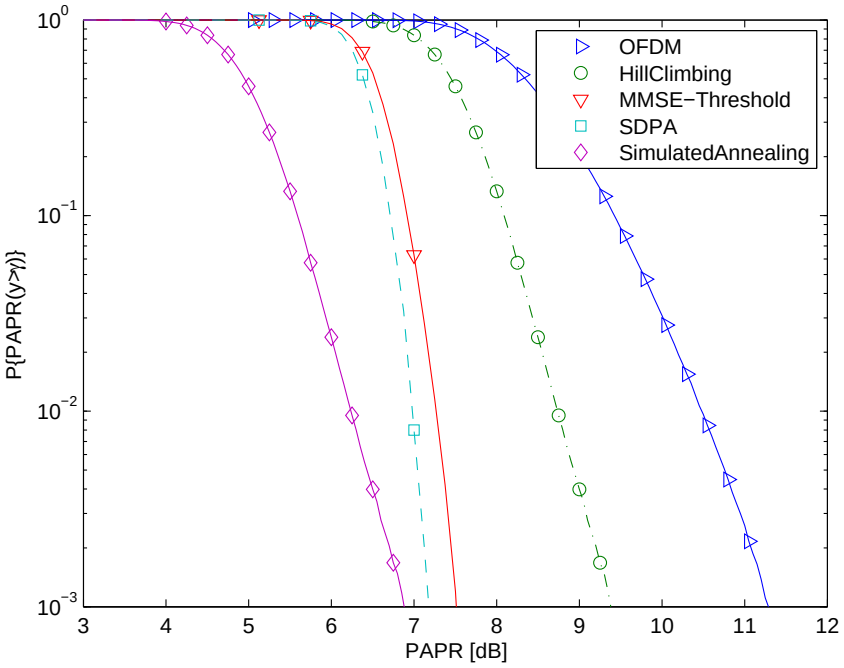

Figure 2.10: Probability that PAPR is greater than PAPR value on the x axis (N=1024)

MMSE-Threshold, semidefinite programming, hill climbing, simulated annealing and exhaustive search. Fig. 2.9 indicates that MMSE-Threshold results in PAPR of $6.7\,dB$ at the probability of 10^{-3}, which corresponds to an improvement of about $4dB$ over the original OFDM system, while SDPA has achieved noticeably better PAPR with higher complexity. The PAPR reduction for exhaustive search, hill climbing and simulated annealing are better than that of MMSE-Threshold technique because they have minimized the PAPR problem (not the quadratic problem defined in (2.15) with much higher complexity compared to MMSE-Threshold.

Fig. 2.10 represents the PAPR results for a system with $N = 1024$ for the following optimization techniques: MMSE-Threshold, semidefinite programming, hill climbing and simulated annealing. The exhaustive search algorithm is not feasible due to the size of the problem in this case. MMSE-Threshold results in PAPR of $7.5\,dB$ at the probability of 10^{-3}, which corresponds to an improvement of about $4\,dB$ over the original OFDM system, while SDPA has noticeable improvement over that of MMSE-Threshold with higher complexity. The PAPR reduction for simulated annealing is better than that of MMSE-Threshold technique because it has minimized the PAPR problem (not the quadratic problem defined in (2.15)) with much higher complexity compared to MMSE-Threshold. The hill climbing performance is significantly inferior compared to the simulated annealing technique

because the number of sub-carriers is large.

In terms of symbol error rate (SER), MMSE-Threshold has negligible degradation compared to the original OFDM signal in additive white Gaussian noise (AWGN) channel and the corresponding graph is not presented.

2.7.1 Comparison with Other Methods

In the following, MMSE-Threshold technique is compared with state of the art methods reported in [6], [96], [69], and [78].

In [6], an orthogonal projection-based partial transmission sequence approach has been presented. The modulation is 16-QAM OFDM with 128, 512, and 1024 sub-carriers. The amount of PAPR at 10^{-3} for the case of $N = 1024$ is $7.5dB$, which is comparable to MMSE-Threshold. In addition, MMSE-Threshold achieves a shaping gain of $1dB$. The complexity of the method in [6] is N trials, each of which performs IFFT with complexity $N \log N$ plus NV complex multiplications where V is the number of sub-blocks (V can be a small or large number). From the complexity point of view, MMSE-Threshold has less complexity. In addition, the system in [6] requires a side information channel to send the phase vector.

In [96], active constellation extension (ACE) with frame interleaving is proposed, where F frames with N sub-carries are combined into one super

frame. The ACE algorithm is then applied to reduce the PAPR of the super frame. Two cases using QPSK modulation are presented: with frame interleaving with the number of sub-carriers equal to $64, 128, 256, 512, 1024$, and 2048 and without frame interleaving with the number of sub-carrier as 128 and 256. The case with frame interleaving can achieve more PAPR reduction than without interleaving at the price of higher complexity.

For the case without interleaving and $N = 256, 1024$, the PAPR values at probability 10^{-3} are $8.5dB$, and $8.4dB$, respectively. The PAPR of MMSE-Threshold is better with PAPR values of $6.7\,dB$ and $7.5\,dB$, respectively. The complexity of the ACE without frame interleaving at the transmitter is $O(FN \log N)$ for small or large values of F. This complexity is bigger than the complexity of MMSE-Threshold method. In addition, MMSE-Threshold method achieves a $1dB$ shaping gain, compared to ACE without frame interleaving which slightly increase the average power ($0.18dB$).

With frame interleaving and $N = 256$, the PAPR at probability 10^{-3} is $8dB$ which is larger than that of MMSE-Threshold. The complexity of this method at the transmitter is $O(NF \log NF)$ which is bigger than the complexity of MMSE-Threshold. In addition, MMSE-Threshold method achieves a $1dB$ shaping gain, compared to ACE with frame interleaving which slightly increases the average power ($0.2dB$). In terms of bit error rate (BER), there is a little degradation in ACE with and without interleaving [96].

In [69], a technique is presented that relies on a cubic constellation called the Hadamard constellation, whose boundary is along the bases defined by the Hadamard matrix. This method is combined with a selective mapping (SLM) technique for further PAPR reduction, at the cost of increasing the complexity.

The system is 16-QAM with $32, 64$, and 128 sub-carriers before application of the SLM technique. The PAPR at probability 10^{-3} is $6.8dB$ for a system with 128 sub-carriers compared to MMSE-Threshold method at $6.7dB$ with 256 sub-carriers. The complexity of the encoding algorithm for the Hadamard constellation is $O(3/2\log(N))$, which is less than that of MMSE-Threshold algorithm ($O(N)$). With the SLM technique for a system with 128 sub-carriers and one redundant bit, the PAPR is around $5.8dB$, doubling the complexity. MMSE-Threshold technique achieves a $1dB$ shaping gain over the technique presented in [69]. In terms of SER, the technique in [69] has negligible degradation compared to original OFDM.

In [78], a new trellis shaping design based on the control of the autocorrelation side lobes of an OFDM data sequence with 256-QAM is presented. Two types of bit mapping referred as Type-I and Type-II are designed to achieve PAPR reduction. Type-I system of constellation mapping forms a symmetry with respect to both of the axes while Type-II system is designed such that all the quadrants of QAM bit mappings are identical.

Type-I system with 256 sub-carriers with sign-bit shaping [12] results in the PAPR at probability 10^{-3} as $7\,dB$, where MMSE-Threshold with 256 sub-carriers has PAPR $6.7\,dB$ at probability 10^{-3}. Type-II system with 256 sub-carriers with sign-bit shaping results in PAPR at probability 10^{-3} of $7.6\,dB$. Also, Type-II system has constellation expansion ratio of 2 for sign-bit shaping compared to that of MMSE-Threshold method which is 1.35. The complexity of both Type-I and II in [78] for encoding are estimated as $O(N^2)$ which is larger than the complexity of the MMSE-Threshold algorithm ($O(N)$). In terms of BER, there is a significant degradation for Type-I system and a small degradation for Type-II system compared to 128-QAM OFDM which achieves the same information rate as sign-bit shaping 256-QAM OFDM. In addition, MMSE-Threshold achieves a $1dB$ shaping gain over the best results presented in [78].

2.8 Conclusion

A new method (MMSE-Threshold) for peak/average power reduction is presented. This technique is derived from the constellation shaping algorithm, where the constellation points with lower average energy are selected from a larger set of points. To select the points with lower peak energy, multiple choices are available for a given sequence of data bits, and this flexibility

is used to reduce the peak to average power ratio (PAPR). Subsequently, this selection algorithm, which is formulated in terms of a zero-one quadratic problem, is optimized by the semidefinite programming algorithm (SDPA).

Simulation results show that the PAPR reduction of MMSE-Threshold is significant compared to the original OFDM. The PAPR reduction of SDPA is noticeably better than MMSE-Threshold, while the complexity of MMSE-Threshold is less than that of SDPA. MMSE-Threshold is compared with some state-of-the-art methods, and the simulation results show that our proposed technique has better or similar PAPR reduction compared to the state-of-the-art techniques and the proposed technique's complexity is low. In addition, the proposed method results in a $1dB$ reduction in average energy.

Chapter 3

Data Traffic Estimation

3.1 Introduction

Accurate forecasting is crucial in many organizations and plays an important role in most decision-making processes [3]. Companies rely on predictions of demand to justify investment and to ensure that resources are correctly allocated [11]. To generate predictions, information about past events, called "time series data", is collected. Time series data fall into four general pattern types: stationary, trend, seasonal, and cyclical [24].

A stationary series is one whose basic statistical properties are constant over time. The trend pattern is the long-term component that represents the growth or decline of the time series over a long period of time. A sea-

sonal pattern corresponds to a series that repeats every year, while a cyclical pattern repeats every two, three, or more years.

Several of naive forecasting techniques use a very simple model to predict future values. Some of these methods simply use the last observed value to predict the future value [24]. For stationary data, the moving average model is a suitable forecasting method, which makes a prediction based on the mean computed from the most recent data [18]. Exponential smoothing is a simple forecasting method, in which the forecast is constructed from an exponentially weighted average of all past observations [87]. The exponential smoothing model can be used to forecast stationary data; however, it does not work well for data with trend and seasonality [87].

Autoregressive integrated moving average (ARIMA) methods are more advanced models that integrate an autoregressive model and a moving average model [100, 61]. This modelling is a very powerful tool that provides accurate forecasts and, compared to the previous models, does not assume any particular pattern in time series data. However, the disadvantage of this model is that it needs a relatively large amount of data to develop the model. In addition, there is no easy way to update the parameters as new data become available, because the model has to be completely refitted.

Another advanced class of methods uses the structural state space models,

which capture observations as a sum of separate components (such as trend and seasonality) and can be used to forecast the stationary, trend, seasonal, and cyclical data [57]. This is a very powerful technique in forecasting that can easily update the parameters of the model as new data become available, which is better than the ARIMA technique that needs to completely refit the model.

3.2 Related Work

Several researchers investigated the problem of predicting the seasonal time series. Magalhaes et al. proposed an approach to combine three forecasting methods to predict the seasonal stream flow time series in southeast Brazil [60]. The three techniques are the periodic autoregressive moving average (PARMA) model and two fuzzy, clustering-based forecasting models. They are combined by a simple neural network trained with the steepest descent algorithm. The work provides results for a four-year prediction using 55 years (660 points) as the training set.

Guang et al. presented a seasonal neural network model to forecast the seasonal internet traffic of eastern China, CERNET [22]. This model consists of three layers: the input layer, the hidden layer, and the output layer. The data samples are divided into different sets such that every set contains

one period of data and are used to train the neural network with a back propagation technique. The length of the training data set is 30 days and the calibration period is 10 days; since the traffic is collected hourly, the total length of the training data is 720.

Young presented an unobserved component model that captures the effect of a vector of exogenous variables [99, 97]. This model is used to restore the non-stationary audio signal of a recorded piano sample. The stochastic varying parameters of the model are described by the generalized random walk process, and estimation of the parameters is performed by maximum log-likelihood estimation using the Kalman filter. Parameter optimization is performed in the frequency domain. An autoregressive spectrum of the observation process is estimated to find the linear least square estimate of the objective function. Kalman filter recursions are performed to restore the signal, where the total number of samples was 3000.

The ARIMA technique is used to forecast NSFNET internet traffic, where the traffic is seasonal and non-stationary [20]. A non-stationary series can often be transformed into a stationary series if the data is subtracted from the data in the previous period. To capture the seasonality, a seasonal form of ARIMA (p, d, q) technique is used, where p is the order of the autoregressive part, d is the order of the differencing, and q is the order of the moving average part, respectively. Data analysis is done by the S-PLUS [86]

statistical package, where the appropriate values of p and q are determined from inspection of the autocorrelation and partial autocorrelation functions. Three diagnostic tests are described in [4], defined as standardized residuals, autocorrelation plot of the residuals, and goodness of fit statistics. The training data are from August 1988 through June 1992, and the forecasting is from July 1992 through June 1993.

Hansen and Nelson proposed a neural network model to forecast revenues for the Utah legislative session in 1997 [25]. Two economic measurements play significant roles in revenue forecasting: the growth rate for non-agricultural employment and the total taxable sales. The first measurement is a cyclical pattern of variable period, and the second measurement is a quarterly seasonal pattern. The prediction is done with time-delay neural network and the application of a genetic algorithm. The growth rate of non-agricultural employment measurement is based on training data from 1981-96 and the forecast is performed for a period of one year. The taxable sale measurement is based on training data from 1978 to 1996 and a prediction of one year.

Tych et al. present a customized package, CRESFOR, for forecasting and signal extraction [91]. This technique was developed at Lancaster University to forecast up to three weeks in advance the hourly telephone calls made to Barclaycard place. The model is an enhancement of the dynamic harmonic

regression (DHR) model [98]. The DHR model is an extension of the BSM model [26] but is different in the nature of the unobserved components, their state space representation, and optimization of hyper-parameters. The training data span one year, 1998-1999, and the prediction period is 70 realization points in one week.

Nelson et al. investigated if a neural network is capable of modelling the data with seasonality [76]. They determined that deseasonalising the data before applying the neural network model is better than using the data directly. Later, Prochazka [82] used a neural network to construct a forecasting method without needing to deseasonalise the data. To process the data, [82] first applied a wavelet transform and then used the neural network to forecast in the wavelet domain.

Econometric models completely different from previous models can forecast economic parameters and are suitable for cyclical data. Since economic data may exist in both a time series and cross-sectional form (data between two countries at the same time), we can also combine these two types of information in the statistical models [38]. There are four basic models: the fixed effect model, the random effect model, the random coefficient model, and the coefficients as functions of exogenous variables model [19, 39, 34, 33]. These techniques are suitable for applications with short time series and that include cross-sectional data. Several other forecasting techniques have also

been proposed to forecast time series data [32, 85, 51, 8, 84, 27].

3.3 Data Traffic Model

3.3.1 Time Series Definition

Time series is a sequence of observations $X_1, X_2, \ldots, X_{t-1}, X_t$ of a random process X at discrete time intervals, where an observation at time t is given by X_t. Time series are often described as having trends and seasonal terms. By plotting the data traffic, the total number of minutes of call per month in Bell Canada network, we can easily inspect a linear trend, and seasonality of 12 since there are 12 months in a year. Then, each observation X_t is defined as the summation of three components: trend, seasonal, and measurement noise, and is given as [35]

$$X_t = \mu_t + \gamma_t + \epsilon_t, \tag{3.1}$$

where μ_t is the trend component at time t, γ_t is the seasonal component at time t, and ϵ_t is the measurement noise at time t. This is the simplest model that can capture the behavior of our traffic. More sophisticated models are tried by making non-linear trend or multiplicative seasonality but the results were not promising.

3.3.2 State Space Model

The observation sequence $X_1, \ldots, X_t$ can be described by a state space model:

$$X_t = \mathbf{h}\boldsymbol{\alpha}_t + \epsilon_t \tag{3.2}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \mathbf{K}\boldsymbol{\theta}_t. \tag{3.3}$$

The state space model is described in terms $\mathbf{h}$, $\boldsymbol{\Phi}$, and $\mathbf{K}$. At time t, the past observations of X are encapsulated by a state vector $\boldsymbol{\alpha}_t$, and any future observations are assumed to be independent of the past observations, given this state vector. The state vector is composed of r state variables $\alpha_t^{(1)}, \ldots, \alpha_t^{(r)}$, such that $\boldsymbol{\alpha}_t = (\alpha_t^{(1)}, \ldots, \alpha_t^{(r)})^T$.

The current state $\boldsymbol{\alpha}_t$ is calculated from the previous state $\boldsymbol{\alpha}_{t-1}$ by the state equation (3.3), where $\boldsymbol{\Phi}$ is a $r \times r$ matrix of the parameters, called the transition matrix, and $\mathbf{K}$ is a $r \times n_j$ matrix, called the state noise coefficient matrix. The state noise $\boldsymbol{\theta}_t$ is an n_j-dimensional, zero-mean Gaussian noise vector with the covariance matrix

$$\boldsymbol{\Sigma} = E\{\boldsymbol{\theta}_t {\boldsymbol{\theta}_t}^T\} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_{n_j}^2). \tag{3.4}$$

The observation X_t is calculated from the state vector $\boldsymbol{\alpha}_t$ by the observation equation (3.2). Here, $\mathbf{h}$ is a $1 \times r$ observation vector and ϵ_t is the measurement noise, defined as a white Gaussian noise with zero mean and variance of σ_ϵ^2.

3.3.3 Basic Structural Model

The time series defined in (3.1) leads to the simplest (i.e. with fewest state variables) corresponding state space model [35]:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \\ \gamma_t &= -\sum_{u=1}^{s-1} \gamma_{t-u} + \omega_t, \end{aligned} \tag{3.5}$$

with the state vector defined as

$$\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-s+1})^T. \tag{3.6}$$

Given the equations in (3.5) and the state vector in (3.6), the corresponding state space matrices $\mathbf{\Phi}$, $\mathbf{K}$ given in (3.3) can be derived. In this model, s is the season period and β_t is the trend slope. The parameters η_t, ζ_t, and ω_t are noise of the trend component, noise of the trend slope, and noise of the seasonality component, respectively. These parameters are mutually uncorrelated white Gaussian random variables with a zero mean and variances of σ_η^2, σ_ζ^2, and σ_ω^2.

This model is well-known in statistical forecasting and has not been used in telecommunications. Our data traffic shows an increasing linear trend with seasonality factor equal to a period of 12 months. This model was selected by inspection since it can capture the linear trend and seasonality of the

data in a simple way. Other models with a higher degree of the recursive equations and larger number of states have also been studied; however, there was not much improvement in performance, although the models were more complicated.

3.3.4 Model Parameter Estimation

The state space model is defined by equations (3.2) and (3.3) and one needs to estimate some or all of the unknown parameters $\mathbf{h}$, σ_ϵ^2, $\mathbf{\Phi}$, and $\mathbf{K}$. To estimate the unknown parameters, the maximum log-likelihood criterion is used. The likelihood of a time series X can be decomposed into the product of the conditional distributions of each X_t on its predecessors, i.e.

$$L = \prod_{t=1}^{N_j} P(X_t | X_1, X_2, \ldots, X_{t-1}), \tag{3.7}$$

where N_j is the total number of observations. Equivalently, we can also calculate the log-likelihood function as

$$\log L = \sum_{t=1}^{N_j} \log P(X_t | X_1, X_2, \ldots, X_{t-1}). \tag{3.8}$$

Since the joint distribution of the series X is multi-variate normal, the distribution of any observation X_t conditioned on any other observation must be normal. Given that the conditional probability density function P is

Gaussian, the log-likelihood function of the time series X is given by

$$\begin{aligned} \log P(X_t|X_1, X_2, \ldots, X_{t-1}) &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log F_t - \frac{1}{2F_t}(X_t - \overline{X}_t)^2 \\ &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log F_t - \frac{V_t^2}{2F_t}, \end{aligned} \tag{3.9}$$

where $\overline{X}_t$ is the mean of the distribution, F_t is the variance, and $V_t = X_t - \overline{X}_t$ is the prediction error. The log-likelihood of the series is then given by

$$\log L = -\frac{N_j}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{N_j}\log F_t - \frac{1}{2}\sum_{t=1}^{N_j}\frac{V_t^2}{F_t}. \tag{3.10}$$

To compute this log-likelihood function, the $\overline{X}_t$ and F_t must be calculated for all $t = 1, \ldots, N_j$. This calculation is achieved by a set of recursions, called the Kalman filter recursions.

These recursive computations are divided into two sets of equations: prediction equations and update equations [41]. From (3.3), we can derive the prediction equations given as

$$\begin{aligned} \mathbf{a}_{t|t-1} &= \mathbf{\Phi}\mathbf{a}_{t-1} \\ \mathbf{C}_{t|t-1} &= \mathbf{\Phi}\mathbf{C}_{t-1}\mathbf{\Phi}^T + \mathbf{K}\mathbf{\Sigma}\mathbf{K}^T, \end{aligned} \tag{3.11}$$

where $\mathbf{a}_{t|t-1}$ is the minimum mean square estimate of the state $\boldsymbol{\alpha}_t$, given all the previous observations at time $t-1$. The corresponding covariance matrix of this estimate is given by $\mathbf{C}_{t|t-1}$.

The update equations for the Kalman filter are given as [35]

$$\begin{aligned}
\overline{X}_t =& \mathbf{h}\mathbf{a}_{t|t-1} \\
F_t =& \mathbf{h}\mathbf{C}_{t|t-1}\mathbf{h}^T + \sigma_\epsilon^2 \\
\mathbf{C}_t =& \mathbf{C}_{t|t-1} - \mathbf{C}_{t|t-1}\mathbf{h}^T\mathbf{h}\mathbf{C}_{t|t-1}/F_t \\
\mathbf{a}_t =& \mathbf{a}_{t|t-1} + \mathbf{C}_{t|t-1}\mathbf{h}^T(X_t - \overline{X}_t)/F_t,
\end{aligned} \tag{3.12}$$

where $\mathbf{a}_t$ is the minimum mean square estimate of the current state $\boldsymbol{\alpha}_t$ at time t, and $\mathbf{C}_t$ is the corresponding covariance matrix of this estimate.

In practice, the recursions are started at $t = 1$, requiring the estimator of the state at $t = 0$, $\mathbf{a}_0$ and the corresponding error covariance matrix $\mathbf{C}_0$. It can be shown that for models that reach a steady state, the state estimates for large t are unaffected by the initial choice of $\mathbf{a}_0$ and $\mathbf{C}_0$. The important advantage of this recursive algorithm is that $\mathbf{a}_{t|t-1}$, $\mathbf{a}_t$ and $\overline{X}_t$ are updated for new available data without needing to restart the computation for all the previous data.

The Kalman filter is first used to compute the log-likelihood function, which must then be maximized to find the unknown parameters. This is an optimization problem, and for our case, $\mathbf{\Sigma} = \text{diag}(\sigma_\eta^2, \sigma_\omega^2)$ and σ_ϵ^2 include the unknown variables in log-likelihood function. In this work, we have implemented optimization techniques such as steepest descent [77], quasi-Newton, and simulated annealing [45] to optimize the log-likelihood function.

The implementation results are similar while steepest descent is a simple method and obtains the minimum at a fast convergence rate.

3.4 Extended Structural Model

To improve performance in traffic prediction, we define the extended structural model by adding extra degrees of freedom to the basic structural model in (3.5). The extended traffic model is given as

$$\begin{aligned}
\mu_t &= b_j \mu_{t-1} + c_j \beta_{t-1} + \eta_t \\
\beta_t &= d_j \beta_{t-1} + e_j \zeta_{t-1} \\
\zeta_t &= f_j \zeta_{t-1} \\
\gamma_t &= -\sum_{u=1}^{s-1} \gamma_{t-u} + \omega_t.
\end{aligned} \tag{3.13}$$

The parameter ζ_t in the model defined in (3.13) is changed from a random variable in the basic structural model in (3.5) to a state variable. The extended model also contains additional variables b_j, c_j, d_j, e_j, and f_j. In this model, the state vector is defined as

$$\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \zeta_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-s+1})^T. \tag{3.14}$$

During the training phase, both the basic structural model parameters Σ

and σ_ϵ^2, as well as the extended structural model parameters b_j, c_j, d_j, e_j, and f_j must be estimated. The data available to build the model (training sequence) is divided into a *model training sequence* and a *model validation sequence*. Note that the model training sequence must be followed by the model validation sequence.

The algorithm first selects initial values for b_j, c_j, d_j, e_j, and f_j. Using these values and the model training sequence, the Kalman filter prediction and update equations in (3.11) and (3.12) are used to estimate the noise variances $\mathbf{\Sigma} = \text{diag}(\sigma_\eta^2, \sigma_\omega^2)$ and σ_ϵ^2 by maximizing the log-likelihood function (3.8).

The estimated noise variances are used by the model to predict future values that are compared to the model validation sequence. The quality of the model is calculated as the mean absolute percentage error (MAPE) between the predicted values and the model validation sequence. The values for the extended model parameters b_j, c_j, d_j, e_j, and f_j are changed according to an optimization search strategy, and the entire process is repeated, leading to a new set of parameters and a new value for the MAPE. In this work, the search follows a steepest descent algorithm; however, other search techniques could be also used.

The parameters b_j, c_j, d_j, e_j, and f_j that resulted in the lowest MAPE are

used by the Kalman filter to build the final model. However, in this case the entire training sequence, i.e. both the model training and model validation, is used. The entire training algorithm is shown in Fig. 3.1. These results are presented in [47, 49, 50].

3.5 Simulation Results and Comparison with Other Methods

The effectiveness of the ESM forecasting model is tested by its prediction of the number of airtime minutes per month on the Bell Canada network. To collect this data, the switch measures the call times for every call, which are then summed for a given month. Since there are 12 months in a year and the traffic follows a yearly seasonal behavior, the seasonality of the model was selected as 12. Using this data, the performance of the ESM is compared to that of both the BSM and the X12-ARIMA [95] software. X12-ARIMA is a seasonal ARIMA package designed by the US census bureau that needs at least three complete cycles of data for prediction to have a small prediction error.

The available data sequence is divided into training and test sequences; the training sequence is used to build the models and the test sequence is used

{*initialization*}

Split the training data into model training and model validation sequences

Select initial values of b_j, c_j, d_j, e_j, and f_j

while

Train the Kalman filter by the model training sequence using equations (3.11) and (3.12)

Maximize the log-likelihood function by steepest descent to find $\mathbf{\Sigma}$ and σ_ϵ^2

Use the Kalman filter on the model validation sequence to generate the predictions

$Error \leftarrow$ MAPE over the model validation sequence between the observed and predicted data

if $Error < minError$ **then**

$minError \leftarrow Error$

$b_{opt} \leftarrow b_j$, $c_{opt} \leftarrow c_j$, $d_{opt} \leftarrow d_j$, $e_{opt} \leftarrow e_j$, $f_{opt} \leftarrow f_j$

end if

if not done

Select new values for b_j, c_j, d_j, e_j, and f_j

else

break

end if

end while

Train the Kalman filter over the model training and model validation sequence using b_{opt}, c_{opt}, d_{opt}, e_{opt}, f_{opt}

Maximize the log-likelihood function by steepest descent to find $\mathbf{\Sigma}$ and σ_ϵ^2

Output the model

Figure 3.1: Extended structural model training algorithm

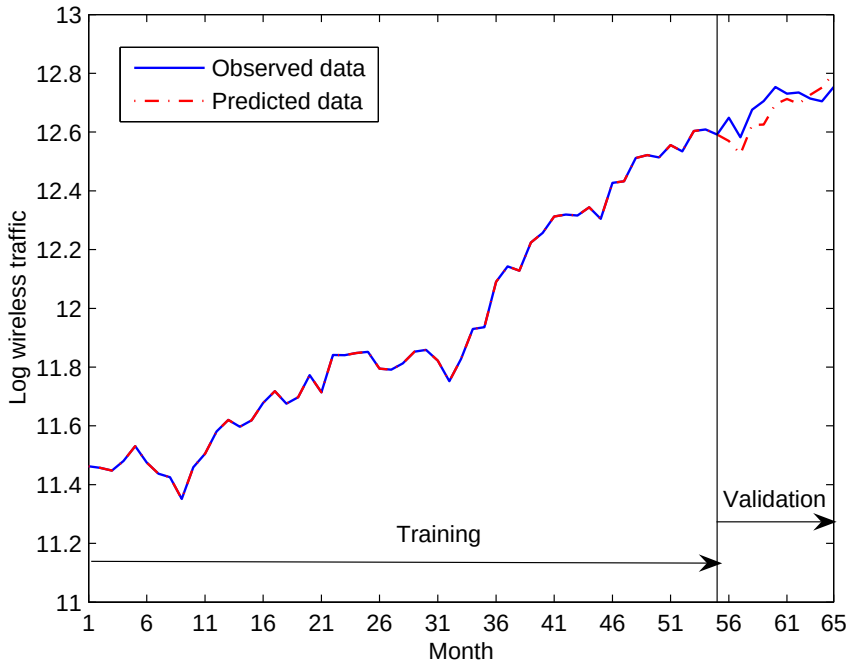

Figure 3.2: Model training and model validation sequences for the ESM model

to measure the performance of the models. Both the BSM and ARIMA use the entire training sequence to build the model, while ESM further divides the training sequence into model training and model validation sequences.

Fig. 3.2 shows the data sequence of 65 points divided into a model training sequence with 55 points and a model validation sequence with 10 points; the entire training algorithm is shown in Fig. 3.1.

Fig. 3.3 shows the wireless traffic prediction using X12-ARIMA, BSM, and

ESM models compared to the actual observed values. In Fig. 3.3, the first 55 points (training sequence) were used to train each model. In the case of ESM, the training sequence was further divided into first 45 points as the model training sequence followed by 10 points of the model validation sequence. Then, the models were used to predict points 56 to 65 (test sequence) and absolute percentage error between the predictions and the actual observations was calculated.

Fig. 3.4 shows similar results for 75 points, where the first 65 points are the training sequence, and points 66 to 75 are the test sequence. In the case of ESM, the training sequence was divided into a 55 point model training sequence and 10 point model validation sequence. Fig. 3.5 and Fig. 3.6 show the corresponding absolute percentage error of the different model predictions.

In both of the figures, we see that the ESM model follows the observed data more closely than either the BSM or the X12-ARIMA models. The improvement of the ESM over the BSM and the X12-ARIMA is better as the prediction period increases. We can also observe an improvement of each model as an extra 10 training points are added.

The absolute percentage error results in Fig. 3.5 and Fig. 3.6 are summarized in Table 3.1, which reports the mean absolute percentage error cor-

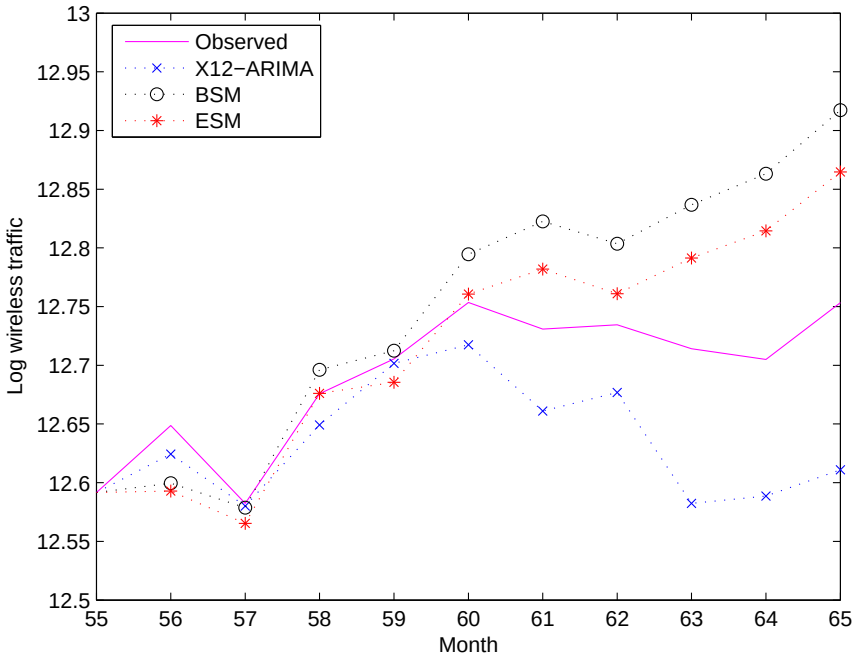

Figure 3.3: Prediction of 10 points for the wireless traffic using different models, 55 training points

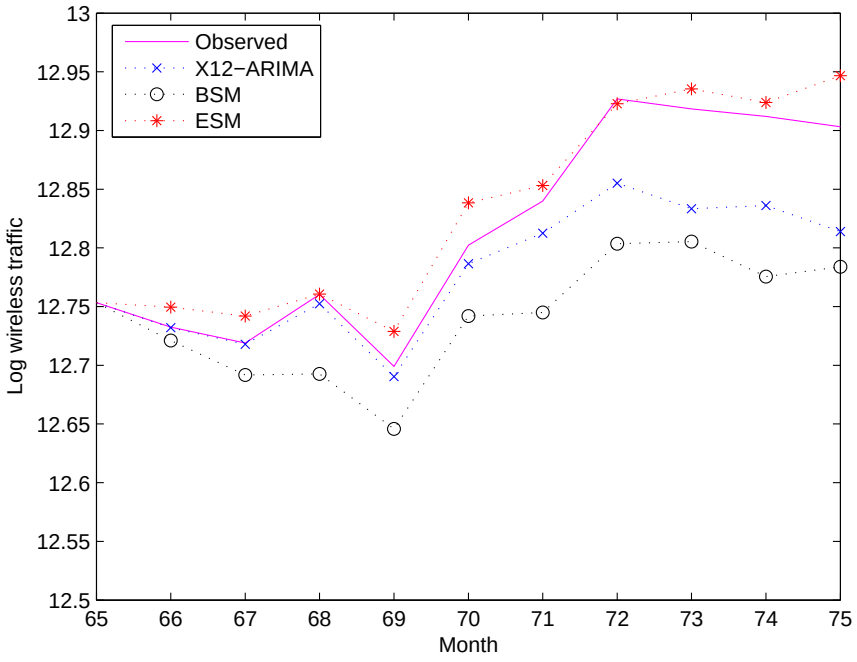

Figure 3.4: Prediction of 10 points for the wireless traffic using different models, 65 training points

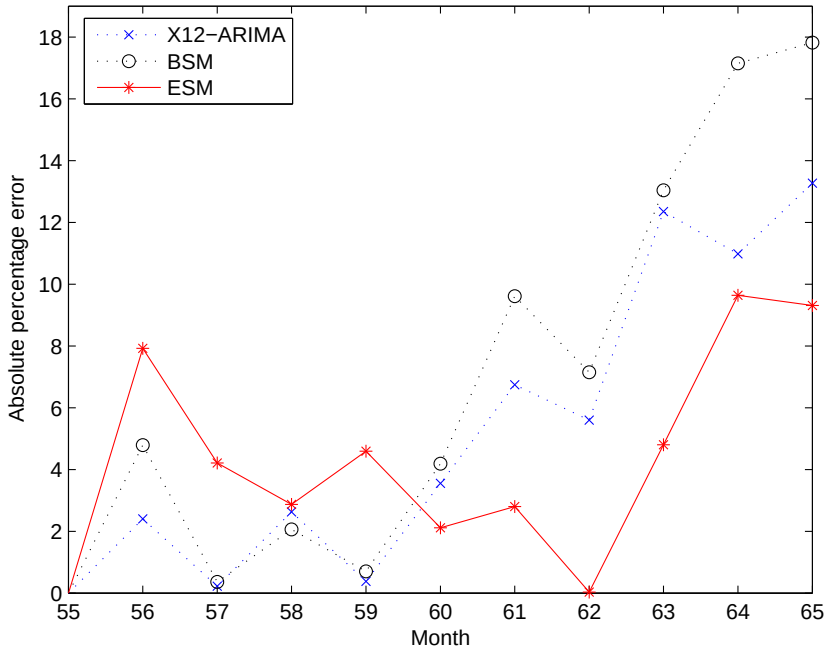

Figure 3.5: Absolute percentage error for 10 point prediction of the wireless traffic using different models, 55 training points

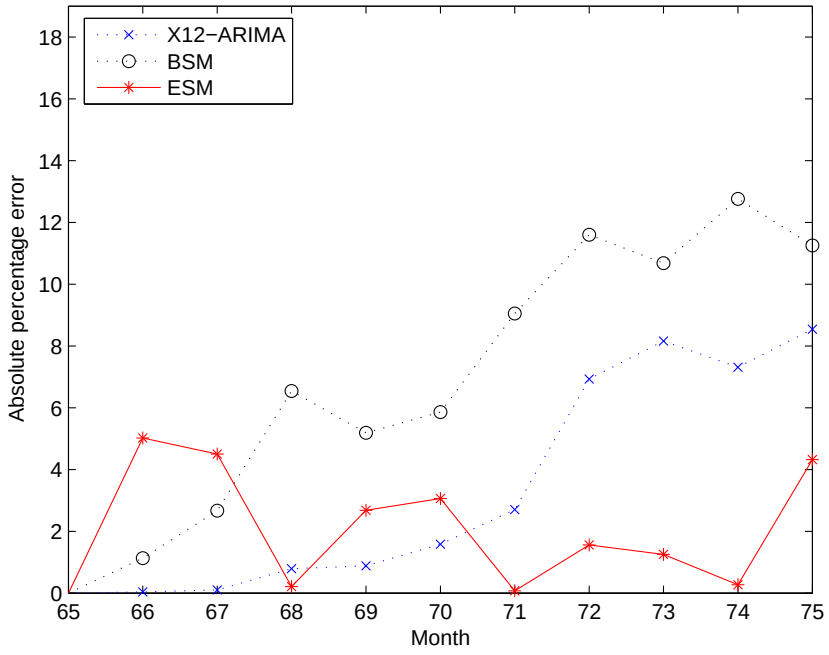

Figure 3.6: Absolute percentage error for 10 point prediction of the wireless traffic using different models, 65 training points

Training	**Test**	**X12-ARIMA**	**BSM**	**ESM**
mean absolute percentage error				
points 1-55	points 56-65	5.81 %	7.69 %	**4.83 %**
points 1-65	points 66-75	3.70 %	7.67 %	**2.29 %**
maximum absolute percentage error				
points 1-55	points 56-65	13.27 %	17.82 %	**9.64 %**
points 1-65	points 66-75	8.54 %	12.76 %	**5.02 %**

Table 3.1: Mean absolute percentage error and maximum absolute percentage error for the different prediction models

responding to the error values in Fig. 3.5 and Fig. 3.6 averaged over the prediction period. The maximum absolute percentage error is the maximum error value reported in Fig. 3.5 and Fig. 3.6. As Table 3.1 shows, the ESM model has the smallest value of mean absolute percentage error and maximum absolute percentage error compared to the BSM and X12-ARIMA models.

3.6 Conclusion

This chapter presented a prediction model of the total number of minutes of wireless airtime per month on the Bell Canada network. The presented

model is the new extended structural model (ESM) that was derived from the basic structural model (BSM). This novel technique is combined with a novel optimization algorithm where maximum likelihood estimation is combined with minimum mean absolute percentage error (MMAPE), to estimate the model parameter over the validation period of the data. This combination prevents a large degree of complexity of the optimization algorithm in maximum likelihood estimation and provides the opportunity to update the model parameters whenever new data become available.

Simulation results show that the ESM model has an improved mean absolute percentage error (MAPE) over the BSM and seasonal ARIMA model. In addition, the maximum of MAPE is also improved ensuring a more accurate forecasting for a longer period of prediction. The improved prediction can significantly reduce the cost for wireless service providers, who need to accurately predict future wireless spectrum requirements.

Chapter 4

Conclusion

4.1 Contributions

This dissertation considers two applications of non-linear programming: PAPR reduction in OFDM systems and airtime traffic estimation which both interest wireless service providers in cost saving. Signal fluctuation is a major disadvantage of OFDM and many other multi-carrier systems where these fluctuations can saturate the amplifiers and result in non-linear activity of the system. This problem is defined as peak to average power reduction (PAPR) in the literature. The proposed PAPR technique is based on constellation shaping algorithm and the improvements are explained later in the following. The second application is the airtime traffic estimation. This is an important

problem for many organizations and accurate forecasting can result in a huge cost saving for the companies. The proposed technique for airtime traffic estimation is based on the basic structural model and the improvements are described in the following.

Peak Reduction Algorithms with Low Complexity

Peak reduction is performed by defining extra bits (one dummy bit in every sub-space) where these dummy bits are used in the shaping algorithm as well. With the extra flexibilities provided by the dummy bits, we are able to reduce the peak energy as defined in section 2.6.2. Also, a new formulation based on MMSE criterion is defined in (2.15) where the complexity is lowered by definition of decision function in (2.16). The MMSE-Threshold technique has an improvement of approximately $4dB$ at 10^{-3} in PAPR reduction for $N = 1024$. In addition, minimization of (2.15) which is a quadratic problem is performed by semidefinite programming algorithm in section 2.6.2 which has a noticeable improvement for SDPA in PAPR reduction over that of MMSE-Threshold with higher complexity.

Application of Heuristic Algorithms for PAPR Reduction

The original PAPR reduction (peak of the signal energy divided by average of the signal energy) is optimized by two heuristic algorithms: hill climbing and simulated annealing. Simulation results presented in Chapter 2 indicate that

for a small number of sub-carriers, hill climbing and simulated annealing can represent the exhaustive search while for a larger number of sub-carriers, simulated annealing has a significant improvement over that of the hill climbing method. Both of these techniques have a large computational complexity.

Extended Structural Model for Airtime Traffic Estimation

A novel model called the extended structural model (ESM) is defined in section 3.4 and captures the behavior of airtime voice traffic. This novel model works with a proposed algorithm in the maximization of likelihood function where it combines the maximum likelihood estimation with mean absolute percentage error (MAPE). Then the model parameters are optimized by the MAPE criterion over a validation interval. The results provided in Chapter 3 indicate a performance improvement in terms of MAPE for ESM over BSM and X-12 ARIMA technique in the test interval and a smaller peak for absolute percentage error. In addition, ESM is capable of updating its parameters when new data become available, but the X-12 ARIMA technique needs to completely rebuild the model.

4.2 Future Work

As explained in Chapter 2, the MMSE-Threshold algorithm works based on constellation shaping, where the points with higher energy are removed. As

future work, we can execute a much smarter selection of points: removing the points with higher average energy and higher peaks and still using the dummy bits for more peak reduction.

In addition, the defined quadratic problem for the MMSE-Threshold technique is simplified to remove the non-linearity. As future work, we can consider the clipping function (non-linearity) and try to minimize it. In addition, we can try another quadratic programming method such as sphere decoding instead of the semidefinite programming, and try to find a lower bound to reduce the complexity.

As a continuation of airtime traffic estimation defined in Chapter 3, we can add the effect of some demographic factors such as gross domestic product (GDP), population, connection charges, and so on. These effects can result in a non-linear model that can not use the Kalman filter in optimization, but we can use the extended Kalman filter. In addition, we can replace the extended Kalman filter with particle filter with the advantage that, with sufficient samples, particle filters approach the Bayesian optimal estimate, so they can be made more accurate than the extended Kalman filter.

Appendix A

List of Abbreviations and Symbols

Abbreviations

2G	Second generation
3G	Third generation
ACE	Active constellation expansion
ADSL	Asymmetric digital subscriber line
ANSI	American national standard institute
AR	Autoregressive

ARIMA	Autoregressive integrated moving average
BER	Bit error rate
BPSK	Binary phase shift keying
BSM	Basic structural model
CDMA	Code division multiple access
CER	Constellation expansion ratio
DAB	Digital audio broadcasting
DHR	Dynamic harmonic regression
DMT	Discrete multi-tone
DS-CDMA	Direct sequence CDMA
DSP	Digital signal processing
DVB	Digital video broadcasting
DVB-T	Digital video broadcasting-terrestrial
ESM	Extended structural model
FDM	Frequency division multiplexing
FFT	Fast Fourier transform
GDP	Gross domestic product

IDFT	Inverse discrete Fourier transform
IFFT	Inverse fast Fourier transform
IMT-2000	International mobile telecommunications-2000
IS-95A	Interim standard 95, version A
IS-95B	Interim standard 95, version B
ISI	Intersymbol interference
MAI	Multiple access interference
MAPE	Mean absolute percentage error
MC-CDMA	Multi-carrier CDMA
ML	Maximum likelihood
MMAPE	Minimum mean absolute percentage error
MMSE	Minimum mean square error
MPSK	M-array phase shift keying
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
OOB	Out-of-band

PAPR	Peak to average power ratio
PAR	Peak to average ratio
PARMA	Periodic autoregressive moving average
PDF	Probability density function
PTS	Partial transmission sequence
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
SARIMA	Seasonal autoregressive integrated moving average
SDP	Semidefinite programming
SDPA	Semidefinite programming algorithm
SER	Symbol error rate
SNR	Signal to noise ratio
UMTS	Universal mobile telecommunications system
UQP	Unconstrained quadratic program
UTRA-FDD	UMTS terrestrial radio access frequency division duplex
VLSI	Very large scale integrated circuit

WCDMA	Wideband code division multiple access
WLAN	Wireless local area network
WMAN	Wireless metropolitan area network

Western Symbols

a_n	in-phase term of the QAM signal
$\mathbf{a}_t$	MMSE estimate of state $\boldsymbol{\alpha}_t$ at time t, $(r \times 1)$
$\mathbf{a}_{t\|t-1}$	MMSE estimate of state $\boldsymbol{\alpha}_t$ at time $t-1$, $(r \times 1)$
b_j	extra parameter defined for the ESM model
b_n	quadrature term of the QAM signal
b_{opt}	optimum value of b_j
c_j	extra parameter defined for the ESM model
c_{opt}	optimum value of c_j
C_1	number of sub-blocks
$\mathbf{C}_t$	covariance matrix of the estimate of $\boldsymbol{\alpha}_t$ at time t, $(r \times r)$
$\mathbf{C}_{t\|t-1}$	covariance matrix of the estimate of $\boldsymbol{\alpha}_t$ at time $t-1$, $(r \times r)$
d	order of the differencing in ARIMA

d_j	extra parameter defined for the ESM model
d_n	original data
d_{opt}	optimum value of d_j
D_1	number of divisions
e_j	extra parameter defined for the ESM model
e_{opt}	optimum value of e_j
$\mathbf{e}$	error vector of the clip function of $\mathbf{z}$
$\mathbf{e_{act}}$	actual clipped error vector
E	energy of the simulated annealing process (PAPR of the system)
f_j	extra parameter defined for the ESM model
f_{opt}	optimum value of e_j
F	number of frame
F_t	variance of V_t
g	index of discrete measurement of x
$\mathbf{h}$	observation vector, $(1 \times r)$
i	index of the sub-space

$J1$	number of constellation shaping modules
k	number of sub-carriers in every sub-space
$K1$	number of bits per 16 OFDM symbols
K_B	Boltzman's constant
$\mathbf{K}$	state noise coefficient matrix, $(r \times n_j)$
l	number of OFDM symbols in t
l_1	index of vector $\mathbf{x}$
L	likelihood function of the time series X_t
m	number of sub-spaces
$m_1, \ldots, m_{16}$	number of rings in shell mapping
n	sub-carrier index
n_1	constant number (vector dimension)
n_j	dimension of the zero-mean Gaussian noise vector
N	total number of sub-carriers
N_j	total number of observations
p	order of the autoregressive part of ARIMA
P	density function

P_c	cut-off probability
q	order of the moving average part in ARIMA
$\mathbf{Q}$	matrix of constants in quadratic function, $(n_1 \times n_1)$
r	number of state variables
R	number of samples of $x(t)$
s	period of seasonality
$\bar{\mathbf{s}}_i$	frequency domain signal vector with dummy bit set to zero in sub-space i
$\mathbf{s}_i$	frequency domain signal vector with dummy bit set to one in sub-space i
t	time index
T	clipping threshold for PAPR proposed algorithm
T_g	guard interval
T_s	useful symbol duration
T_{total}	total symbol duration
$Temp$	temperature in simulated annealing process
u	index of seasonality

$u1$	uncoded bits per OFDM symbol
V	number of sub-blocks
V_t	prediction error of the conditional distribution
$\mathbf{w}$	transmitted time domain signal vector
$\mathbf{w}_f$	frequency domain vector of $\mathbf{w}$
x	OFDM time domain signal
x_p	peak of the OFDM signal
$\mathbf{x}$	vector of binary variables, $(n_1 \times 1)$
X	time series data
X_t	observation of the data traffic at time t
$\overline{X}_t$	mean of the conditional distribution $X_t \vert X_1, X_2, \ldots, X_{t-1}$
$\mathbf{X}$	matrix of rank-1, $(n_1 \times n_1)$
y_i	received point
$\mathbf{y}$	clipped OFDM time domain vector at threshold T
$\mathbf{z}$	OFDM time domain signal vector with all dummy bits set to zero

Greek Symbols

α_i	dummy bit value in sub-space i
$\boldsymbol{\alpha}_t$	state vector, $(r \times 1)$
β_t	trend slope
$\gamma(i)$	decision function of dummy bit in sub-space i
γ_t	seasonal component of the data traffic
Δt	symbol duration of d_n
ϵ_t	measurement noise of the data traffic
$\boldsymbol{\epsilon}_i$	constructed time domain vector in sub-space i
ζ_t	noise of the trend slope
η_t	noise of the trend component
$\boldsymbol{\theta}_t$	state noise vector, $(n_j \times 1)$
μ_t	trend component of the data traffic
$\sigma_\eta^2, \sigma_\omega^2, \sigma_\epsilon^2$	noise variances
$\sigma_1, \ldots, \sigma_{n_j}$	variance of $\boldsymbol{\theta}_t$
$\boldsymbol{\Sigma}$	covariance matrix of the state noise $\boldsymbol{\theta}_t$, $(n_j \times n_j)$
$\boldsymbol{\Phi}$	transition matrix, $(r \times r)$

ω_n	sub-carrier frequency
ω_t	noise of the seasonality component

Bibliography

[1] Simulated Annealing. http://www.cs.sandia.gov/opt/survey/sa.html.

[2] J. Armstrong. Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering. *Electronics Letters*, 38(5):246–247, February 2002.

[3] B. L. Bowerman. *Time series and forecasting: an applied approach*. Duxbury Press, 1979.

[4] G. Box and G. Jenkins. *Time series analysis forecasting and control*. Holden-Day, San Francisco, CA, 1970.

[5] S. Cacopardi, F. Frescura, and G. Reali. Combined OFDM-CDMA configuration for multimedia wireless applications. *IEEE transactions on consumer electronics*, 42(4):865–873, November 1996.

[6] H. Chen and G. J. Pottie. An orthogonal projection-based approach for PAR reduction in OFDM. *IEEE Communications Letters*, 6(5):169–171, May 2002.

[7] K. Choi, K. Kang, and S. Kim. Peak power reduction scheme based on subcarrier scrambling for MC-CDMA systems. *IEE proceedings-communications*, 151(1):39–43, February 2004.

[8] O. Darne, V. Guiraud, and M. Terraza. Forecasts of the seasonal fractional integrated series. *Journal of Forecasting*, 23, issue 1:1–17, 2004.

[9] P. Van Eetvelt, G. Wade, and M. Tomlinson. Peak-to-average power reduction for OFDM schemes by selective scrambling. *Electronics Letters*, 32(21):1963–1964, 1996.

[10] K. Fazel. Performance of convolutionally coded CDMA/OFDM in a frequency-time selective fading channel near-far resistance. In *IEEE international conference on humanity through communications*, volume 3, pages 1438–1442, 1994.

[11] R. Files and V. Kumar. Telecommunications demand forecasting-a review. *International Journal of Forecasting*, 18, issue 4:489–522, 2002.

[12] G. D. Forney. Trellis shaping. *IEEE Transaction on Information Theory*, 38:281–300, March 1992.

[13] M. Friese. Multicarrier modulation with low peak-to-average power ratio. *Electronics Letters*, 32:713–714, April 1996.

[14] M. Friese. Multitone signals with low crest factor. *IEEE Transactions on Communications*, 45(10):1338–1344, 1997.

[15] K. Fujisawa, M. Kojima, K. Nakata, and M. Yamashita. *SDPA User's Manual*. Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, 2004.

[16] Jr. G. D. Forney and L. F. Wei. Multidimensional constellations-part I: Introduction, figures of merit, and generalized cross constellations. *IEEE Journal on Selected Areas in Communications*, 7(6):877–892, August 1989.

[17] A. Gatherer and M. Polley. Controlling clipping probability in DMT transmission. *Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers*, pages 578–584, 1997.

[18] A. G. Gray and P. J. Thomson. On a family of finite moving-average trend filters for the ends of series. *Journal of Forecasting*, 21, issue 2:125–149, 2002.

[19] W. H. Greene. *Econometric Analysis*. Prentice Hall, Inc., 3rd edition, 1997.

[20] N. K. Groschwitz and G. C. Polyzos. A time series model of long-term NSFNET backbone traffic. In *IEEE international conference on communications*, volume 3, pages 1400–1404, 1994.

[21] CDMA Development Group. http://www.cdg.org/technology/index.asp.

[22] C. Guang, G. Jian, and D. Wei. Nonlinear-periodical network traffic behavioral forecast based on seasonal neural network model. In *IEEE international conference on communications, circuits and systems*, volume 1, pages 683–687, 2004.

[23] S. H. Han and J. H. Lee. An overview of peak-to-average power ratio reduction techniques for multicarrier transmission. *IEEE wireless communications*, 12(2):56–65, April 2005.

[24] J. E. Hanke. *Business Forecasting*. Prentice Hall, 7th edition, 2001.

[25] J. W. Hansen and R. D. Nelson. Neural networks and traditional time series methods: a synergistic combination in state economic forecasts. *IEEE Transaction on neural networks*, 8, issue 4:863–873, 1997.

[26] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge, U.K., 1989.

[27] D. I. Harvey and T. C. Mills. Modelling trends in central England temperatures. *Journal of Forecasting*, 22, issue 1:35–47, 2003.

[28] N. Hicheri, M. terre, and B. Fino. OFDM and DS-CDMA approaches, analysis of performances on fading multi-path channels. In *the 13th IEEE international symposium on personal, indoor and mobile radio communications*, volume 4, pages 1498–1501, 2002.

[29] G. Hill and M. Faulkner. Comparison of low complexity clipping algorithms for OFDM. In *The 13th international symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages 227–231, 2002.

[30] W. S. Ho, A. S. Madhukumar, and F. Chin. Peak-to-average power reduction using partial transmit sequences: A suboptimal approach based on dual layered phase sequencing. *IEEE transactions on broadcasting*, 49(2):225–231, June 2003.

[31] C-S Hwang. A peak power reduction method for multicarrier transmission. In *IEEE International Conference on communications*, volume 5, pages 1496–1500, 2001.

[32] N. Hyung and P. H. Franses. Forecasting time series with long memory and level shifts. *Journal of Forecasting*, 24, issue 1:1–16, 2005.

[33] T. Islam and D. G. Fiebig. Modelling the development of supply-restricted telecommunications markets. *Journal of Forecasting*, 20, issue 4:249–264, 2001.

[34] T. Islam, D. G. Fiebig, and N. Meade. Modelling multinational telecommunications demand with limited data. *International Journal of Forecasting*, 18, issue 4:605–624, 2002.

[35] G. Janacek and L. Swift. *Time series: forecasting, simulation, applications*. Ellis Horwood Limited, 1993.

[36] A. E. Jones, T. A. Wilkinson, and S. K. Barton. Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes. *Electronics Letters*, 30(25):2098–2099, 1994.

[37] D. L. Jones. Peak power reduction in OFDM and DMT via active channel modification. In *Asilomar Conference on Signals, Systems, and Computers*, volume 2, pages 1076–1079, 1999.

[38] G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T-C. Lee. *The Theory and Practice of Econometrics*. John Wiley & Sons, 2nd edition, 1980.

[39] D. B. Jun, S. K. Kim, Y. S. Park, M. H. Park, and A. R. Wilson. Forecasting telecommunication service subscribers in substitutive and competitive environments. *International Journal of Forecasting*, 18, issue 4:561–581, 2002.

[40] S. Kaiser and L. Papke. Optimal detection when combining OFDM/CDMA with convolutional coding. In *IEEE international conference on communication*, volume 1, pages 343–348, 1996.

[41] R. Kalman. A new approach to linear filtering and prediction problems. *Transaction of the ASME-Journal of Basic Engineering*, 82:35–45, 1960.

[42] B. Kang, W. Kim, and C. Kim. On the performance of an OFDM-CDMA system under realistic channel conditions. In *IEEE 6th international conference on universal personal communications record*, volume 1, pages 284–288, 1997.

[43] A. K. Khandani and P. Kabal. Shaping multidimensional signal spaces-part I: Optimum shaping, shell mapping. *IEEE transaction on Information Theory*, 39(6):1799–1808, November 1993.

[44] A. K. Khandani and W. Tong. Application of shaping technique to multi-level turbo-coded modulation. Technical Report UW-E&CE# 02-07, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1, 2002.

[45] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[46] G. A. Kochenberger, F. Glover, B. Alidaee, and C. Rego. An unconstrained quadratic binary programming approach to the vertex coloring problem, 2003.

[47] F. Kohandani and A. K. Khandani. Wireless airtime traffic estimation using a state space model. *Communication networks and services research conference*, accepted, 2006.

[48] F. Kohandani, A. K. Khandani, and W. Tong. A low complexity method for peak/average power reduction in OFDM systems. In *Canadian workshop on information theory*, volume 8, pages 131–134, 2003.

[49] F. Kohandani, D. W. McAvoy, and A. K. Khandani. Wireless data traffic estimation using a state space model. *IEEE Transactions on Vehicular Technology*, submitted, 2005.

[50] F. Kohandani, D. W. McAvoy, and A. K. Khandani. Wireless data traffic estimation with kalman filter. Technical Report UW-E&CE# 05-13, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1, 2005.

[51] S. G. Koreisha and Y. Fang. Updating ARMA predictions for temporal aggregates. *Journal of Forecasting*, 23, issue 4:275–296, 2004.

[52] B. S. Krongold and D. L. Jones. An active-set approach for OFDM PAR reduction via tone reservation. *IEEE transactions on signal processing*, 52(2):495–509, February 2004.

[53] H. K. Kwok. *Shape up: Peak-Power Reduction Via Constellation Shaping*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 2001.

[54] E. Lawrey. The suitability of OFDM as a modulation technique for wireless telecommunications, with a CDMA comparison. Technical report, James Cook University, Townsville, QLD, Australia, 1997.

[55] D-W Lim, J-S No, C-W Lim, and H. Chung. A new SLM OFDM scheme with low complexity for PAPR reduction. *IEEE signal processing letters*, 12(2):93–96, February 2005.

[56] M-C Lin, K-C Chen, and S-L Li. Turbo coded OFDM system with peak power reduction. In *The 58th Vehicular Technology Conference*, volume 4, pages 2282–2286, 2003.

[57] H. Liu and S. G. Hall. Creating high frequency national accounts with state space modelling: A monte carlo experiment. *Journal of Forecasting*, 20, issue 6:441–449, 2001.

[58] L. Loyola and T. Miki. An OFDMA/CDMA-based bandwidth resource allocation scheme for future broadband mobile networks. In *IEEE 56th vehicular technology conference*, volume 1, pages 42–46, 2002.

[59] W-K Ma, T. N. Davidson, K. M. Wong, Z-Q Luo, and P-C Ching. Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA. *IEEE transactions on signal processing*, 50(4):912–922, April 2002.

[60] M. H. Magalhaes, R. Ballini, P. Molck, and F. Gomide. Combining forecasts for natural streamflow prediction. In *IEEE annual meeting of the fuzzy information*, volume 1, pages 390–394, 2004.

[61] K. S. Man. Long memory time series and short term forecasts. *International Journal of Forecasting*, 19, issue 3:477–491, 2003.

[62] Adam Marczyk. http://www.talkorigins.org/faqs/genalg/genalg.html.

[63] I. Martoyo, H. Schober, and F. Jondral. CDMA versus OFDM, a performance comparison in selective fading channels. In *IEEE seventh international symposium on spread spectrum techniques*, volume 1, pages 139–143, 2002.

[64] T. May and H. Rohling. Reducing the peak-to-average power ratio in OFDM radio transmission systems. *Proceedings of the IEEE Vehicular Technology Conference*, 3:2474–2478, 1998.

[65] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[66] Z. Michalewicz and D. B. Fogel. *How to solve it: Modern heuristics*. Springer-Verlag, 2000.

[67] I. Miletic, B. Dimitrijevic, and Z. Nicolic. OFDM-CDMA systems with nonlinear power amplifier. In *IEEE wireless communications and networking conference*, volume 3, pages 1167–1171, 1999.

[68] I. Miletic, B. Dimitrijevic, and Z. Nicolic. Performance of OFDM DS/CDMA in fading channels. In *4th international conference on telecommunications in modern satellite, cable and broadcast services*, volume 2, pages 588–591, 1999.

[69] A. Mobasher and A. K. Khandani. Integer-based constellation shaping method for PAPR reduction in OFDM systems. Technical Report UW-E&CE# 05-08, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1, 2005.

[70] A. Mobasher, M. Taherzadeh, R. Sotirov, and A. K. Khandani. A randomization method for quasi maximum likelihood decoding. In *Proceedings of the 9th Canadian workshop on information theory*, pages 135–138, 2005.

[71] S. Muller and J. Huber. OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences. *Proceedings of the IEEE Global Telecommunications Conference*, 33:368–369, February 1997.

[72] K. G. Murty. *Operations research, deterministic optimization models*. Prentice Hall, 1995.

[73] O. Muta and Y. Akaiwa. A peak power reduction scheme based on reversal of parity-bits for a block-coded OFDM signal. In *The 59th vehicular technology conference*, volume 4, pages 1911–1915, 2004.

[74] M. Nayakkankuppam and H. Wolkowicz. http://www-fp.mcs.anl.gov/otc/guide/optweb/continuous/constrained/sdp/, 1996.

[75] R. Van Nee and A. de Wild. Reducing the peak-to-average power ratio of OFDM. *Proceedings of the IEEE Vehicular Technology Conference*, 3:2072–2076, 2002.

[76] M. Nelson, T. Hill, B. Remus, and M. O'Connor. Can neural networks applied to time series forecasting learn seasonal patterns: an empirical investigation. In *IEEE international conference on system sciences*, volume 3, pages 649–655, 1994.

[77] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[78] H. Ochiai. A novel trellis-shaping design with both peak and average power reduction for OFDM systems. *IEEE Transactions on Communications*, 52(11):1916–1926, November 2004.

[79] H. Ochiai and H. Imai. Performance of block codes with peak power reduction for indoor multicarrier systems. *Proceedings of the IEEE Vehicular Technology Conference*, 1:338–342, 1998.

[80] K. Patterson. Generalized Reed-Muller codes and power control in OFDM modulation. *Electronics Letters*, 46:104–120, January 2000.

[81] N. Petersson, A. Johansson, P. Odling, and P.O. Borjesson. A performance bound on PSD-constrained PAR reduction. In *IEEE International conference on communications*, volume 5, pages 3498–3502, 2003.

[82] A. Prochazka. Neural networks and seasonal time-series prediction. In *IEE international conference on artificial neural networks*, volume 440, pages 36–41, 1997.

[83] H-G Ryu, T. P. Hoa, N. T. Hieu, and J. Jianxue. BER analysis of clipping process in the forward link of the OFDM-FDMA communication system. *IEEE transactions on consumer electronics*, 50(4):1058–1064, November 2004.

[84] M. Salvador, J. L. Gallizo, and P. Gargallo. A dynamic principal components analysis based on multivariate matrix normal dynamic linear models. *Journal of Forecasting*, 22, issue 6-7:457–478, 2003.

[85] L. Sarno and G. Valente. Comparing the accuracy of density forecasts from competing models. *Journal of Forecasting*, 23, issue 8:541–557, 2004.

[86] Statistical Sciences, Inc., Seattle, WA. *S-PLUS User's Manual*, 1991.

[87] G. W. Taylor. Smooth transition exponential smoothing. *Journal of Forecasting*, 23, issue 6:385–404, 2004.

[88] J. Tellado and J. Cioffi. *Peak power reduction for multicarrier transmission.* http://www-isl.stanford.edu/people/jtellado/Globecom98notA4.ps.

[89] J. Tellado and J. Cioffi. Efficient algorithms for reducing PAR in multicarrier systems. *Proceedings of the International Symposium on Information Theory*, page 191, 1998.

[90] J. Tellado-Mourelo. *Peak To Average Power Reduction For Multicarrier Reduction*. PhD thesis, University of Stanford, Stanford, California, USA, 1999.

[91] W. Tych, D. J. Pedregal, P. C. Young, and J. Davies. An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system. *International Journal of forecasting*, 18, issue 4:673–695, 2002.

[92] L. Wang and C. Tellambura. A simplified clipping and filtering technique for PAR reduction in OFDM systems. *IEEE signal processing letters*, 12(6):453–456, June 2005.

[93] H. Wolkowicz, R. Saigal, and L. Vanderberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*. Boston, London: Kluwer Academic, 2000.

[94] Y. Wu and W. Y. Zou. Orthogonal frequency division multiplexing: A multicarrier modulation scheme. *IEEE Transactions on Consumer Electronics*, 41(3):392–399, August 1995.

[95] X12-ARIMA. http://www.census.gov/srd/www/x12a/.

[96] Z. Yang, H. Fang, and C. Pan. ACE with frame interleaving scheme to reduce peak-to-average power ratio in OFDM systems. *IEEE Transactions on Broadcasting*, 51(4):571–575, 2005.

[97] P. C. Young. Nonstationary time series analysis and forecasting. *Journal of Progress in Environmental Science*, 1:3–48, 1999.

[98] P. C. Young, D. J. Pedregal, and W. Tych. Dynamic harmonic regression. *Journal of forecasting*, 18, issue 6:369–394, 1999.

[99] P. C. Young, W. Tych, and D. J. Pedregal. Stochastic unobserved component models for adaptive signal extraction and forecasting. In *IEEE signal processing society workshop*, volume 8, pages 234–243, 1998.

[100] H. Zou and Y. Yang. Combining time series models for forecasting. *International Journal of Forecasting*, 20, issue 1:69–84, 2004.