

ABSTRACT

Title of Dissertation: SPATIAL PROCESSING, POWER CONTROL, AND
CHANNEL ALLOCATION FOR OFDM WIRELESS
COMMUNICATIONS

Masoud Olfat, Doctor of Philosophy, 2003

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

OFDM is mainly designed to combat the effect of multipath reception, by dividing the wide-band frequency selective fading channel into many narrow-band flat subchannels. OFDM offers flexibility in adaptation to time-varying channel condition by adopting the parameters at each subcarrier accurately. The purpose of this work is to use this flexibility and study the OFDM systems with power control, multiple transmit and receive antennas, the problem of Peak to Average Power Ratio (PAPR), and the effect of OFDM in providing QoS.

An OFDM uplink multiuser wireless network, combined with power control and receive beamforming is proposed to achieve the desired SINR at each OFDM subchannel. Consequently, better overall BER with the same total power is achieved. To reduce the receiver

complexity, joint time-domain beamforming and power control is also provided. The proposed algorithm is also extended to COFDM.

We use distributed schemes to maximize the maximum achievable data rate for each receiver in a multiuser downlink transmission using MIMO/OFDM, by finding the optimal transmit and receive weight vectors. We propose iterative algorithms to distribute the limited power (per carrier or per user) to multiple streams and multiple antennas in order to maximize the allocated rate per user. The game theoretic analogy of the problem is stated and the convergence of the algorithms are discussed.

To increase the information rate of low PAPR OFDM codes, we propose two frameworks. Super Golay sequences constructed from 16-QAM constellation having PAPR bounded up to 3dB are defined, and constructed by recursive structures. Cyclic-Golay codes are also proposed and constructed by a framework that can be used to obtain the cyclic shift of any code represented by Boolean algebraic functions. These codes are in general a subset of generalized Reed-Muller codes, and have lower error correction capabilities compared to Golay sequences. An extension of the majority logic Reed algorithm for decoding Reed-Muller codes of any order is provided. To reduce decoding complexity, recursive maximum-likelihood decoding schemes are also provided, and the complexity of these algorithms are analyzed.

We also address a scheduling algorithm for wireless networks that provides QoS for mobile users in a shared environment and at the same time utilizes the system resources efficiently. We introduce an income maximization notion, and propose optimal and suboptimal approaches to increase throughput and maintain the QoS for each user, and generate high income for service provider. This notion is used to determine the optimal subcarrier allocation to different users of an OFDMA system based on their required QoS. Optimal and sub-optimal algorithms are presented and their performances and complexities are studied.

SPATIAL PROCESSING, POWER CONTROL, AND
CHANNEL ALLOCATION FOR OFDM WIRELESS COMMUNICATIONS

by

Masoud Olfat

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2003

Advisory Committee:

Professor K. J. Ray Liu, Chairperson/Advisor
Professor Steven Tretter
Professor Carlos Bernestein
Professor Sennur Ulukus
Professor Babis Papadopoulos

© Copyright by

Masoud Olfat

2003

DEDICATION

TO MY PARENTS FOR THEIR ENCOURAGEMENT, GUIDANCE,
SUPPORT, AND INFINITE LOVE,
TO MY LOVELY WIFE MAHNAZ AND MY CHILDREN MAHBOD AND
TARRA WHOSE LOVE IS THE MOTIVATION OF MY BEING ALIVE,
AND TO ALL THOSE WHO HAS CARED FOR ME.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor Professor K.J. Ray Liu for his encouragement, guidance and support. Despite my difficulties, his generous support provided me the opportunity to pursue my ph.D. His patience, and delicate questions and suggestions greatly influenced the performance of my work. His careful comments and criticisms have driven me substantially to achieve more significant contributions. Professor Liu has been more than an academic advisor to me by developing a friendly relationship which is beyond advisor/student relation, and has behaved in such a way that his students are part of his family. I have also had the privilege to use his expertise in teaching high level sophisticated technical courses in a very professional manner.

My sincere gratitude goes to the members of my dissertation committee, Professor Steve Tretter, Professor Carlos Bernestein, Professor Sennur Ulukus, and Professor Babis Papadopoulos for their time and careful consideration of my work and for their great comments and suggestions to improve the performance of my work. Specially, I express my thankfulness to Professor Tretter for his assistance though my studies.

I would like to express my profound gratitude to Professor Hodayoun Hashemi at the Electrical Engineering department of Sharif University of Technology without

whom I have not been able to even pursue my higher educations. His attitude has inspired me of being a better researcher and more importantly a better human.

I am also deeply grateful to my friends Dr. Farrokh Rashid Farrokhi and Dr. Mehdi Alasti who have helped me greatly in developing the ideas and approaches contained in this dissertation. Without their sincere assistances, I would have never been able to achieve this goal.

Special thanks also goes to my friends Professor Alejandra Mercado, Professor Hamid Jafarkhani, Dr. Kamran Etemad, Dr. Javad Razavilar, Dr. Vahid Tabatabaee, Dr. Babak Azimi Sadjadi, Dr. Majid Raissi Dehkordi, Dr. Tahereh Fazel, Dr. Hassan Yaghoobi, and specially Professor Vahid Tarokh for their support, encouragement, and valuable insights, comments and suggestions throughout my ph.D studies.

My sincere gratitude goes to my parents-in-law, my brothers, and my sister who have constantly bestowed me their support, love and encouragements.

Above all, I am profoundly indebted to my parents who made countless personal sacrifices for me and have patiently struggled and suffered from enormous problems I have had throughout my life, and to my beautiful and extremely compassionate wife Mahnaz and my lovely and gracious children Mahbod and Tarra who have been the main motivation of my life. Without all of them, I would have not been who I am. Without their love, my life would have been meaningless.

To all of them and to the future of my beloved country, Iran, I dedicate this dissertation.

TABLE OF CONTENTS

List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Broadband Wireless Communications	1
1.1.1 Standardization and Frequency Bands	2
1.2 Wireless Networks: The Layered Architecture	7
1.3 Physical Layer	8
1.3.1 Modulation Level	8
1.3.2 Interleaving	11
1.3.3 Channel Coding	12
1.4 Multiple Access Control (MAC) Layer	16
1.4.1 MAC Layer in IEEE802.11 WLAN Standard	18
1.5 Network Layer	22
1.6 Quality of Service	24
1.7 Scheduling	25
1.8 Interference and Capacity	25
1.9 Wireless Propagation Models	27

1.9.1	Wireless Channel Model	31
1.10	Multiple Transmit and Receive Antenna	33
1.10.1	Beamforming	36
1.10.2	Space-Time Coding	40
1.10.3	spatial multiplexing	42
1.11	Contribution of this Dissertation	45
1.12	Organization of the Dissertation	47
2	Orthogonal Frequency Division Multiplexing (OFDM)	49
2.1	Motivation for introducing OFDM	49
2.2	OFDM History	51
2.3	Description of OFDM	54
2.3.1	Advantages of OFDM	60
2.3.2	Disadvantages of OFDM	61
2.3.3	Single Carrier versus OFDM Comparison	63
2.4	Loading Algorithms	66
2.5	Peak to Average Power Ratio (PAPR)	69
2.5.1	Statistical Properties of OFDM Signals	72
2.5.2	Techniques for OFDM PAPR Reduction	74
2.6	Orthogonal Frequency Division Multiple Access	79
2.6.1	Channel Allocation	81
3	Power Allocation for OFDM using Adaptive Beamforming over Wireless Networks	84
3.1	Motivation and Previous Works	84
3.2	OFDM with Adaptive Power Control	87

3.2.1	Background	87
3.2.2	System Configuration	90
3.3	Power Control and Frequency-Domain Beamforming	92
3.4	Power Control and Time-Domain Beamforming	97
3.5	Power Control and MMSE Time-Domain Beamforming	102
3.6	Extension to COFDM	105
3.7	Performance Results	110
3.8	Summary of the Chapter	118
4	MIMO-OFDM Systems with Multi-User Interference	120
4.1	Motivation and Previous Works	120
4.2	System Model	123
4.3	Achievable Rate with Known Interference Covariance Matrix	125
4.3.1	Constant Power per Subcarrier	128
4.3.2	Constant power per user	130
4.3.3	Iterative Water-filling	131
4.3.4	Game Theoretic approach for rate maximization	133
4.3.5	Sub-optimal Solution; Same Cell Interference	136
4.4	Single Stream SNR Maximization	138
4.5	Simulation	142
4.6	Summary	147
5	Low Peak to Average Power Ratio with modified Golay Sequences for OFDM Systems	149
5.1	Motivation and Previous Works	149
5.2	Golay Complementary Sequences for equal-power constellations	153

5.2.1	Construction of equal-power Golay Sequences	156
5.2.2	PAPR reduction for the non-equal power constellation	159
5.2.3	Super-Golay 16QAM pairs from QPSK pairs	164
5.2.4	Super-Golay 64QAM pairs from QPSK pairs	166
5.3	Cyclic Golay Sequences	167
5.3.1	Construction of Cyclic Golay Codes	172
5.3.2	Maximum-likelihood Decoding of $RM_{2^h}(r, m)$	184
5.3.3	Performance Results	197
5.3.4	Summary of the Chapter	204
6	Service Level Agreement (SLA) Based Scheduling Algorithms and QoS-provisioned Channel Allocation for OFDMA	207
6.1	Motivation and Previous Works	207
6.2	SLA-Based Scheduling for Wireless Networks	215
6.2.1	System Model	215
6.2.2	Maximum Credit Scheduling (MCS)	217
6.2.3	Maximum Throughput Scheduling (MTS)	219
6.2.4	A Trade-off: SLA-Based Scheduling Algorithms	219
6.2.5	Maximum Income Greedy Scheduling: A Suboptimal Solution	221
6.2.6	Maximum Income Dynamic Programming Scheduling (MIDPS): Optimal Solution	228
6.2.7	System Admission Control Policy, and Pricing	236
6.3	QoS-Provisioned Channel Allocation for OFDMA	237
6.3.1	OFDMA System Model	238
6.3.2	OFDMA Scheduling Algorithms through Subcarrier Allocation	239

6.4	Performance Results	248
6.4.1	Simulation Results for SLA-Based Scheduling	248
6.4.2	Simulation Results for OFDMA Channel Allocation	255
6.5	Summary of the Chapter	259
7	Conclusion and Future Works	262
7.0.1	Summary	262
7.0.2	Future Works	268
	Bibliography	271

LIST OF TABLES

3.1	The COST207 Typical Urban 6-ray power delay profile.	110
5.1	List of repeated Golay sequences under cyclic shifts for $m = 3$ and $h = 2$. . .	179
5.2	List of non-repeated cyclic shifts on Golay sequences for $m = 3$ and $h = 2$. . .	181
5.3	K-Map for 4 variables	185
5.4	K-Map for 5 variables	185
5.5	Computational complexities of Decoders for codes from $RM_4(2, 5)$	196
5.6	Number of constructed 8-valued SGolay Pairs.	197
5.7	Rates and PMEPRs for size 16.	198
5.8	Rates and PMEPRs for size 32	198
5.9	Rate, Hamming and Lee distance of some codes with low PAPR	200

LIST OF FIGURES

1.1	An example of a convolutional encoder.	14
1.2	The MAC Layer in IEEE802.11 as an interface between the PHY and LLC layers.	19
1.3	The IEEE802.11 multiple access structure.	20
1.4	Power fluctuation of the received signal due to different sources of fading	27
1.5	The effect of data rate in ISI.	31
1.6	Narrowband Antenna array and beamformer	35
1.7	Wideband Antenna array and beamformer	36
1.8	Uniform Linear Array	37
2.1	(a) A typical FDM spectrum. (b) A typical OFDM spectrum, where B is the saving in bandwidth.	51
2.2	(a) Spectra of OFDM signal.	53
2.3	Basic Building block of an OFDM transmitter.	54
2.4	Basic Building block of the j th OFDM Receiver.	56
2.5	Partial Transmit Sequences for PAPR reduction in an OFDM.	77
2.6	Selective Mapping for PAPR reduction in an OFDM.	78
2.7	The extension of 16QAM constellation.	79
2.8	OFDMA carrier segmentation.	80
3.1	The i th OFDM transmitter using Adaptive Power Control.	90

3.2	Frequency-Domain Beamforming in the j th OFDM receiver.	92
3.3	Time-Domain Beamforming in the j th OFDM receiver.	98
3.4	The 8-state 8-PSK TCM encoder.	107
3.5	The 8-state 8-PSK TCM trellis.	108
3.6	Bit error rate vs. SINR [dB] for single antenna cases for time-varying channel between power updates.	112
3.7	Bit error rate vs. total network power [dBm] for single antenna cases assuming quasi-static channel.	113
3.8	SINRs of different subchannels in the single antenna cases.	114
3.9	Total network power [dBm] vs. desired SINR [dB] for adaptive power control cases.	115
3.10	Coding gain of the TCM encoder depicted in Fig. 3.5	116
3.11	Total network power vs. desired SINR for coded and uncoded OFDM.	117
3.12	Bit error rate vs. desired SINR for coded and uncoded OFDM.	119
4.1	Cellular structure and distribution of users	142
4.2	Achievable rate CDF, for fixed power per carrier with 16 base stations, 2 mobile per cell, reuse factor of 7 and 2 streams	143
4.3	Achievable rate CDF, for fixed power per carrier with 16 base stations, 2 mobile per cell, reuse factor of 1 and 2 streams	144
4.4	Achievable rate CDF, for fixed power per user with 125 base stations, 1 mobile per cell, reuse factor of 3 and 2 streams	145
4.5	Achievable rate CDF of different number of streams, for fixed power per user with 100 base stations, 2 mobile per cell, and reuse factor of 3.	146
4.6	Achievable rate CDF of different number of streams, for fixed power per user with 16 base stations, 2 mobile per cell, and reuse factor of 3	147

4.7	Achievable rate CDF of different number of streams, for fixed power per user with 16 base stations, 1 mobile per cell, and reuse factor of 3	148
5.1	Bit Error Rate (BER) vs. SNR for AWGN channels	193
5.2	Coding rate vs. SNR for AWGN channels	195
5.3	Bit Error Rate (BER) vs. SNR for Fading channels, when the channel is known at the receiver	199
5.4	Bit Error Rate (BER) vs. SNR for Fading channels, when the channel is not known at the receiver	201
5.5	OFDM transmitter with Golay and cyclic Golay encoder	202
5.6	Bit Error Rate (BER) vs. SNR Threshold for AWGN channels, when $m = 4$	204
5.7	Coding rate vs. SNR Threshold for AWGN channels, when $m = 4$	205
5.8	Bit error rate vs. coding rate for fixed values of SNR for AWGN channels, when $m = 4$	206
6.1	System block diagram	216
6.2	Metric generator block diagram for MIGS	223
6.3	SLA scheduler block diagram	224
6.4	Block diagram of the system.	239
6.5	Scheduling tree of subcarrier allocation.	241
6.6	Viterbi channel assignment.	244
6.7	Throughput versus network load	248
6.8	Minimum assigned relative rate versus network load	249
6.9	Total income vs. network load	250
6.10	Throughput of MIDPS, MIGS, and MPS vs. network load	251
6.11	Minimum assigned relative rate of MIDPS, MIGS, and MPS vs. network load	252

6.12	Income of MIDPS, MIGS, and MPS vs. network load	253
6.13	QoS versus the network load for different penalty functions, $\alpha = 5$	254
6.14	Throughput versus the network load for different penalty functions, $\alpha = 5$	255
6.15	QoS versus the network load for different penalty functions, $\alpha = 500$	256
6.16	Throughput versus the network load for different penalty functions, $\alpha = 500$	257
6.17	CDF's of the optimal exhaustive search algorithm, iterative algorithm with 20 and 80 iterations	258
6.18	Throughput vs. load	259
6.19	Worst case actual to desired throughput vs. load	260
6.20	Total revenue vs. load	261

Chapter 1

Introduction

1.1 Broadband Wireless Communications

In recent years, wireless communication and networking has experienced a rapid growth, and it promises to become a globally important infrastructure. The advances in integrated circuit technology and digital signal processing algorithms have made wireless communication technology accessible to millions of people. The light weight, long operational time, and affordable prices of portable devices have resulted in ever increasing demand for wireless services.

The spectral growth of video, voice, and data communication over the Internet, and equally rapid pervasion of mobile telephony, justify great expectation for mobile multimedia. However, the current wireless communication systems and standards, such as the 3G cellular standard or the WLAN IEEE802.11 standard family, do not fully support the new emerging multimedia applications. The quality of service (QoS) they can provide is not competitive with the QoS wire-line service providers can offer. The large volume and high sensitivity of multimedia data require the development and deployment of wireless communication systems that can guarantee reliable data transmission at high data rates. Research and development

are taking place all over the world to define the next generation of Wireless Broadband Multimedia Communication Systems (WBMCS) consisting of various components at different scales rating from global networks to residential small networks. The demand for wireless mobile, Internet, and multimedia communications is growing exponentially. Therefore it is imperative that wireless, Internet, and multimedia should be brought together. Thus, in the near future, wireless Internet Protocol (IP) and Wireless Asynchronous Transfer Mode (WATM) will play an important role in the development of WBMCS.

While present communication systems are primarily designed for one specific application, such as speech on a mobile telephone or high-rate data in a Wireless Local Area Network (WLAN), the next generation of WBMCS will integrate various functions and applications. Designing wireless communication systems that supporting large data rates with sufficient robustness to radio channel impairments, faces the challenge of devising modulation, coding, and signal processing techniques that can combat the adverse effects of the radio signal propagation environment, such as multipath fading and interference, more effectively.

To implement the wireless broadband communication systems, the following challenges must be considered: Frequency allocation and selection, Channel characterization, Multiple access techniques, Protocol and networks, and finally System development with efficient modulation, coding and smart antenna techniques.

1.1.1 Standardization and Frequency Bands

Inspired by the successful application of the cellular concept, the wireless evolution has so far gone through two generations. First generation (1G) wireless systems (like AMPS, TACS) use analog transmission and support voice services. Second generation (2G) systems (like GSM, IS-95. PDC) employ digital technology and provide circuit-switched data communication services at low speeds in addition to voice. On the other hand, the so-called 2.5G

system (like EDGE/GPRS, HDR), which currently operate in most countries, support more advanced services, such as moderate rate (up to 100 kbps) packet-switched data.

In 1G and 2G systems, the main focus was on increasing system capacity in terms of the number of established connections, which have constant, low rate streams. However, recent evolutions in the telecommunications arena indicate a clear trend towards enhanced, rate-demanding services that are expected to flourish in the next years. The idea of the third generation (3G) systems became evident by the need to support high and diverse data rates for heterogeneous application, such as home-networking, video conferencing, fast wireless/mobile Internet access and multimedia communications. 3G systems, such as UMTS and CDMA2000 are envisioned to support rates in the order of 1 or 2 Mbps [1].

There are several forums for the standardization of wireless broadband systems; namely IEEE802.11 [2], European Telecommunication Standards Institute Broadband Radio Access Networks (ETSI BRAN) [3], Multimedia Mobile Access Communications (MMAC) [4], IEEE802.16/WiMAX [5], and IEEE802.20 [6]. IEEE 802.11 made the first WLAN standard for 2.4 GHZ Industrial, Scientific, and Medical band (ISM), and 5GHZ Unlicensed National Information Infrastructure (UNII) band. The legacy version of WLAN specifies the medium access control and three different physical layers; direct sequence spread spectrum, frequency hopping, and infrared which give a data rate of upto 2Mbps. Later, the committee proposed new versions, namely IEEE802.11b using high speed direct sequence spread sequence physical layer for the speed of 11Mbps, IEEE802.11a with Orthogonal Frequency Division multiplexing (OFDM) for 54Mbps in 5GHZ band, and IEEE802.11g for high the speed of up to 54Mbps in ISM band.

ETSI BRAN and MMAC jointly used OFDM for high speed wireless transmission in 5 GHZ band. ETSI High Performance Local Area Network type 2 (HIPERLAN/2) consists of a family of standards one of which is an OFDM-based standard that is very similar to

IEEE802.11a. MMAC is used in Japan and supports both IEEE802.11a and HIPERLAN/2 standards. Note that Japan has only 100 MHz available in the 5-GHz band, while the United States and Europe provides 300 and 455 MHz, respectively.

Fixed wireless technologies (also called fixed wireless access, wireless broadband access, or broadband wireless access) are not new but because of recent advances, this technology has been successful in rural communities that are out of reach of installed fixed lines. When used at high frequencies, fixed wireless can carry more data but has limited range and requires more complex equipment and line of sight. At lower frequencies, the range is further and the equipment is cheaper, but the transmission rates are low. Multi-point Microwave Distribution Systems (MMDS) and local multi-point distribution systems (LMDS) were viewed as promising technologies, but a lack of uniform standards has hampered their deployment.

IEEE 802.16 and 802.16a are new fixed-wireless standards that should be able to transmit 32-56 km with maximum data rates close to 70 Mbit/s. Again, the higher frequencies require line of sight but it provides high-capacity links. At lower frequencies, line of sight is not required but speeds are lower. This technology is a high-speed wireless backbone designed to link distant ISPs to the Internet. Wireless LAN technologies would then be used for the connection to the user.

The IEEE 802.16 standard is about to revolutionize the broadband wireless access industry. The 802.16 standard, the Air Interface for Fixed Broadband Wireless Access Systems, is also known as the IEEE Wireless-MAN (WMAN) air interface. This technology is designed from the ground up to provide wireless last-mile broadband access in the Metropolitan Area Network (MAN), delivering performance comparable to traditional cable, DSL, or T1 offerings. The principal advantages of systems based on 802.16 are multi-fold: the ability to quickly provision service, even in areas that are hard for wired infrastructure to reach; the avoidance of steep installation costs; and the ability to overcome the physical limitations

of traditional wired infrastructure. Providing a wired broadband connection to a currently underserved area through cable or DSL can be a time-consuming, expensive process, with the result that a surprisingly large number of areas in the US and throughout the world do not have access to broadband connectivity. 802.16 wireless technology provides a flexible, cost-effective, standards-based means of filling existing gaps in broadband coverage, and creating new forms of broadband services not envisioned in a wired world. Drawing on the expertise of hundreds of engineers from the communications industry, the IEEE has established a hierarchy of complementary wireless standards. These include IEEE 802.15 for the Personal Area Network (PAN), 802.11 for the Local Area Network (LAN), 802.16 for the Metropolitan Area Network, and the proposed IEEE 802.20 for the Wide Area Network (WAN). Each standard represents the optimized technology for a distinct market and usage model and is designed to complement the others. A good example is the proliferation of home and business wireless LANs and commercial hot spots based on the IEEE 802.11 standard.

WiMAX (the Worldwide Interoperability for Microwave Access Forum) is a non-profit corporation formed by equipment and component suppliers, including Intel Corporation, to promote the adoption of IEEE 802.16 compliant equipment by operators of broadband wireless access systems. The organization is working to facilitate the deployment of broadband wireless networks based on the IEEE 802.16 standard by helping to ensure the compatibility and interoperability of broadband wireless access equipments.

In an effort to bring interoperability to broadband wireless access, WiMAX is focusing its efforts on establishing a unique subset of baseline features grouped in what is referred to as System Profiles that all compliant equipments must satisfy. These profiles will establish a baseline protocol that allows equipment from multiple vendors to interoperate, and that also provides system integrators and service providers with the ability to purchase equipment from

more than one supplier. System Profiles can address the regulatory spectrum constraints faced by operators in different geographies. WiMAX will establish a structured compliance procedure based upon the proven test methodology results in a complete set of test tools available to equipment developers so they can design in conformance and interoperability during the earliest possible phase of product development. Ultimately, the WiMAX suite will enable service providers to choose from multiple vendors of broadband wireless access equipment that conforms to the IEEE 802.16a standard and that is optimized for their unique operating environment.

By choosing interoperable, standards-based equipment, the operator reduces the risk of deploying broadband wireless access systems. Economies of scale enabled by the standard help reduce monetary risk. Operators are not locked in to a single vendor because base stations will interoperate with subscriber stations from different manufacturers. Ultimately, operators will benefit from lower-cost and higher-performance equipment, as equipment manufacturers rapidly create product innovations based on a common, standards-based platform.

On December 2002, the IEEE Standards Board approved the establishment of IEEE 802.20, the Mobile Broadband Wireless Access (MBWA) Working Group. The mission of IEEE 802.20 is to develop the specification for an efficient packet based air interface that is optimized for the transport of IP based services. The goal is to enable worldwide deployment of affordable, ubiquitous, always-on and interoperable multi-vendor mobile broadband wireless access networks that meet the needs of business and residential end user markets. MBWA Scope is the specification of physical and medium access control layers of an air interface for interoperable mobile broadband wireless access systems, operating in licensed bands below 3.5 GHz, optimized for IP-data transport, with peak data rates per user in excess of 1 Mbps. It supports various vehicular mobility classes up to 250 Km/h in a MAN environment and targets spectral efficiencies, sustained user data rates and numbers of active

users that are all significantly higher than achieved by existing mobile systems.

The 802.20 interface seeks to boost real time data transmission rates in wireless metropolitan area networks to speeds that rival DSL and cable connections (1Mbps or more) based on cell ranges of up to 15 kilometers or more, and it plans to deliver those rates to mobile users even when they are travelling at speeds up to 250 kilometers per hour (155 miles per hour). This would make 802.20 an option for deployment in high-speed trains. The 802.16e project authorization request specifies only that it will support subscriber stations moving at vehicular speeds. Essentially, 802.16e is looking at the mobile user walking around with a PDA or laptop, while 802.20 will address high-speed mobility issues.

1.2 Wireless Networks: The Layered Architecture

The inherent volatility of the wireless medium constitutes the major difficulty in the design of wireless networks. The quality of a narrow-band wireless link between a transmitter and a receiver depends both on radio propagation parameters (path loss, shadow fading, multipath fading) and cochannel interference.

The OSI (Open Systems Interconnection) model defines a layered architecture and the protocols defined in each layer are responsible for communicating with the same peer protocol layer running in the opposite computer, and providing services to the layer above it (except for the top-level application layer). The techniques of layered protocols were developed to logically decompose a complex network into smaller, more understandable parts (layers), to provide standard interfaces between network functions, and to allow each layer to perform the same functions as its counterpart in other nodes of the network,

In the following we will briefly describe the characteristics of the main layers for a wireless networks.

1.3 Physical Layer

Physical layer-based techniques are employed on a link basis, in order to achieve high data rate, while maintaining an acceptable Bit Error Rate (BER) at the receiver, irrespective of link quality. The parameters that are considered as adaptable are modulation and coding, Interleaving, transmission power level, use of multiple antenna, ...

1.3.1 Modulation Level

Modulation is a fundamental component of a digital communications system. It is the process of mapping the digital information to analog form so it can be transmitted over the channel. Consequently every digital communication system has a modulation that performs the task. Closely related to modulation is the inverse process, called demodulation, done by the receiver to recover the transmitted digital information. Modulation is done by changing the amplitude, phase or frequency of the transmitted Radio Frequency (RF) signal. The main design issue of the modulator is the choice of the constellation, which is the set of M points (constellation size) that can be transmitted on a single symbol. This choice affects several important properties of a communication system; for example BER, Peak to Average Power Ratio (PAPR), and RF spectrum shape. Each block of $b = \log_2 M$ bits from the coded bit stream constitutes a symbol and each symbol is mapped to one of M waveforms for transmission over the channel. The single most important parameter for a constellation is the "minimum distance", d_{min} , which is the smallest distance between any two points in the constellation. It depends on several factors; constellation size, average power, and the shape of constellation. The modulation and demodulation can be done either coherently, or non-coherently.

Coherent Modulation Coherent modulation can be used by a communication system that maintains a phase lock between the transmitter and the receiver RF carrier wave. It improves the performance, but requires more complex receiver structure compared to non-coherent systems. The performance gain of coherent modulation is significant when the system uses large constellation. High speed communication systems, like IEEE802.11a are usually coherent. The most common coherent modulations are listed below: (1) Amplitude Shift Keying (ASK), where the information is transmitted by changing the amplitude of the carrier. (2) Phase Shift Keying (PSK) (including QPSK, BPSK 8-PSK) where the information is transmitted by changing the phase of the carrier. (3) M-ary Quadrature Amplitude Modulation (M-QAM), where both amplitude and phase of the carrier change, and is the combination of both ASK and PSK. $b = \log_2 M$ bits are converted to one M-QAM symbol. The symbol error rate of M-QAM constellation is shown to be [7]

$$P_s = 4 \left(1 - \frac{1}{\sqrt{M}} \right) Q \left(\sqrt{\frac{3}{M-1} \frac{E_s}{N_0}} \right), \quad (1.1)$$

where $Q(\cdot)$ is the error function defined as

$$Q(x) = \frac{1}{2\pi} \int_x^\infty e^{-\frac{t^2}{2}} dt, \quad x \geq 0 \quad (1.2)$$

E_s is the symbol energy, and N_0 is the energy of noise. Therefore, the BER of M-QAM depends on the Signal to Noise Ratio (SNR) of the wireless link, and the constellation size. After the receiver has performed all the required synchronization operations, the demodulator tries to detect which of the M symbols has been transmitted.

Non-Coherent Modulations Non-Coherent modulations can be used by a communication system that does not maintain a phase lock between transmitter and receiver, or does not have the knowledge of the amplitude change of the transmitted symbol caused by the channel. This means that the receiver symbols are rotated and scaled arbitrarily compared

to the transmitted symbol. Therefore the ASK, PSK, or QAM modulations cannot be used because they require the received symbol phase and amplitude to be very close to the transmitted phase and amplitude. The solution is to use differential PSK (DPSK) or Differential APSK (DAPSK) modulation. Differential modulations encode the transmitted information to a phase, or phase and amplitude change from one transmitted symbol to the next. This encoding introduces memory to the signal, because transmitted symbol depends on previous symbols. As a consequence, the demodulator has to consider two consecutive symbols when making decisions.

The main benefit of differential encoding is significantly simplified receiver structure. Several of the synchronization algorithms are not needed in a non-coherent receiver. Specifically, phase tracking and channel estimation are not needed, because absolute knowledge of carrier phase and the channel effects is not needed. In an OFDM system (discussed in the next chapter) carrier frequency estimation could also be removed, if the system can tolerate the performance loss due to inter-carrier interference caused by lost orthogonality between the subcarriers. However, many standards do not use this demodulation scheme because of its performance loss associated with differential approaches. In contrast, low data rate systems do use differential techniques, mainly DPSK modulations. In differential Phase Shift Keying (DPSK), the receiver change the carrier phase from its current value according to the data bits.

Differential Amplitude Phase Modulation (DAPSK) combines differential phase and differential amplitude modulation. The differential phase modulation is analogous to the regular DPSK. Differential amplitude modulation, on the other hand, has to change the constellation shape compared to coherent amplitude modulation. The reason is the unknown scaling of the amplitude of the transmitted symbol caused by the channel. A general assumption when differential modulation is used is that the channel and carrier phase are constant during

two consecutive symbols. Therefore, we can cancel the effect of the channel by dividing two consecutive symbols.

The detection of differential modulation is done in two steps. First, the differential encoding is removed from the signal and then a normal demodulation is performed as is done for regular PSK or QAM constellation.

The performance loss of DPSK compared to coherent modulation varies with the size of the modulation; for DBPSK it is 1-2dB, for DPSK about 2.3dB, and for large constellations 3dB.

1.3.2 Interleaving

Interleaving aims to distribute transmitted bits in time or frequency or both to achieve desirable bit error distribution after demodulation. What constitutes a desirable error distribution depends on the used Forward Error Correction (FEC) code. What kind of interleaving pattern is needed depends on the channel characteristics. If the system operates in a purely AWGN environment, no interleaving is needed, because the error distribution cannot be changed by relocating the bits.

Interleaving necessarily introduces delay into the system because bits are not received in the same order as the information source transmits them. The overall communication system usually dictates some maximum delay the system can tolerate, hence restricting the amount of interleaving than can be used. For example, cellular telephone systems usually use time diversity, because the channels are fast fading (discussed later in this chapter). However, the maximum phone to phone delay is usually constrained to 20ms or less, to prevent noticeable degradation in cell quality. This means the maximum interleaving delay must be much less than 20ms to allow for other delay sources in the system.

There are two ways to perform interleaving; Block interleaving and Convolutional inter-

leaving. Block interleaving operates on one block of bits at a time. The number of bits in the block is called interleaving depth, which defines the delay introduced by interleaving. It can be described as a matrix to which data is written in columns and read in rows, or vice versa. Deinterleaving is the opposite operation of interleaving; that is, the bits are put into the original order.

A Convolutional interleaver is another possible interleaving solution that is most suitable for systems that operate on continuous stream of bits. The interleaver operates by writing the bits into a commutator on the left, and reading bits out from the commutator on the right. The main benefit of a convolutional interleaver is that it requires approximately half of the memory required by a block interleaver to achieve the same interleaving depth. This saving can be significant for long interleaver depth. Deinterleaving of convolutional interleaver is achieved by flipping the interleaver along its horizontal axis. The structure is otherwise identical to the interleaver except the longest delay line is at the top, and the no-delay line is last.

In IEEE802.11a has an interleaver depth of one OFDM symbol, because the channel is assumed to be quasi-static; that is, the channel is assumed to stay essentially the same for a duration of a transmitted packet. Therefore it is naturally a block interleaver.

1.3.3 Channel Coding

Channel codes are the most important component of any modern communication system, and they make today's effective and reliable communications possible. The basic measure of channel coding performance is "*coding gain*", which is usually measured in dBs as the reduction of required SNR to achieve a certain symbol error rate in AWGN channel. As an example, IEEE802.11a uses two methods to achieve a 12Mbits/s data rate. The simplest way would be to use uncoded BPSK modulation on each OFDM subcarrier (48 bits worth of

information in each OFDM block). The symbol time is $4\mu s$ or 250000 symbols per second, hence the overall data rate is $250000 \times 48 = 12Mbits/s$.

Another way is to use QPSK and rate 1/2 convolutional codes. This results in a significantly lower required SNR to achieve a good BER performance. At BER of 10^{-5} , the coding gain is about $5.8dB$. This means that to achieve the same performance, the system that does not use channel coding has to spend $5.8dB$ more energy for each transmitted symbol than the system that uses channel coding.

Another important parameter of channel coding is the "coding rate". Code rate is the ratio of bits arrived at the encoder, called the "message word", to the bits exited from the encoder, called the "code word". This ratio is always less than or equal to one. Channel coding always forces the system to use a larger constellation to keep the same data rate as an uncoded system. However, going to larger constellations reduces the "minimum distance of the code", d_{min} ; this implies higher BER at the output of demodulator. However, at the output of channel decoder, the bit error rate is significantly reduced. There are two main types of channel coding; namely "block coding" and "convolutional coding", which will be briefly described next. Note that the performance of channel codes is ultimately limited by the channel capacity formula:

$$C = W \log_2(1 + SNR), \quad (1.3)$$

where W is the channel bandwidth. However, after about 50 years of research, "Turbo codes" [8] have finally emerged as a class of codes that can approach the ultimate limit in performance. Another innovation are Low Density Parity Check (LDPC) codes [9], which also have performance very close to the capacity.

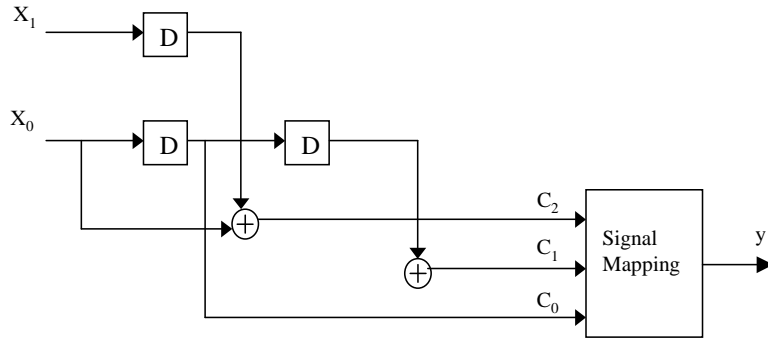


Figure 1.1: An example of a convolutional encoder.

Convolutional Codes

Almost all the major cellular systems (GSM, IS-95), IEEE802.11a and HIPERLAN/2 WLAN standards, and many other standards use convolutional error correcting codes. A convolutional code is defined by a set of connections between stages of one or more shift registers and the output bits of the encoder. If the number of shift registers is k , and the number of output bits is n , the coding rate is $\frac{k}{n}$. For each output bit there are k connections that define how the value of the output bit is calculated from the state of the shift register. The *constraint length* of a convolutional code is the maximum number of bits in a single output stream that can be affected by any input bit. It is the maximum number of taps on the shift registers in the encoder plus one, i.e.

$$K = 1 + \max_i \{m_i\}, \tag{1.4}$$

where m_i the number of shift registers in each branch. Fig. 1.1 depicts a sample convolutional encoder, whose rate is $2/3$, and whose constraint length is 3. The longer the shift registers, the more powerful the code is, but the more complexity is incurred on the decoder.

The performance of a convolutional code is determined by the "*minimum free distance*" of the code. Free distance is defined using the Hamming distance that is equal to the number of

positions in which two codewords are different, and for convolutional codes is the minimum Hamming distance between two different codewords. An asymptotic coding gain at high SNR for a convolutional code can be calculated from the free distance and the rate of the code:

$$\text{coding gain} = 10 \log_{10}(\text{rate} \times \text{free distance}). \quad (1.5)$$

Puncturing [10] is a very useful technique to generate different rates from a single convolutional code. The basic idea behind Rate Compatible Puncturing Convolutional Codes (RCPC) is to avoid transmitting some of the bits output by the convolutional encoder, thus increasing the rate of the code. This increase in rate decreases the free distance of the code.

There are several algorithms that can be used to decode convolutional codes. Viterbi algorithm has reached a dominant position as the method to decode convolutional codes especially in wireless communication. It is a maximum likelihood codeword estimator; it provides the best possible estimate of the transmitted codeword.

Trellis Coded Modulation (TCM) [11] merges channel coding and modulation into a single integrated component. The benefit of this approach is that the code design is optimized for the used constellation. The most significant benefit is reached in AWGN channel and with high spectral efficiencies; in other words with large constellation. It consists of two parts; a convolutional encoder and a modulator.

Block Codes

Block codes are different from convolutional codes in the sense that the code has a definite codeword length nR , instead of variable code word length like convolutional codes. The most popular class of block codes are Cyclic block codes like Reed-Solomon (RS) codes [12], and BCH codes [13]. Another important difference between block codes and convolutional codes is that the block codes are designed using algebraic properties of polynomials or curves over

finite fields, whereas convolutional codes are designed using exhaustive computer search.

Other Codes

Concatenated codes are built by combining an outer code and an inner code. The outer code is usually a Reed-Solomon block code and the inner code a convolutional code. They have reached performances that is only $2.2dB$ from the channel capacity limit.

Turbo codes [14] have a performance only $0.6dB$ from the channel capacity. They are a combination of recursive systematic convolutional codes, interleaving and iterative decoding.

1.4 Multiple Access Control (MAC) Layer

Multiple Access schemes allow the wireless systems to accommodate several users which need to access a common wireless channel, so that the channel is shared efficiently among them. The users could be distinguished either by time, frequency, code, or space. The multiple access methods could be either connection-oriented (fixed assignment), connectionless (random access), or on demand assignment methods. These methods are characterized by a trade-off between overhead they incur and the reliability of transmission.

Connection-oriented multiple access methods are similar to telephone calls, where the connection is dedicated to the call after being established, even if there is nothing to talk about. In this case a separate connection is created for each user and maintained for the duration of the session, even if the user has no more data to be transmitted. The main connection-oriented methods are (1) Frequency Division Multiple Access (FDMA), where the whole bandwidth is divided into non-overlapping carrier frequencies and each user is assigned to one carrier. A special case of FDMA is Orthogonal Frequency Division Multiple Access (OFDMA) in which the subcarriers of an OFDM transmitter are divided into fixed

segments and each segment is assigned to one user. (2) Time Division Multiple Access (TDMA), where a time window is divided into short slots and each slot is assigned to one user. (3) Code Division Multiple Access (CDMA), in which each user transmits all the time over all available frequency band, but is identified by a unique spreading code which is orthogonal to other users' codes to ensure that the receiver could distinguish the user. The code modulates the data and spans it over a wider frequency band as its code rate is much higher than that of the data. (4) Frequency Hopping Multiple Access (FHMA), where carrier frequency of different users are varied in a random fashion. (5) Space Division Multiple Access (SDMA), where the separation of the users is performed in space by directing the emitted energy towards each intended user through directional beams created by multiple antenna arrays.

The connectionless multiple access methods are suitable for low traffic networks, where the streams to be transmitted could be bursty. In Carrier Sense Multiple Access (CSMA) methods, the user listens to media before transmission and tries to capture the control of the channel before transmission. In ALOHA, a user transmits with a certain probability, whenever it has data. These methods reduce the amount of overhead, but increase the risk of collision and interference.

In Demand assignment techniques, the system switches between the random access and fixed assignment methods based on the network load. In heavy loads, the connection-oriented methods are used, while in low traffic the connectionless ones are exploited.

However, the wireless medium has special properties that make the design of MAC protocols different from, and more challenging than, wireline networks, i.e. (1) Collision detection is not possible while sending data and so Ethernet-like protocols cannot be used. (2) Time varying Channel and multipath propagation, necessitates the handshaking between two nodes to test the wireless channel between them. (3) Errors are more likely in wireless

transmissions compared to wireline. Packet loss due to burst errors can be minimized by using either smaller packets, or Forward Error Correcting (FEC) codes, or retransmission methods by using Acknowledgments (ACK) and NACK for detecting packet errors. (4) Carrier sensing is a function of the position of the receiver relative to the transmitter. Hidden node problem is an example of such phenomena.

The metrics that are used to compare the MAC protocols, which can be specified as Quality of Service (QoS) parameters are *delay*, *jitter*, *throughput* as a fraction of channel capacity, *fairness* in sharing the bandwidth among users (by considering the possible priorities), *power consumption*, *Robustness* against channel fading, *multi-stream support* to be able to handle different streams like voice, video, and data which have different requirements, and *stability*.

1.4.1 MAC Layer in IEEE802.11 WLAN Standard

The MAC Layer forms layer 2 as compared with the Open System Interconnection (OSI) Model. The MAC acts as an interface to Logical Link Control Layer (LLC) and Physical layer (PHY) as shown in the Fig. 1.2. The primary function of the MAC Layer is to provide medium access control to applications that contend for medium in such a way as to maximize the utilization of the channel. The MAC layer will also provide the power management and synchronization. The MAC handles three types of messages: data, control, management.

IEEE 802.11 specifies services that are used to support the delivery of MAC Service Data Unit (MSDU) (which are the packets received by the MAC layer from the upper layers) between Stations (STA), and to control WLAN medium access. These services are listed as: (1) Data exchange services that provides reliable transmission of MSDUs between two peer LLC entities, including broadcast and multicast transports. (2) Control services; this service provides handshaking between MAC entities indicating the availability of wireless medium

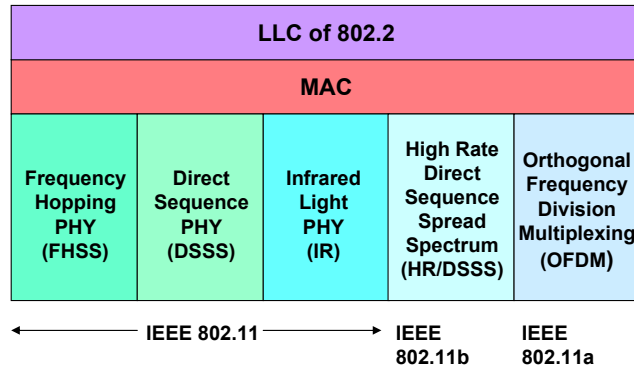


Figure 1.2: The MAC Layer in IEEE802.11 as an interface between the PHY and LLC layers.

(WM) for their communication. The handshaking happens by the exchange of control messages between two entities. The requesting entity sends a control frame Request to send (RTS), asking for confirmation from peer entity and gets an acknowledgement Clear to send (CTS), as a confirmation. This way it sets the environment for reliable communication between two peer MAC entities. (3) Management services; this service provides the services for all management messages like authentication, de-authentication, association, disassociation, re-association, timing and synchronization.

The multiple access scheme that helps the WLAN users to share wireless medium is described in Fig. 1.3. The basic access mechanism, called Distributed Coordination Function (DCF), is a Carrier Sense Multiple Access with Collision Avoidance mechanism (usually known as CSMA/CA). The MAC Layer also incorporates an optional access method called Point Coordination Function (PCF). Both methods are mutually exclusive and operate in different time frames viz. Contention Period (CP) and Contention Free Period (CFP).

- Contention Period (CP); This is the time frame allocated for DCF access mechanism where all the stations (STAs) and Access Point (AP) contend for the wireless medium by control information exchange.

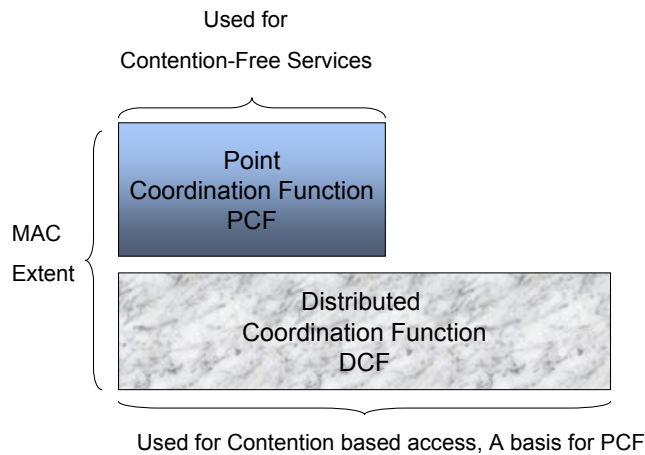


Figure 1.3: The IEEE802.11 multiple access structure.

- Contention Free Period (CFP); This is the time frame allocated for PCF access mechanism where AP becomes the Point Coordinator (PC) or polling master and has the full control of allocating the wireless medium access to different STAs.

Distributed Coordination Function (DCF)

The DCF is implemented in all STAs, for use within all 802.11 network configurations. The basic medium access protocol is a DCF that allows for automatic medium sharing between compatible PHYs through the use of CSMA/CA and a random backoff time following a busy medium condition. In addition, all directed traffic uses immediate positive acknowledgment (ACK frame) where the sender schedules retransmission, if no ACK is received.

CSMA/CA Concept: The CSMA/CA protocol is designed to reduce the collision probability between multiple STAs accessing a medium, at the point where collisions would most likely occur. Just after the medium becomes idle following a busy period is when the highest probability of a collision exists. This is because multiple STAs could have been waiting for the medium to become available again. This is the situation that necessitates a

random backoff procedure (explained later) to resolve medium contention conflicts.

Virtual CSMA concept and Network Allocation vector (NAV): Carrier sense is performed both physically (by PHY) and virtually (by MAC) mechanisms. The virtual carrier-sense mechanism is achieved by distributing reservation information announcing the impending use of the medium. This information is maintained by each STA in the NAV. The exchange of RTS and CTS frames prior to the actual data frame is one means of distribution of the medium reservation information. The RTS and CTS frames contain a duration-ID field that defines the period of time that the medium is to be reserved to transmit the actual data frame and the returning ACK frame. All STAs within the reception range of either the originating STA (which transmits the RTS) or the destination STA (which transmits the CTS) shall update the NAV with this information. Another means of distributing the medium reservation information is the duration-ID field in directed frames (Viz. MAC Headers). This field gives the time that the medium is reserved, either to the end of the immediately following ACK, or in the case of a fragment sequence, to the end of the ACK following the next fragment. The duration information is also available in the MAC headers of all frames sent during the Contention Period (CP) other than Power Save (PS)-Poll Control frames. The NAV may be thought of as a counter, which counts down to zero at a uniform rate. When the counter is zero, the virtual carrier-sense indication is that the medium is idle; This is as shown in the Figure 8. The medium is determined to be busy whenever the STA is transmitting. The time interval between frames is called the Inter Frame Space (IFS). A STA will determine that the medium is idle through the use of the carrier-sense function for the interval specified. There are two main IFS used in DCF, i.e SIFS which is the shortest inter frame spaces. SIFS will be used when STAs have seized the medium and need to keep it for the duration of the frame exchange sequence to be performed. Using the smallest gap between transmissions within the frame exchange sequence prevents other

STAs, which are required to wait for the medium to be idle for a longer gap, from attempting to use the medium. DIFS: A STA using the DCF will be allowed to transmit if its carrier sense mechanism determines that the medium is idle. A STA may transmit after subsequent reception of an errorfree frame, re-synchronizing the STA.

Random Backoff Procedure: Whenever a STA wants to acquire the wireless medium, it checks the state of the medium, as indicated by Physical and Virtual carrier sense mechanism starting from NAV=0. Just after the medium becomes idle following a busy period, is when the highest probability of a collision exists as all the stations find the medium to be idle at the same time. Now each station will wait for a random time duration (back-off procedure) before contending for the medium.

Point Coordination Function (PCF)

The MAC layer software also incorporate an optional access method called PCF which is only usable on infrastructure network configurations. This access method uses a point coordinator (PC), which operates at the Access Point, to determine which STA currently has the right to transmit. The operation is essentially that of polling, with the PC performing the role of the polling master. The PCF distributes information within Beacon management frames to gain control of the medium by setting the network allocation vector (NAV) in STAs.

1.5 Network Layer

Some of the main issues in wireless ad-hoc networks are the *Routing*, and *Traffic Engineering*. The routing issue is a central function in any communication network. The routing protocols meant for wired networks can not be used for mobile ad hoc networks because of the mobility nature of the wireless networks ad hoc networks. These routing protocols can be divided

into two categories: proactive (table-driven) and reactive (on-demand) routings based on when and how the routes are discovered. In table-driven routing protocols each terminal maintains one or more tables containing routing information to every other terminal in the network. All terminals update their tables so as to maintain a consistent and up-to-date view of the network. In reactive routing, the routes are created only when desired by the source host. When the network topology changes, the terminals propagate update messages throughout the network in order to maintain consistent and up-to-date routing information about the whole network.

Traffic engineering (TE) is also a powerful approach for providing quality of service (QoS) over packet networks. The motivation for TE is to distribute traffic flows over the network links to avoid the congestion caused by uneven network utilization [15]. One way to achieve this goal is to use QoS based routing. Given QoS request of a flow or an aggregation of flows, QoS routing obtains the route that is most likely able to meet the QoS requirements [16, 17]. In wired networks, where the network topologies and link capacities are fixed, traffic engineering with QoS based routing finds a good distribution for traffic flows to support the requested qualities [15].

Mobile ad-hoc networks are different from wired networks in the sense that network topology and link capacities vary over time. This causes the TE process to be more complicated. In wireless networks, lack of QoS can be caused by either high bit error rate (BER) on wireless links (low signal to noise and interference ratio) or the congestion of uneven distribution of traffic flows, when some parts of the network could be overloaded while other parts are lightly loaded. Therefore, unlike the wired networks, where the knowledge of overall network traffic was enough to perform traffic engineering, in wireless networks we need to know the link capacities, as well [18]. Similar to wired networks, a QoS based routing algorithm for wireless networks might come up with a longer but lightly loaded route compared to a

heavily loaded shortest path.

1.6 Quality of Service

The primary goal of a wireless communications system is the fulfilment of Quality of Service (QoS) requirements. The QoS parameters vary depending on the network structure and the communication layers. For single user transmission, this could be the physical layer parameters like an acceptable Signal to Noise Ratio (SNR) level or Bit Error Rate (BER) at the receiver. In Data Link Layer or MAC layer it could be expressed as Packet Error Rate (PER), or minimum achievable rate, or maximum tolerable delay guaranteed to each user. QoS could also be interpreted as throughput, delay, and jitter in a session based, or even fairness in rate allocation to different users (for one-to-many transmission).

The ability of the network infrastructure to satisfy such QoS requirements and ultimately enhance system capacity depends on procedures and mechanisms which span several communication layers. For shared media, an efficient multiple access mechanism must be employed. At the MAC layer, QoS could be guaranteed by appropriate scheduling strategies, as well as resource management and reuse methods. At the physical layer, adaptive transmission techniques provide the potential to adjust parameters such as transmission power, modulation level, and symbol rate to maintain acceptable quality on a link level. The employment of multiple transmit and/or receive antenna (both in downlink and uplink) is one of the most important means for increasing capacity, and therefore making the job of providing QoS guarantees more efficient.

1.7 Scheduling

One of the main characteristics that make the packet scheduling over mobile wireless networks distinct from wireline networks is the fact that the wireless channel is time-varying due to multipath fading. In delay tolerant data systems it is thus possible, with the aid of channel condition feedback from the users, to schedule transmission to users when their fading conditions are favorable thereby achieving multi-user diversity [19]. They have shown that with single transmit antenna, transmitting to a single best user during each scheduling interval, is an efficient strategy the performance of any scheduling algorithm critically depends on the transmission rates achieved in each scheduling interval which in turn depends on coding, modulation, and number of transmit antenna employed.

1.8 Interference and Capacity

Interference is one of the major factors that limits the performance of wireless networks. Interference in a receiver can be caused by any communication system which leaks energy into the frequency band the receiver is working in. Interference can degrade Signal to Interference and Noise Ratio (SINR) and therefore causes higher error probability and possibly termination of a transmission session. Interference has been recognized as a major bottleneck in increasing the capacity of a cellular system. In a cellular system the interference can be categorized as co-channel or adjacent channel.

To exploit the limited bandwidth, it is possible that in a given coverage area several cells use the same set of frequency channels. The interference between signals from these cells creates the cochannel interference. Unlike thermal noise which can be overcome by increasing the transmitter power, cochannel interference cannot be compensated by simply increasing the transmitter power. This is due to the fact that although an increase in one

transmitter's power can increase its SINR, it could increase the interference on other cells. To reduce cochannel interference, we can either physically separate the cochannel cells by a minimum distance, or use some signal processing means for interference cancellation, such as power control, or multiple receive antenna beamforming.

In almost all wireless networks, the amount of transmission power is a key performance issue, for two reasons. First, the lower the power, the less interference it causes on other receivers, and therefore a better link quality is achieved. Second, the limited battery life of mobile units forces us to be conservative in the amount of transmission power. On the other hand, a mobile must transmit enough power such that the SINR at its corresponding receiver is in an acceptable level. As a result, power control is a key element of such wireless systems that require shared bandwidth and time slots (cochannel transmission), like CDMA wireless networks [20–22].

As the demand for wireless services increases, the number of channels assigned to a cell (in cellular systems) or service sets (in WLAN) eventually becomes insufficient to support the required number of users. At this point, careful designs are needed to accommodate more users using the limited resources. There are several techniques in cellular systems to increase the capacity, like cell splitting (dividing congested cells into smaller cells), and cell sectoring (using multiple directional antenna at the base station). However these methods suffer from increasing the handoff rates. Moreover, the directional antennas are fixed and cannot adjust themselves to the changing environment. Therefore, we can employ adaptive beamforming techniques, rather than fixed antenna patterns. Array processing is a powerful technique which calls for replacing the single omni-directional antenna or directional sector with an array of antenna elements at the base station. Using adaptive beamforming techniques, one can place an antenna beam toward the and place antenna nulls toward other cochannel interference sources. This results in huge reduction in interference in the received signal and

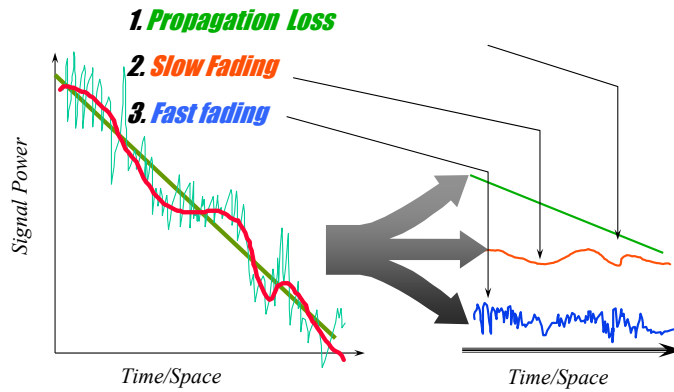


Figure 1.4: Power fluctuation of the received signal due to different sources of fading .

increases significantly the SINR for the signal of interest. This technique is called space diversity combining method. we will discuss the increase in system capacity using different multiple transmit and receive antenna structure in Section 1.10.

1.9 Wireless Propagation Models

A signal transmitted in a wireless channel will experience three separable effects: *path loss*, *shadow fading*, and *multipath fading*. Path loss, or the mean propagation loss, comes from wave propagation, absorption, and vertical multipath. Shadow fading, or slow fading, is caused by large obstacles, such as buildings or hills, and is characterized by log-normal distribution. Multipath fading or fast fading, results from multipath scattering. If there is a non-fading direct path component it is modelled by Ricean distribution; otherwise it is modelled by Rayleigh distribution.

Figure 1.4 illustrates the power fluctuation of the received signal due to different sources of fading around the mean power.

Path Loss: In free space the received power decays as a logarithmic function of the transmitter-receiver separation. The power received by a receiver antenna, in a distance d

from the transmitter antenna, in free space is given by [20]

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} \quad (1.6)$$

where P_t is the transmitter power, G_t is the transmitter antenna gain, G_r is the receiver antenna gain, d is the transmitter-receiver separation distance in meters, and λ is the wavelength in meters. This implies that the received power decays at a rate of 20 dB/decade with distance.

The Path Loss (PL) is defined as the difference between the effective transmitted power and the received power in dB, and is for free space is given by

$$PL = 10 \log_{10} \frac{P_t}{P_r} = -10 \log_{10} \frac{G_t G_r \lambda^2}{(4\pi)^2 d^2} \quad (1.7)$$

Although Eq. (1.6) applies only to free space, in other environments the received power decays by the distance raised to some power n , i.e.

$$P_r(d) = \frac{P_r(d_0)}{(d/d_0)^n}, \quad (1.8)$$

where $P_r(d_0)$ is the received power at the reference distance d_0 . Eq. (1.8) states that the path loss is always proportional to an exponent n of d/d_0 . The value of n depends on the specific propagation environment and it ranges from 2 to 5. Path loss is also referred to as large-scale fading.

Shadow Fading: Slow fading known also as log-normal shadow fading is a result of diffraction and shadowing of the transmitted signal caused by a large object such as buildings, hills, cars, mountains, or other terrain configurations in mobile wireless environment. The shadow fading follows a log-normal distribution (or Gaussian in dB) with mean zero and variance σ^2 . Accordingly, if K shows the effect of such fading, Eq. (1.8) is changed to

$$P_r(d) = K \frac{P_r(d_0)}{(d/d_0)^n}. \quad (1.9)$$

The values of n and σ are computed from measured data, using linear regression and mean square error methods. Their typical value in urban cellular wireless environments is $n = 2.7$ and $\sigma = 11.8$ dB [20].

Multipath Fading: In mobile radio communications, fading occurs due to reflection from scatterers, or bigger objects. This causes the receivers to receive a number of copies of the transmitted signal which have been reflected and diffracted by buildings and other urban obstacles. When the signals from various paths sum constructively at the BS antenna, the received signal level is enhanced. A serious condition occurs when the multipath signals, i.e. the transmitted signal arriving via many paths, effectively sum to a small value. When this happens the received signal is said to be in a fade and the phenomenon is called *multipath fading*. The delay spread of the channel may be considered as the length of the received pulse when an impulse is transmitted through the channel [7, 20].

A scatterer is a small object that reflects a wireless signal. If there is no object in the line of sight from the transmitter to the receiver, Line of Sight (LOS) might exist, too. If scatterers and the mobile move in relative to each other, the received power fluctuates; that is the wireless channel causes time varying fading. This variation in the channel response gives rise to random frequency modulation due to the different Doppler shifts in each path. Because of the inherent randomness in the phase, amplitude and time delay of the different multipath components, they could be combined at the receiver either constructively or destructively. The scatterer could be local to the transmitter, local to the receiver, or far from both of them. Local scatterers along with the movement of the mobile causes Doppler spread, or time selective fading. The delay spread due to local scattering is negligible. Local scatterers to the receiver cause angle spread or space selective fading. Both angle and delay spread could be caused by remote scatterers; they cause frequency selective fading.

The impulse response of a frequency selective fading channel has a multipath delay spread

that is greater than the time duration of the transmitted signal waveform, i.e., $\tau \gg T_s$. Signal transmitted through a wireless channel undergoes frequency selective fading if $T_s \ll \tau$. Depending on the signal bandwidth, symbol period, channel delay and Doppler spread, the signal may experience different types of fading. If we transmit data at a slow rate, the data can easily be resolved at the receiver. Causing the fading to be frequency non-selective or flat. This is because the extension of a data pulse due to the multipath is completed before the next impulse is transmitted. However, if we increase the data transmission rate, a point will be reached where each data symbol significantly spreads into adjacent symbols, a phenomenon known as Inter-Symbol Interference (ISI). This phenomena is illustrated in Fig. 1.5. ISI without equalization results in very high Bit Error Rate (BER). In the frequency domain, this means certain frequency components in the received signal spectrum have greater gains than others. In other words, some frequency components of the transmitted signal is attenuated severely in frequency domain. Frequency selective fading channels are much more difficult to model than flat fading channels. For frequency selective fading, the spectrum $S(f)$ of the transmitted signal has a bandwidth which is greater than the channel bandwidth B_C . The channel becomes frequency selective, if the wireless channel gain is different for different frequency components. Frequency selective fading is caused by multipath delays which approach or exceed the symbol period of the transmitted signal.

Frequency selective channels are also known as wideband channels since the bandwidth of the signal $s(t)$ is wider than the bandwidth of the channel impulse response $h(t, \tau)$. As time varies, the channel varies in gain and phase across the spectrum of $s(t)$, resulting in time varying distortion in the received signal $r(t)$.

For flat fading the frequency components of the signal stay unchanged, while the time domain signal undergoes variation. Most commonly, the amplitude is modelled according to

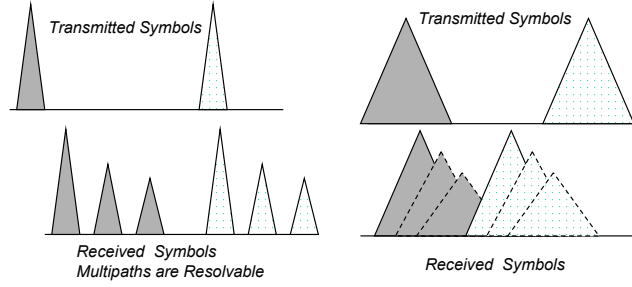


Figure 1.5: The effect of data rate in ISI.

a Rayleigh distribution. The pdf of a Rayleigh distributed random variable x is

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{\sigma^2}\right). \quad (1.10)$$

If the Doppler spread is larger than the signal bandwidth, the channel impulse response will vary rapidly during the symbol period, and the signal will undergo a fast fading. If the Doppler spread is much less than the bandwidth of the signal, the channel becomes static during the signal period. In this case the signal undergoes slow fading.

The amplitude of the received signal in frequency selective fading depends on whether there is a non-fading LOS or not. If there is not such a link, the amplitude is Rayleigh distributed (Eq. (1.10)), otherwise it has a Ricean distribution

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2+A^2}{2\sigma^2}} I_0\left(\frac{Ax}{\sigma^2}\right) u(x), \quad (1.11)$$

where I_0 is the Bessel function of first kind and zero-order, A denotes the peak amplitude of the dominant signal, and $u(x)$ is the step function. As the amplitude of the dominant path decreases, the Ricean distribution approaches to a Rayleigh distribution.

1.9.1 Wireless Channel Model

Let the transmitted signal be

$$s(t) = e^{j2\pi f_c t} \sum_n b_n g(t - nT), \quad (1.12)$$

where f_c is the carrier frequency, b_n is the n^{th} transmitted symbol, T is the symbol duration, and $g(\cdot)$ is the pulse shaping waveform. An example of pulse shaping waveforms is the rectangular pulse. Another example is the square root raised cosine pulse, given by

$$g(t) = \frac{\sin(\pi t)/T}{\pi/T} \left(\frac{\cos(\alpha\pi/T)}{1 - (2\alpha t/T)^2} \right), \quad (1.13)$$

where α represents the excess time beyond T to avoid ISI. The received signal $\tilde{r}(t)$ is:

$$\tilde{r}(t) = \sqrt{\rho(t)} \sum_n \sum_{l=1}^L \sqrt{G_l \alpha_l} b_n g(t - \tau_l - nT) e^{j[2\pi(f_c + f_d \cos \phi_l)t - 2\pi f \tau_l]} + \tilde{n}(t), \quad (1.14)$$

where v is the speed of mobile, ϕ_l is direction from the l^{th} propagation path, $f_d = \frac{v}{\lambda}$ is the maximum Doppler frequency, λ is the wavelength, L is the number of paths, G_l represents the path loss, α_l is the l^{th} path fading, τ_l is the propagation path delay, $\rho(t)$ is the shadow fading component, and $\tilde{n}(t)$ is the thermal noise.

The delay spread is $\tau = \max \tau_l - \min \tau_l$. In the small spread case, τ is much smaller than the symbol period, or $\tau \ll \frac{1}{B}$, where B is the bandwidth of the signal. In this case, different replicas of the pulse shaping waveform could be considered the same and therefore the baseband equivalent of 1.14 can be approximated by

$$\tilde{r}(t) = \sqrt{\rho(t)G} e^{-j2\pi f \tau_0} \sum_n b_n g(t - \tau_0 - nT) \sum_{l=1}^L \sqrt{\alpha_l} e^{j[2\pi f_d \cos \phi_l t]} + \tilde{n}(t), \quad (1.15)$$

where τ_0 is an approximation for the delay, and we have assumed the same path loss for all paths. From (1.14), the impulse response of a wireless wideband channel, $h(\tau, t)$, is represented by

$$h(\tau, t) = \sqrt{\rho(t)} \sum_{i=1}^L \sqrt{G_i \alpha_i(t)} \delta(\tau - \tau_i) e^{j[2\pi(f_d \cos \phi_i)t - 2\pi f \tau_i]}. \quad (1.16)$$

In the case of flat fading (small spread), the impulse response is given by

$$h(\tau, t) = \sqrt{\rho(t)G} \delta(\tau - \tau_0) e^{-j2\pi f \tau_0} \sum_{l=1}^L \sqrt{\alpha_l} e^{j[2\pi(f_d \cos \phi_l)t]}. \quad (1.17)$$

1.10 Multiple Transmit and Receive Antenna

A smart antenna system combines multiple antenna elements with a signal-processing capability to optimize its radiation and/or reception pattern automatically in response to the signal environment.

Antennas have been the most neglected of all the components in personal communications systems. Yet, the manner in which energy is distributed into and collected from surrounding space has a profound influence on the efficient use of spectrum, the cost of establishing new networks, and the service quality provided by those networks.

In this section, we will briefly describe the essential concepts of smart antenna systems and their important advantages over conventional omnidirectional approaches.

Omnidirectional Antennas: Since the early days of wireless communications, there has been the simple dipole antenna, which radiates and receives equally well in all directions. To find its users, this single-element design broadcasts omni-directionally in a pattern resembling ripples radiating outward in a pool of water. While adequate for simple RF environments where no specific knowledge of the users' whereabouts is available, this unfocused approach scatters signals, reaching desired users with only a small percentage of the overall energy sent out into the environment. This strategy impacts the spectral efficiency, limiting frequency reuse.

Directional Antennas: A single antenna can also be constructed to have certain fixed preferential transmission and reception directions. This can be done by sectorizing the 360 area of the cell into three 120 subdivisions, each of which covered by one directional antenna. Sector antennas provide increased gain over a restricted range of azimuths as compared to an omnidirectional antenna. This is commonly referred to as antenna element gain and should not be confused with the processing gains associated with smart antenna systems.

While sectorized antennas multiply the use of channels, they do not overcome the major disadvantages of standard omnidirectional antenna broadcast such as cochannel interference.

Smart Antenna: Instead of one or more directional antennas, a system of antenna can become an antenna system that can be designed to shift signals before transmission at each of the successive elements so that the antenna has a composite effect. This concept is known as the phased array antenna. The array can be exploited by creating sectorized antenna systems, that take a traditional cellular area and subdivide it into sectors that are covered using directional antennas looking out from the same base station location.

The system can also incorporate two antenna elements at the base station, the slight physical separation (space diversity) of which has been used historically to improve reception by counteracting the negative effects of multipath. This diversity offers an improvement in the effective strength of the received signal by using either switched diversity (each antenna directs to one direction and the system continually switches between them), or diversity combining, in which the power of both signals coming from two paths can be combined effectively to produce gain. Maximal Ratio Combining (MRC) is also another kind that combines the outputs of all the antennas to maximize the ratio of combined received signal energy to noise.

As a matter of fact, antennas are not smart, but antenna systems are smart. Generally a smart antenna system combines an antenna array with a digital signal-processing capability to transmit and receive in an adaptive, spatially sensitive manner. In other words, such a system can automatically change the directionality of its radiation patterns in response to its signal environment. This can dramatically increase the performance characteristics (such as capacity) of a wireless system [23]. There are two major categories of smart antennas regarding the choices in transmit strategy: one is switched beam, with a finite number of fixed, predefined patterns or combining strategies (sectors). The second is the adaptive

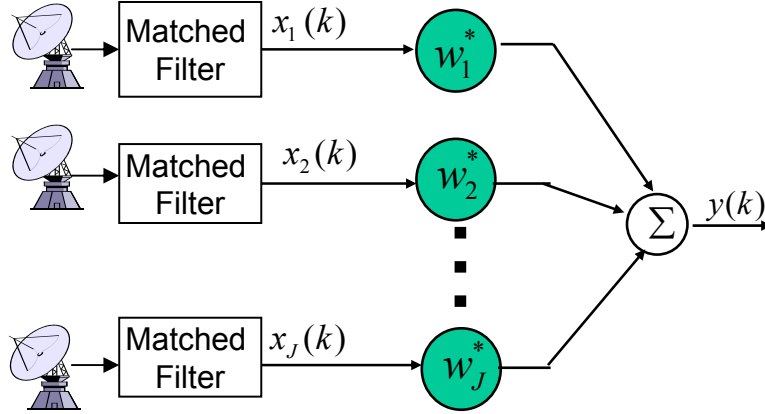


Figure 1.6: Narrowband Antenna array and beamformer

antenna array, with an infinite number of patterns that are adjusted in real time. Using a variety of new signal-processing algorithms, the adaptive system takes advantage of its ability to effectively locate and track various types of signals to dynamically minimize interference and maximize intended signal reception.

Both adaptive and switched systems attempt to increase gain according to the location of the user; however, only the adaptive system provides optimal gain while simultaneously identifying, tracking, and minimizing interfering signals. The benefits of adaptive antenna array are *the ability to obtain signal gain*, and therefore a better SNR, *increasing the transmission coverage area*, by focusing the energy in specific directions, *interference rejection*, *spatial diversity*, that minimizes the detrimental effects of multipath fading, *power efficiency*, and reduced expense. Moreover, exploiting adaptive antenna array, we can develop some techniques like determining the direction and location of a transmitter (using Multiple Signal Classification (MUSIC) [24] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [25]). Other than these applications, multiple antennas are used in three major scenarios, **beamforming**, **space-time coding** (to gain diversity), and **spatial**

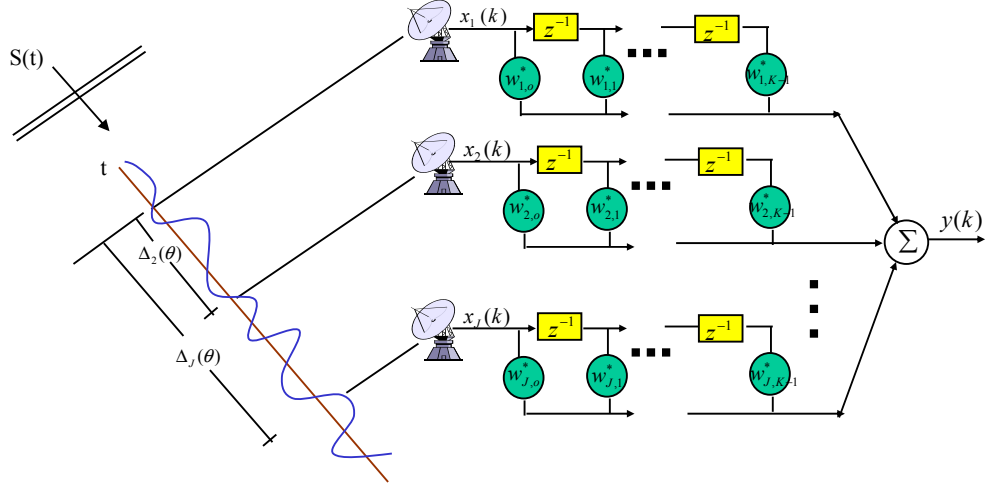


Figure 1.7: Wideband Antenna array and beamformer

multiplexing.

1.10.1 Beamforming

In receive beamforming, the antenna array are used to receive signals radiating from some specific directions and attenuate signals radiating from other directions of no interest. Fig. 1.6 shows the narrowband beamformer in which the beamforming is performed in one time instant, and 1.7 depicts the wideband beamformer. The outputs of array elements are weighted and added by a beamformer as shown in Fig. 1.6 to place nulls in the directions of sources of interference, and steer to the direction of the target signal by maintaining constant gain at this direction.

Assuming an Uniform Linear Array (ULA) as shown in Fig. 1.8, if the number of antenna elements is K , and d is the adjacent antenna separation, and θ is the direction of arrivals of a signal, the response of the i^{th} ($i = 0, \dots, K - 1$) antenna element, $\nu_i(\theta)$ is

$$\nu_i(\theta) = e^{-\frac{j2\pi d \sin \theta}{c}}, \quad i = 0, 1, \dots, K - 1 \quad (1.18)$$

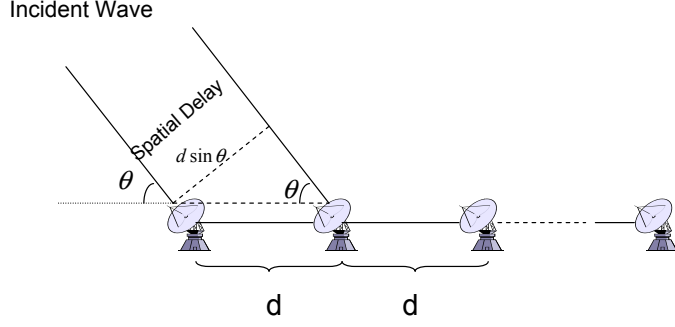


Figure 1.8: Uniform Linear Array

where c is the propagation velocity. The antenna array is therefore $\nu(\theta) = [\nu_0(\theta), \nu_2(\theta), \dots, \nu_{K-1}(\theta)]$.

Now consider a cochannel set consisting of M transmitter and receiver pairs. Let the receiver i be assigned to transmitter i .

Assuming negligible delay spreads, and slow fading channel, where the channel response can be assumed constant over several symbol intervals, the received vector at the i^{th} array can be written as

$$\mathbf{x}_i(t) = \sum_{j=1}^M \sqrt{P_j G_{ji}} \sum_{l=1}^L \alpha_{ji}^l \nu_j(\theta_l) s_j(t - \tau_j) + \mathbf{n}_i(t), \quad (1.19)$$

where $s_j(t)$ is the message signal transmitted from the j^{th} user, τ_j is the corresponding time delay, $\mathbf{n}_i(t)$ is the thermal noise vector at the input of antenna array at the i^{th} receiver, and P_j is the power of the j^{th} transmitter. The $K \times 1$ vector \mathbf{a}_{ji} , given by

$$\mathbf{a}_{ji} = \sum_{l=1}^L \alpha_{ji}^l \nu_j(\theta_l), \quad (1.20)$$

is called the spatial signature or the array response or the steering vector of the i^{th} receiver to the j^{th} source.

The received signal after performing sampling at the symbol intervals and matched filtering in the receiver i is given by

$$\mathbf{x}_i(n) = \sum_{j=1}^M \sqrt{P_j G_{ji}} \mathbf{a}_{ji} b_j(n) + \mathbf{n}_i(n), \quad (1.21)$$

The output of the beamformer and the average output power at the i^{th} receivers are given by

$$\begin{aligned}\mathbf{e}_i(n) &= \mathbf{w}_i^H \mathbf{x}_i(n), \\ \epsilon_i &= \mathbf{w}_i^H \mathbb{E}\{\mathbf{x}_i(n)\mathbf{x}_i^H(n)\}\mathbf{w}_i = \mathbf{w}_i^H \mathbf{R}_i \mathbf{w}_i,\end{aligned}\tag{1.22}$$

where \mathbf{R}_i is the correlation matrix of the received vector $\mathbf{x}_i(n)$. Assuming that the message signals $s_j(t)$ are uncorrelated and zero mean, the correlation matrix \mathbf{R}_i is given by

$$\mathbf{R}_i = \left[\sum_{j \neq i} P_j G_{ji} \mathbf{a}_{ji} \mathbf{a}_{ji}^H + N_i \mathbf{I}_M \right] + P_i G_{ii} \mathbf{a}_{ii} \mathbf{a}_{ii}^H,\tag{1.23}$$

in which the term inside the bracket is the received interference plus noise, and the second term is the energy of desired signal.

The goal of beamforming is to find a weight vector \mathbf{w}_i that minimizes the total received energy (Eq. (1.22)) subject to a constant response toward the desired signal ($\mathbf{w}_i^H \mathbf{a}_{ii} = 1$). This optimization problem readily tries to minimize the interference plus noise, and is called Minimum Variance Distortionless Response (MVDR) beamforming. MVDR is equivalent to placing main lobes toward the desired mobile and nulls toward the interferers. It can be shown that the unique solution to this problem is given by [26]

$$\mathbf{w}_i = \frac{\mathbf{R}_i^{-1} \mathbf{a}_{ii}}{\mathbf{a}_{ii}^H \mathbf{R}_i^{-1} \mathbf{a}_{ii}}.\tag{1.24}$$

The antenna gain for the signal of interest is unity. That is, the desired signal is unaffected by beamforming.

Here we assumed that the array response to the source of interest, given by (1.20), is known. The array response can be obtained by the estimation of the Direction Of Arrival (DOA) for the signal of interest from different paths. In wireless networks usually the number of cochannels and multipath signals are much larger than the number of array elements. As

a result, conventional DOA estimation methods like ESPRIT and MUSIC are not applicable. However, there exist some schemes that can be used to estimate the array response in non-spread spectrum [27], and spread spectrum systems [28], without the need to estimate the DOA. Further, if we use a training sequence there is no need to estimate the array response.

Transmit beamforming is also performed by finding the antenna weight vectors such that (1) *the transmitted energy toward the desired mobile is maximized*, and (2) *the transmitted energy toward other mobiles is minimized to avoid interference*. Note that a base station may transmit to more than one mobile with different beamforming weight vectors. Denote the diversity vector for the i^{th} mobile by \mathbf{v}_i . The received signal at each mobile is a superposition of the transmitted signal from different base stations and their delayed versions through the multipath channel. The k^{th} received symbol at the i^{th} mobile is then given by

$$y_i(k) = \sum_{b=1}^M \sum_{n=0}^{N-1} \mathbf{v}_b^H \mathbf{h}_i b(n) \sqrt{P_b} s_b(k-n) + n_i(k), \quad (1.25)$$

where s_b is the message signal transmitted from the b^{th} base station to its associated mobile, and $n_i(k)$ is the thermal noise at the i^{th} mobile. P_b can be considered as the signal power before the beamformer. Instead of absorbing this factor into the beamforming weight vector, we use it to adjust the level of the transmit power.

Similar to the receive diversity case, we can show that the desired signal power at the i^{th} receiver is given by $P_i \mathbf{v}_i^H G_{ii} \mathbf{v}_i$, and the interference power from the b^{th} base is given by $P_b \mathbf{v}_b^H G_{ib} \mathbf{v}_b$, where G_{ii} and G_{ib} are the channel gain matrices. The SINR at this receiver is given by

$$\Gamma_i = \frac{P_i \mathbf{v}_i^H G_{ii} \mathbf{v}_i}{\sum_{b \neq i} P_b \mathbf{v}_b^H G_{ib} \mathbf{v}_b + N_i}, \quad (1.26)$$

where N_i is the thermal noise power at the i^{th} mobile. Again, the transmit weight vectors \mathbf{v}_i could be chosen such that the SNR at the desired mobile is maximized.

1.10.2 Space-Time Coding

The design of codes for multiple-antenna systems (space-time codes [29]) is attracting considerable attention. Assume that a space-time code is used with block length N . The transmitted signal is represented by the $t \times N$ matrix X , $(\mathbf{x}[1], \dots, \mathbf{x}[N])$. The code, which we denote \mathcal{X} , has $|\mathcal{X}|$ words. The row index of X indicates space, while the column index indicates time: the i^{th} component of the t -vector $x[n]$, denoted $x_i[n]$, is a complex number describing the signal transmitted by the i^{th} antenna at discrete time n (for $i = 1, \dots, t, n = 1, \dots, N$). The received signal is the $r \times N$ matrix $Y = HX + Z$, where Z is the matrix of zero-mean complex Gaussian random variables with zero mean and independent real and imaginary parts with the same variance $N_0/2$ (i.e., circularly-distributed). Thus, the noise affecting the received signal is spatially and temporally independent, with $E[ZZ^H] = NN_0I_r$, where I_r denotes the size r identity matrix. The channel is described by the $r \times t$ matrix H , which is independent of both X and Z , and assuming quasi static channels, it remains constant during the transmission of an entire code word, and its realization (the channel state information, or CSI) is known at the receiver. Notice that the variance of the elements of H are chosen such that the total power received by r antennas from each transmit antenna remains constant as r varies.

Under the assumption of CSI perfectly known by the receiver, and of additive white Gaussian noise, maximum likelihood detection and decoding corresponds to choosing the codeword X which minimizes the squared Frobenius norm $\|Y - HX\|^2$, where for matrix A , we define, $\|A\|^2 = \text{Tr}(AA^H)$.

ML detection and decoding corresponds to the minimization of the quantity

$$\|Y - HX\|^2 = \sum_{n=1}^N \sum_{i=1}^r \left| y_{in} - \sum_{j=1}^t h_{ij}x_{jn} \right|^2. \quad (1.27)$$

Using the union bound to error probability

$$P(e) < \frac{1}{|\mathcal{X}|} \sum_{X \in \mathcal{X}} \sum_{\tilde{X} \in \mathcal{X} \setminus \{X\}} P(X \mapsto \tilde{X}), \quad (1.28)$$

the pairwise error probability $P(X \mapsto \tilde{X})$ is given by

$$P(X \mapsto \tilde{X}) < E \left[Q \left(\frac{\|H(X - \tilde{X})\|}{\sqrt{2}N_0} \right) \right] \leq E \left[\exp \left(-\|H(X - \tilde{X})\|^2 / 4N_0 \right) \right]. \quad (1.29)$$

Since $\|H(X - \tilde{X})\|^2 = \text{Tr} \left(H^H H (X - \tilde{X})(X - \tilde{X})^H \right)$, and using the properties of Rayleigh flat fading channels, we can change the inequality to

$$P(X \mapsto \tilde{X}) \leq \det \left[\mathbf{I}_t + (X - \tilde{X})(X - \tilde{X})^H / 4N_0 \right]^{-r}. \quad (1.30)$$

By writing the RHS of this inequality in terms of the product of the eigenvalues λ_j of the matrix $(X - \tilde{X})(X - \tilde{X})^H$, we have

$$P(X \mapsto \tilde{X}) \leq \left(\prod_{j=1}^{\rho} \lambda_j \right)^{-r} \left(\frac{\gamma}{4} \right)^{-r\rho}, \quad (1.31)$$

where γ is the signal-to-noise ratio, and ρ the number of nonzero eigenvalues. From this expression we see that the total diversity order of the coded system is $r\rho_{min}$, where ρ_{min} is the minimum rank of $(X - \tilde{X})(X - \tilde{X})^H$ across all possible pairs X, \tilde{X} (diversity gain). Moreover, the pairwise error probability depends on the power r of the product of eigenvalues of $(X - \tilde{X})(X - \tilde{X})^H$. This does not depend on the SNR γ , and displaces the error probability curve rather than changing its slope. This is called the coding gain.

In a rapidly-changing mobile environment, or when long training sequences are not allowed, the assumption of perfect CSI at the receiver may not be valid. In the absence of CSI at the receiver, [30] advocates unitary space-time modulation, a technique which circumvents the use of training symbols (which for maximum throughput should occupy half of the transmission interval, as seen before). Here the information is carried on the subspace that

is spanned by orthonormal signals that are sent to the transmit antennas. This subspace survives multiplication by the unknown channel-gain matrix. A scheme based on differential unitary space-time signals using algebraic techniques, is presented in [31].

1.10.3 spatial multiplexing

The basic principle of spatial multiplexing is to transmit independent data from each antenna, albeit with FEC coding. Then at the receiver the data from each antenna is separated by appropriate signal processing, usually involving a combination of linear decorrelation/MMSE detection and non-linear interference cancellation. In BLAST (Bell Labs Layered Space-Time), unlike beamforming techniques, we assume a rich scattering environment, and use multiple transmitters and receivers, each with its own antenna carrying independent data. All the transmitted signals occupy the same bandwidth simultaneously, so spectral efficiency is roughly proportional to the number of streams. Here the data is demultiplexed into as many separate streams as there are transmit antennas, each of which is separately coded, and fed to a separate antenna. However, the mapping of code stream to antenna is cyclically rotated every few code symbols. In this way if any particular transmit antenna is subject to particularly severe fading, only those symbols are affected, and the code should be able to overcome it. At the receiver, BLAST uses a combination of linear and nonlinear (joint MMSE) detection techniques to disentangle the mutually interfering co-channel signals. The results of decoding one stream are used in succession to remove interference from the others, while interference from those streams which cannot be cancelled (because they have not been decoded yet) is minimized in the MMSE algorithm.

Using this approach, the aggregate theoretical capacity of the subchannels can considerably exceed the capacity obtained when the channel is treated conventionally, i.e. as a single (scalar) channel. Under the assumption of independent Rayleigh scattering, the information

theoretic capacity of the BLAST architecture grows roughly linearly with the number of antennas, even when the total transmitted power is held constant [32]. This is translated into a tremendous capacity improvement. In effect, Foschini showed that in the high SNR regime, the capacity of a channel with t transmit, r receive antennas and i.i.d. Rayleigh faded gains between each antenna pair is given by:

$$C(\gamma) = \min\{t, r\} \log \gamma + O(1), \quad (1.32)$$

where γ is the SNR of the environment. However, there are a number of technical issues to be addressed before BLAST can be deployed in a mobile wireless cellular environment. First, both transmitter and receiver are required to have multiple antennas, which increases size and cost of mobile devices. Second, BLAST assumes rich scattering environments, which may not always exist in outdoor environments. Finally, BLAST requires computationally intensive processing.

There are different versions of BLAST, namely Diagonal BLAST (D-BLAST), where each successive block of information is sent from different antenna, Vertical BLAST (V-BLAST), in which one data stream is sent from each transmit antenna, ...

Here, we briefly describe Zero Forcing V-BLAST for $r \geq t$, when r and t are the number of receive and transmit antennas, respectively [33].

The channel matrix H is first decomposed as $H = QR$ using the QR factorization, where R is an upper triangular $t \times t$ matrix (normalized so that the diagonal elements are positive), and Q is an $r \times t$ matrix with $Q^H Q = I_t$ (notice that if $r = t$ then Q becomes a unitary matrix, i.e., $Q^H Q = Q Q^H = I_t$). Observe that this factorization implies an ordering of the transmit antennas, which can be performed in $t!$ ways.

The receiver uses the feed-forward filter matrix Q to obtain

$$\tilde{Y} \triangleq Q^H Y = Q^H (QRX + Z) = RX + \tilde{Z}, \quad (1.33)$$

where \tilde{Z} is a $t \times N$ matrix whose elements have the same distribution as those of Z , the covariance of noise. If interface processing were stopped at this stage (i.e., no cancellation took place), the metric would be equivalent to ML:

$$\|\tilde{Y} - RX\|^2 = \|Q^H Y - Q^H HX\|^2 = \|Y - HX\|^2. \quad (1.34)$$

Further processing (the cancellation step) by the ZF-BLAST interface is done by the (nonlinear) feedback filter. This removes the remaining spatial interference resulting from the off-diagonal terms of R , which is achieved by decoding the subcode transmitted by antenna t first, then subtracting its decoded values from the signal received from antenna $t - 1$, and so on. Specifically, we first obtain the estimate \hat{x}_t of x_t by decoding $R_{t,t}x_t + \tilde{z}_t$. Next we obtain \hat{x}_{t-1} by decoding $R_{t-1,t-1}x_{t-1} + R_{t-1,t}\hat{x}_t + \tilde{z}_{t-1}$, etc. The statistics of $R_{i,i}$ depend on the value of i : in particular, the expected value of $R_{i,i}^2$ decreases as i increases, which means that the first decoding steps are more at risk of entailing error propagation to subsequent steps. To avoid this error propagation, a possible strategy consists of choosing the ordering of rows of H which is most favorable [34].

It is worthwhile to conclude this section with the tradeoff between the diversity and spatial multiplexing gain [35]. Given a MIMO channel, both gains can in fact be simultaneously obtained, but there is a fundamental tradeoff between how much of each type of gain any coding scheme can extract: higher spatial multiplexing gain comes at the price of sacrificing diversity. A scheme is said to have a spatial multiplexing gain r and a diversity advantage d if the rate of the scheme scales like $r \log \gamma$, and the average error probability decays like $(1/\gamma)^d$. The diversity-multiplexing tradeoff is essentially the tradeoff between the error probability and the data rate of a system, by allowing both of them to scale with the SNR.

1.11 Contribution of this Dissertation

The contributions of this dissertation are as follows:

- A joint iterative frequency-domain power control and receive beamforming is used in an OFDM system to achieve the desired SINR with the minimum total power at each OFDM subchannel, such that we can achieve a better overall error probability with the same total transmission power.
- To reduce the computational complexity of the joint frequency-domain beamforming and power control for OFDM, an iterative joint time-domain beamforming and power control for an OFDM uplink transmission is proposed. Although, this scheme is sub-optimal in terms of total transmission power, its complexity is significantly lower. The convergence of this scheme is also discussed.
- For practical implementations, joint time-domain MMSE beamforming and power control for OFDM is also proposed.
- The same schemes mentioned in the previous items, are also extended to Coded OFDM (COFDM) and their results are compared.
- Two iterative scheme is proposed for a downlink multi-user multi-cell multi-stream multiple transmit and receive antenna OFDM (MIMO OFDM) system is proposed to maximize the overall mutual information for all users. Multiple stream are used in order to achieve spatial multiplexing and increase the maximum achievable data rate.
- A non-cooperative theory is established for maximization of achievable data rate, where the players are different users and the parameters to be selected are the transmit and receive weight vectors.

- The single stream MIMO-OFDM is also considered and an iterative algorithm is proposed to maximize the actual transmission data rate when the transmission power is constrained.
- The concept of super Golay codes is introduced and their effect on the Peak to Average Power Ratio (PAPR) of OFDM systems using 16QAM sequences is discussed.
- The concept of cyclic-Golay codes is introduced such that their PAPR is as low as Golay codes, but their coding rate is higher.
- A construction method for creating cyclic-Golay codes is proposed. This method could be applied to find the cyclic shifts of any code presented by means of generalized Boolean functions.
- Two decoding schemes, one non-recursive and one recursive, for $RM_{2^h}(r, m)$ are devised. The non-recursive scheme is an extension of the majority logic Reed algorithm used for decoding of binary Reed-Muller codes. The complexities of both methods are analyzed and it is shown that the recursive method has lower complexity.
- A scheduling algorithm for providing QoS for wireless mobile users and utilizing the system resources efficiently is devised. The algorithm performs a trade-off between the QoS and throughput by maximizing the income a Service Level Algorithm (SLA)-based networks.
- Both a short-term optimal greedy algorithm (input maximization for one time slot) and a long-term optimal scheme using dynamic programming are introduced.
- A QoS-provisioned optimal subcarrier allocation to different users of an Orthogonal Frequency Division Multiple Access (OFDMA) system using the notion of revenue

maximization is proposed. Optimal and sub-optimal algorithms are also presented and their performances and complexities are studied.

1.12 Organization of the Dissertation

In Chapter 2 of this dissertation, we describe the basic principals of an Orthogonal Frequency Division Multiplexing (OFDM) system, its history, and its advantages and disadvantages. We will also compare OFDM with single carrier modulation schemes. Next we move to present the loading algorithms, and its comparison with the use of coding in OFDM system. We will also describe the problem of Peak to Average Power Ratio (PAPR) in an OFDM system, and different solutions proposed to resolve this issue. Finally, we will briefly talk about the Orthogonal Frequency Division Multiple Access (OFDMA) scheme, which is used in IEEE802.16 Fixed wireless broadband systems, and the problem of channel allocation in these systems.

In Chapter 3, first we will describe the power control problem in a multi-user cellular system and then outline our proposal for using joint power control and receive beamforming in multi-user OFDM systems. To reduce the complexity of this scheme, we will also propose joint power control and time-domain receive beamforming in an OFDM system. In situations where the array response is not known at the receiver, we will propose MMSE joint beamforming and power control, both in time and frequency domain. We will apply these scheme to Coded-OFDM (COFDM), and finally the performances of these schemes are compared in this Chapter.

Chapter 4 considers multiple antennas both at the transmitter and receiver of an multi-user, multi-cell, and multi-stream OFDM system. It proposes the use of iterative water-filling to divide a fix power among different streams at different subcarriers, to achieve

maximum achievable rates. A game theoretic approach for this problem is presented and the convergence of these schemes is also discussed. Another algorithm for single stream transmission is also presented.

In Chapter 5, we will focus on the problem of PAPR reduction in OFDM system by coding across subcarriers. In an attempt to increase the coding rate of well-known Golay codes, first, we propose the use of super-Golay codes when the symbols are selected from a non-equal energy like 16-QAM. Next, we propose the use of Cyclic Golay codes, and analyze the PAPR achieved by these codes. Then, we propose a construction method to obtain these codes out of Golay codes. The proposed scheme can be used to generate the cyclic shift of any code described by means of Boolean functions. We will show that these codes are in general a subset of r^{th} order Reed-Muller codes ($RM_{2^h}(r, m)$), and propose two scheme, one recursive and one non-recursive to decode generalized Reed-Muller codes. The complexities of these schemes are discussed and compared with some existing methods.

Chapter 6 serves two purposes. First, it proposes a novel scheduling algorithm when there is a Service Level Agreement (SLA) between the administrator and users of a wireless network. This scheme is based on the notion of network income and tries to achieve a meaningful tradeoff between the throughput of the system and guaranteed QoS for users. Both greedy (optimization for one time slot) and long-term solutions are proposed and their performances are compared to each other and to those of some other scheduling algorithms. This concept also builds a foundation for a channel allocation scheme proposed in the second part of this Chapter. This scheme is used to allocate the subcarriers of an OFDM system to different users, while a meaningful tradeoff between the overall throughput and the QoS is aimed.

Finally, Chapter 7 concludes the dissertation and outlines some possible future works for the subject covered in this dissertation.

Chapter 2

Orthogonal Frequency Division Multiplexing (OFDM)

2.1 Motivation for introducing OFDM

The most detrimental effect in wireless communication is the fading caused by multipath propagation. Other problems are Inter-Symbol Interference (ISI), shadowing, and interference. Further constraints are limited bandwidth, low power consumption, and network management. Because of multipath propagation, many reflected signals from trees, hills, building, people, cars, etc. arrive at the receiver at different times. Fading and ISI are caused by the combination of these echoes. This combination could be either constructive or destructive. Because of this fading, some frequencies are enhanced, whereas others are attenuated, and therefore the channel is frequency selective. If the bandwidth of the signal is great, some parts of the signal may suffer from constructive interference and be enhanced in level, whereas others may suffer from destructive interference and be attenuated. In general, frequency components that are close together will suffer variations in signal strength that are strongly correlated. The correlation (or coherence) is used as a measure of this phenomenon. For a narrowband signal, distortion is usually minimized if the bandwidth is less than the coherence bandwidth of the channel, because all frequencies in the band are usually

distorted in the same way. However, a signal which occupies a wider bandwidth, greater than coherence bandwidth, will be subject to more distortion, but will suffer less variation in total received power, even if it is subject to significant levels of multipath propagation. This comes from the fact that variation averages out if the bandwidth is much larger than the coherence bandwidth, because different parts of the band suffer different levels of distortion. One can often find following formula for coherence bandwidth CBW [7], where τ_{rms} is the RMS value of delay spread (not average):

$$CBW \simeq \frac{1}{\tau_{rms}}. \quad (2.1)$$

There are many echoes present in the time-domain response of the channel. These number of echoes are different for different environments, like outdoor/indoor areas. This range of delay can be measured and then processed to get statistical parameters. Different studies use the total range of delay, or the average delay. Whichever is chosen, the inverse of this leads to a good approximation for the coherence bandwidth.

The spread spectrum techniques have proven to be robust against fading and interference, but they set forth impossible demands on the existing technology for instance, if a user needs a speed of $20Mb/s$ on air and the spreading factor is 128, this results in $2.56GB/s$ which have to be processed in real-time and have impracticably large bandwidth. Besides that, they have difficulty with the near-far effect and have a large power-consumption. Single-carrier techniques are vulnerable to fading and multipath propagation, especially in the case of very high bit rates. Improvements can be made with frequency equalization and directional antennas, which can also be used to improve multicarrier techniques.

Orthogonal Frequency Division Multiplexing (OFDM) is a wideband modulation scheme that is designed to cope with the problems of the multipath reception. Essentially, the wideband frequency selective fading channel is divided into many narrow-band subchannels. If

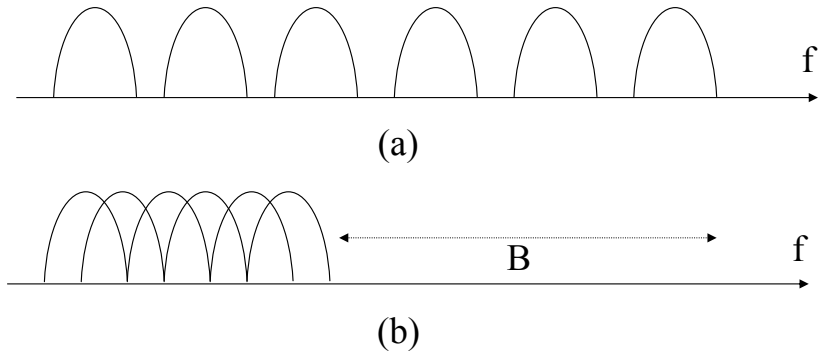


Figure 2.1: (a) A typical FDM spectrum. (b) A typical OFDM spectrum, where B is the saving in bandwidth.

the number of subchannels is high enough, each subchannel could be considered as flat. This is because we transmit many narrowband overlapping digital signals in parallel, inside one wide band. Increasing the number of parallel transmission channels reduces the data rate that each individual carrier must convey, and that lengthens the symbol period. Therefore, the delay time of reflected waves is suppressed to within 1 symbol time.

2.2 OFDM History

The idea of using parallel data transmission by frequency division multiplexing (FDM) was published in mid 60s [36]. A U.S. patent was filed and issued in January, 1970. The idea of an U.S. patent, which was filed in 1970 and was for military application, was to use parallel data streams and FDM with overlapping subchannels to avoid complicated equalization and to combat frequency domain noise, and multipath distortion. The term Discrete Multi-Tone (DMT), multichannel modulation and Multi-Carrier Modulation (MCM) are used interchangeably with OFDM. The main specification of OFDM which is not necessary in MCM is that each carrier is orthogonal to all other carriers. OFDM is an optimal version of multicarrier transmission schemes.

For a large number of subchannels, the arrays of sinusoidal generators and coherent demodulators required in a parallel system become unreasonably expensive and complex. The receiver needs precise phasing of the demodulating carriers and sampling times in order to keep crosstalk between subchannels acceptable. Weinstein and Ebert [37] applied the Discrete Fourier Transform (DFT) to parallel data transmission system as part of the modulation and demodulation process. In addition to eliminating the banks of subcarrier oscillators and coherent demodulators required by FDM, a completely digital implementation could be built around special-purpose hardware performing the Fast Fourier Transform (FFT). Fig. 2.1 compares the bandwidth utilization of FDM and OFDM. Fig. 2.2 shows the spectrum of OFDM signals, where every one of the $\text{sinc}(\cdot)$ functions represent the spectra of one OFDM tone. This figure shows that the OFDM signal is indeed the multiplexing of individual spectras with a frequency spacing equal to the transmission speed of each subcarrier. It shows that at the center frequency of each subcarrier, there is no crosstalk from other carriers. Therefore, if we use DFT at the receiver and calculate correlation values with the center frequency of each subcarrier, we can recover the transmitted data with no Inter-Channel Interference (ICI). In addition, using the DFT-based multicarrier technique, frequency division multiplexing is achieved not by bandpass filtering but by baseband processing.

Recent advances in VLSI technology enable making of high-speed chips that can perform large size FFT at affordable price. In the 1980s, OFDM has been studied for high-speed modems, digital mobile communications and high-density recording. One of the systems used a pilot tone for stabilizing carrier and clock frequency control and trellis coding was implemented. Various fast modems were developed for telephone networks. In 1990s, OFDM has been exploited for wideband data communications over mobile radio FM channels, High-bit-rate Digital Subscriber Lines (HDSL) with a speed up to 1.6Mb/s , Asymmetric Digital Subscriber Lines (ADSL) for $1,536\text{Mb/s}$, Very High-speed Digital Subscriber Lines (VHDSL)

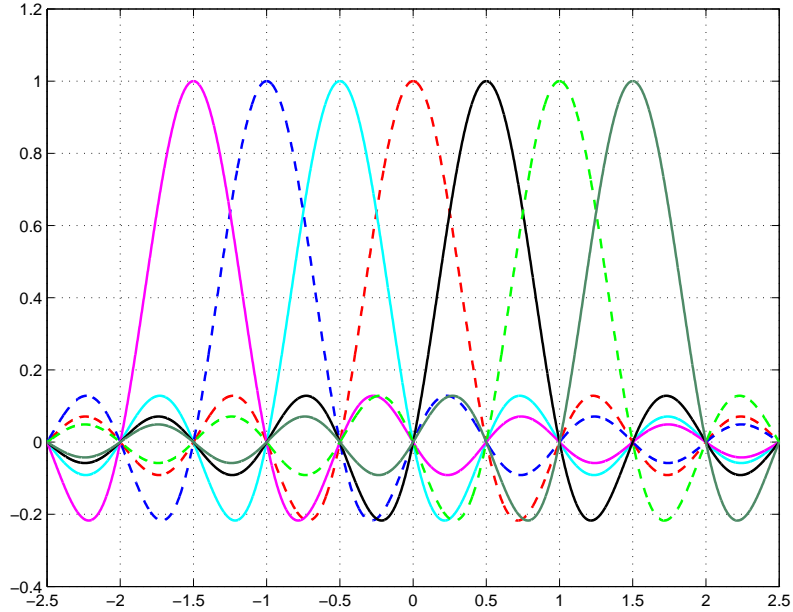


Figure 2.2: (a) Spectra of OFDM signal.

for 100Mb/s , Digital Audio Broadcasting (DAB), and HDTV terrestrial broadcasting. In 1999, the Wireless LAN (WLAN) standard committee adopted OFDM as the physical layer of IEEE802.11a that was supposed to support up to 54Mb/s in 5GHz band. Other WLAN standards like the more update IEEE802.11g, ETSI HIPERLAN/2, Mobile Multimedia Access Communication (MMAC) also have accepted OFDM as their physical layer specification. These WLAN systems also incorporate coding with OFDM to combat dispersing channels. It has been shown that Coded-OFDM over dispersing channels can improve the reliability of the transmission. IEEE802.16, the standard for fixed wireless has adopted, and IEEE802.20, the standard for broadband wireless systems which is aimed to replace the wideband systems like 3G also exploit OFDM.

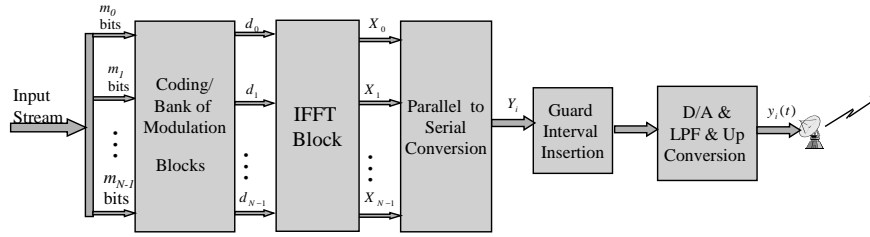


Figure 2.3: Basic Building block of an OFDM transmitter.

2.3 Description of OFDM

In a conventional serial data system, the symbols are transmitted sequentially, with the frequency spectrum of each data symbol allowed to occupy the entire available bandwidth.

In a parallel data transmission system several symbols are transmitted at the same time, that offers possibilities for alleviating many of the problems encountered with serial systems.

In OFDM, the data is divided among large number of closely spaced carriers. This accounts for the frequency division multiplex part of the name. This is not a multiple access technique, since there is no common medium to be shared. The entire bandwidth is filled from a single source of data. Instead of transmitting in serial way, data is transferred in a parallel way. OFDM can be simply defined as a form of multicarrier modulation where its carrier spacing is carefully selected so that each subcarrier is orthogonal to the other subcarriers. As is well known, orthogonal signals can be separated at the receiver by correlation techniques; hence, ISI among channels can be eliminated. Orthogonality can be achieved by carefully selecting carrier spacing, such as letting the carrier spacing be equal to the reciprocal of the useful symbol period.

Unlike FDM, the carriers in an OFDM signal are arranged so that the sidebands of the individual carriers overlap and the signals can still be received without adjacent carrier interference. In order to do this, the carriers must be orthogonal. The receiver acts as a

bank of demodulators, translating each carrier down to DC, the resulting signal then being integrated over a symbol period to recover the raw data. If the other carriers all beat down to frequencies which, in the time domain, have a whole number of cycles in the block period (T_b), then the integration process results in zero contribution from all these carriers. Thus, the carriers are linearly independent (i.e. orthogonal) if the carrier spacing is a multiple of $1/T_b$.

Mathematically, it is very easy to see that the sinusoidal series, $e^{\frac{j2\pi nt}{N\Delta t}}$, $n = 0, 1, \dots, N - 1$ constitutes an orthogonal series, in the sense that

$$\int_0^{N\Delta t} e^{\frac{j2\pi mt}{N\Delta t}} e^{-\frac{j2\pi nt}{N\Delta t}} dt = \begin{cases} N\Delta t, & \text{for } m = n; \\ 0, & \text{for } m \neq n. \end{cases} \quad (2.2)$$

where the interval $[0, N\Delta t]$ is a symbol period. Figs. 2.3 and 2.4 depict a conventional OFDM transmitter and receiver. The complex continuous wave N -carrier OFDM signal after the FFT block is:

$$S_c(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{j2\pi(f_0+n\Delta f)t}, \quad (2.3)$$

where f_0 is the carrier frequency, Δf is the carrier spacing, and X_n is the complex symbol to be transmitted at carrier n . A window of incoming data bits are selected (the size of window is equal to the OFDM block period T_b) and are converted from serial to parallel substreams. X_n could be one chosen among the constellation points of some modulation scheme, like QPSK or 16QAM, or ... selected by the substream bits. It also could be generated by some coding scheme applied to the incoming windowed bits.

Eq. (2.3) is the DFT of the symbols X_n ($n = 0, 1, \dots, N - 1$). The DFT is a variant on the normal transform in which the signals are sampled in both time and the frequency domains. By definition, the time waveform must repeat continually, and this leads to a frequency spectrum that repeats continually in the frequency domain [38].

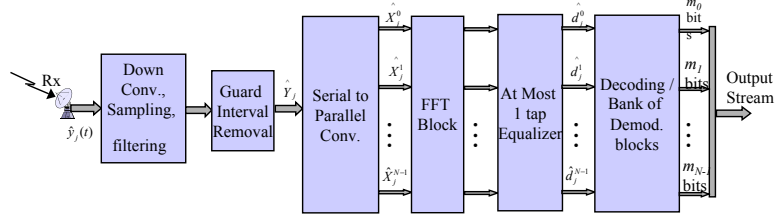


Figure 2.4: Basic Building block of the j th OFDM Receiver.

The fast Fourier transform (FFT) is merely a rapid mathematical method for computer applications of DFT. It is the availability of this technique, and the technology that allows it to be implemented on integrated circuits at a reasonable price, that has permitted OFDM to be developed efficiently. The process of transforming from the time domain representation to the frequency domain representation uses the Fourier transform itself, whereas the reverse process uses the inverse Fourier transform.

A natural consequence of using FFT is that it allows us to generate carriers that are orthogonal. The members of an orthogonal set are linearly independent.

The baseband complex discrete OFDM symbols are

$$D_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{j2\pi nk/N} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{j2\pi f_n t_k},$$

where $f_n = n/(N\Delta t)$, $t_k = k\Delta t$, and Δt is an arbitrarily chosen symbol duration of the serial data sequence X_n . The in-phase and quadrature components of these symbols are applied to a low-pass filter to obtain the baseband continuous wave OFDM signal as

$$y(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{-j2\pi nt/N\Delta t}. \quad 0 \leq t \leq N\Delta t = T_b$$

Here, it is assumed that the pulse-shaping filter is normalized to 1. If the signal $y(t)$ is up-converted to carrier frequency, the signal in (2.3) is obtained. The transmitted OFDM signal is the real part of $S_c(t)$.

By dividing the input data stream into N subcarriers, the symbol duration is made N times smaller, which also reduces the multipath delay spread, relative to the symbol time, by the same factor. The orthogonality of subchannels in OFDM can be maintained and individual subchannels can be completely separated by the FFT at the receiver when there are no ISI and ICI introduced by transmission channel distortion. In practice these conditions can not be obtained. Since the spectra of an OFDM signal is not strictly band limited ($\text{sinc}(f)$ function), linear distortion such as multipath cause each subchannel to spread energy into the adjacent channels and consequently cause ISI. A simple solution is to increase symbol duration or the number of carriers so that distortion becomes insignificant. However, this method may be difficult to implement in terms of carrier stability, Doppler shift, FFT size and latency.

One way to prevent ISI is to create a cyclically extended guard interval, where each OFDM symbol is preceded by a periodic extension of the signal itself. The total symbol duration is $T_{total} = T_g + T_b$, where T_g is the guard interval and T_b is the useful symbol duration. When the guard interval is longer than the channel impulse response, or the multipath delay, the ISI can be eliminated. However, the ICI, or in-band fading, still exists. ICI is the crosstalk between different subcarriers, which means they are no longer orthogonal. The ratio of the guard interval to useful symbol duration is application-dependent. Since the insertion of guard interval will reduce data throughput, T_g is usually less than $T_b/4$. Other reasons to use a cyclic prefix for the guard interval are:

- To maintain the receiver carrier synchronization; some signals instead of a long silence must always be transmitted;
- Cyclic convolution can still be applied between the OFDM signal and the channel response to model the transmission system.

To explain the second reason of using cyclic prefix, note that the received signal is the linear convolution of channel impulse response, $h(t)$ and the time domain OFDM signal, plus noise, $n(t)$:

$$r(t) = y(t) * h(t) + n(t), \quad (2.4)$$

where $*$ represents linear convolution. If we assume that channel response has L taps, then by insertion of $(L - 1)\Delta t$ guard time between successive transmission blocks, we can make sure that the symbol duration is made larger than the time dispersion caused by the channel. However, adding the time domain samples $y_{N-P}, y_{N-P+1}, \dots, y_{N-1}$, ($P \geq L - 1$), to the block, as a cyclic prefix, causes the linear convolution shown in (2.4) to be performed between the channel and the augmented symbol. However since the augmented signal is cyclic, this is equivalent the circular convolution of the original OFDM sequence, y_n with the channel coefficients. Note that

$$DFT(y \circledast h) = DFT(y).DFT(h), \quad (2.5)$$

where \circledast shows the circular convolution. This relation is not valid for linear convolution. At the receiver, the signal is translated to base-band and is sampled at the multiples of Δt . After removal of cyclic prefix, it is fed to a DFT module in order to yield the received frequency domain symbols $\hat{X}_m, m = 0 \dots N - 1$.

To simplify the discussion, assume that multipath fading channel is constant over one OFDM symbol (quasi-static), and therefore can be shown by

$$h(t) = \sqrt{G} \sum_{l=1}^L \alpha_l \delta(t - \tau_l), \quad (2.6)$$

where G accounts for the path loss and shadow fading. The signal after down conversion is

$$\begin{aligned}
r(t) &= \sqrt{G} \sum_{l=1}^L \alpha_l e^{-j2\pi f_c \tau_l} y(t - \tau_l) + n(t) \\
&= \frac{\sqrt{G}}{\sqrt{N}} \sum_{l=1}^L e^{-j2\pi f_c \tau_l} \sum_{n=0}^{N-1} X_n e^{\frac{j2\pi n(t-\tau_l)}{N\Delta t}} \alpha_l + n(t) \\
&= \frac{\sqrt{G}}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{\frac{j2\pi n t}{N\Delta t}} \sum_{l=1}^L \alpha_l e^{-j2\pi(f_c + 2\pi n/N\Delta t)\tau_l} \alpha_l + n(t), \tag{2.7}
\end{aligned}$$

where f_c is the carrier frequency. After sampling at time points $k\Delta t$ ($k = 0, 1, \dots, N-1$), the k^{th} sample is given by:

$$r_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n X_n e^{j2\pi n k/N} + n_k, \quad k = 0, 1, \dots, N-1 \tag{2.8}$$

where n_k are noise samples. H_n is the channel frequency response at subcarrier n , and is given by:

$$H_n = \sqrt{G} \sum_{l=1}^L \alpha_l e^{-j2\pi(f_c + 2\pi n/N)\tau_l}. \tag{2.9}$$

These samples are applied to a DFT block, and because of orthogonality and Eq. (2.5), the resulting frequency domain symbols are

$$\hat{X}_m = H_m X_m + n_m, \tag{2.10}$$

which says that the received OFDM symbols are the scaled versions of the transmitted ones plus thermal noise.

In order to retrieve the transmitted symbol, the receiver needs to know the channel state information (CSI). Frequency-domain channel estimation can be performed with pilot symbols that are interspread with the transmitted OFDM symbols. A pilot e consists of known symbols e_m , $m = 0, \dots, N-1$. The received pilot symbol at subcarrier m after DFT is $z_m = e_m H_m + n_m$. Then the Minimum Mean Square Error (MMSE) estimate of the

complex gain H_m is obtained by

$$\tilde{H}_m = \frac{z_m}{e_m} = H_m + \frac{n_m}{e_m}.$$

These estimates are used for Frequency Domain Equalization (FEQ). If the estimations are accurate enough, the maximum likelihood detector makes its decision based on the statistics R_m/\tilde{H}_m . If the channel is slowly time varying, the transmitter can obtain reliable CSI as feedback from the receiver.

2.3.1 Advantages of OFDM

Compared to FDM, the overlapping spectra of subcarriers in OFDM yields better spectral utilization. Only a small amount of the data is carried on each subcarrier, and by this lowering of the bit rate per subcarrier (not the total bit rate), the influence of ISI is significantly reduced. In principle, many modulation schemes could be used to modulate the data at a low bit rate onto each subcarrier. It is an important part of the OFDM system design that the bandwidth occupied is greater than the coherence bandwidth of the fading channel. Then, although some of the carriers are degraded by multipath fading, the majority of the carriers should still be adequately received. OFDM can effectively randomize burst errors caused by Rayleigh fading, which comes from interleaving due to parallelisation. So, instead of several adjacent symbols being completely destroyed, many symbols are only slightly distorted. Because of dividing an entire channel bandwidth into many narrow subbands, the frequency response over each individual subband is relatively flat. Since each subchannel covers only a small fraction of the original bandwidth, equalization is potentially simpler than in a serial data system. A simple equalization algorithm can minimize mean-square distortion on each subchannel, and the implementation of differential encoding may make it possible to avoid equalization altogether [37]. This allows the precise reconstruction of majority of them, even

without forward error correction (FEC). In addition, by using a guard interval the sensitivity of the system to delay spread can be reduced [39]. Furthermore, OFDM provides additional flexibility in transmission adaptation to varying channel conditions, by allowing modulation level and power adjustment for each symbol in a subcarrier [40]. It also accommodates multi-users by allocating different subcarriers to different users (OFDMA). OFDM can be used in conjunction with bit loading techniques to improve the capacity of a highly frequency selective channel.

2.3.2 Disadvantages of OFDM

In what follows, we will describe the main disadvantages of an OFDM system very briefly:

High sensitivity to carrier frequency offset: The offset in carrier frequency results in reducing desired symbol amplitude and introduces ICI. When comparing OFDM to a conventional single carrier system, it is in orders of magnitude more sensitive to frequency offset and Wiener phase noise [41]. This problem is more expressed in mobile applications, since a mobile channel has a time-varying nature. This offset causes the received frequency domain subcarriers to be shifted. The subcarriers are still mutually orthogonal, but the received data symbols, which were mapped to the OFDM spectrum, are in the wrong position in the demodulated spectrum, resulting in increased BER.

Sensitivity to time-domain synchronization errors and Sensitivity to phase noise: If the receiver's FFT window is shifted with respect to that of the transmitter, then at the receiver side, the timing mis-alignment introduces phase error between adjacent subcarriers. The influence of phase error on OFDM system depends on the modulation scheme; (1) *Coherent Modulation*: In case of imperfect time synchronization, phase correction mechanisms are very crucial for coherent modulation. (2) *Pilot symbol assisted Modulation*: Pilots are inter-spread with the data symbols in the frequency domain and the receiver can esti-

mate the evolving phase error from the received pilots' phases. (3) *Differential Modulation*: Differential encoding can be implemented both between corresponding subcarriers of consecutive OFDM blocks or between adjacent subcarriers of the same OFDM block to alleviate the effect of phase error.

Large Peak to Average Power Ratio (PAPR): will be described later.

Non-linear distortion: Since an OFDM signal generally consists of a large sum of independently modulated subcarriers, then we might see large signal excursions and large PAPRs. Therefore, an OFDM signal is very vulnerable to non-linear distortion caused by any non-linear element in the system such as High Power Amplifiers (HPAs). The effect of a non-linearity on an OFDM signal can be seen from two different point of views. First, the large signal excursions occasionally reach the non-linear region of any non-linear element in the system (saturation point) and hence the non-linearity output is a distorted replica of the input. Second, any nonlinear element in the system introduces severe harmonic distortion and inter-modulation distortion due to the multi-carrier nature of an OFDM signal. For example, the response of a third-order or cubic non-linearity to a sum of sine waves with frequencies $f_i, i = 1, \dots, N$, corresponds to a sum of sine waves consisting of the original sine waves with frequencies $f_i, i = 1, \dots, N$, harmonics with frequencies $3f_i, i = 1, \dots, N$, and other frequencies of the type $f_i + f_j \pm f_k$ and $2f_i \pm f_j$, where i, j and k are integers that distinguish the different input frequencies. Generally, the response of an n^{th} order non-linearity to a signal consisting of a sum of sine waves corresponds to a signal also consisting of a sum of sine waves with frequencies corresponding to every possible combination of sums and/or differences of n input frequencies. These newly generated frequencies can fall in-band or out-of-band. Frequencies that fall in-band, are associated with signal error probability degradation and frequencies that fall out-of-band are associated with signal spectral spreading. Note that the two perspectives discussed above are related as the large signal excursions

are the consequence of the multi-carrier nature of the OFDM signal.

2.3.3 Single Carrier versus OFDM Comparison

The main difference between the single carrier and OFDM are their robustness to fading and synchronization errors (both frequency and timing). Assuming perfect synchronization, the performances of a single carrier system and an OFDM system are equivalent for AWGN and flat fading channels. Consider a received signal for a single carrier system

$$y_i = h_i s_i + n_i \quad 0 \leq i \leq N - 1, \quad (2.11)$$

where h_i is a complex random variable, s_i is the baseband representation of the i^{th} modulation symbol, and n_i is complex additive Gaussian noise sample in the i^{th} signal interval. The equivalent OFDM received signal (assuming no ISI and ICI) is

$$Y_k = H_k S_k + N_k \quad 0 \leq k \leq N - 1, \quad (2.12)$$

where H_k , X_k , and N_k are the frequency domain representations of r_i , x_i , and n_i , respectively. Since the noise power of n_i and N_k are equivalent by Parseval Theorem [38], there is no advantage in detecting the signal using either of these equations. Now, consider reception of a signal over a frequency-selective channel. For the single carrier system, the received signal becomes

$$\mathbf{y}(i) = \mathbf{h}(i) * \mathbf{s}(i) + \mathbf{n}(i), \quad (2.13)$$

where $\mathbf{y}(i) = [y_i, y_{i+1}, \dots, y_{L+i-1}]^H$, A^H represents the Hermitian (complex transpose) of matrix A , L is the number of taps, and $*$ represents the linear convolution. $\mathbf{h}(i)$, $\mathbf{s}(i)$, and $\mathbf{n}(i)$ are defined correspondingly. The single carrier system requires an equalizer $\mathbf{g}(i)$ to compensate for the channel effect, by the deconvolution operation on the received signal, i.e.,

$$\hat{\mathbf{y}}(i) = \mathbf{g}(i) * \mathbf{y}(i) = \mathbf{s}(i) + \mathbf{h}(i) * \mathbf{n}(i) + \epsilon(i), \quad (2.14)$$

where $\epsilon(i)$ is the residual error which is resulted because normally the equalizer cannot perfectly inverse the effects of the channel. Note that the deconvolution process also enhances the noise amplitude in some samples.

In contrast, if we assume that the circular convolution has removed the effect of ISI, the OFDM system performs equalization in the same manner as the case of the flat fading channel, i.e.,

$$\begin{aligned} Y_k &= H_k S_k + V_k \\ \hat{Y}_k &= S_k + \Phi_k V_k, \end{aligned} \quad (2.15)$$

where

$$\Phi_k = [1/h_i, 1/h_{i+1}, \dots, 1/h_{p_k+i+1}]^H, \quad (2.16)$$

and p_k is the number of taps in carrier k . Equalization in OFDM systems is subject to the same impairments such as residual error and noise enhancement as the single carrier system; thus, theoretically, the two systems have equivalent performance. Yet, the complexity of the equalizer for the OFDM system is substantially less than that for the single carrier system. The reason is that OFDM systems employ a bank of single-tap equalizers while single carrier systems employ multi-tap equalizers. Further, the complexity of the of the equalizer grows as the square of the number of taps.

Synchronization Errors

Synchronization errors can be either timing, frequency, or both. The single carrier system is much more sensitive to timing errors than the OFDM system. On the other hand, the OFDM system is more sensitive to frequency errors.

Effect of Timing Errors: Even with the training intervals, the demodulator reference circuitry may not be able to recover completely the timing at the transmitter. Without tim-

ing synchronized, the SNR at the output of the detection filter is degraded. For a particular sampling time $T_{optimal}$, the output SNR is depends on the autocorrelation, $A(\tau)$ [42], where τ is a random variable that represents the delay between the optimum sampling instance $T_{optimal}$ and the associated symbol timing for the received signal, and is estimated in the presence of noise. The variability of this delay is called *timing jitter*. As an example, the autocorrelation function for the band-limited signals is given by [43]:

$$A(\tau) = \frac{1}{N} \left(\frac{\sin(\pi N W \tau)}{\sin(\pi W \tau)} \right), \quad (2.17)$$

where W is the bandwidth of the band-limited signal. $A(\tau)$ can describe the single carrier system, but OFDM system is described as time-limited signal. For single carrier systems, the timing error or jitter causes a phase error for the bandpass signal. However, in OFDM systems, we can transmit pilot symbols on some reserved tones and therefore we can estimate residual phase errors. The single carrier system does not have a mechanism to archive this compensation [44].

Effects of Frequency Errors: When there is relative motion between the transmitter and receiver, a Doppler shift of the RF carrier results and introduces frequency error. Also, any error in the oscillators either at the transmitter or at the receiver can result in residual frequency error. In either case, there are well-known carrier recovery schemes available for single carrier systems such as a first Costa loop [45]. The important result in [46] says that although the carrier can be recovered, the phase may be unknown and random.

In summary, the single carrier systems are more robust to frequency offset error, but more vulnerable (SNR degradation) to timing error. OFDM systems are robust to timing error, but their SNR could be degraded by frequency offset error. These properties come from the fact that, supporting the same data rate, OFDM symbol duration is N times longer than the single carrier system. The complexity of the equalizer and the system performance

in flat fading is the same for both, while OFDM has much better performance and much simpler equalizer in frequency-selective fading environments.

2.4 Loading Algorithms

The key advantage of OFDM is that each subchannel is relatively narrowband and is assumed to have flat fading. However, it is entirely possible that a given subchannel has a low gain, resulting in a large BER. Thus, it is desirable to take advantage of subchannels having relatively good performance; this is the motivation for adaptive modulation. In the context of time-varying channels, there is a decorrelation time associated with each frequency-selective channel instance. Thus, a new adaptation must be implemented each time the channel decorrelates.

The number of bits that can be carried by each of the N subcarriers is given by:

$$b_i = \log_2 \left(1 + \frac{SNR_i}{\Gamma_i} \right),$$

where the SNR Gap, Γ_i at subcarrier i is used to measure the reduction of SNR with respect to capacity. It depends on the objective bit error rate and the modulation used. If the objective BER is the same for all sub-channels, then $\Gamma_i = \Gamma$ for all i . If coding is being used, SNR gap can be modified by introducing additional coding gain factors [47].

The optimal adaptive transmission scheme, which achieves the Shannon capacity for a fixed transmit power is the water-filling distribution of power over the frequency-selective channel. However, while the water-filling distribution will indeed yield the optimal solution, it is difficult to compute. Moreover, we need infinite granularity in the constellation size, which is not practically realizable.

Here, we will present the OFDM bit loading algorithms used in [48, 49], which optimize the power and rate based on the knowledge of the subchannel gains.

In the discrete bit loading algorithm of [48], the amount of energy necessary to transmit b bits on subchannel n at the desired probability of error using a given coding scheme is represented by a convex monotonic function $e_n(b)$, which is initialized to zero for all n .

The problem of energy minimization and bit allocation can be formulated as:

$$\begin{aligned} & \min_{b_n} \sum_{n=1}^N e_n(b_n) & (2.18) \\ \text{subject to } & \begin{cases} \sum_{n=1}^N b_n = B, \\ b_n \in \mathcal{Z}, b_n \geq 0, n = 1, 2, \dots, N. \end{cases} \end{aligned}$$

To initialize the bit allocation, first the number of bits for the i^{th} $i = 0, 1, \dots, N - 1$ subchannel, $b(i)$, is computed based on (2.4), and is rounded down to the integer value $\tilde{b}(i)$. Note that, since we are using MQAM, the number of bits are restricted to the values $0, 1, 2, 4, 6, 8$, based on the chosen constellation. Then, the energy for the i^{th} subchannel is computed by rewriting (2.4) as:

$$e_i(\tilde{b}(i)) = \left(2^{\tilde{b}(i)} - 1\right) * \Gamma / SNR_i. \quad (2.19)$$

Then, we need to form a table of energy increments for each subchannel. For the i^{th} subchannel

$$\Delta e_i(b) = e_i(b) - e_i(b - 1) = \frac{2^{b-1} * \Gamma}{SNR_i} \quad (2.20)$$

This table provides incremental energies required for each subchannel to transition from supporting $b - 1$ bits to b bits, given the channel gain and noise Power Spectral Density (PSD). The required energy increment to increase from the maximum supportable bits is set to infinity (or a very high value). Note that, since we are restricted to a limited set ($\{0, 1, 2, \dots, \log_2 M\}$), the energy for other number of bits is averaged using the energy increments for allowed number of bits. This averaging could be applied to all but the final bit to

even out the total number of bits on that subchannel. [48] has also proposed an algorithm to resolve the last bit issue, too.

After the initialization is performed, the authors in [48] take the initial bit allocation, b , the total number of bits to be allocated, B as the input, and find the optimized bit allocation. The algorithm starts with $B' = \sum_n b(n)$. If $B' > B$, it finds the subcarrier that maximizes the energy to decrement from b_n bits to $b_n - 1$ bits, $\Delta e_n(b_n)$, and decrease one bit from that subcarrier. It then updates B' , accordingly. The algorithm continues as long as $B' \neq B$.

However, if at the beginning, $B' < B$, it finds the subcarrier that minimizes the energy decrement from $b_n + 1$ bits to b_n bits, $\Delta e_n(b_n + 1)$, and increments the bit allocation of that subcarrier. It then updates B' , accordingly. The algorithm continues as long as $B' \neq B$.

These algorithms constitute a complete characterization of the bit loading procedure proposed in [48, 49] for a given frequency selective channel.

Hughes-Hartogs [50], proposed another method in which the bits are assigned one by one to the subchannel with the lowest power increment, until a pre-specified target rate R_T is reached. Obviously this is a very slow procedure and requires lots of sorting and searching.

Another loading algorithm has been proposed in [51] where the distribution of bits is adapted to the shape of the transfer function of each subchannel. This is basically very close to what has been proposed in [49], except a different approach is used to obtain the bit capacity of each subchannel:

$$C'_{QAM} \simeq 0.31 \left(10 \log \frac{S}{N} - 0.67 \right) \quad (2.21)$$

This is rounded to the maximum integer m_i , which is smaller than C'_{QAM} . It is assumed that the required number of bits per OFDM symbol (M) is fixed and in each iteration, depending on the relation between M and $m_\sigma = \sum_i m_i$, the number of bits in each individual subchannel is reduced or increased.

Fischer and Huber [52] exploit the fact that the signal power and the rate at each subchannel are related. They minimize the BER at each subchannel with a constant data rate and transmission power. He characterizes the subchannels as AWGN and use the following relation to find the symbol error rate at each subchannel (assuming QAM)

$$p_e(i) = K_i Q \left(\sqrt{\frac{d_i/4}{N_i/2}} \right) \quad (2.22)$$

where the term under the square root is equal to the constant SNR at each subchannel, $Q(\cdot)$ is the Gaussian pdf tail, d_i is the minimum Euclidean distance between signal points in the constellation and $N_i/2$ is the noise power.

2.5 Peak to Average Power Ratio (PAPR)

A major hurdle to the widespread use of OFDM is the high Peak to Average Power Ratio(PAPR) of OFDM signals. An OFDM signal consists of a number of independently modulated subcarriers, which can give rise to a large Peak to Average Power Ratio (PAPR), when added up coherently. When N equi-amplitude signals are added with the same phase, they produce a peak power that is N times the average power. The peak power is defined as the power of a sine wave with an amplitude equal to the maximum envelope value. Hence, an unmodulated carrier has a PAPR ratio of 0 dB. An alternative measure of the envelope variation of a signal is the crest factor, which is defined as the maximum signal value divided by the rms signal value. For an unmodulated carrier, the crest factor is $3dB$. This $3dB$ difference between PAPR and crest factor also holds for other signals, provided that the center frequency is large in comparison with the signal bandwidth. Usually the transmitters are constrained to a limited peak power. If the peak envelope power is subject to a design or regulatory limit, then this has the effect of reducing the mean envelope power

allowed under OFDM, relative to that allowed under constant envelope modulation. This reduces the effective range of the OFDM transmissions and is particularly acute in mobile applications where battery power is a constraint. Moreover, to prevent signal distortions and spectral growth due to non-linearities inherent in electronic components, power amplifiers must operate below their compression point, where power is converted most efficiently. This results in more expensive and inefficiently used components. The high PAPR or high crest factor could cause problems when the signal is applied to a non-linear device such as a power amplifier, since it results in-band distortion and spectral spreading. To counteract these effects, the amplifier needs to be highly linear or operate with a large back-off. Both approaches result in a severe power efficiency penalty and are expensive [53]. An analysis of non-linear amplifiers, however, points out that an acceptable performance regarding the in-band distortion is obtained with an analysis of non-linear amplifiers, however, points out that an acceptable performance regarding the in-band distortion is obtained with only a small back-off [54].

The tolerable out-of-band radiation or spectral spreading sets the bound on the back-off that is needed [53]. As a conclusion, simply dimensioning the system components to be able to cope with the worst-case signal peaks is practically impossible. That's why many researchers have been trying over the years, to counteract the PAR problem.

PAPR can be defined in different ways. It can be defined in discrete domain in complex continuous domain, or in real continuous domain. The definitions in continuous domain could be also extended to baseband as well as the pass-band OFDM signals.

Assume that the sequence x_n , $n = 0, 1, \dots, N - 1$ is applied to the IFFT block in an OFDM transmitter. The discrete-domain PAPR of this codeword is

$$PAPR(x) = \frac{1}{P_x} \max_k \{p_x[k]\} = \frac{1}{P_x} \max_k \left[\sum_{i=0}^{N-1} x_i x_u^* e^{\frac{j2\pi(i-u)k}{N}} \right]^2, \quad (2.23)$$

where $P_x = \|x\|^2$ is the energy of the codeword x . The complex domain continuous pass-band OFDM signal is

$$s_x(t) = \sum_{i=0}^{N-1} x_i \exp(j2\pi(f_c + i\Delta f)t). \quad (2.24)$$

The instantaneous envelope power of the signal is

$$p_x(t) = |s_x(t)|^2 = \sum_{i=0}^{N-1} \sum_{u=0}^{N-1} x_i x_u^* \exp(j2\pi(i-u)\Delta f t). \quad (2.25)$$

The PAPR of the sequence x is then defined as

$$PAPR(x) \triangleq \frac{1}{P_x} \max_{0 \leq t \leq T} \{p_x(t)\}, \quad (2.26)$$

where T is the length of OFDM block.

Some papers have introduced another term, Peak to Mean Envelope Power Ratio (PMEPR) to describe the statistical definition of PAPR. PAPR is also defined for a code applied to an OFDM system. Let \mathcal{C} be a code that maps blocks of k input bits into blocks of N constellation symbols from a constellation \mathcal{Q} with 2^a elements. The rate of \mathcal{C} is defined to be $R = k/aN$. The codeword $c = [c_0 \ c_1 \ \dots \ c_{N-1}]$ is applied to IFFT block and the transmitted signal can be derived by $\text{Re}(s_c(t))$ for $0 \leq t \leq T$, where $\text{Re}(\cdot)$ denote the real part of a signal. The relation between the quantities Δf and T depends on whether a guard interval is appended or not. However, we note that $\Delta f = 1/T$ is commonly assumed in an ideal situation.

Assuming that $f_c/\Delta f \gg 1$, it is well known that the average envelope power of $s_c(t)$ is $\|c\|^2 = \sum_{i=0}^{N-1} |c_i|^2$. Thus, the PAPR of the signal is given by

$$PAPR(c) = \max_{0 \leq t \leq T} \frac{[\text{Re}(s_c(t))]^2}{P_{av}}, \quad (2.27)$$

where $P_{av} = E(\|c\|^2) = \sum_{c \in \mathcal{C}} \|c\|^2 p(c)$, and $p(c)$ is the probability of the transmission of the codeword c . The value P_{av} is referred to as the average square length of the transmitted

codewords. The PAPR of the codebook \mathcal{C} is defined to be

$$PAPR(\mathcal{C}) = \max_{c \in \mathcal{C}} [PAPR(c)]. \quad (2.28)$$

We also let $P_{max} = \max_{c \in \mathcal{C}} \|c\|^2$, and refer to it as the maximum square length of the transmitted codewords. The maximum square length P_{max} is the maximum of the average envelope power of signals corresponding to different codewords. Thus, P_{max}/P_{av} is not to be mixed up with $PAPR(\mathcal{C})$. For instance, if an equal-energy constellation is used, then $P_{max}/P_{av} = 1$, but the $PAPR(\mathcal{C})$ can be as large as N .

To evaluate the PAPR of continuous signals, we need to sample the signal with some specific rate. If the sampling period is $1/N\Delta f = T/N$, the continuous PAPR and discrete PAPR would be the same. However, to obtain more accurate value for continuous PAPR, we need to oversample the OFDM signal with a rate that is normally a multiple of $N\Delta f$ [55].

2.5.1 Statistical Properties of OFDM Signals

This section describes the statistical properties of OFDM signals assuming that the discrete PAPR is considered (critical sampling). Let's assume that the symbols $x_n, n = 0, 1, \dots, N-1$ are i.i.d complex Gaussian random variables with zero mean and unit variance. It can be shown that the time domain symbols X_k are also i.i.d Gaussian random variables with zero mean and unit variance. Even if x_n 's are not Gaussian, using Central Limit Theorem (CLT) it is easy to see that the time domain symbols which are a linear combination of N frequency domain symbols, are still i.i.d Gaussian, with unit variance, since IDFT operation

is an orthogonal transformation. To obtain independence, we can see that

$$\begin{aligned}
E [X_r X_s^*] &= \frac{1}{N} E \left[\left(\sum_{l=0}^{N-1} x_r e^{j2\pi lr/N} \right) \left(\sum_{k=0}^{N-1} x_s e^{j2\pi ks/N} \right)^* \right] \\
&= \frac{1}{N} \sum_{n=0}^{N-1} E [x_r x_s^*] e^{j2\pi lr/N} e^{-j2\pi ks/N} \\
&= \delta(r - s).
\end{aligned} \tag{2.29}$$

In most practical applications, the data is randomized prior to modulation and the frequency domain symbols can be approximated as independent discrete uniform random variables, typically MQAM, MPSK, or APSK. For these cases, the symbol samples X_n are linear combination of N discrete uniform random variables. For the OFDM cases, all symbols x_k are chosen from the same constellation and thus the N discrete uniform random variables are i.i.d. Since the symbols x_n are independent, the symbol samples X_k 's are still uncorrelated. Moreover, for large N 's, the CLT leads to the common assumption that the symbol samples are approximately i.i.d random variables. With this assumption, the Commutative Distribution Function(CDF) of the random variables $PAPR(x)$ has a simple closed distribution:

$$\begin{aligned}
Prob [PAPR(x) < \gamma^2] &= Prob \left[\frac{|X_0|^2}{P_x} < \gamma^2, \dots, \frac{|X_{N-1}|^2}{P_x} < \gamma^2 \right] \\
&= \left(Prob \left[\frac{|X_0|^2}{P_x} < \gamma^2 \right] \right)^N
\end{aligned} \tag{2.30}$$

For real baseband symbols, we have

$$Prob [PAPR(x) < \gamma^2] = (1 - 2Q(\gamma))^N. \tag{2.31}$$

For the complex OFDM symbols the CDF is turned out to be

$$Prob [PAPR(x) < \gamma^2] = (1 - \exp(-\gamma^2))^N. \tag{2.32}$$

Therefore, for complex signals, the Complementary Cumulative Distribution Function (CCDF), or the probability that the PAPR exceeds some value γ^2 is [56]:

$$\begin{aligned} Prob [PAPR(x) \geq \gamma^2] &= 1 - (1 - \exp(-\gamma^2))^N \\ &= 1 - \exp\left(-\sqrt{\frac{\pi}{3}}N\gamma e^{-\gamma^2}\right). \end{aligned} \quad (2.33)$$

γ could be considered as a clipping point. Setting the value of γ at such a level that the clipping noise is negligible (e.g. 50dB below the signal level is obtained when $\gamma > 4$ [57]) is not optimal. Lowering the dynamic range of the A/D and D/A converters for a constant number of bits reduces the quantization noise significantly. The clipping noise however increases, as the clipping probability is larger. So, we need to make a trade-off between the quantization distortion and the clipping distortion to minimize the overall distortion [58]. Alternatively, for a constant SNR, the number of bits in the D/A and A/D can be decreased, lowering the implementation cost.

2.5.2 Techniques for OFDM PAPR Reduction

There are three main classes of methods to reduce the PAPR. The methods based on block coding, the methods based on clipping, and the probabilistic methods.

PAPR Reduction with Distortion (Clipping)

These methods clip the OFDM signal, if it is above the dynamic range of the A/D and D/A, or in some cases above a predetermined threshold. However, clipping causes in-band (for oversampled) and out-of band distortion (for unoversampled or analog signals). These methods try to reduce the effect of clipping. There are three main clipping schemes. One is block scaling in which an optimal clipping threshold is determined out of a limited set. This threshold

determination is performed on a symbol-by-symbol basis [59]. The selected threshold is then transmitted to the receiver in a reserved tone.

In the second scheme, clipping is performed at the transmitter, but the receiver tries to compensate for some of its effects. The receiver needs to estimate the size and the location of the clip [60]

Finally, the third method applies some signal processing to reduce the effect of clipping. Two kinds of processing are applied, namely peak windowing in which the large peaks are multiplied by a small window like Kaiser or Hamming [61], and adding correction function in the vicinity of the clip [62]. Both approaches decrease the out-of-band distortion by smoothing the hard limiting effect.

Block Coding schemes

By block coding we limit the set of possible signals that can be transmitted. Only those signals with low amplitude are allowed. Therefore no clipping occurs and therefore no distortion is implied. These codes not only offer a low PAPR, but have the error-correcting capabilities. As a matter fact the CCDF in Eq. (2.33) is the fraction of symbols that are discarded for a given γ . This expression shows that as N increases, the proportion of sequences to be discarded goes up. As a matter of fact, no good codes for practical values of $N > 64$ are known. In other words, for large N the code rate is very low, and this is an inherent property of coding methods. One simple strategy is to perform exhaustive search over all possible codewords and use a table lookup [63]. Other options restrict the phase possibilities of certain tones [64] or only use part of the bits in a differential phase modulation scheme [65].

Some codes have the property that they have a small PAPR by an instantaneous power that is most of the time close to the average power. Thus, the spectrum of the code is almost

flat, or alternatively an impulse-like autocorrelation [66, 67].

Two codes based on this criterion are Golay complementary sequences, which we describe in details in Chapter 5, and m-sequences [67].

The m-sequences are a class of $(2^m - 1, m)$ linear cyclic codes that create OFDM block length of $N = 2^m - 1$. They are used for generating pseudo-noise sequences for spread spectrum communications, because their autocorrelation functions is an impulse and therefore their spectrum is almost flat, implying a very low PAPR. Their autocorrelation function is

$$A_c(j) = \sum_{i=0}^{N-1} c_i c_{[i+j \bmod N]} = \begin{cases} 2^m - 1, & j = 0; \\ -1, & 1 \leq j \leq 2^{m-1} - 2. \end{cases} \quad (2.34)$$

All these block codes provide a low PAPR (typically below $3dB$). However, the most important drawback of these codes is that their code rate for large N is not acceptable. This drawback dramatically limits their usefulness with regard to real applications.

Probabilistic methods, Parameter Optimization

These methods do not reduce PAPR, instead they reduce the probability of the occurrence of large peaks. This causes the CCDF shown in (2.33) to reduce for any given γ . The basic way to reduce this probability is by performing a linear transformation on the input vector of IFFT block [68]

Partial Transmit Sequence (PTS): In PTS, the input vector x is now subdivided into V non-overlapping subvectors x_v of size N/V [69]. Each carrier in the subvector x_v is multiplied with a rotation factor θ_v . The rotation factors of the different subvectors are statistically independent. This scheme corresponds to a linear transformation, where the additive vector is all-zero. Because of linearity of IFFT, this is equivalent to applying IFFT to each subvector (by setting other elements to zero) and then multiplying the result by θ_v . Fig. 2.5 depicts a block diagram of this method. One advantage of this method is that it

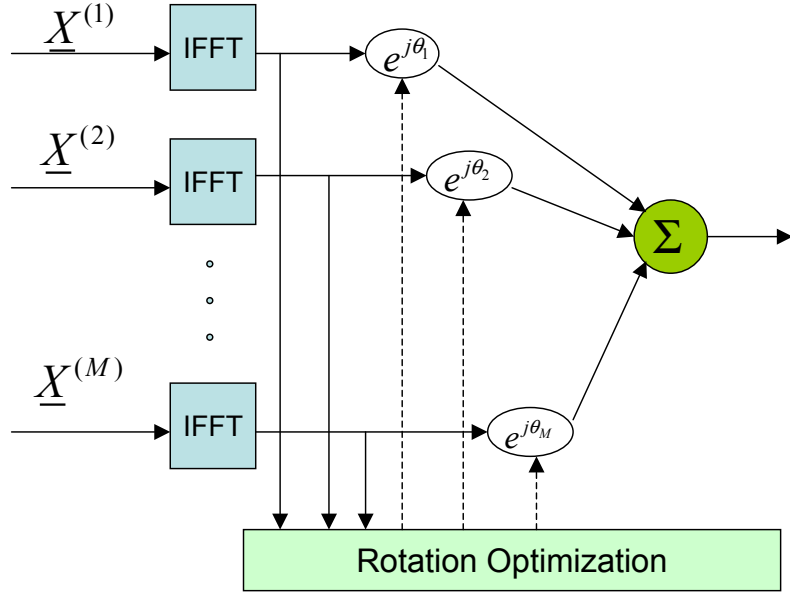


Figure 2.5: Partial Transmit Sequences for PAPR reduction in an OFDM.

reduces the complexity of IFFT operation, because the length of each IFFT is N/V . If the rotation factors θ_v are chosen from a set of size L , the total number of phases is L^V . For each set of rotations, the PAPR is computed and the one with the lowest PAPR is transmitted. The index for the chosen set of phases must be transmitted to the receiver using $V \log_2 L$ bits. It is also possible to use differential modulation for each subvector and in this case no side information is sent.

Selective Mapping: The basic idea is to have L statistically independent vectors to represent the same information [70]. The vector with the lowest PAPR is selected for transmission. Since these vectors are independent, the probability that PAPR exceeds some value γ is therefore equal to CCDF in (2.33) to the power of L , and therefore is significantly lower. Fixed, independent rotation vectors are used to create these L independent vectors. Fig. 2.6 depicts the building block of SLM.

The authors in [71] use L cyclically non-correlated m-sequences to generate these L

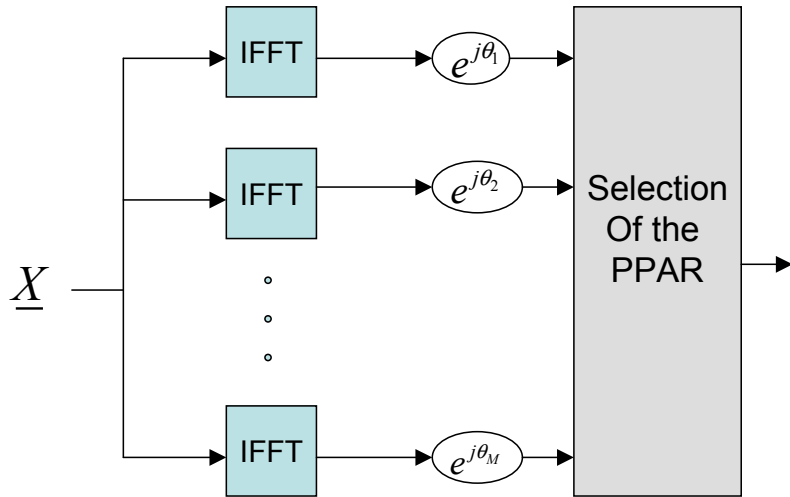


Figure 2.6: Selective Mapping for PAPR reduction in an OFDM.

replicas, rather than rotations. This is different from using m-sequences as in Section 2.5.2. Here, The input vector is multiplied by an m-sequence and therefore a more flat spectrum is obtained and therefore the amplitude peaks are reduced.

Tone Reservation and Injection

The idea of tone reservation is to use a shaping function such that the clipping noise is concentrated on unused tones. Unused tones are the tones in which the SNR is low and therefore the bit loading algorithms allocate no data (or very small number of bits) to those subcarriers. This problem is equivalent to finding the closure of some convex sets [60].

In tone injection, the QAM constellation is extended such that the same data point corresponds to multiple possible constellation points [72, 73]. Fig.fig:TI illustrates an example of such an extension.

The advantage of this scheme is that receiver does not need any side information from the transmitter, but it increases the energy of some point compared to the original constellation points. The optimization problem that works per tone and repeats iteratively for all tones

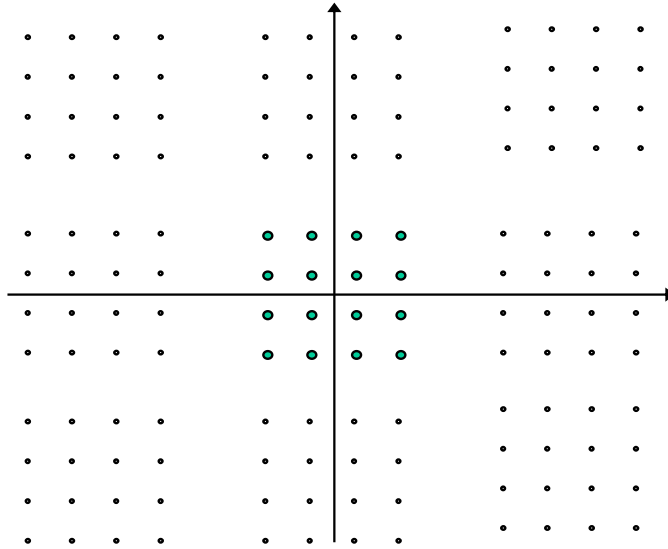


Figure 2.7: The extension of 16QAM constellation.

is to choose one of the constellation points corresponding to the tone's substream such that the reduction in PAPR is maximized while the energy of the symbol is as low as possible.

2.6 Orthogonal Frequency Division Multiple Access

There are many multiple access schemes to accommodate multi user systems. In Section 1.4 we briefly listed some of these schemes. In summary, in TDMA, time is the resource that is shared among different users, but every user uses the whole bandwidth. In FDMA, the whole available spectrum band is divided into several slots and each user modulates its data over one band. However, all user send their data simultaneously. In CDMA, all users send their data simultaneously using the whole spectrum, but are distinguished by the spreading code they use. Orthogonality of the codes, and power control are very essential in this system. In SDMA, users are separated by their location and the space diversity is essential in this system.

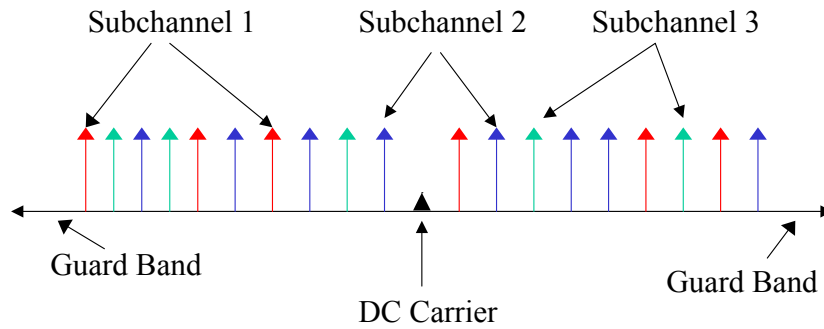


Figure 2.8: OFDMA carrier segmentation.

Orthogonal Frequency Division Multiple Access (OFDMA) is another multiple access system that uses the subcarriers of an OFDM transmitter to differentiate different users. The functionality of OFDMA is different in down-link or up-link transmission. Here, the active carriers (the carriers that are aimed to carry data, not pilot or NULL) are divided into subsets of carriers, each subset is termed a subchannel [5]. In the down-link, a subchannel may be intended for different groups of receivers, while in the up-link, a transmitter may be assigned one or more subchannels, several transmitters may transmit in parallel. Note that in the up-link, all users are transmitting at the same time (unlike TDMA), using the same frequency band (unlike FDMA), without using any orthogonal spreading code (unlike CDMA), and assuming no ISI or ICI, there is no interference among all users. Because of this separation, performing power control is not very crucial in OFDMA. However, if the number of users is high, and all of them are required to be assigned distinct subchannels, the transmission rate will fall significantly. For this reason, we might need to allow some users to share one or more subchannels. As a result we might need to use power control and/or multiple transmit antennas to reject interference. The concept of carrier segmentation in OFDMA is shown in Fig. 2.8.

In addition to the support for multiple access, OFDMA allows for scalability, and the use of multiple antenna signal processing capability.

The up-link OFDMA block transmitted from user n is

$$s(t) = \Re \left\{ e^{j2\pi f_c t} \sum_{N_{start}+1}^{N_{start}+N_{used}-1} c_k e^{j2\pi k \Delta f (t-T_g)} \right\}, \quad (2.35)$$

where t is the time elapsed since the beginning of the subject OFDM block, with $0 < t < T_g$, c_k is the complex symbol to be transmitted at subcarrier k , T_g is the guard time, T_S is the OFDM symbol duration, including guard time, Δf is the carrier frequency spacing. The carrier N_{start} is excluded to transfer the DC component.

One slot in IEEE802.16 is defined as a group of contiguous subchannels, in a group of contiguous OFDMA blocks. This allocation can be seen as a two dimensional rectangle, where the horizontal component is the time domain, in which a specific number of OFDMA blocks are included, and the vertical component includes the number of subchannels. The whole slot is called a data region.

2.6.1 Channel Allocation

Channel allocation is an integral part of OFDMA. Channel allocation is also the most important part of any multiple access scheme. In TDMA time slot allocation, in FDMA carrier frequencies, in CDMA spreading codes, in SDMA antenna beams, and finally in OFDMA subcarriers (or a group of subcarriers named subchannel). The performance of any channel allocation algorithm is measured by the number of users it can accommodate given a limited resource, and the quality of transmission in each channel (capacity of the system). Channel allocation algorithms aim to maximize the reuse factor. However, some procedures might be taken to cancel the effect of possible interference. One goal in channel allocation is to increase the total system throughput, if users have different rate requirements.

Channel allocation could be done either dynamically or in a fixed manner. In Fixed

Channel Allocation (FCA), a set of channels are allocated permanently to a cell and all users in that cell use available channels. New users are admitted to the cell if there is any channel available. In Dynamic Channel Allocation (DCA), each cell has a reuse distance constraints. Channels can be reused in different cells if they are well separated and the amount of interference is low enough.

Another scheme is the reassignment of channels to different users to accommodate newly coming user. The performance of these methods which are called Maximum Packing (MP) has the upper bound on the performance of other algorithms in terms of the number of users it can accommodate [74].

Channel Allocation in OFDMA

The carriers forming one subchannel may, but need not be adjacent. Except the guard tones and the DC carriers, other carriers are allocated to pilot and data both in down-link and up-link. In down-link the pilot tones are allocated first and the rest are grouped in subchannels and allocated to data, and therefore there is a common fixed pilot carriers. However, in up-link the set of used carriers are first grouped into subchannels, and then the pilot carriers are allocated from within each subchannel, so each subchannel contains its own pilot symbols. The reason for this set up is that in down-link OFDM every subscriber receives the signal, but in up-link each subchannel may be transmitted from a different user. Note that in each case, there are two kinds of pilot carriers, fixed location, and variable location pilots, which shift their location every block repeating every 4 blocks for down-link and 13 blocks for up-link. For down-link, the remaining carriers are first divided into groups of contiguous carriers, and then each subchannel consists of one carrier from each of these groups. The partitioning of subcarriers into subchannels is performed according to the following equation,

which is called a permutation formula:

$$carrier(n, s) = N_{subchannels} \cdot n + \{p_s[n_{mod}(N_{subchannels})] + ID_{cell} \cdot ceil[(n+1)/N_{subchannels}]\}_{mod}(N_{subchannels}), \quad (2.36)$$

where s is the index number of a subchannel, selected from the set $[0, \dots, N_{subchannels} - 1]$, n is the carrier-in-subchannel index from the set $[0, \dots, N_{subcarriers} - 1]$, $carrier(n, s)$ is the carrier index of carrier n in subchannel s , $N_{subchannels}$ is the number of subchannels, $p_s[j]$ is the series obtained by rotating a permutation cyclically to the left s times, $ceil[.]$ is the function that rounds its argument up to the next integer (ceiling), ID_{cell} is a positive integer assigned by the MAC to identify this particular base station sector.

Chapter 3

Power Allocation for OFDM using Adaptive Beamforming over Wireless Networks

3.1 Motivation and Previous Works

Orthogonal Frequency Division Multiplexing (OFDM) is a parallel data transmission scheme. If the width of each subchannel is smaller than the coherence bandwidth of the channel, it converts the wideband frequency selective fading channel to a series of narrowband flat fading subchannels [75]. Depending on the carrier spacing, data rate, and the coherence bandwidth of the channel, there is no need for sophisticated equalization methods [76]. One of the disadvantages of OFDM is the worst subchannel domination [51, 77]. In an uncoded OFDM system with fixed modulation scheme for all subchannels, the error probability of the whole system is dominated by the subchannel with the highest attenuation [51, 77]. If the SINR fluctuates over subchannels, the ones with the worst SINR would affect the overall BER the most. As a result, in the case of frequency selective fading channels, the performance of the whole system in terms of error probability will improve slowly by increasing the transmitted power. In order to obtain a minimum overall error probability, the optimum algorithm, in a fixed total power policy, is to have a uniform error probability for all of the subchannels

[77].

Several methods have been proposed to combat the aforementioned problem. One solution is to use Coded OFDM (COFDM) [78–80]. Other methods try to adjust the bit and power distribution among subchannels according to their link gains and are mostly called “loading algorithms” [49–52]. In most of these algorithms, the bit allocation in each subchannel is adapted to its capacity and therefore, a fixed modulation scheme is not considered for all subchannels. Some of the loading algorithms were introduced in Section 2.4.

However, most of these methods have been proposed for a single transmitter with a single antenna without considering the effect of interferences. In this chapter, we are looking at the problem of power allocation for different subchannels from a different point of view. In a mobile environment, each user’s signal can affect others and this in turn, results in more interference. The loading algorithms proposed in the previous works do not consider such phenomena and therefore cannot reach the optimum solution in the sense of minimum total transmission power. In contrast, we use an adaptive power allocation scheme to distribute the powers at each subchannel based on the interference from other users at the same subchannel.

Furthermore, we exploit antenna arrays to further reduce the interference. We will consider the antenna arrays in two situations. First we assume that the base station knows the array response and perform both frequency-domain and time-domain beamforming. Then we relax this assumption and use MMSE beamforming in which training sequences are exploited to update the weight vectors and minimize the interference. These training sequences are transmitted to the receiver to estimate the channel response. The rate at which the training sequences are transmitted depends on the speed of channel variation. In each method, total interference and noise is calculated and fed back to the transmitter through a feedback channel.

It should be noted here that sometimes because of deep frequency selective fading, the

tone to tone SNR differences are very large and therefore we might need tremendous amount of power adjustment to compensate for SNR fluctuations. In this case, it might be better to discard the subchannel, instead of allocating a large portion of power to it. However, this can result in data rate reduction. Besides, it is not always possible to discard some parts of the signal. As an example in IEEE802.11a wireless LAN protocol, we use some of the subchannels to transmit pilot signals, or we transmit the preambles using OFDM symbols, which are very important to recover. In these situations, we have to do proper bit or power allocation to save the signal. Another approach is to limit the transmitted power at each subchannel.

In COFDM, by coding across subchannels, the BER is averaged over all of the subchannels. However, if by exploiting the power control and beamforming, the SINR at all of the subchannels can be increased, the overall BER is decreased. Moreover, using power control and beamforming can reduce the total network power. We will compare the uncoded OFDM using our proposed algorithms with COFDM using power control and beamforming and also with a COFDM system with no power control or beamforming .

In this chapter we will assume that each mobile uses all of the subchannels. However, with a slight modification, the same formulations and the same algorithms can be applied to OFDMA, where different subchannels could be assigned to different users.

This chapter is organized as follows: In Section 3.2 we will review the concept of power control and propose the power control for OFDM receivers. Section 3.3 proposes the OFDM joint power control and frequency-domain beamforming. Joint time-domain beamforming and power control is proposed in Section 3.4. In Section 3.5 we will use MMSE approach to perform the beamforming. Section 3.6 extends the proposed algorithm to COFDM. Section 3.7 presents some simulation results, and finally Section 3.7 concludes the chapter.

3.2 OFDM with Adaptive Power Control

3.2.1 Background

The objective of power control in wireless networks is to minimize the transmitted power while some target error probabilities are met [81,82]. Consider a network of M mobiles trying to access the same channel. We denote the power link gain between the i^{th} mobile and the b^{th} base station by a real number G_{ib} , and the i^{th} mobile transmitted power by P_i . we assume that one base station is assigned to each mobile. Moreover, these base stations are co-channel in the sense that they use either the same frequency band in FDMA, or the same time slot like in TDMA, or spreading code like in CDMA. As a result they all suffer from co-channel interference. The SINR at the b^{th} receiver is given by:

$$\Gamma_b = \frac{G_{bb}P_b}{\sum_{\substack{i=0 \\ i \neq b}}^{M-1} G_{ib}P_i + N_b}, \quad (3.1)$$

where N_b is the noise power at the b^{th} base station. The objective is to maintain the total transmitted power as low as possible, while the SINRs are kept above a threshold. If we denote the minimum acceptable SINR at base b by γ_b , The requirement for acceptable link quality is

$$\Gamma_b \geq \gamma_b, \quad 1 \leq b \leq M,$$

or in matrix form:

$$[\mathbf{I} - \mathbf{DF}]\mathbf{P} \geq \mathbf{u}, \quad (3.2)$$

where \mathbf{I} is an $M \times M$ identity matrix, \mathbf{P} is the power column vector, \mathbf{D} is a diagonal matrix whose b^{th} diagonal element is $\frac{\gamma_b}{G_{bb}}$, and $[\mathbf{u}]_b = \frac{\gamma_b N_b}{G_{bb}}$, and

$$[\mathbf{F}]_{ij} = \begin{cases} 0 & \text{if } j = i \\ G_{ji} & \text{if } j \neq i \end{cases}, \quad (3.3)$$

If there is at least one power vectors that satisfies (3.2), the SNR thresholds could be achieved. The power control problem now is defined as follows:

$$\begin{aligned} & \text{minimize} && \sum P_i, \\ & \text{subject to} && [\mathbf{I} - \mathbf{DF}]\mathbf{P} \geq \mathbf{u}. \end{aligned}$$

The matrix \mathbf{DF} is called the gain matrix. Let's call the spectral radius of this matrix $\rho(\mathbf{DF})$. The Perron-Frobenius theorem [83] says that any positive definite (or irreducible) matrix has a positive real eigenvalue λ^* such that $\lambda^* = \max\{|\lambda_i|\}_{i=1}^M$, where λ_i 's are the eigenvalues of the matrix. It has been shown in [84] that if the spectral radius of the gain matrix is less than unity, then the matrix $\mathbf{I} - \mathbf{DF}$ is invertible and positive. Such a network is called a "feasible network". For a feasible network, the lowest possible total power is obtained when all of the SINRs are equal to the threshold, i.e.

$$\Gamma_b = \gamma_b, \quad b = 0, \dots, M - 1. \quad (3.4)$$

This is translated to

$$\hat{\mathbf{P}} = [\mathbf{I} - \mathbf{DF}]^{-1}\mathbf{u}. \quad (3.5)$$

The above formula can be rewritten as

$$\hat{\mathbf{P}} = \mathbf{DF}\hat{\mathbf{P}} + \mathbf{u} \quad (3.6)$$

This can lead us to the following iterative equation

$$\mathbf{P}^{n+1} = \mathbf{DFP}^n + \mathbf{u} \quad (3.7)$$

A distributed power update scheme is proposed in [85] that achieves the optimal solution

for (3.1). The b^{th} mobile power at the n^{th} stage of iteration is updated by

$$P_b(n+1) = \frac{\gamma_b}{G_{bb}} \left(\sum_{\substack{i=0 \\ i \neq b}}^{M-1} G_{ib} P_i(n) + N_b \right). \quad b = 0, \dots, M-1 \quad (3.8)$$

The right hand side in (3.8) is a function of the noise and interference at the b^{th} base station (the term inside parenthesis), the link gain G_{bb} , and the target SINR. All of these can be measured locally and transmitted through a feedback channel to the corresponding mobile [81]. So, transmitters need not to know all the existing path gains and transmitter powers. At each iteration, transmitters update their powers using the total interference and link gain to its corresponding receiver, that is fed back by the receiver.

If the network is feasible, the above iteration converges to the optimal power vector $\hat{\mathbf{P}}$. It has been shown in [86] that starting from any arbitrary power vector, this solution converges to the optimal solution $\hat{\mathbf{P}}$.

In the following, we consider this scheme in a multiuser environment using multicarrier transmission. Our objective is to optimize the power allocation at each subchannel for all of the mobiles, such that:

1. The SINR at all of the subchannels for all of the mobiles are close to each other and they are above a SINR threshold. This causes the error probability to decrease faster by increasing the transmitter power, compared to that of unbalanced SINRs.
2. The total power used to achieve the aforementioned objective is minimized.

The basic idea is to allocate less power to the subchannels with less interference, and more power to the subchannels with lower SINR.

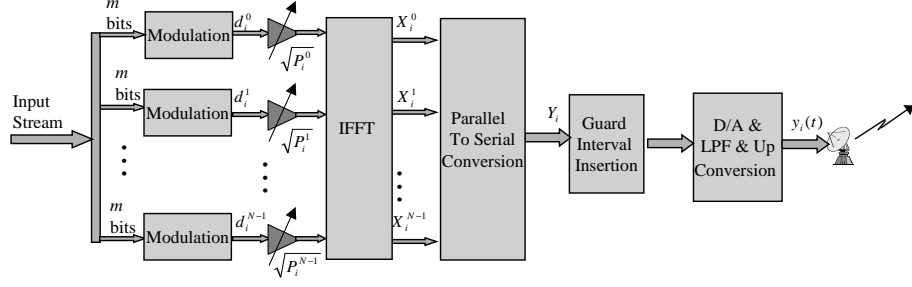


Figure 3.1: The i th OFDM transmitter using Adaptive Power Control.

3.2.2 System Configuration

Fig. 3.1 depicts the transmitter proposed in OFDM systems using adaptive power control at each subchannel. If the maximum number of paths between the i^{th} user and the b^{th} base station is assumed to be L , the corresponding link can be modelled by the following impulse response (we have ignored the doppler effect):

$$h_{ib}(t) = \sqrt{G_{ib}} \sum_{l=0}^{L-1} \alpha_{ib}^l \delta(t - \tau_{ib}^l), \quad (3.9)$$

where α_{ib}^l denote the l^{th} path fadings that are independent complex Gaussian variables with variance $\sigma_{ib}^l{}^2$ (their amplitudes are Rayleigh). τ_{ib}^l 's are the delays of the corresponding paths. G_{ib} is a real random variable representing the log-normal shadow fading and path loss. The frequency response of the channel is

$$H_{ib}(f) = \sqrt{G_{ib}} \sum_{l=0}^{L-1} \alpha_{ib}^l e^{-j2\pi f \tau_{ib}^l}. \quad (3.10)$$

In this chapter the vectors are shown by bold underline letters. Moreover, the transmitted and received signals at the time domain are shown by uppercase and the same values at the frequency domain by lowercase letters. Let's assume that N denotes the number of subchannels, T_s is the symbol period, and f_0 is the carrier frequency. If N is large enough, each subchannel can be modelled as a flat fading channel [75] and so the link gain at subchannel

c , H_{ib}^c can be calculated simply by replacing f with $f_c \triangleq f_0 + \frac{c}{NT_s}$ in (3.10), i.e.

$$H_{ib}^c = \sqrt{G_{ib}} \sum_{l=0}^{L-1} \alpha_{ib}^l e^{-j2\pi f_c \tau_{ib}^l}. \quad (3.11)$$

Without loss of generality, we can assume the path loss and shadowing for different paths to be the same, and any difference can be absorbed in fading coefficients

In this chapter, we assume that a proper guard interval has been inserted in time domain such that the effect of Inter-Symbol Interference (ISI) can be ignored. Moreover, the guard interval has the form of cyclic prefix and therefore the interaction of the received signals at different subchannels in the frequency domain is zero. This is due to the cyclic convolution performed between the channel and the transmitted signal. The modulated data at subchannel c for user i is d_i^c whose energy is assumed to be unity (e.g. using a fixed MPSK on all subchannels).

We assume that all of the subchannels use the same modulation and so due to the fact that subchannel link gains are different, by distributing the power equally among them, the SINRs would become unbalanced. Now, we perform a power control algorithm at each subchannel separately. If we denote the power allocated to mobile i at subchannel c by P_i^c and define $W_N \triangleq e^{-j2\pi/N}$, the k^{th} sampled received signal ($k = 0 \dots N-1$) at the b^{th} receiver (after down conversion, guard interval removal, proper matched filtering, and sampling at intervals T_s) will be

$$\hat{X}_b^k = \frac{1}{\sqrt{N}} \sum_{i=0}^{M-1} \left(\sum_{c=0}^{N-1} \underbrace{H_{ib}^c \sqrt{P_i^c} d_i^c}_{t_{ib}^c} W_N^{-ck} \right) + n_b^k \quad (3.12)$$

$$= \frac{1}{\sqrt{N}} \sum_{c=0}^{N-1} \left(\sum_{i=0}^{M-1} t_{ib}^c + \hat{n}_b^c \right) W_N^{-ck}, \quad k = 0 \dots N-1 \quad (3.12')$$

where n_b^k is the k^{th} noise sample, and \hat{n}_b^c is the Fourier transform of the noise samples. Since the noise samples are uncorrelated, these two variables have the same power, N_b .

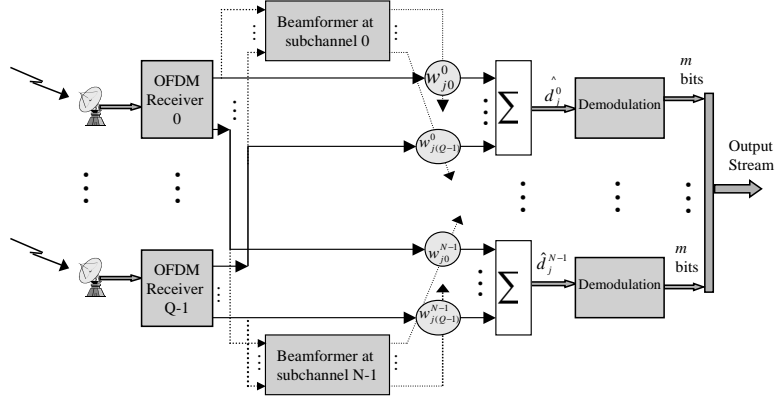


Figure 3.2: Frequency-Domain Beamforming in the j th OFDM receiver.

The signal at subchannel c for base station b in frequency domain is

$$\hat{d}_b^c = \sum_{i=0}^{M-1} t_{ib}^c + \hat{n}_b^c = t_{bb}^c + \left[\sum_{i \neq b}^{M-1} t_{ib}^c + \hat{n}_b^c \right], \quad i = 0, \dots, N-1 \quad (3.13)$$

where the first part is the desired signal attenuated by the link gain, and the term inside the bracket is the sum of the interferences and thermal noise. The SINR at the c^{th} subchannel, Γ_b^c , is given by (3.1) as a function of link gain, power value, and noise at the c^{th} subchannel.

Our goal is to maintain Γ_b^c above a target value, γ_b^c while the sum of allocated powers is minimized. To achieve this goal, we apply the power control algorithm, described in the previous subsection, to each subchannel independently. Since the subchannels are assumed to be orthogonal, this guarantees that the SINR at each subchannel is at least γ_b^c [81, 82].

3.3 Power Control and Frequency-Domain Beamforming

Now consider an uplink OFDM system where adaptive beamforming is deployed at each subchannel of all OFDM receivers. Fig. 3.2 depicts the receiver with frequency-domain

beamforming at each subchannel which ensures that the subchannels can still be considered independently. The k^{th} sampled received vector at the b^{th} base station at the time domain is given by (using the notations introduced in (3.12))

$$\underline{\mathbf{x}}_b^k = \frac{1}{\sqrt{N}} \sum_{i=0}^{M-1} \left(\sum_{c=0}^{N-1} t_{ib}^c W_N^{-ck} \right) \underline{\mathbf{a}}_{ib} + \underline{\mathbf{n}}_b^k, \quad (3.14)$$

where $\underline{\mathbf{n}}_b^k$ is the noise vector (with dimension Q , the number of antennas) whose elements are the noise samples at the input of each antenna, and the Q -dimensional vector $\underline{\mathbf{a}}_{ib}$ is the array response at the b^{th} receiver for the i^{th} transmitter.

The received signal at each antenna is passed through an OFDM receiver. The resultant c^{th} outputs of the FFT blocks create the vector

$$\hat{\underline{\mathbf{d}}}_b^c = \sum_{i=0}^{M-1} t_{ib}^c \underline{\mathbf{a}}_{ib} + \hat{\underline{\mathbf{n}}}_b^c, \quad (3.15)$$

where $\hat{\underline{\mathbf{n}}}_b^c$ is the Fourier transform of $\underline{\mathbf{n}}_b^k$. The output of the beamformer at subchannel c is then given by $e^c = \underline{\mathbf{w}}_b^{cH} \hat{\underline{\mathbf{d}}}_b^c$.

If we assume that the receiver knows the array response to the desired user, we can use Minimum Variance Distortionless Response (MVDR) beamforming [87]. In MVDR, the weight vector is calculated in order to minimize the total energy at the beamformer output, when the gain toward the desired direction is unity. The joint beamforming and power control algorithm is performed at each subchannel separately, assuming perfect orthogonalization. The energy of the beampattern at subchannel c is $\epsilon^c = E[e^c e^{c*}] = \underline{\mathbf{w}}_b^{cH} E \left[\hat{\underline{\mathbf{d}}}_b^c \hat{\underline{\mathbf{d}}}_b^{cH} \right] \underline{\mathbf{w}}_b^c$. Assuming that the noise is zero mean, white Gaussian process, and the transmitted symbols are independent and have average unity energy (see the assumptions in Section 3.2), we obtain

from (3.15) that

$$\begin{aligned}\epsilon^c &= \sum_{i=0}^{M-1} \left(P_i^c |H_{ib}^c|^2 \underbrace{\mathbf{w}_b^{cH} \mathbf{a}_{ib} \mathbf{a}_{ib}^H \mathbf{w}_b^c}_{u_{ib}^c} \right) + \frac{N_b}{2} |\mathbf{w}_b^c|^2 \\ &= \left[\sum_{i \neq b} (P_i^c |H_{ib}^c|^2 u_{ib}^c) + \frac{N_b}{2} |\mathbf{w}_b^c|^2 \right] + P_b^c |H_{bb}^c|^2 u_{bb}^c,\end{aligned}\quad (3.16)$$

in which the term inside the bracket is the interference plus noise and the second term is the power of the signal coming from the desired direction.

The SINR at the output of the beamformer at subchannel c is given by

$$\Gamma_b^c = \frac{P_b^c |H_{bb}^c|^2 |\mathbf{w}_b^{cH} \mathbf{a}_{bb}|^2}{\sum_{i \neq b} \left(P_i^c |H_{ib}^c|^2 |\mathbf{w}_b^{cH} \mathbf{a}_{ib}|^2 \right) + N_b |\mathbf{w}_b^c|^2}, \quad c = 0, \dots, N-1. \quad (3.17)$$

The MVDR solution for beamforming optimization will be [87]

$$\mathbf{w}_b^c = \frac{(\mathbf{R}_b^c)^{-1} \mathbf{a}_{bb}}{\mathbf{a}_{bb}^H (\mathbf{R}_b^c)^{-1} \mathbf{a}_{bb}}, \quad c = 0, \dots, N-1 \quad (3.18)$$

where the data correlation matrix at the c^{th} subchannel of base station b is

$$\mathbf{R}_b^c = \sum_{i \neq b} (P_i^c |H_{ib}^c|^2 \mathbf{a}_{ib} \mathbf{a}_{ib}^H) + N_b \mathbf{I}. \quad (3.19)$$

Note that in (3.19) we have used $i \neq b$.

Lemma 3.3.1. *To find the optimum receive beamforming weight vector, we can drop the restriction the $i \neq b$ in the definition of autocorrelation matrix in MVDR beamforming (Eq. (3.19)).*

Proof. Take \mathbf{R}_b^c as defined in Eq. (3.19), and define

$$\Phi = \sum_i (P_i^c |H_{ib}^c|^2 \mathbf{a}_{ib} \mathbf{a}_{ib}^H) + N_b \mathbf{I}. \quad (3.20)$$

We want to show that

$$\frac{\mathbf{R}_b^{c-1} \mathbf{a}_{bb}}{\mathbf{a}_{bb}^H \mathbf{R}_b^{c-1} \mathbf{a}_{bb}} = \frac{\Phi^{-1} \mathbf{a}_{bb}}{\mathbf{a}_{bb}^H \Phi^{-1} \mathbf{a}_{bb}}. \quad (3.21)$$

The Matrix Inversion Lemma (MIL) [87] says that for any three matrices \mathbf{P} , \mathbf{M} , \mathbf{Q} , we have

$$[\mathbf{P}^{-1} + \mathbf{M}^H \mathbf{Q}^{-1} \mathbf{M}]^{-1} = \mathbf{P} - \mathbf{P} \mathbf{M}^H [\mathbf{M} \mathbf{P} \mathbf{M}^H + \mathbf{Q}]^{-1} \mathbf{M} \mathbf{P}. \quad (3.22)$$

Take $\mathbf{A} = P_{bb}^c |H_{bb}^p|^2 \mathbf{a}_{bb} \mathbf{a}_{bb}^H$, $\mathbf{P} = \mathbf{A}^{-1}$, $\mathbf{M} = \mathbf{a}_{bb}^H$, $\mathbf{Q} = \frac{1}{P_{bb}^c |H_{bb}^p|^2}$, and considering $\Phi = \mathbf{A} + \mathbf{R}_b^c$, we need to show

$$\frac{(\mathbf{P}^{-1} + \mathbf{M}^H \mathbf{Q}^{-1} \mathbf{M}) \mathbf{M}^H}{\mathbf{M} (\mathbf{P}^{-1} + \mathbf{M}^H \mathbf{Q}^{-1} \mathbf{M})^{-1} \mathbf{M}^H} = \frac{\mathbf{P} \mathbf{M}^H}{\mathbf{M} \mathbf{P} \mathbf{M}^H}. \quad (3.23)$$

Using MIL, and noting that $\alpha \triangleq [\mathbf{M} \mathbf{P} \mathbf{M}^H + \mathbf{Q}]^{-1}$ and $\beta = \mathbf{M} \mathbf{P} \mathbf{M}^H$ are two scalar numbers, we have

$$\begin{aligned} \frac{(\mathbf{P}^{-1} + \mathbf{M}^H \mathbf{Q}^{-1} \mathbf{M}) \mathbf{M}^H}{\mathbf{M} (\mathbf{P}^{-1} + \mathbf{M}^H \mathbf{Q}^{-1} \mathbf{M})^{-1} \mathbf{M}^H} &= \frac{\left\{ \mathbf{P} - \mathbf{P} \mathbf{M}^H [\mathbf{M} \mathbf{P} \mathbf{M}^H + \mathbf{Q}]^{-1} \mathbf{M} \mathbf{P} \right\} \mathbf{M}^H}{\mathbf{M} \left\{ \mathbf{P} - \mathbf{P} \mathbf{M}^H [\mathbf{M} \mathbf{P} \mathbf{M}^H + \mathbf{Q}]^{-1} \mathbf{M} \mathbf{P} \right\} \mathbf{M}^H} \\ &= \frac{\mathbf{P} \mathbf{M}^H - \alpha \beta \mathbf{P} \mathbf{M}^H}{\alpha - \alpha^2 \beta} \\ &= \frac{\mathbf{P} \mathbf{M}^H (1 - \alpha \beta)}{\alpha (1 - \alpha \beta)} \\ &= \frac{\mathbf{P} \mathbf{M}^H}{\alpha} = \frac{\mathbf{P} \mathbf{M}^H}{\mathbf{M} \mathbf{P} \mathbf{M}^H}. \end{aligned}$$

□

By considering the fact that the MVDR constraint enforces $|\underline{\mathbf{w}}_b^{cH} \mathbf{a}_{bb}|^2 = 1$, we solve (3.17) in terms of P_b^c and adopt the iterative scheme presented in [81]. As a result, the mapping

at each iteration is the combination of (3.18) and the following equation:

$$P_b^c(n+1) = \min_{\underline{\mathbf{w}}_b^c} \frac{\gamma_b^c}{|H_{bb}^c|^2} \left[\underbrace{\sum_{i \neq b} (|H_{ib}^c|^2 P_i^c(n) u_{ib}^c)}_{I_b^c(n)} + \frac{N_b}{2} |\underline{\mathbf{w}}_b^c|^2 \right].$$

subject to $\underline{\mathbf{w}}_b^{cH} \underline{\mathbf{a}}_{bb} = 1$ (3.24)

The following algorithm achieves the jointly optimal power allocations and adaptive beamforming at each subchannel:

Algorithm I [*Joint Power control and Frequency-Domain Beamforming for OFDM*]

1. At step $n = 0$, the b^{th} base station sets $P_b^c(n) = 0$ ($c = 0 \dots N - 1$) for its mobile.
2. For each subchannel, the b^{th} base station calculates the autocorrelation matrix \mathbf{R}_b^c , and uses (3.18) to compute the weight vector $\underline{\mathbf{w}}_b^c$.
3. The base station calculates the interference and noise at subchannel c , $I_b^c(n)$, as given in (3.16), and transmits it to the transmitter through the feedback channel.
4. The mobile transmitter updates the power at each subchannel according to

$$P_b^c(n+1) = \frac{\gamma_b^c}{|H_{bb}^c|^2} I_b^c(n). \quad (3.25)$$

5. If $P_b^c(n+1) > P_{max}$, we set $P_b^c(n+1) = P_{max}$, where P_{max} is a pre-determined maximum subchannel power. This prevents the subchannels in deep fade to consume a tremendous amount of power
6. If $\sum_{p=0}^{N-1} |P_b^p(n+1) - P_b^p(n)|^2 \leq \mu$, when μ is a threshold that defines the speed of convergence, the base station stops, otherwise sets $n = n + 1$ and goes back to step 2.

Note that instead of imposing an upper bound on each subchannel's power (P_{max}), we can discard the subchannel that is in deep fade. However, this results in data rate reduction.

It has been shown in [81] that if there is a solution for the joint power control and beamforming problem for single carrier users, this algorithm will converge to the optimum solution and this solution is unique. Here, we have used the same algorithm at each subchannel independently. Therefore, the convergence and uniqueness of the solution can also be applied to the proposed algorithm.

In this work, we assumed that each mobile is using all available subchannels. However in Orthogonal Frequency Division Multiple Access (OFDMA), users are grouped and a subset of OFDM subchannels is assigned to each group of users. In this case, the proposed algorithms can be applied to the co-channel users in each group by using N_i for the number of subchannels assigned to group i . The modulation schemes can be different in different groups and this allows us to have a separate desired SINR for each group.

3.4 Power Control and Time-Domain Beamforming

In Fig. 3.2, to be able to consider the subchannels individually, we had to apply the weight vectors after the OFDM receiver. If an Q -element antenna array is deployed and we use N subchannel FFT blocks at each antenna, the complexity (number of multiplications) at the receiver would be in the order of $QN \log N + NQ^4$. Let us consider a system as depicted in Fig. 3.3. In this case, the weight vectors will pass through the FFT block and the subchannels cannot be considered independently anymore. Note that different beamformers are used at different subchannels prior to the FFT block and the data applied to the beamformer is called to be in time domain. This means they are not the data in different subchannels and so their optimum weight vectors need not to be different. In this system the beamforming

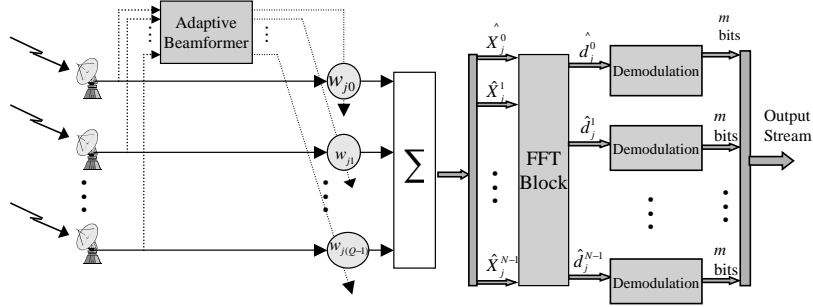


Figure 3.3: Time-Domain Beamforming in the j th OFDM receiver.

is performed in the time domain and unlike Fig. 3.2 in which N set of weight vectors are calculated at each iteration, only one set of weight vectors is calculated at each iteration. Apparently, this system has less complexity compared to that of Fig. 3.2. Using the same parameters, the complexity of Fig. 3.3 is $N \log N + Q^4$. Using typical values of 4 for Q and 128 for N , the complexity of Fig. 3.2 is in the range of 33847, while that of Fig. 3.3 is 525, which means a complexity decrease of an order of 64.

Using the system depicted in Fig. 3.3, we are no longer able to consider the joint beamforming and power control at each subchannel independently. In an OFDM system, the symbol decisions are made at the FFT output. The error and weight vector calculations have to be done in the frequency domain. If a time-domain beamformer is to be used, we need to relate the frequency domain error to that quantity in the time domain. In the sequel, we present the relationship between the time-domain beamforming error and the same quantity in the frequency domain. One way to look at this problem is to minimize the energy of \hat{D}_b , the output of the beamformer in Fig. 3.3. Using the Parseval equation, this is equivalent to minimizing the sum of the energies of the subchannels at the output of the FFT block. If Q antennas are employed at each receiver and within one symbol period made up of N samples, and if within one symbol period made up of N samples, we denote the k th sample

input to the q th antenna at base station b as x_{bq}^k , the k th input to the FFT block will be

$$\hat{X}_b^k = \sum_{q=0}^{Q-1} x_{bq}^k w_{bq}^*. \quad k = 0 \dots N-1. \quad (3.26)$$

Using (3.14) and (3.26), the received signal at the p th subchannel will be

$$\begin{aligned} \hat{d}_b^p &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{X}_b^k W_N^{kp} \\ &= \frac{1}{N} \sum_{q=0}^{Q-1} w_{bq}^* \sum_{i=0}^{M-1} a_{ib}^q \sum_{c=0}^{N-1} t_{ib}^c \sum_{k=0}^{N-1} W_N^{k(p-c)} + \frac{1}{\sqrt{N}} \sum_{q=0}^{Q-1} w_{bq}^* \sum_{k=0}^{N-1} W_N^{kp} n_{bq}^k \\ &= \sum_{i=0}^{M-1} t_{ib}^p \underbrace{\sum_{q=0}^{Q-1} a_{ib}^q w_{bq}^*}_{v_{ib}^* \triangleq \underline{\mathbf{w}}_b^H \mathbf{a}_{ib}} + \sum_{q=0}^{Q-1} w_{bq}^* \hat{n}_{bq}^p. \end{aligned} \quad (3.27)$$

Note that in (3.27) we have used the property of Comb Sequences, which states $\sum_{k=0}^{N-1} W_N^{kp} = 0$ for all $p \neq 0$. Using the fact that the input symbol energy is unity, the signal energy at subchannel p is obtained by

$$e^p = E \left[\hat{d}^p \hat{d}^{p*} \right] = \sum_{i=0}^{M-1} P_i^p |H_{ib}^p|^2 |v_{ib}|^2 + \frac{N_b}{2} |\underline{\mathbf{w}}_b|^2. \quad (3.28)$$

It is not possible to minimize the energy of all of the subchannels simultaneously, thus we use a metric which is a positive combination of all e^p 's. Since each e^p is actually an energy quantity, we simply minimize the sum of the energies which is equivalent to the energy at the output of the beamformer, \hat{D}_b , as illustrated in Fig. 3.3. So our optimization problem becomes

$$\begin{aligned} \underline{\mathbf{w}}_b &= \arg \min_{\underline{\mathbf{w}}_b} \left\{ \sum_{i=0}^{M-1} |v_{ib}|^2 \sum_{p=0}^{N-1} P_i^p |H_{ib}^p|^2 + \frac{NN_b}{2} |\underline{\mathbf{w}}_b|^2 \right\}, \\ &\text{subject to } \underline{\mathbf{w}}_b^H \mathbf{a}_{bb} = 1 \end{aligned} \quad (3.29)$$

in which the term inside the bracket is equal to $\sum_{p=0}^{N-1} e^p$.

This is very similar to a normal beamforming process. The solution for the vector $\underline{\mathbf{w}}_b$ is similar to (3.18) with \mathbf{R}_b matrix defined as

$$\mathbf{R}_b = \sum_{i \neq b} \left(\underline{\mathbf{a}}_{ib} \underline{\mathbf{a}}_{ib}^H \sum_{p=0}^{N-1} P_i^p |H_{ib}^p|^2 \right) + \frac{NN_b}{2} \mathbf{I} = \sum_{p=0}^{N-1} \mathbf{R}_b^p, \quad (3.30)$$

where \mathbf{R}_b^p is the autocorrelation matrix at subchannel p as defined in (3.19). As in the case of frequency-domain beamforming, using the Matrix Inversion Lemma, the term $i \neq b$ in (3.30) can be dropped.

By replacing u_{ib}^c with $u_{ib} = \frac{\gamma_b^c}{|H_{bb}^c|^2} \underline{\mathbf{w}}_b^H \underline{\mathbf{a}}_{ib}$, the SINR at the c th subchannel is similar to (3.17).

Again, by adopting the iterative scheme presented in [81, 82], the iteration for power calculation at each step will be

$$P_b^c(n+1) = \min_{\underline{\mathbf{w}}_b} \frac{\gamma_b^c}{|H_{bb}^c|^2} \underbrace{\left[\sum_{i \neq b} \left[u_{ib} \sum_{p=0}^{N-1} \left(|H_{ib}^p|^2 P_i^c(n) \right) \right] + \frac{NN_b}{2} |\underline{\mathbf{w}}_b|^2 \right]}_{I_b^c(n)}. \quad (3.31)$$

subject to $\underline{\mathbf{w}}_b^H \underline{\mathbf{a}}_{bb} = 1$

At each receiver the algorithm will be as follows:

Algorithm II [*Joint Power control and Time-Domain Beamforming for OFDM*]

1. At step $n = 0$, the b th base station sets $P_b^p(n) = 0$, ($p = 0 \dots N - 1$) for its mobile.
2. The b th base station calculates the sum of autocorrelation matrices of all subchannels $\sum_{p=0}^{N-1} \mathbf{R}_b^p$.
3. The base station uses the quantities computed in Step 2 to find the weight vector, $\underline{\mathbf{w}}_b$.

Note that here the weight vector is not calculated for each subchannel.

4. The base station measures the interference represented by $I_b^c(n)$ in (3.31) at each subchannel locally, and transmits these values to the b th mobile through a feedback channel.
5. The b th mobile uses (3.25) to re-calculate the power at each subchannel.
6. If $P_b^c(n+1) > P_{max}$, we set $P_b^c(n+1) = P_{max}$, where P_{max} is a pre-determined maximum subchannel power. This prevents the subchannels in deep fade to consume a tremendous amount of power
7. If

$$\sum_{c=0}^{N-1} |P_b^c(n+1) - P_b^c(n)|^2 \leq \mu,$$

when μ is a pre-determined threshold that defines the speed of convergence, the base station stops, otherwise sets $n = n + 1$ and goes back to step 2.

Let's assume that the gain matrix is denoted by $\mathbf{F}(\underline{\mathbf{w}})$ whose $(ib)^{th}$ element is $\frac{\gamma_b^c |H_{ib}^c|^2 |\underline{\mathbf{w}}_b^H \mathbf{a}_{ib}|^2}{|H_{bb}^c|^2}$ for $i \neq b$ and 0 for $i = b$. The gain matrix is an irreducible non-negative matrix and by Perron-Frobenius theorem [83] has a positive real eigenvalue that is larger than the amplitude of all other eigenvalues (spectral radius of the matrix). If the spectral radius of the gain matrix is less than unity, there is a solution for the algorithm [81, 82]. Let's call the mappings defined by the modified version of (3.25) (replacing $\underline{\mathbf{w}}_b^c$ with $\underline{\mathbf{w}}_b$) $m^w(\mathbf{p}^n)$, and the mapping defined by the combination of (3.29) and (3.25) as $m(\mathbf{p}^n)$.

Theorem 3.4.1. *For any two power vectors \mathbf{p}_1^l and \mathbf{p}_2^l at subchannel l such that $\mathbf{p}_1^l \leq \mathbf{p}_2^l$ the following holds:*

$$\begin{aligned}
(a) \quad & m(\mathbf{p}_1^l) \leq m^w(\mathbf{p}_1^l), \quad \text{for all } w \\
(b) \quad & m^w(\mathbf{p}_1^l) \leq m^w(\mathbf{p}_2^l), \quad \text{for all } w \\
(c) \quad & m(\mathbf{p}_1^l) \leq m(\mathbf{p}_1^l).
\end{aligned} \tag{3.32}$$

Moreover, using Algorithm II, the fixed point the mapping $m(\mathbf{p}^n)$ is unique.

Proof. Note that the coefficients of the power values in the mappings are positive. As a result, the proof of (a), (b), and (c), and also the convergence and optimality of Algorithm II, and the uniqueness of the fixed point of $m(\mathbf{p}^n)$ is essentially very similar to the proof of Theorem 1 in [81]. \square

Therefore, if the link gains and steering vectors are such that there exists a solution for this joint power control and beamforming problem, the above mentioned algorithm will always converge to a unique optimal solution. If there is a solution to the iterative algorithm, the application of the upper bound to each subchannel's power (P_{max}) will ease the convergence.

Like frequency domain, in time-domain beamforming, only one real value is exchanged through the feedback channel from the receiver to the transmitter for each link per update. Therefore, the required bandwidth for the feedback channel is the same for both methods.

3.5 Power Control and MMSE Time-Domain Beamforming

If the base stations do not have full knowledge of the array responses, \mathbf{a}_{lb} , we must use a training sequence which is correlated with the desired signal. The weight vector is obtained by minimizing the mean square error between the estimated signal and the training sequence. This is called Minimum Mean Squared Error (MMSE) approach [26]. MMSE can be applied to both frequency and time-domain beamforming. Here we only show the MMSE time-domain beamforming. If we call the k^{th} sample at the q^{th} antenna at base station b by X_{bq}^k ,

from (3.27) we obtain

$$\hat{d}_b^c = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{X}_b^k W_N^{kc} = \sum_{q=0}^{Q-1} w_{bq}^* \left[\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_{bq}^k W_N^{kc} \right] = \underline{\mathbf{w}}_b^H \underline{\mathbf{s}}_b^c, \quad (3.33)$$

where the sequence s_{bq}^c ($c = 0 \dots N-1$) is the Fourier transform of the sequence X_{bq}^k ($k = 0 \dots N-1$).

The objective in MMSE beamforming is to minimize ε_{cb}^2 given by

$$\varepsilon_{cb}^2 = E[|\hat{d}_b^c - t_b^c|^2] = P_b^c |H_{bb}^c|^2 + \underline{\mathbf{w}}_b^H \mathbf{R}_{ss}^p \underline{\mathbf{w}}_b - 2\text{Re}\{\underline{\mathbf{w}}_b^H \underline{\mathbf{r}}_{st}^c\}, \quad (3.34)$$

where $t_b^c = \sqrt{P_b^c} H_{bb}^c d_b^c$, and d_b^c is the training sequence at the b^{th} transmitter whose power is assumed to be unity. Moreover, $\mathbf{R}_{ss}^c = E[\underline{\mathbf{s}}_b^c \underline{\mathbf{s}}_b^{cH}]$ and $\underline{\mathbf{r}}_{st}^c = E[\underline{\mathbf{s}}_b^c t_b^{c*}]$.

Here we have N subchannels and the weight vector is the same for all of them, so the criteria in MMSE time-domain beamforming is to minimize $\sum_{c=0}^{N-1} \varepsilon_{cb}^2$, where

$$\sum_{p=0}^{N-1} \varepsilon_{pb}^2 = \sum_{p=0}^{N-1} P_b^p |H_{bb}^p|^2 + \underline{\mathbf{w}}_b^H \left(\sum_{p=0}^{N-1} \mathbf{R}_{ss}^p \right) \underline{\mathbf{w}}_b - 2\text{Re}(\underline{\mathbf{w}}_b^H \sum_{p=0}^{N-1} \underline{\mathbf{r}}_{st}^p). \quad (3.35)$$

This is a typical MMSE optimization problem and if $\sum_{c=0}^{N-1} \mathbf{R}_{ss}^c$ is nonsingular, its solution will be the well-known Wiener-Hopf equation [26] given by

$$\underline{\mathbf{w}}_b^{\text{opt}} = \left(\sum_{c=0}^{N-1} \mathbf{R}_{ss}^c \right)^{-1} \sum_{c=0}^{N-1} \underline{\mathbf{r}}_{st}^c. \quad (3.36)$$

The time-domain beamformer does not minimize the individual error at each subchannel. Therefore, the principal of orthogonality, valid for the Wiener-Hopf solution, is not satisfied here. However, in the following lemma, we will prove that this solution is equivalent to the MVDR solution up to a constant coefficient and therefore results in the same SINR [26].

Theorem 3.5.1. *If the training sequences transmitted from different mobiles are uncorrelated, the MMSE weight vector presented in (3.36) is equivalent to the MVDR weight vector, as expressed in (3.18) with autocorrelation matrix defined as in (3.30), up to a constant coefficient.*

Proof. Using (3.14) and the definition of vector $\underline{\mathbf{s}}_b^c$, we have

$$\begin{aligned}
\mathbf{r}_{st}^p &= E[\underline{\mathbf{s}}_b^p t_b^{p*}] = E\left[\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \underline{\mathbf{x}}_b^k W_N^{kp} t_b^{p*}\right] \\
&= \frac{1}{N} \sum_{i=0}^{M-1} \sum_{c=0}^{N-1} \mathbf{a}_{ib} E[t_i^c t_b^{p*}] \sum_{k=0}^{N-1} W_N^{k(p-c)} + \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} E[\underline{\mathbf{n}}_b t_b^{p*}] W_N^{kp} \\
&= \frac{1}{N} \sum_{i=0}^{M-1} \mathbf{a}_{ib} E[t_i^p t_b^{p*}] = E[|t_b^p|^2] \mathbf{a}_{bb} = P_b^p |H_{bb}^p|^2 \mathbf{a}_{bb}.
\end{aligned} \tag{3.37}$$

Therefore

$$\sum_{p=0}^{N-1} \mathbf{r}_{st}^p = \mathbf{a}_{bb} \sum_{p=0}^{N-1} P_b^p |H_{bb}^p|^2. \tag{3.38}$$

On the other hand by using the definition of \mathbf{R}_{ss}^p , we get

$$\begin{aligned}
\mathbf{R}_{ss}^p &= E[\underline{\mathbf{s}}_b^p \underline{\mathbf{s}}_b^{pH}] = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{k'=0}^{N-1} E[\underline{\mathbf{x}}_b^k \underline{\mathbf{x}}_b^{k'H}] W_N^{kp} W_N^{-k'p} \\
&= \frac{1}{N^2} \sum_{i=0}^{M-1} \sum_{i'=0}^{M-1} \mathbf{a}_{ib} \mathbf{a}_{i'b}^H \sum_{c=0}^{N-1} \sum_{c'=0}^{N-1} E[t_i^c t_{i'}^{c'*}] \sum_{k=0}^{N-1} W_N^{k(p-c)} \sum_{k'=0}^{N-1} W_N^{-k'(p-c')} \\
&\quad + \frac{1}{N} \sum_{k=0}^{N-1} \sum_{k'=0}^{N-1} E[\underline{\mathbf{n}}_b^k \underline{\mathbf{n}}_b^{k'H}] W_N^{(k-k')p} \\
&= \sum_{i=0}^{M-1} \sum_{i'=0}^{M-1} \mathbf{a}_{ib} \mathbf{a}_{i'b}^H E[t_i^p t_{i'}^{p*}] + \frac{1}{N} \sum_{k=0}^{N-1} E[\underline{\mathbf{n}}_b^k \underline{\mathbf{n}}_b^{kH}] \\
&= \sum_{i=0}^{M-1} \mathbf{a}_{ib} \mathbf{a}_{ib}^H E[|t_i^p|^2] + \frac{N_0}{2} \mathbf{I} \\
&= \sum_{i=0}^{M-1} \mathbf{a}_{ib} \mathbf{a}_{ib}^H P_i^p |H_{ib}^p|^2 + \frac{N_0}{2} \mathbf{I} = \mathbf{R}_b^p.
\end{aligned} \tag{3.39}$$

Therefore

$$\sum_{p=0}^{N-1} \mathbf{R}_{ss}^p = \sum_{p=0}^{N-1} \mathbf{R}_b^p. \tag{3.40}$$

This is the same as the autocorrelation for MVDR defined in (3.30). Using (3.37) and (3.39), the weight vectors for MVDR and MMSE time-domain beamforming are the same.

□

The interference at subchannel c equals the difference between the received power, $E[|\hat{d}_b^c|^2]$, and the power of the desired signal and is given by

$$\begin{aligned} I_b^p &= \underline{\mathbf{w}}_b^{optH} \mathbf{R}_{ss}^p \underline{\mathbf{w}}_b^{opt} - P_b^p |H_{bb}^p|^2 \left| \underline{\mathbf{w}}_b^{optH} \underline{\mathbf{a}}_{bb} \right|^2 \\ &= \underline{\mathbf{w}}_b^{optH} \left(\mathbf{R}_{ss}^p - \frac{1}{P_b^p |H_{bb}^p|^2} \mathbf{r}_{st}^p \mathbf{r}_{st}^{pH} \right) \underline{\mathbf{w}}_b^{opt}. \quad p = 0 \dots N - 1. \end{aligned} \quad (3.41)$$

Therefore, the MMSE algorithm is outlined as follows:

Algorithm III [*Joint Power control and MMSE Time-Domain Beamforming for OFDM*]

1. At step $n = 0$, the b^{th} base station sets $P_b^c(n) = 0$, ($c = 0 \dots N - 1$) for its mobile.
2. Using (3.36), the b^{th} base station finds the weight vector $\underline{\mathbf{w}}_b^{opt}$.
3. The base station uses (3.41) to find the interference at each subchannel and transmits these values to the b^{th} mobile through the feedback channel.
4. The b^{th} mobile uses (3.25) to re-calculate the power at each subchannel. If $P_b^c(n+1) > P_{max}$, we set $P_b^c(n+1) = P_{max}$.

5. If

$$\sum_{c=0}^{N-1} |P_b^c(n+1) - P_b^c(n)|^2 \leq \mu, \quad (3.42)$$

when μ is a pre-determined threshold that defines the speed of convergence, the base station stops, otherwise sets $n = n + 1$ and goes back to step 2.

3.6 Extension to COFDM

Theoretically, the bit and power allocation obtained by the loading algorithms meet the desired bit error rate as long as the time variation of the channel is very limited. Performing bit loading on time-varying channels requires a mechanism to adapt to the channel variation.

Moreover, as we mentioned in Section 3.1, the interference from other mobiles could be severe and cause performance degradation. Many practical OFDM systems use coding across subchannels (in frequency domain) to achieve better immunity to the frequency-selective fading channels. This provides a link between bits transmitted on separate subchannels and is done in such a way that the information conveyed by the subchannels in deep fade can be reconstructed by the information received through the ones which are not in deep fade. Therefore, coding applied to OFDM can be seen as a tool to average fading across subchannels. Block or convolutional codes are used either by their own or combined together (as the inner and outer code) and possibly with interleaving. Here, we will use Trellis Coded Modulation (TCM) at the OFDM transmitters. In TCM, convolutional coding is combined with modulation and results in higher coding gain. Mostly, TCM is based on the set partitioning performed by the Ungerboeck's encoder [11] in which m information bits map into a signal from the 2^{m+1} -ary constellation. k of these bits are encoded by a rate $k/(k+1)$ convolutional encoder to select one of the 2^{k+1} partitions at the $(k+1)^{th}$ level of the constellation's partition tree. The remaining $m-k$ bits are used to select one point within the designated partition. Adaptive TCM (ATCM) uses MQAM constellation and has a coding gain of at least $7dB$ over simple TCM (see [79] for details).

The TCM used in this chapter is the 8-state 8-PSK encoder depicted in Fig. 3.4 and its trellis is depicted in Fig. 3.5. This encoder has a coding gain of $3.6dB$ for high SNR environments [88].

Although the COFDM averages the fading on all of the subchannels, in an environment with low or moderate high SNR, the BER depends on the SNR of each subchannels. This dependence could be better seen for TCM, through the following inequality [7, 89]:

$$N_{d_f} Q \left(\sqrt{d_f^2 E_s / (2N_0)} \right) \leq P_e \leq \sum_{d=d_f}^{\infty} N_d Q \left(\sqrt{d^2 E_s / (2N_0)} \right), \quad (3.43)$$

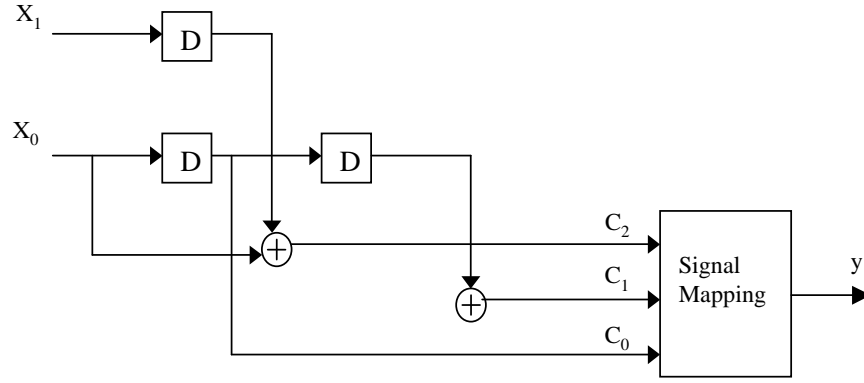


Figure 3.4: The 8-state 8-PSK TCM encoder.

where N_d is the average number of paths in the trellis having the squared Euclidian distance of d^2 from the all-zero path, d_f^2 is the normalized square free distance of the code, and $Q(\cdot)$ is the error function.

Moreover, in a multiuser environment it is not only the fading that determines both the performance of each subchannel and the overall performance of a single link. The effect of interferences plays a detrimental role on the overall bit error rate, and therefore increasing the SINR at each subchannel can improve the overall bit error rate. From a system level point of view, COFDM applied in a single user only averages the fades over different subchannels of the same user, but cannot optimize the allocation of resources in a multiuser environment such that the effect of interferences is minimized. Consequently, applying beamforming to each subchannel can improve the performance of the system. Moreover, by beamforming at each subchannel, we would be able to decrease the power consumption for achieving the same performance. The SINR fluctuation is amplified by the spatial processing and therefore the dependence of the overall bit error rate on the SINR of each subchannel becomes more severe. Power control can compensate for this fluctuation.

In this chapter, we will consider four COFDM systems using TCM. Those are a COFDM system with no power control or beamforming, a COFDM system using the frequency-

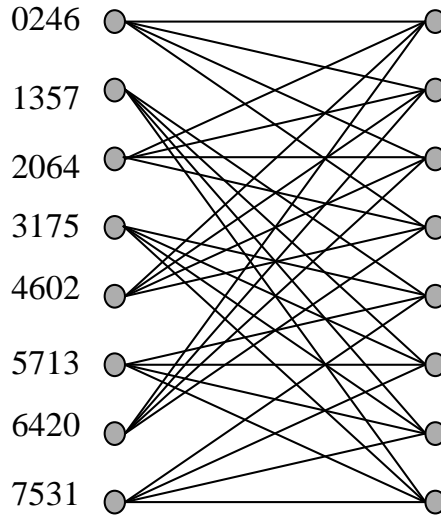


Figure 3.5: The 8-state 8-PSK TCM trellis.

domain beamforming to increase the SINR with the same amount of total network power, a system with joint power control and frequency-domain beamforming per subchannel, and finally a COFDM system where power control is performed per user and beamforming per subchannel. In the second and third system, the per subchannel SINR is measured at the symbol level, before decoding (or demodulation). In the last system, the following iterative algorithm is used to achieve power control per user and beamforming per subchannel. This algorithm tries to adapt the total user power by the equivalent SINR of the COFDM system derived from the BER of the receiver. The equivalent SINR of the COFDM system is defined as the SINR of an uncoded-OFDM system achieving the same BER, minus the coding gain of the code.

Algorithm IV [*Joint Power control per user and Frequency-Domain Beamforming for COFDM*]

1. At step $n = 0$, all mobiles start with equal powers at all subchannels. The weight vectors have one component as one and the rest as zero.

2. Each base station calculates the BER using a fixed number of frames.
3. Each base station calculates the SINR of the equivalent uncoded system using the relationship:

$$BER \simeq \frac{2}{k} Q \left(\sqrt{2\gamma_{uncoded}} \sin \left(\frac{\pi}{M} \right) \right), \quad (3.44)$$

where $M = 2^k$ is the constellation size, and $\gamma_{uncoded}$ is the SINR of the equivalent uncoded system [7]. This statement is an approximation of the bit error probability of AWGN channels for MPSK modulations and we use it here because we have assumed that the channel is known at the receiver, and by central limit theorem, the interference can be considered Gaussian.

4. The equivalent SINR of the COFDM system is calculated as

$$\gamma_{coded} = \gamma_{uncoded} - C, \quad (3.45)$$

where C is the coding gain of the coding scheme and the SINRs are evaluated in dB.

5. The following relationship is used to calculate the total power of each mobile that is distributed equally among subchannels:

$$P(n+1) = P(n) \frac{\gamma_{desired}}{\gamma_{coded}}. \quad (3.46)$$

6. Each base station uses (3.18) to find the beamforming weight at each subchannel.
7. The algorithm is repeated until convergence.

Spatial processing improves the SINR in each subchannel and the amount of improvement depends on the channel response, spatial signature and the interference in each subchannel. Since some subchannels get more benefits from the spatial processing, the power control per subchannel saves more power compared to a system where the power control is performed per user.

<i>Path</i> (l)	0	1	2	3	4	5
<i>Delay</i> (τ_l), μs	0.0	0.2	0.5	1.6	2.3	5.0
<i>Power</i> (δ_l^2)	0.189	0.379	0.239	0.095	0.061	0.037

Table 3.1: The COST207 Typical Urban 6-ray power delay profile.

3.7 Performance Results

Simulation Setup

We use a wireless network consisting of 36 base stations placed in a hexagonal pattern. We assume that all of the base stations belong to the same co-channel set and each cell contains one base station and one mobile. This model can be used by adopting any multiple access scheme to distinguish the mobiles in a cell. The formulations presented in this chapter is applicable to the cases where multiple users are assigned to each base station, by allocating different indices to the same base stations associated with different mobiles. Users are randomly distributed in a cell according to a uniform distribution. We use an OFDM system with 32 subchannels for transmission. The communication channel is assumed to follow the COST207 Typical Urban 6-ray channel model whose power delay profile is given in Table 3.1 [90].

The maximum channel delay spread is $5\mu s$ and so the channel coherence bandwidth is $200KHz$. Link gains are calculated by considering $2.5dB$ for the variance of shadow fading and the path loss exponent to be four. We assume a quasi-static channel where the channel is assumed to be fixed over multiple OFDM symbols. Note that the subchannel link gains for each user are correlated according to (3.11). A one-tap frequency-domain equalizer is assumed at the receiver such that the channel between the base station and its desired mobile can be estimated. Each OFDM symbol is assumed to be $1\mu s$ long which corresponds

to a bandwidth of $1MHz$. Therefore, the subcarrier spacing is $32KHz$ which is smaller than the coherence bandwidth of the channel and so the fading at each subchannel can be considered flat. The average power of the signal at each subchannel at each transmitter is assumed to be unity. A white Gaussian noise whose power is $60dB$ below the received signal power is added to the signal at the receiver. QPSK modulator and demodulator are used at all of the subchannels in the transmitter and receiver. The desired SINR at each subchannel for uncoded systems varies over a range of $-5dB$ to $15dB$. For the systems using beamforming a four-element antenna array is deployed at each receiver. When the adaptive power control is performed, the SINR at all of the subchannels for all of the mobiles are almost the same and therefore, it does not matter which base station we pick for the calculation of bit error rate and SINR. However, when we divide the total network power among all of the subchannels and all mobiles equally (uniform power policy), the SINRs are different. Therefore we pick several base stations based on the average SINRs of their subchannels to calculate the performances (e.g. the base stations labelled as "best base", "worst base", "base 0" and "average base" in Figs. 3.6 and 3.7). The variable "total network power" appeared in some plots represents the sum of the powers of all subchannels of all mobiles.

Numerical Results

In Fig. 3.6 a single antenna configuration is used to perform adaptive and uniform power policy, when the channel is assumed to change from one OFDM block to another, but is fixed during one OFDM symbol, and the total network power in the uniform policy experiment is the same as in the adaptive one. It is clear that the BER of all of the base stations for the adaptive case is close to the base station having average SINR. The BER vs. total network power is plotted in Fig. 3.7 when the channel follows a quasi-static model. In other words, in Fig. 3.7 the channel is assumed to be constant between two successive power updates, but

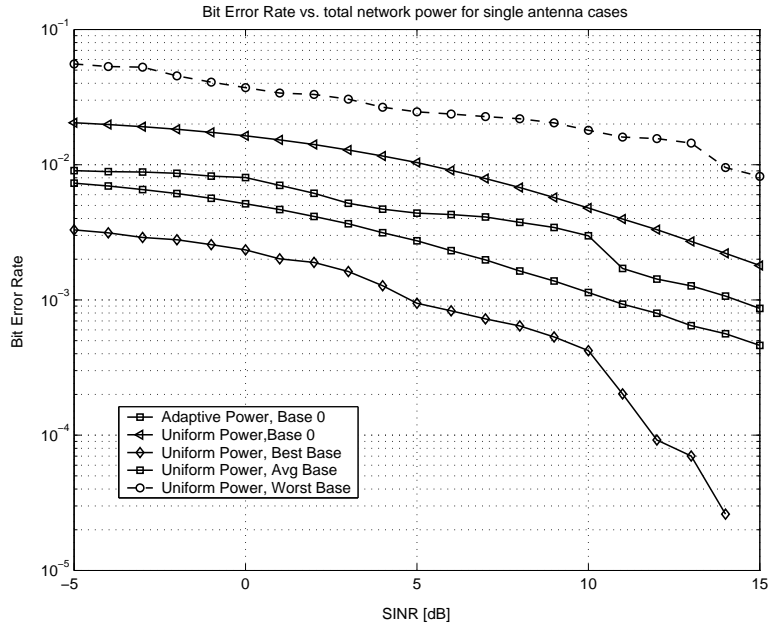


Figure 3.6: Bit error rate vs. SINR [dB] for single antenna cases for time-varying channel between power updates.

it could vary from one to another, while in Fig. 3.6 the channel coefficients are only allowed to change between two successive power updates.

We expect that in adaptive power policy all of the subchannels perform close to the target SINR, while in uniform policy, because of different link gains at different subchannels, the SINRs are expected to be different. This fact is illustrated in Fig. 3.8, where the SINRs for different subchannels of a base station using both policies are shown. This figure clearly shows that in adaptive policy, all of the subchannels perform at the same level of SINR and therefore the error probability of the receiver improves better by increasing the transmitted power [77]. By power allocation, we force the SINRs at all of the subchannels of all mobiles to be in the vicinity of a desired value, while the total transmission power is minimized. The target SINR at this experiment is $5dB$. The decrease in the SINR of the adaptive policy for the 6^{th} subchannel is due to the upper bound enforced on each subchannel's transmitted

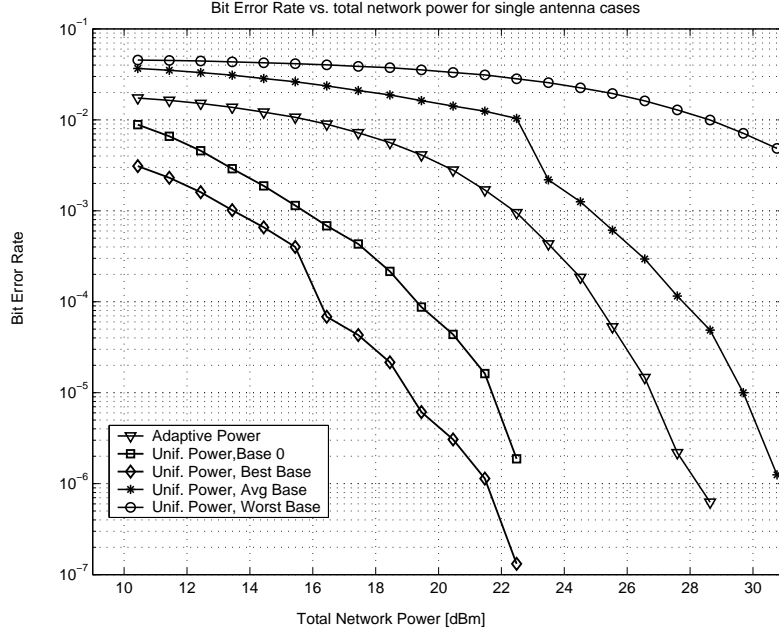


Figure 3.7: Bit error rate vs. total network power [dBm] for single antenna cases assuming quasi-static channel.

power (P_{max}). The average SINR for the mobile chosen in uniform policy is about $5.1dB$.

Figs. 3.9 compares the three adaptive power control methods proposed in this chapter. This figure shows that by using frequency-domain beamforming, we can achieve lower total network power for the same target SINR. For example, the $10dB$ threshold SINR is achieved by reducing about $4dBm$ in total network power compared to the single antenna case. It is also shown that with the channel parameters we have used, the time-domain beamforming, although not optimal, performs somewhere between the single antenna system and the system utilizing the frequency-domain beamforming at each receiver. In this case, for the same target SINR the total network power is about $3dBm$ lower compared to the single antenna case. This amount clearly depends on the channel parameters. Obviously, the frequency-domain beamforming performs better than the time-domain beamforming. In turn, we can significantly decrease complexity by using the time-domain beamforming. In all of these

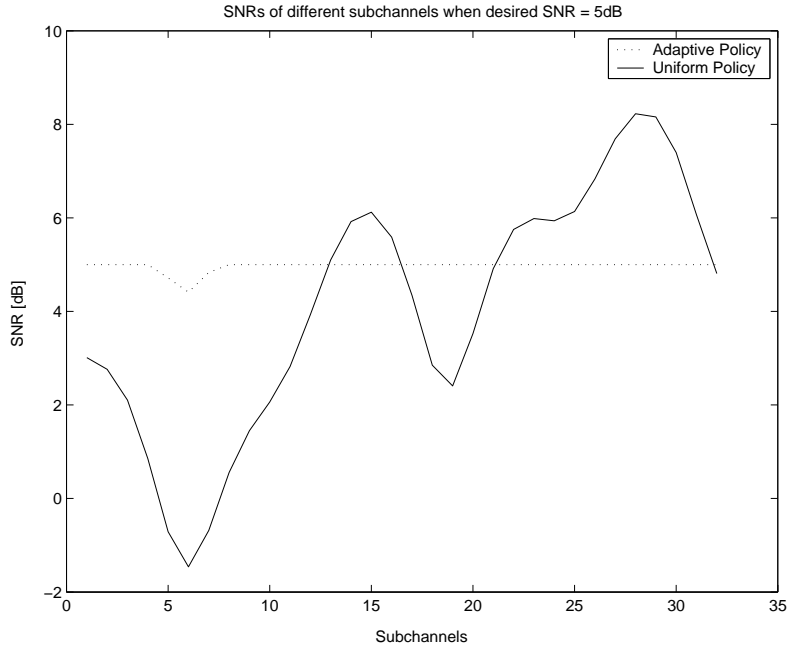


Figure 3.8: SINRs of different subchannels in the single antenna cases.

cases the uncoded-OFDM is used and by using the adaptive power control scheme, we have guaranteed the SINR at all of the subchannels to be close to the desired SINR. Since we have assumed a fixed modulation scheme at all of the subchannels, we expect to achieve similar bit error rates in all of these cases. The simulation results have confirmed our expectation.

Figs. 3.11 and 3.12 compare the uncoded-OFDM system with the rate 2/3 COFDM system using the TCM represented in Fig. 3.5. This is an Ungerboeck 8-state 8-PSK TCM encoder, whose minimum free distance d_{free}^2 is equal to 4.568 (no parallel transition) and the asymptotic coding gain (the coding gain at high SNR) is $\gamma = 2.29$ (3.6dB). This scheme is optimized for AWGN channels [13], and since we assume the perfect knowledge of the channel at the receiver we will use it in this simulation.

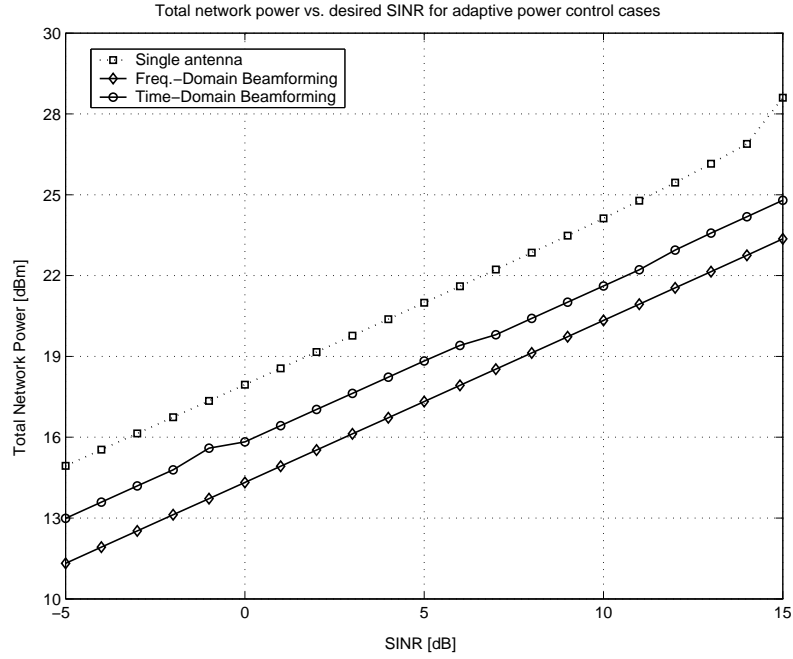


Figure 3.9: Total network power [dBm] vs. desired SINR [dB] for adaptive power control cases.

The generator matrix of the encoder is

$$\begin{aligned}
 g_{11} &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} & g_{21} &= \begin{bmatrix} 0 & 0 \end{bmatrix} \\
 g_{12} &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} & g_{22} &= \begin{bmatrix} 1 & 0 \end{bmatrix} \\
 g_{13} &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} & g_{23} &= \begin{bmatrix} 0 & 1 \end{bmatrix}.
 \end{aligned}$$

Viterbi decoding is used at each receiver. The equivalent uncoded system uses QPSK modulation. The SINR range for comparing the COFDM systems is chosen to be 0 – 30dB. Fig. 3.10 is used to evaluate the coding gain at different SINRs. This figure is obtained by calculating the performance of a single carrier system using the same TCM encoder and Viterbi decoder (the arrows show the coding gain at $BER = 10^{-5}$). In Fig. 3.11, the total network power of an uncoded-OFDM system is compared with a COFDM system with per

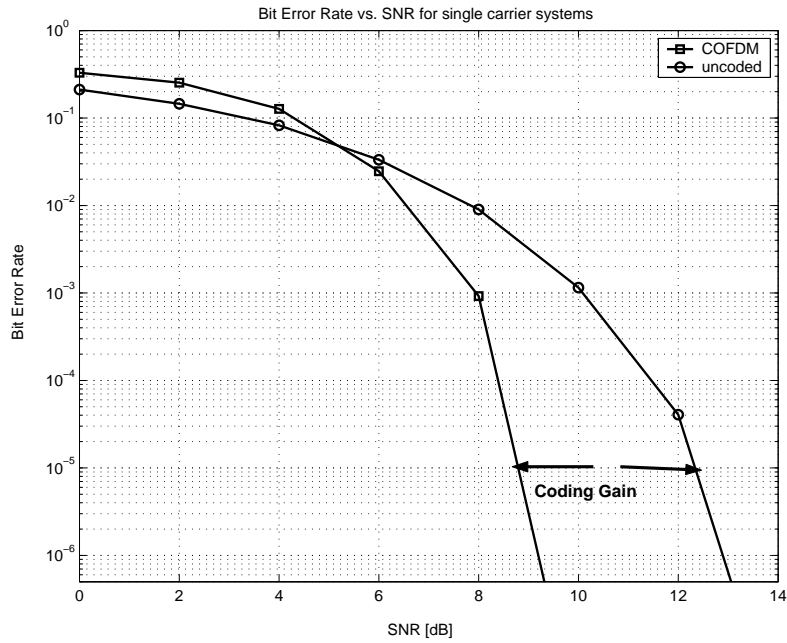


Figure 3.10: Coding gain of the TCM encoder depicted in Fig. 3.5

user power control and per subchannel beamforming. Note that the total network power of a COFDM system with per subchannel power control and beamforming is the same as that of the uncoded-OFDM system with power control and beamforming per subchannel. The total power for uncoded system is lower than COFDM system for low or moderate SNR, but is higher for high SNR. This is compensated by lower BER shown in Fig. 3.12 where we compare the BER vs. desired SINR for different OFDM systems. As can be seen from these curves, the COFDM system without any power control and beamforming has the lowest performance compared to other systems. A COFDM system where the transmitted powers are equal at all subchannels but the frequency-domain beamforming is performed at each subchannel, has a better performance compared to a COFDM system with no power control or beamforming. The curve marked by diamonds shows the BER of a COFDM system in which the per user power control jointly with per subchannel frequency-domain beamforming

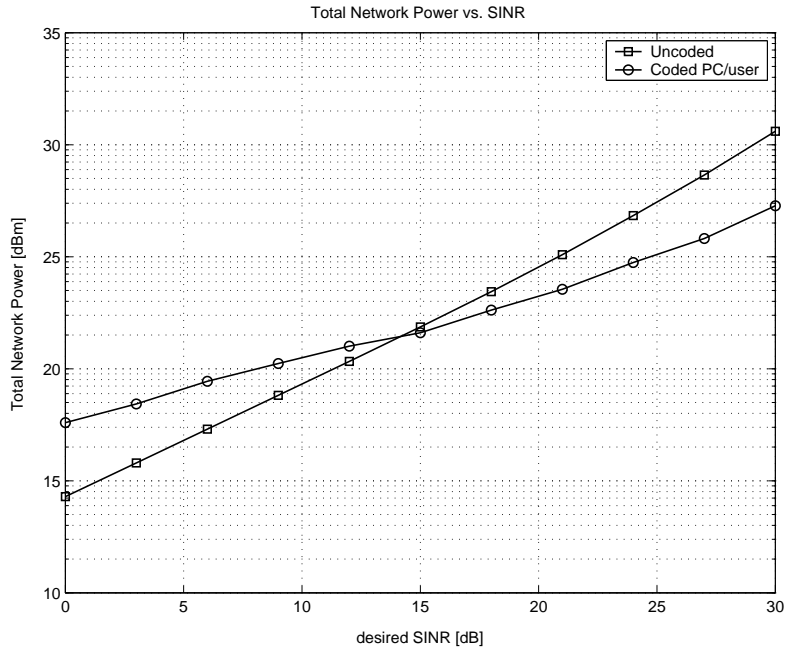


Figure 3.11: Total network power vs. desired SINR for coded and uncoded OFDM.

(the algorithm mentioned in Section 3.6) is performed. This figure shows that if the joint power control and frequency-domain beamforming is performed at each subchannel, both the uncoded and coded system have better performances compared to other configurations. For low SINR environments, the uncoded system achieves lower BER, while the performance of the coded system is better for the moderate and high SINR environments. As can be seen, these two curves intersect when the desired per subchannel SINR is 7dB . As the SINR is increased the coded system performs better. For low BERs, the OFDM coding gain is about 3.6dB , which is consistent with the asymptotic coding gain of the trellis depicted in Fig. 3.5. Since in both cases power control and frequency-domain beamforming is performed at the symbol level, we expect the uncoded system to have a better performance in low SINRs, while in moderate or high SNR the coded system performs better.

3.8 Summary of the Chapter

We considered iterative joint power control and beamforming for wireless networks using OFDM. Our study showed that we can force the SINR at all of the subchannels at all mobiles to be at least equal to a target value, while the total network power to achieve the above goals is minimized. To reduce the complexity of the OFDM receivers, we performed the array processing in the time domain and provided an iterative algorithm to distribute the power among subchannels. We also proposed MMSE time-domain beamforming jointly with power control, for the cases when the angle of arrivals are unknown.

We observed that an uncoded-OFDM system with the proposed algorithms performs better than the simple COFDM system, a coded system with per subchannel beamforming with equal powers across subchannels and a COFDM system with per user power control and per subchannel beamforming. If the proposed algorithm is applied to COFDM, the BER is improved for moderate and high SINRs.

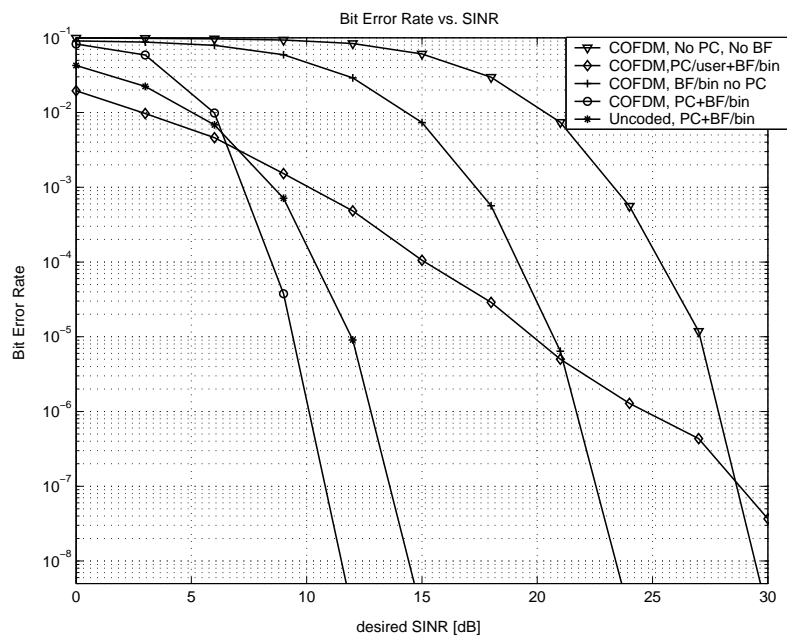


Figure 3.12: Bit error rate vs. desired SINR for coded and uncoded OFDM.

Chapter 4

MIMO-OFDM Systems with Multi-User Interference

4.1 Motivation and Previous Works

Recent information theoretic results suggest that there is significant capacity improvement for wireless communication systems using multiple antenna transmission [23]. The use of multiple antennas at both ends of a wireless link has been shown to have the potential to achieve tremendous improvements in bit rates [32, 91]. This bit rate increase is obtained without the necessity of using additional power or bandwidth. Extensive study of on the capacity of flat fading (deterministic and stochastic) multiple-input multiple-output (MIMO) channels can be found in [23, 32, 92–94]. Multiple antennas are used to perform beamforming [26], gain space diversity by using the well-known space time codes [29], or achieve spatial multiplexing [32]. In BLAST (Bell-Labs Layered Space- Time architecture) multiple data streams are transmitted simultaneously and in the same frequency band, and can be separated using receiver signal processing because of distinct spatial signatures at the transmit antennas. Spatial multiplexing can significantly benefit from transmit processing when the channel is known at the transmitter side in addition to the receiver side [91]. Many of these researches have assumed a single user system and transmit multiple data streams. Optimal

or suboptimal strategies that maximizes the information theoretic sum capacity of vector multiple access channels have been studied in [95].

In a multiple access system, if the data rates are constant for different users, normally there would be a trade-off between the throughput of the overall network and the spectral efficiency of one particular user. The reason is the interference each user might incur on other co-channel receivers. In the systems where the data rate is homogeneous, power control can be used, while in non-homogeneous systems data rate adaptation is necessary for increasing the system spectral efficiency [96].

Transmit and receive beamforming have also been studied for multiuser systems when each user is transmitting a single data stream [97, 98] by maximizing the Signal to Interference and Noise Ratio (SINR). The authors in [99] have proposed an algorithm that performs optimal transmit and receive processing by performing system-wide Mean Square Error (MSE).

So far, most of the research in this context has considered a narrowband flat fading channel. However, Orthogonal Frequency Division Multiplexing (OFDM) has proven to be an efficient transmit and receive scheme, which is prone to the detrimental effect of multipath delays. Different subchannels in an OFDM system experience different attenuations. So, how to adaptively allocate the powers and transmitting rates to fully utilize the spectrum is a hot topic in recent research works. As a matter of fact, this specification of OFDM systems offers an efficient tool for performing data rate adaptation. The exploitation of multiple transmit and receive antennas (Multiple-Input, Multiple-Output or MIMO) both in the form of beamforming [26] and space-frequency coding[29] has recently found much attention. The authors in [91, 100], provide expressions for the ergodic capacity and the outage capacity of OFDM-based MIMO systems, with unknown channel at the transmitter, and studied the effect of channel parameters on the capacity. In [91] the authors have used

the spatial multiplexing for DMT when the powers are distributed equally among all eigenmodes. Many of these schemes have considered a single transmitter and receiver. In addition to the effect of multipath fading, in a multi-user environment, the interference from other transmitters, plays a detrimental role in degrading the system performance.

In this chapter, we will consider MIMO-OFDM systems in a multi-user environment, where multiple data streams are transmitted from each transmitter. Unlike some previous works in this context, we will consider different power constraints at each transmitter. Two scenarios are considered in this chapter. First, we restrict ourselves to the case where the transmission power is fixed at each subcarrier of an OFDM system. This case can happen when, by performing power control, the subcarrier power is restricted to a maximum value. In the second assumption, we assume that the overall transmit power of each user is fixed. In each case, we will quantify the overall OFDM bit rate with respect to transmit and receive weight vectors, and use iterative water-filling to find these vectors to maximize the OFDM overall bit rate. We present an extension of BLAST, which is optimal in the sense that it maximizes the spectral efficiency of different links, assuming that the covariance of interference and the channel conditions are fed back to the transmitter from the receiver. The transmitter sends multiple streams through the eigenmodes of the channel and interference. We will also consider the rate maximization in MIMO/OFDM systems, when one single stream is transmitted. We will establish a game theoretic analogy of the problem. The rest of this chapter is organized as follows:

Section 4.2 describes the system model. Section 4.3 quantifies the overall OFDM data rate with the transmit and receive weight vectors in both cases, and provide iterative algorithms to maximize this quantity. We will also present the game theoretic approach of MIMO/OFDM problem in this section. In Section 4.4, a single mode MIMO/OFDM is considered. Section 4.5 presents some numerical results, and finally the chapter is concluded in Section 4.6

concludes the chapter.

4.2 System Model

We consider a multi-user OFDM system consisting of M cells, one base station per cell, and N_m mobiles in cell number m . Total number of mobiles is considered to be N , and therefore $\sum_m^M N_m = N$. That is, there are N co-channel link pairs for uplink and downlink transmission. A downlink communications is considered here, where N_t transmit antennas is employed at each base station and each mobile user is equipped with N_r receive antennas. However the same analysis can be applied to uplink transmission. T independent data streams are transmitted from each base station, where a different transmit weight vector is calculated for transmitting each stream to the corresponding mobile. Assuming that the number of OFDM subcarriers, K , is large enough, each subchannel is assumed to follow a flat fading quasi static model, where the subchannel is constant over a few OFDM blocks. The frequency domain subchannel link gains from base station m to mobile n at the k^{th} subcarrier is denoted by an N_r by N_t matrix \mathbf{H}_{mn}^k , whose (i, j) -th element $[\mathbf{H}_{mn}^k]_{ij}$, represents the channel link gain from the transmit antenna j to receive antenna i . Note that we will consider a general model in which all mobile receivers in a cell are supposed to receive all of the streams transmitted from the base station. In this case, a specific mobile considers the data transmitted from its base station for other mobiles as interference. However, in a more practical model, specific data streams might be intended for specific users. In those cases, the powers of the eigen-modes corresponding to those streams for those mobiles are considered to be zero.

Consider an OFDM system where the received signal at the k^{th} carrier of the i^{th} mobile

is given by

$$\mathbf{x}_i^k = \mathbf{W}_i^{kH} \left(\sum_{m=1}^N \mathbf{H}_{mi}^k \mathbf{V}_m^k \mathbf{s}_m^k + \mathbf{n}_i^k \right), \quad (4.1)$$

where \mathbf{s}_m is the data vector of size T that is intended to be transmitted to user m , and its covariance Φ_m^k is defined by

$$\Phi_m^k = E\{\mathbf{s}_m^k (\mathbf{s}_m^k)^H\}. \quad (4.2)$$

\mathbf{n}_i^k is the thermal noise vector of size N_r at the i^{th} receiver, whose spatial covariance matrix is diagonal with equal power $(\sigma_i^k)^2$ per antenna. Note that in (4.1), we have considered the sum of N transmitted signals rather than M . The reason is that each base station (say base station m) transmits independent data streams to N_m receivers, and the data aimed to each mobile is considered interference at other receivers. The $N_r \times T$ matrix \mathbf{W}_i^k and $N_t \times T$ matrix \mathbf{V}_m^k represent the receive and transmit weight vectors, respectively.

If s_{mt}^k is the t^{th} stream transmitted from transmitter t at subcarrier k , the transmitted power of stream t , from antenna i at subcarrier k is $E[|s_{mt}^k|^2] |V_{mt,i}^k|^2$, and therefore the total transmit power at the k^{th} subcarrier is

$$\begin{aligned} P_m^k &= \sum_{t=1}^T \sum_{i=1}^{N_t} [E[|s_{mt}^k|^2] |V_{mt,i}^k|^2] \\ &= \sum_{i=1}^{N_t} \left[\sum_{t=1}^T (V_{mt,i}^k)^* E[|s_{mt}^k|^2] V_{mt,i}^k \right] \\ &= \text{tr} \left(\mathbf{V}_m^{kH} \Phi_m^k \mathbf{V}_m^k \right) \end{aligned} \quad (4.3)$$

Assuming that the transmitted streams from different users are independent of each other and also independent of noise samples, the total covariance matrix at receiver i is given by

$$\begin{aligned} \mathbf{X}_i^k &= E\{\mathbf{x}_i^k (\mathbf{x}_i^k)^H\} = \mathbf{W}_i^{kH} \left[\sum_{m=1}^N \mathbf{H}_{mi}^k \mathbf{V}_m^k \Phi_m^k \mathbf{V}_m^{kH} \mathbf{H}_{mi}^{kH} + \sigma_i^{k2} \mathbf{I}_{N_r} \right] \mathbf{W}_i^k \\ &= \mathbf{W}_i^{kH} \mathbf{H}_{ii}^k \mathbf{V}_i^k \Phi_i^k \mathbf{V}_i^{kH} \mathbf{H}_{ii}^{kH} \mathbf{W}_i^k + \mathbf{W}_i^{kH} \mathbf{Q}_i^k \mathbf{W}_i^k, \end{aligned} \quad (4.4)$$

where

$$\mathbf{Q}_i^k = \sum_{m \neq i} \mathbf{H}_{mi}^k \mathbf{V}_m^k \Phi_m^k \mathbf{V}_m^{kH} \mathbf{H}_{mi}^{kH} + \sigma_i^{k2} \mathbf{I}_{N_r}, \quad (4.5)$$

is the covariance of the interference and noise at the k^{th} subcarrier of the i^{th} receiver.

4.3 Achievable Rate with Known Interference Covariance Matrix

We assume that the interference has a Gaussian distribution. This happens, for example, when the transmit signal has a Gaussian distribution or the number of users in the system is large according to Central Limit Theorem (CLT). Furthermore, we follow two scenarios. In the first one, we assume a fix transmit power policy per subcarrier where every user adapts its data rate with the total transmit power across its antennas held constant. This happens in situations like when we use power control and therefore each user is assigned a fixed power value per subcarrier. In the second scheme, the transmit power is fixed per user. In the following, we will find the expressions for the achievable rate for each subcarrier at the receiver i . For notation simplicity, we will drop the subcarrier index k , and reuse it wherever needed. Assuming Gaussian interference, The mutual information at subcarrier k can be expressed as [101]:

$$\begin{aligned} \mathcal{I}_i &= \log_2 \det [\mathbf{W}_i^H (\mathbf{H}_{ii} \mathbf{V}_i \Phi_i \mathbf{V}_i^H \mathbf{H}_{ii}^H + \mathbf{Q}_i) \mathbf{W}_i] - \log_2 \det [\mathbf{W}_i^H \mathbf{Q}_i \mathbf{W}_i] \\ &= \log_2 \det \left[\mathbf{W}_i^H \mathbf{H}_{ii} \mathbf{V}_i \Phi_i \mathbf{V}_i^H \mathbf{H}_{ii}^H \mathbf{W}_i (\mathbf{W}_i^H \mathbf{Q}_i \mathbf{W}_i)^{-1} + I_T \right] \end{aligned} \quad (4.6)$$

So the optimization problem for each user i knowing the data covariance matrix, Φ_i is

defined as finding how to find the transmit and receiver weight vector so as to

$$\begin{aligned} & \max_{\mathbf{W}_i, \mathbf{V}_i} \mathcal{I}_i, \\ \text{subject to } & \text{tr} \left(\mathbf{V}_m^k \mathbf{H}_m^k \Phi_m^k \mathbf{V}_m^k \right) \leq P_m^k \end{aligned} \quad (4.7)$$

To this end we will use a well-known relation in linear algebra which states that if matrix A is m by n , and B is n by m , then [102]

$$\det(I_m + AB) = \det(I_n + BA), \quad (4.8)$$

where I_m and I_n are the identity matrices. Using (4.8), we can change the mutual information to

$$\mathcal{I}_i = \log_2 \det \left[\mathbf{H}_{ii}^H \mathbf{W}_i (\mathbf{W}_i^H \mathbf{Q}_i \mathbf{W}_i)^{-1} \mathbf{W}_i^H \mathbf{H}_{ii} \mathbf{V}_i \Phi_i \mathbf{V}_i^H + I_{N_t} \right]. \quad (4.9)$$

The MMSE receiver weight vectors are given by

$$\mathbf{W}_i = \mu \mathbf{Q}_i^{-1} \mathbf{H}_{ii} \mathbf{V}_i, \quad (4.10)$$

where μ is a real constant. It is well known [103] that this choice can maximize the SINR (and therefore maximizes the mutual information). Using this expression for the receive weight vectors, we have

$$(\mathbf{W}_i^H \mathbf{Q}_i \mathbf{W}_i)^{-1} = \mu^{-2} (\mathbf{V}_i^H \mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii} \mathbf{V}_i)^{-1}. \quad (4.11)$$

The mutual information becomes

$$\begin{aligned} \mathcal{I}_i &= \log_2 \det \left[\mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii} \mathbf{V}_i (\mathbf{V}_i^H \mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii} \mathbf{V}_i)^{-1} \mathbf{V}_i^H \mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii} \mathbf{V}_i \Phi_i \mathbf{V}_i^H + \mathbf{I}_{N_t} \right] \\ &= \log_2 \det \left[\mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii} \mathbf{V}_i \Phi_i \mathbf{V}_i^H + \mathbf{I}_{N_t} \right], \end{aligned} \quad (4.12)$$

which is the capacity obtained by optimum receiver processing. Therefore the chosen MMSE receiver weight vector is optimum.

The eigenvalue decomposition the matrix $\mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii}$ is

$$\mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H, \quad (4.13)$$

where the columns of unitary matrix \mathbf{U} and the diagonal elements of the diagonal matrix $\mathbf{\Lambda}$ are the eigenvectors and eigenvalues of the matrix $\mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii}$, respectively. Since $\mathbf{H}_{ii}^H \mathbf{Q}_i^{-1} \mathbf{H}_{ii}$ is positive definite, all of these eigenvalues are non-negative. Note that the number of eigenmodes of the system is determined by the rank of this matrix that we call r_i . From (4.5), the matrix \mathbf{Q}_i is the sum of many independent matrices, and assuming that the number of transmitters is high enough, we can assume it is full rank (rank of N_r). By the inequalities [102]

$$\text{rank}(\mathbf{A}\mathbf{B}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})), \quad (4.14)$$

we have

$$r_i \leq \min(N_t, N_r). \quad (4.15)$$

Using this decomposition, the mutual information becomes

$$\begin{aligned} \mathcal{I}_i &= \log_2 \det [\mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \mathbf{V}_i \Phi_i \mathbf{V}_i^H + I_{N_t}] \\ &= \log_2 \det [\mathbf{\Lambda}^{1/2} \mathbf{U}^H \mathbf{V}_i \Phi_i \mathbf{V}_i^H \mathbf{U} \mathbf{\Lambda}^{1/2} + I_{N_t}]. \end{aligned} \quad (4.16)$$

The Hadamard inequality for an m by m matrix \mathbf{A} says [102]

$$\det(\mathbf{A}) \leq \prod_{i=1}^m a_{ii}, \quad (4.17)$$

where a_{ii} is the i^{th} diagonal elements of matrix \mathbf{A} . Therefore, $\det(\mathbf{A})$ is maximized if matrix \mathbf{A} is either diagonal or upper or lower triangular. The mutual information, \mathcal{I}_i is maximized if

$$\mathbf{U}^H \mathbf{V}_i \Phi_i \mathbf{V}_i^H \mathbf{U} = \tilde{\mathbf{P}}_i, \quad (4.18)$$

is diagonal with non-negative elements $p_{i1}, p_{i2}, \dots, p_{iN_t}$. In this case

$$\mathcal{I}_i = \log_2 \det \left[\tilde{\mathbf{P}}_i \boldsymbol{\Lambda} + I_{N_t} \right] = \sum_j \log_2 (1 + p_{ij} \lambda_j), \quad (4.19)$$

where λ_j is the j^{th} eigen-values of $\boldsymbol{\Lambda}$. Since \mathbf{U} is unitary, and $\mathbf{tr}(AB) = \mathbf{tr}(BA)$, we have

$$\mathbf{tr}(\mathbf{V}_i \Phi_i \mathbf{V}_i^H) = \mathbf{tr}(\tilde{\mathbf{P}}_i) = \sum_{l=1}^{N_t} p_{il} \leq P_i^k. \quad (4.20)$$

4.3.1 Constant Power per Subcarrier

If we assume that the total power of eigen-modes at each subcarrier is fixed, our problem becomes the maximization of \mathcal{I}_i , with the constraint $\mathbf{tr}(\tilde{\mathbf{P}}_i) \leq P_i^k$. The answer is the well-known water-filling solution [101], which states that for non-zero λ_j 's

$$p_{ij} = \left(\nu - \frac{1}{\lambda_j} \right)^+, \quad (4.21)$$

where ν is chosen in such a way that $\sum_j p_{ij} = P_i^k$, and $(x)^+ = \max(x, 0)$. If $\lambda_j = 0$, we take $p_{ij} = 0$.

To find the optimum transmit weight vectors, we need to solve $\mathbf{U}^H \mathbf{V}_i \Phi_i \mathbf{V}_i^H \mathbf{U} = \tilde{\mathbf{P}}_i$ or $(\mathbf{V}_i \Phi_i^{1/2}) (\mathbf{V}_i \Phi_i^{1/2})^H = \mathbf{U} \tilde{\mathbf{P}}_i \mathbf{U}^H$ for \mathbf{V}_i . To this end, we consider 3 cases:

1. $N_t < T$: In this case

$$\mathbf{U} \begin{bmatrix} \tilde{\mathbf{P}}_i^{1/2} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{P}}_i^{1/2} \\ \mathbf{0} \end{bmatrix} \mathbf{U}^H = \mathbf{U} \tilde{\mathbf{P}}_i \mathbf{U}^H. \quad (4.22)$$

So, $\mathbf{V}_i \Phi_i^{1/2} = \mathbf{U} \underbrace{\begin{bmatrix} \tilde{\mathbf{P}}_i^{1/2} & \mathbf{0} \end{bmatrix}}_{\mathbf{B}}$, or $\mathbf{V}_i = \mathbf{U} \mathbf{B} \Phi_i^{-1/2}$.

2. $N_t > T$: Note that $\text{rank}(\mathbf{V}_i \Phi_i \mathbf{V}_i^H) \leq \min(T, N_t) = T$, therefore $\text{rank}(\tilde{\mathbf{P}}_i) \leq T$. Since $\tilde{\mathbf{P}}_i$ is diagonal, without loss of generality, we can assume that the first r_i eigenvalues

are non-zero, and therefore we can write

$$\tilde{\mathbf{P}}_i = \begin{pmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (4.23)$$

when \mathbf{P}_1 is an $T \times T$ diagonal matrix. In this case,

$$\mathbf{V}_i = \mathbf{U} \underbrace{\begin{bmatrix} \mathbf{P}_1^{1/2} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{B}} \Phi_i^{-1/2}. \quad (4.24)$$

3. $N_t = T$: In this case $\mathbf{B} = \tilde{\mathbf{P}}_i^{1/2}$, and again $\mathbf{V}_i = \mathbf{U}\mathbf{B}\Phi_i^{-1/2}$.

Note that in all cases, we have $\mathbf{B}\mathbf{B}^H = \tilde{\mathbf{P}}_i$. We assume that all of the data streams are independent, and therefore the matrix Φ_i is diagonal. Therefore

$$\begin{aligned} \mathbf{x}_i^k &= \mathbf{W}_i^{kH} \left(\mathbf{H}_{ii}^k \mathbf{V}_i^k s_i^k + \underbrace{\sum_{m \neq i} \mathbf{H}_{mi}^k \mathbf{V}_m^k s_m^k}_{q} + \mathbf{n}_i^k \right) \\ &= \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^{kH} \mathbf{Q}_i^{-1} \mathbf{H}_{ii}^k \mathbf{U} \mathbf{B} \Phi_i^{-1/2} s_i^k + \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^{kH} \mathbf{Q}_i^{-1} q \\ &= \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \mathbf{U} \mathbf{B} \Phi_i^{-1/2} s_i^k + \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^{kH} \mathbf{Q}_i^{-1} q \\ &= \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{\Lambda} \mathbf{B} \Phi_i^{-1/2} s_i^k + \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^k \mathbf{Q}_i^{-1} q \\ &= \mu^* \Sigma \Phi_i^{-1} s_i^k + \mu^* \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^{kH} \mathbf{Q}_i^{-1} q, \end{aligned} \quad (4.25)$$

where $\Sigma = \mathbf{B}^H \mathbf{\Lambda} \mathbf{B}$ is a diagonal matrix. This result shows that at the receiver, and at each subchannel the multiple streams are orthogonal to each other. The receiver covariance

matrix is

$$\begin{aligned}
E \left[\mathbf{x}_i^k \mathbf{x}_i^{kH} \right] &= \mu^2 \Sigma \Phi_i^{-1} E[s_i^k s_i^{kH}] \Phi_i^{-1} \Sigma + \mu^2 \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^{kH} \mathbf{Q}^{-1} E[qq^H] \mathbf{Q}^{-1} \mathbf{H}_{ii}^k \mathbf{U} \mathbf{B} \Phi_i^{-1/2} \\
&= \mu^2 \Phi_i^{-1} \mathbf{B}^H \underbrace{\mathbf{\Lambda} \mathbf{B} \mathbf{B}^H}_{\tilde{\mathbf{P}}_i} \mathbf{\Lambda} \mathbf{B} + \mu^2 \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{H}_{ii}^{kH} \mathbf{Q}^{-1} \mathbf{H}_{ii}^k \mathbf{U} \mathbf{B} \Phi_i^{-1/2} \\
&= \mu^2 \mathbf{B}^H \mathbf{\Lambda}^2 \tilde{\mathbf{P}}_i \mathbf{B} \Phi_i^{-1} + \mu^2 \Phi_i^{-1/2} \mathbf{B}^H \mathbf{U}^H \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \mathbf{U} \mathbf{B} \Phi_i^{-1/2} \\
&= \mu^2 \mathbf{B}^H \mathbf{\Lambda}^2 \tilde{\mathbf{P}}_i \mathbf{B} \Phi_i^{-1} + \mu^2 \Phi_i^{-1} \Sigma. \tag{4.26}
\end{aligned}$$

1. $N_t < T$: In this case $\mathbf{B}^H \mathbf{D} \mathbf{B} = \begin{bmatrix} \mathbf{D} \tilde{\mathbf{P}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, for each diagonal matrix \mathbf{D} , and therefore

$$E \left[\mathbf{x}_i^k \mathbf{x}_i^{kH} \right] = \mu^2 \tilde{\mathbf{P}}_i^2 \mathbf{\Lambda}^2 \Phi_{i1}^{-1} + \mu^2 \tilde{\mathbf{P}}_i \mathbf{\Lambda} \Phi_{i1}^{-1}, \tag{4.27}$$

where Φ_{i1} is the first diagonal $N_t \times N_t$ subblock of Φ_i . So the SNR for the first N_t streams are $p_{ij} \lambda_j$.

2. $N_t > T$: In this case $\mathbf{B}^H \mathbf{D} \mathbf{B} = \mathbf{D}_1$, where \mathbf{D}_1 is the first diagonal $T \times T$ subblock of the diagonal matrix \mathbf{D} . Therefore

$$E \left[\mathbf{x}_i^k \mathbf{x}_i^{kH} \right] = \mu^2 \tilde{\mathbf{P}}_{i1}^2 \mathbf{\Lambda}_1^2 \Phi_i^{-1} + \mu^2 \tilde{\mathbf{P}}_{i1} \mathbf{\Lambda}_1 \Phi_i^{-1}. \tag{4.28}$$

So the SNR for the j^{th} stream is $p_{ij} \lambda_j$.

3. $N_t = T$: In this case

$$E \left[\mathbf{x}_i^k \mathbf{x}_i^{kH} \right] = \mu^2 \tilde{\mathbf{P}}_i^2 \mathbf{\Lambda}^2 \Phi_i^{-1} + \mu^2 \tilde{\mathbf{P}}_i \mathbf{\Lambda} \Phi_i^{-1}. \tag{4.29}$$

So the SNR for the j^{th} stream is $p_{ij} \lambda_j$.

4.3.2 Constant power per user

In the previous subsection, we assumed that the total transmit power of each user at each subcarrier is a known fixed value. However, a more practical arrangement is to consider a

known constant power for each user, rather than each subcarrier. To find the optimal weight vectors in this case, we notice that the overall mutual information between the i^{th} receiver and its corresponding transmitter can be obtained from (4.16) by

$$\mathcal{I}_i = \sum_{k=1}^K \mathcal{I}_i^k = \sum_{k=1}^K \log_2 \det \left[\Lambda_i^{k/2} \underbrace{\mathbf{U}_i^{kH} \mathbf{V}_i^k \Phi_i^k \mathbf{V}_i^{kH} \mathbf{U}_i^k}_{\tilde{\mathbf{P}}_i^k} \Lambda_i^{k/2} + I_{N_i} \right]. \quad (4.30)$$

If $\tilde{\mathbf{P}}_i^k$ for each k is diagonal, by Hadamard inequality, each term of the above sum is maximized regardless of any power constraint. Since each term is positive, the sum will also be maximized. Therefore, we want to maximize

$$\begin{aligned} \mathcal{I}_i &= \sum_{k=1}^K \log_2 \prod_{t=1}^T (1 + p_{it}^k \lambda_{it}^k) \\ &= \sum_{r=1}^{KT} \log_2 (1 + p_{ir} \lambda_{ir}). \end{aligned} \quad (4.31)$$

constrained to $\sum_{r=1}^{KT} p_{ir} = P_i$, where P_i is the power allocated to user i . The index r is used to represent the powers over all streams and all subcarriers. For a given r , the index of the stream is given by $j = (r \bmod T) + 1$, and the index of subcarrier by $k = \lfloor (r/T) \rfloor$. Again, the answer is the water-filling solution presented in (4.21). The SNRs are again $p_{ir} \lambda_{ir}$ at corresponding subcarriers and streams.

4.3.3 Iterative Water-filling

Note that each transmitter optimizes its transmit spectrum independently of the other transmitters, but knowing the interference covariance matrix. Individual link optimal solutions are not stable points for the network, since they would be affected by any change in the spatial signatures on other links, and so they need to be re-optimized. We use these link solutions to define a system-wide algorithm, wherein at each iteration the link solution is readjusted to new interference environment. Since each transmitter is repeating the same

process, the interference at the receiver is going to be changed and the above steps should be repeated until the transmit and interference covariance matrices converge. Therefore, the iterative water-filling [95] is used to find the powers at each eigen-mode at each subcarrier of each transmitter. Note that the authors in [95] have used the iterative water-filling for a single-cell single-carrier system, while we are proposing this scheme for multi-cell multicarrier multiple antenna systems.

For the first scheme in which the constant power per subcarrier is considered, the iterative water-filling at subcarrier $k(k = 1 \dots K)$ is proposed as follows:

Algorithm I [Multi-cell iterative water-filling, constant power per subcarrier]

1. For receiver n ($n = 1 \dots N$) set the elements of the first row of V_n^k to 1, and the rest to 0. Set $n = 1$
2. Find the interference from Eq. (4.5).
3. Find the eigenvalue decomposition of $\mathbf{H}_{nn}^k H \mathbf{Q}_n^{k-1} \mathbf{H}_{nn}^k = \mathbf{U}_n^k \mathbf{\Lambda}_n^k \mathbf{U}_n^{kH}$ (as in (4.13)).
4. Find the power of each eigen-mode with nonzero eigenvalue from (4.21) where λ_j 's are the diagonal elements of the diagonal matrix $\mathbf{\Lambda}_n^k$. For $\lambda_j = 0$, take $p_{ij} = 0$. Create a diagonal matrix $\tilde{\mathbf{P}}_n^k$, whose diagonal elements are p_{ij} .
5. Find the transmit weight vectors from $\mathbf{V}_n^k = \mathbf{U}_n^k \mathbf{B} \Phi_n^{k-1/2}$, when \mathbf{B} is $\begin{bmatrix} \tilde{\mathbf{P}}_n^{1/2} & \mathbf{0} \end{bmatrix}$, or $\begin{bmatrix} \mathbf{P}_1^{1/2} \\ \mathbf{0} \end{bmatrix}$ (look at (4.23)), or $\tilde{\mathbf{P}}_n^{1/2}$, depending on the relation between N_t and T .
6. Set $n = n + 1$ and continue from step 2 until convergence.

For the second scheme where the constant power per user is considered, the iterative water-filling is proposed as follows:

Algorithm II [Multi-cell iterative water-filling, constant power per user]

1. For the receiver n ($n = 1 \dots N$) and all subcarriers k ($k = 1 \dots K$), set the elements of the first row of V_n^k to 1, and the rest to 0. Set $n = 1$.
2. For each carrier k , find the interference from Eq. (4.5).
3. For each carrier k , find the eigenvalue decomposition of $\mathbf{H}_{nn}^k H \mathbf{Q}_n^{k-1} \mathbf{H}_{nn}^k = \mathbf{U}_n^k \mathbf{\Lambda}_n^k \mathbf{U}_n^{kH}$ (as in (4.13)).
4. For $r = 1 \dots KT$, find the power of each eigen-mode with nonzero eigenvalue from (4.21) where λ_{ir} 's are the diagonal elements of the corresponding subcarrier and stream. For $\lambda_{ir} = 0$, take $p_{ir} = 0$. For each subcarrier, create a diagonal matrix $\tilde{\mathbf{P}}_n^k$, whose diagonal elements are corresponding p_{ir} 's.
5. Find the transmit weight vectors at each subcarrier k , from $\mathbf{V}_n^k = \mathbf{U}_n^k \mathbf{B} \Phi_n^{k-1/2}$, when \mathbf{B} is introduced in the previous Algorithm.
6. Set $n = n + 1$ and continue from step 2 until convergence.

For each iterative water-filling scheme, the choice of receive weight vectors at a receiver has no effect on the amount of interference at other users. Therefore, the receive weight vectors are evaluated using Eq. (4.10), after the iterative algorithms are converged.

Note that the above iteration is independent of the receiver processing. We saw that a receiver can achieve the maximum achievable rate by applying the optimum processing (MMSE receive weight vector (4.10)). In the simulation result of multi-cell system, we will avoid the receiver processing by using the above capacity as the achievable rate.

4.3.4 Game Theoretic approach for rate maximization

As described in Section 4.3, each user tries to maximize its mutual information or the maximum achievable rate, by allocating powers to different streams at different subcarriers of an

OFDM system. However, it is apparent that the choice of transmit weight vectors by user i will affect the interference on other users, and therefore no centralized solution is available for this problem, unless we define a different utility function, like the total rate of all users, constrained to some fairness conditions. This problem might be very difficult to approach, and therefore we proposed the distributed iterative solution.

The structure of the optimization problem suggests that it can fit into the context of game theory [104, 105]. The players of the game are the transmitters that try to select the transmit and receive weight vectors, and the utility function to be optimized is the mutual information or the maximum achievable rates for each user. Therefore by defining $\mathcal{W}_i = \{\mathbf{W}_i^1, \mathbf{W}_i^2, \dots, \mathbf{W}_i^K\}$, $\mathcal{V}_i = \{\mathbf{V}_i^1, \mathbf{V}_i^2, \dots, \mathbf{V}_i^K\}$, the non-cooperative game for the fixed power per user can be established as a zero-sum (players with opposed preferences) strictly competitive game as:

$$\begin{aligned} \max_{\mathcal{W}_i, \mathcal{V}_i} \left\{ u_i = \sum_{k=1}^K \mathcal{I}_i^k = \sum_{k=1}^K \log_2 \prod_{t=1}^T (1 + p_{it}^k \lambda_{it}^k) = \sum_{r=1}^{KT} \log_2 (1 + p_{ir} \lambda_{ir}) \right\} \\ \text{subject to } \sum_{k=1}^K \text{tr} \left(\mathbf{V}_i^{kH} \Phi_i^k \mathbf{V}_i^k \right) \leq P_m, \end{aligned} \quad (4.32)$$

u_i is the utility function for user i that needs to be maximized. This game can be formulated as an extensive game in which there is perfect (or imperfect) information from other players game. The selected strategy of a player in an extensive game depends on the previously selected strategy of other users, as opposed to the strategic games (pure or mixed), in which a user has absolutely no information about other player's strategy, or all the users choose their strategies simultaneously (For two users, it is called Stackelberg game [104, 105]). The saddle point of an extensive game is different from the Nash equilibrium obtained for strategic games. However, in some specific situations (like game of a player and that nature, where the player moves first) this saddle point is the same as the Nash equilibrium. So we use the concept of Nash equilibrium:

Definition 1. *Define*

$$\mathcal{V}_i^- = \{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_{i-1}, \mathbf{V}_{i+1}, \dots, \mathbf{V}_M\} \quad (4.33)$$

and

$$\mathcal{W}_i^- = \{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{i-1}, \mathbf{W}_{i+1}, \dots, \mathbf{W}_N\} \quad (4.34)$$

as the transmit and receive strategies of all players other than user i . Then the Nash Equilibrium Saddle Point (NESP) is defined as a point in which

$$\mathcal{I}_i(\mathbf{V}_i, \mathcal{V}_i^-, \mathbf{W}_i, \mathcal{W}_i^-) \geq \mathcal{I}_i(\mathbf{V}'_i, \mathcal{V}_i^-, \mathbf{W}'_i, \mathcal{W}_i^-) \quad (4.35)$$

where \mathbf{V}'_i , and \mathbf{W}'_i belong to the set of admissible matrices (those who satisfy the constraints defined in (4.32)).

In other words, the NESP is a point in which given other users' weight vectors, none of the transmitters can increase their mutual information by modifying their strategies alone.

The maximization problem is performed in a distributed fashion, and we proposed the iterative water-filling to approach the optimal weight vector allocations. Simulation results have shown that, starting from different initial values for powers, and transmit weight vectors, the proposed algorithm always converge to a unique solution. This fact motivates us to propose the following Conjecture, which we will try to prove in the future. At the moment, the exact necessary conditions are not quite apparent.

Conjecture 4.3.1. *Assuming that there exists a solution for $\text{tr}(\mathbf{V}_i^{kH} \Phi_i^k \mathbf{V}_i^k) \leq P_m$, (the set of feasible solutions is nonempty) the iterative water-filling proposed in Algorithm II converges to the unique NESP for the game defined in (4.32). The NESP is the maximum achievable rate solution for the multiuser OFDM system.*

An insight into the proof comes from the fact that there exists an NESP for a non-cooperative game if the set of feasible transmit weight vectors is non-empty, convex, and compact and if the mutual information function \mathcal{I}_i is continuous in \mathbf{V}_i and \mathbf{W}_i .

Note that the capacity of the channel is the infimum of the maximum achievable rates (\mathcal{I}_i), when the infimum is taken over all possible channel gain matrices.

The same game can be established for the case when the power per carrier is fixed.

4.3.5 Sub-optimal Solution; Same Cell Interference

As was mentioned in the Section 4.2, we have considered a cellular system in which there are M co-channel cells, each cell containing one base station and N_m mobiles, performing downlink transmission, where the base stations are the transmitters and the mobiles are the receivers. In this case the interference at each mobile can be divided into two categories; the interference the signals transmitted from the same cell base station to other mobiles in the cell. The second category consists of the interference coming from other base stations. Due to the shorter distance of the same cell base station compared to other base stations, the former of these two interferences is more significant. Therefore, the performance degradation would not be very significant if we modify either Algorithm I or II to consider only the same cell interference. It is worthwhile to mention that in Algorithms I and II, the weight vectors are calculated in such a way that the achievable rates are maximized at each receiver, considering all other signals as interference. In other words the m^{th} base station transmits a sum of $N_m T$ symbols simultaneously. Every T streams are aimed to be transmitted to one user in the cell, and the weight vectors are adjusted to the channel from the base station and the desired mobile.

Algorithms III and IV in the following, outline the iterative suboptimal solution for the maximum achievable rates for fixed power per subcarrier and per user, respectively.

Algorithm III [Single-cell iterative water-filling, constant power per subcarrier]

1. For receiver n ($n = 1 \dots N$) set the elements of the first row of V_n^k to 1, and the rest to 0. Set $n = 1$
2. Assume that the receiver n is in the cell determined by base station m . Define $S_m = \{1 + \sum_{m'=1}^{m-1} N_{m'}, \dots, \sum_{m'=1}^m N_{m'}\}$ as the set of the indices of all mobiles in cell m . Find the interference at the k^{th} subcarrier from the following equation, which is the modification of Eq. (4.5).

$$\mathbf{Q}_n^k = \sum_{\substack{i \in S_m \\ i \neq n}} \mathbf{H}_{mi}^k \mathbf{V}_m^k \Phi_m^k \mathbf{V}_m^{kH} \mathbf{H}_{mi}^{kH} + \sigma_n^{k2} \mathbf{I}_{N_r}, \quad (4.36)$$

3. Find the eigenvalue decomposition of $\mathbf{H}_{nn}^{kH} \mathbf{Q}_n^{k-1} \mathbf{H}_{nn}^k = \mathbf{U}_n^k \mathbf{\Lambda}_n^k \mathbf{U}_n^{kH}$.
4. Find the power of each eigen-mode with nonzero eigenvalue from (4.21) Create a diagonal matrix $\tilde{\mathbf{P}}_n^k$, whose diagonal elements are p_{ij} .
5. Find the transmit weight vectors from $\mathbf{V}_n^k = \mathbf{U}_n^k \mathbf{B} \Phi_n^{k-1/2}$, when \mathbf{B} is defined appropriately.
6. Set $n = n + 1$ and continue from step 2 until convergence.

Algorithm IV [Single-cell iterative water-filling, constant power per user]

1. For the receiver n ($n = 1 \dots N$) and all subcarriers k ($k = 1 \dots K$), set the elements of the first row of V_n^k to 1, and the rest to 0. Set $n = 1$.
2. Find the single-cell interference form (4.36).
3. For each carrier k , find the eigenvalue decomposition of $\mathbf{H}_{nn}^{kH} \mathbf{Q}_n^{k-1} \mathbf{H}_{nn}^k = \mathbf{U}_n^k \mathbf{\Lambda}_n^k \mathbf{U}_n^{kH}$ (as in (4.13)).

4. For $r = 1 \dots KT$, find the power of each eigen-mode with nonzero eigenvalue from (4.21). For each subcarrier, create a diagonal matrix $\tilde{\mathbf{P}}_n^k$, whose diagonal elements are corresponding p_{ir} 's.
5. Find the transmit weight vectors at each subcarrier k , from $\mathbf{V}_n^k = \mathbf{U}_n^k \mathbf{B} \Phi_n^{k-1/2}$, when \mathbf{B} is defined appropriately.
6. Set $n = n + 1$ and continue from step 2 until convergence.

In each case, the optimal processing is assumed at each receiver. Note that although we only consider the in-cell interference to perform the iterative algorithm, the overall interference is taken into account to evaluate the achievable rate at each receiver.

4.4 Single Stream SNR Maximization

If a single stream is transmitted from each base station to its corresponding mobiles, from (4.15), the rank of the system is 1 and therefore, all of the powers allocated to a subcarrier is given to that stream. As a result Algorithms I-IV are not applicable in this case. For this reason, we will consider the problem of determining transmit and receive weight vector from a different point of view. In the previous section, we tried to maximize the maximum achievable rate at each receiver, while in this section we look at the actual achieved rate. It is well known that the throughput of the transmission link from a transmitter to the i^{th} receiver at the k^{th} subcarrier of an OFDM is

$$r_i^k = \log_2 \left(1 + \frac{\gamma_i^k}{\Gamma} \right), \quad (4.37)$$

where γ_i^k is the Signal to Noise Ratio (SNR) at the k^{th} subcarrier, and Γ is the "SNR Gap" which converts the capacity (obtained from Shannon formula) to an achievable rate. This gap

is a function of the coding and modulation scheme, and the desired Bit Error Rate (BER). As a result, since the \log_2 function is a monotonically increasing function, maximizing the SNR is equivalent to maximizing the achieved throughput, at the subcarrier. To this end, we consider frequency domain beamforming, where the transmitter and receiver each have a beamforming weight vector at each OFDM subchannel. The received signal at subchannel k , at the n^{th} receiver is given by

$$\begin{aligned} x_n(k) &= \mathbf{w}_n^H(k) \mathbf{x}_n(k) \\ &= \mathbf{w}_n^H(k) \sum_{m=1}^N \sqrt{P_m(k)} \mathbf{H}_{mn}(k) \mathbf{v}_m(k) s_m(k) + \mathbf{w}_n^H(k) \mathbf{n}_n(k) \end{aligned}$$

where we have considered having N rather than M transmitters to reflect the fact that each base station transmits independent data streams to different receivers in its cell simultaneously, and for each receive, one transmit beamforming weight vector is calculated. Here $\mathbf{v}_m(k)$ and $s_m(k)$ are the transmitter beamforming vector and the message at the m^{th} transmitter, $\mathbf{w}_n(k)$ is the received beamforming weight vector, and $\mathbf{n}_n(k)$ is the noise vector at the receiver n , all in subchannel k . The noise samples are considered to be independent with zero mean and variance σ^2 . $P_m(k)$ is the transmitting power of transmitter m at subchannel k . By defining $f_{mn}(k) = \mathbf{w}_n^H(k) \mathbf{H}_{mn}(k) \mathbf{v}_m(k)$, the SINR at subchannel k is given by

$$\gamma_n(k) = \frac{P_n(k) |\mathbf{w}_n^H(k) \mathbf{H}_{nn}(k) \mathbf{v}_n(k)|^2}{\sum_{m \neq n} P_m(k) |\mathbf{w}_n^H(k) \mathbf{H}_{mn}(k) \mathbf{v}_m(k)|^2 + \sigma^2 \mathbf{w}_n^H(k) \mathbf{w}_n(k)} \quad (4.38)$$

Let's assume that the transmitting power at each subchannel is a known value $P_m(k) \|\mathbf{v}_m(k)\|^2 = P_0$.

It is straightforward to see that the receive weight vector that maximizes the SNR can be obtained either by MVDR approach or MMSE approach and is given by [26]

$$w_n(k) = \mu_n(k) \mathbf{Q}_n^{-1}(k) \mathbf{H}_{nn}(k) \mathbf{v}_n(k), \quad (4.39)$$

where $\mu_n(k)$ is a constant coefficient. In the case of MVDR, the weight vector is given by

$$w_n(k) = \frac{\mathbf{Q}_n^{-1}(k)\mathbf{H}_{nn}(k)\mathbf{v}_n(k)}{\mathbf{v}_n^H(k)\mathbf{H}_{nn}^H(k)\mathbf{Q}_n^{-1}(k)\mathbf{H}_{nn}(k)\mathbf{v}_n(k)}, \quad (4.40)$$

where

$$\mathbf{Q}_n(k) = \sum_{m \neq n} [P_m(k)\mathbf{H}_{mn}(k)\mathbf{v}_m(k)\mathbf{v}_m^H(k)\mathbf{H}_{mn}^H(k)] + \sigma^2\mathbf{I}. \quad (4.41)$$

In this case, $\mu_n(k)$ is equal to the inverse of denominator of (4.40). Given this receive weight vector, the SNR is obtained from

$$\begin{aligned} \gamma_n(k) &= P_n(k)\mathbf{v}_n^H(k) (\mathbf{H}_{nn}^H(k)\mathbf{Q}_n^{-1}(k)\mathbf{H}_{nn}(k)) \mathbf{v}_n(k) \\ &= P_0\mathbf{v}'_n{}^H(k) (\mathbf{H}_{nn}^H(k)\mathbf{Q}_n^{-1}(k)\mathbf{H}_{nn}(k)) \mathbf{v}'_n(k), \end{aligned} \quad (4.42)$$

where $\|\mathbf{v}'_m(k)\|^2 = 1$. With the constant length constraint for $\mathbf{v}'_m(k)$, it is proven that (4.42) is maximized when $\mathbf{v}'_m(k)$ is the principal eigenvector of the matrix $\mathbf{H}_{nn}(k)\mathbf{Q}_n^{-1}(k)\mathbf{H}_{nn}^H(k)$ [82, 98, 106]. The principal eigenvector of an irreducible matrix A is the eigenvector corresponding to the $\rho(A)$, which is the real positive eigenvalue of A with the maximum modulus (spectral radius).

If the network enforces a fixed transmit power per user, the weight vectors (transmit and receive) of different subchannels cannot be calculated independently. In this case, we have $\sum_{k=0}^{K-1} P_m(k)|\mathbf{v}_m(k)|^2 \leq P_0$. The total user power is divided among the subchannels in such a way to maximize the overall data rate of each user. Therefore, the optimization problem in this case is

$$\begin{aligned} &\max \left\{ r_i = \sum_{k=1}^K r_i^k \right\}, \\ &\text{subject to } \sum_{k=1}^K P_m(k)|\mathbf{v}_m(k)|^2 \leq P_0, \end{aligned} \quad (4.43)$$

where r_i^k is the data rate at subcarrier k at user i , obtained from 4.37), and r_i is user i 's overall rate.

Assuming that we know the SNR at each subcarrier, it is well known that under a fixed overall power, water-filling [101] achieves the maximum overall rate. However, for the case of multiuser environment, the power allocation among different subchannels for one user affects the SINR of other users. Therefore, the following distributed iterative algorithm is proposed for all users, simultaneously:

1. For each subchannel k ($k = 0 \dots K - 1$), initialize the transmission power, transmit and receive weight vectors
2. For the k^{th} subcarrier of user $m = 1, \dots, N$, calculate the receive weight vector so as to maximize the achievable SINR, γ_m^k , using either MVDR or MMSE approach both represented by Eq. (4.39).
3. Use the water-filling algorithm to find the power at each subcarrier from the following relation:

$$P_m(k) = \left(\nu_m - \frac{\Gamma}{\gamma_m(k)} \right)^+, \quad (4.44)$$

where $(.)^+$ is zero when the argument is negative, Γ is the SNR gap, and the constant ν_m is chosen such that the total transmit power is equal to P_0 :

$$\sum_k P_m(k) = P_0.$$

4. Fix the receive weight vectors, and calculate the transmit weight vector as the principal eigenvector of the matrix $\mathbf{H}_{mm}^H(k)\mathbf{Q}_m^{-1}(k)\mathbf{H}_{mm}(k)$, where $\mathbf{Q}_m(k)$ is the k^{th} subcarrier interference defined as in (4.41).
5. Repeat from step 2, until convergence.

As in the case of multiple steam MIMO/OFDM, it is possible to establish the problem in the framework of game theory. Observing the simulation results suggest that we can propose

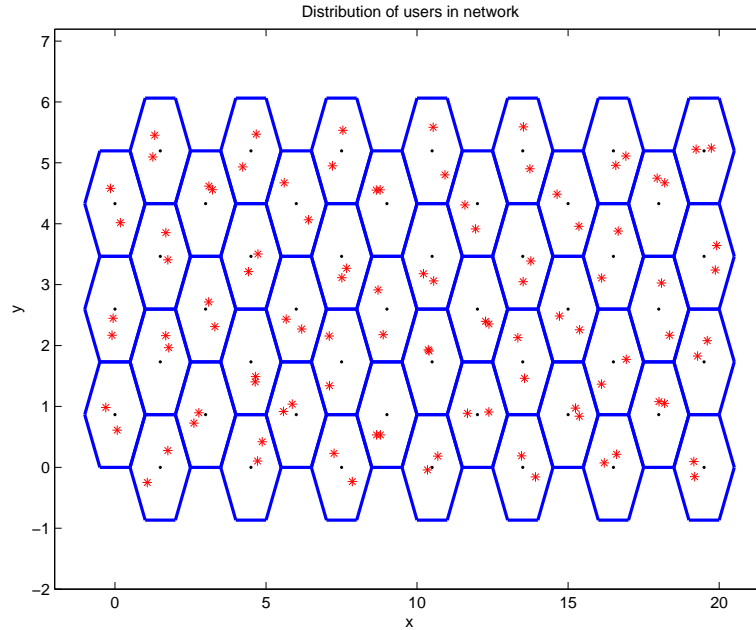


Figure 4.1: Cellular structure and distribution of users

a Conjecture similar to (4.3.1) to show that this algorithm converges to the NESP of this game.

4.5 Simulation

We use a wireless network consisting of M base stations placed in a hexagonal pattern. We assume that all of the base stations belong to the same co-channel set and each cell contains one base station and N_m mobile. The cellular pattern is shown in Figure 4.1, depicted for $M = 16$, and $N_m = 2$. Users are randomly distributed in a cell according to a uniform distribution. We use an OFDM system with 8 subchannels for transmission. The communication channel is assumed to follow the COST207 Typical Urban 6-ray channel model with average path delays of $\{0.0, 0.2, 0.5, 1.6, 2.3, 5.0\}$ measured in μs and path fading powers of $\{0.189, 0.379, 0.239, 0.095, 0.061, 0.037\}$ [90]. The maximum channel delay spread

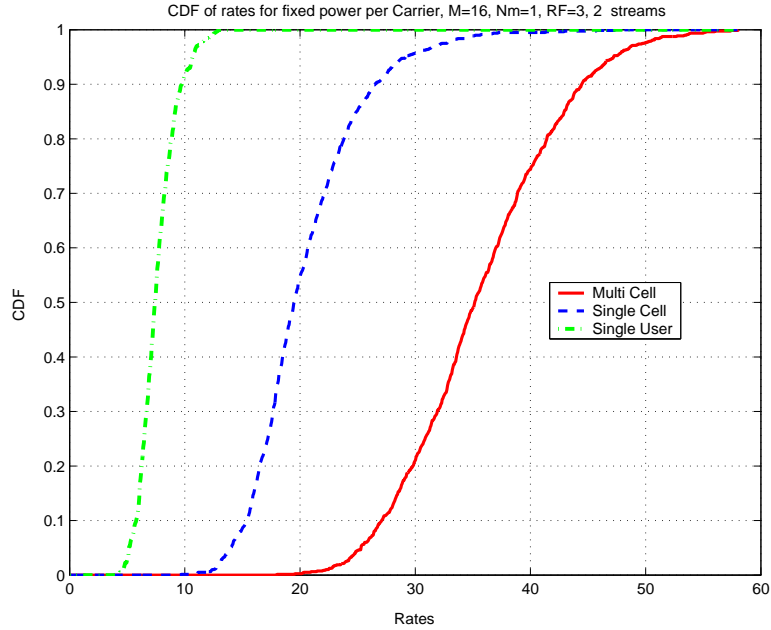


Figure 4.2: Achievable rate CDF, for fixed power per carrier with 16 base stations, 2 mobile per cell, reuse factor of 7 and 2 streams

is $5\mu s$ and so the channel coherence bandwidth is $200KHz$. Link gains are calculated by considering $6 - 8dB$ for the variance of shadow fading and the path loss exponent to be four. We assume a quasi-static channel where the channel is assumed to be fixed over multiple OFDM symbols. Channel frequency response can be obtained simply by taking the Fourier transform of the time-domain channel impulse response, and sampling this response at the carrier frequency, mf_c , where f_c is the subcarrier separation; i.e.

$$H^m = \sqrt{G} \sum_{l=0}^{L-1} \alpha_n^l e^{-j2\pi m f_c \tau_n^l}, \quad (4.45)$$

where G captures the effect of path loss and shadow fading which we consider the same for different paths. Any difference can be absorbed in fading coefficients. L is the number of multipaths, τ^l the delay of each path, and α^l is the fading of path l . Note that the subchannel link gains for each user are correlated according to (4.45).

A one-tap frequency-domain equalizer is assumed at the receiver such that the channel

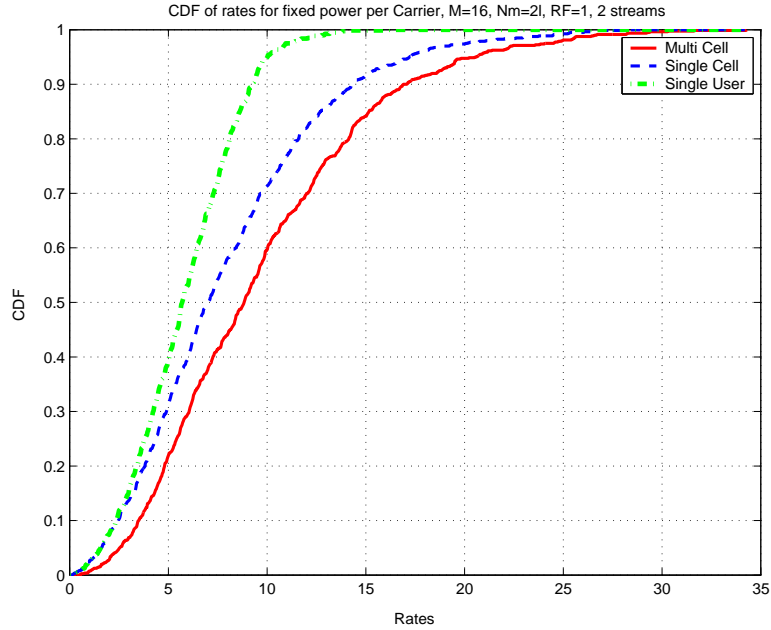


Figure 4.3: Achievable rate CDF, for fixed power per carrier with 16 base stations, 2 mobile per cell, reuse factor of 1 and 2 streams

between the base station and its desired mobile can be estimated. Each OFDM symbol is assumed to be $2\mu s$ long which corresponds to a bandwidth of $500KHz$. Therefore, the subcarrier spacing is $16KHz$ which is smaller than the coherence bandwidth of the channel and so the fading at each subchannel can be considered flat. The average power of the signal at each subchannel at each transmitter is assumed to be unity. The white Gaussian thermal noise power at each receiver is calculated based on a noise figure of 3dB and the receiver bandwidth, which is assumed to be $500KHz$ bandwidth. Each base station uses four transmit antenna and each mobile uses four receive antennas. Multiple data streams are transmitted from each base stations to each mobile in its corresponding cell.

We have simulated Algorithms I, II and III. Note that in Algorithm I, we consider multi-cell optimization, in which the interference from all base stations are considered to find the transmit and receive weight vectors, and the power is fixed per subcarrier. Algorithm II

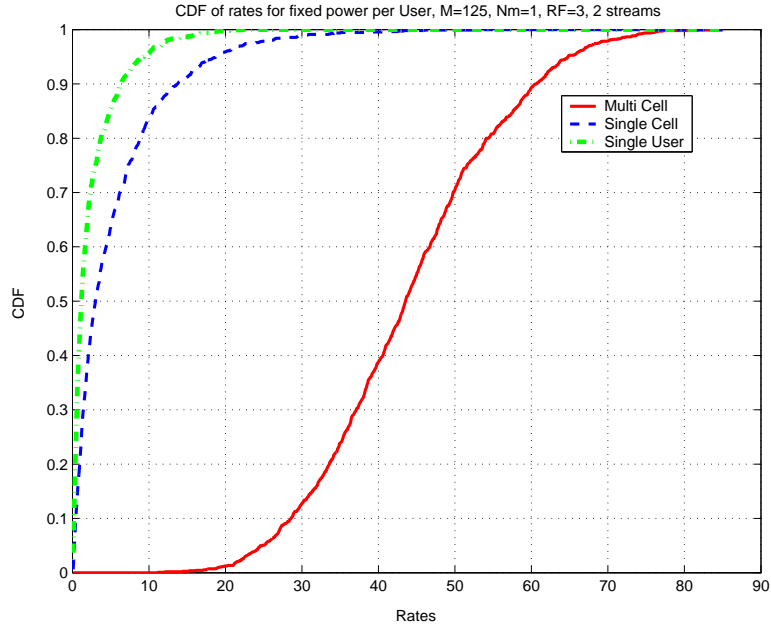


Figure 4.4: Achievable rate CDF, for fixed power per user with 125 base stations, 1 mobile per cell, reuse factor of 3 and 2 streams

considers the same configuration, but with fixed power per user. In this case, we needed to adapt the power at each subcarrier. In Algorithms III only the in-cell interference is considered to perform the optimization, while multi-cell interference is included at the final step to find the OFDM rate, and the power is fixed per user. Here, we have ignored the out of cell interference in performing the iterative optimization. Note that the same experiment has been simulated when the power per carrier is fixed (which we call it Algorithm III'). For reference, we have simulated Algorithm III (and III'), by only considering the thermal noise covariance in optimization. Again, the multi-cell interference is considered for evaluating the final OFDM rate. In each case the appropriate noise and interference covariance matrix (the one used for optimization) is measured at each receiver and is fed back to the base station for transmit weight optimization.

Figure 4.2 shows the CDF of the achievable rate for the Algorithms II, III, and noise only

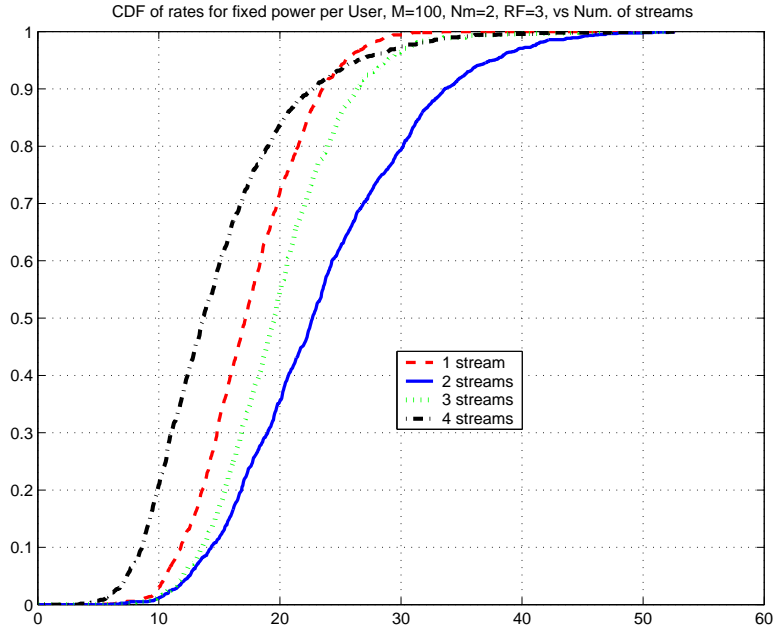


Figure 4.5: Achievable rate CDF of different number of streams, for fixed power per user with 100 base stations, 2 mobile per cell, and reuse factor of 3.

optimization, with 16 base stations, 16 base stations, 2 mobile per cell, reuse factor of 1 and 2 streams. Figure 4.2 considers the same configuration with a reuse factor of 1. Both of these figures show that the maximum achievable rate for multi-cell optimization outperforms those of single-cell and also noise-only optimizations. Moreover, comparing these two figures, infers the expectable result that increasing the reuse factor, reduces the interference and therefore results in higher capacities. However, this improvement is more significant in the case of multi-cell optimization. The same argument is valid for fixed power per user, as is seen in Figure 4.4.

Fig. 4.5 depicts the CDF of achievable rates for 100 base stations, 2 mobile per cell, reuse factor 3, for different number of streams when the power is fixed per user. The same quantities are depicted for 16 cells, two mobiles per cell in Fig. 4.6 and 16 cells, one mobiles per cell in Fig. 4.7. When there is only one mobile per cell (Using some multiple-access

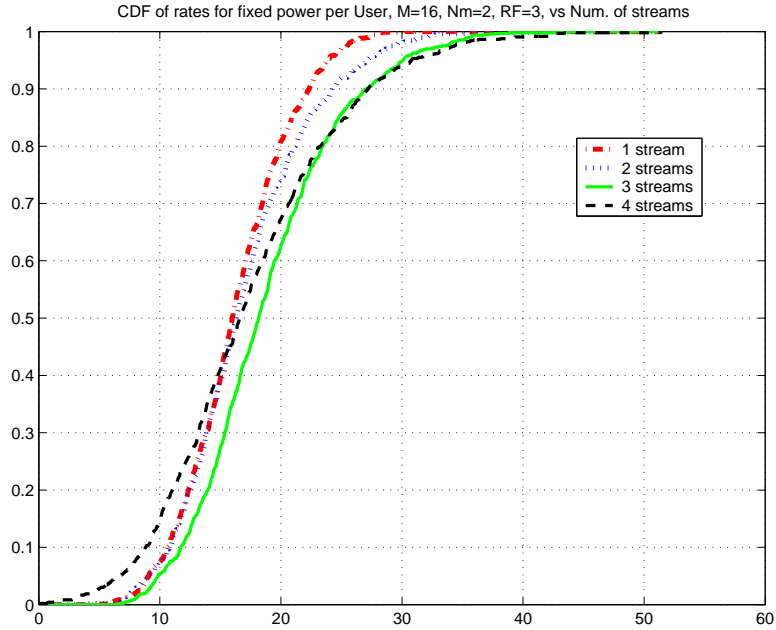


Figure 4.6: Achievable rate CDF of different number of streams, for fixed power per user with 16 base stations, 2 mobile per cell, and reuse factor of 3

methods, like TDMA, or CDMA, the mobiles per cell are considered no interfering), as we increase the number of streams, more bandwidth can be assigned per user and the maximum achievable rate per user is increased. When we increase the number of co-channel mobiles per cell, more processing is needed to combat the effect of in-cell interferences, and therefore the performance of two streams is better than that of four streams. When we increase the number of cells, it is clearly shown that multiple stream will degrade the maximum achievable rate.

4.6 Summary

We have proposed iterative water-filling solutions for multi-user multi-cell wireless systems where multiple antennas are deployed at both transmitters and receivers. The proposed

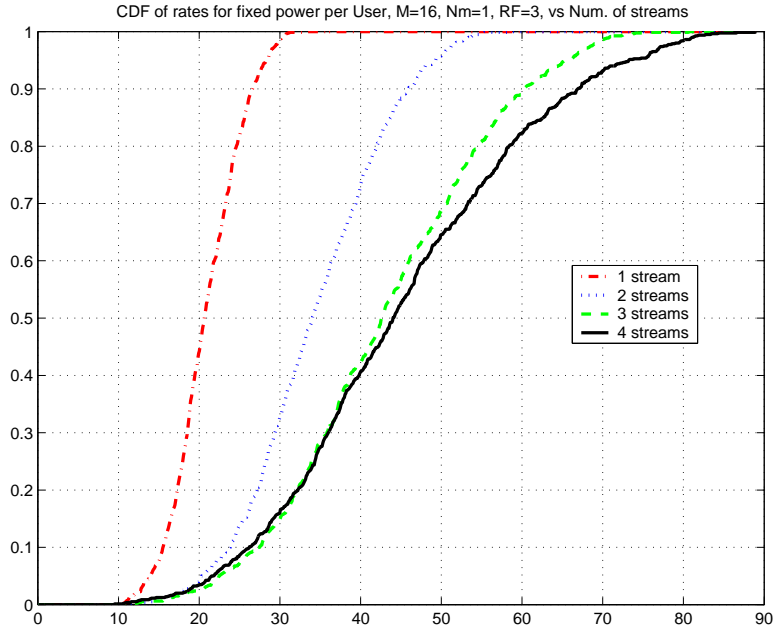


Figure 4.7: Achievable rate CDF of different number of streams, for fixed power per user with 16 base stations, 1 mobile per cell, and reuse factor of 3

algorithm assigns multiple independent substreams for each user to increase data rate for each user and is performed as a distributed scheme. We have established a non-cooperative game theoretic analogy for the MIMO/OFDM problem and proposed a Conjecture that the iterative algorithms proposed in this work will converge to the Nash equilibrium saddle point of game. The iterative algorithm that considers single stream transmission, and tries to maximize the actual transmission data rate of each user by performing transmit and receive beamforming. Through numerical analysis, we observed that if the number of co-channel mobiles per cell is increased, it is better to limit the number of streams.

Chapter 5

Low Peak to Average Power Ratio with modified Golay Sequences for OFDM Systems

5.1 Motivation and Previous Works

One of the major hurdles to the widespread use of OFDM is the high Peak to Average Power Ratio (PAPR) of OFDM signals [107]. A large PAPR may force the amplifier to operate in the nonlinear region, or back-off the saturation point. These might reduce the mobile battery lifetime, or degrade the signal quality.

Many approaches have been proposed to overcome this barrier [69, 70, 108–120]. In Section 2.5 we categorized these approaches in three classes, and describe each class briefly. We will focus more on the block coding approach in this dissertation. Cyclic coding has been used in [110] and involves adding extra carriers in which the phase of every fourth carrier is calculated from the phase of previous three information carriers. Lawrey and Kikkert [119] presented a technique that combines SLM and Cyclic coding. They added an extra carrier, referred to as Peak Reduction Carriers (PRC), whose phase and amplitude is varied to minimize overall PAPR. Another block coding approach was proposed in [120] using codewords drawn from offsets of a linear code. The work in [121] goes further and proposes a com-

putationally efficient algorithm which, given any code and a maximum-likelihood decoding algorithm for that code, finds good offsets. However this method does not guarantee to obtain significant PAPR reduction all of the times.

In one of the most recent and efficient works done in this area, Golay Complementary Sequences (GCS) [122] are used, to control the modulation of carrier information, resulting in OFDM signals with PAPR of at most 2. The correlation properties of these codes have made them a suitable choice for several applications like multi-code CDMA systems [123]. Recently, the IEEE802.11 standard committee has adopted Complementary Code Keying (CCK) signals (that are in principal QPSK Golay complementary sequences with size 8) as the physical layer of Wireless LAN IEEE802.11b standard. Davis and Jedwab in [118] obtained a large set of length 2^m binary Golay Complementary Pairs (GCP) from certain second order cosets of the first order Reed-Muller (RM) codes [13]. They combined block coding schemes (with all of its properties, like efficient encoding and decoding, error correction capabilities) with the use of GCS (with their attractive power control properties). They also went one step further and found 2^h -ary GCP from cosets of an appropriate generalization of the Reed-Muller codes. As a result, they found good binary, quaternary and octary OFDM codes with good error correcting capabilities, efficient encoding and decoding, and a tightly controlled PAPR. By allowing higher PAPR, they were able to guarantee higher coding rates.

Paterson [112] generalized Davis and Jedwab's results in two ways. First, he used q -ary instead of 2^h -ary alphabets (with q even), related to the general q -ary second order cosets of the first order Reed-Muller codes. Secondly, instead of Golay complementary pairs, he defined the concept of Golay complementary sets of size 2^{k+1} . He showed that any sequence lying in a Golay complementary set of size 2^{k+1} has PAPR at most 2^{k+1} , and found upper bounds and lower bounds on the PAPR of complete second-order cosets of $RM_q(1, m)$ (the

first order Reed-Muller code defined over \mathcal{Z}_q). Moreover, he gave an explicit, non-recursive construction of q -phase Golay complementary sets. However, Paterson and Tarokh [115] performed a theoretical analysis for a general coding scheme, having a constellation with equal energy symbols. They found a trade-off among the PAPR, the data rate, and the minimum distance of the codebook, regardless of the coding scheme. Given the minimum distance of the code, they found a lower bound for the PAPR that increases with increasing data rate. They also showed that, in general, if the PAPR is bounded up by a constant value, the data rate decreases with decreasing minimum distance of the code. Davis and Jedwab [118] have provided several trade-offs among different codes with different PAPR and rates.

In the first part of this chapter, we will look at the general problem of PAPR reduction in multicarrier systems and specifically try to overcome the limitation on PAPR reduction imposed by coding rate [115]. Here, we relax the assumption of having an equal energy constellation and use QAM for modulation. We define a new version of Golay Complementary Sequences to support these codes. The scheme presented in this chapter uses a recursive procedure to build the SGolay 16-QAM sequences. We generalize the recursive schemes introduced in [122] for the case of binary Golay sequences. Using QAM modulation allows a better error correcting capabilities over QPSK, too. Tarokh and Chong in [124] designed a construction method for low PAPR 16-QAM that uses the Jedwab's construction for QPSK Golay sequences [118]. The coding rate achieved by this construction is low (about 3% for 256 subchannels). It should be pointed out that their constructions does not guarantee an upper bound of $3dB$ for the OFDM PAPR. If they want the PAPR to be bounded up to $3dB$, their coding rate will be cut in half. Tarokh and Roessing, in a related work [117], designed another construction for 16-QAM low PAPR codes. In this work a QAM symbol is represented as a weighted sum of two QPSK symbols. The constructed sequences are not Golay, but their PAPR is bounded up to 3.6 ($=5.6dB$), and the coding rate is twice as the

rate for Golay QPSK codes found in [118]. For example for $N = 256$ it results in a rate of $2 * 50 / (4 * 256) = 9.7\%$ while the PAPR is guaranteed to be bounded up to 3.6 (5.5dB).

In the second part of this chapter we will set an upper bound for the PAPR ($3dB$) and provide some trade-offs among the coding rate and the error correction capabilities of the OFDM code, by developing the concept of cyclic Golay codes. We will present a method for constructing these codes out of Golay codes. The construction method we will propose in this chapter, is suitable for generating the cyclic shift of any code described by means of Boolean functions. We will show that If these codes are used as the codewords applied to the IFFT block in an OFDM system the PAPR is bounded up to $3dB$, and present a trade-off between among the coding rate and minimum distance of these codes. Note that if the signal to noise ratio in an environment is above a threshold, we might be able to tolerate low distance codes with the expense of higher rates. Although, this work does not offer a very significant rate increase over the Golay codes, it outlines a methodology to generate the cyclic shift of any code, presented by Boolean algebraic functions, even if the generated code is not linear (the construction of linear cyclic codes is well known in the literature).

We will show in this chapter that cyclic Golay codes are in principal special cases of Generalized Reed-Muller codes. Decoding of the first order RM codes is done efficiently by exploiting the fast Hadamard transformations [13]. Higher order binary RM codes are decoded using majority logic Reed algorithm [13]. Grant *et. al.* [125], and Davis *et. al.* [118], in different works, proposed second order coset decoding of first order Generalized RM codes defined over alphabet \mathcal{Z}_{2^h} . Paterson and Jones [126] presented several optimal and suboptimal decoding algorithms, both hard and soft decision decoding, and both in signal and coding domain for second order generalized RM codes. They also mentioned about the extension of their reduction decoding scheme, proposed for second order codes, to higher order Generalized RM codes. In this chapter, we will focus on both soft and hard-decision

decoding of generalized RM codes of any order in both complex and coding domain, from a different perspective. We will provide two decoding algorithms for $RM_{2^h}(r, m)$. The first one is a generalization of Reed algorithm for binary Reed-Muller codes. We will restate this algorithm using the concept of Karnaugh Map or K-Map [127], and present it for $RM_{2^h}(3, 4)$. The decoding steps for other sizes and orders can be obtained similarly. We also propose a recursive decoding algorithm to avoid the complexity of higher dimensional K-Maps.

The remainder of this chapter is organized as follows: We first introduce the concept of Golay Complementary Sequences (GCS) and their impact on the PAPR of OFDM systems using equal-power constellations. Then Section 5.2.2 outlines a proposed structure to achieve low PAPR for unequal power constellations. In Section 5.2.3 a seed for our recursive procedure is introduced. Section 5.3 outlines the definition of cyclic Golay codes and their PAPR properties. In Section 5.3.1, we will find construction methods for building these codes out of Davis-Jedwab Golay codes. Section 5.3.2 outlines two decoding algorithms for $RM_{2^h}(r, m)$ (in particular, cyclic Golay codes), one recursive and one non-recursive. In Section 5.3.3 we present some simulation results for both parts, and finally Section 5.3.4 concludes the chapter.

5.2 Golay Complementary Sequences for equal-power constellations

Channel coding is a mean to perform the PAPR reduction and some error correction for OFDM systems simultaneously. If the total number of bits assigned to one OFDM symbol is Nm , when N is the number of subchannels, then we choose a N -valued sequence from a codebook and this sequence is fed into the FFT block. if x is submitted codeword to the

IFFT block, the transmitted OFDM signal is

$$s_x(t) = \sum_{i=0}^{N-1} x_i \exp[j2\pi(f_0 + i\Delta f)t], \quad (5.1)$$

where Δf is the subcarrier separation, and f_0 is the carrier frequency. The instantaneous envelope power of the signal is

$$p_x(t) = |s_x(t)|^2 = \sum_{i=0}^{N-1} \sum_{u=0}^{N-1} x_i x_u^* \exp[j2\pi(i-u)\Delta f t]. \quad (5.2)$$

Let's define the cross-correlation of two sequences and auto-correlation of a sequence as follows:

Definition 2. *The Cross-Correlation of two N -valued complex codes x and y with replacement $-N \leq l \leq N-1$ is defined as*

$$C_l(x, y) = \begin{cases} \sum_{i=0}^{N-l-1} x_{i+l} y_i^* & \text{if } 0 \leq l \leq N-1, \\ \sum_{i=0}^{N+l-1} x_i y_{i-l}^* & \text{if } -N \leq l \leq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.3)$$

In other words, for non-negative l 's we have

$$C_l(x, y) = \sum_{i=0}^{N-l-1} x_{i+l} y_i^*, \quad (5.4)$$

and $C_{-l}(x, y) = (C_l(y, x))^*$.

Definition 3. *The Auto-Correlation of an N -valued complex codes x with nonzero replacement l is defined as*

$$A_l(x) = C_l(x, x). \quad (5.5)$$

In other words,

$$A_l(x) = \begin{cases} \sum_{i=0}^{N-l-1} x_{i+l} x_i^*, & \text{for } l \geq 0; \\ A_{-l}^*(x), & \text{for } l \leq 0. \end{cases} \quad (5.6)$$

Let's start from (5.2) to restate the instantaneous envelope power of the OFDM signal as follows:

When i and u go from 0 to $N - 1$, if we express $l = i - u$, then l goes from $-(N - 1)$ to $N - 1$. For each $l > 0$, we keep u and the other index, i goes from 0 to $N - 1 - l$, and for each For each $l < 0$, we keep i and the other index, u goes from 0 to $N - 1 - l$. This fact could be easily seen by writing the indices i and u as a matrix. Therefore, using Definition 3, we have

$$\begin{aligned}
p_x(t) &= \sum_{u=0}^{N-1} \sum_{i=0}^{N-1} x_i x_u^* \exp[j2\pi(i - u)\Delta ft]. \\
&= \sum_{i=0}^{N-1} x_i^* x_i + \sum_{l=1}^{N-1} \sum_{u=0}^{N-1-l} x_u^* x_{u+l} e^{j2\pi l \Delta ft} + \sum_{l=-(N-1)}^{-1} \sum_{i=0}^{N-1-l} x_i x_{i-l}^* e^{j2\pi l \Delta ft} \\
&= A_0(x) + \sum_{l=1}^{N-1} A_l(x) e^{j2\pi l \Delta ft} + \sum_{l=-(N-1)}^{-1} A_{-l}^*(x) e^{j2\pi l \Delta ft} \\
&= A_0(x) + \sum_{l=1}^{N-1} A_l(x) e^{j2\pi l \Delta ft} + \sum_{l=-(N-1)}^{-1} A_l(x) e^{j2\pi l \Delta ft} \\
&= A_0(x) + \sum_{|l|=1}^{N-1} A_l(x) e^{j2\pi l \Delta ft} \tag{5.7}
\end{aligned}$$

Note that the last two equalities in (5.7) can be rewritten as

$$\begin{aligned}
p_x(t) &= A_0(x) + \sum_{l=1}^{N-1} A_l(x) e^{j2\pi l \Delta ft} + \sum_{l=-(N-1)}^{-1} A_{-l}^*(x) e^{j2\pi l \Delta ft} \\
&= A_0(x) + \sum_{l=1}^{N-1} A_l(x) e^{j2\pi l \Delta ft} + \sum_{l=1}^{N-1} A_l^*(x) e^{-j2\pi l \Delta ft} \\
&= A_0(x) + 2 \operatorname{Re} \sum_{l=1}^{N-1} A_l(x) e^{j2\pi l \Delta ft} \tag{5.8}
\end{aligned}$$

The average power of the OFDM signal is $A_0(x)$, which is the same as the power of the code x (by Parseval equation) and is denoted by P_x . The maximum possible value for PAPR

in an OFDM system is equal to the number of subchannels, N (for the cases when all of the symbols applied to the FFT block are equal).

Definition 4. *Two N -valued complex sequences x and y are called Golay Complementary Pairs (GCP) if*

$$A_l(x) + A_l(y) = 0 \quad \forall l \neq 0.$$

We show this by $x \sim y$. Each of the sequences x and y is called Golay Complementary Sequence (GCS).

Assume that the sequences x and y are GCP and have the same power ($A_0(x) = A_0(y)$). Since the instantaneous envelope power is non-negative at all times, we can see that

$$PAPR(x) \triangleq \frac{\max_t \{p_x(t)\}}{P_x} \leq 2 = 3dB \quad (5.9)$$

Therefore, if we choose the codes from a set of Golay sequences, the PAPR is bounded up to 3dB.

5.2.1 Construction of equal-power Golay Sequences

A Boolean function is a function f from $\mathcal{Z}_2^m = \{(x_1, x_2, \dots, x_m) | x_i \in \{0, 1\}\}$ to \mathcal{Z}_2 . We regard each 0 – 1 variable x_i as itself being a Boolean function $f_i(x_1, x_2, \dots, x_m) = x_i$ and consider the 2^m monomials

$$1, x_1, \dots, x_m, x_1x_2, x_1x_3, \dots, x_{m-1}x_m, \dots, x_1x_2 \dots x_m. \quad (5.10)$$

Any Boolean function f can be uniquely expressed as a linear combination over \mathcal{Z}_2 of these monomials, where the coefficient of each monomial belongs to \mathcal{Z}_2 . We specify a sequence \mathbf{f} of length 2^m corresponding to f by listing the values taken by $f(x_1, x_2, \dots, x_m)$ as

x_1, x_2, \dots, x_m ranges over all its 2^m combinations in lexicographic order. In other words, if (i_1, i_2, \dots, i_m) is the binary representation of integer i , i.e.

$$i = \sum_{k=1}^m i_k 2^{k-1}, \quad (5.11)$$

the i^{th} element of the sequence \mathbf{f} is $f(i_1, i_2, \dots, i_m)$. For example, with $m = 2$, the Boolean function $f(x_1, x_2)$ defines the following codeword:

$$\mathbf{f}(x_1, x_2) = \left[\underbrace{f(0,0)}_0 \quad \underbrace{f(1,0)}_1 \quad \underbrace{f(0,1)}_2 \quad \underbrace{f(1,1)}_3 \right].$$

Note that, the first symbol in each element of the codeword represents the least significant bit of the lexicographic representation of the index of that element, and the last one represents the most significant bit.

We define a generalized Boolean function to be a function f from \mathcal{Z}_2^m to \mathcal{Z}_{2^h} , where $h \geq 1$. It is straightforward to show that any such function can be uniquely expressed as a linear combination over \mathcal{Z}_{2^h} of the monomials (5.10), where the coefficient of each monomial belongs to \mathcal{Z}_{2^h} . As above, we specify a sequence f of length 2^m corresponding to the generalized Boolean function f . For example, for $h = 2$ and $m = 3$ we have $3x_1 = (0 \ 0 \ 0 \ 0 \ 3 \ 3 \ 3 \ 3)$, $2x_1x_2x_3 = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 2)$, and $x_1x_2 + 3x_2x_3 + 2.1 = (2 \ 2 \ 2 \ 1 \ 2 \ 2 \ 3 \ 2)$. (Technically, for such expressions to be valid we must embed the range space \mathcal{Z}_2^m of the monomials (5.10) in $\mathcal{Z}_{2^h}^m$.)

We assume that the elements of the codeword to be transmitted using an OFDM system, are chosen from an equal-energy constellation like QPSK or 8-PSK. Therefore, without loss of generality, the elements of a codeword y can be written as

$$y_i = \exp\left(\frac{j2\pi}{2^h} a_i\right), \quad (5.12)$$

where a_i is chosen from a 2^h -ary alphabet, \mathcal{Z}_{2^h} . As a result, constructing the sequence a , will provide the sequence y . Using these notations, the main result of [118] is stated as follows:

Theorem 5.2.1. *Let*

$$f \triangleq f(x_1, x_2, \dots, x_m) = \left[2^{h-1} \sum_{k=1}^{m-1} x_{\pi(k)} x_{\pi(k+1)} + \sum_{k=1}^m c_k x_k + c \right] \text{ mod } 2^h, \quad (5.13)$$

where π is a permutation of the symbols $\{1, 2, \dots, m\}$, and $c, c_k \in \mathcal{Z}_{2^h}$. Then, the sequences f and $(f + 2^{h-1} x_{\pi(1)} + c' \text{ mod } 2^h)$ comprise a Golay pair of length 2^m over \mathcal{Z}_{2^h} , for any $c' \in \mathcal{Z}_{2^h}$.

Note that it is not the sequence a with elements $a_i = f(i_0, \dots, i_{m-1})$, when i and the vector (i_0, \dots, i_{m-1}) are related as in (5.11), that is a Golay sequence, but the sequence y defined as in (5.12) is Golay.

Now, let's consider the definition of Reed-Muller codes as follows:

Definition 5. *For $h \geq 1$ and $0 \leq r \leq m$, the r^{th} order linear Reed-Muller code of length 2^m over \mathcal{Z}_{2^h} , $RM_{2^h}(r, m)$, is defined to be generated by all monomials in the x_i of degree at most r .*

Using this definition, the authors in [118] have restated Theorem 5.2.1 as:

Theorem 5.2.2. *Each of the $\frac{m!}{2}$ cosets of $RM_{2^h}(1, m)$ having a coset representation of the form*

$2^{h-1} \sum_{k=1}^{m-1} x_{\pi(k)} x_{\pi(k+1)}$ *comprises $2^{h(m+1)}$ Golay sequences over \mathcal{Z}_{2^h} of length 2^m where π is a permutation of the symbols $\{1, 2, \dots, m\}$.*

By varying c_k 's and c over \mathcal{Z}_{2^h} in (5.13), this theorem generates $\frac{m!}{2} 2^{h(m+1)}$ Golay sequences of length 2^m . So, the coding rate of Davis-Jedwab construction is $\frac{h(m+1) + \log_2(\frac{m!}{2})}{h 2^m}$.

Using these theorems, in Corollary 2.5 of [118], which we repeat here, Davis and Jedwab introduced a construction for building at least $2^{h(m+2)} \frac{m!}{2}$ Golay complementary pairs over \mathcal{Z}_{2^h} of length 2^m .

Corollary 5.2.3. *Let f be defined as in (5.13). Then any sequence in set A forms a Golay complementary pair over \mathcal{Z}_{2^h} of length 2^m with any sequence in set B , when*

$$A = \{f + c, f + 2^{h-1}(x_{\pi(1)} + x_{\pi(m)}) + c \mid c \in \mathcal{Z}_{2^h}\}$$

$$B = \{f + 2^{h-1}x_{\pi(1)} + c', f + 2^{h-1}x_{\pi(m)} + c' \mid c' \in \mathcal{Z}_{2^h}\}$$

5.2.2 PAPR reduction for the non-equal power constellation

Given the length of the code, minimum Euclidian distance, and maximum PAPR, Tarokh and Paterson [115] found a lower bound for achievable coding rate. On the other hand, given the length of the code, coding rate and minimum Euclidian distance, they found a lower bound for the PAPR. The lower bound for the PAPR increases by increasing the coding rate.

The need for low PAPR and at the same time overcoming the lower bound of rate for equal-power codes has motivated us to investigate the non-equal power codes that achieve low PAPR. To do this we define a special case of GCSs.

Definition 6. *Two N -valued complex sequences x and y are called Super Golay Complementary Pairs (SGCP) if*

- *They are Golay complementary pairs.*
- *If P_{av} is the average power of the constellation,*

$$P_x + P_y \leq 2NP_{av},$$

We show this by $x \approx y$. Each of the sequences x and y is called Super Golay Complementary Sequence (SGCS)

For a special case, It has been proved in [117] that if the 16-QAM codes are realized as a sum of two sequences chosen from an equiprobable set of QPSK codewords, then the mean envelope power of the transmitted OFDM symbol is P_{av} , the average power of the 16-QAM constellation. This fact can be easily generalized to our structures introduced in Section 5.2.3.

Theorem 5.2.4. *The PAPR achieved by any SGCS is bounded up by 3dB.*

Proof. By virtue of (5.23) and the fact that the instantaneous power of each code is always non-negative, we have

$$PAPR(x) = \frac{\max_t\{p_x(t)\}}{P_x} \leq \frac{p_x(t) + p_y(t)}{NP_{av}} = \frac{P_x + P_y}{NP_{av}} \leq 2 = 3dB. \quad (5.14)$$

□

Next, we would like to find a construction method for SGolay codes. If x and y are two N -valued sequences, we show the inverse of x by \hat{x} , the element-wise conjugate of x by x^* , the concatenation of x and y by $x|y$, and the interleaving of x and y by $x \downarrow y$. Also we show the sequence $(x_1, -x_2, x_3, \dots, (-1)^{N-1}x_N)$ by x' .

Theorem 5.2.5. *The property of being super Golay complementary pairs is invariant under the following transformations:*

- a .Reflection w.r.t the origin.*
- b .Reflection w.r.t both axes.*
- c .Multiplication of one or both sequences by a complex number with magnitude 1.*
- d .Reflection w.r.t the bisectors of all regions.*
- e .Rotation.*

Proof. We will prove each item separately:

a . Using Definition 3,

$$A_l(-x) = A_l(x). \quad (5.15)$$

b . The reflection of x w.r.t the real axis is x^* . Using Definition 3,

$$A_l(x^*) = [A_l(x)]^*. \quad (5.16)$$

The reflection of a sequence x w.r.t the imaginary axis is $-x^*$.

c . If α is an arbitrary complex number , then

$$A_l(\alpha x) = |\alpha|^2 A_l(x). \quad (5.17)$$

Therefore if $x \approx y$ and $|\alpha| = 1$, then $x \approx \alpha y$ and $\alpha x \approx \alpha y$.

d . The reflection of x w.r.t the bisector of the first and third regions is jx^* . Also, the reflection of x w.r.t the bisector of the second and fourth regions is $-jx^*$.

e . Rotation of a sequence with the angle θ is equivalent to multiplying the sequence by $e^{j\theta}$.

Note that, in all of these cases, the power of the sequences are preserved. □

Theorem 5.2.6. *If $x \approx y$ then*

$$a. x' \approx y'. \quad b. \hat{x} \approx \hat{y} \quad c. x \approx \hat{y}^*$$

Proof. If $x \approx y$ then

a . The k th member of the sequence x' is $x'_k = (-1)^k x_k$, therefore using Definition 3.

$$A_l(x') = (-1)^l A_l(x). \quad (5.18)$$

b . The k th member of the sequence \hat{x} is $\hat{x}_k = x_{N-1-k}$, therefore using Definition 3.

$$\begin{aligned} A_l(\hat{x}) &= \sum_{i=0}^{N-l-1} \hat{x}_{i+l} \hat{x}_i^* = \sum_{i=0}^{N-l-1} x_{N-l-i-1} x_{N-i-1}^* \\ &= \sum_{k=0}^{N-l-1} x_{k+l}^* x_k = (A_l(x))^* \end{aligned} \quad (5.19)$$

c Using (5.16) and (5.19) the statement is concluded. □

Theorem 5.2.7. *If $x \approx y$ then*

$$a. \ x|y \approx x| - y \qquad b. \ x \downarrow y \approx x \downarrow -y$$

Proof. The items are proved separately:

a . It is easy to see that

$$A_l(x|y) = A_l(x) + A_l(y) + \sum_{i=0}^{l-1} x_{N-1-i}^* y_{l-1-i}$$

and therefore

$$A_l(x|y) + A_l(x| - y) = 2(A_l(x) + A_l(y)) = 0.$$

b . if $l = 2k$ then

$$A_l(x \downarrow y) = A_{l/2}(x) + A_{l/2}(y) = 0. \quad (5.20)$$

Using (5.15) and since $x \approx -y$, $A_l(x \downarrow -y) = 0$. Therefore, $A_l(x \downarrow y) + A_l(x \downarrow -y) = 0$.

If $l = 2k + 1$ then

$$A_l(x \downarrow y) = \sum_{i=0}^{N-1-\frac{l-1}{2}} x_i^* y_{i+\frac{l-1}{2}}. \quad (5.21)$$

and therefor $A_l(x \downarrow y) + A_l(x \downarrow -y) = 0$.

□

By applying the transformations defined in Theorems 5.2.5 and 5.2.6 to the statements of Theorem 5.2.7 we can build a set of structures that create $2N$ -valued super Golay pairs from N -valued ones. Specifically, if $x \approx y$ each with size N , then the following sequences are super Golay pairs ($[j]$ means multiplying the sequence by j is optional) :

- 1) $\pm [j](x|y) \approx \pm [j](x| -y)$
- 2) $\pm [j](x \downarrow y) \approx \pm [j](x \downarrow -y)$
- 3) $\pm [j](x|y) \approx \pm [j](\hat{y}^*| -\hat{x}^*)$
- 4) $\pm [j](x \downarrow y) \approx \pm [j](\hat{y}^* \downarrow -\hat{x}^*)$
- 5) $\pm [j](x \downarrow -y) \approx \pm [j](\hat{y}^* \downarrow \hat{x}^*)$

However, because of the special structure of 16-QAM constellation, many of these constructions yield similar sequences. For example reversing the role of x and y will not yield new pairs. if the number of N -valued pairs is M , the first structure yields $4M$ of $2N$ -valued SGolay pairs and this is true for the second structure too. We have performed a simulation for the pairs with size 8 and obtained the same result. In general each pair with size N yields 32 pairs each with size $2N$. This is very similar to the Reed-Muller codes used for equal-power Golay sequences. Reed-Muller codes of degree $r + 1$ and length 2^{m+1} can be constructed from two 2^m -length Reed-Muller codes, one in degree r and one in degree $r + 1$. The exact statement of the theorem is [13]

Theorem 5.2.8.

$$RM(r + 1, m + 1) = \{f|f + g \quad \forall f \in RM(r + 1, m) \quad \text{and} \quad g \in RM(r, m)\}.$$

These similarities can lead us to a new definition for a modified Reed-Muller codes in the context of non-equal power constellations like 16-QAM.

5.2.3 Super-Golay 16QAM pairs from QPSK pairs

In this section we will look at an important question which is how to start the recursive construction. To this end, we will use the relation that Tarokh and Roessing [117] used. They represented a 16-QAM symbol by a weighted sum of two QPSK symbols. We will generalize their observation and find a construction that builds 16-QAM SGolay sequences from QPSK Golay sequences. Let's define QPSK symbols as the set

$$QPSK = \{exp[j(\frac{k\pi}{2} + \frac{\pi}{4})], |k \in \mathcal{Z}_{2^n}\}$$

Using Definitions 3 and 2, the following theorem can be proved easily,

Theorem 5.2.9. *For any two sequences x and y and any two complex numbers α and β*

$$A_l(\alpha x + \beta y) = |\alpha|^2 A_l(x) + |\beta|^2 A_l(y) \\ + \alpha\beta^* C_l(x, y) + \alpha^* \beta C_l(y, x).$$

Theorem 5.2.10. *If x and y are N -valued QPSK Golay complementary pairs, and α and β are two arbitrary complex numbers with $|\alpha| = |\beta|$, then each of the following pairs are 16-QAM super Golay sequences:*

$$1 . \quad c = \alpha(x + 2y) \quad \text{and} \quad t = \beta(-2x + y)$$

$$2 . \quad c = \alpha(x - 2y) \quad \text{and} \quad t = \beta(2x + y)$$

$$3 . \quad c = \alpha(x + 2jy) \quad \text{and} \quad t = \beta(2jx + y)$$

$$4 . \quad c = \alpha(x - 2jy) \quad \text{and} \quad t = \beta(-2jx + y)$$

Proof. We will prove the result for the third item. The rest can be proved similarly. Using

Theorem 5.2.9, for each nonzero l ,

$$\begin{aligned}
A_l(c) + A_l(t) &= |\alpha|^2 A_l(x) + 4|\alpha|^2 A_l(y) - 2j|\alpha|^2 C_l(x, y) + 2j|\alpha|^2 C_l(y, x) \\
&\quad + 4|\beta|^2 A_l(x) + |\beta|^2 A_l(y) + 2j|\beta|^2 C_l(x, y) - 2j|\beta|^2 C_l(y, x) \\
&= 5|\alpha|^2 (A_l(x) + A_l(y)) = 0
\end{aligned}$$

Therefore c and t are Golay complementary pairs. It is easy to see that each of these sequences are actually a 16-QAM sequence, when the average power of the constellation is $P_{av} = 5|\alpha|^2$. If we denote the Hermitian of x by x^H , and considering the fact that the power of both x and y is N , then

$$\begin{aligned}
P_c + P_t &= |\alpha|^2 (\|x + 2jy\|^2 + \|2jx + y\|^2) = |\alpha|^2 [(\|x\|^2 + 4\|y\|^2 + 2jx^H y - 2jy^H x) \\
&\quad + (\|y\|^2 + 4\|x\|^2 - 2jx^H y + 2jy^H x)] = 5|\alpha|^2 (\|x\|^2 + \|y\|^2) = 10N|\alpha|^2 = 2NP_{av}.
\end{aligned}$$

Therefore, by Definition 6., c and t are super Golay sequences. \square

Theorem 5.2.10 suggests a starting point for the proposed recursive construction. If we limit ourselves to an 16-QAM construction with $P_{av} = 10$, then α and β can be chosen from the set $\{\sqrt{2}, -\sqrt{2}, j\sqrt{2}, -j\sqrt{2}\}$, and therefor for each of the $2^{h(m+2)} \frac{m!}{2}$ Golay complementary pairs over \mathcal{Z}_{2^h} of length 2^m , there are 64 super Golay 16-QAM pairs. However, some of these pairs are repeated. As an example if $(c \sim t)$, then $(-c \sim -t)$ and therefore we do not need to multiply the two sequences in the first construction of Theorem 5.2.10 by $-\sqrt{2}$. Eliminating these repeated sequences, the number of 16-QAM super Golay pairs generated from each QPSK Golay pair is 16. Therefore we can build $2^{4+h(m+2)} \frac{m!}{2}$ distinct QAM super Golay pairs over \mathcal{Z}_{2^h} of length 2^m . For $m = 2$ and QPSK symbols ($h = 2$), this translates to 4096 pairs. Through exhaustive search, we have found that there are exactly 12032 super Golay pairs and Theorem 5.2.10 builds 4096 of them.

We were able to come up with some structures that build in average 32 new $2N$ -valued SGolay pairs from one N -valued SGolay pair. Therefore, starting from 4-valued codes, our construction is able to achieve at least

$$R = \frac{12 + 5 \log_2 N/4}{4N}. \quad (5.22)$$

code rate for N OFDM subchannels.

For 128 subchannels, this achieves 7.3% code rate. Although not still acceptable, this is about 11% improvement over Tarokh-Chong's work [124].

The achievable code rate is about 20% below the equal power Golay codes constructed by Jedwab and Davis [118]. However, because of using 16-QAM constellation, the information rate achieved by these structures is twice as the information rate achieved by Jedwab's construction, while the error correction properties of the code is maintained.

5.2.4 Super-Golay 64QAM pairs from QPSK pairs

The structure we proposed is for general super Golay codes, regardless of the constellation. However, we have focused on 16-QAM constellation for the sake of simulation. This scheme can be generalized to higher order QAM constellations, like 64-QAM which is used in IEEE WLAN standards like IEEE802.11a .

To find a construction method for low PAPR 64-QAM sequences we can use the concept of Golay sets. The Golay sets are defined as follows:

Definition 7. *The set of N -valued complex sequences $\{x^i \mid i = 1 \dots n\}$ is called a Golay set if*

$$\sum_{i=1}^n A_l(x^i) = 0 \quad \forall l \neq 0.$$

It is easy to see that if a Golay complementary set is taken from an equal-energy constellation, the PAPR of each sequence is bounded up to n . Paterson has shown that [112] the

2^{k+1} -size equal power Golay set can be represented by some certain cosets of $RM_{2^h}(1, m)$ in $RM_{2^h}(2, m)$.

The following lemma which can be proved easily using the Definitions 2 and 3 is a generalization of Lemma 5.2.9.

Lemma 5.2.11. *For any n sequences x^i and any n complex numbers α^i , $i = 1, \dots, n$*

$$A_l\left(\sum_{i=1}^n \alpha_i x^i\right) = \sum_{i=1}^n |\alpha_i|^2 A_l(x^i) + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n \alpha_i \alpha_j^* C_l(x^i, x^j).$$

Using this Lemma, the following theorem can be used to generate 64-QAM Super Golay sets (The sum of the powers is nNP_{av}).

Theorem 5.2.12. *If $\{x^i | i = 1 \dots 4\}$ comprises a N -valued QPSK Golay set, and $|\alpha_i|$ is constant for all $i = 1, \dots, 4$, then the following is a 64-QAM Super Golay set and therefore, the PAPR of each element is at most 4(6dB):*

$$\{\alpha_1(4x^1 + jx^2 - jx^3 - x^4), \alpha_2(4x^2 + jx^1 - x^3 - jx^4), \\ \alpha_3(4x^3 - jx^1 + x^2 - jx^4), \alpha_4(4x^4 + x^1 - jx^2 - jx^3)\}.$$

5.3 Cyclic Golay Sequences

Eq. (5.2) showed the instantaneous power of continuous OFDM signal. This power envelope is obtained after passing the OFDM channel symbols through a low-pass filter and up-converting the signal to the carrier frequency. If we consider coded OFDM with codebook \mathcal{C} , the Peak to Mean Envelope Power Ratio (PMEPR) of the code is defined as

$$PMEPR(\mathcal{C}) = \max_{0 \leq t \leq T} \frac{p_x(t)}{P_{av}},$$

and the PAPR is defined using the maximum of the real part of the OFDM signal in the time domain taken over all admissible codewords. PAPR measures the peak of the RF signal,

but PMEPR measures the peak of the baseband signal. The actual peak of OFDM time domain signal depends on the pulse shaping and the low-pass filter we use after the IFFT block. It is obvious that this peak is in general different from the samples of the OFDM signal at the multiples of $\frac{1}{N\Delta f}$. However, there is a direct relation between the peak of the continuous OFDM signal and the maximum of the discrete OFDM symbols sampled at the multiples of $\frac{1}{N\Delta f}$. The PAPR of discrete OFDM symbols is an indication of the PMEPR of the continuous OFDM signal. By some proper time shaping, we can consider the PAPR of the discrete sequence obtained after IFFT operation, as a measure of the PAPR of the OFDM signal. In the sequel, we define the PAPR of a codeword x to be

$$PAPR(x) = \frac{1}{P_x} \max_k \{p_x[k]\} = \frac{1}{P_x} \left\{ \sum_{i=0}^{N-1} x[i]x^*[u]e^{\frac{j2\pi(i-u)k}{N}} \right\}, \quad (5.23)$$

where $P_x = \|x\|^2$ is the power of the codeword x . Using the definition of the auto-Correlation of an N -valued complex sequence x as in Definition 3, the power of the k^{th} OFDM channel symbol can be restated as

$$p_x[k] = A_0(x) + \sum_{|l|=1}^{N-1} A_l(x)w_k^l, \quad (5.24)$$

where $w_k \triangleq e^{\frac{j2\pi k}{N}}$, is the N^{th} root of unity. Note that $A_0(x)$ is actually the same as P_x , the power of the code x , and by Parseval equation, this is the same as the average power of OFDM channel symbols. Therefore, the PAPR of the codeword x is:

$$PAPR(x) = \frac{1}{A_0(x)} \max_k \{p_x[k]\}. \quad (5.25)$$

For simplicity, we will denote " $k \bmod n$ " by " $k \% n$ " in the following definition and what comes hereafter.

Definition 8. *The Cyclic Auto-Correlation of an N -valued complex sequence x , with nonzero replacement l , is defined as*

$$CA_l(x) = \sum_{i=0}^{N-1} x^*[i]x[(i+l) \% N].$$

Using this definition, one can see that

$$\begin{aligned}
CA_l(x) &= \sum_{i=0}^{N-1} x^*[i]x[(i+l) \% N] \\
&= \sum_{i=0}^{N-l-1} x^*[i]x[i+l] + \sum_{i=N-l}^{N-1} x^*[i]x[i+l-N] \\
&= A_l(x) + \sum_{i'=0}^{l-1} x^*[i'+N-l]x[i'+N-l+l-N] \\
&= A_l(x) + \sum_{i'=0}^{N-(N-l)-1} x[i']x^*[i'+N-l] \\
&= A_l(x) + \sum_{i'=0}^{N-(N-l)-1} x[i']x^*[i'+N-l] \\
&== A_l(x) + A_{N-l}^*(x)
\end{aligned} \tag{5.26}$$

Definition 9. Two N -valued complex sequences x and y are called cyclic Golay complementary pairs if

$$CA_l(x) + CA_l(y) = 0 \quad \forall l \neq 0.$$

Each of the sequences x and y is called Cyclic Golay Complementary Sequence (CGCS).

Using (5.26), it is obvious that if two sequences are Golay pairs, they are cyclic Golay pairs, too. In other words, if \mathcal{G} is the set of Golay sequences, and \mathcal{GC} is the set of cyclic Golay sequences, then

$$\mathcal{G} \subseteq \mathcal{GC}. \tag{5.27}$$

Theorem 5.3.1. The PAPR achieved by cyclic Golay sequences is upper bounded by 3dB.

Proof. The power of OFDM channel symbols in (5.24), can be restated as

$$\begin{aligned}
p_x(k) &= A_0(x) + \sum_{|l|=1}^{N-1} A_l(x)w_k^l \\
&= A_0(x) + \sum_{l=1}^{N-1} (A_l(x)w_k^l + A_{-l}(x)w_k^{-l}) \\
&= A_0(x) + \sum_{l=1}^{N-1} (A_l(x)w_k^l + A_l^*(x)w_k^{-l}) \\
&= A_0(x) + \sum_{l=1}^{N-1} A_l(x)w_k^l + \sum_{l'=1}^{N-1} A_{N-l'}^*(x)w_k^{-(N-l')} \\
&= A_0(x) + \sum_{l=1}^{N-1} A_l(x)w_k^l + \sum_{l=1}^{N-1} A_{N-l}^*(x)w_k^l w_k^{-N} \\
&= A_0(x) + \sum_{l=1}^{N-1} (A_l(x) + A_{N-l}^*(x))w_k^l = CA_0(x) + \sum_{l=1}^{N-1} CA_l(x)w_k^l. \tag{5.28}
\end{aligned}$$

Note that, to obtain (5.28), we have used the equation $A_0(x) = CA_0(x)$, and the fact that $A_l^*(x) = A_{-l}(x)$ from Eq. (5.6). Considering the fact that the power of each OFDM channel symbol is always non-negative, combining (5.25) and (5.28) deduces that the PAPR of each CGCS is upper bounded by 3dB. \square

Theorem 5.3.1 and equation (5.27) state that the number of codewords achieving a PAPR at most equal to 3dB (in discrete domain), is more than just the number of Golay sequences. This is translated to higher coding rates.

If x is an N -sized sequence, we denote its cyclic l -shift by x^l ($0 \leq l \leq N - 1$). The k^{th} element of x^l is

$$x^l[k] = x[(k+l) \% N]. = \begin{cases} x[k+l] & \text{if } k \leq N-l-1, \\ x[k+l-N] & \text{if } k > N-l-1 \end{cases} \tag{5.29}$$

Lemma 5.3.2. *The property of being cyclic Golay is preserved under any cyclic l -shift of a sequence with size N .*

Proof.

$$\begin{aligned}
CA_u(x^l) &= \sum_{i=0}^{N-1} x^{l*}[i]x^l[(i+u) \% N] \\
&= \sum_{i=0}^{N-l-1} x^*[i+l]x[(i+l+u) \% N] + \sum_{i=N-l}^{N-1} x^*[i+l-N]x[(i+l+u) \% N] \\
&= \sum_{k=l}^{N-1} x^*[k]x[(k+u) \% N] + \sum_{k'=0}^{l-1} x^*[k']x[(k'+u+N) \% N] \\
&= \sum_{k=l}^{N-1} x^*[k]x[(k+u) \% N] + \sum_{k=0}^{l-1} x^*[k]x[(k+u) \% N] \\
&= \sum_{k=0}^{N-1} x^*[k]x[(k+u) \% N] = CA_u(x). \tag{5.30}
\end{aligned}$$

Therefore, if two sequences are cyclic Golay sequences, their shifted version by any replacement l are cyclic Golay, too. \square

Theorem 5.3.1 can be generalized for non-Golay sequences in the following way:

Theorem 5.3.3. *The PAPR achieved by any cyclic shifted versions of a sequence x is the same as the PAPR achieved by the sequence x itself.*

Proof. $A_0(x)$ is the power of the sequence x , which is preserved under any cyclic shift. Therefore, by (5.23) and (5.30)

$$p_{x^l}[k] = A_0(x^l) + \sum_{i=1}^{N-1} CA_i(x^l)w_k^i = A_0(x) + \sum_{i=1}^{N-1} CA_i(x)w_k^i = p_x[k]. \tag{5.31}$$

So, the PAPR of the sequence x^l is the same as the PAPR of x . \square

In Section 5.2.2, we listed several transformation that when performed on Golay pairs, the resultant pairs are still GCP. These transformation are reflection with respect to the origin, with respect to both axes, with respect to the bisectors of all regions, and the rotation of one or both sequences. The same was true for the concatenation and interleaving of Golay

pairs, reversing each sequence in Golay pairs, and alternatively multiplying the elements of each sequence by -1 . Considering equation (5.26), we can deduce that the property of being cyclic Golay pairs is invariant under any of these transformations.

5.3.1 Construction of Cyclic Golay Codes

If we start from one of the their Golay sequences with size 2^m over \mathcal{Z}_{2^h} , and make a cyclic shift, the resultant sequence is cyclic Golay. Our construction can create 2^m cyclic Golay sequences out of each Golay sequence by l -shifting the original sequence with $0 \leq l \leq 2^m - 1$. However, some of these newly generated sequences are also part of original Golay sequences that can be created by different values of c_k 's and c in (5.13). As a result, we need to carefully develop a structure for constructing the cyclic Golay sequences. In what follows, we will design a framework for obtaining the cyclic shifts of a sequence, presented by Boolean functions. To this end, we start from the field \mathcal{Z}_2 with the addition defined modula 2.

Cyclic Shift of Binary Codes

The k^{th} element of each basis sequence, $\{x_n \mid n = 1 \dots m\}$, is given by:

$$x_n[k] = \begin{cases} 0 & \text{if } k \% 2^n < 2^{n-1}, \\ 1 & \text{if } k \% 2^n \geq 2^{n-1}. \end{cases} \quad (5.32)$$

By considering the relation $(t \% ab) \% b = t \% b$, we can represent the cyclic l -shift of x_n ($1 \leq n \leq m$) as:

$$x_n^l[k] = \begin{cases} 0 & \text{if } [(k+l) \% 2^m] \% 2^n < 2^{n-1}, \\ 1 & \text{if } [(k+l) \% 2^m] \% 2^n \geq 2^{n-1} \end{cases} = \begin{cases} 0 & \text{if } (k+l) \% 2^n < 2^{n-1}, \\ 1 & \text{if } (k+l) \% 2^n \geq 2^{n-1}. \end{cases} \quad (5.33)$$

For example, with $m = 3$, we have

$$\begin{aligned}
 x_1 &= [0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1] \\
 x_2 &= [0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1] \\
 x_3 &= [0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1] \\
 x_2^3 &= [1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1].
 \end{aligned} \tag{5.34}$$

Moreover, by xy , we mean the Hadamard product of two codewords x and y . The Hadamard product of two $p \times q$ matrices A and B , is another $p \times q$ matrix C , whose ij^{th} element is $c_{ij} = a_{ij}b_{ij}$. The second order monomials in the definition of Reed-Muller codes, and also in Theorem 5.2.1, are equivalent to the Hadamard product of two basis codeword. In the sequel, all additions are performed modula 2 and all codeword products are Hadamard product, unless otherwise stated.

Lemma 5.3.4. *The cyclic 1-shift of a basis codeword x_n , $1 \leq n \leq m$ is*

$$x_n^1 = x_n + \prod_{i=0}^{n-1} x_i,$$

where $x_0 = \mathbf{1}$ is an all one codeword.

Proof. Examining (5.32), it is easy to see that

$$\left(\prod_{i=0}^{n-1} x_i \right) [k] = \begin{cases} 1 & \text{if } k \% 2^{n-1} = -1, \\ 0 & \text{if } k \% 2^{n-1} \neq -1. \end{cases} \tag{5.35}$$

If $k \% 2^{n-1} = -1$, then $k = q(2^{n-1}) - 1$. Depending on whether q is even or odd, we have

$k \% 2^n = 2^n - 1$ or $k \% 2^n = 2^{n-1} - 1$. Therefore,

$$\begin{aligned} \left(x_n + \prod_{i=0}^{n-1} x_i \right) [k] &= \begin{cases} 0 & \text{if } k \% 2^n < 2^{n-1} - 1, \\ 0 & \text{if } k \% 2^n = 2^{n-1} - 1, \\ 1 & \text{if } 2^{n-1} \leq (k \% 2^n) < 2^n - 1, \\ 1 & \text{if } k \% 2^n = 2^n - 1 \end{cases} + \begin{cases} 0 & \text{if } k \% 2^n < 2^{n-1} - 1, \\ 1 & \text{if } k \% 2^n = 2^{n-1} - 1, \\ 0 & \text{if } 2^{n-1} \leq (k \% 2^n) < 2^n - 1, \\ 1 & \text{if } k \% 2^n = 2^n - 1 \end{cases} \\ &= \begin{cases} 0 & \text{if } k \% 2^n < 2^{n-1} - 1 \\ 1 & \text{if } k \% 2^n = 2^{n-1} - 1 \\ 1 & \text{if } 2^{n-1} \leq (k \% 2^n) < 2^n - 1 \\ 0 & \text{if } k \% 2^n = 2^n - 1 \end{cases} = \begin{cases} 0 & \text{if } (k+1) \% 2^n < 2^{n-1}, \\ 1 & \text{if } (k+1) \% 2^n \geq 2^{n-1} \end{cases} = x_n^1. \end{aligned}$$

□

Equation (5.35) can be generalized for $s > 0$ as:

$$\left(\prod_{i=s}^{n-1} x_i \right) [k] = \begin{cases} 1 & \text{if } -2^s \leq k \% 2^{n-1} < 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.36)$$

Using (5.36) and the same argument as in Lemma 5.3.4, we can generalize Lemma 5.3.4 and obtain the following result:

Lemma 5.3.5. *For $n > 1$ and $0 \leq k < n - 1$, the cyclic 2^k -shift of a basis codeword x_n , is*

$$x_n^{2^k} = x_n + \prod_{i=k+1}^{n-1} x_i.$$

Moreover, $x_n^{2^{n-1}} = 1 + x_n$, and $x_n^{2^k} = x_n$ for $k \geq n \geq 0$.

Theorem 5.3.6. *The following relations hold:*

- For $2^{n-1} \leq l \% 2^n$, the l -shift of x_n is $x_n^l = 1 + x_n^{l \% 2^{n-1}}$.

- For $l \geq 2^n$, the l -shift of x_n is $x_n^l = x_n^{l \% 2^n}$.
- If $l = 2^a + 2^b$ with $n - 1 > a > b$ we have

$$x_n^l = x_n + \left(x_{a+1} + \prod_{i=b+1}^{a+1} x_i + \prod_{i=b+1}^a x_i \right) y_n,$$

where

$$y_n = \begin{cases} \prod_{i=a+2}^{n-1} x_i & \text{if } a < n - 2, \\ 1 & \text{if } a = n - 2. \end{cases} \quad (5.37)$$

Proof. we will prove each item separately,

- If $2^{n-1} \leq l \% 2^n$, then $l \% 2^n = 2^{n-1} + l \% 2^{n-1}$. Therefore,

$$(k + l) \% 2^n = (k \% 2^n + 2^{n-1} + l \% 2^{n-1}) \% 2^n = [(k + l \% 2^{n-1}) \% 2^n + 2^{n-1}] \% 2^n.$$

As a result, if $(k + l \% 2^{n-1}) \% 2^n < 2^{n-1}$, then $(k + l) \% 2^n \geq 2^{n-1}$, and vice versa.

Using (5.33), the relation $x_n^l = 1 + x_n^{l \% 2^{n-1}}$ is obtained.

- It is easy to see that $(k + l) \% 2^n = (k + l \% 2^n) \% 2^n$. Using (5.33), the relation $x_n^l = x_n^{l \% 2^n}$ is obtained for $2^n \leq l$.
- If $l = 2^a + 2^b$ with $a > b$, we have $x_n^l = \left(x_n^{2^b} \right)^{2^a}$. Then, by virtue of Lemma 5.3.5, we have

$$\begin{aligned} x_n^l &= \left(x_n + \prod_{i=b+1}^{n-1} x_i \right)^{2^a} \\ &= x_n + \prod_{i=a+1}^{n-1} x_i + x_{b+1}^{2^a} x_{b+2}^{2^a} \cdots x_a^{2^a} x_{a+1}^{2^a} \cdots x_s^{2^a} \cdots x_{n-1}^{2^a} \\ &= x_n + \prod_{i=a+1}^{n-1} x_i + x_{b+1} x_{b+2} \cdots x_a (1 + x_{a+1}) \cdots \left(x_s + \prod_{i=a+1}^{s-1} x_i \right) \cdots \left(x_{n-1} + \prod_{i=a+1}^{n-2} x_i \right) \end{aligned}$$

For $a = n - 2$, we have

$$x_n^l = x_n + x_{a+1} + (1 + x_{a+1}) \prod_{i=b+1}^a x_i = x_n + x_{a+1} + \prod_{i=b+1}^a x_i + \prod_{i=b+1}^{a+1} x_i.$$

If $a > n - 2$, for each $s \geq a + 2$, we have

$$\left(x_s + \prod_{j=a+1}^{s-1} x_j \right) (1 + x_{a+1}) = x_s (1 + x_{a+1})$$

. Therefore,

$$\begin{aligned} x_n^l &= x_n + \prod_{i=a+1}^{n-1} x_i + \left(\prod_{i=b+1}^a x_i \right) \left(\prod_{i=a+2}^{n-1} x_i \right) (1 + x_{a+1}) \\ &= x_n + \left(x_{a+1} + \prod_{i=b+1}^a x_i + \prod_{i=b+1}^{a+1} x_i \right) \prod_{i=a+2}^{n-1} x_i. \end{aligned}$$

Note that in the last equation we have used the fact $\left(\prod_{j=l_1+1}^{i-1} x_j \right) (1 + x_{l_1-1}) = 0$ for each i .

□

As an application of Lemma 5.3.6, for $l = 3 = 2^1 + 2^0$, we have

$$x_n^3 = x_n + \left(\prod_{i=3}^{n-1} x_i \right) (x_1 + x_2 + x_1 x_2).$$

The following Corollary is a direct consequence of Theorem 5.3.6.

Corollary 5.3.7. *If $l = 2^a + b$, the Boolean function for x_n^l can be found by replacing each $x_{n'}$ in x_n^b , by $x_{n'} + \prod_{i=a+1}^{n'-1} x_i$.*

All of the cyclic shifts we have discussed up to this point, have been left shifts. Now, we turn our attention to right shifts. The right cyclic shift of x_n is defined as:

$$x_n^{-l}[k] = \begin{cases} 0 & \text{if } (k-l) \% 2^n < 2^{n-1}, \\ 1 & \text{if } (k-l) \% 2^n \geq 2^{n-1}. \end{cases} \quad (5.38)$$

Examining (5.32), one can see that

$$\left(\prod_{i=s}^{n-1} (1+x_i) \right) [k] = \begin{cases} 1 & \text{if } 0 \leq k \% 2^{n-1} < 2^s - 1, \\ 0 & \text{otherwise.} \end{cases} \quad (5.39)$$

Using the same arguments as in Lemma 5.3.4 and 5.3.5, one can further prove the following lemma:

Lemma 5.3.8. *For $n \geq 1$ and $0 \leq k < n-1$, the right cyclic 2^k -shift of a basis codeword x_n , is*

$$x_n^{-2^k} = x_n + \prod_{i=k+1}^{n-1} (1+x_i).$$

Moreover, $x_n^{-2^{n-1}} = 1+x_n$ and $x_n^{-2^k} = x_n$ for $k \geq n \geq 0$.

Finally, the following Lemma relates the left cyclic $(2^{n-1}-l)$ -shift of x_n to the right cyclic l -shift of the basis codewords.

Lemma 5.3.9. *For $1 \leq l < 2^{n-1}$, the left cyclic $(2^{n-1}-l)$ -shift of x_n , is*

$$x_n^{2^{n-1}-l} = 1+x_n^{-l}.$$

Proof. It is easy to see that $(k+2^{n-1}-l) \% 2^n = [(k-l) \% 2^n + 2^{n-1}] \% 2^n$. So, if $(k+2^{n-1}-l) \% 2^n < 2^{n-1}$, then $(k-l) \% 2^n \geq 2^{n-1}$ and vice versa. This means $x_n^{2^{n-1}-l} = 1+x_n^{-l}$. \square

Cyclic Shift of non-Binary Codes

We have built a framework to find the Boolean function representation of a large set of cyclic shifts of basis codewords. This framework defines the cyclic shift of any codeword defined over \mathcal{Z}_2 by means of Boolean functions. However, if the code is defined over \mathcal{Z}_{2^h} , there is an ambiguity over modula-2 addition and modula- 2^h additions. This ambiguity can be avoided by using the following relation:

$$c[(x + y) \% 2] = (cx + cy - 2cxy) \% 2^h, \quad (5.40)$$

where c is a constant number defined over \mathcal{Z}_{2^h} and x and y are two binary codewords. This equation can be easily verified by examining 3 cases for every element of x and y . The cases are when the k^{th} components of x and y are both 0, both are 1, or one is 0 and the other one is 1. By defining $S_m = \{1, 2, \dots, m\}$, induction can be used to generalize equation (5.40) for m sequences f_i ($i = 1, \dots, m$) as:

$$c \left[\left(\sum_{i=1}^m f_i \right) \% 2 \right] = \sum_{i=1}^m c f_i - \left(\sum_{\substack{(i,j) \in S_m^2 \\ i \neq j}} 2c f_i f_j \right) \cdots + \begin{cases} (-2)^{m-1} c \prod_{i=1}^m f_i & \text{if } m \leq h, \\ \sum_{\substack{(i_1, i_2, \dots, i_h) \in S_m^h \\ i_1 \neq i_2 \neq \dots \neq i_h}} (-2)^{h-1} c \prod_{l=1}^h f_{i_l} & \text{otherwise,} \end{cases} \quad (5.41)$$

where the additions and subtraction in the right hand side of (5.41) are in modula 2^h , and the codewords f_i represent any binary codeword, not only the basis codewords.

As an example of this framework, we will construct the cyclic Golay code with size 8 over \mathcal{Z}_4 . The coset representations of $RM_4(1, 3)$ that belongs to $RM_4(2, 3)$, coming from Theorem 5.2.1, are $2(x_1x_2 + x_2x_3)$, $2(x_1x_2 + x_1x_3)$, and $2(x_1x_3 + x_2x_3)$. It is worthwhile to mention that the additions in the coset representations in (5.13) are in modula- 2^h . Table 5.1 shows the Golay sequences that are repeated under cyclic shifts. The sequences are categorized based

<i>Shifts/Coset Rep.</i>	$x_1x_2 + x_2x_3$	$x_1x_3 + x_2x_3$	$x_1x_3 + x_1x_2$
$l = 1$	$c_3 = 0 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$ $c_3 = 2 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$		$c_3 = 0 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$ $c_3 = 2 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$
$l = 2$	$c_3 \% 2 = 0$		$c_3 \% 2 = 1$
$l = 3$	$c_3 = 0 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$ $c_3 = 2 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$		$c_3 = 0 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$ $c_3 = 2 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$
$l = 4$	all		
$l = 5$	$c_3 = 0 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$ $c_3 = 2 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$		$c_3 = 0 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$ $c_3 = 2 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$
$l = 6$	$c_3 \% 2 = 0$		$c_3 \% 2 = 1$
$l = 7$	$c_3 = 0 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$ $c_3 = 2 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$		$c_3 = 0 \ \& \ (c_2 = 1 \ \text{or} \ c_2 = 3)$ $c_3 = 2 \ \& \ (c_2 = 0 \ \text{or} \ c_2 = 2)$
# of non-Golay sequences	1024	1024	1024

Table 5.1: List of repeated Golay sequences under cyclic shifts for $m = 3$ and $h = 2$

on the coset representative and the coefficient values from \mathcal{Z}_4 . These values are obtained by applying the above mentioned framework to each basis codeword. For example, when the coset representative is $2(x_1x_2 + x_2x_3)$, the 3^{rd} cyclic shift of the Golay sequence represented in (5.13) is:

$$\begin{aligned}
& c_1x_1^3 + c_2x_2^3 + c_3x_3^3 + c + 2(x_1^3x_2^3 + x_2^3x_3^3) = c_1(1 + x_1) + c_2(1 + x_1 + x_2) + \\
& c_3(x_3 + x_1 + x_2 + x_1x_2) + c + 2(1 + x_1 + x_2 + x_3 + x_1x_3 + x_2x_3) = \\
& c_1 + c_2 + c + 2 + (2 + c_3 - c_1 - c_2)x_1 + (2 + c_3 - c_2)x_2 + (2 + c_3)x_3 + \\
& (-c_3 - 2c_2)x_1x_2 + (2 - 2c_3)x_1x_3 + (2 - 2c_3)x_2x_3 - 2c_3x_1x_2x_3.
\end{aligned}$$

The requirements for this expression to be a Golay sequence (and therefore the 3^{rd} cyclic shift does not create a new codeword) are:

- $2c_3 = 0 \ \text{mod} \ 4$ to delete the term $x_1x_2x_3$,

- One of the three terms " $-c_3 - 2c_2$ " or " $2 - 2c_3$ " or " $2 - 2c_3$ " is equal to zero to create a valid second order coset.
- The other two nonzero terms are equal to 2.

These requirements are translated into $c_3 \% 2 = 0$. Moreover, if $c_3 = 0$ then $c_2 \% 2 = 0$, and if $c_3 = 2$ then $c_2 \% 2 = 1$. Therefore, out of 256 Golay sequences generated by this coset, the 3^{rd} cyclic shift of only 64 sequences are repetitions of Golay codes. So, the 3^{rd} cyclic shift of this coset generates 192 cyclic Golay sequences that are not Golay. The last row of Table 5.1 shows the total number of cyclic Golay sequences created by each cost representative that are not Golay. The total number of new sequences created by $2(x_1x_2 + x_2x_3)$ is 1024. Considering all 3 cosets, the cyclic shifts generate 3072 new sequences. However, some of the newly generated sequences by a coset representative can be created by some other cosets and different values of c_k 's. For example, when $c_3 = 3$, all of the sequences generated by cyclic shifts of Golay codewords are similar to the ones created by $c_3 = 1$, with different values for c_1 , c_2 , and c , and different coset representatives. To find this out, we need to create a table containing all of the coefficients for each coset representatives, and each shift, for different values of c_3 . By deleting the similar columns, we can find the non-repeated cyclic Golay sequences. We have performed such inspection for $m = 3$ and $h = 2$. The result shows that 1024 of these cyclic Golay sequences are non-repeated and can be generated from Golay sequences, using the procedure presented in Table 5.2. This is the procedure that we use at the encoder to encode these sequences.

<i>Coeff. / Coset</i>	$x_1x_2 + x_2x_3$	$x_1x_3 + x_2x_3$	$x_1x_3 + x_1x_2$
$c_3 = 0; c_2 \% 2 = 0$	None	$l = 1, 2, 3$	$l = 2, 3$
$c_3 = 0; c_2 \% 2 = 1$	None	$l = 2$	$l = 1, 2$
$c_3 = 1$	$l = 1, 2, 3, 5, 6, 7$	None	$l = 1, 3, 5, 7$
$c_3 = 2; c_2 \% 2 = 0$	None	$l = 2$	None
$c_3 = 2; c_2 \% 2 = 1$	None	$l = 1, 2, 3$	None
$c_3 = 3$	None		

Table 5.2: List of non-repeated cyclic shifts on Golay sequences for $m = 3$ and $h = 2$.

The rate of size 8 Golay sequences over \mathcal{Z}_4 is

$$\frac{\log_2 \left(\frac{3!}{2} 2^{2(3+1)} \right)}{2 \times 2^3} = 0.599,$$

while the rate of size 8 cyclic Golay sequences over \mathcal{Z}_4 is

$$\frac{\log_2 \left(1024 + \frac{3!}{2} 2^{2(3+1)} \right)}{2 \times 2^3} = 0.690.$$

As the size of the codewords increases, more basis vectors are considered and therefore, the cyclic shifts of (5.13) create higher order monomials. As a result, the number of newly generated cyclic Golay codewords is increased. The same thing happens as the dimension of the field from which we pick the coefficients c_i is increased. Therefore, as we move toward higher dimension fields and higher size codes, the rate increase is more significant. As an example for $m = 4$, if we shift each of the codewords from the coset represented by $x_1x_2 + x_2x_3 + x_3x_4$ one position, none of the newly generated codewords are repeated Golay sequence. This is because the Boolean function representation of these codewords not only includes some of the 3^{rd} order monomials, but also the second order coset is not in the form shown in (5.13). Therefore, none of these cyclic shifts can be obtained from (5.13). However,

the number of newly generated cyclic Golay sequences from (5.13) cannot be more than the number of Golay sequences times the size of each codeword (2^m). It is worthwhile to mention that the total number of size 2^m cyclic Golay sequences over \mathcal{Z}_{2^h} , using exhaustive search, is greater than the number we just found, using (5.13) and our framework.

The cyclic Golay code we generated, is clearly non-linear. Even if we subtract the coset representatives from (5.13) to obtain a linear $RM_{2^h}(1, m)$ code and then perform the cyclic shifts, the resultant code is not linear. Therefore, this is not a linear cyclic code and we cannot define a minimal generator polynomial for generating this code. It is known that a linear cyclic code created by polynomials cannot be of size 2^m and our code is of size 2^m . Massey [128] has introduced cyclic Reed-Muller codes by puncturing the first column of the generator matrix of $RM(r, m)$ and reordering the first order rows to create m -sequences. The size of the codewords in this case is $2^m - 1$. However, even with those special cosets defined in (5.13), cyclic Reed-Muller code is not Golay anymore and therefore, does not have the low PAPR property.

The major drawback of the cyclic Golay code is their low Hamming and Lee distances of the code. The Hamming weight of a codeword x of size m over \mathcal{Z}_{2^h} , denoted as $wt_H^h(x)$, is equal to the number of nonzero components of x , while the Lee weight of x over \mathcal{Z}_{2^h} , denoted by $wt_L^h(x)$, is defined to be $\sum_{j=0}^{m-1} \min\{x_j, 2^h - x_j\}$ (The sum is performed regularly, not in modula 2^h). The Hamming distance of two codewords x and y is defined as $d_H^h(x, y) = wt_H^h(x - y \text{ mod } 2^h)$, and the Lee distance is $d_L^h(x, y) = wt_L^h(x - y \text{ mod } 2^h)$. The Hamming distance measures the number of positions in which x and y are different, whereas the Lee

distance takes into account the magnitude of the difference over \mathcal{Z}_{2^h} at each position. These two coincide in the binary case, $h = 1$. For example, the Hamming distance between the pair of codewords $(5, 7, 0, 1)$ and $(3, 7, 7, 6)$ over \mathcal{Z}_{2^3} is 3, while their Lee distance is $2 + 0 + 1 + 3 = 6$. The minimum Hamming distance of a code \mathcal{C} , which is taken over all distinct pair of codewords in \mathcal{Z}_{2^h} is defined to be $d_H^h(\mathcal{C}) = \min_{\substack{x, y \in \mathcal{C} \\ x \neq y}} \{d_H^h(x, y)\}$, and similarly, the minimum Lee distance of the code is $d_L^h(\mathcal{C}) = \min_{\substack{x, y \in \mathcal{C} \\ x \neq y}} \{d_L^h(x, y)\}$.

The minimum Hamming or Lee distance of a code is a measure of its error correction capability; that is if the minimum Hamming or Lee distance is d , then we can always correct error of (Hamming or Lee) weights less than $d/2$. If the transmission channel renders all $H - 1$ possible error for a given codeword positions equally likely, then the traditional Hamming distance metric is suitable. However, if error involve a transition between adjacent values in \mathcal{Z}_{2^h} are much more likely than other error in a given position, then Lee distance metric is more appropriate [129]. Both measures are appropriate metrics for evaluating the error correction capability for OFDM transmission. It is proved in [13, 118] that the Hamming and Lee distance of the $RM_{2^h}(r, m)$ are both 2^{m-r} . Therefore, the Hamming and Lee distance of the code defined in (5.13) are both 2^{m-2} .

The cyclic Golay code defined in this section, is a subset of $RM_{2^h}(m, m)$, and therefore, in general has a low distance. However, by reducing the coding rate, we can increase the Hamming and Lee distances, while maintaining the same upper bound for the PAPR. This trade-off can be explained using an example. Take $m = 4$, and $h = 2$. If we start from (5.13) with the coset representation $x_1x_2 + x_2x_3 + x_3x_4$, the Boolean function representation of the

cyclic 1-shift of each codeword is:

$$c + c_1 + (c_2 - c_1)x_1 + (2 + c_2)x_2 + c_3x_3 + c_4x_4 + (-2c_2 + c_3 + 2)x_1x_2 + 2x_3(x_1 + x_2 + x_4) + (-2c_3 + c_4)x_1x_2x_3 + 2x_1x_2x_4 - 2c_4x_1x_2x_3x_4.$$

It is observed from this expression that if we avoid the cyclic 1-shift of the codewords for $c_4 \bmod 2 = 1$, the resulting code is in $RM_4(3, 4)$, instead of $RM_4(4, 4)$. Moreover, it can be easily checked that by applying some restriction on c_3 , all of the codewords can be considered either as a subset of $RM_4(2, 4)$ or a coset of $RM_4(2, 4)$ with coset representation $2x_1x_2x_4$.

5.3.2 Maximum-likelihood Decoding of $RM_{2^h}(r, m)$

We assume that the cyclic Golay code is constructed to be a subset of $RM_{2^h}(r, m)$. Our goal in this section is to find a decoding scheme for these codes. To this end, we provide two decoding algorithms for $RM_{2^h}(r, m)$. The first one is a generalization of Reed algorithm which is stated in [13] for binary Reed-Muller codes. We will restate this algorithm using the concept of Karnaugh Map or K-Map [127] for generalized non-binary Reed-Muller codes. In this chapter, we will focus particularly on $RM_{2^h}(3, 4)$. The decoding steps for other sizes and orders can be easily obtained. The second algorithm is a recursive algorithm based upon the first scheme and other existing maximum likelihood decoding schemes for the first order generalized Reed-Muller codes. In the Subsection 5.3.2, we will compare the complexities of these two algorithms.

		00	01	$\overbrace{11 \quad 10}^{x_2}$			
00	0	1	3	2			
01	4	5	7	6			
x_4 { 11	12	13	15	14		} x_3	
10	8	9	11	10			
		x_1					

Table 5.3: K-Map for 4 variables

		00	01	$\overbrace{11 \quad 10}^{x_2}$							
00	0	1	3	2			00	16	17	19	18
01	4	5	7	6			01	20	21	23	22
x_4 { 11	12	13	15	14		} x_3	11	28	29	31	30
10	8	9	11	10			10	24	25	27	26
		x_1						x_5			

Table 5.4: K-Map for 5 variables

Maximum-likelihood Decoding of $RM_{2^h}(r, m)$ using Karnaugh Maps

K-Map is a set of squares, each representing a number in the range $\{0, 1, \dots, 2^m - 1\}$, with the property that the binary representation of every two neighboring squares are different in exactly one bit position. The K-Maps for $m = 4$ and $m = 5$ are depicted in Tables 5.3 and 5.4. As it is apparent from these tables, two squares are neighbors if they are either adjacent, or circularly adjacent (like 0 and 2 in Table 5.3), or they are at the same position in different blocks (like 15 and 31 in Table 5.4). Let's define $S_k = \{1, 2, \dots, k\}$. The transmitted sequence in $RM_{2^h}(3, 4)$ can be written as:

$$f(x_1, x_2, x_3, x_4) = \left[\alpha + \sum_{i \in S_4} \alpha_i x_i + \sum_{\substack{(i,j) \in S_4^2 \\ i \neq j}} \alpha_{ij} x_i x_j + \sum_{\substack{(i,j,k) \in S_4^3 \\ i \neq j \neq k}} \alpha_{ijk} x_i x_j x_k \right] \text{ mod } 2^h, \quad (5.42)$$

where α 's represent the Reed-Muller coefficients. Note that, decoding of Generalized Reed-Muller codes is equivalent to determining the coefficients of the monomials. Using the expansion in (5.42), we can find the components of the transmitted vector. For example, by setting all of x_i 's equal to 1, we can obtain f_{15} , the 15th component of f . In the following we show f_{15} , f_{12} , and f_1 :

$$\begin{aligned}
f_{15} &= \left[\alpha + \sum_{i \in S_4} \alpha_i + \sum_{\substack{(i,j) \in S_4^2 \\ i \neq j}} \alpha_{ij} + \sum_{\substack{(i,j,k) \in S_4^3 \\ i \neq j \neq k}} \alpha_{ijk} \right] \text{ mod } 2^h, \\
f_{14} &= \left[f_{15} - \alpha_1 - \sum_{i \in S_4 - \{1\}} \alpha_{1i} - \sum_{\substack{(i,j) \in (S_4 - \{1\})^2 \\ i \neq j}} \alpha_{1ij} \right] \text{ mod } 2^h \\
f_{13} &= [c_{15} - \alpha_2 - \sum_{i \in I_4 - \{2\}} \alpha_{2i} - \sum_{(i,j) \in \{S_4 - \{2\}\}^2} \alpha_{2ij}] \text{ mod } 2^h \\
f_{12} &= \left[f_{15} - \alpha_1 - \alpha_2 - \alpha_{12} - \sum_{i \in S_4 - \{1,2\}} (\alpha_{1i} + \alpha_{2i} + \alpha_{12i}) - \sum_{\substack{(i,j) \in (S_4 - \{1,2\})^2 \\ i \neq j}} (\alpha_{1ij} + \alpha_{2ij}) \right] \text{ mod } 2^h, \\
f_1 &= [\alpha + \alpha_1] \text{ mod } 2^h.
\end{aligned}$$

If we rewrite these expressions for all of the components of the encoded codeword, both α_{123}^1 and α_{123}^2 , shown in (5.43), could be considered as the estimates of α_{123} :

$$\begin{aligned}
\alpha_{123}^1 &= [f_{15} - (f_{11} + f_{13} + f_{14}) + (f_9 + f_{10} + f_{12}) - f_8] \text{ mod } 2^h \\
\alpha_{123}^2 &= [f_7 - (f_3 + f_5 + f_6) + (f_1 + f_2 + f_4) - f_0] \text{ mod } 2^h.
\end{aligned} \tag{5.43}$$

By examining Table 5.3, the estimation of α_{123} is obtained from the K-Map in the following manner:

- Specify the block of squares containing x_1 , x_2 , and x_3 (squares 15 and 7). Write the corresponding terms with a positive sign in different equations.
- For each of the squares found in the previous step (e.g. f_{15}), write the adjacent squares with negative signs (f_{11} , f_{13} , and f_{14}). Exclude the squares you found in the previous step.
- Continue step 2, by alternating the negative and positive signs, until all of the squares are covered.

If α_{123}^1 and α_{123}^2 are the same, we can estimate the common value as α_{123} . Note that, the Hamming distance of $RM_{2^h}(r, m)$ is 2^{m-r} [13]. For $RM_{2^h}(3, 4)$, this distance is 2 and therefore, we can only detect one symbol error. This fact is apparent from our algorithm, too. Any one symbol error might cause the values for α_{123} , found from these two equations, to be different. Therefore, we can detect one symbol error, but cannot correct it. This is true for the Lee distance, too. The Lee distance of $RM_{2^h}(r, m)$ is also 2^{m-r} . Other 3^{rd} order coefficients like α_{124} , α_{134} , and α_{234} can be found in the same way. For the lower order coefficients, we first need to subtract the higher order monomials from the original codeword. If we denote the result of subtracting the third order monomials from the transmitted codeword by f^1 , we have

$$f^1 = \left[f - \sum_{\substack{(i,j,k) \in S_4^3 \\ i \neq j \neq k}} \alpha_{ijk} x_i x_j x_k \right] \text{ mod } 2^h. \quad (5.44)$$

Then, the following set of equations can be obtained from the K-Map for α_{24} .

$$\begin{aligned}
\hat{\alpha}_{24} &= f_{15}^1 - (f_7^1 + f_{13}^1) + f_5^1 = f_{14}^1 - (f_6^1 + f_{12}^1) + f_4^1 \\
&= f_{11}^1 - (f_3^1 + f_9^1) + f_1^1 = f_{10}^1 - (f_2^1 + f_8^1) + f_0^1 \pmod{2^h}.
\end{aligned}$$

Again, α_{24} can be found using a majority voting. Likewise, we can obtain the hard-decision estimation of other second order coefficients. By subtracting the second order monomials from f^1 and obtaining f^2 , we can write a set of equations for decoding the first order monomials, like:

$$\hat{\alpha}_1 = f_1^2 - f_0^2 = f_3^2 - f_2^2 = f_5^2 - f_4^2 = f_7^2 - f_6^2 = f_9^2 - f_8^2 = f_{11}^2 - f_{10}^2 = f_{13}^2 - f_{12}^2 = f_{15}^2 - f_{14}^2 \pmod{2^h}$$

The same set of equations can be obtained for other first order coefficients. Finally, the constant term coefficient is the result of majority voting over all of the remaining components. It can be verified that the majority logic scheme, shown above using K-Maps, is an 2^{m-r} Hamming and Lee distance decoder.

To perform the soft-decision decoding in the complex domain, if the received codeword is $\{y_0, y_1, \dots, y_{15}\}$, the maximum-likelihood estimation of α_{123} is a number $\hat{\alpha}_{123} \in \mathcal{Z}_{2^h}$, that maximizes the real part of the following expression:

$$w^{-\hat{\alpha}_{123}} [y_{15}(y_{11}y_{13}y_{14})^*(y_9y_{10}y_{12})y_8^* + y_7(y_3y_5y_6)^*(y_1y_2y_4)y_0^*]. \quad (5.45)$$

Note that, the conjugate of $y_{11}y_{13}y_{14}$ in (5.45) performed in complex domain, corresponds to the subtraction in (5.43) performed in code domain. Other coefficients can be found likewise.

Recursive Maximum-likelihood Decoding of $RM_{2^h}(r, m)$

The K-Map scheme is appropriate, if m is not very large. So, we introduce a recursive scheme to decompose the decoding of $RM_{2^h}(r, m)$ into decoding of lower order or lower size Reed-Muller codes. To this end, we first restate the maximum-likelihood soft-decision decoding scheme for $RM_{2^h}(1, m)$ presented in [125]. If the information symbol to be transmitted is shown by a vector $[u_0 \ u] \in (\mathcal{Z}_2^h)^{m+1}$, then the transmitted sequence from $RM_{2^h}(1, m)$ would be $w^{-u_0}w^{-A^T u}$, where $w = e^{-\frac{2\pi j}{2^h}}$ is the 2^h -th root of unity, A is an $m \times 2^m$ binary matrix whose columns are the 2^m binary vectors of length m , and the exponentiation is considered component-wise. The received vector through an AWGN channel is $y = w^{-u_0}w^{-A^T u} + n$, where n is a complex white Gaussian noise vector with $E[nn^H] = \sigma^2 I_{2^m}$, where I_{2^m} is the size 2^m identity matrix. The maximum-likelihood decoder finds the vector u that maximizes the correlation between the received vector y and all of the first order Reed-Muller codewords in \mathcal{Z}_{2^h} . Specifically, it maximizes $|w^{-u_0}y^T w^{-A^T u}|$, with the constraint that $-\frac{\pi}{2^h} \leq \arg(w^{-u_0}y^T w^{-A^T u}) \leq \frac{\pi}{2^h}$ [125]. This maximization can be done using the modified fast Hadamard transformation. The Hadamard matrix, H_m , is a $2^m \times (2^h)^m$ matrix. Each row corresponds to a binary vector of size m , and each column corresponds to a 2^h -ary vector of size m . For any two vectors $a \in (\mathcal{Z}_{2^h})^m$ and $b \in (\mathcal{Z}_2)^m$, the element in the intersection of the corresponding row and column is $H_m[a, b] = w^{-a^T b}$. The matrix H_m can be generated recursively using the following equation, which is very similar to the recursive equation for

binary Hadamard matrices [13]:

$$H_m = \prod_{i=1}^m I_{2^{m-i}} \otimes H_1 \otimes I_{2^{h(i-1)}},$$

where \otimes denotes the Kronecker product. The algorithm finds the fast Hadamard transformation of the received word, $Y = H_m^H y$, and locates the element of Y , say Y_i , with the largest magnitude, $|Y_i|$. Then, it finds $u_0 \in \mathcal{Z}_{2^h}$, such that $|w^{-u_0} Y_i|$ is maximized. If the i^{th} column of H_m corresponds to the 2^h -ary vector u with size 2^m , the decoder outputs the vector $[u_0 \ u]$.

Another soft-decision decoding algorithm for $RM_{2^h}(1, 3)$ is presented in [107]. This algorithm is based on the Reed majority logic decoder for binary Reed-Muller codes. This algorithm can be easily generalized for the first order Reed-Muller code of any size. IEEE Wireless LAN standard committee has adopted this structure for IEEE802.11b physical layer. For size 8 complementary codes, at the transmitter, 4 phases ϕ_1 , ϕ_2 , ϕ_3 , and ϕ_4 are calculated for each input symbol, and based on these phases, the sequence

$$\{e^{j(\phi_1)}, -e^{j(\phi_1+\phi_2)}, e^{j(\phi_1+\phi_3)}, e^{j(\phi_1+\phi_2+\phi_3)}, -e^{j(\phi_1+\phi_4)}, e^{j(\phi_1+\phi_2+\phi_4)}, e^{j(\phi_1+\phi_3+\phi_4)}, e^{j(\phi_1+\phi_2+\phi_3+\phi_4)}\}$$

is transmitted. At the receiver the sequence y is received and the phase ϕ_2 is estimated from

$$\hat{\phi}_2 = \frac{1}{4}(y_0^* y_1 + y_2^* y_3 + y_4^* y_5 + y_6^* y_7).$$

Other phases can be similarly computed. This algorithm can be easily generalized to decode any $RM_{2^h}(1, m)$ code, for an arbitrary m .

Another efficient algorithm for hard-decision decoding of $RM_{2^h}(1, m)$ is presented in [118] that performs the binary fast Hadamard decoding, h times. This algorithm is not a maximum-likelihood decoding scheme, but is a 2^{m-1} Hamming and Lee distance decoder.

All of these algorithms can be used for decoding $RM_{2^h}(1, m)$. However, to decode higher order Reed-Muller codes, we will restate the theorem presented in [13] for constructing higher order non-binary Reed-Muller codes from lower order ones, defined in \mathcal{Z}_{2^h} . The proof is basically the same as the binary case, by performing the additions in modula 2^h . The steps of the proof are used in our decoding scheme, and therefore, the proof is stated here for the case of $RM_{2^h}(r, m + 1)$. In the following, $x|y$ means the concatenation of two sequences x and y .

Theorem 5.3.10. *The r^{th} order generalized Reed-Muller code of size 2^{m+1} can be constructed by the following recursive equation:*

$$RM_{2^h}(r, m + 1) = \{(f|f + g), \forall f \in RM_{2^h}(r, m) \ \& \ \forall g \in RM_{2^h}(r - 1, m)\}.$$

Proof. If c is a sequence from the r^{th} order generalized Reed-Muller code of size 2^{m+1} , it can be written as:

$$c(x_1, x_2, \dots, x_m, x_{m+1}) = [f(x_1, x_2, \dots, x_m) + 0 \cdot (x_{m+1})] + (x_{m+1}) \cdot g(x_1, x_2, \dots, x_m) \text{ mod } 2^h, \quad (5.46)$$

where $f(x_1, x_2, \dots, x_m)$ is an r^{th} order generalized Reed-Muller sequence of size 2^m and $g(x_1, x_2, \dots, x_m)$ is an $(r - 1)^{th}$ order generalized Reed-Muller codeword of size 2^m , both over \mathcal{Z}_{2^h} , and the addition is modula 2^h . Equation (5.46) can be written as $c = (f|f) + (0|g) = f|f + g$. Note that, the coefficients of all of the monomials containing x_{m+1} are determined by g . □

Since the transmitted codeword is equal to w^c , in order to obtain the partial sequence g , we have to multiply the second half of the received codeword by the conjugate of the first half, component-wise. Theorem 5.3.10 provides the following algorithm to decode $RM_{2^h}(r, m+1)$:

Algorithm I [*Recursive Decoding of $RM_{2^h}(r, m+1)$*]

1. The word $y = \{y_0, y_1, \dots, y_{2^{m+1}-1}\}$ is received.
2. For each $i = 0, 1, \dots, 2^m - 1$, find $g_i = y_{i+2^m} y_i^*$. This is equivalent to finding $g(x_1, x_2, \dots, x_m)$ or all of the monomials containing x_{m+1} , in (5.46). The resultant g is in $RM_{2^h}(r-1, m)$.
Call the first half of the original sequence, f .
3. Use either this recursive algorithm or any other non-recursive algorithms (mentioned in the previous or this subsection) to find the coefficients corresponding to g .
4. Subtract (in code domain) g from the second half of y , to find an estimate of f , which is in $RM_{2^h}(r, m)$.
5. Apply either this recursive algorithm or any other non-recursive algorithms (presented in the previous or this subsection) to the average of f and the estimate of f found in the previous step. Instead of averaging, we can find the equations corresponding to the coefficients in f and its estimate separately, and then for each coefficient, perform the same maximum-likelihood decoding, similar to (5.45).

Note that, we can terminate the recursive algorithm at any step by using any one of the non-recursive algorithms mentioned in this subsection (for the case of the first or second

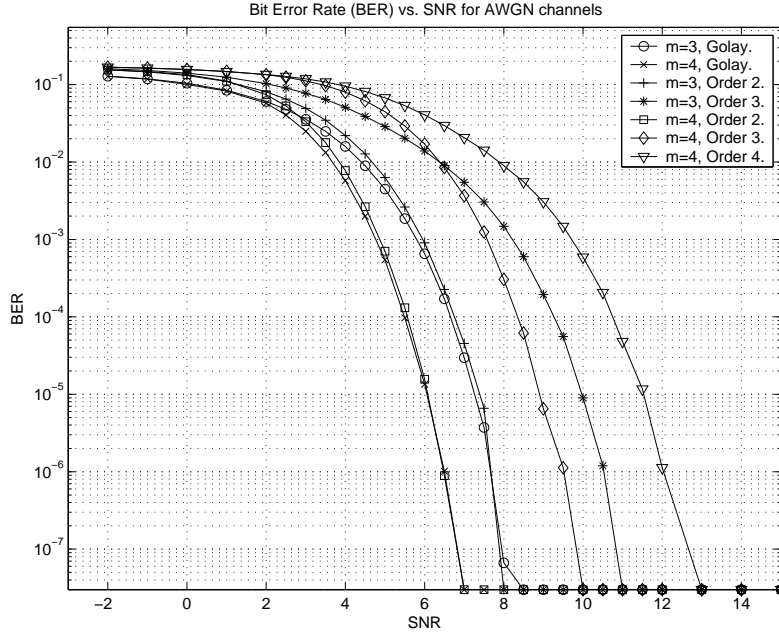


Figure 5.1: Bit Error Rate (BER) vs. SNR for AWGN channels

order generalized Reed-Muller code) or the K-Map method presented in the previous section (for higher order codes).

Comparison of the Complexities of Decoding Algorithms

We measure the complexity of these decoders by enumerating the number of complex multiplications and additions needed to perform the decodings. For the non-recursive scheme, to estimate one of the k^{th} order coefficients, we need to perform $2^{m-k}(2^k - 1)$ complex multiplication to create equations similar to (5.43), 2^h complex multiplications to perform equations like (5.45), and $2^{m-k} - 1$ complex additions. Since there are exactly $\binom{m}{k}$ coefficients of order k , decoding of $RM_{2^h}(r, m)$ requires $\sum_{k=0}^r \binom{m}{k} [2^h + 2^m (1 - 2^{-k})]$ complex multiplications and

$\sum_{k=0}^r \binom{m}{k} (2^{m-k} - 1)$ complex additions. For $m = 5$, $r = 3$ and $h = 2$, these numbers are calculates as 736 complex multiplications and 232 complex additions.

As for the non-recursive scheme, we measure the complexity of one iteration assuming that at the end of this iteration, the K-Map method is used to decode the partial sequences. Assuming that we start from the generalized $RM_{2^h}(r, m + 1)$, separation of f and g (as defined in Theorem 5.3.10) requires 2^m complex multiplications, decoding of $g \in RM_{2^h}(r - 1, m)$ requires $\sum_{k=0}^{r-1} \binom{m}{k} [2^h + 2^m (1 - 2^{-k})]$ complex multiplications and $\sum_{k=0}^{r-1} \binom{m}{k} (2^{m-k} - 1)$ complex additions, and decoding of $f \in RM_{2^h}(r, m)$ and its estimate requires $\sum_{k=0}^r \binom{m}{k} [2^h + 2 \cdot 2^m (1 - 2^{-k})]$ complex multiplications and $\sum_{k=0}^r \binom{m}{k} (2^{m+1-k} - 1)$ complex additions.

In the following theorem, we will prove that one iteration of the recursive algorithm is less complex compared to the K-Map method.

Theorem 5.3.11. *The decoding complexity of $RM_{2^h}(r, m + 1)$, in terms of the number of complex multiplications and additions, associated to the K-Map decoding method presented in Subsection 5.3.2 is higher than the one associated to one iteration of the recursive Algorithm I.*

Proof. We will consider decoding of $RM_{2^h}(r, m + 1)$.

a. Multiplications:

From our discussion in Subsection 5.3.2, the total number of complex multiplications

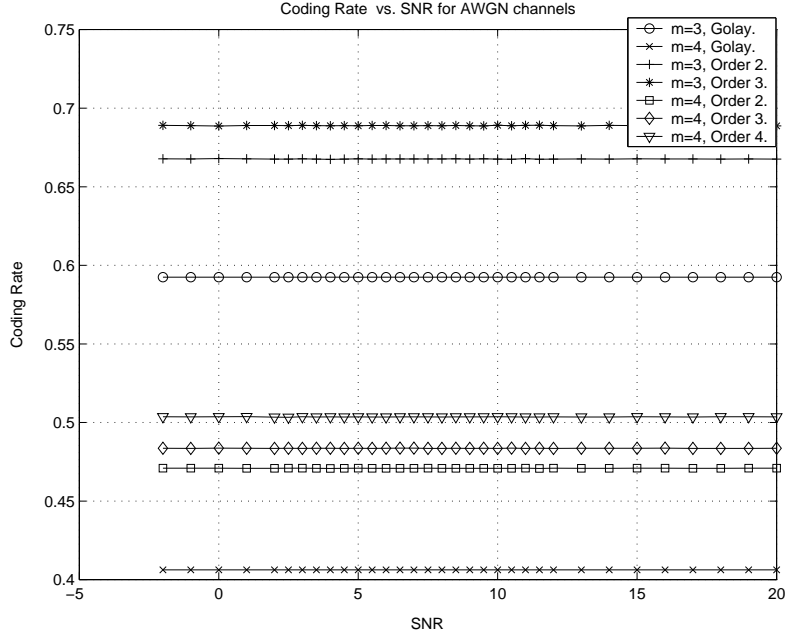


Figure 5.2: Coding rate vs. SNR for AWGN channels

for the non-recursive algorithm minus the ones for the recursive algorithm is given by:

$$\sum_{k=0}^r \binom{m+1}{k} [2^h + 2^{m+1} (1 - 2^{-k})] - \left[2^m + \sum_{k=0}^{r-1} \binom{m}{k} [2^h + 2^m (1 - 2^{-k})] + \sum_{k=0}^r \binom{m}{k} [2^h + 2 \cdot 2^m (1 - 2^{-k})] \right].$$

Using the relation $\binom{m+1}{k} = \binom{m}{k} + \binom{m}{k-1}$, and some changes of indices, the above statement becomes

$$2^{m+1} \sum_{k=0}^{r-1} \binom{m}{k} (1 - 2^{-(k+1)}) - 2^m - \sum_{k=0}^{r-1} \binom{m}{k} (1 - 2^{-k}) = 2^m \sum_{k=0}^{r-1} \binom{m}{k} - 2^m \geq 0$$

The equality holds if $r = 1$.

b. Addition:

From our discussion in Subsection 5.3.2, the total number of complex additions for the

Operation	[125], Alg. 1	[118], Alg. 17	[126], Alg. 6	[126], Alg. 8	K-Map	Recursive Alg.
+	19200	9600	1960	5800	1184	1088
×	0	0	720	840	1248	1104

Table 5.5: Computational complexities of Decoders for codes from $RM_4(2, 5)$

non-recursive algorithm minus the ones for the recursive algorithm is given by:

$$\begin{aligned}
& \sum_{k=0}^r \binom{m+1}{k} (2^{m+1-k} - 1) - \left[\sum_{k=0}^{r-1} \binom{m}{k} (2^{m-k} - 1) + \sum_{k=0}^r \binom{m}{k} (2^{m+1-k} - 1) \right] \\
& \sum_{k=0}^r \binom{m}{k} (2^{m+1-k} - 1) + \sum_{k=0}^{r-1} \binom{m}{k} (2^{m-k} - 1) - \\
& \sum_{k=0}^{r-1} \binom{m}{k} (2^{m-k} - 1) - \sum_{k=0}^r \binom{m}{k} (2^{m+1-k} - 1) = 0
\end{aligned}$$

□

The algorithms presented in [118, 125, 126] are proposed for the cosets of the first order Reed-Muller codes that are in the second order Reed-Muller codes, and our algorithms are presented for general $RM_{2^h}(r, m)$. However, we have compared the complexity of our algorithms for $m = 5$, $h = 2$, with the known schemes ones for the cosets of $RM_4(1, 5)$ in $RM_4(2, 5)$, in terms of the number of real additions and multiplications in Table 5.5. It is worthwhile to mention that although our algorithms are proposed for the generalized Reed-Muller code of any order, they still have comparable complexities with other schemes. Another point is that other schemes consider special second order monomials as the cosets, while our algorithms consider any second order monomials, and therefore, our algorithms perform higher number of calculations. As a matter of fact, the basis of comparison is not accurate enough and is against our algorithms.

<i>Scheme</i>	<i>size 4 SGP</i> s	<i>size 8 SGP</i> s built by Theorem 5.2.7
1	12032	385020
2	4096	131720

Table 5.6: Number of constructed 8-valued SGolay Pairs.

5.3.3 Performance Results

Simulation Results for super-Golay Codes

We have performed a search to find all of QPSK Golay pairs and 16-QAM SGolay pairs of size 4. The total number of QPSK Golay pairs is 256 which is exactly compliant by the formula given in [118] ($2^{2(m+2)} \frac{m!}{2}$ distinct QPSK Golay pairs with size 4). Using the representations in Theorem 5.2.10, We have been able to build 4096 SGolay pairs with size 4. The total number of 4-valued SGolay pairs is 12032. Table 5.6 compares the number of 8-valued SGolay pairs obtained from these two sets. In both cases, each Golay pair with size 4 yields 32 new SGolay pairs with size 8. In this table, scheme 1 means the total number of SGolay pairs and scheme 2 means the number of SGolay pairs built by Theorem 5.2.10.

Table 5.7 and 5.8 compare the code rate, information rate and the achievable PMEPR of the proposed recursive structure with the same values reported in [112, 118, 124] using QPSK constellation for 16 and 32 subcarriers. In the rows represented by “Jedwab1” the second order Reed-Muller cosets of the form $2^{h-1} \sum_{k=0}^{m-1} x_{\pi(k)} x_{\pi(k+1)}$ are used and therefore the constructed codes are Golay and their PMEPR is bounded up by 3dB. However, in the scheme represented by “Jedwab2” other forms of second order Reed-Muller cosets are also

<i>Scheme</i>	<i>Max PMEPR</i>	<i>Code rate</i>	<i>Information rate</i>
Recursive1	3dB	0.3438	1.375
Recursive2	3dB	0.375	1.5
Jedwab1	3dB	0.31	0.62
Jedwab2	6dB	0.47	0.94
Paterson	6dB	0.563	1.126
Chong1	3dB	0.2954	1.1817
Chong2	5.56dB	0.3053	1.2212

Table 5.7: Rates and PMEPRs for size 16.

<i>Scheme</i>	<i>PMEPR</i>	<i>Code rate</i>	<i>Information rate</i>
Recursive1	3dB	0.2109	0.84375
Recursive2	3dB	0.2266	0.9063
Jedwab1	3dB	0.19	0.38
Jedwab2	6dB	0.31	0.62
Paterson	6dB	0.375	0.75
Chong1	3dB	0.1835	0.7341
Chong2	5.56dB	0.3053	1.2212

Table 5.8: Rates and PMEPRs for size 32

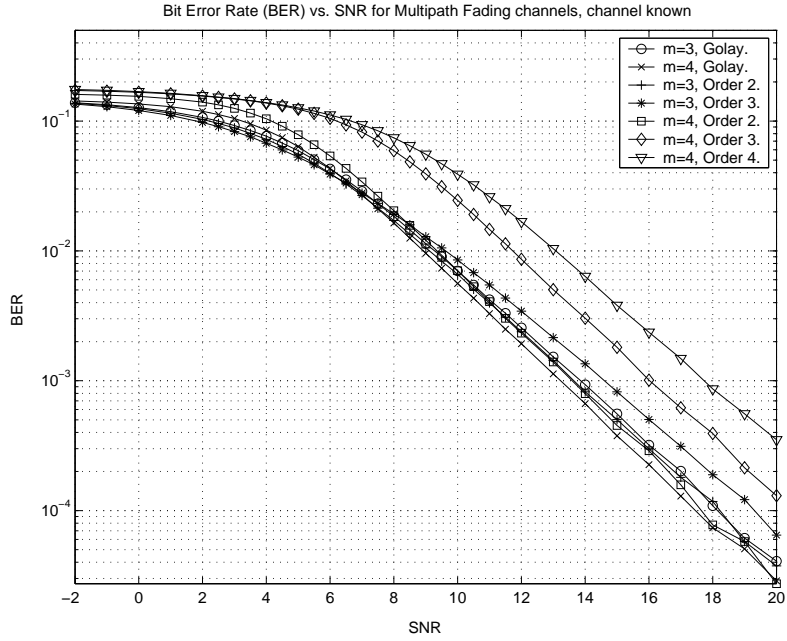


Figure 5.3: Bit Error Rate (BER) vs. SNR for Fading channels, when the channel is known at the receiver

allowed. This causes the PMEPR of the code to lie above 3dB. At the row represented by “Paterson”, the concept of Golay sets are used. It is proved in [112] that a Golay set of size 2^{k+1} achieve the maximum PMEPR of 2^{k+1} . The rows represented by “Recursive1” is using the 4-valued 16-QAM SGolay pairs generated from QPSK pairs as the seed, while “Recursive2” scheme uses the total 4-valued 16-QAM codes generated by exhaustive search, as the seed.

Simulation Results for Cyclic Golay Codes

Table 5.9 compares the rate and distances of Golay and cyclic Golay QPSK codes for $m = 3$ and $m = 4$. As we expect, the Hamming and Lee distances of the cyclic Golay code is in

<i>Code /Size</i>	$m = 3$	$m=4$
Golay	$R = 0.59, H = 2, L = 4$	$R = 0.42, H = 4, L = 8$
m^{th} order Cyclic	$R = 0.68, H = 2, L = 2$	$R = 0.51, H = 2, L = 2$
Exhaustive cyclic	$R = 0.73, H = 2, L = 2$	$R = 0.58, H = 2, L = 2$

Table 5.9: Rate, Hamming and Lee distance of some codes with low PAPR

general low, while their coding rate is higher than the Golay code.

For the comparison of bit error rate, we used an OFDM system with $m = 3$ and 4, corresponding to 8 and 16 subchannels. The same discussion applies to larger number of subchannels. A white Gaussian noise with variance 0.5 per dimension was assumed at each subchannel. The symbols were chosen from \mathcal{Z}_4 , corresponding to $h = 2$. The bit error rate versus SNR for Golay and cyclic Golay schemes is shown in Fig. 5.1 for AWGN channels. It is obvious that the cyclic Golay codes with $m = 4$, order 4 and $m = 3$, order 3 have the worst bit error rates, and the Golay code with $m = 4$ has the best performance. The second order cyclic Golay code with $m = 4$ performs better than other codes. As the difference between the dimension of the code and the order is decreased, the bit error rate of the code increases. This is attributed to the fact that the cyclic Golay code is a subset of $RM_{2^h}(r, m)$ and also the distance of the generalized Reed-Muller codes depends on the difference between the dimension and the order of the code. Note that, the difference between the BER's are better observed at high SNRs. Fig. 5.2 shows the rate we obtained for these codes. The following facts can be seen from this figure. First, the rates shown in this figure are a little lower

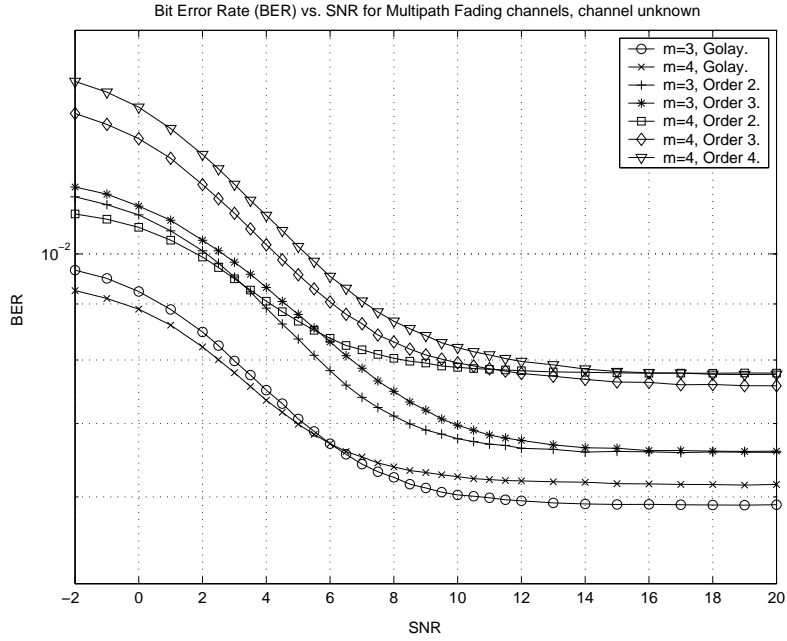


Figure 5.4: Bit Error Rate (BER) vs. SNR for Fading channels, when the channel is not known at the receiver

than Table 5.9. This is because for example for $m = 4$, only $8(= 2^3)$ out of $12(= \frac{4!}{2})$ coset representatives are used in simulation. Second, the codes with $m = 4$ normally have lower rate than the cases with $m = 3$. Third, for each dimension, the pure Golay code has the worst coding rate. Fourth, as the difference between the dimension and the order of cyclic Golay code decreases, the code can achieve better coding rate. Finally, although, the cyclic Golay code can result in different rates at different OFDM block, the rate of the code in a long frame size is almost constant for all values of SNR. These two figures clearly show the trade-off between the coding rate and the coding distance, while the PAPR is bounded up to $3dB$.

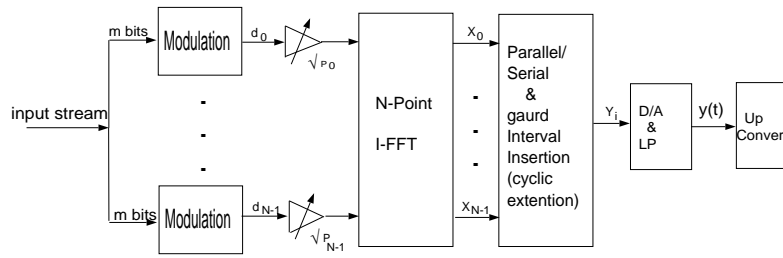


Figure 5.5: OFDM transmitter with Golay and cyclic Golay encoder

The results for the two path fading channels are shown in Figs. 5.3 and 5.4. Both paths have a Raleigh envelope and the delay equal to the OFDM symbol duration divided by the number of subchannels. We have also considered the same white Gaussian noise as in Fig. 5.1 in these figures. In Fig. 5.3 we assume that the receiver knows the channel characteristics, and in Fig. 5.4, the receiver has no knowledge of the channel condition. We can observe the same trend as in Fig. 5.1, of course with lower bit error rates. When the channel is unknown at the receiver, the BER tends to converge at a constant level for high SNR. This is due to the fact that in this case, the SNR is the quotient of the signal power to the noise power. At high SNRs, the effect of fading characteristics prevails the effect of noise, and therefore the BER doesn't change significantly with increasing SNR.

We saw that the cyclic Golay code can perform higher coding rate at the expense of lower distance. However, in high SINR environment, we might be able to tolerate lower distance codes. The system depicted in Fig. 5.5 sets a SINR threshold level. If the SINR is above the threshold, we use cyclic Golay codes at the transmitter to transmit higher data rates. If the SINR is below the threshold, we switch to the Golay codes having better error correcting property, but lower rates. The threshold can be chosen according to the upper bound of

BER, we can tolerate. We use this system along with $m = 4$, for our simulation purposes in Figs. 5.6 and 5.7. We assume that the SNR of the received signal changes randomly according to a uniform distribution. These figure compare four schemes. In the first and second schemes, when the SNR is below a threshold, we use the Golay code, while the cyclic Golay code is used when the SNR is above the threshold. In the first scheme the second order and in the second scheme the fourth order cyclic Golay codes are used. In the other two schemes, we have two thresholds. If the SNR crosses the first threshold, we switch from Golay to the 2^{nd} order cyclic Golay, and when the SNR is above the second threshold, we switch to the 4^{th} order cyclic Golay code. The difference between the thresholds are different in these two schemes. Notice that, the x -axis in these figures attributes to the SNR threshold on which we switch the coding scheme, not the actual SNR. The actual SNR is changed from one OFDM block to another, according to a uniform distribution. When the threshold is set to the lowest value, the Golay codes are never used, and therefore we expect poor BER performance with a better coding rate. As the threshold becomes larger, the cyclic Golay code is less used and therefore the BER tends to be lower. When the SNR is set at the highest value, we use the Golay code, all the time.

Fig. 5.8 shows the changes in bit error rate versus the coding rates for $m = 4$, when the SNR is kept constant. The figure is in agreement with our expectation that for a fixed SNR, as the rate increases the codes performs higher bit error rate. For example, by fixing the SNR at $6dB$, as the coding rate goes from 0.40625 to 50358, the bit error rate is increased from 5.4×10^{-4} to 6.7×10^{-2} .

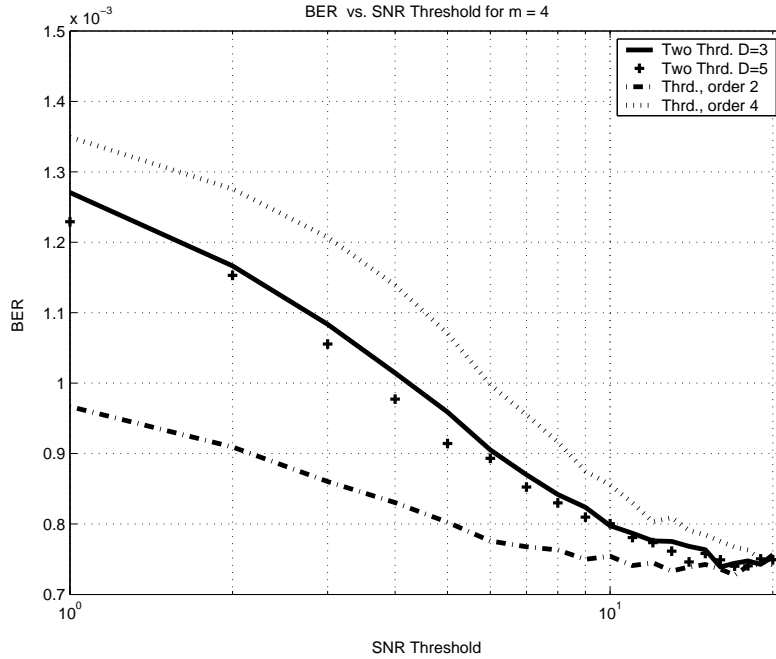


Figure 5.6: Bit Error Rate (BER) vs. SNR Threshold for AWGN channels, when $m = 4$

5.3.4 Summary of the Chapter

In this chapter we tried to make some modification to Golay complementary codes to increase their achievable rates. First, we modified the concept of Golay codes to cover non-equal energy constellations, and to this end we defined super-Golay codes. We proposed recursive construction schemes that allows us to build super Golay codes with a specific size from lower size codes. The construction started from QPSK Golay sequences which are efficiently created using 2nd order cosets of $RM_{2^h}(1, m)$.

Next, we introduced the concept of cyclic Golay codes and shown that, with appropriate time shaping, they maintain the same level of PAPR in the discrete domain (critical sampling) as the Golay codes. Moreover, we have shown that the set of cyclic Golay codes is

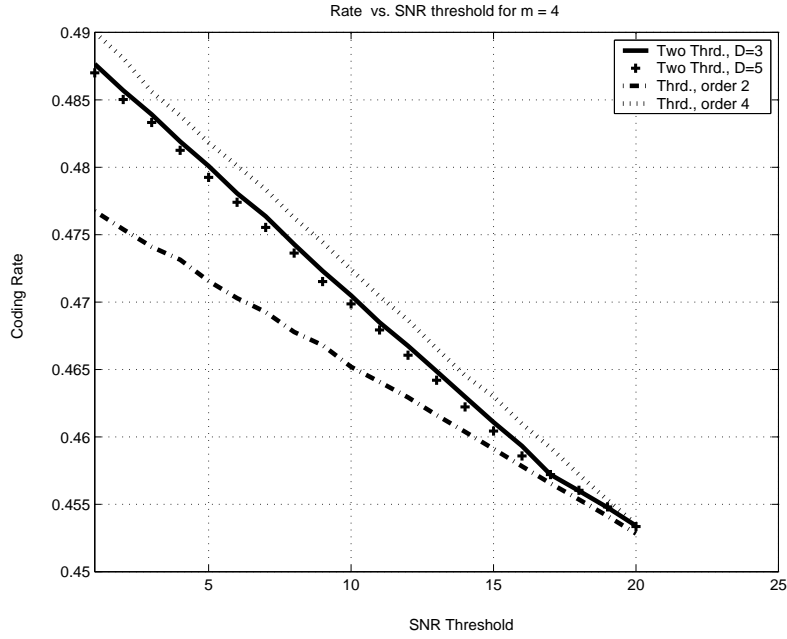


Figure 5.7: Coding rate vs. SNR Threshold for AWGN channels, when $m = 4$

a super-set of Golay codes and therefore results in higher coding rate. We have designed a construction method to find the cyclic shift of any code represented by Boolean algebraic forms. These codes introduce a trade-off between the coding rate and the distance of the code. Two decoding methods for $RM_{2^h}(r, m)$ are also introduced. A generalization of majority logic decoding approach, using Karnaugh maps, both for soft decision, and hard-decision decoding of $RM_{2^h}(r, m)$ is proposed. To reduce the complexity of decoders, a recursive approach is introduced that brings the decoding procedure down into the decoding of lower size and lower order codes.

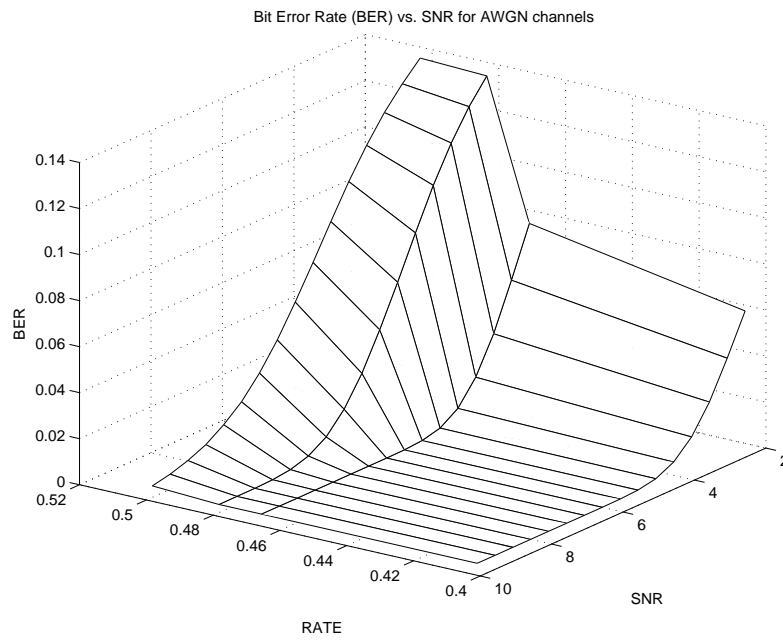


Figure 5.8: Bit error rate vs. coding rate for fixed values of SNR for AWGN channels, when $m = 4$

Chapter 6

Service Level Agreement (SLA) Based Scheduling Algorithms and QoS-provisioned Channel Allocation for OFDMA

6.1 Motivation and Previous Works

With the rapid growth in broad-band wireless data networks and increasing demand for multimedia applications, the future wireless networks should be able to provide services for heterogeneous traffic with diverse quality of service (QoS) requirements. Also, the fast increase of traffic volume, both the number of users and bandwidth requirements of the emerging applications, needs efficient utilization of limited spectrum of wireless networks. The development of new technologies (including packet scheduling) for wireless networks has been receiving a lot of attention in both research and industry. A packet scheduling scheme is the mechanism that resolves contention between the packets and manages bandwidth among

the mobile users. Scheduling algorithms that support QoS and maintain high throughput are crucial to the development of broad-band wireless networks.

Although many developed scheduling algorithms are available for wireline networks, they cannot be directly carried over to wireless networks due to major differences in medium. These include location-dependent and time-varying wireless link capacity, scarce and limited bandwidth, high error-rate, and user mobility. The time varying nature of wireless channels introduces some discontinuity in the availability of a user when the channel is in a bad condition. The very same nature of wireless channel provides opportunities for the transmission of large amount of information when the channel is in a good condition. An efficient scheduler has to take advantage of the time variation of wireless channel. On the other hand, if a scheduler operates independent of channel condition, it might allocate bandwidth to users in deep fade, where most of data is lost and bandwidth is wasted, while at the same time deprive users with good channel from taking advantage of their instantaneous large capacity. Therefore, there is a need to develop new packet scheduling technologies which account for the special characteristics of wireless networks, to support QoS (differentiation and guarantees) and to provide high network utilization.

Quality of service (QoS) refers to the capability of a network to provide certain service to selected network traffic. The four important attributes of QoS in packet networks are *dedicated bandwidth, controlled jitter/latency, and controlled loss characteristics*. Through the notion of effective bandwidth, it can be shown that a certain QoS level can be translated into a bandwidth guaranteed to a user [16]. In other words, given traffic characteristics

and its QoS requirements, there is a required bandwidth (effective bandwidth) needed to be provided to the traffic for supporting the requested QoS. For instance, let us assume that the traffic is a two state Markov process (ON/OFF model) with exponential state transition rates. The ON state durations and OFF state durations are exponentially distributed with parameters β_M and α_M , respectively. When in OFF state, the source is idle, while in ON state the source generates packets at the peak rate λ_M . For this traffic, we assume that the QoS requirement is to guarantee the probability that the delay of the packets in buffer exceeds a certain threshold D_M is less than ϵ_M , i.e.

$$P[d > D_M] < \epsilon_M. \quad (6.1)$$

Given the traffic parameters $(\alpha_M, \beta_M, \lambda_M)$ and the QoS parameters (D_M, ϵ_M) , the effective bandwidth of this traffic would be [130]:

$$C_d = \frac{\lambda_M(\alpha_M D_M - \ln \epsilon_M)}{(\alpha_M + \beta_M) D_M - \ln \epsilon_M}. \quad (6.2)$$

For other QoS parameters and traffic patterns, there are similar expressions that relate QoS to effective bandwidth. As a consequence, in this work QoS is represented by a reserved bandwidth, guaranteed to a user.

In wireless scheduling algorithms, there is a trade off between throughput maximization, which relates to the efficiency in utilizing bandwidth, and supporting QoS, which indicates how resources are shared among users. Throughput is defined as the total bandwidth supported for the aggregate of the users, regardless of how much bandwidth is assigned to each user. On the other hand, QoS refers to the ability of providing a minimum guaranteed band-

width to a user, under different rate assignment. Let us define *QoS region* as the summation of the reserved rates across all the users for all the feasible rate assignments. It is important to notice that for any given scheduler, the QoS region cannot be higher than the throughput. To maximize spectrum utilization, a good scheduler identifies and schedules users with high instantaneous channel capacities. However, this methodology could be biased to mobile users who are closer to the base station. In other words, this scheduler allows only users close to the base station to access the channel, but the users away from the base station will not receive the reserved bandwidth necessary for the required QoS. There exist scheduling mechanisms able to provide high utilizations but they sacrifice the user satisfaction. Also, there exist other scheduling schemes that sacrifice network throughput to support QoS. Our main objective is to obtain a scheduling algorithm that maximizes the system throughput while it provides a QoS region very close to the system throughput.

Channel state dependent packet scheduling (CSDPS) defers transmission of packets on links experiencing bursty errors [131]. A link status monitor, checks the channel condition for all mobile hosts, and when it determines that a channel is in a bad state, the scheduler does not serve the user associated with that link. Any one of the known wireline scheduling algorithms, e.g., round robin, earliest deadline first, and longest queue first, could be used as the service policies for this scheduling algorithm. However, CSDPS does not have any mechanism for supporting QoS (to guarantee bandwidth) for a mobile user. CSDPS with class based queuing (CSDPS-CBQ) groups users into different classes and each class is committed with a certain amount of bandwidth [132]. It keeps track of the amount of service received

by each class in a certain time interval window. However, this scheduling algorithm does not have an explicit mechanism for compensating those mobile users who have not received the promised service.

Idealized wireless fair queuing (IWFQ) is a modified version of weighted fair queuing (WFQ) scheduling algorithm for wireless networks [133]. It uses an error-free WFQ as a reference system and tries to approximate the real service to the ideal error-free system. In IWFQ, each flow is assigned a queue, and the m^{th} packet of the n^{th} queue is assigned a *start time* $S_{n,m}$ and a *finish time* $F_{n,m}$, where:

$$\begin{aligned} S_{n,m} &= \max\{v(A(t)), F_{n,m-1}\} \\ F_{n,m} &= S_{n,m} + \frac{L_{n,m}}{r_n} \end{aligned} \quad (6.3)$$

Here, $L_{n,m}$ is the packet size, $v(A(t))$ is the system virtual time defined in WFQ, and r_n is the rate reserved by user n [133]. This algorithm provides some appealing properties in fairness and QoS guarantees. However, when a user is compensated for its previous lagged service, all other users with good channels will not be served at all.

Service Level Agreement (SLA) is a contract between a user and its service provider. An SLA defines the service (QoS) requested by the user, the price that the user must pay for the service, the penalty if the agreement is violated, and etc. This chapter considers SLA as the reference point between the network and the network users. In this work, we introduce a notion of income maximization, by which the scheduler is rewarded when total network throughput is increased, and penalized when SLA for each user is violated.

We will show that by properly choosing penalty, as a function of SLA, and reward, as a function of network throughput, the trade-off between the throughput and user satisfaction is performed efficiently. We will also show that our algorithms, meet the QoS and utilize network resources efficiently. We propose a greedy solution and a dynamic programming approach for the problem.

This scheme is used for channel allocation in an OFDMA system. In OFDM systems a high data rate stream is split into a number of lower rate data streams. Each substream is modulated separately on one of the orthogonal subcarriers. One way of applying OFDM to a multi-user scenario is through OFDM-TDMA or OFDM-CDMA [134], where different users are allocated different time slots or different frequency spreading codes. However, each user has to transmit its signal over the entire spectrum. This leads to an averaged-down effect in the presence of deep fading and narrowband interference. Alternatively, one can divide the total bandwidth into traffic channels (one or a cluster of OFDM subcarriers) so that multiple access can be accommodated in an orthogonal frequency division multiple access (OFDMA) fashion. An OFDMA system is defined as one in which each user occupies a subset of subcarriers, and each carrier is assigned exclusively to only one user at any time. One advantage of OFDMA over OFDM-TDMA and OFDM-CDMA is elimination of intra-cell interference, thus avoiding the need of CDMA type of multi-user detection. In OFDMA, each transmitter is assigned to a subset of the entire subcarriers to resolve the effect of interference caused by the multi-access environment. This enables the network to perform a flexible resource allocation with the goal of increasing the overall network throughput depending

on different users' traffic load. This capability leads to increased system throughput and spectral efficiency, when the allocation of subcarriers to different users is performed carefully [135, 136]. Another advantage of OFDMA is that it can exploit multiuser diversity, when an user avoids the channels that are in deep fade or have narrowband interference. Since different users face different channel qualities, a badly faded channel for one user may still be favorable to other users. By careful subcarrier allocation, the spectral efficiencies of the system can outperform interference-averaging techniques significantly [137].

Clearly, careful subchannel allocation in an OFDMA system is very crucial in determining the performance of the overall system. However optimum channel allocation in OFDMA is a fundamentally difficult problem. In practice, additional constraints, e.g., individual users rate requirement, further complicates the problem. In [136], the problem of subcarrier allocation for generic multiple access systems with orthogonal subchannels in a multi-cell system is studied. In [138] the authors propose an model wherein a single network access point serves a number of terminals, which require varying data rates. They also use OFDMA so that different terminals are allocated a different number of subcarriers depending on their data rate requirement. In this manner all terminals can transmit simultaneously without collisions. The use of an IFFT in an OFDM modulator greatly simplifies this because subcarriers that are not used can be simply zeroed out at the IFFT input. In [139] a dynamic resource allocation scheme for OFDMA-based wireless broadband networks is presented. They formulate the problem of maximizing the total packet throughput subject to individual users outage probability constraint, by assuming a finite buffer for the arrival packets and dynamically

allocates the radio resource based on users channel characteristics, traffic patterns and QoS requirements. A Lagrangian relaxation algorithm is first introduced in [140] to minimize the total power consumption with constraints on transmission rate for users requiring different classes of services. Linear programming is used in [141] to solve the subcarrier allocation by linearizing the function of rate in term of power.

Section 2.6 briefly described the OFDMA and channel allocation in IEEE802.16a. In summary, a fixed subset of subcarriers in consecutive time slots are assigned to each user according to a static assignment [5].

The network has to maintain a required Quality of Service (QoS) for each user, which is not necessarily in line with maximizing the network's total throughput. The main goal of this chapter is to allocate a subset of subcarriers to each user such that the QoS is satisfied for each user and at the same time the overall network throughput is maximized. In order to increase resource utilization in the network while providing QoS, a scheduler has to adjust the allocation of subcarriers to users based on user demands and channel conditions. The scheduler has to achieve this purpose with reasonable performance and low complexity.

As a result, we use the concept of revenue maximization is then used in an OFDMA system, in which the allocation of subcarriers to different users is performed in such a way that the overall network revenue is maximized. As a result, throughput is maximized while QoS is maintained above the required level.

This chapter is divided in two parts. In the first part (Section 6.2) we present the SLA-based scheduling in wireless networks. The system model is introduced in Subsection

6.2.1. Subsections 6.2.2 and 6.2.3 describe two extreme cases, where only throughput or QoS is the objective, and finally Subsection 6.2.4 describe our proposed trade-off between QoS and throughput through the notion of “income maximization”. The second part covers the application of income maximization in OFDMA channel allocation (Section 6.3). The OFDMA system model is brought in Subsection 6.3.1, and different scheduling algorithms are presented in Subsection 6.3.2. In Section 6.4 the performance of the proposed algorithms both for TDMA, and OFDMA are compared through numerical studies. Finally, Section 6.5 summarizes the chapter.

6.2 SLA-Based Scheduling for Wireless Networks

In the first part of this chapter, we will propose our framework for SLA-based scheduling. This scheme can be applied to any multiuser scheme like CDMA, TDMA, and OFDMA. In the second part of this chapter, we will apply this framework to OFDMA.

6.2.1 System Model

In the first part, a single cell wireless network is investigated. We consider a time-slotted system, where time is the resource to be shared among all mobile users by a central processor. At any given time, only one user can be scheduled to occupy a given channel within a cell. The scheduling algorithm decides a time slot should be assigned to which user. At the down-link, a separate queue is assigned to each mobile user at the base station, and we assume

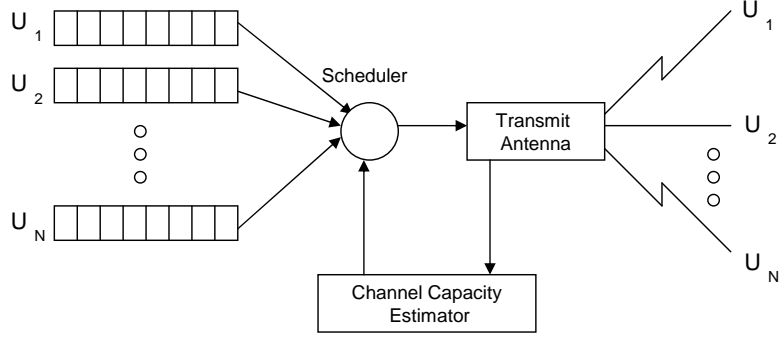


Figure 6.1: System block diagram

that the scheduler at the base station has the full knowledge of the status of the queues.

The block diagram of this system is shown Figure 6.1.

Let r_n denote the effective bandwidth (rate) rate reserved by user n ($n = 1, 2, \dots, N$), which is a fraction of the total available bandwidth ($0 \leq r_n \leq 1$). In fluid model, user n expects to receive a fraction of a time slot, r_n . However, in this work, we do not consider the fluid model and assume that a time slot is assigned only to one user. Define $Y_n(t)$ as:

$$Y_n(t) = \begin{cases} 1 & \text{if the scheduler selects user } n \text{ at time } t, \\ 0 & \text{otherwise.} \end{cases} \quad (6.4)$$

Also, assume that the indicator $I_n(t)$ is defined as:

$$I_n(t) = \begin{cases} 1 & \text{if queue of user } n \text{ is non-empty at time } t, \\ 0 & \text{otherwise.} \end{cases} \quad (6.5)$$

We assume that the link between each user and the base station is a wireless fading channel. In a power controlled system, the average power in each link is maintained at a fixed level and the instantaneous power follows a Rayleigh fading distribution. The signal

to noise ratio (SNR) for the n^{th} user is a function of the received power P_n and the noise power N_n . The capacity per unit bandwidth for this user, B_n , is given by

$$B_n = \log_2(1 + P_n/N_n). \quad (6.6)$$

Since thermal noise at each receiver is fixed, the SNR at each user follows a Rayleigh distribution.

We assume that the link capacity is quantized to a limited number of levels. Let us assume that the channel capacity for user n at time t is denoted by $g_n(t)$, which is a fractional number. Therefore, the service received by user n at time t is $g_n(t)Y_n(t)$.

Next, we consider two scheduling algorithms. The first one is proposed to support QoS and the other one provides high network throughput.

6.2.2 Maximum Credit Scheduling (MCS)

In order to support QoS, a scheduler monitors and allocates resources in such a way that users' effective rates stay within a satisfactory range. A credit based mechanism can be used to measure and control the service provided to each user; the user n is assigned a credit, denoted by $C_n(t)$ ($n = 1, 2, \dots, N$). A user's credit represents how much service the network owes to the user.

The credit for user n at time t evolves as follows:

$$C_n(t) = C_n(t-1) + I_n(t)r_n - g_n(t)Y_n(t). \quad (6.7)$$

The second term on the right hand side of the above equation, $I_n(t)r_n$, represents the

service reserved by user n . If the n^{th} queue is non-empty, this term is the requested service. The third term, however, represents the service received by user n . Starting from $C_n(-1) = 0$ for all users, by induction, it is straightforward to see that (6.7) leads to the following non-recursive expression for the credits:

$$C_n(t) = r_n \sum_{s=0}^t I_n(s) - \sum_{s=0}^t g_n(s) Y_n(s). \quad (6.8)$$

In the above equation, the first and second terms are the reserved and received service by user n up to time t , respectively. A negative credit means that a user has received a better service than the reserved service. On the other hand, a positive credit implies that the network owes service to the user. Therefore, credit is a measure of how much QoS is delivered or how much network owes to a user. To support QoS, a scheduler must keep the credits of all the users as small as possible. In this case, for users with non-empty queues the service delivered to each user is close to its reserved service.

In order to minimize user credits, a maximum credit scheduler (MCS), assigns the available bandwidth to the user with maximum credit [142]:

$$Y_n(t) = \begin{cases} 1 & \text{if } n = \arg \max_k \{C_k(t)\}, \\ 0 & \text{otherwise.} \end{cases} \quad (6.9)$$

Since this scheduling is based on the credit values at time t that is independent of the channel capacities at time t , the total system throughput with this algorithm is equal to the average channel capacity, i.e., $\mathbf{E}(g_n(t))$.

6.2.3 Maximum Throughput Scheduling (MTS)

It has been proved that to maximize the network throughput, a scheduler selects the user with the best capacity or with the lowest fading among all the users [143], i.e.,

$$Y_n(t) = \begin{cases} 1 & \text{if } n = \arg \max_k \{g_k(t)\}, \\ 0 & \text{otherwise.} \end{cases} \quad (6.10)$$

The total throughput in this algorithm is equal to the expectation of maximum channel capacity, i.e., $\mathbf{E}(\max_n g_n(t))$. However, this algorithm has no mechanism for supporting QoS.

6.2.4 A Trade-off: SLA-Based Scheduling Algorithms

In MTS, a user that is trapped in a bad channel state, does not receive a service as long as its channel stays at that state. For this user, QoS or SLA is not satisfied. Thus, supporting QoS and maximizing network throughput cannot necessarily be achieved at the same time.

MCS does not face this trade-off in wired networks since in those networks channel responses of all users are equally good, i.e., $g_n(t) = 1$ for all n, t . Therefore, by selecting the user with the highest credit, the scheduler maintains the credits as small as possible. However, in wireless networks, attempting to support QoS for users with bad channel response may result in reducing network throughput and it does not guarantee QoS for those users.

A more efficient scheduler may ignore users with the most eligible QoS that are in deep fade, in favor of users with better channel response with the hope that in the future the capacity of those ignored users would improve. In this work, we study schedulers that

provide QoS, and at the same time maximize network throughput.

In this work, we propose scheduling algorithms that resolves the trade-off between throughput and QoS based on the users SLA. SLA includes QoS, pricing for the service provided and penalty when the agreement is violated. Let us denote $d_n(t)$ to be the income of the network from user n at time t . Also, let us assume that for the service received by user n , i.e., $g_n(t)Y_n(t)$, the network charges the user n by $\alpha_n g_n(t)Y_n(t)$, where α_n is the rate that the n^{th} user pays for the service. On the other hand, if the user n has not received its requested QoS, the credit assigned to it will increase, and so the network is penalized by $f_n[C_n(t)]$. We assume that $f_n[\cdot]$ is a real, positive and continuous function with $f_n[x] = 0$ for $x \leq 0$. Both function $f_n[\cdot]$ and α_n are defined through SLA between the network and user n . Then, we obtain:

$$\begin{aligned} d_n(t) &= \alpha_n g_n(t)Y_n(t) - I_n(t)f_n[C_n(t)] \\ &= \alpha_n g_n(t)Y_n(t) - I_n(t)f_n[C_n(t-1) + r_n - g_n(t)Y_n(t)]. \end{aligned}$$

The total income of the network at time t is given by

$$D(t) = \sum_{n=1}^N d_n(t),$$

where N is the total number of users. An SLA-based scheduler selects the user that increases the total income.

The penalty function has a significant role in the performance of a SLA-based algorithm. This function is chosen in such a way that a user with negative credit does not penalize the system since this user has received its requested QoS; therefore, $f_n[x] = 0$ for $x \leq 0$. Also, if

a user has accumulated a big credit, i.e., has received a poor QoS, it might be beneficial to disconnect this connection and pay the corresponding penalty. Moreover, we will expect that the penalty increase to be more significant for high credits. This means that $f[\cdot]$ needs to be convex. We will see that for some special case, the convexity of f is necessary. Besides, the effect of penalty functions is more described in the simulation results. One special example for $f_n[\cdot]$ is:

$$f_n[x] = \begin{cases} \gamma_n x^2 & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (6.11)$$

where γ_n is a positive number.

6.2.5 Maximum Income Greedy Scheduling: A Suboptimal Solution

The maximum income greedy scheduling (MIGS) algorithm selects the user that maximizes the total system income at each time slot t :

$$D(t) = \sum_{n=1}^N d_n(t) = \sum_{n=1}^N \alpha_n g_n(t) Y_n(t) - \sum_{n=1}^N I_n(t) f_n[C_n(t)].$$

Without loss of generality, from now on, we assume all users have non-empty queues, i.e., $I_n(t) = 1$. The following lemma summarizes the MIGS algorithm.

Theorem 6.2.1. *The maximum income greedy scheduling algorithm selects the user that maximizes the following quantity over all the users with non-empty queues:*

$$H_p(t) = \alpha_p g_p(t) + f_p[C_p(t-1) + r_p] - f_p[C_p(t-1) + r_p - g_p(t)]. \quad (6.12)$$

Proof. If user k is selected at time t , then $Y_k(t) = 1$ and $Y_n(t) = 0$, $n \neq k$. We denote the total income at time t if user k is selected by $D^k(t)$, i.e.:

$$D^k(t) = \alpha_k g_k(t) - \sum_{\substack{n=1 \\ n \neq k}}^N f_n[C_n(t-1) + r_n] - f_k[C_k(t-1) + r_k - g_k(t)]. \quad (6.13)$$

The MIGS selects the user p that maximizes the total income:

$$p = \arg \max_k \{D^k(t)\}. \quad (6.14)$$

That is, $Y_p(t) = 1$, and $Y_k(t) = 0$, $k \neq p$ and $D^p(t) \geq D^k(t)$, $k \neq p$. Therefore, we get

$$\begin{aligned} \alpha_p g_p(t) - \sum_{\substack{n=1 \\ n \neq p}}^N f_n[C_n(t-1) + r_n] - f_p[C_p(t-1) + r_p - g_p(t)] > \\ \alpha_k g_k(t) - \sum_{\substack{n=1 \\ n \neq k}}^N f_n[C_n(t-1) + r_n] - f_k[C_k(t-1) + r_k - g_k(t)]. \end{aligned}$$

After simple manipulations, we obtain

$$\begin{aligned} \alpha_p g_p(t) + f_p[C_p(t-1) + r_p] - f_p[C_p(t-1) + r_p - g_p(t)] \geq \\ \alpha_k g_k(t) + f_k[C_k(t-1) + r_k] - f_k[C_k(t-1) + r_k - g_k(t)]. \end{aligned} \quad (6.15)$$

Therefore, the scheduling algorithm maximizing the total income at time t maximizes the right side of the above inequality. \square

Figure 6.2 displays the metric generator block. This block receives α_n , $f_n[\cdot]$ and r_n from the customer SLA, and $g_n(t)$ from the channel estimator and generates $H_n(t)$. The state variable of this block is the user credit, $C_n(t)$. The state variable is updated by the feedback

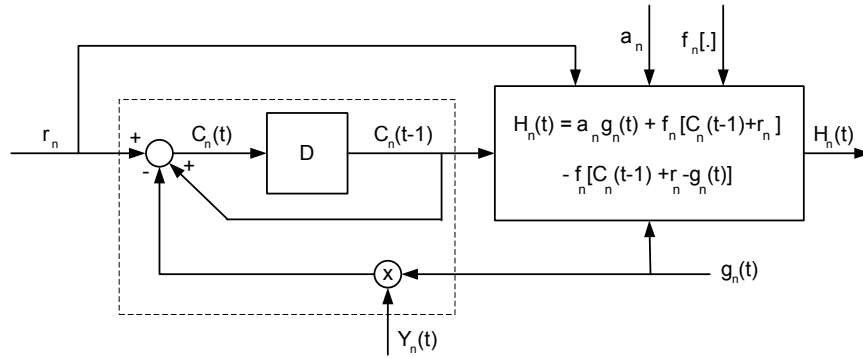


Figure 6.2: Metric generator block diagram for MIGS

circuit, shown in Figure 6.2. As illustrated in this figure, the decision of the scheduler ($Y_n(t)$) contributes to update the credit, ($C_n(t)$). It is interesting to notice that this module depends only on the parameters associated with user n , and is independent of the other users. Figure 6.3 shows the SLA scheduler. As shown in this figure, the SLA scheduler is only a maximization that selects the user with the highest metric, ($H_n(t)$). The output of the scheduler is the binary vector $\{Y_n(t)\}_{n=1}^N$ that decides which user is selected in this time slot.

Special Cases

In the following, we will consider two extreme cases where in one case only the system throughput is important, and in the other case, only the QoS matters. We will show that the scheduling algorithm mentioned in this section approaches the two cases, so this algorithm addresses the throughput-QoS trade-off.

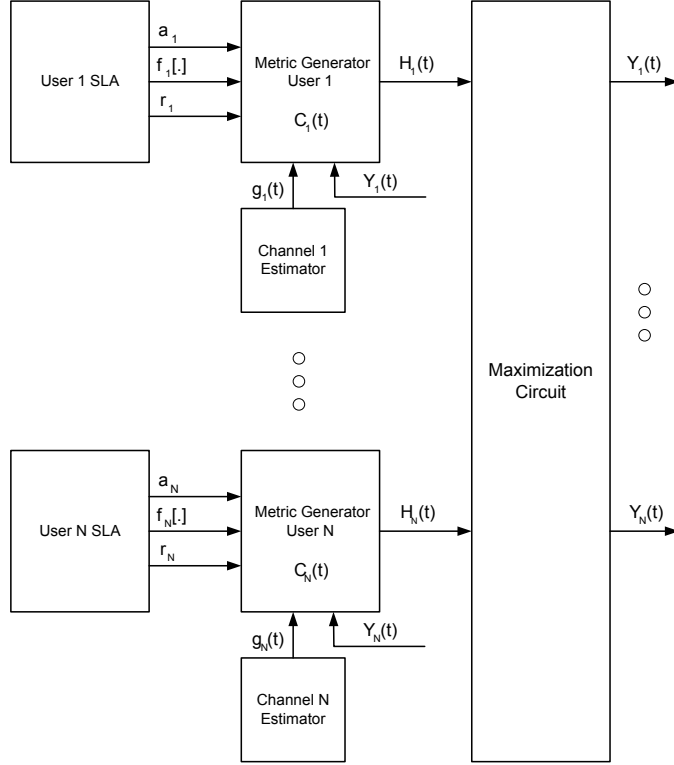


Figure 6.3: SLA scheduler block diagram

Case I: Deep Fading: Let us assume that at time t , the k^{th} user is in deep fade, (i.e., $g_k(t) = \epsilon \ll 1$). Intuitively, we expect this user not to be scheduled, because scheduling this user results in bandwidth loss at time slot t . Moreover the small allocated rate ($g_k(t)$) is not large enough to support the QoS for this user. Because of continuity of the function $f_k[.]$, we can write

$$f_k[C_k(t-1) + r_k] - f_k[C_k(t-1) + r_k - g_k(t)] = \delta,$$

for a small positive δ . Thus, the metric associated with this user would be:

$$H_k(t) = \alpha_k g_k(t) + f_k[C_k(t-1) + r_k] - f_k[C_k(t-1) + r_k - g_k(t)] = \alpha_k \epsilon + \delta,$$

which is a small positive number. Therefore, a user in deep fade has a small metric and is not expected to be scheduled.

Case II: Maximum Throughput Scheduling (MTS): Let us assume that the system would not be penalized for not supporting QoS, i.e., $f_k[\cdot] = 0$. Also, we assume that the system charges different users with the same rate, i.e., $\alpha_k = \alpha$. As we discussed in Section 6.2.3, in this case, the system tries to maximize the system throughput and the scheduler selects the user with the best capacity. With these assumptions, the optimal user is selected as follows

$$p = \arg \max_k \{\alpha g_k(t)\} \equiv \arg \max_k \{g_k(t)\}. \quad (6.16)$$

Hence, if QoS is not an issue, our SLA based scheduling algorithm is the maximum throughput scheduling (MTS).

Case II: Minimum Penalty Scheduling (MPS): Here, let us assume that the only goal of the system is to deliver QoS to the users, and the system throughput is not important. In this case $\alpha_n = 0$ for all the users. Also, we assume that the penalty function is the same for all users, i.e., $f_k[\cdot] \equiv f[\cdot]$. Then, the SLA based scheduling process will be:

$$p = \arg \max_k \{f[C_k(t-1) + r_k] - f[C_k(t-1) + r_k - g_k(t)]\}. \quad (6.17)$$

We expect this scheduling algorithm (we call it minimum penalty scheduling-MPS) to support QoS for wireless networks.

Now, let us assume that $g_n(t) = 1$, as in wireline networks. In this case, MPS decision is

obtained by

$$p = \arg \max_k \{f[C_k(t-1) + r_k] - f[C_k(t-1) + r_k - 1]\}. \quad (6.18)$$

Now we move ahead to show that for , some special case, if the penalty function $f[.]$ is a positive, continues, increasing and convex function, then the minimum penalty scheduling (MPS) is equivalent to MCS. To this end, we start with the following lemma.

Lemma 6.2.2. *If $f(.)$ is a convex function, then for any $x \geq z \geq y \geq w$ when $x+w = z+y$, we have:*

$$f(z) + f(y) \leq f(x) + f(w)$$

Proof. It is obvious that

$$0 \leq a \triangleq \frac{z-y}{z-w} \leq 1 \Rightarrow$$

$$y = aw + (1-a)z.$$

Moreover, because of convexity of $f[.]$:

$$f(y) \leq \frac{z-y}{z-w}f(w) + \left(1 - \frac{z-y}{z-w}\right) f(z) \quad (6.19)$$

Also,

$$0 \leq b \triangleq \frac{z-y}{x-y} \leq 1 \Rightarrow$$

$$z = bx + (1-b)y$$

However,

$$x + w = z + y \Rightarrow a = b \Rightarrow$$

$$f(z) \leq \frac{z - y}{z - w} f(x) + \left(1 - \frac{z - y}{z - w}\right) f(y) \quad (6.20)$$

Adding equations (6.19) and (6.20) yields:

$$f(z) + f(y) \leq \frac{z - y}{z - w} f(x) + \frac{z - y}{z - w} f(w) + \left(1 - \frac{z - y}{z - w}\right) f(z) + \left(1 - \frac{z - y}{z - w}\right) f(y) \Rightarrow$$

$$f(z) + f(y) \leq f(x) + f(w)$$

□

Corollary 6.2.3. *For convex function $f[.]$ and $x \geq y$, we have:*

$$f(y) - f(y - \Delta) \leq f(x) - f(x - \Delta)$$

Proof. No matter if $x - \Delta \geq y$ or $x - \Delta < y$, using the previous fact, we have:

$$f(x - \Delta) + f(y) \leq f(x) + f(y - \Delta) \Rightarrow$$

$$f(y) - f(y - \Delta) \leq f(x) - f(x - \Delta)$$

□

Now we move to prove the main theorem which says

Theorem 6.2.4. *If $f[.]$ is a positive, continuous, increasing and convex function, then the minimum penalty scheduling (MPS) with $g_n(t) = 1$ is similar to MCS (mentioned in Section*

6.2.2) as follows:

$$p = \arg \max_k \{C_k(t-1) + r_k\}$$

Proof. With wireline assumption, the MPS is based on:

$$p = \arg \max_k \{f[C_k(t-1) + r_k] - f[C_k(t-1) + r_k - g_k(t)]\}. \quad (6.21)$$

Using the previous corollary we have:

$$p = \arg \max_k \{C_k(t-1) + r_k\}$$

which is similar to MCS. □

6.2.6 Maximum Income Dynamic Programming Scheduling (MIDPS):

Optimal Solution

Background: Dynamic Programming

In this section, we review the concept of Dynamic Programming (DP) and mention some of its relevant results. DP is used to find the long-term optimal scheduling policy in this work. The DP problem is categorized into two main categories; Finite Horizon (FH) and Infinite Horizon (IH). We first summarize the DP for FH, and then move to IH.

Let $\{X_t\}_{t=0}^{N-1}$ be a discrete time process, taken from a countable state space, S_k , which can be denoted by a set of non-negative integers $\{0, 1, 2, \dots, K-1\}$. At each time instant $k \in \{0, 1, 2, \dots, N-1\}$, we are required to choose an “action” μ_k , $\mu_k \in M$, where M is the given set of all possible actions. The dynamic of the system can be stated as

$$X_{k+1} = f_k(X_k, u_k, w_k), \quad (6.22)$$

where u_k is a control element taken from a space C_k and depends on the present state, and the random noise (disturbance) w_k is taken from another space D_k , and is denoted by a probability distribution $P_k(\cdot|X_k, u_k)$. This distribution depends on X_k and u_k , but not on the previous values of this random variable w_0, w_1, \dots, w_{k-1} . We define a policy, π , as a rule for choosing the sequence of actions μ_k , for $k \in \{0, 1, 2, \dots, N-1\}$. The policy is called admissible, if each action $\mu_k(X_k)$ belongs to the admissible set of control elements.

If at time k we are in state X_k and choose an action μ_k , a cost of $g_k(X_k, \mu_k, w_k)$ is incurred. The cost function $g(\cdot)$ is a mapping of the aggregate set $\{0, 1, 2, \dots, K-1\} \times M \mapsto \mathbf{R}$, where \mathbf{R} denotes the set of real numbers. Given the initial state X_0 , we want to find the optimal policy π^* that minimizes the overall cost function; i.e.

$$J_{\pi^*}(X_0) = J^*(X_0) = \min_{\pi \in \Pi} E_{w_k} \left\{ g_N(X_N) + \sum_{k=0}^{N-1} g_k(X_k, \mu_k(X_k), w_k) \right\} \quad (6.23)$$

$J^*(\cdot)$ is a function that assigns each initial state to an optimal cost and is called optimal cost function or optimal value function.

The theory of dynamic programming uses the principal of optimality, which states that if at time instant i we are at state X_i , and assuming that $\pi^* = \{\mu_0^*, \mu_1^*, \dots, \mu_{N-1}^*\}$ is the optimal policy, then the truncated policy $\pi^* = \{\mu_i^*, \mu_{i+1}^*, \dots, \mu_{N-1}^*\}$ is optimal in the sense that it minimizes the "cost-to-go" from time i to time N obtained from (6.23) by evaluating the cost per step from time $k = i$ to $N-1$, rather than $k = 0, \dots, N-1$.

The DP algorithm is stated as [144]:

Theorem 6.2.5. *For any initial state X_0 , the optimal cost $J^*(X_0)$ is equal to the cost of the last step of the backward problem stated as*

$$J_N(X_N) = g_N(X_N),$$

$$J_k(X_k) = \min_{u_k, w_k} E \{g_k(X_k, u_k, w_k) + J_{k+1}(f_k(X_k, u_k, w_k))\} \quad (6.24)$$

Moreover, if $u_k^* = \mu_k^*(X_k)$ minimizes the right hand side of (6.24) for each X_k and k , the policy $\pi^* = \{\mu_0^*, \mu_1^*, \dots, \mu_{N-1}^*\}$ is optimal.

The IH problem is different from the FH problem in the sense that the number of stages is not finite (instead of N we have ∞), and also the system is stationary; that is the noise distribution, the cost per stage, and the dynamic of the system do not depend on k (do not change with time). In other words, the action taken at time k depends only on the state of the process at time k . Therefore, a stationary policy is, in effect, a mapping f from the state space to the action space; i.e., given the current state we can determine the current action uniquely. Such a process is called a Markov Decision Process (MDP) [144–146]. The stationary policy can be considered admissible if it is in the form $\{\pi, \pi, \dots, \pi\}$. The IH problem can be classified in four principal classes, where we consider one of them in this chapter, which is "discounted problems with bounded cost per stage". If the state space is finite or countable, then by a theory in real analysis [147], it is known that any mapping defined over such a finite or countable space is bounded.

Given the evolution of the process $\{X_k\}_{t=0}^{\infty}$, the dynamic programming tries to choose $\{\mu_k\}_{k=0}^{\infty}$ such that

$$J^{\pi}(i) \triangleq E_i^{\pi} \left[\sum_{k=0}^{\infty} \beta^k g(X_k, \mu_k) \right], \quad (6.25)$$

is minimized. Here E_i^{π} denotes the expectation under the policy π , $s_0 = i$ is the initial state, and $\beta > 0$ is the discounting factor. The above cost reflects the fact that while choosing the action μ_k at time k , we would like to take into account the effect of this action on the future. How much into the future we wish to look before taking any action can be controlled by choosing an appropriate value for β . If $\beta < 1$, the use of the discount factor is motivated by the fact that a cost to be incurred in the future is less important than one incurred at the present time instant. In general, if no restrictions are placed on the nature of the set of allowable policies, the action could in principle depend on the history of the process up to and including the present time.

In order to ensure the existence of the expected infinite horizon discounted cost, it suffices to have a uniformly bounded cost function $g(\cdot)$ and $0 < \beta < 1$. Note that, for stationary policies, given the current state, the current action can be taken uniquely, regardless of the time instant.

Then the optimal discounted cost will be

$$J(i) = \min_{\pi} J^{\pi}(i), \quad i \geq 0, \quad (6.26)$$

assuming that the minimum exists, where the minimization is over *all* policies π 's. Let π^*

denote an optimal policy which achieves the minimum in (6.26). Then

$$J^{\pi^*}(i) = J(i), \quad i \geq 0. \quad (6.27)$$

The main result for discounted infinite horizon DP problem states that [144, 145]

Theorem 6.2.6. *If $p_{ij}(u)$ is the probability of going from state i to state j , when the action u is taken, then the sequence*

$$J_{k+1}(i) = \min_{u \in U(i)} \left\{ g(i, u) + \beta \sum_{j=0}^{K-1} p_{ij}(u) J_k(j) \right\}, \quad i = 0, 1, \dots, K-1 \quad (6.28)$$

converges to the optimal costs $J^(i)$ for all i , starting from arbitrary initial conditions $J_0(0), J_0(1), \dots, J_0(K-1)$. Moreover, the optimal costs $J^*(0), J^*(1), \dots, J^*(K-1)$ uniquely satisfy the Bellman's equations:*

$$J^*(i) = \min_{u \in U(i)} \left\{ g(i, u) + \beta \sum_{j=0}^{K-1} p_{ij}(u) J^*(j) \right\}, \quad i = 0, 1, \dots, K-1 \quad (6.29)$$

The algorithm stated by the iterative equation (6.28) is called *the value iteration*. It requires an infinite number of iterations. However, in practical applications it terminates finitely when the change in the cost value is below a threshold. The value iteration can be presented by a stationary policy μ , where the costs $J_\mu(0), \dots, J_\mu(K-1)$ are the unique solutions of the equation

$$J_\mu(i) = g(i, \mu(i)) + \beta \sum_{j=0}^{K-1} p_{ij}(\mu(i)) J_\mu(j), \quad i = 0, 1, \dots, K-1 \quad (6.30)$$

Furthermore, given any initial condition, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = g(i, \mu(i)) + \beta \sum_{j=0}^{K-1} p_{ij}(\mu(i)) J_k(j), \quad i = 0, 1, \dots, K-1 \quad (6.31)$$

converges to the optimal cost $J_\mu(i)$ for each i .

An alternative to the value iteration is called *policy iteration*, which starts from a stationary policy μ^0 and generates a sequence of new policies μ^0, μ^1, \dots . The policy evaluation step is given by

$$\mu^{k+1}(i) = \arg \min_{u \in U(i)} \left\{ g(i, u) + \beta \sum_{j=0}^{K-1} p_{ij}(u) J_{\mu^k}(j) \right\} \quad .i = 0, 1, \dots, K - 1 \quad (6.32)$$

It is proven in [144] that the policy iteration is an improving sequence of policies and terminates with an optimal policy.

MIDPS

The algorithms presented in the previous sections, maximize the total income locally. In this section, the objective is to globally maximize the system income. In order to do so, dynamic programming algorithms are used to predict the future to make the decisions at the present time. In this framework, the optimization can be done within a finite horizon or infinite horizon [144]. We focus on the infinite horizon problem since it provides the steady state policy which is independent of time. Define the expected total income as follows:

$$D = E \left\{ \sum_{t=0}^{\infty} \beta^t \left[\sum_{n=1}^N D_n(t) \right] \right\},$$

where $0 < \beta \leq 1$ is the discount factor to keep the total income bounded.

For simplicity, we assume that $I_n(t) = 1$ for all users. Then the credit update equation is given by

$$C_n(t + 1) = C_n(t) + r_n - g_n(t)Y_n(t).$$

Let us define

$$\begin{aligned}\underline{C}(t) &= \left[C_1(t), C_2(t), \dots, C_N(t) \right]^T \\ \underline{Y}(t) &= \left[Y_1(t), Y_2(t), \dots, Y_N(t) \right]^T \\ \underline{g}(t) &= \left[g_1(t), g_2(t), \dots, g_N(t) \right]^T \\ \underline{r} &= \left[r_1, r_2, \dots, r_N \right]^T, \quad \underline{\alpha} = \left[\alpha_1, \alpha_2, \dots, \alpha_N \right]^T.\end{aligned}$$

Then,

$$\underline{C}(t+1) = \underline{C}(t) + \underline{r} - \underline{g}(t) \otimes \underline{Y}(t),$$

where \otimes denoted Hadamard product. In infinite horizon case, we maximize the total system income as follow:

$$\max_{\underline{Y}} \mathbf{E} \left\{ \sum_{s=0}^{\infty} \beta^s \left[\sum_{n=1}^N \{ \alpha_n g_n(t) Y_n(t) - f_n [C_{n+1}(t)] \} \right] \right\}.$$

Now, we define:

$$\begin{aligned}G(\underline{C}(t), \underline{Y}(t), \underline{g}(t)) &\triangleq \\ \sum_{n=1}^N \{ \alpha_n g_n(t) Y_n(t) - f_n [C_n(t) + r_n - g_n(t) Y_n(t)] \}.\end{aligned}$$

Let S_t denote the set of vectors $\underline{X}(t) = [\underline{C}(t), \underline{g}(t)]$ at time t . Therefore, if at time $t = 0$ the system state is $\underline{C}(0) = \underline{C}^i$ and the capacity vector is \underline{g} , then we have

$$\begin{aligned}J^*(\underline{C}^i, \underline{g}) &\triangleq \\ \max_{\underline{Y}} \mathbf{E} \left\{ \sum_{t=0}^{\infty} \beta^t G(\underline{C}(t), \underline{Y}(t), \underline{g}(t)) | \underline{C}(0) = \underline{C}^i, \underline{g}(0) = \underline{g} \right\}.\end{aligned}$$

We would like to obtain the optimal policy $\underline{Y}^*(t) = \mu^*(\underline{C}(t), \underline{g}(t))$ that maximizes the above income function. Note that the scheduler knows the channel capacities at the decision time, and therefore, the channel capacities is a part of the state vector. We can rewrite the optimal income in the form of Bellman's equation for discounted infinite horizon problem [144] as follows:

$$J^*(\underline{C}^i, \underline{g}) = \max_{\underline{Y}} \left\{ G(\underline{C}^i, \underline{Y}, \underline{g}) + \beta \mathbf{E} \left\{ J^*((\underline{C}^i + \underline{r} - \underline{g}\underline{Y}), \underline{g}(t+1)) \right\} \right\}.$$

If we denote the probability of $\underline{g}(t+1) = \underline{g}^k$ by \hat{p}_{g^k} , we obtain:

$$J^*(\underline{C}^i, \underline{g}) = \max_{\underline{Y}} \left\{ G(\underline{C}^i, \underline{Y}, \underline{g}) + \beta \sum_k \hat{p}_{g^k} J^*((\underline{C}^i + \underline{r} - \underline{g}\underline{Y}), \underline{g}^k) \right\}.$$

If the system state is \underline{C}^i and the capacity vector is \underline{g} , the scheduler selects the user that maximizes:

$$\max_{\underline{Y}} \left\{ \sum_{n=1}^N \left\{ \alpha_n g_n Y_n - f_n [C_n^i + r_n - g_n Y_n] \right\} + \beta \sum_k \hat{p}_{g^k} J^*((\underline{C}^i + \underline{r} - \underline{g}\underline{Y}), \underline{g}^k) \right\}. \quad (6.33)$$

The first two terms represent the current income (as seen in the greedy algorithm-MIGS) and the third term represents the income-to-go. In a similar fashion as in Theorem 6.2.1, it can be shown that this maximization is equivalent to selecting the user n in the following

maximization problem:

$$\max_n \{ \alpha_n g_n + f_n [C_n^i + r_n] - f_n [C_n^i + r_n - g_n] + \beta \sum_k \hat{p}_{g^k} J^*(\underline{C}^i + \underline{r} - \begin{bmatrix} 0 \\ \cdot \\ g_n \\ \cdot \\ 0 \end{bmatrix}, \underline{g}^k) \}.$$

6.2.7 System Admission Control Policy, and Pricing

Admission control policy in a network controls the network load by controlling either the number of users or the load of users in such a way that a criteria is met. For example, the performance criteria could be either QoS, or total income. Obviously, for large network load, supporting QoS is a challenge and the network prefers to deal with lower network loads and so not to pay too much penalty for QoS violation. One reasonable scheme for admission control policy in our system is to keep the network load below a threshold in such a way that the relative assigned rate is above some predetermined factor, e.g. 0.95. Another scheme for admission control policy is to keep network load below a threshold such that the system income is above some expected value.

Pricing is another factor that could be considered in this work. The problem of pricing can be observed from two perspectives. In the first perspective, the network earning from the users who have received the service (α_n) and the penalty function could be adjusted based

on the reserved rate by that user. It has economical justification that whoever requires more services (higher r_n) needs to pay more (higher α_n), and penalize the network more for the lack of service (higher order for f_n). Another aspect of pricing relates to the problem of supply and demand. In other words, when the load of the network is low, the system administrator, might be willing to decrease the prices (lower α_n) to encourage the users to use the system (require higher r_n), while in the case of heavy loads (high ρ), the prices might be adjusted accordingly.

6.3 QoS-Provisioned Channel Allocation for OFDMA

The main goal of this Section is to allocate a subset of OFDM subcarriers to each user such that the QoS is satisfied for each user and at the same time the overall network throughput is maximized. The allocation is performed based on each users's users demands and channel conditions. The scheduler has to achieve this purpose with reasonable performance and low complexity. The concept of revenue maximization is used in an OFDMA system, in which the allocation of subcarriers to different users is performed in such a way that the overall network revenue is maximized. As a result, throughput is maximized while QoS is maintained above the required level.

6.3.1 OFDMA System Model

A single cell multiuser OFDMA system with N users and M subcarriers is considered in this chapter (see Figure 6.4). In each OFDM block period, the network can assign any subset of subcarriers to each user. The maximum achievable rate per unit bandwidth at the m^{th} subcarrier for the n^{th} user is given by

$$g_n^m = \log_2(1 + |H_n^m|^2 P_n^m / (N_n^m \Gamma)), \quad (6.34)$$

where H_n^m is the n^{th} user channel frequency response at the m^{th} subcarrier, P_n^m is the transmit power, N_n^m is the noise power, and Γ is the SNR gap [77]. We assume that the base station has the knowledge of channel condition for each user. In other words, g_n^m 's are known in every time slot [57].

Traffic from different users is directed to their assigned queues and each queue is served according to its users QoS. Let r_n denote the rate reserved by user n ($n = 1, 2, \dots, N$), i.e., assigned to queue n . It is well known that QoS of a user can be translated to a minimum guaranteed rate (i.e., r_n) through the notion of effective bandwidth [16]. We assume that the queues of all users are back-logged, so they have packets to transmit at all times. Also, we assume that at each time slot a subcarrier can be assigned to only one user. We denote the set of subcarriers assigned to user n at time t by $S_n(t)$. Obviously, the total rate assigned to user n at time t is $\sum_{m \in S_n(t)} g_n^m(t)$.

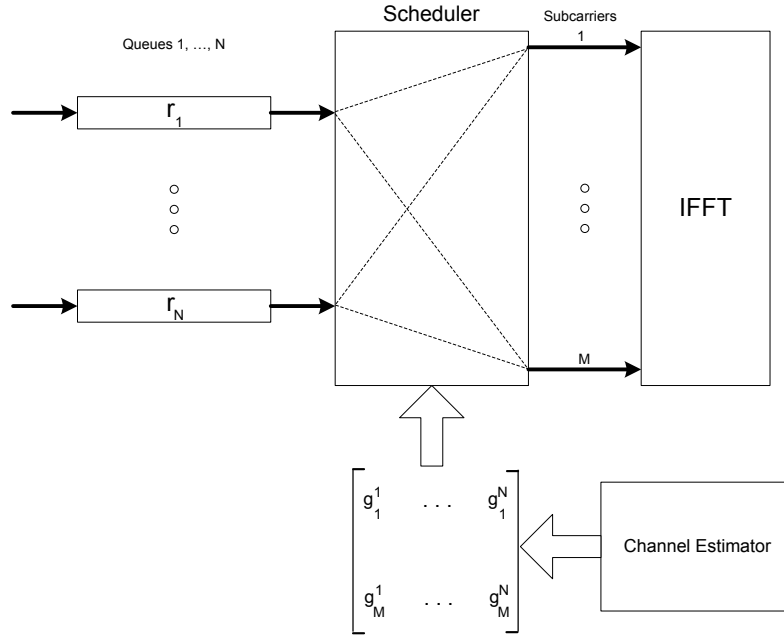


Figure 6.4: Block diagram of the system.

6.3.2 OFDMA Scheduling Algorithms through Subcarrier Allocation

We introduce a revenue model where the network charges the users based on the throughput it provides for users, and is penalized if the QoS defined in SLA for any user is violated. We assume that for the service received by user n at time t , $\sum_{m \in S_n(t)} g_n^m(t)$, the network charges the user by $\alpha_n \sum_{m \in S_n(t)} g_n^m(t)$. Here, α_n is the rate that the n^{th} user pays for the service and is defined in its SLA.

To measure QoS delivered to users, we use the notion of credit. The network assigns a

credit to user n denoted by $C_n(t)$ that evolves as follows:

$$C_n(t) = C_n(t-1) + r_n - \sum_{m \in S_n(t)} g_n^m(t)$$

where r_n and $\sum_{m \in S_n(t)} g_n^m(t)$ are the n^{th} user's reserved rate and received service up to time t , respectively. Credit is introduced in Section 6.2.2. It is a measure of how much service the network has provided to a user [142]; i.e. if the network provides the reserved rate to a user, the user's credit is close to zero. We measure the QoS provided to users by the credits, so if a user has not received the requested QoS, its credit is high (Section 6.2.4). In this case, the network is penalized by $f[C_n]$, where $f[\cdot]$ is a real, positive, convex and continuous function with $f[x] = 0$ for $x \leq 0$.

Let us denote $d_n(t)$ to be the income of the network from user n at time t . Here, we perform scheduling in one OFDM symbol period, and maximize the total income at time t using a greedy algorithm [148]. Therefore, from now on and without loss of generality, we drop the index of time (t) in our discussions. Let us assume that C_n is the credit of user n at the beginning of current OFDM symbol period, the new credit after assigning all subcarriers are $C_n + r_n - \sum_{m \in S_n} g_n^m$. As a result, we obtain:

$$d_n = \alpha_n \sum_{m \in S_n} g_n^m - f \left[C_n + r_n - \sum_{m \in S_n} g_n^m \right] \quad (6.35)$$

The total income of the network is given by $D = \sum_{n=1}^N d_n$.

We assume that in any time slot, the SLA-based scheduler knows r_n 's and C_n 's ($n = 1, \dots, N$) and also g_n^m 's ($n = 1, \dots, N, m = 1, \dots, M$), and it assigns subcarriers to the users such that the total income (D) is maximized. The penalty function shown in (6.11) is

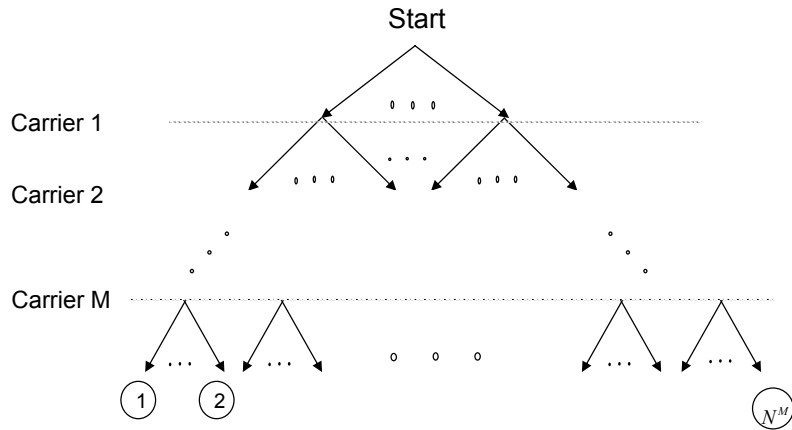


Figure 6.5: Scheduling tree of subcarrier allocation.

used here.

In what follows, we present an optimal and suboptimal scheduling algorithms for subcarrier allocation in OFDMA systems.

Optimal Solution

An exhaustive search among all possible assignments can achieve the optimal solution that maximizes the total income, $D = \sum_{n=1}^N d_n$. There are N different users that can be assigned to M subcarriers. Therefore, the total number of assignments is N^M . The set of all possible assignments can be illustrated by scheduling tree as shown in Figure 6.5. The leaf labelled with 1 shows the choice of allocating all subcarriers to user 1, in leaf N^M all are allocated to user N and in leaf 2, carriers $1 \dots M - 1$ are assigned to user 1 and carrier M to user 2. Other leaves are labelled accordingly. The exhaustive search algorithm evaluates the income for each leaf of the scheduling tree (all possible assignments) and selects the one with the

maximum income.

If the complexity of evaluating d_n is bounded up by L , the complexity of performing the exhaustive search will be LN^{M+1} . Since the computational complexity of this algorithm grows exponentially with the number of subcarriers, exhaustive search may not be a practical solution. However, it still can be used as a performance reference for all of the other algorithms. In the next sections, we propose a number of lower complexity suboptimal algorithms. The algorithms are presented in the increasing order of complexity and optimality.

Sequential Assignment (SA)

As the simplest suboptimal algorithm, we can assign users to carriers, one at a time. Assume that at the k^{th} step, there is a pool of subcarriers left. Now, we introduce a notation of income at step k , d_n^k similar to Equation (6.35) by replacing S_n with S_n^k , the set of subcarriers assigned to user n at the end of the k^{th} step. Also, assume that at the k^{th} step, a carrier denoted by m_k is to be assigned. The best user for this carrier is determined by:

$$\hat{n}_k = \arg \max_n \left\{ \alpha_n \sum_{m \in S_n^{k-1}} g_n^m + \alpha_n g_n^{m_k} - f \left[C_n + r_n - \sum_{m \in S_n^{k-1}} g_n^m - g_n^{m_k} \right] \right\} \quad (6.36)$$

The above algorithm is summarized as:

- Start with a subcarrier, calculate the income for each user and find the best user with the best income according to (6.36).
- Assign the subcarrier to the best user found in the previous step. Remove the carrier

from the pool.

- Proceed to the next subcarrier.

The carrier selection order can be random or fixed. However, we observed through numerical studies that the performances in both cases are similar. In other words, the performance of SA is independent of subcarrier assignment order. While having low complexity, this scheme performs far from the optimal solution, because in this algorithm each subcarrier is selected independent of assignments in the next steps.

Viterbi Algorithm (VA)

As mentioned in earlier sections, each channel assignment is represented by a path on the scheduling tree. We can observe that the scheduling tree can be translated into a scheduling trellis. In scheduling trellis, rows represent the users and columns represent the subcarriers. For simplicity, we can add a dummy initial node and a dummy terminal node to the scheduling trellis. Therefore, every path connecting the initial node to the terminal node through the scheduling trellis is an assignment. For instance, if the carrier m is assigned to the user n , the n^{th} node of the m^{th} column in the scheduling trellis will be in the assignment path. For every path in the scheduling tree, there is one (and exactly one) corresponding path in the scheduling trellis.

Now, the optimal assignment can be translated to finding a path in the scheduling tree with the highest income. If we consider the income of every assignment (path) in the trellis as the weight or length of the path, then the optimization problem is to find the longest

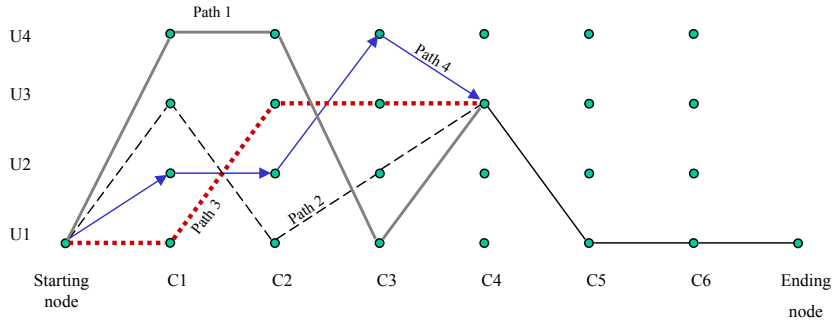


Figure 6.6: Viterbi channel assignment.

path in the trellis. The longest path in a trellis can be obtained by using the low complexity Viterbi Algorithm, if the length of a path has additive property. In other words, by adding a link to a path at step $k + 1$, the length of the link is added to the length of the path at step k . Unfortunately, in this problem, the length of a path does not have the additive property. The length of a path at step k depends on the assignment (S_n^k) through the non-linear function $f[\cdot]$. Therefore, applying the Viterbi algorithm to the very same problem does not necessarily provide the longest (the highest income) path in the scheduling trellis. However, we expect the Viterbi Algorithm to provide a sub-optimal solution.

In the Viterbi Algorithm, at each column, each node computes the cost of the N emerging links, and picks the one with the highest length as the survival path. Hence, at each stage, there are N survival paths and after M steps, the algorithm chooses one path as the final survival path. In each column there are N nodes, and N paths compete at each node. Thus, the complexity of computing a path length is upper bounded by NL , so the complexity of this algorithm is bounded up by MN^3L . Figure 6.6 displays the survival of a path using the Viterbi Algorithm in the scheduling trellis for 4 users and six subcarriers.

Iterative Algorithm (IA)

The sequential assignment algorithm has a very low complexity but its performance is sub-optimal. This is because the assignments in the future steps would change the income for the current assignment. That is, the assignment at each step must not be considered independently from other steps, including the future assignments. As a result, the Viterbi and sequential Algorithms are sub-optimal. In this section, we modify the sequential assignment algorithm to achieve close to optimal solution. For this purpose, we repeat the carrier assignment and refine the set of carriers assigned to each user in order to maximize income until they converge. The iterative subcarrier assignment algorithm is as follows:

1. Assume that all of the subcarriers are assigned initially to different users, by some algorithms (like fixed assignments).
2. Take the k^{th} subcarrier. We want to reassign the this carrier to the locally optimum user. In this step the incomes for all users are checked for a possibly new assignment.

The user that maximizes the total income will obtain this carrier. That is,

$$\hat{n}_k = \arg \max_n \left\{ \alpha_n \sum_{m \in S_n^{k-1}} g_n^m + \alpha_n g_n^{m_k} - f \left[C_n + r_n - \sum_{m \in S_n^{k-1}} g_n^m - g_n^{m_k} \right] \right\}$$

3. Assume that in the initial assignment or previous iterations, the k^{th} subcarrier was

assigned to another user, say \tilde{n} . Then the assignments are updated as

$$\begin{cases} S_n^k = S_n^{k-1} & n \neq \tilde{n} \text{ or } n \neq \hat{n}, \\ S_{\tilde{n}}^k = S_{\tilde{n}}^{k-1} - \{k\} \quad , \\ S_{\hat{n}}^k = S_{\hat{n}}^{k-1} + \{k\} \quad . \end{cases} \quad (6.37)$$

4. Update all incomes for effected users, \tilde{n} , and \hat{n} .
5. Repeat steps 2-4 for the remaining subcarriers until all of the subcarriers are reassigned.
6. Repeat steps 2-5 until the income does not increase anymore.

The above algorithm converges to a fixed point. The reason is that in the above algorithm the income will increase at each step of the algorithm. Since the total income has an upper bound which is the one for the optimal full search assignment, this algorithm will converge to a fixed point which may not be the globally optimum assignment. Depending on the initial assignments, it will converge to a local or the global optimum. In order to improve the performance, we start the iteration from different and random initial points, and then pick the assignment with the maximum overall income. As we increase the number of random initial points, the probability that the algorithm achieves the globally optimal assignment increases. In order to speed up the convergence we try reassignment of carriers in random order.

We will show through numerical studies that the achieved performance of this algorithm is close to the optimal solution with much lower complexity. However, its complexity is

higher than that of sequential algorithm or Viterbi solution. If we assume that we repeat the iterative algorithm from Q different initial points, and the average number of iterations to reach a fixed point is P , and the average complexity of evaluating the income for one user is L , the overall complexity of this scheme is $QPNML$. Considering typical numbers for N and M , this value is much less than LN^{M+1} for the exhaustive search method.

Subcarrier Clustering

We observed that there is a trade-off between the complexity and the performance of the algorithms. One way to reduce the complexity of subcarrier assignments is to bundle the subcarriers in a cluster. For instance, if the system has M subcarriers, and if each cluster has k subcarriers, then we obtain M/k clusters. The capacity matrix of the new system can be easily derived by adding the rows of subcarriers in each cluster, and it will be a matrix with dimensions of $N \times M/k$. Now, every one of the algorithms mentioned in this chapter can be used for the new system, by allocating clusters rather than subcarriers. By clustering, we can reduce the complexity, however, it is clear that this process degrades the system performance.

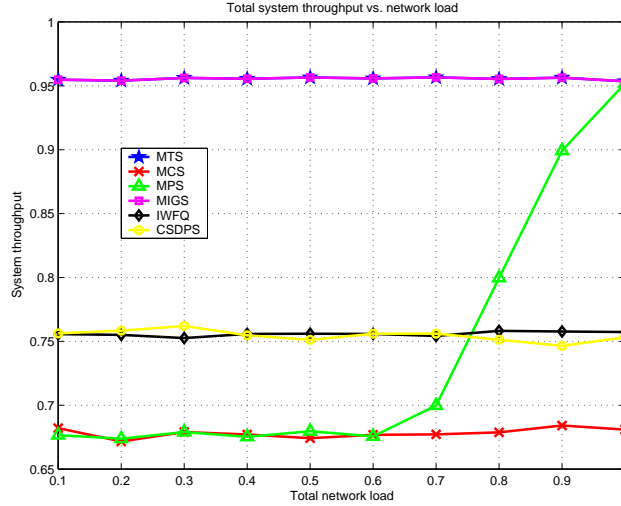


Figure 6.7: Throughput versus network load

6.4 Performance Results

6.4.1 Simulation Results for SLA-Based Scheduling

In order to evaluate the performance of our algorithms, we have simulated a single-cell wireless system where users are randomly distributed. We assume that path loss and shadow fading are compensated by a power allocation mechanism and the channel follows a Rayleigh fading distribution. By considering the same noise level at all receivers, the received signal power also follows a Rayleigh distribution. We have quantized the Signal to Noise Ratio (SNR) of each link into four distinct levels, and for each SNR level, we have calculated the channel capacity according to Equation (6.6). The quantized levels of channel capacities and their probabilities are $\{1.0, 0.6, 0.4, 0.2\}$ and $\{0.43, 0.24, 0.19, 0.14\}$, respectively.

If R_n is the assigned rate to user n (or the proportional time assigned to user n), and r_n

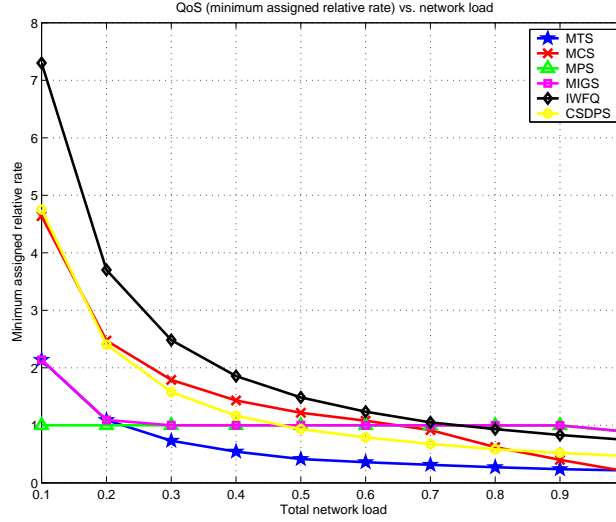


Figure 6.8: Minimum assigned relative rate versus network load

is the reserved rate by that user, we define the *minimum assigned relative rate* over all users to be:

$$\eta = \min_n \left\{ \frac{R_n}{r_n} \right\}. \quad (6.38)$$

This value can be considered as the measure of QoS; to support QoS for all users, we want $\eta \geq 1$.

First, we present the simulation results for MIGS and compare its performance with MTS, MCS, MPS, IWFQ and CSDPS for a system with four users. The reserved rates of the four users are $\rho[0.1, 0.2, 0.3, 0.4]$, where $0 \leq \rho \leq 1$ is the network load (the sum of the reserved rates is ρ). Also, we assume that $\alpha = 1000$ for all users.

Throughput, minimum assigned relative rate (η), and total income are plotted in Figures 6.7-6.9, respectively. The penalty function in the simulations is selected as in (6.11) with $\gamma = 1$. The horizontal axis in all these figures shows the network load (ρ).

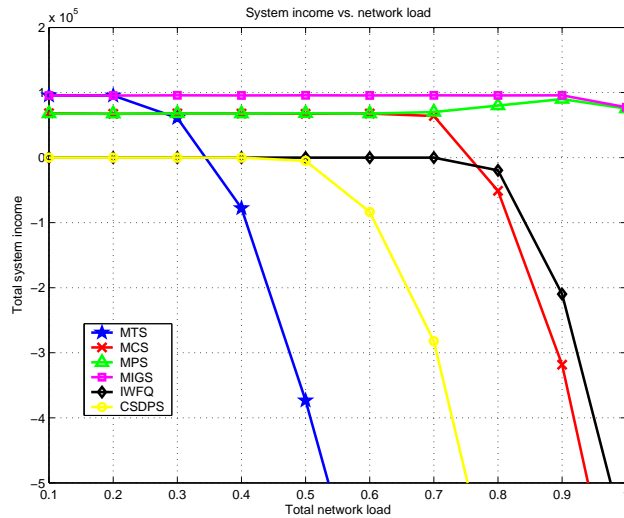


Figure 6.9: Total income vs. network load

As illustrated in Figure 6.7, MTS and MIGS achieve the maximum throughput (the expectation of maximum link capacity, $\mathbf{E}\{\max(g) = 0.96\}$). It should be noticed that the maximum achievable throughput is $\mathbf{E}\{\max(g) = 0.96\}$ and therefore, no throughput above this value is plausible. As can be seen from these figures, IWFQ and CSDPS provide a lower throughput compared to MIGS. It is interesting to notice that the throughput of IWFQ and CSDPS are almost equal. The reason is that in both of these schedulers, if the link capacity of a specific user is below a threshold which is the same for both schemes, the scheduler does not schedule that user. MCS achieves a flat throughput which is equal to the average link capacity ($\mathbf{E}(g) = 0.68$). At low network loads, MPS tries to satisfy each user with its requested bandwidth; that is, throughput is minimally allocated to satisfy each user. As network load increases, the system throughput increases and it approaches to that of MTS and MIGS.

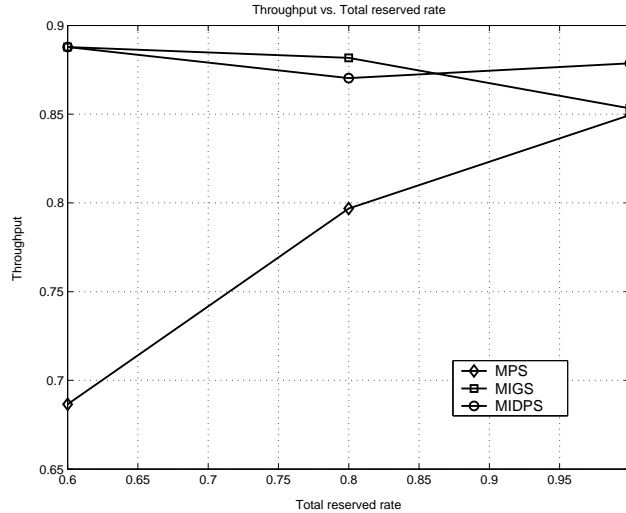


Figure 6.10: Throughput of MIDPS, MIGS, and MPS vs. network load

As illustrated in Figure 6.8, at low network loads all the algorithms support QoS. However, as the network load increases, MPS and MIGS try to maintain QoS for all users, while MTS and CSDPS fail to do so. This result is expected, since MTS and CSDPS are not designed to provide QoS. Since MCS does not utilize bandwidth as efficiently as MIGS, it fails at high network loads due to the lack of available channel bandwidth. IWFQ has a mechanism for supporting QoS. However, since its throughput does not go beyond 0.76, it fails to support QoS after this network load. This is a general rule, which is reflected in Remark 6.4.1.

Remark 6.4.1. *If the Maximum achievable throughput by a scheduling scheme is C_{max} , then the scheme fails to support the QoS ($\eta \geq 1$) for all loads above C_{Max}*

Therefore, IWFQ cannot maintain QoS as strongly as MIGS. It is important to notice that it is not possible to find a scheduler to provide QoS for network load of one.

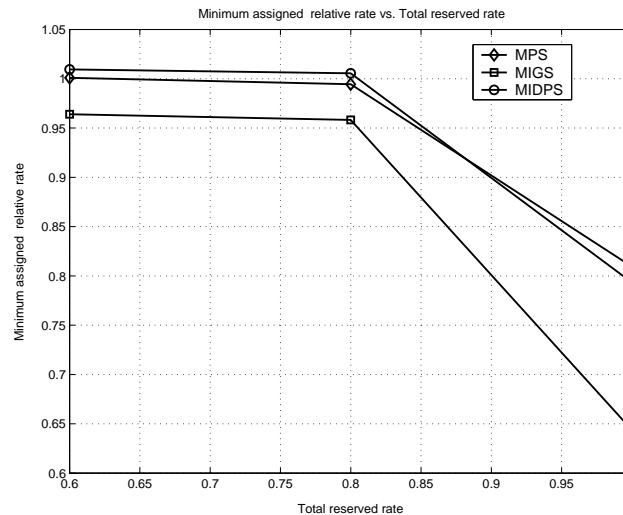


Figure 6.11: Minimum assigned relative rate of MIDPS, MIGS, and MPS vs. network load

As it was mentioned earlier, total income is a combination of system throughput and penalty when QoS requirement is not met. This quantity is shown in Figure 6.9. MIGS generates the highest income since the throughput and penalty are optimized jointly. The income for MTS, CSDPS, MCS, and IWFQ drop at high network loads, since both all to meet QoS after certain loads. MTS and CSDPS fail to meet QoS at lower network loads compared to MCS and IWFQ, and as a result, their incomes drop faster. Total income for MPS increases as load increases, since it tries to minimize penalty independent of load, while at large loads, the effect of throughput prevails. At large loads, the throughputs grows and increases the income. At high network loads, MPS income approaches that of MIGS, since both achieve similar throughput at high loads.

Next, we evaluate the performance of MIDPS and compare its performance with those of MIGS and MPS (See Figures 6.10, 6.11 and 6.12). However, because of complexity issue of

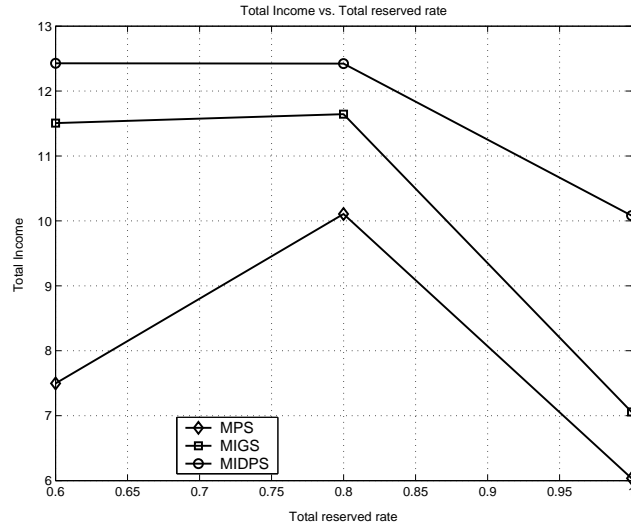


Figure 6.12: Income of MIDPS, MIGS, and MPS vs. network load

DP algorithm, we consider only three users, limit the credits of users to be between -1 and 2, and perform the simulations for three cases where the reserved rates are a:[0.2, 0.2, 0.2], b:[0.2, 0.2, 0.4], and c:[0.2, 0.2, 0.6]. The penalty function is described by Equation 6.11, and $\alpha_1 = 1$, $\alpha_2 = 2$, and $\alpha_3 = 4$. Figures 6.10 and 6.11 show that the system throughput and QoS with MIDPS are as good as the throughput and QoS with MIGS. Therefore, MIDPS can support QoS and provide high system throughput. However, as shown in Figure 6.12, the total income with MIDPS is better than the sub-optimal MIGS. We have to mention that when we increase the range of credits, sub-optimal solution MIGS performs close to the optimal solution MIDPS.

Optimizing of the System Parameters: The system performance, i.e., QoS, system throughput and total system income depend on the values of α_n 's and the penalty functions, $f_n[\cdot]$'s. The system charges the user n based on the value of α_n and the service provided to

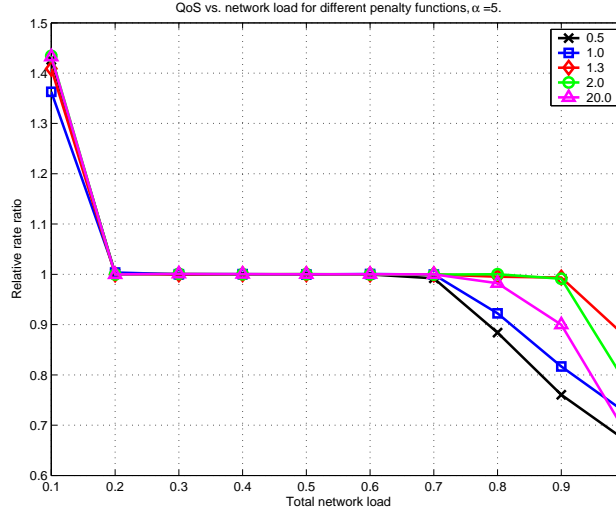


Figure 6.13: QoS versus the network load for different penalty functions, $\alpha = 5$

this user. On the other hand, the user n penalizes the system if its SLA is violated based on the penalty function $f_n[\cdot]$. In this subsection, we would like to investigate the effect of these parameters on the system performance.

As the value of α_n increases, the throughput becomes more significant. For the penalty function, we consider the polynomial function $f_n(x) = x^v$. For high values of v , the penalty function increases for large values of credits, therefore, QoS becomes more significant. As mentioned earlier in this chapter, a scheduler cannot provide QoS above its throughput. Therefore, low throughput implies lack of QoS for large network loads.

In our simulations, we use $a[0.1, 0.1, 0.2, 0.6]$ as the reserved rate assignment. Figure 6.13 displays the relative assigned rate ratio and Figure 6.14 shows the throughput for $\alpha_n = 5$ and for different penalty functions. As shown in these figures, for this value of $\alpha_n = 5$, the exponent of between $1.1 \leq v \leq 3$ is an appropriate value for good performance. Also, there

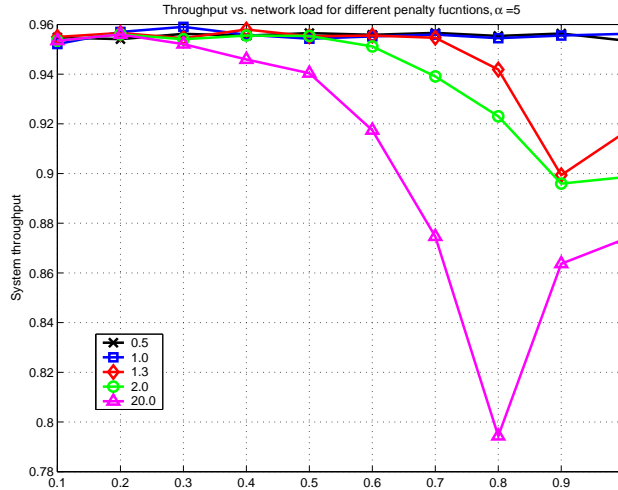


Figure 6.14: Throughput versus the network load for different penalty functions, $\alpha = 5$

is an optimal value of $v^* = 1.3$ that optimizes the system performance in terms of both QoS and throughput. For small values of v , the system does not pay enough attention to QoS and for large values of v , the system throughput reduces and the system fails to support QoS for large network loads. Figures 6.15 and 6.16 show the QoS and throughput for different values of penalty function for $\alpha_n = 500$. We expect the optimal solution to change for this case. As shown in this figure, the optimal value for v moves, i.e., $v^* = 3$.

6.4.2 Simulation Results for OFDMA Channel Allocation

We evaluate the performance of the proposed algorithms in a single cell system where four users ($N = 4$) are randomly distributed in the cell, and each user can be assigned to any of $M = 32$ subcarriers. We consider a multipath channel model with $R = 4$ distinct paths.

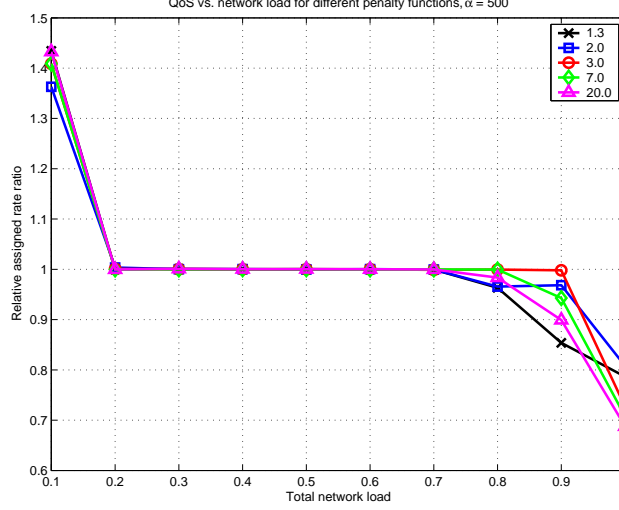


Figure 6.15: QoS versus the network load for different penalty functions, $\alpha = 500$

The channel response for the n^{th} user can be represented as:

$$h_n(t) = \sqrt{G_n} \sum_{r=0}^{R-1} \alpha_n^r \delta(t - \tau_n^r), \quad (6.39)$$

where G_n includes log-normal shadow fading and path loss, τ_n^l and α_n^l denote the l^{th} path delay and fading, respectively. Path delay fading follows a complex Gaussian distribution, so the received signal amplitude has a Rayleigh distribution. The baseband channel frequency response can be represented simply by the Fourier transform of $h_n(t)$ sampled at the carrier frequency, $m f_c$, where f_c is the subcarrier separation:

$$H_n^m = \sqrt{G_n} \sum_{l=0}^{L-1} \alpha_n^l e^{-j2\pi m f_c \tau_n^l}. \quad (6.40)$$

We assume that the path loss and shadowing for different paths are the same, and any difference can be absorbed in fading coefficients.

For performance evaluation, we consider the sequential assignment (SA), Viterbi Algo-

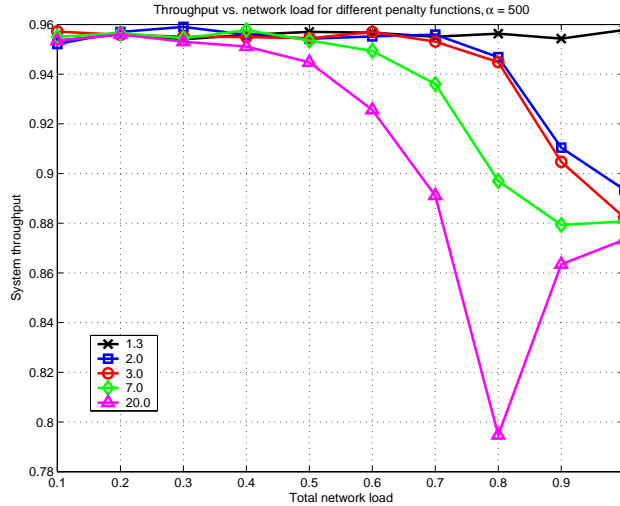


Figure 6.16: Throughput versus the network load for different penalty functions, $\alpha = 500$

rithm (VA), iterative algorithm (IA), and fixed assignment (FA). In the fixed assignment, we assume that the network assigns a set of subcarriers to each user for the whole duration of the simulations. In the other approaches, the subcarrier assignment is performed for one time slot, and it changes in every time slot. Obviously, FA and the exhaustive search have the lowest and the highest complexities, respectively.

Our numerical results reveal that the IA performs close to the optimal exhaustive search algorithm. To see this, we compare the performance of both approaches through simulations. Figure 6.17 shows the cumulative distribution functions (CDF) of the total income for the exhaustive search algorithm along with IA with 20 and 80 iterations. As shown in this figure, the performance of IA is close to the optimal one even with 20 iterations. Therefore, from now on, we can consider IA is as the reference algorithm.

The total throughput versus the network load is shown in Figure 6.18. As it is illustrated

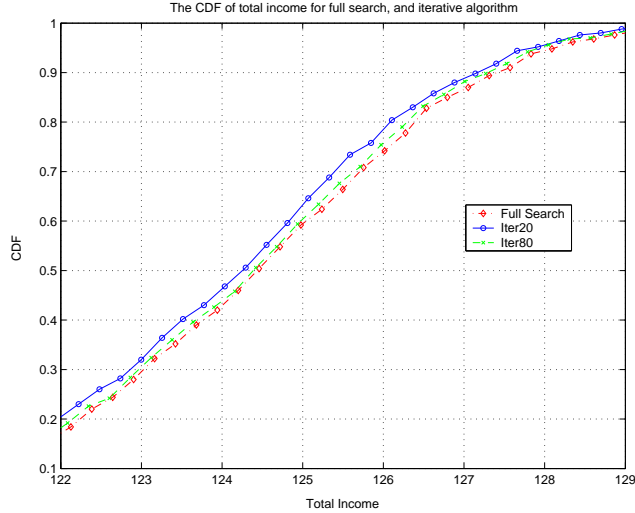


Figure 6.17: CDF's of the optimal exhaustive search algorithm, iterative algorithm with 20 and 80 iterations

in this figure, the IA achieves the maximum throughput (the expectation of maximum link capacity, $E\{\sum_m \max(g_n^m)\}$). VA achieves close to the maximum throughput, but is outperformed by IA. FA achieves close to the average capacity of the system, $E\{g_n^m\}$. At low network loads, SA achieves the maximum throughput. However, its performance drops very fast as load is increased. This is because this algorithm assigns the subcarrier independently; therefore, the assignments at the early stages of the algorithm limit performance of the later stages.

We use the minimum assigned relative rate defined in 6.38 to present QoS. This value is displayed versus the network load in Figure 6.19. Again IA and VA satisfy QoS requirement for almost all loading values while the SA and FA fail to meet the QoS requirement for large loadings.

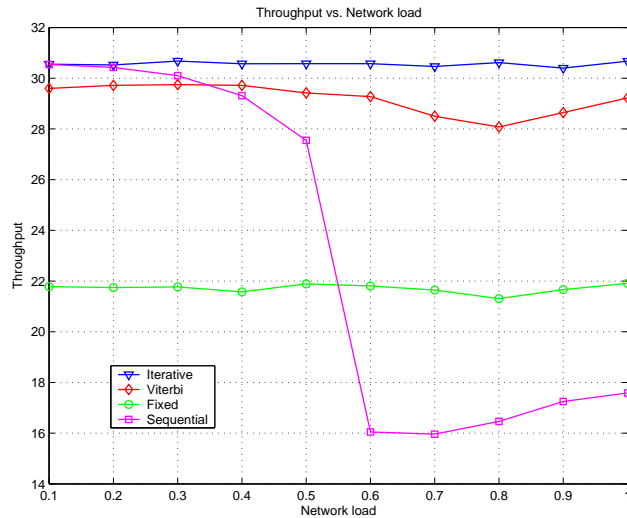


Figure 6.18: Throughput vs. load

The total income is depicted versus network load in Figure 6.20. As shown in this figure, the income of IA is higher than that of the other algorithms since IA provides the highest throughput and it supports the QoS, thus, the penalty of IA for violating the QoS is the lowest. As shown in this figure, VA provides a suboptimal performance. The performance of SA and FA drop significantly as the network load increases.

6.5 Summary of the Chapter

In the first part of this chapter we proposed Service Level Agreement (SLA) based scheduling schemes. We introduced a notion of income maximization where throughput is the objective of maximization with the constraint that the scheduler is penalized when the QoS or SLA is violated. We proposed a greedy approach and a dynamic programming approach to solve

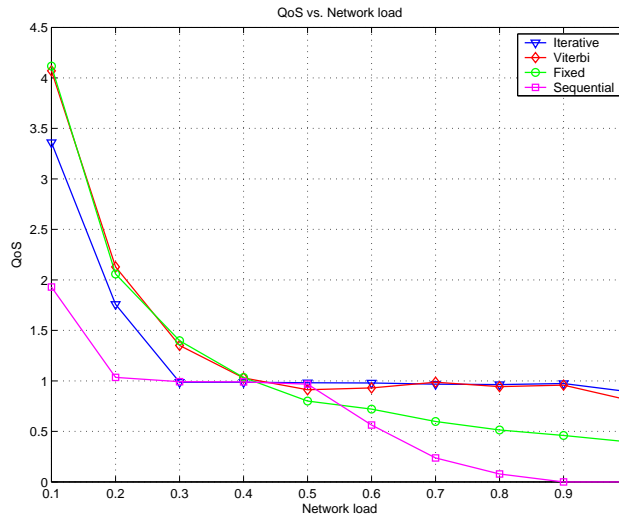


Figure 6.19: Worst case actual to desired throughput vs. load

the problem. Our results show that the performance of the algorithm is superior to cases where only throughput or QoS is considered in the scheduling process.

In the second part of the chapter, we presented scheduling algorithms that maximize OFDMA system throughput for QoS sensitive users, using the notion of revenue maximization. The OFDM subcarriers were allocated to different users based on their capacities, users' required rates, and the total income of the system. We have proposed low complexity sub-optimal scheduling algorithms to reduce the computational complexities.

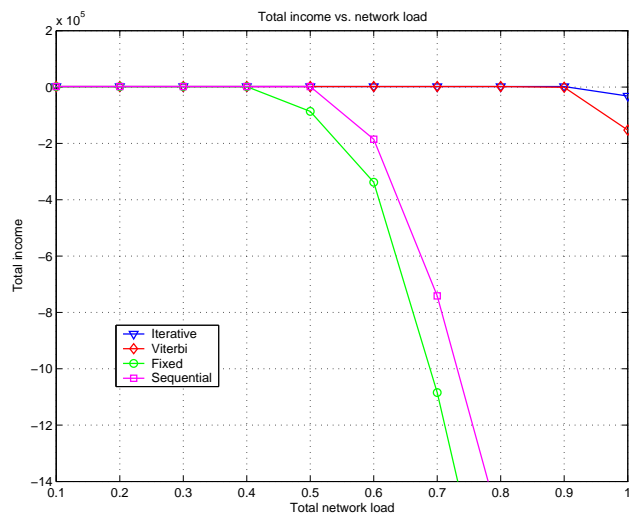


Figure 6.20: Total revenue vs. load

Chapter 7

Conclusion and Future Works

7.0.1 Summary

The main focus of this dissertation was on the development of efficient algorithms for increasing the performance of OFDM systems via power control, multiple transmit and receive antenna, and provide the required QoS's to the users of a network who share a Wireless Media (WM).

In Chapter 2 we outlined the basic concepts of OFDM and explained some of the existing problems of an OFDM system. The problem of equalizing the performance of different subcarriers, increasing the capacity of the system, high Peak to Average Power Ratio (PAPR) of OFDM systems, and different applications of multiple transmit and receive antenna in a wireless system was described.

In Chapter 3, we considered iterative joint power control and beamforming, both in frequency-domain and in time-domain for OFDM wireless networks. We could achieve the

following goals, simultaneously:

- The performance of all of the subchannels of an OFDM receiver are very close to each other.
- The SINRs at all subchannels of all mobiles are at least equal to a pre-defined threshold value.
- The total network power to achieve the above goals is minimized.

To reduce the complexity of the OFDM receivers, we performed the array processing in the time domain and provided an iterative algorithm to distribute the power among subchannels. This scheme could provide a sub-optimal performance in terms of the total network power, with a complexity at the receiver which is about 64 times less (for 128 subchannels and 4 antennas) than the optimal frequency-domain beamforming. This reduction in complexity was achieved by paying a price of having higher total network power for the same target bit error rate. We also considered MMSE time-domain beamforming jointly with power control for practical situations where the angle of arrivals are unknown. We proved that the MMSE time-domain beamforming solution has the same SINR as the MVDR solution. Finally we applied the proposed joint power control and frequency-domain beamforming to the COFDM systems. We observed that an uncoded-OFDM system with the proposed algorithms performs better than the simple COFDM system, a coded system with per subchannel beamforming with equal powers across subchannels and a COFDM system

with per user power control and per subchannel beamforming. If the proposed algorithm is applied to COFDM, the BER is improved for moderate and high SINRs.

In Chapter 4, we proposed iterative water-filling solutions for downlink multi-user multi-cell wireless systems where multiple antennas are deployed at both transmitters and receivers. The proposed algorithm assigned multiple independent substreams for each user to increase the maximum achievable data rate for each user and is performed as a distributed scheme. Each transmitter has only the knowledge of interference at its own receiver and the channel response in that link. By simulation, we have observed that the algorithm converges to a fixed point solution where the data rate for each user is locally optimized. We have established a non-cooperative game theoretic analogy for the MIMO/OFDM problem and proposed a Conjecture that the iterative algorithms proposed in this work will converge to the Nash equilibrium saddle point of the game. We have also proposed another iterative algorithm that considers single stream transmission, and tries to maximize the actual transmission data rate of each user by performing transmit and receive beamforming. The performance of the proposed iterative schemes have also been evaluated through numerical analysis. We saw that if the number of co-channel mobiles per cell is increased, it is better to limit the number of streams. For example, if two co-channel mobiles per cell is used, with four transmit antennas, two streams outperforms the capacity of four streams.

Chapter 5 focused on the problem of reducing PAPR of OFDM systems. We tried to make some modification to Golay complementary codes to increase their achievable rates. We introduced the concept of super-Golay codes, which are the extension of Golay Com-

plementary codes to non-equal energy constellations. Specifically we focused on 16-QAM constellation. We constructed a recursive scheme that allowed us to build all of the super Golay codes with a specific size. These codes were derived from non-equal energy constellations and had the same property that Golay complementary codes derived from equal energy constellation had. We formed some structures to obtain these codes from super Golay Codes with half size. This helped us to establish a modified Reed-Muller code that covers the structure for super Golay complementary. The construction started from QPSK Golay sequences, created from the 2nd order cosets of $RM_{2^h}(1, m)$. the future directions of this work is to search for more structures. The more structure we obtain, the bigger number of SGolay codes we can cover and the higher coding rate we can achieve.

The construction started from QPSK Golay sequences which are efficiently created using 2nd order cosets of $RM_{2^h}(1, m)$. Although the information rate is higher than the existing works in this context, it is not still an acceptable rate. Another future direction of this work is to find the possible trade-off between the coding rate and PAPR.

The structure we propose is for general super Golay codes, regardless of the constellation. However, we have focused on 16-QAM constellation for the sake of simulation. This scheme can be generalized to higher order QAM constellations, like 64-QAM which is used in IEEE WLAN standards like IEEE802.11a.

Then, in the second part of this chapter, we introduced the concept of cyclic Golay codes and showed that with appropriate time shaping, and at the discrete domain (critical sampling), they maintain the same level of PAPR as the Golay codes. Moreover, we have shown

that these constitute a super-set of Golay codes and therefore results in higher coding rate. However, the increase in the coding rate was not very significant (22% for $m = 4$). We designed a construction method to find the cyclic shift of any code represented by Boolean algebraic forms. The cyclic shifts of the Golay second order cosets of the first order Reed-Muller codes generated by our construction has a low Hamming and Lee distance. However, we introduced a trade-off between the coding rate and the distance of the code. An OFDM system, with a SNR threshold is introduced that according to SNR of the received signal switches between Golay code, and its cyclic shifts with different orders. We demonstrated that increasing the threshold drops both the BER and the coding rate. If the SNR threshold was set as the lowest value ($-2dB$ in these figures) the system always used the cyclic Golay code. However, if the threshold was set to the maximum value ($20dB$), the Golay code is always used and that is the point of convergence shown in these figures. We observed that the cyclic Golay codes were in general a subset of $RM_{2^h}(r, m)$. So, we proposed two decoding algorithms for these codes. The first one was a generalization of majority logic decoding approach, using the Karnaugh maps. It was discussed that the decoding scheme is an Hamming and Lee distance decoder for generalized $RM_{2^h}(r, m)$. A maximum-likelihood equivalent of the scheme was introduced for soft-decision decoding in complex domain. We also devised a recursive approach that reduces the decoding procedure into the decoding of lower size and lower order codes. We analyzed the complexities of both decoding algorithms in terms of the number of complex multiplications and additions. We have observed that although our algorithms are useful for generalized $RM_{2^h}(r, m)$, their complexities are comparable to the

existing decoding schemes for the second order cosets of first order generalized Reed-Muller codes. We proved that the complexity of the recursive scheme is lower than the one for the non-recursive scheme.

In Chapter 6, we proposed Service Level Agreement (SLA) based scheduling schemes. We introduced the notion of income maximization where the objective was to maximize the throughput of the network with the constraint that the scheduler is penalized when the QoS or SLA is violated. We proposed a greedy short-term approach that considered the optimization problem over one time-slot, and also a dynamic programming approach that tried to solve the problem optimally over a long period of time (infinity in this case). This scheme provided the scheduling decision for each time slot using the fading characteristic of each wireless link. Our results showed that the performance of these algorithms were superior to cases where only throughput or QoS is considered in the scheduling process.

Finally, in this chapter we presented scheduling algorithms that maximize OFDMA system throughput for QoS sensitive users. We used the same notion of revenue maximization to balance throughput optimization and QoS. The OFDM subcarriers were allocated to different users based on their capacities, users' required rates, and the total income of the system. We proposed low complexity sub-optimal scheduling algorithms to reduce the computational complexities. Through simulation, we observed that the iterative sequential assignment, while having significantly lower complexity, can achieve the fair trade-off between the QoS and total throughput through the notion of total income.

7.0.2 Future Works

The works presented in this dissertation could be extended in the following directions

- To devise the suboptimal time domain time-domain MIMO OFDM to increase the overall mutual information with lower complexities
- To prove the Conjecture 4.3.1 presented in Chapter 4, and find the more accurate necessary conditions for this Conjecture.
- To search for more structures, that constitute higher dimension super Golay codes from lower dimension codes. The more structure we obtain, the bigger number of super Golay codes we can cover and the higher coding rate we can achieve.
- To find the possible trade-off between the coding rate and PAPR of super Golay codes. Although the information rate of super Golay codes were higher than the existing works in this context, they didn't present acceptable rates for practical applications.
- To devise systematic encoding and decoding methods for super Golay codes.
- To create a relation between the discrete domain PAPR based on sampling of the OFDM signal, and the PMEPR of the continuous time OFDM signal. The $3dB$ upper bound for the cyclic Golay codes is only applicable for discrete sampled OFDM time domain symbols. However, they are an indication of how well the code performs.
- To find the relation between the PAPR of some extended second order cosets of $RM_{2^h}(1, m)$ (who has PAPR above $3dB$) and the PAPR of their cyclic shifts.

- To devise coding schemes with low PAPR for space-frequency OFDM codes. Many PAPR reduction codes have considered single antenna transmission. In a paper published in [149], we devised a code that achieved full diversity gain over OFDM systems. We plan to extend these codes to those with the same diversity and coding advantage and low PAPR. There are 3 approaches we are following, namely rotation of the symbols, using cosets of first order Reed-Muller codes, and finally row interleaving.
- To apply the framework proposed in Chapter 6 for SLA-based scheduling, to other multiple access protocols like CDMA and FDMA
- To devise a pricing scheme based on demand of network resources and their availability in terms of the overall capacity of the wireless network. The choice of revenue and penalty parameters in defining the total income of the network has some effect on the trade-off between the QoS and the system throughput, and also the income of the network. We plan to choose these parameters based on demand and supply. Another aspect of pricing is the dependence of α_n and f_n on the required rates of users (r_n). It has an economical justification that whoever requires more services (higher r_n) needs to pay more (higher α_n), and penalize the network more for the lack of service (higher order for f_n).
- To devise an admission control policy to adjust the maximum achievable network loads in such a way that the relative assigned rate is above some predetermined factor. Another scheme for admission control policy is to keep network load below a threshold

such that the system income is above some expected value.

- To find an efficient long-term scheduling decision obtained from dynamic programming. The algorithm presented in Chapter 6 is obtained using the value iteration. However, we will seek a decision driven algorithm long-term scheduling
- To prove the convergence of the sub-optimal schemes proposed for channel allocation in OFDMA systems.
- To combine the channel allocation scheme with power control. In the second part of Chapter 6, we assumed that the powers are fixed for each subchannel. However it is possible to allocate powers intelligently, such that the overall throughput is higher, while the required QoS for all users are fulfilled.
- To apply the framework proposed in Chapter 6 to devise scheduling algorithms for mobile Ad-Hoc networks that use multiple antenna transmission.

BIBLIOGRAPHY

- [1] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Communications Magazine*, vol. 38, pp. 54–64, Jan 2000.
- [2] wirelessLAN802.11 standard committee, "IEEE802.11 wireless lan standards," <http://grouper.ieee.org/groups/802/11/>.
- [3] E. T. S. I. (ETSI), "Etsi broadband radio access networks," <http://www.etsi.org/>.
- [4] J. Kruys, "Standardization of wireless high speed premises data networks adaptation techniques in wireless packet data services," *Wireless ATM workshop, Espoo, Finland*, Sep. 1996.
- [5] wirelessLAN802.16 standard committee, "IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems," <http://standards.ieee.org/getieee802/download/802.16-2001.pdf>, *IEEE Std 802.16-2001*, 2001.
- [6] wirelessLAN802.20 standard committee, "IEEE802.20 mobile broadband wireless access (MBWA)," <http://grouper.ieee.org/groups/802/20/>.

- [7] J. G. Proakis, *Digital Communications*. McGraw Hill, third ed., 1995.
- [8] C. Heegard and S. B. Wicker, *Turbo Codes*. Kluwer Academic Publishers, 1999.
- [9] A. Shokrollahi, “LDPC Codes: an Introduction,” <http://www.ipm.ac.ir/IPM/homepage/Amin2.pdf>, Apr. 2003.
- [10] J.B.Cain, J. G.C.Clark, and J.M.Geist, “Punctured convolutional codes of rate $(n-1)/n$ and simplified maximum likelihood decoding,” *IEEE Transaction on Information Theory*, pp. 97–100, Jan 1979.
- [11] G. Ungerboeck, “Trellis Coded Modulation with Redundant Signal Set,” *IEEE Comm. Magazine*, vol. 27, pp. 5–21, February 1987.
- [12] I. S. Reed and G. Solomon, “Polynomial codes over certain finite fields,” *SIAM Journal on Applied Mathematics*, vol. 8, Jan 1960.
- [13] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Prentice Hall, New Jersey, 1995.
- [14] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo-codes,” *IEEE International Conference on Communications, Geneva*, vol. 2, 1993.
- [15] C. Villamizar and T. Li, “IS-IS optimized multipath,” *Internet draft, draft-villaizar-isis-omp-00.tex*, Oct 1998.

- [16] R. Guerin, H. Ahmadi, and M. Naghsineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High-speed Networks," *IEEE Journal of Selected Areas in Communications*, Sep 1991.
- [17] A. Orda, "Routing with end-to-end qos guarantees in broadband networks," *Tech. Rep., Technion, Iseael*.
- [18] S. Abhyankar, R. Roshawal, C. Cordeiro, and D. Agrawal, "On the application of traffic engineering over bluetooth ad hoc networks," *Department of ECECS, University of Cincinnati*, Oct 1998.
- [19] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijaykumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, pp. 150–154, Feb 2001.
- [20] T. Rappaport, *Wireless Communications*. IEEE Press, 1996.
- [21] J. Gibson, *The Mobile Communications Handbook*. IEEE Press, 1996.
- [22] R. Steele, *Mobile Radio Communications*. IEEE Press, 1992.
- [23] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in Fading Environment When using Multi-element Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.
- [24] R. O. Schmidt, "Multiple emitter location and signal parameters estimation," *IEEE Trans. on Antenna and Propagations*, vol. 34, pp. 276–280, March 1986.

- [25] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 34, pp. 1340–1342, Oct 1986.
- [26] R. A. Monzingo and T. Miller, *Introduction to Adaptive Arrays*. Wiley, New York, 1980.
- [27] B. Ottersten, R. Roy, and T. Kailath, "Signal waveform estimation in sensor array processing," *Proc. of the 23rd Asilomar Conference on Signals, Systems, and Computers*, Nov. 1989.
- [28] B. Suard, A. Naguib, G. Xu, and A. Paulraj, "Performance analysis of CDMA mobile communication systems using antenna array," *IEEE Proc. Intn. Conf. on Acoustic, Speech and Signal Processing, ICASSP' 93*, vol. 4, pp. 153–156, April 1993.
- [29] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [30] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communication in rayleigh flat-fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, March 2000.

- [31] B. Hassibi, B. M. Hochwald, A. Shokrollahi, and W. Sweldens, "Representation theory for high-rate multiple antenna code design," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2335–2367, Sept. 2001.
- [32] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment When using Multi-element Antennas," *Bell Labs Tech Journal*, vol. 1, pp. 41–59, Fall 1996.
- [33] E. Biglieri and A. T. G. Taricc and, "Coding and signal processing for multiple-antenna transmission systems: A review," *Seventh International Workshop on Digital Signal Processing Techniques for Space Communications*, Oct 2001.
- [34] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communications employing multi-element arrays," *IEEE Journal Select. Areas Comm. JSAC*, vol. 17, p. 18411852, November 1999.
- [35] L. Zheng and D. N. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, May 2003.
- [36] B. Salzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. Commun. Technol.*, vol. COM-15, pp. 805–813, Dec 1967.

- [37] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 628–634, Oct 1971.
- [38] A. Oppenheim and R. Schaffer, *Discrete -time signal processing*. Prentice-Hall International, 1989.
- [39] J. L.J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. on Commun.*, vol. COM-33, pp. 665–675, Jul 1985.
- [40] T. Keller and L. Hanzo, "Adaptive modulation techniques for duplex OFDM transmission," *IEEE Trans. on Vehicular Technologies Commun.*, vol. 49, pp. 1893–1905665–675, Sep. 2000.
- [41] T. Pollet, M. Bladel, and M. Moeneclaey, "Sensitivity of OFDM systems to carrier frequency offset and wiener phase noise," *IEEE Transaction on Communications*, vol. 43, pp. 191–193, Apr 1995.
- [42] B. Sklar, *Digital Communications: Fundamentals and Applications*. Prentice Hall, 1988.
- [43] A. J. Viterbi, *CDMA, Principles of Spread Spectrum Communication*. Addison Wesley, 1996.

- [44] J. Heiskala and J. Terry, *OFDM Wireless LANs: A theoretical and practical guide*. Sams, 2002.
- [45] A. N. D. Andra and U. Mengali, *Synchronization Techniques for Digital Receivers*. Plenum Press, New York, 1997.
- [46] V. K. Bhargava, D. Haccoun, R. Matyas, and P. P. Nuspl, *Digital Communications by satellite: Modulation, Multiple Access, and Coding*. Kreiger, 1991.
- [47] L. M. C. Hoo, J. Tellado, and J. M. Cioffi, "Dual QoS loading algorithms for multi-carrier systems offering different CBR services," *Proc. of PIMCR*, vol. 1, 1998.
- [48] J. C. de Souza, "Discrete bit loading for multicarrier modulation systems," *PhD Thesis*, May 1999.
- [49] J. A. C. Bingham, J. M. Cioffi, and P. S. Chow, "A practical discrete multitone transmitter loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. On Communications*, pp. 773–775, February 1995.
- [50] D. Hughes-Hartogs, "Ensembled Modem Structure for Imperfect Transmission Media," *U.S. Patent Notes 4679226*, July 1987.
- [51] A. Czylik, "Adaptive OFDM for Wideband Radio Channels," *Proc. of IEEE GlobeCom'96*, vol. 1, pp. 713–718, 1996.
- [52] R. F. H. Fischer and J. B. Huber, "A New Loading Algorithm for Discrete Multitone Transmission," *Proc. of IEEE GlobeCom'96*, pp. 724–728, February 1996.

- [53] S. H. Muller, R. W. Ba, R. F. Fischer, and J. B. Huber, "OFDM with Reduced Peak to Average Power Ratio by Multiple Signal Representation," *Annals of Telecommunications*, vol. 52, no. 1/2, pp. 58–67, 1997.
- [54] C.-L. Liu, "The effect of nonlinearity on a QPSK-OFDM-QAMsignal," *IEEE Trans. on Consumer Electronics*, vol. 43, no. 3, pp. 443–447, 1997.
- [55] M. Sharif, M. G. Alkhansari, and B. H. Khalaj, "On the peak-to-average power of ofdm signals based on oversampling," *IEEE Trans. on Comm.*, vol. 51, pp. 72–78, Jan. 2003.
- [56] J. Tellado, *Multicarrier Modulation with Low PAPR, Application to DSL and Wireless*. Kluwer Academic Publishers, 2000.
- [57] J. Gross and F. Fitzek, "Channel State Dependent Scheduling Policies for an OFDM Physical Layer Using a Binary State Model," *Tech. Rep. TKN-01-009, Telecommunication Networks Group, Technische Universityat Berlin*, Jun 2001.
- [58] D. Mestdagh and P. Spruyt, "A method to reduce the probability of DMT-based transceivers," *IEEE Trans. on Comm.*, vol. 44, no. 10, pp. 1234–1238, 1996.
- [59] J. Chow, J. Bingham, and J. Flowers, "Mitigating clipping noise in multi-carrier systems," *Proceedings ICC97, Montreal, Canada*, pp. 715–719, 1997.
- [60] A. Gatherer and M. Polley, "Controlling clipping probability in DMT transmission," *Proc. Conf. on Signals, Systems and Computers, Pacific Grove, CA*, pp. 578–584, 1997.

- [61] R. V. Nee and A. D. Wild, “Reducing the peak-to-average power ratio of OFDM,” *Proceedings VTC98, Ottawa, Canada*, pp. 2072–2076, 1998.
- [62] T. May and H. Rohling, “Reducing the peak-to-average power ratio in OFDM radio transmission systems,” *Proceedings VTC98, Ottawa, Canada*, pp. 2474–2478, 1998.
- [63] T. A. Wilkinson and A. E. Jones, “Minimization of the peak to mean envelope power ratio of multicarrier transmission schemes by block coding,” *IEEE 45th Vehicular technology Conf., Chicago, IL*, pp. 825–829, Jul. 1995.
- [64] A. Kamerman and A. Krishnakumar, “OFDM encoding with reduced crest factors,” *Symp. On Comm. and Vehicular Technology in the Benelux, Louvain-La-Neuve, Belgium*, pp. 182–186, 1994.
- [65] M. Friese, “Multicarrier modulation with low peak-to-mean average power ratio,” *Electronic Lett.*, vol. 33, pp. 713–714, 1996.
- [66] H. Ochiai and H. Imai, “Block Coding Scheme Based on Complementary Sequences for Multicarrier Signals,” *IEICE Trans. Fundamentals*, pp. 2136–2143, Nov. 1997.
- [67] X. Li and J. Ritcey, “M-sequences for OFDM peak-to-average power ratio reduction and error correction,” *Electronics Letters*, vol. 33, no. 7, pp. 554–555, 1997.
- [68] C. Schurgers and M. B. Srivastava, “A systematic approach to peak-to-average power ratio in OFDM,” *SPIE’s 47th Annual Meeting, San Diego, CA*, pp. 454–464, Jul 2001.

- [69] S. H. Muller and J. B. Huber, "OFDM with Reduced Peak-to-Average Power Ratio by Optimum Combination of Partial Transmit Sequences," *Electronic Lett.*, vol. 33, pp. 368–369, Feb. 1997.
- [70] R. W. Bauml, R. F. H. Fischer, and J. B. Huber, "Reducing the Peak-to-Average Power Ratio of Multicarrier Modulation by Selective Mapping," *Electronic Lett.*, vol. 32, pp. 2056–2057, Oct. 1996.
- [71] P. V. Eetvelt, G. Wade, and M. Tomlinson, "Peak to average power reduction for OFDM schemes by selective scrambling," *Electronics Letters*, vol. 32, pp. 1963–1964, Jul 1996.
- [72] J. Tellado and J. Cioffi, "Peak power reduction for multicarrier transmission," *Proceedings Globecom98, Sydney, Australia*, 1998.
- [73] D. Jones, "Peak power reduction in OFDM and DMT via active channel modification," *Proceedings of 1999 Asilomar Conference, Pacific Grove, CA*, pp. 1076–1079, 1999.
- [74] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," *IEEE Journal of Selected Areas in Comm., JSAC*, vol. 7, pp. 1172–1180, Oct 1989.
- [75] J. T. E. McDonnell and T. A. Wilkinson, "Comparison of Computational Complexity of Adaptive Equalization and OFDM for Indoor Wireless Networks," *Proc. of Personal, Indoor and Mobile Radio Comm. (PIMRC)*, pp. 1088–1091, 1996.

- [76] Y. Wu and W. Y. Zou, "Orthogonal Frequency Division Multiplexing: A Multicarrier Modulation Scheme," *IEEE Trans. on Consumer Electronics*, vol. 41, pp. 392–398, June 1995.
- [77] J. A. C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea Whose Time has Come," *IEEE Communication Magazine*, pp. 5–14, May 1990.
- [78] ETS300-401(1994), "Radio Broadcast Systems: Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers," <http://www.ets.fr/>, 1994.
- [79] S. K. Lai, R. S. Cheng, K. Letaief, and R. D. Murch, "Adaptive Trellis Coded MQAM and Power Optimization for OFDM Transmission," *Proc. of IEEE Vehicular. Tech. Conf.*, vol. 49, 1999.
- [80] H. Sari, G. Karam, and I. Jeanclaude, "Transmission Techniques for Digital Terrestrial TV Broadcasting.," *IEEE Comm. Magazine*, 1995.
- [81] F. R. Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint Optimal Power Control and Beamforming in Wireless Networks Using Antenna Arrays," *IEEE Trans. on Communications*, vol. 46, pp. 1313–1324, October 1998.
- [82] F. R. Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit Beamforming and Power Control for Cellular Wireless Systems," *IEEE Journal of Selected Areas in Communications, Special Issue on Signal Processing for Wireless Communications*, vol. 16, pp. 1437–1450, October 1998.

- [83] F. R. Gantmacher, *The Theory of Matrices*. Chelsea, New York, third ed., 1990.
- [84] J. Zander, "Performance of Optimum Transmitter Power Control in Cellular Radio Systems," *Proc. of IEEE Vehicular. Tech. Conf.*, vol. 41, pp. 57–62, February 1992.
- [85] G. J. Foschini, "A Simple Distributed Autonomous Power Control Algorithm and its Convergence," *Proc. of IEEE Vehicular. Tech. Conf.*, vol. 42, pp. 641–646, November 1993.
- [86] J. Zander, "Distributed Cochannel Interference Control in Cellular Radio Systems," *Proc. of IEEE Vehicular. Tech. Conf.*, vol. 41, pp. 305–311, August 1992.
- [87] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, third ed., 1996.
- [88] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis Coded Modulation with Applications*. Macmillan, 1991.
- [89] G. Ungerboeck, "The State of the Art in Trellis Coded Modulation," *Proc. of the 5th Tirrenia Intl. workshop on Digital Communications, Tirrenia, Italy*, September 1991.
- [90] A. Mehrotra, *GSM System Engineering*. Artech House Inc., Boston, 1996.
- [91] G. G. Raleigh and J. M. Cioffi, "Spatio-Temporal Coding for Wireless Communications," *IEEE Transaction on Communications*, vol. 46, no. 3, pp. 357–366, 1998.
- [92] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Bell Labs Tech Journal, Tech Rep. num. BL0112170-0950615-07TM*, 1995.

- [93] T. L. Marzetta and B. Hochwald, "Capacity of a Mobile Multi-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Transaction on Information Theory*, vol. 45, pp. 139–157, January 1999.
- [94] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading Correlation and its Effect on the Capacity of Multi-Element Antenna Systems," *IEEE Transaction on Communications*, vol. 48, pp. 502–513, Mar 2000.
- [95] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative Water-Filling for Gaussian Vector Multiple Access Channels," *IEEE International Symposium on Information Theory (ISIT2001)*, June 2001.
- [96] M. S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *Transaction on Vehic. Tech.*, vol. 48, pp. 1047–1066, July 1999.
- [97] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input, multiple-output (MIMO) transmission systems," *IEEE Transactions on Information Theory*, vol. 42, Dec. 1994.
- [98] K. K. Wong, R. S. K. Cheng, K. B. Latief, and R. D. Murch, "Adaptive antennas at the mobile and base station in an OFDM/TDMA system," *IEEE Transactions on Information Theory*, vol. 49, Jan 2001.

- [99] S. Serbetli and A. Yener, “Iterative transceiver optimization for multiuser MIMO systems,” *Proceedings of 40th Allerton Conference on Communications, Control and Computing, Monticello, Illinois*, Oct. 2002.
- [100] H. Bölcskei, D. Gesbert, and A. J. Paulraj, “On the capacity of ofdm-based spatial multiplexing systems,” *IEEE Trans. on Communications*, vol. 50, pp. 225–234, February 2002.
- [101] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1990.
- [102] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, 1985.
- [103] J. R. T. Compton, *Adaptive Antennas Concepts and Performance*. New Jersey: Prentice Hall, 1988.
- [104] M. Osborne and A. Rubinstein, *A course in Game Theory*. Cambridge, MA: MIT Press, 1994.
- [105] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1992.
- [106] H. B. Jovanovich, *Linear Algebra and Its Application*. 1988.
- [107] R. van Nee, “OFDM Codes for Peak-to-Average Power Reduction and Error Correction,” *Proc. of IEEE Globecom’96*, pp. 740–744, 1996.

- [108] R. Dinis, P. Montezuma, and A. Gusmao, "Performance Trade-offs with Quasi-Linearly Amplified OFDM through a Two Branch Combining Technique," *Proc. of IEEE Vehicular. Tech. Conf.*, pp. 899–903, May 1996.
- [109] X. Li and L. J. Cimini, "Effects of Clipping and Filtering on the Performance of OFDM," *Proc. of IEEE Vehicular. Tech. Conf.*, pp. 1634–1638, May 1997.
- [110] D. Wulich, "Reduction of Peak to Mean Power Ratio of Multicarrier Modulation Using Cyclic Coding," *Electronic Lett.*, vol. 32, pp. 432–433, Feb. 1996.
- [111] H. Ochiai and H. Imai, "Block Coding Scheme Based on Complementary Sequences for Multicarrier Signals," *Proc. of IEEE Int'l. Conf. on Comm. (ICC)*, 1998.
- [112] K. G. Paterson, "Generalized Reed-Muller Codes and Power Control in OFDM Modulation," *IEEE Trans. on Comm.*, vol. 46, pp. 104–120, Jan. 2000.
- [113] K. G. Paterson and A. E. Jones, "Efficient Decoding Algorithms for Generalized Reed-Muller Codes," *IEEE Trans. on Comm.*, vol. 48, p. 1272, Aug. 2000.
- [114] K. G. Paterson, "Coding Techniques for Power Controlled OFDM," *Proc. of 9th Intl. Symp. on Personal, Indoor and Mobile Radio Comm. (PIMRC '98)*, vol. 2, pp. 801–805, 1998.
- [115] K. G. Paterson and V. Tarokh, "On the Existence and Construction of Good Codes with Low Peak-to-Average Power Ratios," *IEEE Trans. on Information Theory*, vol. 46, pp. 1974–1987, 2000.

- [116] 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 Scheme," *Electronic Lett.*, vol. 47, pp. 2098–2099, Dec. 1994.
- [117] C. Rößing and V. Tarokh, "A Construction of OFDM 16-QAM Sequences Having Low Peak Powers," *IEEE Trans. on Information Theory*, vol. 47, pp. 2091–2093, Jul. 2001.
- [118] J. A. Davis and J. Jedwab, "Peak-to-Mean Power Control in OFDM, Golay Complementary Sequences, and Reed-Muller Codes.," *IEEE Trans. on Information Theory*, vol. 45, pp. 2397–2417, Nov. 1999.
- [119] E. Lawrey and C. J. Kikkert, "Peak to Average Power Ratio Reduction of OFDM Signals Using Peak Reduction Carriers (PRC)," *the Fifth Intl. Symp. on Signal Processing and its Applications. Brisbane, Queensland, Australia (ISSPA '99)*, vol. 542, pp. 737–740, Aug. 1999.
- [120] A. E. Jones and T. A. Wilkinson, "Combined Coding for Error Control and Increased Robustness to System Nonlinearities in OFDM," *IEEE 46th Vehicular Tech. Conf., Atlanta, GA*, pp. 904–908, Apr/May. 1996.
- [121] V. Tarokh and H. Jafarkhani, "On the computation and reduction of the peak to average power ratio in multicarrier communications," *IEEE Transactions on Communications*, vol. 48, pp. 37–44, Jan. 2000.

- [122] M. J. E. Golay, "Complementary Series," *IRE Trans. on Information Theory*, pp. 82–87, Apr. 1961.
- [123] K. G. Paterson, "On the Codes with Low Peak-to-Average Power Ratio for Multi-Code CDMA," *IEEE International Symposium on Information Theory (ISIT)*, pp. 49–49, 2002.
- [124] C. V. Chong and V. Tarokh, "Two constructions of 16-QAM Golay Complementary Sequences," <http://www.mit.edu/vahid>.
- [125] A. J. Grant and R. D. V. Nee, "Efficient Maximum-Likelihood Decoding of Q-ary Modulated Reed-Muller Codes," *IEEE Trans. on Comm.*, vol. 2, pp. 134–136, May 1998.
- [126] K. G. Paterson and A. E. Jones, "Efficient Decoding Algorithms for Generalized Reed-Muller Codes," *IEEE Trans. on Comm.*, vol. 48, pp. 1272–1285, August 2000.
- [127] M. M. Mano, *Digital Design*. Third Edition, Prentice Hall, New Jersey, 2002.
- [128] J. L. Massey, "The Ubiquity of Reed-Muller Codes," *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (Eds. S. Boztas and I. E. Shparlinski)*, New York, Springer, no. 2227, pp. 1–12, 2001.
- [129] W. W. Peterson and E. J. W. Jr., *Error Correcting Codes*. Cambridge: MIT Press, 2 ed., 1972.

- [130] L. Yanyu, Z. Zhifei, and T. Pinghui, “Equivalent bandwidth estimation under delay constrains,” *IEEE International Conference on Communication Technology Proceedings, 2000. WCC - ICCT 2000*, vol. 2, pp. 1465–1469, 2000.
- [131] P. Bhagwat, A. Krishna, and S. Tripathi, “Enhancing Throughput over Wireless LANs Using Channel State Dependent Packet Scheduling,” *Proceedings of IEEE INFOCOM96*, vol. 2, pp. 1133–1140, Mar 1996.
- [132] C. Fragouli, “Controlled Multimedia Wireless Link Sharing via Enhanced Class-based Queuing with Channel-State Dependent Packet Scheduling,” *Proceedings of IEEE INFOCOM’98*, vol. 2, pp. 572–580, Mar 1998.
- [133] Y. Cao, V. O. K. Li, and Z. Cao, “Scheduling Delay Sensitive and Best Effort Traffic in Wireless Networks,” *Proc. of IEEE Int’l. Conf. on Comm. (ICC)*, May 2003.
- [134] H. Rohling and R. Grunheid, “Performance comparison of different multiple access schemes for downlink of an ofdm communication system,” *IEEE 47th Vehicular Technology Conference*, vol. 3, pp. 1365–1369, May 1997.
- [135] W. Rhee and J. Cioffi, “Increase in Capacity of Multiuser OFDM System Using Dynamic Subchannel Allocation,” *Proc. Vehicular Technology Conference (VTC)*, p. 1085–1089, 2000.

- [136] I. Koutsopoulos and L. Tassiulas, "Channel State Adaptive Techniques for Throughput Enhancement in Wireless Broadband Networks," *Proceedings of IEEE INFOCOM*, pp. 757–766, 2001.
- [137] G. J. Pottie, "System design choices in personal communications," *IEEE Personal Communication*, vol. 2, pp. 50–67, Oct. 1995.
- [138] B. McFarland, G. Chesson, C. Temme, and T. Meng, "The 5-UPTM protocol for unified multi-service wireless networks," *IEEE Communications Magazine*, pp. 74–80, November 2001.
- [139] G. Li and H. Liu, "Dynamic resource allocation with finite buffer constraint in broadband OFDMA networks," *Proc. of IEEE Conf. on Acoustics, Speech, and Signal Processing, ICASSP '03*, vol. 4, pp. 197–200, April 2003.
- [140] Y. W. Cheong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in communication*, vol. 17, pp. 1747–1758, Oct. 1999.
- [141] K. Inhyoung, L. L. Hae, K. Beomsup, and Y. H. Lee, "On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM," *IEEE Global Telecommunications Conference*, vol. 6, pp. 3648–3652, 2001.

- [142] A. C. Kam and K. Y. Siu, “Linear Complexity Algorithms for Bandwidth Reservations and Delay Guarantees in Input Queued Switches with no Speedup,” *Proc. International Conference on Network Protocols 98, Austin TX*, pp. 2–11, Oct 1998.
- [143] M. Grossglauser and D. Tse, “Mobility Increases the Capacity of Wireless Adhoc Networks,” *Proceedings of IEEE INFOCOM’01*, Apr 2001.
- [144] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.
- [145] S. M. Ross, *Stochastic Dynamic Programming*. Academic Press, 1983.
- [146] P. P. Varaiya, *Notes on Optimization*. New York: Van Nostrand Reinhold, 1972.
- [147] H. L. Royden, *Real Analysis*. New York: McMillan, 1968.
- [148] M. Alasti, F. R. Farrokhi, M. Olfat, and K. J. R. Liu, “Service Level Agreement (SLA) Based Scheduling Algorithms for Wireless Networks,” *Submitted to International Conference on Communications, ICC*, 2004.
- [149] W. Su, Z. Safar, M. Olfat, and K. J. R. Liu, “Obtaining Full Diversity Space Frequency Codes from Space-Time Codes via Mapping,” *IEEE Trans. on Signal Processing, special issue on Signal Processing for Multiple-Input Multiple-Output (MIMO) Wireless Communications systems, Dec 2002*, vol. 51, pp. 2905–2916, Nov. 2003.