The Pennsylvania State University
The Graduate School
Department of Electrical Engineering

HIGH-SPEED ACCESS OVER COPPER WIRING

A Dissertation in
Electrical Engineering
by
Ali Enteshari

© 2009 Ali Enteshari

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2009
The dissertation of Ali Enteshari was reviewed and approved* by the following:

Mohsen Kavehrad  
Professor in Electrical Engineering  
Dissertation Advisor  
Chair of Committee

John D. Mitchell  
Professor in Electrical Engineering

Ali R. Hurson  
Professor in Computer Science and Engineering

William Waterhouse  
Professor in Mathematics

W. Kenneth Jenkins  
Professor in Electrical Engineering  
Head of the Department of Electrical Engineering

*Signatures are on file in the Graduate School
This dissertation focuses on conceptual design, development and demonstration of a 40 Gigabits per second (Gbps) transmission system for distribution of a digital signal over Ethernet copper cables serving ultra high-speed interconnection in e.g. data centers, local area networks (LAN), consumer multimedia, teleconferencing, telemedicine, and many others. Although the actual implementation may not be feasible at the time of writing this dissertation due to technology limitations and high production costs, we have been trying to address the analysis, modeling and design of sub-systems in a framework leading to a practical implementation of beyond 10G, namely 40Gbps, in a not too distant future with the ever increasing speed of technology advances.

In the commercial market, a new challenge is the extension of fiber into the access network in small business and dense metropolitan areas. It has long been known that a major bottleneck in delivering multimedia services to the computer users is the low-capacity of LANs. With ever-increasing demand for higher capacities, the need for broadband access is transformed from a convenience into a necessity. So far, data communication has been the main driving force behind increased traffic on the communications networks. To keep up with this explosive growth, ultra high-capacity networks were required, and thus optical networks with terabit capacities dominated the network core. To enable the end user to take full advantage of this core, reliable high-speed LAN access is required. Providing service in a broadband access LAN, using a copper cable approach, has the advantages of the network being highly-dependable and cost-effective. This will benefit the providers of service over campus settings like hospitals, industry compounds or universities with facilities spread over several buildings that a quick service upgrade could extend new service offerings. Also, within server farms and data centers, short copper connectors are preferable.
After release of the 10GBASE-T, which supports data rates of 10Gbps up to a distance of 100 meters (for connecting work areas to a telecommunications room), many IEEE members recognized the potential for higher speed and are currently thinking of ways to deliver tens of Giga bits per second over copper cables. Researchers have started to study the technical feasibility, broad market potential, and economic feasibility of speeds beyond 10Gbps over copper. In this dissertation, we evaluate the possibility of 40Gbps data transmission (40GBASE-T) over horizontal balanced CAT-7A cables up to a distance of 50 meters. The objective of IEEE 40GBASE-T is to create an application that is capable of transporting data at a rate of 40Gbps over at least 10 meters of copper cable. Our capacity analysis of this cable shows that 40Gbps transmission is practical up to 50m over this cable at the cost of elaborate digital signal processing and complex coding schemes. Besides, the data center geography statistics show that 80~90% of links are shorter than 45m. Accordingly, by the aid of upcoming new fast and low power CMOS technologies, we consider designing a system for 40Gbps transmission over 50m CAT-7A cable.

From a signal processing standpoint, 40Gbps applications on copper have two main disadvantages:

- Computational complexities are substantially increased.
- Circuits require significant innovations in analog mixed-signal design, and DSP design.

In this dissertation, we discuss methods and algorithms that can contribute to alleviate these problems. The main contribution of this dissertation is in the technical feasibility assessment and system design for a data rate of 40Gbps over copper wire. We start by multi-input multi-output (MIMO) modeling and present formulas for capacity bounds. These bounds are good performance measures for a channel impaired by background noise and crosstalk signals. We also prove, by means of multiple access channel theorem, that single-input single-output
(SISO) implementation will perform as well as MIMO implementation and this is due to low far-end crosstalk (FEXT) level in CAT-7 cables. We used the specification of CAT-7A cable, an enhanced version of CAT-7 with a better performance and engineering design, to analyze, model and finally optimize the parameters for the given system requirements. We also consider a more elaborate pulse shaping filter than regular zero-ISI raised-cosine filters. Different equalization and precoding methods are evaluated. Finally multi-dimensional trellis and Low-density parity-check codes are considered as potential candidates for channel coding. We designed a low-power, low-latency LDPC assisted coded modulation which can provide coding gains up to 6dB, sufficient to satisfy the required constraints in terms of system margin and error probability. With some optimization algorithms, we came up with an idea to substantially reduce the power through multichannel data transmission.
Mathematical Notations and Symbols

The mathematical methods used in each chapter may differ slightly. However, the following convention for the names of variables and filters is followed in the dissertation. Scalar variables are denoted by plain lower case letters (e.g. \( x \)), bold face letters (e.g. \( \mathbf{x}, \mathbf{y} \)) denote column-vectors (with elements \( x_i, y_i \)). Capital letters (A, B) denote \( q \times q \) matrices (with elements \( a_{ij}, b_{ij} \)) and \((H, G)\) denote matrix-valued functions. We shall use these notations both for time and frequency domain, and distinction will be clear from the context without any confusion. Polynomial matrices in this dissertation refer to polynomial matrices in \( z \), or \( D \). The \( k \)-th delay term matrix coefficients of a polynomial matrix \( H(D) \) will be denoted as \( H_k \). Some frequently used notations are as follows:

- \( \mathbf{I}_N \) is the \( N \times N \) identity matrix.
- \( \mathbb{E}\{x\} \) is the statistical expectation of random variable \( x \).
- \( \text{det}\{X\} \) is the determinant of a matrix \( X \).
- \( \text{tr}\{X\} \) is the trace of a matrix \( X \).
- \( \text{diag}\{x_1, \ldots, x_n\} \) is an \( n \times n \) matrix with diagonal elements \( x_1, \ldots, x_n \) and zero elsewhere.
- \( \mathbf{x}_{n_1:n_2} \) denotes the column vector \( [x_{n_1}, x_{n_1+1}, \ldots, x_{n_2}]' \).
- \( \mathbf{e}_j \) is the unit vector having a 1 in position \( j \).
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................................................................. ix

LIST OF TABLES .................................................................................................................... xiii

ACKNOWLEDGEMENTS ....................................................................................................... xiv

Chapter 1 Introduction ......................................................................................................... 1

1.1 Motivations of This Study ........................................................................................... 1
1.2 Applications ............................................................................................................... 6
1.3 Outline of Dissertation ............................................................................................... 8

Chapter 2 Multi-Gigabit Ethernet ....................................................................................... 11

2.1 Introduction ............................................................................................................... 11
2.2 Important definitions ................................................................................................. 16

Chapter 3 Capacity Analysis ............................................................................................. 19

3.1 Introduction ............................................................................................................... 19
3.2 System Model ............................................................................................................. 20
3.3 Capacity Analysis ...................................................................................................... 22
3.3.1 SISO Single-Carrier and Water-filling Bounds .................................................. 23
3.4 MIMO Capacity, Coordinated Multi-Channel ........................................................... 26
3.5 The Loss Between SISO and MIMO Implementations ............................................. 28
3.6 Rate Optimization for Beyond-10G ........................................................................... 29
3.6.1 Minimizing the Probability of Error .................................................................. 29
3.11.2 Maximizing the System Margin ..................................................................... 31
3.12 Simulations and Results ........................................................................................... 31
3.8 40GBASE-T Full and Short Reach Modes ............................................................... 35

Chapter 4 Equalization and Precoding .............................................................................. 40

4.1 Introduction ............................................................................................................... 40
4.2 Pulse Shaping ............................................................................................................. 40
4.3 The ISI-channel Model .............................................................................................. 43
4.3.1 MMSE Linear Equalizer ................................................................................. 44
4.4 Decision-Feedback Equalizer ..................................................................................... 46
4.5 Tomlinson-Harashima Precoding .............................................................................. 49
4.18.1 MMSE-THP .................................................................................................. 51
4.6 Simulations and Results ............................................................................................. 52

Chapter 5 Interference Cancellation .................................................................................... 57

5.1 Introduction ............................................................................................................... 57
5.2 Echo Cancellation ...................................................................................................... 58
5.3 Channel Shortening Filter Design .............................................................................. 63
LIST OF FIGURES

Figure 1.1: 10GBSE-T system block diagram ................................................................. 3
Figure 1.2: Evolution of Ethernet by driving applications ............................................... 4
Figure 1.3: Cabling geography, left: commercial office, right: data center ..................... 5
Figure 1.4: Shielded twisted-pair structure used in CAT7 and CAT8 cables .................... 6
Figure 1.5: Global subscriber access traffic .................................................................... 7
Figure 2.1: Simplified model of a LAN served by Gigabit Ethernet switches .................. 13
Figure 2.2: Estimated power of 40G and 100G LAN Controllers .................................... 14
Figure 2.3: Expected die area of 40G and 100G LAN Controllers ................................... 14
Figure 3.1: Insertion Loss, Return Loss, NEXT and FEXT measurements for 50m CAT-7A ......................................................................................................................... 22
Figure 3.2: Block diagram of MIMO MMSE-DFE system .............................................. 27
Figure 3.3: Decomposition of H into diagonal and off-diagonal elements to represent it as a MIMO multiple-access channel ................................................................. 29
Figure 3.4: SISO Shannon, single-carrier, and water-filling capacity bounds .................... 32
Figure 3.5: MIMO capacity and R1 rate for 20 and 50 meters cables .............................. 33
Figure 3.6: Minimization of probability of error for two 40 and 100GBASE-T systems .... 34
Figure 3.7: Maximization of margin for two 40 and 100GBASE-T systems ..................... 35
Figure 3.8: System margin of CAT-7A cable (g_c=6dB). The full-reach mode is defined when the system margin is above the required 6dB margin ........................................ 36
Figure 3.9: 40GBASE-T short-reach mode block diagram .......................................... 37
Figure 3.10: System margin of CAT-7A cable for short-reach mode system (g_c=0dB) ..... 37
Figure 3.11: Bound on insertion loss to support 40Gbps. Different TX power levels give different bound. CAT7A 50m and CAT6A 50m, both have insertion loss above this bound for 10dBm ................................................................. 39
Figure 4.1: Equivalent discrete-time channel ................................................................. 43
Figure 4.2: Block diagram of unbiased MMSE linear equalizer ..................................... 44
Figure 4.3: Block diagram of decision feedback equalizer......................................................47
Figure 4.4: Block diagram of linear pre-equalization and Tomlinson-Harashima precoding..............................................................................................................................49
Figure 4.5: Block diagram of MMSE Tomlinson-Harashima precoding. ..............................52
Figure 4.6: Optimum asymmetric (left) and raised-cosine (right) FIR pulse shapes. 
\( N = 24, \beta = 0.1 \) ............................................................................................................52
Figure 4.7: Power spectrum estimate of random binary signal using raised-cosine & optimum asymmetric FIR pulse shapes. ....................................................................................53
Figure 4.8: Insertion loss measurements of 50m CAT-7A cable.............................................54
Figure 4.9: Impulse response of 50m CAT-7A cable, \( R_s = 3200 \text{MHz} \) ..................................54
Figure 4.10: Signal-to-noise ratio of unbiased MMSE decision-feedback equalizer versus forward taps.......................................................................................................................55
Figure 4.11: Initial convergence of adaptive decision-feedback equalizer..............................55
Figure 4.12: Performance of decision-feedback equalizer and Tomlinson-Harashima precoding.............................................................................................................................56
Figure 5.1: Interference sources in general multi-wire communication systems......................58
Figure 5.2: Near-end and far-end echo scenarios. ....................................................................59
Figure 5.3: Block diagram of full-duplex data link and echo cancellation............................60
Figure 5.4: The echo impulse response of 50m CAT-7A with 1550MHz bandwidth sampled at Nyquist rate..................................................................................................................61
Figure 5.5: ERLE versus number of least square estimated echo canceller. ..........................62
Figure 5.6: ERLE versus sparsity of corresponding FIR echo canceller. The canceller obtained by least-square estimation has a total of 1000 taps. .........................................................63
Figure 5.7: Shortening impulse response block diagram..........................................................64
Figure 5.8: Block diagram of MMSE joint channel and echo impulse response shortening...66
Figure 5.9: Block-convolution implementation of an FIR filter..............................................77
Figure 5.10: Complexity ratio between direct and block implementations of a 1024-tap FIR filter at different block lengths.........................................................................................78
Figure 5.11: Crosstalk from the local transmitter (near-end or NEXT).................................78
Figure 5.12: Conventional NEXT and FEXT cancellation ......................................................... 79

Figure 5.13: Crosstalk signal from far-end transmitter (FEXT). ................................................. 86

Figure 5.14: Type Caption Here ............................................................................................. 87

Figure 5.15: Power spectrum density of far-end crosstalk signal in a 50m CAT-7a cable. ......... 87

Figure 5.16: Original and shortened impulse response of a 50m CAT-7A cable. ................. 89

Figure 5.17: Return Loss ($S_{11}$) of 50m CAT-7A cable versus frequency. ......................... 90

Figure 5.18: Channel and echo impulse responses shortened by unit-norm joint shortening filter. .............................................................................................................. 91

Figure 5.19: Simulation results of block-convolution implementation of echo canceller. ....... 92

Figure 6.1: Capacity of multi-level PAM signaling .................................................................. 94

Figure 6.2: Set partitioning of a 16-QAM signal set into subsets with increasing minimum distance. The final partition level used by the encoder is the fourth level - that is, the subsets with two signal points each. .......................................................... 97

Figure 6.3: Block diagram of trellis encoder. ........................................................................... 98

Figure 6.4: Bipartite graph for a (3, 4) regular LDPC code ................................................... 100

Figure 6.5: Tree representation of bit nodes and check nodes connectivity for the first and second tiers ........................................................................................................... 104

Figure 6.6: Plot of $f(x)$ function .......................................................................................... 106

Figure 6.7: Construction of block coded modulation. ............................................................ 109

Figure 6.8: Data path between a variable node and one of its check nodes ......................... 112

Figure 6.9: Parity check architecture. ..................................................................................... 113

Figure 6.10: Variable node architecture .................................................................................. 114

Figure 6.11: System margin of 40GBASE-T system over 50m CAT-7A cable ...................... 116

Figure 6.12: Block diagram of multi-channel system .............................................................. 118

Figure 6.13: Simulation results of a 2-D 8-state trellis coded modulation. ......................... 119

Figure 6.14: Simulation results of a 4-D 8-state trellis coded modulation. ......................... 120

Figure 6.15: Histograms of the bit-to-check information for various decoder iterations. ....... 121
Figure 6.16: EXIT chart for iterative decoding of a regular (dλ =3, dρ = 6) LDPC code on a BEC with p =0.4 ................................................................................................................122

Figure 6.17: 6dB set partitioning on 64-QAM constellation..................................................122

Figure 6.18: Scaled LLR values of coded bits ........................................................................124

Figure 6.19: Bit error probability (BER) of coded and uncoded systems.................................124

Figure 7.1: Minimum SNR per symbol vs. bits per symbol for M-PAM transmission............130

Figure 7.2: Block diagram of a fix-point implementation of 16-PAM modulator/demodulator.................................................................................................................................131

Figure 7.3: Admissible ADC ENOB vs. signal bandwidth....................................................132

Figure 7.4: ADC ENOB vs. sampling frequency uncertainty..................................................133

Figure 7.5: Estimated partitioning of 40GBASE-T chip area....................................................138

Figure 7.6: Clock frequency of LDPC decoder versus block length......................................139

Figure 7.7: Latency of LDPC decoder versus block length for 40GBASE-T system. ............139
LIST OF TABLES

Table 3.1: Parameters for capacity analysis ................................................................. 32
Table 3.2: Capacity bounds for 50m cable and 3dBm Tx power ........................................ 33
Table 6.1: Number of levels versus bandwidth ............................................................... 123
Table 6.2: Parameters of 2-channel implementation ....................................................... 126
Table 7.1: 10GBASE-T FIR filters .................................................................................. 135
Table 7.2: 40GBASE-T FIR filters .................................................................................. 136
Table 7.3: LDPC codes’ parameters .............................................................................. 137
ACKNOWLEDGEMENTS

I would like to thank all people who helped and inspired me during my doctoral study. I am especially indebted to my advisor, Prof. Mohsen Kavehrad, for his support and guidance during my research and study at The Pennsylvania State University. His ceaseless energy and enthusiasm in research have always motivated me. In addition, he was always available and willing to help. This was truly rewarding for me. Professors J. Mitchell, A. Hurson and W. Waterhouse deserve special thanks as my thesis committee members and advisors for their time, encouragement and help. I was delighted to interact with Prof. Nirmal Bose by attending his classes and having many fruitful conversations with him. His insights into signal processing are second to none.

All my friends and colleagues at Sharif University of Technology and the Center for Information and Communications Technology Research at Penn State University have helped me a lot by providing an inspiring and enjoyable atmosphere. Special thanks go to my friend Jarir M. Fadlullah; apart from our many discussions, he has helped me a lot by reading my manuscripts and providing constructive criticism on my papers and technical reports. I would also like to thank my dear friend Shahram Takyar for being a caring friend and for providing a different perspective on life through the past 14 years.

Needless to say, my parents and my sisters (and my little niece) have been very important for me. Here, I would like to thank my mother, who has devoted her life to the family and supported her children by every possible way that she could. My deepest gratitude goes to my sisters for their everlasting love for me.

Lastly, I am forever indebted to my beloved wife Maryam Kherad Pezhouh. I dedicate the thesis to Maryam for her continuing support, encouragement, faith and love. Thanks for your endless patience when it was most required.
To my wife,

my parents,

and to the memory of my grandmother.
Chapter 1

Introduction

1.1 Motivations of This Study

This Ph.D. dissertation focuses on the design, development and demonstration of 40 and 100 Gigabits per second (Gbps) digital signal transmission systems over Ethernet copper cables serving ultra high-speed interconnection in data centers, local area networks (LAN), and many others.

In the commercial market, a new challenge is the extension of fiber into the access network in small business and dense metropolitan areas. It has long been known that a major bottleneck in delivering multimedia services to the computer users is the low-capacity of LANs. With ever-increasing demand for higher data rates, the need for broadband access is transformed from a convenience to a necessity. So far, data communication has been the main driving force behind increased traffic on the communication networks. Applications stemming from a wide range of disciplines such as high-performance computing, consumer multimedia, teleconferencing and telemedicine are just few examples that require data rates in the gigabits per second range. To keep up with this explosive growth, ultra high-capacity networks were required, and thus optical networks with terabit capacities dominated the network core. To enable the end user to take full advantage of this core, reliable high-speed LAN access is required. Providing service in a broadband access LAN, using a copper cable approach, has the advantages of the network being highly-dependable and cost-effective. This will benefit the service providers over campus settings like hospitals, industry compounds or universities with facilities spread over
several buildings in that a quick service upgrade could extend new service offerings. Also, within server farms and data centers, short copper connectors are preferable.

Copper LAN is intended to be a cost-effective competitor to the traditional LANs based on wireless, coaxial cable, and optical fiber. In such a system, the copper cable link would serve as an extending arm to fiber reach. This would be accomplished by receiving signal packets coming from the optical backbone edge to a local feeder and delivering these signals to remote access points through copper cable. A copper LAN serves the same purpose as that of other LANs: to convey information among the devices attached to the LAN. With the increasing popularity of multimedia services supplied over a fixed network, services such as web browsing, video conferencing and video on demand, it is for sure only a matter of time before users will demand higher bandwidth LAN access. Advances in signal processing and fast CMOS processing power have also made it possible for users to afford high-resolution visual services.

The 10GBASE-T, or IEEE 802.3an-2006, is a standard released in 2006 to provide 10Gbps connections over conventional unshielded or shielded twisted pair cables (UTP and STP, respectively), over distances up to 100m, i.e., connecting work areas to a telecommunications room. With existing Category 6 cables (CAT6), 10GBASE-T will work up to 55m (180 ft). After introducing a new partitioned augmented Category 6 or "6A" cable specification, specifically designed to reduce crosstalk between cables (known as "alien crosstalk"), this could be extended to the usual 100m (328 ft). Under this condition, it is extremely hard to design a system operating at this speed and meet the required quality of service (probability of error better than $10^{-12}$ according to IEEE 802.3). Sophisticated protocols, advanced forward error correction coding, multidimensional signal space and robust equalization are involved in this system. The actual cost of this system comes from a complex mixed-signal VLSI implementation and not its engineering work. The block diagram of this system is shown in Fig. 1.1.
Demand for high-density switches and servers with multi-gigabit cable connectivity continues. This will extend the 10GBASE-T systems beyond data centers to small local area networks and even home networking. Many IEEE members recognized this potential and are thinking of ways to deliver tens of Giga bits per seconds over balanced cabling. After successful implementation of IEEE 10GBASE-T system in 2006, people started to study technical feasibility, broad market potential, and economic feasibility of data transmission over copper beyond 10Gbps. Ethernet applications have evolved from 10BASE-T, 100BASE-T, 1000BASE-T and recently 10GBASE-T. It is over a decade that the so called Fast Ethernet 100BASE-TX networking physical connectivity standard was established. In 2007, IEEE higher speed study group announced their persuasion of 40 and 100GBASE-T applications [53], a thousand times faster than Fast Ethernet (surprisingly this falls into the Moors law very well, Fig. 1.2 [36])! This is in fact in correspondence with explosive growth of demand for higher speed in high-performance computing, consumer multimedia, teleconferencing and telemedicine, etc. 100GBASE-T, fact or fiction, it is only a matter of time before the bandwidth of a 100G cabling system will be needed. The 10GBASE-T standard is complete now, but it does not mean that the Ethernet Alliance and the IEEE can rest -- 100GBASE-T is just over the horizon.

Figure 1.1: 10GBSE-T system block diagram.
IEEE is interested in the next generation networking physical connectivity standard over twisted pair copper cables (40 and 100GBASE-T). This coincides with Category-7 cable (CAT7) coming to market that offers a much better performance at almost the same price as CAT6 cables. We shall investigate and evaluate the feasibility of standardizing 40 and 100Gbps data rates over horizontal balanced cabling. The objective of IEEE 40 and 100GBASE-T would be to create an application that is capable of transporting data at rates 40 and 100Gbps, respectively, over 10 meters of balanced twisted-pair cable. The cabling industry, is promoting the use of Category 7A cable (or more convenient CAT8) to support these demanding applications. The capacity analysis of this cable shows that 40Gbps transmission is practical up to 50m over this cable, thanks to its excellent shielding and engineering design, at the cost of elaborate digital signal processing and complex coding schemes. Besides, the data center geography statistics show that 80–90% of links are shorter than 45m (Fig. 1.3) [36]. Accordingly, by the aid of upcoming new fast and low power CMOS technologies, we consider designing a system for 40Gbps transmission over 50m CAT-7A cable.

Figure 1.2: Evolution of Ethernet by driving applications.
We shall consider the technical feasibility and system design parameters for a data rate of 40/100Gbps over copper. We use the specification of an enhanced version of Category 7 (CAT-7A) cable to analyze, model and finally design a conceptual system that can address the implementation of this system with future advances in technology. Fig. 1.4 illustrates the structure of CAT7 and CAT8 shielded twisted-pairs. The characteristics of copper cable and the regulatory constraints create significant challenges in defining the physical layer. Fortunately, digital communications techniques can be used to design robust transceivers that are capable of achieving reliable operations. Other Ethernet connectivity standards like 1000BASE-T and 10GBASE-T take advantage of several of these techniques to transform the desired bit rate into an acceptable baud rate for operation over 4-pair copper cable. Signal equalization is used to compensate for signal distortion introduced by communications channel. Forward error correction (FEC) provides a second level of protection that helps to recover the transmitted symbols in the presence of high noise and crosstalk.

From a signal processing point of view, 40 and 100Gbps applications on copper have two main disadvantages:

- Computational complexities are substantially increased.
- They require significant innovations in analog mixed-signal design, and DSP design.
In this dissertation, we shall discuss methods and algorithms that can contribute to alleviate these problems. We address the analysis, modeling and design of sub-systems in a framework leading to a practical implementation of beyond-10Gbps within a next few years with the ever increasing technology advances. The approach has applications in local area networking as well as for intra-chip networking and communications.

1.2 Applications

Over the past 10 years, the demand for higher bandwidths has increased exponentially. The growth of number of IP-based applications definitely continues. Video and voice transmission are the main factors of growing network traffic. It is estimated that by year 2010, over 250 terabytes (TB) of information are carried over the global network among which video streaming and voice over IP (VoIP) are the dominant traffics [36] (Fig. 1.5).

One of the most implications of high-speed copper cabling is in Data Centers and High Performance Computing (HPC) centers. A data center – according to WIKIPEDIA – is a facility

Figure 1.4: Shielded twisted-pair structure used in CAT7 and CAT8 cables.
used to house mission-critical computer systems and associated components, usually owned and maintained by an organization for the purpose of handling the data necessary for its operations.

The Telecommunications Industry Association (TIA/EIA), in charge of the standardization of telecommunication systems, in 2005 adopted the TIA/EIA-942 final document defining the telecommunications infrastructure for data centers. The standard specifies the types and lengths of structured cabling that can be used to wire a data center as well as the different areas regrouping the equipment interconnected by this cabling.

There are many services that a data center can provide. Storage can fulfill multiple roles and purposes within the corporate environment. The software applications have different requirements in terms of how fast, how often, how many replicas, how much data needs to be stored; how fast, how often and how many of these replicas need to be accessed for retrieval of the stored data. The most basic application is storage device pooling, or aggregation. It consists of managing several independent storage devices together, as a pool of storage, so that adjustments can be made dynamically on the storage capacity assigned to each server with access to this pool. Usually the group of servers is dedicated to one and the same software application, e.g. providing

Figure 1.5: Global subscriber access traffic.
web files, or archiving data. Pooling is a particular aspect of storage virtualization. Data sharing is the use of stored data in common by several software applications – possibly running on the same server. Data sharing is commonly divided between two types of usage: data copying and real time data sharing.

1.3 Outline of Dissertation

The dissertation is organized as follows. In Chapter 2, a brief history of gigabit Ethernet evolution, 1000BASE-T and later on multi gigabit Ethernet 10GBASE-T, is presented. Gigabit Ethernet has become the unifying technology enabling communications via the Internet and other networks using Internet Protocol (IP). Due to its proven low cost, known reliability, and simplicity, the majority of today’s internet traffic starts or ends on a gigabit Ethernet connection. Two major problems faced during the development of 1000BASE-T and 10GBASE-T standards are: power consumption and latency. In chapter 2, we also explain how these challenges motivated us during the course of this research. Some important definitions concerned in the high-speed physical connectivity are also presented in this chapter.

In Chapter 3, we investigate the theoretical limitations of wireline communication systems. We present both single-input/single-output (SISO) and multi-input/multi-output (MIMO) system modeling. Shannon capacity analysis, as an important benchmark, is used to confirm the technical feasibility of emerging 40 and 100Gbps cabling systems with future IEEE 802.3 compliant networking equipment. Two important practical capacity bounds, single-carrier and water-filling, are extensively discussed and formulated. These bounds determine the maximum rate of data transmission over a dispersive channel under practical limitation and constraints. Also, we develop a set of optimization techniques and tools, essentially to understand the trade-offs between design parameters. We pursue two major optimization procedures to
minimize probability of error and maximize system margin. The *how-far how-fast* model for data transmission over CAT-7A cable is developed in this chapter. We also discuss a hypothetical cable model appropriate for emerging 40 and 100GBASE-T applications. The cable manufacturers can use this model to design a better cable in order to achieve the convenient 100m reach mode.

In Chapter 4 we start by reviewing Nyquist criteria for data transmission over a bandwidth-limited channel. We also consider a more elaborate pulse shaping filter than regular zero-ISI raised-cosine filters. Different equalization and precoding methods are considered and evaluated for over desired wireline channel, CAT-7A cable. The performance of a finite-length equalizer is compared to the baseline, ideal, infinite-length equalizer. The precoding techniques that alleviate the problem of error propagation associated to the conventional data-aided equalizers are also considered in this chapter.

In Chapter 5, the degradations caused by different types of interference signal presented in balanced twisted-pair cables are examined. The cancellation levels for these crosstalks to minimize system debasement are determined. Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) models for echo signal are developed and corresponding echo cancellations are designed. We present a low-complex method for joint echo cancellation and channel equalization. We extensively analyze the echo cancellation techniques in terms of complexity, ease of implementation, speed of convergence, etc. We specially develop a MIMO cancellation scheme for cables with poor connectivity at near-end (high near-end crosstalk).

In Chapter 6, the forward error correcting scheme is pursued to fulfill the required quality service offered by 40 and 100GBASE-T systems, i.e. the probability of error better than $10^{-12}$. Also, in this chapter, the perspectives of combined coding and precoding are our concern to resolve some of the system design issues. Trellis coded modulation is briefly reviewed as it shares the fundamental principles with our proposed block coded modulation. An algorithm is developed
to construct a high-rate, high-gain code from a low-rate, high-gain LDPC code. Also, a multi-channel system is designed and developed to reduce the complexity of the system in terms of chip area and power consumption.

In Chapter 7, we discuss the main causes of the gap between ideal and practical systems. The quantization noise of mixed-signal circuits, apparently the dominant noise in wireline Ethernet communications, is greatly investigated. The dynamic and static performance of analog-to-digital (ADC) and digital-to-analog (DAC) converters are investigated to obtain a lower bound for the number of quantization levels. The impacts of echo cancellation are also considered in this study.

Chapter 8 concludes the dissertation and describes the future tasks needs to be considered for further developments in this research.
Chapter 2
Multi-Gigabit Ethernet

2.1 Introduction

Gigabit Ethernet over unshielded twisted pair cable is a term that refers to 1000BASE-T physical layer standard. It serves as backbone links on high-capacity networks to support thousands of Ethernet/fast Ethernet users (these terms refer to 10BASE-T and 100BASE-T Ethernet). 1000BASE-T was originally implemented for server-type computer, but it quickly became a built-in feature in many computers. Fig. 2.1 presents a simple block diagram of a LAN served by 1000BASE-T connectivity.

After the 1000BASE-T successful story, to meet the growing demands of packet-switched networks, 10GBASE-T standard was pursued and finally released in 2006. But soon, in 2006, the IEEE 802.3 working group formed the Higher Speed Study Group (HSSG) and found that the Ethernet ecosystem needed a communication medium faster than 10 Gigabit Ethernet. The growth in bandwidth for network aggregation applications was found to be outpacing the capabilities of networks employing link aggregation with 10 Gigabit Ethernet. As the HSSG studied the issue, it was determined that computing and network aggregation applications were growing at different rates. For the first time in the history of Ethernet, a Higher Speed Study Group determined that two new rates were needed: 40 gigabit per second for server and computing applications and 100 gigabit per second for network aggregation applications. The objectives of these applications are listed below:
Support full-duplex operation only

Preserve the 802.3 / Ethernet frame format utilizing the 802.3 media access controller (MAC)

Preserve minimum and maximum frame size of current 802.3 standard

Support a bit error rate (BER) better than or equal to $10^{-12}$ at the MAC/physical layer service interface

Provide appropriate support for optical transport network (OTN)

Support a MAC data rate of 40 gigabit per second

Provide physical layer specifications which support 40 gigabit per second operation over:

- at least 10km on single mode fiber (SMF)
- at least 100m on OM3 multi-mode fiber (MMF)
- at least 10m over a copper cable assembly (40GBASE-T)
- at least 1m over a backplane

Support a MAC data rate of 100 gigabit per second

Provide physical layer specifications which support 100 gigabit per second operation over:

- at least 40km on SMF
- at least 10km on SMF
- at least 100m on OM3 MMF
- at least 10m over a copper cable assembly (100GBASE-T)

We are interested in 40 and 100GBASE-T physical connectivity standards. Unfortunately, due to some practical issues encountered in the course of 10GBASE-T implementation, there has not been much research around these applications. The research in [97] is the only serious work about the technical feasibility of 40 and 100GBASE-T systems. The
implementation costs and power consumption of 40/100GBASE-T seem to be the main reasons that researchers have not followed this research yet, and all the studies regarding the 40/100Gbps transmission involve fiber optic communication. But our works on technical feasibility of 40/100Gbps transmission on Category 7 copper cables has revealed that, theoretically, 40Gbps is quite feasible for up to 50 meters.

The work reported in [97] considers the complexity and implementation of these systems; the estimated power consumption and die area of the VLSI chip for 40 and 100GBASE-T are evaluated using the 10GBASE-T as the baseline. The results of this report are shown in figures 2.2 and 2.3. These figures illustrate how the power consumption and die area (indirectly

Figure 2.1: Simplified model of a LAN served by Gigabit Ethernet switches.
refers to cost) are anticipated to be downsized through advances in VLSI technology (and possibly some innovative ideas) during a 10-year time span.

Figure 2.2: Estimated power of 40G and 100G LAN Controllers.

Figure 2.3: Expected die area of 40G and 100G LAN Controllers.
There are two major concerns regarding these new ultra fast applications as future IEEE high-speed physical connectivity. We explain them in the context of 10GBASE-T system as follows.

1. The 10GBASE-T modem is readily in the market, but this system is power hungry. This can cause a serious problem in data centers and high performance computing centers where hundreds or thousands of these modems are to operate. As energy efficiency continues to be of great concern, standards initiatives, such as the IEEE P802.3az Energy Efficient Ethernet, will introduce ways to map power consumption to usage and network loads. Additional power savings modes will be incorporated at both the component level and in end-user products. The IEEE P802.3az task force is on target to ratify requirements in the second half of 2010, shaping the way components operate and are designed.

2. In real-time data sharing, applications work on the same data set. This saves storage space and removes the necessity of managing several versions of the same data. In compensation, real time data sharing requires synchronization between the different applications that access the common data. If the applications run on the same computer, the issue is already solved, since all applications have a common clock, given by the CPU. However in most cases, the applications run on different computers, each of them availing itself of the full processing power of a CPU. In this case, the computers work together as a cluster, exchanging messages in order to synchronize with each other. Hence, for real-time data sharing applications, transmission latency – the delay sustained by the data between two communicating CPUs – is a critical parameter.

In the next chapter, we will design the key blocks of these systems and try to identify the bottle necks and present some design ideas that can alleviate the complexity of these systems.
2.2 Important definitions

It is worthwhile here to review some definitions and parameters of Category cables. This overview sheds some light on what the research entails.

**Attenuation to Crosstalk Ratio, Far-End (ACRF):** Pair-to-pair far-end crosstalk (FEXT) loss quantifies undesired signal coupling between adjacent pairs at the far-end (the opposite end from the transmit-end) of cabling or a component. ACRF is calculated by subtracting the measured insertion loss from the measured far-end crosstalk loss and yields a normalized value that can be used to compare cable and cabling performance independent of length. Poor ACRF levels can result in increased bit error rates and/or undeliverable signal packets. Note that, near-end crosstalk (NEXT) loss margin alone is not sufficient to ensure proper ACRF performance.

**Attenuation to Crosstalk Ratio (ACR):** A critical consideration in determining the capability of a cabling system is the difference between insertion loss and near-end crosstalk (NEXT) loss. This difference is known as the attenuation to crosstalk ratio (ACR). Positive ACR calculations mean that transmitted signal strength is stronger than that of near-end crosstalk. ACR can be used to define a signal bandwidth (i.e. 200MHz for Category 6) where signal-to-noise ratio values are sufficient to support certain applications. It is interesting to note that digital signal processing techniques can perform crosstalk cancellation allowing some applications to expand usable bandwidth up to and beyond the point at which the calculated ACR equals zero. Even so, the maximum frequency for which positive ACR is assured provides a benchmark to assess the usable bandwidth of twisted-pair cabling systems.

**Balance:** Twisted-pair transmission relies on signal symmetry or "balance" between the two conductors in a pair. Maintaining proper balance ensures that cabling systems and components do not emit unwanted electromagnetic radiation and are not susceptible to electrical
noise. Component balance requirements are specified for Cat 6/Class E cabling. Component and cabling balance requirements are specified for category 6A/class EA and higher grades of cabling. Balance may be characterized by longitudinal conversion loss (LCL), longitudinal conversion transfer loss (LCTL), transverse conversion loss (TCL), or equal level transverse converse transfer loss (ELTCTL).

**Insertion Loss (Attenuation):** Insertion loss is a measure of decrease in signal strength along the length of a transmission line. Ensuring minimal signal attenuation is critical because digital signal processing techniques cannot compensate for excessive signal loss.

**Near-End Crosstalk Loss (NEXT):** Pair-to-pair near-end crosstalk loss quantifies undesired signal coupling between adjacent pairs at the near-end (the same end as the transmit-end) of cabling or a component. Excessive NEXT loss can be detrimental to applications that do not employ crosstalk cancellation techniques.

**Power Sum:** All pair-to-pair crosstalk parameters can be expressed as a power summation, which approximates the level of undesired internal signal coupling present when all pairs are energized. Power sum NEXT loss, ACRF, ANEXT loss, and AACRF characterization confirm that the cabling is significantly robust to minimizing crosstalk from multiple origins. This type of characterization is necessary to ensure cabling compatibility with applications that use all four pairs for transmitting and receiving signals, simultaneously.

**Propagation Delay and Delay Skew:** Propagation delay is the amount of time between transmission and reception of a signal traversing a cabling channel. The effect is akin to the time delay between lightning and thunder. Delay skew is the arrival time difference between the pair with the least delay and the pair with the most delay. Transmission errors that are associated with excessive delay and delay skew include increased jitter and bit error rates.

**Return Loss:** Return loss is a measure of the signal reflections occurring along a transmission line and relate to impedance mismatch in a cabling channel. Because emerging
applications such as 1000Base-T and 10GBase-T rely on full duplex transmission encoding schemes (transmit and receive signals are superimposed over the same conductor pair), they are sensitive to errors that may result from marginal return loss performance.

**Alien Crosstalk** Unwanted signal coupling from one component, channel, or permanent link to another is defined as alien crosstalk. Since alien crosstalk is an indicator of differential (or balanced) signal coupling, alien crosstalk cannot be adversely affected by common mode noise such as that from motors or fluorescent lights present in the environment. Alien crosstalk is only specified by the standards as a power sum parameter for components and cabling; it approximates the energy present when all pairs are energized. Power sum alien crosstalk measured at the near-end is called power sum alien near-end crosstalk loss (PSANEXT loss) while power sum alien crosstalk measured at the far-end is called power sum alien attenuation to crosstalk ratio, far-end (PSAACRF). High-power sum alien crosstalk levels can compromise 10GBase-T operation.

**Transfer Impedance** Shield effectiveness characterizes the ability of screened (F/UTP) and fully shielded (S/FTP) cables and connecting hardware to maximize immunity from outside noise sources and minimize radiated emissions. Transfer impedance is a measure of shield effectiveness; lower transfer impedance values correlate to better shield effectiveness.
Chapter 3
Capacity Analysis

3.1 Introduction

The capacity of a communication media is an important quantity that relates the physical properties of the channel to the quality of the service offered over the channel. There are a few capacity bounds defined under different conditions and processing power necessary to detect the signal available at the receiver. The Shannon bound of a channel determines the fastest possible data rate when infinite memory and processing are assumed for the receiver. Single-carrier bound reflects the fact that all the resources, e.g., memory, transmit power, signal processing, etc., are limited. Another important capacity bound is water-filling which includes the assumption of the availability of channel information at the transmitter side as well. In this chapter, we define these capacity bounds for the twisted-pair channel and present the formulations both for single-input/single-output channel (SISO) and multi-input/multi-output channel (MIMO). We will use these bounds to obtain a set of parameters for practical implementation of 40GBASE-T system. Furthermore, we develop a set of optimization techniques and tools, basically to understand the trade-offs between design parameters. We pursue two major optimization procedures to minimize probability of error and maximize system margin. To extend the reach mode, in 40GBASE-T application, a hypothetical cable model is derived which could be used as the benchmark by cable manufacturers. Simulations are performed to assess the feasibility and practicality of 40 and 100GBASE-T systems over 50 and 20m CAT-7A cables, respectively.
3.2 System Model

The CAT-7A cable is made of four doubly shielded horizontal balanced twisted-pairs, labeled by Blue, Brown, Orange and Green colors, and is shown in Fig. 1.3. One can set up a communication system consisting of four transceivers at each end. A data stream is subsequently shaped by a pulse filter \( p(t) \) and mapped onto a vector waveform \( \mathbf{x}(t) = [x^1(t), \cdots, x^4(t)]^T \), where \( x^n(t) = \sum_m x^n_k p(t - mT), \ n = 1, \cdots, 4 \) is the input signal to the \( n \)-th channel. Typically, signal constellations \( x_k = [x_k^1, \cdots, x_k^4]^T \) are taken from the scaled lattice \( \mathbb{Z}^4 \) (or \( \mathbb{D}^4 \)) [25] to maximize the minimum Euclidean distance. Category cables are designed in a way that all twisted pairs behave more or less similarly, so the total rate is divided among them equally. Therefore, \( x^n_k \) is drawn from a finite alphabet set (e.g. M-PAM signaling) that contains equidistant real symbols centered on the origin, for example

\[
\mathcal{A} = d\{\pm 1, \pm 3, \cdots, \pm(M - 1)\}
\]  

(3.1)

where \( 2d \) is the minimum distance between symbols.

Assuming that the receiver synchronously samples the output of the vector channel, at time \( kT \), the tap-weight matrices and tap-voltage vectors are defined as

\[
\mathbf{r}_k = \sum_{m=0}^{n_H} \mathbf{H}_m \mathbf{x}_{k-m} + \sum_{m=0}^{n_L} \mathbf{G}_m \mathbf{z}_{k-m} + \mathbf{v}_k
\]  

(3.2)

where

- \( \mathbf{x}_k = [x_k^1, \cdots, x_k^4] \) is the \( k \)-th sample of vector signal \( \mathbf{x}(t) = [x^1(t), \cdots, x^4(t)]^T \), the input signal to the channel \( \mathbf{H}(t) \).
- \( \mathbf{z}_k = [z_k^1, \cdots, z_k^4] \) is the \( k \)-th sample of vector signal \( \mathbf{z}(t) = [z^1(t), \cdots, z^4(t)]^T \), the interfering signal from the near-end transmitters.
• $H_k = H(kT)$ is the discrete matrix impulse response of the vector channel that describes the signal attenuation (insertion loss) and electromagnetic coupling between the twisted-pairs (FEXT).

• $G_k = G(kT)$ is the discrete matrix impulse response of interfering channels representing return loss (RL) and near-end crosstalk (NEXT).

• It is assumed that, without loss of generality, all the elements of $H$ (G) have the same channel order $n_H$ ($n_G$), and

$$H_k = \begin{bmatrix} h_{1,1}[k] & \ldots & h_{1,4}(k) \\ \vdots & \ddots & \vdots \\ h_{4,1}[k] & \ldots & h_{4,4}[k] \end{bmatrix}, G_k = \begin{bmatrix} g_{1,1}[k] & \ldots & g_{1,4}(k) \\ \vdots & \ddots & \vdots \\ g_{4,1}[k] & \ldots & g_{4,4}[k] \end{bmatrix}$$ (3.3)

• Generally, $x_k$ and $z_k$ share the same statistical properties; they are both zero-mean wide-sense stationary processes, and the $4 \times 4$ autocorrelation matrices of $x_k^n$ and $z_k^n$ are given by $R_{xx} = R_{zz} = \sigma^2_s I_4$, which means $x_k^n$ (and $z_k^n$) are uncorrelated, both spatially and temporally with variance $\sigma^2_s$. We also assume that the noise samples are uncorrelated, both temporally and spatially (i.e. $R_v = \sigma^2_v I_4$).

• Due to proper shielding of CAT-7A cable, the alien near-end crosstalk is negligible (although, if its power is significant, it can be simply included into our model).

Fig. 3.1 depicts the insertion loss and return loss measurements of the Blue pair ($h_{1,1}$ and $g_{1,1}$) along with the NEXT and FEXT interference signals from the Brown pair ($h_{1,2}$ and $g_{1,2}$) as functions of the frequency.

Although our results are based on actual measurements and characterization made by Nexans, for interested readers who want to evaluate/duplicate some of the results, we adopted the
following equation which can model the insertion loss of 50m CAT-7 cable very well up to 3GHz ($f$ is in MHz in this equation).

$$\text{IL}_{\text{dB}} = \begin{cases} 975 \sqrt{f} + 0.0025 f + 0.125 \sqrt{f} - 0.08 \sqrt{f} & f < 2000 \\ -77 + \sqrt{3000 - f} & f > 2000 \end{cases}$$  

(3.4)

![Figure 3.1: Insertion Loss, Return Loss, NEXT and FEXT measurements for 50m CAT-7A.](image)

3.3 Capacity Analysis

In this section, we first review the capacity of the general Additive White Gaussian Noise (AWGN) channel, and introduce the Single-Carrier, Water-filling, SISO and MIMO capacity notations for Category cables.
3.3.1 SISO Single-Carrier and Water-filling Bounds

As a first step toward designing any communication system, the capacity limitation of the communication medium (twisted-pair cable) has to be determined. Shannon’s theorem shows how to compute a channel capacity from a statistical description of a channel, and establishes that given a noisy channel with capacity $C$ and information transmitted at a rate $R$, if $R < C$ there exists a coding technique which allows the probability of error at the receiver to be made arbitrarily small. This means that it is theoretically possible to transmit information almost without error up to nearly a limit of $C$ bits per second. The channel capacity $C$ can be calculated from the physical properties of a channel. The Shannon-Hartley theorem [9,10,11,57] states the theoretical maximum rate of error-free data that can be sent with a given signal power spectrum $S(f)$ through an analog communications channel subject to additive (possibly colored) Gaussian noise with power spectrum $V(f)$. According to this theorem, the capacity of a band-limited channel with Gaussian noise is

$$C_{\text{SH}} = \int_0^W \log_2 \left( 1 + \frac{S(f)}{V(f)} \right) df \approx \sum_{n=1}^N \frac{W}{N} \log_2 \left( 1 + \frac{S(f_n)}{V(f_n)} \right)$$

where $C_{\text{SH}}$ is the channel capacity in bits per second, $W$ is the bandwidth of the channel in Hz, and finally $f$ is the frequency in Hz.

Over a highly dispersive channel, an efficient way to utilize the allocated bandwidth is to treat the channel through $N$ independent sub-channels which have a nearly flat frequency response by making $N$ large enough. It is desirable to have all sub-channels with the same probability of error $p_e$. Constant $p_e$ can take place when all sub-channels use the same class of codes with a constant SNR gap $\Gamma$ [56]. In this case, a single performance measure characterizes a multi-channel transmission system. This measure is a geometric SNR that can be compared to the
detection SNR of equalized transmission systems [56]. The asymptotic capacity of this multi-channel system is considered as *single-carrier* bound and is obtained by:

$$C_{SC} = \lim_{N \to \infty} W \sum_{n=1}^{N} \frac{1}{N} \log_2 \left( 1 + \frac{\text{SNR}_n}{\Gamma} \right) = W \log_2 \lim_{N \to \infty} \prod_{n=1}^{N} \left( 1 + \frac{\text{SNR}_n}{\Gamma} \right)^{1/N}$$  \hspace{1cm} (3.6)

The limit can be calculated as:

$$\lim_{N \to \infty} \prod_{n=1}^{N} \left( 1 + \frac{\text{SNR}_n}{\Gamma} \right)^{1/N} = \exp \left( \lim_{N \to \infty} \sum_{n=1}^{N} \ln \left( 1 + \frac{\text{SNR}_n}{\Gamma} \right)^{1/N} \right)$$ \hspace{1cm} (3.7)

$$= \exp \left( \frac{1}{W} \int_{0}^{W} \ln \left( 1 + \frac{\text{SNR}(f)}{\Gamma} \right) df \right)$$

Literally, this limit is related to the so-called *Salz* SNR, which is often used in practical system implementations to estimate the system noise margin (required SNR subtracted from achievable SNR). The Salz SNR is defined as

$$\gamma_{\infty}^W \{U(f)\} = \exp \left\{ \frac{1}{W} \int_{0}^{W} \ln \left( U(f) \right) df \right\}$$ \hspace{1cm} (3.8)

where $U(f) = 1 + \Gamma^{-1} \text{SNR}(f)$. Therefore,

$$C_{SC} = W \log_2 \gamma_{\infty}^W \left[ 1 + \frac{\text{SNR}(f)}{\Gamma} \right]$$ \hspace{1cm} (3.9)

In fact, this bound indicates the ultimate throughput of a real implementation of a system with finite coding gain and signal processing for any communication medium. Two such implementations are the minimum mean-squared error decision feedback equalizer (MMSE-DFE) [22,23,89], and Tomlinson-Harashima Precoding (THP) [93][49].

Maximizing the data rate, for a set of parallel channels when the symbol rate is fixed, requires maximization of the achievable $C = \sum_n c_n$ over $E_n$, the average power of each sub-
channel. This is summarized as the following maximization problem, where \( H_{kk}(f_n) \) represents the \( n \)th sub-channel transfer function of \( k \)-th channel

\[
\lim_{N \to \infty} \left\{ \maximize_{E_n} \sum_{n=1}^{N} \log_2 \left( 1 + \frac{E_n \left| H_{kk}(f_n) \right|^2}{\Gamma N_n} \right) \right. \\
\left. \text{subject to } \sum_{n=1}^{N} E_n = \sigma_s^2 \right. \\
\] (3.10)

We should mention here that the optimization is performed separately for each twisted pair. This corresponds to our previous assumption that twisted pairs are more or less similar. One can perform a joint optimization in the case where the characteristics of twisted pairs differ significantly. A natural solution of this optimization problem is to use Lagrange multipliers [86,26,45]. In this paper, we refer to the maximum value of this function, denoted by \( C_{WF} \), as the water-filling bound [26].

For parallel twisted-pair channels, if the individual channels are treated and equalized separately, and the interference signals from other channels are considered as noise (although the power of these interfering signals are attenuated by proper crosstalk cancellers), then the total SISO single-carrier capacity reads as

\[
C_{SISO-SC} = \sum_{k=1}^{4} W \log_2 \gamma_\infty \left( 1 + \frac{\sigma_a^2 \left| D_{k,k}(f) \right|^2}{\Gamma N_k(f)} \right) \\
\] (3.11)

where \( N_k(f) = \sigma_c^2 + \sum_{l=1, l \neq k}^{4} \Gamma_{kl}^F \left| H_{k,l}(f) \right|^2 + \sum_{l=1}^{4} \Gamma_{kl}^N \left| G_{k,l}(f) \right|^2 \) and \( \Gamma_{kk}^N, \Gamma_{kl}^N, \text{ and } \Gamma_{kl}^F \) are attenuation factors of the corresponding RL, NEXT, and FEXT crosstalk cancellers, respectively. Similar water-filling definition and formulation can be presented for parallel twisted-pair channels.
3.4 MIMO Capacity, Coordinated Multi-Channel

In this section, we assume that the effect of $G$ in model Eq. (3.2) is well reduced by proper NEXT and echo cancellers. In [85], it is shown that in the case of strictly monotonous decreasing channel attenuation, a constant power density in the first Nyquist set of frequencies $f \in [-1/2T, 1/2T]$ ($T$ symbol rate) is optimum. Therefore, as a generalization of the Shannon-Hartley theorem, the capacity of the MIMO system can be evaluated as

$$C_{\text{MIMO}} = \int_0^W \log_2 \det \left( I_N + \frac{\sigma^2_a}{\sigma^2_v} H(f)H^\dagger(f) \right) df$$

where $W$ is the available bandwidth.

The MIMO single-carrier (MIMO-SC) capacity can be bounded by

$$C_{\text{MIMO-SC}} = W \log_2 \left( \frac{1}{1 - \frac{1}{\Gamma}} \int_0^W \det \left( I_N + \frac{\sigma^2_a}{\sigma^2_v} H(f)H^\dagger(f) \right) df \right)$$

In sequel, we show that this bound corresponds to mean squared error (MSE) minimization of a MIMO-MMSE-DFE system, depicted in Fig. 3.2. In this diagram, $W(z)$ and $B(z)$ are feedforward and feedback filters, respectively. The estimation error, $e(z)$, is defined as the difference between the input and output variables of the decision maker device. For optimum filters, that minimize MSE, $e(z)$ is a white sequence [59,60,98], i.e. $R_{ee}(z) = \sigma^2_a \sigma^2_v \Lambda^{-1}$, where

$$S_e(z) \triangleq \sigma^2_a H^\dagger(1/z^*) H(z) + \sigma^2_v I = Q^\dagger(1/z^*) \Lambda Q(z).$$

The monic\(^1\) matrix polynomial $Q(z) = \sum_{m=0}^{\infty} Q_m z^{-m}$ is causal and minimum-phase, and $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_4\}$ is a real-valued diagonal matrix. Also [100],

\(^1\) A polynomial with the zero index value equal to one is said to be monic.
This leads to signal-to-noise ratio of channel \( n \) as:

\[
\text{SNR}_{\text{MMSE-DFE}}^n = \frac{\sigma_e^2}{\sigma_v^2} = \frac{\lambda_n}{\sigma_v^2}
\]  

The capacity in this coordinated multi-channel reads

\[
C' = W \sum_{n=1}^{4} \log_2 \left( 1 + \frac{\lambda_n}{\Gamma \sigma_v^2} \right) \approx W \sum_{n=1}^{4} \log_2 \left( \frac{\lambda_n}{\Gamma \sigma_v^2} \right) \]

\[
= W \log_2 \left( \frac{1}{\Gamma} \det \left( \frac{\Lambda}{\sigma_v^2} \right) \right)
\]  

where in this approximation we assumed that all the channels are well behaved, i.e. \( \lambda_n \gg \sigma_v^2, n = 1, \ldots, 4 \). This is, indeed, significantly related to the correct past decisions in the derivation of MIMO-DFE. This reveals that Eq. (3.13) is a lower bound for \( C' \), the capacity of coordinated multi-channel system.
3.5 The Loss Between SISO and MIMO Implementations

It is quite common in practice that in multi-channel systems, the individual channels are equalized independently, and crosstalk terms from other channels are removed by fixed or adaptive cancellers. Therefore, the channel matrix \( \mathbf{H} \) can be rewritten as

\[
\mathbf{H} = \mathbf{D} + \mathbf{F}
\]

where \( \mathbf{D} \) and \( \mathbf{F} \) are the polynomial matrices containing the diagonal and off-diagonal elements of \( \mathbf{H} \), respectively (i.e. \( \mathbf{D} = \text{diag} \{ \mathbf{H}_{1,1}, \mathbf{H}_{2,2}, \mathbf{H}_{3,3}, \mathbf{H}_{4,4} \} \) and \( \mathbf{F} = \mathbf{H} - \mathbf{D} \)).

The decomposition of \( \mathbf{H} \) in Eq. (3.17) can be interpreted differently. One can consider this system as a multiple access channel (MAC) with two users, as shown in Fig. 3.3. If the detection starts with user 1, the maximum rate of this user is [101]

\[
C_1 = \int_0^W \log_2 \det \left( \mathbf{I}_4 + P(f)\mathbf{D}(f)\mathbf{R}_{nn}^{-1}(f)\mathbf{D}^\dagger(f) \right) df
\]

(3.18)

where \( \mathbf{R}_{nn} = \sigma_v^2 \mathbf{I}_4 + P(f)\mathbf{F}(f)\mathbf{F}^\dagger(f) \) and \( \int_0^W P(f)df = \sigma_v^2 \). Therefore, if the rate of user 1 fulfills \( R_1 < C_1 \), it can be detected without errors and, hence, removed from the received signal. The remaining signal used by user 2 is now only impaired by the thermal noise, leading to its maximum rate:

\[
R_2 = C_2 = \int_0^W \log_2 \det \left( \mathbf{I}_4 + \frac{P(f)}{\sigma_v^2} \mathbf{F}(f)\mathbf{F}^\dagger(f) \right) df
\]

(3.19)

Recall from multi-user detection theory [101] that \( R_1 + R_2 \) is bounded above by the capacity of the channel, \( C_{\text{MIMO}} \). This leads to the conclusion that \( R_1 \leq C_{\text{MIMO}} - R_1 \), which means the interference terms must be attenuated enough to achieve high-capacity SISO implementation for
user 2. By this method, one can achieve reliable bounds for FEXT attenuation factors, $\Gamma^F$. A similar approach can be used to determine the proper attenuation level for NEXT, $\Gamma^N$.

The above analysis is based on MIMO and SISO Shannon capacities. For single-carrier capacities, one can easily replace the corresponding bounds and follow the same analysis.

### 3.6 Rate Optimization for Beyond-10G

In ultra high-speed applications where the trade-offs of power consumption, implementation complexity, and reliability are dramatically challenging, it is of considerable importance to study the problem of input symbol rate optimization under practical realizability constraints. We obtain the optimum specifications for systems equalized by an ideal (no error propagation) infinite-length MMSE-DFE.

#### 3.6.1 Minimizing the Probability of Error

Here, we are interested in achieving a fixed target bit rate while keeping the probability of error $p_e$ as small as possible. This can be achieved by performance analysis of the coded transmission system.
system and link budget analysis for decision feedback equalizer (DFE) implementation. We assume the same signal constellation $\mathcal{A}$ (Eq. 3.1) is used for each individual channel (dimension). We also suppose the power is equally divided among the transmitters. Under these conditions, in CAT-7A cable, it is fairly reasonable to assume same average error probability for individual channels. The union bound estimate of the probability of symbol decoding error associated with each of these constellations is \[39\]

$$p_e \simeq K_{\text{min}} Q\left(\sqrt{3\text{SNR}_{\text{norm}}} \frac{\gamma_c(\Lambda)\gamma_s(\Lambda)}{\gamma_m}\right) \quad (3.20)$$

where $K_{\text{min}}$ is the multiplicity of codewords with minimum weight, $\gamma_c(\Lambda)$ is the nominal coding gain associated with set partitioning, $\gamma_s(\Lambda)$ is the shaping gain, and $\gamma_m$ is the desired system margin. The $\text{SNR}_{\text{norm}}$ is the normalized SNR and signifies how far a system is operating from the Shannon limit (the gap to capacity). More importantly, this quantity is independent of constellation size for large signal spaces. This, in fact, significantly simplifies the underlying analysis. For our QAM baseline, we have $\gamma_s(\Lambda) = 1$ and $\gamma_c(\Lambda) = d_{\text{min}}^2(\Lambda)$. Therefore, $p_e$ reduces to:

$$p_e = K_{\text{min}} Q\left(\sqrt{3\text{SNR}_{\text{norm}}} \times \frac{d_{\text{min}}^2(\Lambda)}{\gamma_m}\right) \quad (3.21)$$

We should recall that the $\text{SNR}_{\text{norm}}$ defined in this section is the same SNR gap $\Gamma$ used in Eq. (3.9), (3.10) and (3.11). From Eq. (3.20), it is clear that $p_e$ is minimized when the $\text{SNR}_{\text{norm}}$ is maximized (assuming $d_{\text{min}}^2(\Lambda)$ and $\gamma_m$ are fixed). One can calculate the SNR gap $\Gamma$ at each frequency from Eq. (3.15) and substitute into Eq. (3.20) to obtain the error probability versus bandwidth.
3.6.2 Maximizing the System Margin

Alternatively, the system designer may want to choose a specific reliability level, and seek to maximize the system margin to account for unforeseen sources of performance degradation. By rearranging Eq. (3.20), one can define the system margin in terms of error probability, coding gain, and gap to capacity as

\[
\gamma_m = \frac{3\text{SNR}_{\text{norm}} s_i^2}{\left(Q^{-1}(p_e / K_{\text{min}})\right)^2}
\]

(3.22)

This means that, assuming a fixed coding gain, the two optimization scenarios, one that minimizes \( p_e \) for fixed \( \gamma_m \), and the other that maximizes \( \gamma_m \) for fixed \( p_e \), lead to the same optimum bandwidth. In fact, in both cases \( \Gamma \) is maximized.

3.7 Simulations and Results

The SISO single-carrier and water-filling bounds are calculated for 20m and 50m CAT-7A cables. Although these cables are much better in terms of isolation and thermal noise compared to other UTP cables, we set the background noise level to -146dBm/Hz in our simulations [53]. This is mainly because the noise from analog front-end is dominant in these systems. The capacity bounds and the corresponding optimum bandwidths can be obtained from Fig. 3.4. The MIMO capacity and user 1 admissible rate are also calculated for these cables and are shown in Fig. 3.5. The parameters for these simulations are listed in Table 3.1. Also, echo interference is assumed to be attenuated by 65dB while no NEXT and FEXT cancellation are applied.
Table 3.1: Parameters for capacity analysis.

<table>
<thead>
<tr>
<th></th>
<th>Tx Power [dBm]</th>
<th>Margin [dB]</th>
<th>Probability of error</th>
<th>Coding gain [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20m</td>
<td>10</td>
<td>6.0</td>
<td>$10^{-12}$</td>
<td>0.0</td>
</tr>
<tr>
<td>50m</td>
<td>10</td>
<td>6.0</td>
<td>$10^{-12}$</td>
<td>6.0</td>
</tr>
</tbody>
</table>

There are a few remarks about Fig. 3.5 explained as follows. First, for 50m cable, the maximum data rate of user 1 is only 3Gbps less than the total MIMO capacity of this channel which indicates that the amount of information carried by the FEXT channels is negligible. Second, the SISO capacity of this cable has a maximum of 93Gbps without any FEXT and NEXT cancellation which is about the same as the maximum rate of user 1. This proves that the channels are isolated from each other very well, and they perform almost as well as isolated parallel channels. Finally, despite the fact that MIMO outperforms SISO system, it results in a very minor improvement over SISO implementation but at a higher cost and power consumption which may not be acceptable.

Figure 3.4: SISO Shannon, single-carrier, and water-filling capacity bounds.
Some designers may argue about high transmit power and the stresses that it may cause in future submicron CMOS technologies and try to reduce the transmit power at the cost of more complex and sophisticated codes. We repeated the SISO capacity analysis for 50m cable with 5dBm transmit power, which causes less non-linearity in the line driver. The capacity bounds obviously drop from their maximum points in Fig. 3.4. We also repeated this analysis for 9dB coding gain. To avoid the extra cost of this complex code, one can cut back the cable length by a few meters to keep the less complex code with 6dB gain. The capacity bounds corresponding to these two codes at 5dBm transmit power in presented in Table 3.2. The capacity bound values in this table are in Gbps.

Table 3.2: Capacity bounds for 50m cable and 3dBm Tx power.

<table>
<thead>
<tr>
<th>g_c</th>
<th>Shannon bound</th>
<th>Single-Carrier bound</th>
<th>Water-filling bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>6dB</td>
<td>79.76</td>
<td>43.69</td>
<td>44.35</td>
</tr>
<tr>
<td>9dB</td>
<td>79.76</td>
<td>51.08</td>
<td>51.75</td>
</tr>
</tbody>
</table>

Figure 3.5: MIMO capacity and R_1 rate for 20 and 50 meters cables.
Now, we consider two communication systems transporting data at rates 40 and 100Gbps over 50m and 20m, respectively, of horizontal balanced CAT-7A cables. These systems are equalized by an ideal infinite-length MMSE-DFE. Fig. 3.6 depicts the variation of \( p_e \) as a function of bandwidth \( W \) for a target system margin of 0dB (other parameters are kept fixed).

![Figure 3.6: Minimization of probability of error for two 40 and 100GBASE-T systems.](image)

The system margins versus bandwidth of these systems are shown in Fig. 3.7. It is observed that, the system transmitting data at 40Gbps over 50m passes the 6dB margin requirement (conventionally, a 6dB margin is considered for multi-gigabit transmission over copper) if a 6dB coding gain is available at bandwidth around 1.6GHz. A 3dB coding gain for a system transmitting data at 100Gbps results in 10dB margin around 5.0GHz.

Some remarks are worth mentioning here. First, the probability of error minimization and margin maximization occur at the same optimum bandwidth. In general, this bandwidth can be different from the bandwidth obtained under different optimization criteria, e.g., power minimizing bandwidth of the MMSE-DFE. Second, figures 3.6 and 3.7 do not show any symmetry about the optimum bandwidth. More precisely, one can, in general, over-estimate the optimal bandwidth by a few percent instead of under-estimating it without any serious
degradation (the flatness of curves around the optimum point differ in different scenarios). Finally, it is apparent from figures 3.6 and 3.7 that the DFE can suffer significant performance degradation when the transmission bandwidth is not optimized. Therefore, for high data rate applications, the process of rate optimization becomes extremely important.

![Figure 3.7: Maximization of margin for two 40 and 100GBASE-T systems.](image)

### 3.8 40GBASE-T Full and Short Reach Modes

One of the primary objectives of this research is to obtain (or define) the reach mode of 40GBASE-T system, meaning the range of cable length in which data is properly transmitted for a given cost and complexity. Usually two modes of operation are considered in Ethernet applications: short-reach mode and full-reach mode. In full-reach mode, all the signal processing blocks required for proper data transmission over maximum cable length under practical considerations, i.e. single-carrier transmission, are employed. This mode usually involves sophisticated signal processing and subsequently higher power and latency. In short-reach mode, most of the complicated blocks are reduced in order to establish a low cost, low power and low latency communication link over a shorter range. In 10GBASE-T, full-reach mode and short-
reach mode are defined as 100m and 30m, respectively. As we already demonstrated, 40GBASE-T is feasible up to 50m over CAT-7A cable with the corresponding assumptions. Fig. 3.8 illustrates the full-reach mode of this system at different transmit power level when a coding gain of 6dB is applied and at least 6dB system margin is considered. Obviously the lower the transmit power level, the shorter full-reach mode. This can be improved exploiting higher coding gains with more complex architectures. We should emphasize that these bounds are a little optimistic since we assumed that there are no residual effects from echo, and NEXT/FEXT crosstalk cancellers.

![Figure 3.8: System margin of CAT-7A cable ($g_c=6$dB). The full-reach mode is defined when the system margin is above the required 6dB margin.](image)

The block diagram of 40GBASE-T short-reach mode is shown in Fig. 3.9 which only includes the channel equalizer, echo canceller, line drivers and the external interface. Since the FEXT and NEXT crosstalk levels are already low enough to eliminate the corresponding cancellers, we just need to find out the minimum cable length at which the system margin is satisfied without any coding scheme.
We performed the margin versus length analysis for CAT7A cable; the result is depicted in Fig. 3.10. Increasing the transmit power may allow us to extend this short-reach mode by few more meters.

![Figure 3.9: 40GBASE-T short-reach mode block diagram.](image)

![Figure 3.10: System margin of CAT-7A cable for short-reach mode system (g_c=0dB).](image)

Besides, to obtain a bound on insertion loss for a cable that would be able to support 40Gbps up to 100m, we deploy an algorithm to derive a hypothetical cable model. This mode can be used as the benchmark by cable manufacturers for further improvements in current CAT-7A cable or developing new cables.
In this algorithm, we assume certain characteristics for insertion loss, and through a reverse approach we find a minimum requirement on insertion loss (as it is the main factor in determining the capacity bounds) for a system supporting 40GBAS-T. The model that we adopted to obtain this bound is as follows:

$$IL(f, p)_{[dB]} = p_0 \left( p_1 \sqrt{f} + p_2 f + \frac{p_3}{\sqrt{f}} \right)$$  \hspace{1cm} (3.23)

where \( f \) is frequency in MHz, and \( p = [p_0, \ldots, p_3] \) is a set of unknown constants to be determined. If \( \gamma_m(IL, K) \) represents the margin function for a cable characterized by its insertion loss \( IL \) and a set of parameters \( K \), then we would like to have

$$\max \left\{ \gamma_m(IL(f, p), K) \right\} \geq \gamma_c, \quad f \in [0, W]$$  \hspace{1cm} (3.24)

where \( W \) is some frequency usually greater than the optimum bandwidth, e.g., 2-3 GHz for 40GBASE-T system, and \( \gamma_c \) is the desirable system margin. We turn this problem into the following more tractable convex optimization problem:

$$\begin{align*}
\text{minimize} \quad & s \\
\text{subject to} \quad & \gamma_m(IL(f, p), K) + s = \gamma_c \\
& s \geq 0
\end{align*}$$  \hspace{1cm} (3.25)

A 6dB coding gain, 6dB system margin and error probability of \( 10^{-12} \) are assumed. The results of this optimization are shown in Fig. 3.11 for two different transmit power levels. From this figure, one can conclude that, for example CAT-7A and CAT-6A cables, both can deliver 40Gbps with up to 50m, with 10dBm transmit power. However, CAT6A cable fails to support 40Gbps if transmit power level is reduced to 5dBm.
Figure 3.11: Bound on insertion loss to support 40Gbps. Different TX power levels give different bound. CAT7A 50m and CAT6A 50m, both have insertion loss above this bound for 10dBm.
Chapter 4
Equalization and Precoding

4.1 Introduction

Digital transmissions over bandwidth-limited channels are generally more restricted by intersymbol interference (ISI) rather than by the additive noise. Additive noise is usually mitigated through some channel coding techniques, while ISI can seriously degrade the reliability of high-speed data transmission even at very high SNR. Pulse shaping and equalization are commonly deployed to cope with this problem.

In this chapter, we aim to provide some background materials regarding the pulse shaping and channel equalization. Filter-based equalizations due to their low-complexity are our focus. While equalizers aids the ISI mitigation at the receiver side, precoding techniques do the same at the transmitter side. This generally prevents error propagation that is usually associated to data-aided equalization methods. Tomlinson-Harashima precoding technique is discussed and some guidelines are provided. Finally simulation results are provided for 40GBASE-T system.

4.2 Pulse Shaping

In digital transmission over bandwidth-limited channels, transmitter and receiver filters with finite impulse response (FIR) are used. These filters are designed to achieve the optimum spectral concentration in the transmission bandwidth of the channel, and zero intersymbol interference (ISI) is obtained when the filters operate in cascade. In order to maximize signal-to-noise ratio in the presence of additive noise and for practical considerations, matched transmitter
and receiver filters are considered. This means that identical filter coefficients occur in both filters, but in time-reversed order with respect to one another. For baseband transmission, the maximum spectral concentration should occur in the region \([- (1 + \beta)/2T, (1 + \beta)/2T]\) where \(1/T\) denotes the symbol rate and \(\beta\) determines the excess bandwidth allowed. Parameters of the filters are the number of coefficients \(N\) and the sampling rate \(\mu/T\), where \(\mu\) is an integer oversampling factor. By allowing the impulse response to exceed \(T\), better spectral concentration can be obtained. However, it then becomes necessary to enforce zero ISI explicitly in the overall response of the filters.

Raised cosine pulse spectrum, which satisfies the Nyquist condition for zero intersymbol interference [72][86], has been widely used in practice in digital communication systems. The raised cosine pulse shaping is given by

\[
p(t) = \mathcal{F}^{-1}\left\{ \frac{T}{T \sqrt{1 - \frac{1}{2} \sin \left( \frac{2\pi}{\alpha} \left( |f| - \frac{1}{2T} \right) \right)}} \right\}, \quad \left| f \right| \leq \frac{1-\alpha}{2T},
\]

where \(\alpha\), \(0 \leq \alpha \leq 1\), indicates excess bandwidth and \(\mathcal{F}^{-1}\) is the inverse Fourier transform. Residual ISI decreases with increasing \(N\) and \(\alpha\). In [Error! Not a valid link.], the same joint optimization of matched transmitter and received filters under the constraint of overall zero ISI is considered. However, it is recognized that asymmetric filter responses, i.e., responses with nonlinear phase characteristic, provide a better spectral concentration. We follow [Error! Not a valid link.] to obtain the asymmetric optimal pulse shape filter.

Consider the baseband data transmission system. The \(N\)-tap transmitter filter is characterized by the coefficient vector \(g = [g_0, g_1, \ldots, g_{N-1}]^T\) and is clocked at rate \(\mu/T\) corresponding to an oversampling factor \(\mu\). Every \(T\) seconds an information symbol \(x_k\) enters
the transmitter filter followed by \( \mu - 1 \) zeros. Passing the output signal of the transmitter filter through a receiver filter with matched coefficient vector \( \tilde{g} = [g_{N-1}, g_{N-2}, \ldots, g_0]^T \) leads to

\[
y[\mu n + N - 1] = \sum_l x_{k-l} \sum_{i=0}^{N-1} g_i g_{i+\mu l}
\]  

(4.2)

The cascaded filters introduce a delay of \((N-1)T/\mu\) seconds. It is shown in [Error! Not a valid link.] that optimum pulse shape \( g \) having maximum spectral concentration within \([-(1+\beta)/2T, (1+\beta)/2T]\) can be obtained by following optimization problem:

\[
\begin{align*}
\text{maximize} & \quad g^T R g \\
\text{subject to} & \quad g^T S_l g = \delta_l, \quad |l| \leq (N-1)/\mu
\end{align*}
\]  

(4.3)

where

- \( R \) is a positive-definite \( N \times N \) Toeplitz matrix with elements

\[
r_{ik} = \frac{1+\beta}{T} \begin{cases} 
1 & i=k \\
\frac{\sin \left[ \pi(1+\beta)(i-k)/\mu \right]}{\pi(1+\beta)(i-k)/\mu} & i \neq k
\end{cases}
\]

- \( S_l \) is a "shift" matrix which its elements are zero, except \( s_{l,ik} = 1 \) for \( k-i = \mu l \).

- \( \delta_l \) is the Kronecker symbol, and \( \lfloor t \rfloor \) denotes the integer portion of \( t \), from below.

Numerical solutions are obtained by the iterative projected gradient method \[71][13]. For a given excess bandwidth factor \( \beta \) and oversampling factor \( \mu \), the achievable spectral concentration increases with the filter length. The optimum coefficients are asymmetric. Symmetric coefficients can easily be enforced in the filter design discussed.
4.3 The ISI-channel Model

The linear ISI channels can be modeled as shown in Fig. 4.1. In this model, the cascades of continuous-time transmit filter $H_T(\omega)$, channel $H_C(\omega)$, receiver filter $H_R(\omega)$ and symbol-rate sampler are represented by the equivalent discrete-time filter $H(z)$. The data symbols $\{x_k\}$, transmitted at the rate $R = 1/T$, are modeled as a sequence of independent, identically distributed (i.i.d.) random variables, with mean zero and variance $\sigma_x^2$. The additive noise samples $\{\nu_k\}$ are modeled as a zero-mean, wide-sense stationary, white Gaussian random process with total variance $\sigma_\nu^2$.

![Figure 4.1: Equivalent discrete-time channel.](image)

The frequency response of this discrete-time channel is obtained by letting $z = e^{j\omega T}$, and it is the folded-spectrum of the total continuous-time system, i.e.,

$$H(\omega) = \frac{1}{T} \sum_m H_T\left(\omega - \frac{2m\pi}{T}\right) H_C\left(\omega - \frac{2m\pi}{T}\right) H_R\left(\omega - \frac{2m\pi}{T}\right) \quad |\omega| \leq \pi/T \quad (4.4)$$

The discrete-time equivalent channel model is conveniently described by its $z$-transform

$$H(z) = \sum_{\ell=0}^{v_h} h_{\ell} z^{-\ell},$$

where we assumed the ISI channel is modeled as a finite impulse response (FIR) filter with $v_h+1$ coefficients $\{h_0, \ldots, h_{v_h}\}$. Therefore, the ISI-channel can be modeled as:
The ISI and noise are unwanted signals that act to distort the information being transmitted. The receiver estimates the transmitted data sequence from a set of sufficient statistics \( \{y_k\} \). A lower bound on the performance of any receiver in the presence of ISI is called the **matched filter bound** (MFB). This bound is obtained for an optimum receiver under ISI-free transmission. The SNR corresponding to the matched filter bound is defined as:

\[
\text{SNR}_{MFB} = \frac{\sigma_s^2 \|h\|_2^2}{\sigma_v^2} = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{\sigma_s^2}{\sigma_v^2} |H(\omega)|^2 d\omega
\]  

In this dissertation, we shall focus on filter-based equalizers due to their ease of implementation. The optimum receiver, maximum-likelihood sequence detection (MLSD), has an exponentially increasing complexity which prohibits them for our ultra high-speed applications.

### 4.3.1 MMSE Linear Equalizer

The block diagram of MMSE linear equalizer is depicted in Fig. 4.2. The output sequence of the noisy ISI channel \( y_k \) is passed through a linear filter. The optimum transfer function of the MMSE linear equalizer results from the minimization of the MSE \( \varepsilon^2 = \mathbb{E}\{\|\tau_k - x_k\|^2\} \). Unfortunately, the MMSE linear equalizer is a biased estimator that should be compensated in order to reduce the probability of error.

Figure 4.2: Block diagram of unbiased MMSE linear equalizer.
The MMSE linear equalizer is, in fact, the well-known Wiener filter expressed as:

\[ W(z) = \frac{\sigma_s^2 H^* \left( \frac{1}{z} \right)}{\sigma_s^2 H^* \left( \frac{1}{z} \right) + \sigma_v^2} \]  \hfill (4.7)

The corresponding minimum MMSE is given by:

\[ \varepsilon^2_{\text{min}} = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} R_{\text{ee}}(e^{j\omega}) d\omega = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{1}{1 + \frac{\sigma_v^2}{\sigma_s^2 |H(e^{j\omega})|^2}} d\omega \]  \hfill (4.8)

where \( R_{\text{ee}}(e^{j\omega}) \) is the characteristic function of error sequence \( e_k = r_k - x_k \). For a finite-length linear equalizer with \( N_f \) taps, i.e. \( w = [w_0, \ldots, w_{N_f-1}]^T \), the equalized sample at time \( k \) is given by \( r_k = w^T y_{kk-N_f} \). Minimizing the mean squared error \( J = \mathbb{E}\{|e_k|^2\} \) yields \([58]\)

\[ w_{\text{opt}} = R_{yy}^{-1} R_{yx} \]  \hfill (4.9)

where the following notations are introduced:

\[ R_{yy} = \mathbb{E}\{y_{kk-N_f} y_{kk-N_f}'\} = \sigma_s^2 HH' + \sigma_v^2 \mathbf{I}_{N_f+1} \]  \hfill (4.10)

\[ R_{yx} = \mathbb{E}\{y_{kk-N_f} x_{k-\Delta}\} = \sigma_s^2 H e_{\Delta} \]  \hfill (4.11)

and \( H \) is the \( N_f \times (N_f + v_h) \) channel convolution matrix defined as:

\[ H = \begin{bmatrix} h_0 & h_1 & \cdots & h_{v_h} & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_{v_h} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_{v_h} \end{bmatrix} \]  \hfill (4.12)

The corresponding MSE is given by:

\[ J_{\text{MMSE-LE}} = \sigma_s^2 R_{yx}^{-1} R_{yy}^{-1} R_{yx} \]  \hfill (4.13)
We note that the MSE and hence the output SNR are functions of the delay $\Delta$. With infinite-length filters, the need for such a delay does not enter the mathematics because the infinite-length filters are not realizable in any case. This parameter should be also optimized to achieve the highest performance of this equalizer. The optimum delay is obtained as follows. In Eq. (4.13), $J_{\text{MMSE-LE}}$ is a function of the delay $\Delta$ through $R_{yx}$, i.e.,

$$
J_{\text{MMSE-LE}}(\Delta) = \sigma_s^2 - R_{yx}^t R_{yy}^{-1} R_{yx} \\
= \sigma_s^2 - \sigma_s^2 e_\Delta^t H^t R_{yy}^{-1} \sigma_s^2 H e_\Delta \\
= \sigma_s^2 (1 - \sigma_s^2 e_\Delta^t K e_\Delta)
$$

where $K = H^t R_{yy}^{-1} H$. Note that $e_\Delta^t K e_\Delta = K_{\Delta\Delta}$. Therefore, calculation of optimum delay can be summarized as:

$$
\Delta_{opt} = \arg\min_{0 \leq \Delta \leq N_f + n_f + 1} K_{ii}
$$

### 4.4 Decision-Feedback Equalizer

Another important class of channel equalization is decision-feedback equalization. The decision-feedback equalizer (DFE) uses the fact that in detecting the symbol $x_k$, prior detected symbols can be employed as these symbols contain information that is correlated with the ISI term in $y_k$. Decision feedback equalizers attempt to implement this strategy and are therefore better suited for channels with pronounced ISI. The block diagram of this equalizer is shown in Fig. 4.3. In addition to using a filter $w$ in the feedforward path, as in linear equalizer, a DFE employs a feedback filter in order to feed the previous decisions back into the decision device and uses them to reduce the ISI. The key step in deriving the optimum solution for feedforward and
feedback filters is to apply the spectral factorization on the z-transform $R_{yy}(z)$ of the autocorrelation of the received observations $\{y_k\}$ as:

$$R_{yy}(z) = \sigma_s^2 H(z)H^*(1/z^*) + \sigma_v^2 = \gamma_0 F(z)F^*(1/z^*)$$  \hspace{1cm} (4.16)

where $R_{yy}(z)$ is factored into the product of a monic minimum-phase filter $F(z)$ with a monic maximum-phase filter $F^*(1/z^*)$. The MMSE optimum transfer functions for the feedforward and feedback filters are then given by:

$$W(z) = \frac{\sigma_v^2 H^*(1/z^*)}{\gamma_0 F^*(1/z^*)}$$  \hspace{1cm} (4.17)

$$B(z) = F(z) - 1$$  \hspace{1cm} (4.18)

The corresponding MSE is given by:

$$J_{\text{MMSE-DFE}} = \frac{\sigma_s^2 \sigma_v^2}{\gamma_0} = \sigma_s^2 \exp \left\{ -\frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \log \left( 1 + \frac{\sigma_s^2 |H(e^{j\omega})|^2}{\sigma_v^2} \right) d\omega \right\}$$  \hspace{1cm} (4.19)

The analysis of finite-length decision feedback equalizer (FIR DFE) is very similar to the finite-length MMSE linear equalizer, except that we now augment the feedforward section with a feedback section [2, Error! Not a valid link., 5]. The MSE for the FIR DFE case, as long as $w$ and $b$ are sufficiently long, is:

$$\text{MSE} = \mathbb{E}\{ |x_{k-\Delta} - w^t y_{kk-\Delta-1} - b^t x_{k-\Delta-bk-\Delta-N_e} |^2 \}$$  \hspace{1cm} (4.20)
where it is assumed that after a delay of $\Delta + 1$ symbols, estimates are fed into the feedback path. The MMSE-optimum filters are given by [2]

$$
\tilde{b}_{opt} = L e_\Delta
$$

$$
\tilde{w}_{opt} = \frac{\sigma_s^2 H(L')^{-1} e_\Delta}{d_\Delta}
$$

where

- $L$ is an $(N_f + v_h) \times (N_f + v_h)$ lower triangular matrix, with 1’s on the main diagonal, obtained by the Cholesky factorization $\sigma_s^2 H' H + \sigma_\nu^2 I = L D L'$.  

- $D$ is an $(N_f + v_h) \times (N_f + v_h)$ diagonal matrix with positive real entries, i.e.  
  $$D = \text{diag}(d_0, \ldots, d_{N_f + v_h - 1})$$.

- the augmented feedback vector $\tilde{b}$ is defined as $\tilde{b} = [0_{v \Delta} \ 1 \ b' \ 0_{v_s}]$, and $s$ satisfies  
  $$N_f + v_h = \Delta + 1 + N_b + s.$$  

The minimum SNR at the output of equalizer is:

$$
J_{\text{MMSE-FIR-DFE}} = \frac{\sigma_s^2 \sigma_\nu^2}{d_\Delta}
$$

The delay parameter $\Delta$ has to be optimized to achieve the best performance for the finite-length MMSE DFE. It can be shown that optimum delay is [2]:

$$
\Delta_{opt} = \arg \min_{0 \leq i \leq N_f + v_h - 1} d_i
$$
4.5 Tomlinson-Harashima Precoding

The decision-feedback equalizer is very attractive on severely ISI distorted channels mostly for its optimality and simple structure. Unfortunately, DFE suffers from the error propagation, and channel coding cannot be applied in a straightforward manner. In DFE (or other equalization schemes) the primary task of the receiver is to remove the ISI and basically convert the ISI channel to an overall flat front-end channel. The ISI-channel model given by Eq. (4.5) can be written as

\[ y_k = \sum_{\ell=0}^{v_h} h_\ell x_{k-\ell} + \nu_k = x_k + \sum_{\ell=1}^{v_h} h_\ell x_{k-\ell} + \nu_k \]  

(4.25)

where, without loss of generality, \( h \) is normalized so that \( h_0 = 1 \). The second term in Eq. (4.25), that is \( f_k = \sum_{\ell=1}^{v_h} h_\ell x_{k-\ell} \), corresponds to the intersymbol interference, which is uniquely determined by the preceding transmitted symbols \( x_{k-\ell}, \ell = 1, \ldots, v_h \). Therefore, if one chooses the transmitting symbols as \( \tilde{x}_k = x_k - f_k \), the channel output reduces to \( y_k = x_k + \nu_k \). Thus, the estimations of original symbols \( x_k \) can be made without disturbance from intersymbol interference. An implementation corresponding to Eq. (4.25) is shown in Fig. 4.4(a), where the transfer function of the feedback filter is given by \( \delta_k - h_k \). This scheme is known as linear precoding.

![Figure 4.4: Block diagram of linear pre-equalization and Tomlinson-Harashima precoding.](image-url)
In linear precoding, despite its simple structure, there is a serious problem that the peak value of transmitted pulse sequence $\tilde{x}_k$ tends to increase or sometimes diverge to infinity. In the Tomlinson-Harashima precoding technique [93][49], which will be described, this difficulty is solved by insertion of an additional nonlinear signal transformation $\Gamma(.)_{2M}$, an arithmetic modulo operation, as shown in Fig. 4.4(b). The signal transformation $\Gamma(.)_{2M}$ is applied so that the sequence $\tilde{x}_k$ is peak-limited, i.e., $-M \leq \tilde{x}_k \leq +M$. The modulo-$2M$ operation can be characterized in various ways. By definition, the transmitted symbol $\tilde{x}_k$ is a unique number that satisfies the constraint:

$$\tilde{x}_k = \Gamma \left( x_k - \sum_{l \geq 1} h_l \tilde{x}_{k-l} \right)_{2M} \quad (4.26)$$

In other words, the transmitter finds the unique integer $d_k$ such that

$$\tilde{x}_k = x_k - \sum_{l \geq 1} h_l \tilde{x}_{k-l} + 2M d_k \quad (4.27)$$

is in the interval $(-M, M]$. In q-transform notation: $\tilde{x}_k = x_k - (h(q)-1)\tilde{x}_k + 2M d_k$. This implies $\tilde{x}_k = \left[x_k + 2M d_k\right]/h(q)$. Consequently, the received signal is:

$$y_k = h(q)\tilde{x}_k + \nu_k = x_k + 2M d_k + \nu_k \quad (4.28)$$

Thus, in the absence of noise, the received sequence is the modified data sequence, which may be detected on a symbol-by-symbol basis to give the estimated sequence $\hat{x}_k$. There is no ISI, and the error probability is the same as if the original data sequence $x_k$ were sent on the same channel and detected with a zero-forcing DFE. The estimates $\hat{x}_k$ of the original data symbols $x_k$ can then be retrieved by reducing $y_k$ to the interval $(-M, M]$ with a modulo-$2M$ operation.
Therefore, using this method allows complete removal of interference due to preceding symbols under the constraint that the peak signal power is limited. It can be shown that this system is able to attain asymptotically the performance bound predicted by information theory [93].

In general, in the TH precoding technique, the levels of the transmitting sequence are not necessarily discrete but rather are distributed continuously between $-M$ and $+M$. If, in particular, $h_k$ is a set of integers with the greatest common divisor equal to 1, the levels of the transmitting sequence $x_k$ reduce to $M$ discrete values. This special case is easily shown to be identical to the correlative level coding technique developed by Lender [63] or the partial-response technique by Kretzmer [61]. For example, $\{h_0, h_1\} = \{1, 1\}$ corresponds to the duobinary technique or class 1 ($n = 2$) partial response.

In the conventional techniques, $\{h_k\}$ must be integers, but in the TH precoding technique one can transmit the data over the channels where $h_k$ are not necessarily integers. Further, the TH precoding technique allows transmission of analog data so far as the constraint of peak limitation of data is satisfied.

### 4.5.1 MMSE-THP

The precoding schemes originally proposed by Tomlinson [93] and Harashima [49] (independently though), are indeed equivalent to Zero-Forcing (ZF) equalization. A better performance is obtained by minimizing the mean square error (MMSE). The block diagram of MMSE-THP is shown in Fig. 4.5.
$B(q)$ must be causal and monic so that the feedback filter $1 - B(q)$ in Fig. 4.5 requires only previous values of $\hat{x}_k$. $W(q)$ may be any linear time-invariant filter. The unbiased MMSE-THP is obtained by choosing $W(q)$ and $B(q)$ to minimize $E\{|\tilde{n}_k + r_k|^2\}$ under the constraint that $B(q)$ and $W(q)$ are monic. Spectral factorization techniques (or Cholesky factorization for finite-length filters) provide the desired filters, which are exactly the values of $B(q)$ and $W(q)$ used in unbiased MMSE-DFE.

### 4.6 Simulations and Results

Fig. 4.6 shows the impulse responses of optimum symmetric and asymmetric filters for $N=24$, $\mu=4$, and $\beta=0.1$.

Figure 4.6: Optimum asymmetric (left) and raised-cosine (right) FIR pulse shapes. ($N=24$, $\beta=0.1$)
The spectra of the optimum asymmetric filter with $N = 24$, $\mu = 4$ and $\beta = 0.1$, and a conventional raised-cosine filter obtained by Eq. (4.1) with $N = 24$, $\mu = 4$, and $\alpha = 0.1$ are compared in Fig. 4.7. Residual mean-square ISI is 31.2dB below signal level when two filters are in cascade. The optimum asymmetric filter leads to a better spectral concentration with a guaranteed zero ISI. The conventionally designed filter requires larger excess bandwidth for acceptable ISI, but then achieves smaller side-lobes. For the optimum asymmetric filter, comparable side-lobe magnitude can be obtained by increasing $\beta$. The optimum asymmetric filter then still has the advantages of a slightly better spectral concentration and guaranteed zero ISI. The improvement in spectral concentration achieved by allowing asymmetric filter coefficients is not very large, but neither is negligible. In fact, the asymmetric filter leads to about 30 percent less energy in the stop-band region than the symmetric filter.

Figure 4.7: Power spectrum estimate of random binary signal using raised-cosine & optimum asymmetric FIR pulse shapes.

The measurements of insertion loss of twisted-pairs of a CAT-7A cable (50m) are shown in Fig. 4.8. The corresponding impulse response of the blue pair sampled at 3.2GHz is shown in Fig. 4.9. The SNR of MMSE linear equalizer is only 31.36dB with 200 taps. In Fig. 4.10, the performance of finite-length DFE versus the number of feedforward taps is calculated for this channel and compared to the optimum infinite-length decision-feedback equalizer.
We observe that the performance of FIR DFE improves very little for feedforward taps more than 100. This can be explained easily as follows. In DFE, the feedforward filter conceptually shapes the discrete-time channel into a causal transfer function, i.e. eliminates the post-cursor part of the impulse response. The impulse response of our example is very close to a unilateral impulse response (up to a finite delay), meaning the post-cursor part of the channel has very small power. Therefore, even short feedforward filters could easily invert this channel into a transfer function with very small post-cursor taps.

Figure 4.8: Insertion loss measurements of 50m CAT-7A cable.

Figure 4.9: Impulse response of 50m CAT-7A cable, R_s=3200MHz.
The Leas-Mean-Square (LMS) algorithm was used to adaptively equalize this channel. The number of feedforward and feedback taps for the equalizer were $N_f = 200$ and $N_b = 160$. Fig. 4.11 illustrates the initial convergence characteristics of the LMS algorithm for $\mu = 0.001$. Other values of step-size result in different convergence rates and excess mean square errors [51].

![Figure 4.10: Signal-to-noise ratio of unbiased MMSE decision-feedback equalizer versus forward taps.](image)

The performance of FIR MMSE-DFE in terms of bit error probability is simulated for the same channel. The feedforward filter is 200 taps while 160 taps are used in the feedback path. 4 bits/sym is used for this simulation, i.e., PAM-16. The MMSE Tomlinson-Harashima precoding with same parameters is also implemented for the sake of comparison. The results of bit error probabilities versus bit energies are presented in Fig. 4.12.

![Figure 4.11: Initial convergence of adaptive decision-feedback equalizer.](image)
These results illustrate the superiority of the Tomlinson-Harashima precoding over the decision-feedback equalizer. It is apparent that this gain can be explained by the effect of decision errors in DFE performance.

Figure 4.12: Performance of decision-feedback equalizer and Tomlinson-Harashima precoding.
Chapter 5

Interference Cancellation

5.1 Introduction

Reliable transmission of data across a four-pair cable channel is mainly degraded by insertion loss, return loss (echo), near-end crosstalk (NEXT) and far-end crosstalk (FEXT). For shorter cables, far-end echo might be added to this list. FEXT describes the coupled signals that originate from the opposite end of the affected receiver, while NEXT describes the coupled signals that originate from the same end as the affected receiver. Full-duplex transmission schemes have the advantage that both transmit and receive signals occupy the low frequency portion of the channel, thereby minimizing channel attenuation. However, the disadvantage of full-duplex schemes is that the transmit and receive spectra overlap, causing interference between transmit and receive signals. This interference is known as echo. This is best understood by Fig. 5.1 where two twisted pairs and their corresponding interference terms are depicted.

To remove the effects of Echo, NEXT and FEXT interference signals, we assumed that the local transmitter has access to data of far-end and near-end transmitters. Therefore, symbol-interval data-driven adaptive filters can be applied to remove these interference terms. This assumption implies that the local receiver is synchronized with the far-end and near-end transmitters. One should note that this assumption may not be satisfied. For example in echo canceller, the local receiver's clock is locked to the data stream coming from distant transmitter. If the echo canceller, locked to the local transmitter's clock, can produce outputs at Nyquist intervals, an interpolator can drive samples at the sampling times, locked to the distant transmitter, demanded by the local receiver. A line-signal-driven canceller satisfies this
requirement, but at a formidable cost in complexity. Therefore, in designing echo, NEXT and FEXT we assumed that the local receiver is synchronized to the local and distant transmitters.

Furthermore, we should note that the near-end and far-end crosstalks are not random noises uncorrelated with transmitted signal, and it is reasonable to assume, however, the crosstalk is a linear effect. Therefore, it can be modeled as a linear time-invariant filter (possibly slowly varying). Since the measurements are affected by random and systematic errors specially for FEXT and NEXT, and also they could change at different configurations and installations, the best strategy for designing cancellers is to consider the worst case situation, meaning that the corresponding impulse responses are upper bounded.

### 5.2 Echo Cancellation

Here, we only consider the crosstalk signals caused by local transmitters (Fig. 5.2). To remove the echo interference, the *data-driven* echo canceller is considered where driving signal is the data stream at the input to the local transmitter rather than the line signal at its output. The *symbol-interval data-driven* model is used in the following analysis.
The echo impulse response is denoted by $p$, which consists of $v_p + 1$ taps. The near-end (echo) data sequence $\{z_k\}$ is real, zero-mean, and has autocorrelation matrix $R_{zz}$. The noise sequence $\nu_k$ is assumed real, zero-mean, independent of the input and echo sequences, and has an autocorrelation matrix $R_{\nu\nu}$. At time $kT$, we define the tap-weight and the tap-voltage vectors as:

$$
\hat{p} = [\hat{p}_0, \cdots, \hat{p}_{\nu_p}]^t \\
z_{k,k-\nu} = [z_k, \cdots, z_{k-\nu}]^t
$$

The constant echo channel impulse response is

$$
p = [p(\tau), p(T + \tau), \cdots, p(v_p T + \tau)]^t
$$

where $\tau$ is the sampling epoch chosen by the receiver. Note that, the transmitter pulse shaping is included in $p(t)$. The sample at time $kT$ of the received signal is

$$
r_k = z^t p + \nu_k
$$

where $\nu_k$ is the desired far-end signal plus the Gaussian noise at time $kT$. The structure of echo canceller is shown in Fig. 5.3, where the impulse response of the canceller that creates a replica of echo signal is $\hat{p}$. The optimum tap-weight vector $\hat{p}_{opt}$ minimizes the mean-square error $J = \mathbb{E}\{e_k^2\}$. We define the error signal at time $kT$ as:
The mean square error is then summarized to,

\[ e_k = r_k - \hat{r}_k = \sum_{l=0}^{v_p} P_l z_{k-l} + \nu_k - \sum_{l=0}^{\hat{v}_p} \hat{P}_l \hat{z}_{k-l} \]

\[ = \sum_{l=0}^{\hat{v}_p} (P_l - \hat{P}_l) z_{k-l} + \nu_k + \sum_{l=\hat{v}_p+1}^{v_p} P_l z_{k-l} \]

\[ = (p_{\hat{v}_p} - \hat{p})^{\dagger} z_{k-\hat{v}_p} + u_k \tag{5.4} \]

Thus, it only remains to obtain an estimate of \( \hat{p}_{\hat{v}_p} \) from the noisy received samples. The well known least-squares or least mean-squares (LMS) estimations can obtain this estimate during a training sequence transmission. The solution to the LS estimate is \([58][59]\)

\[ p_{lS} = (Z^{\dagger} Z)^{-1} Z^{\dagger} r_{\hat{v}_p + LN} \tag{5.7} \]
where the Toeplitz matrix $Z$ of dimension $(N - \hat{\nu}_p)(N - \hat{\nu}_p + 1)$ contains the last $N - \hat{\nu}_p$ symbols of a block of $N$ transmitted training symbols. If the additive Gaussian noise is not white, the LS estimate becomes:

$$ p_{LS} = (Z' R_{\nu\nu}^{-1} Z)^{-1} R_{\nu\nu}^{-1} Z' r_{\nu + LN} $$  \hspace{1cm} (5.8) $$

The echo return-loss enhancement is defined as:

$$ \text{ERLE} = 10 \log \frac{\mathbb{E}\{e_r^2\}}{\mathbb{E}\{e_k^2\}} $$  \hspace{1cm} (5.9) $$

The echo impulse response of a typical CAT-7A cable (50m) is shown in Fig. 5.4. For this impulse response, the ERLE versus $\hat{\nu}_p$ is calculated for 40GBASE-T application over 50m, and illustrated in Fig. 5.5. Even long estimators with 1000 taps can not satisfy the cancellation level requirement of 40GBASE-T, which is estimated to be about 60-65dB [30].

![Figure 5.4](image.png)

Figure 5.4: The echo impulse response of 50m CAT-7A with 1550MHz bandwidth sampled at Nyquist rate.

A veracious inspection of the echo impulse response reveals that most of the energy is concentrated at the beginning of the impulse response and the rest is distributed over a large time span. However, most of these taps contain low energy and can be considered zero as long as their total effect is below some margin level. Therefore, a zero-tap detection process can find the
negligible taps and omits the corresponding multipliers in convolution operation. This reduces the implementation complexity of the corresponding circuit in terms of gate counts, and consequently consumes less power. The zero-tap detection process is applied to a 1000-tap echo canceller obtained by Eq. (5.7) at different threshold levels. Fig. 5.5 illustrates the resulting ERLE versus sparsity of the corresponding FIR filter, i.e. the number of zero-tap coefficients. This figure reveals that a 6-dB back-off from optimal point can reduce the complexity by about 20% (200 taps out of 1000).

The process of zero-tap detection can be modified by a more rigorous and accurate method stated as follows. One can obtain the least-square estimate (5.7) by solving the following \(\ell_2\)-norm optimization [13]:

\[
\minimize \|Zp - r_{\hat{v}_{p+L}}\|_2^2
\]  

(5.10)

We define the class of \(k\)-sparsity containing vectors \(x \in \mathbb{R}^n\) with at least \(k\) zero coordinates. One can find a sparse solution of the minimization problem Eq. (5.10) by imposing the \(k\)-sparsity constraint, i.e.,

\[
\minimize \|Zp - r_{\hat{v}_{p+1}}\|_2^2 \\
\text{subject to} \quad p_j = 0, \quad j = j_1, j_2, \ldots, j_k, \quad 1 \leq j_i \leq n
\]  

(5.11)
Unfortunately, we can not predetermine what sparsity pattern, i.e. the subset \( \{j_1, j_2, \ldots, j_k\} \), gives the best estimation. An exhaustive search towards the optimum sparsity patterns requires examining \( n!/k!(n-k)! \) different combinations, which is quite impractical even for moderate size filter lengths. A heuristic approach for solving this problem is discussed in [13].

![Graph](image.png)

Figure 5.6: ERLE versus sparsity of corresponding FIR echo canceller. The canceller obtained by least-square estimation has a total of 1000 taps.

### 5.3 Channel Shortening Filter Design

One solution to the long echo canceller is the shortening impulse response (SIR) technique as proposed for xDSL applications. Channel shortening can be investigated with various objectives in mind; depending on the criterion adopted different methods for shortening the channel will be employed. Minimum mean square error (MMSE), maximum shortening signal-to-noise-ratio (MSSNR), minimum ISI, maximum bit rate, and others are some of the criteria extensively reviewed in the literature [7, 21, 31].

The most common approach in designing the shortening impulse response filter is the MMSE shortening which was first proposed by Falconer and Magee [31] in the context of
maximum likelihood receiver design. They succeeded in substantially decreasing the computational complexity of the Viterbi algorithm by shortening the channel impulse response, hence reducing the system memory. Later on, the idea was applied to multicarrier modulation, essentially to reduce the overhead of added cyclic prefix for long channel responses. Melsa et al [74,7675] tackled the problem by maximizing the shortening SNR and extended the idea to jointly equalize the channel and shorten the echo impulse response.

The concept of shortening impulse response is illustrated in Fig. 5.7. The function of the linear equalizer $w_k$ is to make the equalizer output appear like the output of the second filter, which is the convolution of the desired channel shaping function $b_k$ and the known channel input sequence $x_k$. The target channel $b_k$ has a length of $v_b + 1$ samples that is usually much less than the length of the original channel impulse response $h_k$. The optimum filters $w_k$ and $b_k$ are determined by maximizing the equalizer performance. The minimum mean-square error between the equalizer output and the desired channel shaping is a good measure of equalizer performance.

![Figure 5.7: Shortening impulse response block diagram.](image)

The error

$$e_k = z_k - \hat{z}_k$$

$$= w' y_{k-k-v_b} - b' x_{k-k-v_b-v_\Delta}$$

(5.12)
The solution to the minimum mean square error problem can be summarized as:

\[
\text{MMSE} = \min_{b, \Delta} b' R_{ee}(\Delta) b
\]  

(5.13)

where \( \Delta \) is the delay parameter and \( R_{ee}(\Delta) \) is the autocorrelation matrix of error signal \( e_k \). The solution of this problem is easily shown to be the eigenvector of \( R_{ee}(\Delta) \) that corresponds to the minimum eigenvalue. Thus:

\[
b_{\text{opt}} = cv_{\text{min}}(\Delta)
\]  

(5.14)

where \( v_{\text{min}}(\Delta) \) is the eigenvector corresponding to the minimum eigenvalue of \( R_{ee}(\Delta) \). In echo impulse response shortening case, the filter \( w_k \) affects both direct channel and echo channel. Thus, the MMSE optimization has to be done to jointly equalize the channel and shorten the echo impulse response.

### 5.3.1 Joint Channel Equalization and Echo Impulse Response Shortening

We consider the joint MMSE channel and echo impulse response shortening scenario depicted in Fig. 5.8. We follow the notation of Al-Dhahir in [1]. The problem and formulation presented in [1] are briefly stated as follows. The channel impulse response is denoted by \( h \), which consists of \( v_h + 1 \) taps. The echo impulse response is as what was described earlier in section 5.2.
Our objective is to shorten both channel and echo impulse responses, through a linear equalization, to the FIR filters $b_k$ and $c_k$, which consist of $v_b + 1$ and $v_c + 1$ taps, respectively. Also, we assume $v_h \gg v_b$ and $v_p \gg v_c$, otherwise the shortening does not make any sense. The far-end sequence data $\{x_k\}$ is also assumed real, zero-mean, with autocorrelation matrix $R_{xx}$.

The input-output relationship is given by:

$$y_{k+N-1:k} = H x_{k+N-1:k} + P z_{k+N-1:k} + v_{k+N-1:k}$$

where $H$ and $P$ are the Toeplitz convolution matrices. The error sequence subject to mean square minimization is:

$$e_k = \tilde{b}^t x_{k+N-1:k} + \tilde{c}^t z_{k+N-1:k} - w^t y_{k+N-1:k}$$

$$= \begin{bmatrix} \tilde{b}^t \\ \tilde{c}^t \end{bmatrix} \begin{bmatrix} x_{k+N-1:k} \\ z_{k+N-1:k} \end{bmatrix} - w^t y_{k+N-1:k}$$

$\tilde{b}$ is a concatenation of $\Delta_b$ leading zeros with $b$, followed by $s_1$ tail zeros, where $s_1 = N + v_h - v_b - \Delta_b$. $\tilde{c}$ is defined in a same way, i.e. concatenation of $\Delta_c$ leading zeros with $c$,
followed by $s_2 = N + v_p - v_c - \Delta_c$ tail zeros. The optimal shortening filter is derived by applying the orthogonality principle, i.e. $E\{e_k y_{k+N-1}^t \} = 0$. This results in an optimum shortening filter as:

$$w_{opt}^t = [b_{opt}^t, c_{opt}^t] \begin{bmatrix} R_{xy} & R_{zy} \end{bmatrix} R_{yy}^{-1}$$  \hspace{1cm} (5.17)

The mean square error (MSE) is given by

$$\text{MSE} = E\{e_k^2\} = \begin{bmatrix} b^t & c^t \end{bmatrix} R \begin{bmatrix} b \r{c} g Rg$$  \hspace{1cm} (5.18)

where

$$R = q' \begin{bmatrix} R_{xy} & -R_{xy} R_{yy}^{-1} R_{yz} \\ -R_{yx} R_{yy}^{-1} R_{yz} & R_{yx} \end{bmatrix} \simeq q' S q$$  \hspace{1cm} (5.19)

and

$$q' = \begin{bmatrix} 0_{(v_0 + v_1 + \Delta_c) \times \Delta_c} & B' 0_{(v_0 + v_1 + 2 \times (v_1 + \Delta_c)} & C' 0_{(v_0 + v_1 + 2) \times \Delta_c} \end{bmatrix}$$  \hspace{1cm} (5.20)

The optimum solution to this quadratic optimization using brute-force search is extensively formulated in [1]. To avoid the trivial all-zero solution, the unit-tap constraint was suggested and imposed on this optimization, i.e., one of the coefficient taps of $b$ and $c$ is set to one.

Although this calculation can be done once, it requires large matrix multiplications and inversions. The sparse structure of the matrix $q$ suggests further investigations to reduce possible redundancies. If we rewrite the matrix $S$ as the following block matrix:
The sizes of sub-matrices $S_{ij}$ are as follows:

- $S_{1j} : \Delta_b \times \Delta_b$
- $S_{2j} : (v_b + 1) \times (v_b + 1)$
- $S_{3j} : (s_1 + \Delta_c) \times (s_1 + \Delta_c)$, $1 \leq j \leq 5$
- $S_{4j} : (v_c + 1) \times (v_c + 1)$
- $S_{5j} : s_2 \times s_2$

Then, after simple block-matrix multiplications, the matrix $q' R_q$ is reduced to:

$$q'Sq = \begin{bmatrix} B' & C' \\ S_{22} & S_{24} \\ S_{42} & S_{44} \end{bmatrix} \begin{bmatrix} B \\ C \end{bmatrix} \begin{bmatrix} B' \\ C' \end{bmatrix} = A \begin{bmatrix} B \\ C \end{bmatrix}$$  \hspace{1cm} (5.23)

This can be reduced further to $A$ by noticing that:

$$\begin{bmatrix} B' \\ C' \end{bmatrix} = I_{(v_b + v_c + 2)}$$  \hspace{1cm} (5.24)

Therefore, the MSE is reduced to $g' A g$ where the moderate size matrix $A$ will greatly reduce the computational complexities of the original minimization problem. However, it is known and reported several times in the literature that the unit-norm constraint can achieve a better performance than unit-tap constraint in MMSE shortening filter design \[6\]. Here, we impose the unit-norm constraint on target channel impulse response $b$, i.e.,

$$\begin{align*}
\text{minimize} & \quad J_1(b, c) = \begin{bmatrix} b' & c' \end{bmatrix} A \begin{bmatrix} b \\ c \end{bmatrix} \\
\text{subject to} & \quad \|b\|_2 = 1, \quad c_j = 1, \quad \text{for } \exists \ 1 \leq j \leq m
\end{align*}$$  \hspace{1cm} (5.25)
We should emphasize here that the matrix $A$ is not necessarily symmetric, but we conjecture that $A$ is positive semidefinite (PSD) even though PSD is commonly defined for symmetric matrices \cite{47,52}. One can simply resolve this issue by replacing matrix $A$ with $0.5(A + A^t)$, which is obviously symmetric. Therefore, hereon, we shall assume $A$ is a PSD matrix.

If we expand the MSE in terms of properly sized block matrices $A_{11}$, $A_{12}$, and $A_{22}$,

$$J_1(b, c) = b^tA_{11}b + 2b^tA_{12}c + c^tA_{22}c$$

(5.26)

This problem can be solved iteratively by solving it first with respect to $c$ assuming $b$ is fixed. The solution to this unit-tap minimization problem, $c^*(j)$, is presented in Appendix A for a general quadratic programming (QP). The minimizer of this quadratic form is then replaced in our original objective function $J_1$.

$$J_2(b) = J_1(b, c^*(j)) = b^t(A_{11} - A_{12}A_{22}^{-1}A_{12}^t)b + (1 + b^tA_{12}A_{22}^{-1}(\cdot, j))^2$$

(5.27)

The resulting objective function is another quadratic form that can be minimized with the unit-norm constraint. The solution to this problem is discussed in Appendix B. This process can be iterated with respect to index $j$ to find out the global minimizers $b^*$ and $c^*$. This is summarized in the Algorithm 5.1.
Algorithm 5.1: **Unit-norm and unit-tap joint optimization**

\[ J_{\text{min}} = +\infty \]

for \( j = 1 : m \)

\[ A = A_{11} - A_{12} A_{22}^{-1} A_{12}^\top + \cdots \]

\[ A_{12} A_{22}^{-1}(:, j) [A_{22}^{-1}(:, j)]^\top A_{12}^\top / A_{22}^{-1}(j, j) \]

\[ b = [A_{22}^{-1}(:, j)]^\top A_{12} / A_{22}^{-1}(j, j) \]

minimize \( \mathbf{x}' \mathbf{A} \mathbf{x} + 2 b' \mathbf{x} + 1 / A_{22}^{-1}(j, j) \)

subject to \( \| \mathbf{x} \|_2 = 1 \)

\( x(j) = x^* \)

\( y(j) = A_{22}^{-1}[(1 + x'^* A_{12} A_{22}^{-1}(:, j) e_j / A_{22}^{-1}(j, j) - A_{12}^\top x^*)) \]

\( J_j = J(\mathbf{x}(j), \mathbf{y}(j)) \)

if \( J_j \leq J_{\text{min}} \)

\( \mathbf{x}_{\text{opt}} = \mathbf{x}(j) \)

\( \mathbf{y}_{\text{opt}} = \mathbf{y}(j) \)

\( J_{\text{min}} = J_j \)

end

end

We now show that the norm-constraint on \( b \) is sufficient to avoid trivial solutions. In fact, having only norm-constraint on target channel impulse response, automatically results in a non-zero solution for target echo impulse response. The optimization problem can be stated as:

\[
\begin{align*}
\text{minimize} & \quad J_3(b) = \begin{bmatrix} b' \ c' \end{bmatrix} A \begin{bmatrix} b \\ c \end{bmatrix} \\
\text{subject to} & \quad \|b\|_2 = 1
\end{align*}
\]

(5.28)

**Lemma:** The global minimizer of \( J_3(b) \) results in a non-zero answer for column vector \( c \).

**Proof:** The second-order cone \( \Omega = \{ c \mid 2 b' A_{12} c + c' A_{22} c \leq 0 \} \) [13] is a non-empty subspace, therefore \( J_1(b, c = 0) = b' A_{11} b \) is bigger than \( J_1(b, c) \) for some non-zero \( c \). Therefore, if \( b^* \) is the global minimizer of \( J_3(b) \), then \( J_3(b^*) \leq J_1(b, c = 0) \). This concludes that a single unit-norm constraint is sufficient to avoid a non-zero solution for target echo impulse response. □
To solve the problem (5.28), we incorporate the unit-norm constraint into an auxiliary function:

\[ L(b, c, \lambda) = \begin{bmatrix} b' & c' \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} + \lambda \left(1 - b'b\right) \]  

(5.29)

Applying the optimality conditions yields:

\[ \frac{\partial L}{\partial b} = A_{11}b + A_{12}c - \lambda b = 0 \]
\[ \frac{\partial L}{\partial c} = A_{12}'b + A_{22}c = 0 \]

(5.30)

Given \( A_{22} \) is invertible, the second optimality condition gives:

\[ c = -A_{22}^{-1}A_{12}'b \]  

(5.31)

This equation explicitly indicates that, given \( b \) is non-zero, the solution for \( c \) is non-zero. Substituting \( c \) back into the first optimality condition results in the following equation:

\[ \left( A_{11} - A_{12}A_{22}^{-1}A_{12}' \right)b = \lambda b \]  

(5.32)

This is a standard eigenvalue-eigenvector problem; however, it needs some extra normalization and manipulations to obtain the minimizer of our quadratic optimization. The details are presented in the following algorithm. We incorporated the MATLAB notations for the sake of implementation ease.
Algorithm 5.2: **Unit-norm joint optimization**

\[
M = A_{11} - A_{12} A_{22}^{-1} A_{12}^t \\
[X \ D] = \text{eig}(M) \\
\text{for } k = 1: \text{size}(X, 2), \\
\quad X(:, k) = X(:, k)/\sqrt{X(:, k)}^2 \\
\text{end} \\
Y = -A_{22}^{-1} A_{12}^t X \\
Z = [X ; Y] \\
Z = [Z - Z] \\
J_{\text{min}} = +\infty \\
\text{for } k = 1: \text{size}(Z, 2) \\
\quad J(k) = Z(:, k)^t A Z(:, k) \\
\quad \text{if } J(k) \leq J_{\text{min}} \\
\quad \quad J_{\text{min}} = J(k) \\
\quad \quad b_{\text{opt}} = Z(1:m, k) \\
\quad \quad c_{\text{opt}} = Z(m+1:end, k) \\
\quad \quad k_{\text{min}} = k \\
\quad \text{end} \\
\text{end}
\]

### 5.3.2 MSSNR

Here, for the sake of completeness and comparison, we overview the maximum shortening SNR approach for the joint shortening filter design. We adopt our notation to the formulation derived in [75]. The equalized channel \( h_{eq} = H w \), can be partitioned as

\[
h_{\text{win}} = H(\Delta_b + 1:v_b + \Delta_b + 1, :) w \overset{\triangle}{=} H_{\text{win}} w 
\]

(5.33)

and

\[
h_{\text{wall}} = H([1: \Delta_b, \Delta_b + v_b + 1], :) w \overset{\triangle}{=} H_{\text{wall}} w 
\]

(5.34)

where
Hence, \( h_{\text{win}} \) denotes the part of the equalized channel inside the shortening window, and
\( h_{\text{wall}} \) represents the part of the equalized channel outside the shortening window. \( p_{\text{win}} \) and \( p_{\text{wall}} \) are defined similarly.

Here, instead of minimizing the mean square error, the optimization criterion is to maximize the ratio of signal energy inside the shortening window to the signal energy spreaded outside the shortening window. In fact, the equalizer performance depends on how well it has concentrated the signal into a target window. The corresponding energies can be found from

\[
\begin{align*}
\mathcal{E}_{\text{win}} &= \beta h_{\text{win}}^t h_{\text{win}} + (1-\beta)p_{\text{win}}^t p_{\text{win}} = w^t B(\beta) w \\
\mathcal{E}_{\text{wall}} &= \alpha h_{\text{wall}}^t h_{\text{wall}} + (1-\alpha)p_{\text{wall}}^t p_{\text{wall}} + \sigma_u^2 = w^t A(\alpha) w
\end{align*}
\]

where \( \sigma_u^2 \) is the power of the filtered noise \( u_t = w^t n_{k+N-Lk} \) and

\[
\begin{align*}
B(\beta) &= \beta H_{\text{win}}^t H_{\text{win}} + (1-\beta)P_{\text{win}}^t P_{\text{win}} \\
A(\alpha) &= \alpha H_{\text{wall}}^t H_{\text{wall}} + (1-\alpha)P_{\text{wall}}^t P_{\text{wall}} + R_{vw}
\end{align*}
\]
The parameters $\alpha$ and $\beta$ balance the energy relationship between the energy in and out of the window for both the channel and echo. The shortening SNR is defined as:

$$\text{SSNR} = \frac{\mathcal{E}_{\text{win}}}{\mathcal{E}_{\text{wall}}} = \frac{w^T B(\beta) w}{w^T A(\alpha) w} \quad (5.38)$$

The SSNR can be maximized either by minimizing $w^T A w$ subject to unit-energy constraint, i.e., $w^T B w = 1$, or by maximizing $w^T B w$, subject to $w^T A w = 1$. If $B$ is not singular, the solution for the first optimization problem is

$$w_{\text{opt}} = \left( \sqrt{B} \right)^{-1} e_{\text{min}} \quad (5.39)$$

where $\sqrt{B}$ is the Cholesky decomposition of $B$, and $e_{\text{min}}$ is the eigenvector corresponding to the minimum eigenvalue of matrix $C = \left( \sqrt{B} \right)^{-1} A \left( \sqrt{B} \right)^T$. The maximum shortening signal-to-noise-ratio (MSSNR) is:

$$\text{MSSNR} = \text{SNR}_{\text{max}} = -10 \log_{10} \left( \lambda_{\text{min}} (C) \right) \quad (5.40)$$

This analysis is based on non-singularity assumption of matrix $B$. Note that $B$ is singular if the channel shortening filter length is larger than the length of the shortening window, i.e., $v_w > v_b$, which is the case, generally. Some details about this minimization problem and the case where $B$ is singular is provided in Appendix D.

Although there is no clear connection between MMSE and MSSNR impulse response shortening filters, it is also worth to mention that in the noiseless scenario and when data is white stationary random variable, both MMSE and MSSNR approaches are equivalent [27].
5.4 Block-Convolution Filter

In this section the usual convolution and recursion that implement FIR discrete-time filters are reformulated in terms of vectors and matrices. We show that the implementation complexity of long FIR filters can be hugely reduced by employing a very efficient Fast Fourier Transform (FFT). The idea is to group the input and output sequences and perform the operation on a group of samples instead of running the whole convolution operation per each input sample. The idea goes back to the original works of Burrus [14][15], Mitra and Clark [24][73], trying to enhance the speed and reduce the complexity of conventional FIR filters.

Suppose $h$ is a causal FIR filter and of length $v_h+1$. In direct implementation, to obtain each output sample at time instance $k$, the convolution is carried out as:

$$y_k = \sum_{m=0}^{k} h_{k-m} x_m$$  \hspace{1cm} \text{(5.41)}

This convolution can be expressed as a matrix operation by:

$$
\begin{bmatrix}
y_0 \\
y_1 \\
y_2 \\
\vdots
\end{bmatrix} =
\begin{bmatrix}
h_0 & 0 & 0 & \cdots & 0 \\
h_1 & h_0 & 0 & \cdots & 0 \\
h_2 & h_1 & h_0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
x_2 \\
\vdots
\end{bmatrix}
$$

\hspace{1cm} \text{(5.42)}

Now we group the input and output sequences into vectors of length $L$ and rewrite the original convolution in terms of these vectors. In vector form, Eq. (5.42) can be expressed in term of block matrices $H_0, H_1, H_2, \ldots$ as:

$$
\begin{bmatrix}
y_0 \\
y_1 \\
y_2 \\
\vdots
\end{bmatrix} =
\begin{bmatrix}
H_0 & 0 & 0 & \cdots & 0 \\
H_1 & H_0 & 0 & \cdots & 0 \\
H_2 & H_1 & H_0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
x_2 \\
\vdots
\end{bmatrix}
$$

\hspace{1cm} \text{(5.43)}

where
This is, indeed, a vector-convolution represented by:

\[
\begin{align*}
x_k &= [x_{kL} \ x_{kL+1} \ \cdots \ x_{(k+1)L-1}]' \\
y_k &= [y_{kL} \ y_{kL+1} \ \cdots \ y_{(k+1)L-1}]'
\end{align*}
\] (5.44)

where one can express it compactly by z-transforms as:

\[
y_k = \sum_{m=0}^{k} H_{k-m} x_m
\] (5.45)

It can be shown that the polynomial matrix \( H(z) \) can be decomposed into two matrices as

\[
H(z) = P(z)Q(z) \quad [14][15],
\]

where \( P(z) \) is given by

\[
P(z) = \begin{bmatrix} p_0(z) & p_0(z) & \cdots & p(z) & 0 & 0 & \cdots & 0 \\
p_0(z) & p_1(z) & \cdots & p(z) & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & p_0(z) & p_1(z) & p(z) & 0 \\
0 & 0 & \cdots & 0 & p_0(z) & p_1(z) & \cdots & p(z) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & p_1(z) \\
p_1(z) & p_2(z) & 0 & 0 & 0 & \cdots & 0 & p_0(z) \\
\end{bmatrix}
\] (5.47)

and \( Q(z) = [I_L \ z^{-1}I_L] \). \( P(z) \) is the product of \([I_L \ 0_{L,L}]\) with a circulant matrix \( C(z) \). The eigenvalues of a circulant matrix of a given size are the column of the discrete Fourier transform matrix of the same size defined by \( F_N = [f_{mn}] \), \( f_{mn} = e^{-2\pi in/m} \). Thus the matrix \( F_N \) diagonalizes \( C(z) \), i.e., \( F_N C(z) F_N^* = \text{diag} \{ F_N e \} \) [47]. Therefore, \( C(z) = F_N^* D(z) F_N \), where \( D(z) \) is \( 2L \times 2L \) diagonal matrix function with entries \( d_0(z), \ldots, d_{2L-1}(z) \). Each of these entries \( d_k(z) \) is an FIR filter with \((v_h + 1)/L\) taps. The diagram of block-convolution implementation of this FIR filter is shown in Fig. 5.9. More details of this implementation can be found in [51].
The complexity of this system can be represented by the number of multiplications in the algorithm. There are:

- \( L \log_2(2L) \) multiplications for each FFT or IFFT operations.
- \((v_h + 1)/L\) multiplication for each of 2L filters.

They totally add up to \( 2L \log_2(2L) + 2(v_h + 1) \) multiplications to compute L output samples. Besides, the filters \( d_k(z) \) themselves need to be determined form \( H(z) \). Nevertheless, this calculation is done once unless there is an adaptation in the system which has to be carried on for every block of L samples. This step also amounts to a cost of \((v_h + 1) \log_2(2L)\) multipliers.

Therefore the total complexity of block-convolution implementation of FIR filter \( h \) reads to:

\[
2(v_h + 1) + (v_h + 1 + 2L) \log_2(2L)
\]  \hspace{1cm} (5.48)

The complexity ratio between direct and block implementations of an FIR filter with 1024 taps is shown in Fig. 5.10 where the block length is varied from 4 to 1024. It is noticed that, for this example, the complexity is substantially reduced for block lengths larger than 8 samples. Although increasing the block length reduces this ratio, one should note that higher block lengths introduce larger delays in the data path that may not be tolerable in some applications.
In Category 7A cables, the near-end interference (NEXT) is caused by impedance mismatching and the cable connector, and it is quite independent of cable length. Fortunately, as long as the cable is set up, its transfer function for NEXT is completely static over a long period and the requisite adaptation can be removed or carried out in a fairy long time period. Therefore, the only remaining task is to estimate the NEXT transfer function during the start-up process and employ it to mitigate the near-end crosstalks.

5.5 NEXT Cancellation

In Category 7A cables, the near-end interference (NEXT) is caused by impedance mismatching and the cable connector, and it is quite independent of cable length. Fortunately, as long as the cable is set up, its transfer function for NEXT is completely static over a long period and the requisite adaptation can be removed or carried out in a fairy long time period. Therefore, the only remaining task is to estimate the NEXT transfer function during the start-up process and employ it to mitigate the near-end crosstalks.

Figure 5.10: Complexity ratio between direct and block implementations of a 1024-tap FIR filter at different block lengths.

Figure 5.11: Crosstalk form the local transmitter (near-end or NEXT).
We assume that the local receiver is synchronized to adjacent and distant transmitters and has access to their data (training mode). This synchronous conjecture is satisfied for near-end transmitters as they are usually derived by a master clock.

The classical approach to cancel near-end crosstalk is depicted on the left side of Fig. 5.12. The NEXT cancellation can be done simply by subtracting the output of crosstalk canceller filters from the sampled received signal. Rather than this straightforward approach, we follow the MIMO crosstalk cancellation that can incorporate joint processing and reduce any kind of redundancy.

Consider a MIMO system with 4 inputs and 4 outputs. We also assume that the MIMO system is linear and time-invariant. NEXT cancellation consists of identifying 4 echo paths at each local transmitter so that in total, 16 echo paths need to be estimated. The output signal is obtained by

\[ y_k = \sum_{\ell=0}^{\ell_{\max}} G_{\ell} z_{k-\ell} + \nu_k = G_{0_{\max}} z_{k_{\max}} + \nu_k \]  

(5.49)
where we introduced the matrix $G_{0:v_G}$ as (with some abuse of notations$^2$):

$$G_{0:v_G} = [G_0, G_1, \ldots, G_{v_G}]$$  \hfill (5.50)

We define the error signal at time $k$ as:

$$e_k = y_k - \hat{y}_k$$

$$= \sum_{l=0}^{v_G} G_l z_{k-l} + \nu_k - \sum_{l=0}^{v_G} \hat{G}_l z_{k-l}$$

$$= (G_{0:v'_G} - \hat{G}_{0:v'_G})^t z_{k-k_{v'_G}} + u_k,$$  \hfill (5.51)

where $u_k = \nu_k + G_{v'_G + v_G} z_{k-v'_G - v_G}$.

Having introduced the error signal, we now define the least-squares error criterion with respect to the modeling filter:

$$J = tr \left[ E \left\{ e_k e_k^t \right\} \right]$$

$$= tr \left[ (G_{0:v_G} - \hat{G}_{0:v_G})^t R_{zz} (G_{0:v'_G} - \hat{G}_{0:v'_G}) + R_{uu} \right]$$  \hfill (5.52)

This is a quadratic form in the tap weight matrix $\hat{G}_{0:v'_G}$, and hence there is a unique minimum. The minimization of Eq. (5.52) leads to the unique minimizer:

$$\hat{G}_{0:v'_G} = G_{0:v'_G}$$  \hfill (5.53)

The preceding equations have expressed $y_k$ as a linear combination of the transmitted symbols $z_k$. We introduce the vectorized notation $g_k = vec(G_k)$ for the channel matrix tap $G_k$.

Alternatively, we can write the convolution operation given by Eq. (5.49) as a linear operation on the channel coefficient vector $g_{0:v_G} = [g_0^t, \ldots, g_{v_G}^t]^t$, which gives

$\hat{G}_{0:v'_G}$

$\hat{G}_{0:v'_G}$

\[\hat{G}_{0:v'_G} = G_{0:v'_G}\]

Throughout this dissertation, we generally use $x_{l:N}$ notation to express a vector of signal samples, unless otherwise indicated in the text. In the case of matrices, $G_{0:v_G}$ represents the horizontal concatenation of a sequence of matrices $G_0, \ldots, G_{v_G}$.
\[ y_{kk-N_i} = (Z_k \otimes I_4) g_{0v_{k_i}} + \nu_{kk-N_i} \]  \hspace{1cm} (5.54)

where \( \otimes \) is the Kronecker product and \( Z_k \) is the block Toeplitz matrix of transmit symbols.

The maximum likelihood (ML) is the standard approach for this channel estimation, which has the advantage that the knowledge of the noise variance is not required. The ML channel estimate related to Eq. (5.54) is obtained by solving the following optimization problem [40]:

\[
g_{0v_{k_i}}^{\text{ML}} = \arg \min_{g_{0v_{k_i}}} \sum_{\ell=0}^{N_i} \left\| g_{k-k-\ell} - (Z_{k-k-\ell} \otimes I_4) g_{0v_{k_i}} \right\|^2
\]

\[
= \arg \min_{g_{0v_{k_i}}} \left\| g_{kk-N_i} - (Z_{kk-N_i} \otimes I_4) g_{0v_{k_i}} \right\|^2
\]

This is a standard least-squares (LS) problem, whose solution is given in terms of a pseudo-inverse \( \left(Z_{kk-N_i} \otimes I_4\right)^\dagger \) which after some manipulation reduces to \( Z_{kk-N_i}^\dagger \otimes I_4 \). Assuming \( Z_{kk-N_i}^\dagger \) has full column-rank, we obtain:

\[
g_{0v_{k_i}}^{\text{ML}} = \left(Z_{kk-N_i}^\dagger \otimes I_4\right) y_{kk-N_i}
\]

\[
= \left(\left(Z_{kk-N_k}^\dagger Z_{kk-N_k}\right)^{-1} Z_{kk-N_k}^\dagger \right) y_{kk-N_i}
\]

\[
= \left(Z_{kk-N_k}^\dagger Z_{kk-N_k}\right)^{-1} \otimes I_4 \left(\sum_{\ell=0}^{N_i} Z_{k-k-\ell}^\dagger \otimes I_4\right) y_{k-k-\ell}
\]

The covariance of the channel estimation error is given by:

\[
E\left\{ (g_{0v_{k_i}}^{\text{ML}} - g_{0v_{k_i}})(g_{0v_{k_i}}^{\text{ML}} - g_{0v_{k_i}})^\dagger \right\} = \sigma^2 \left( \sum_{\ell=0}^{N_i} (Z_{k-k-\ell}^\dagger)^\otimes I \right)^{-1}
\]

\hspace{1cm} (5.57)
The mean square error (MSE) of the channel estimate is summarized to:

\[
J = \mathbb{E} \left[ \left\| \hat{g}_{\text{ML}} - g \right\|^2 \right] = 4\sigma^2 \text{tr} \left[ \left( Z_{V_{k} - N} \right)^{-1} \right]
\]  \hspace{1cm} (5.58)

The time-dispersion or frequency-selectivity of near-end channels can be estimated by Eq. (5.56) and proper crosstalk cancellers are designed accordingly. However, the computational complexity of this estimation becomes massively large even for moderate channel lengths. In the following, we present a less complex training based MIMO channel estimation.

In our developed MIMO channel model, between every transmit node \( m \) and every receive node \( n \), there is a single-input single-output (SISO) channel impulse response \( g^{n,m} \) of length \( v_G + 1 \), described by the vector (assuming the same order \( v_G \) for all channels):

\[
g^{n,m} = [g_{0}^{n,m}, g_{1}^{n,m}, \ldots, g_{v_G}^{n,m}]^t
\]  \hspace{1cm} (5.59)

This frequency selective MIMO channel can be described by \( v_G + 1 \) channel matrices:

\[
G_{k} = \begin{bmatrix}
g_{k}^{1,1} & g_{k}^{1,2} & g_{k}^{1,3} & g_{k}^{1,4} \\
g_{k}^{2,1} & g_{k}^{2,2} & g_{k}^{2,3} & g_{k}^{2,4} \\
g_{k}^{3,1} & g_{k}^{3,2} & g_{k}^{3,3} & g_{k}^{3,4} \\
g_{k}^{4,1} & g_{k}^{4,2} & g_{k}^{4,3} & g_{k}^{4,4}
\end{bmatrix}
\]  \hspace{1cm} (5.60)

To reduce the complexity of the estimation, the channel is modeled as a superposition of 4 multiple-input single-output (MISO) channels. Each MISO channel has \( 4 \times (v_G + 1) \) unknowns.

The received signal \( y_{k}^{n} \) of the \( n \)-th MISO channel can be expressed as:

\[
y_{k}^{n} = \sum_{m=1}^{4} \sum_{l=0}^{v_G} g_{l}^{n,m} z_{k-l}^{m} + \nu_{k} \quad (5.61)
\]

\(^{3}\) The pseudo-inverse \( A^{\dagger} \) of a matrix \( A \) is defined by four properties: (1) \( AA^{\dagger}A = A \), (2) \( A^{\dagger}AA^{\dagger} = A^{\dagger} \), (3) \( (AA^{\dagger})^{\dagger} = AA^{\dagger} \), (4) \( (A^{\dagger}A)^{\dagger} = A^{\dagger}A \).
The number of training symbols per transmit node and per frame is $N_{t} + 1$. We form the matrix $Z_{m}$ of dimension $(N_{T} - v_{G}') \times (v_{G}' + 1)$, which contains the transmitted training symbols of transmit node $m$, given by:

$$
Z_{m} = \begin{bmatrix}
    z_{v_{G}}^{m} & z_{v_{G} - 1}^{m} & \cdots & z_{0}^{m} \\
    z_{v_{G} + 1}^{m} & z_{v_{G}}^{m} & \cdots & z_{1}^{m} \\
    \vdots & \vdots & \ddots & \vdots \\
    z_{N_{T}}^{m} & z_{N_{T} - 1}^{m} & \cdots & z_{N_{T} - (v_{G} + 1)}^{m}
\end{bmatrix}
$$

(5.62)

The received vector $y_{0N_{t}}^{m} = [y_{0}^{m}, \ldots, y_{N_{T}}^{m}]^{T}$ of the receive node $n$, can be written as

$$
y_{0N_{t}}^{n} = Z_{1}g_{1}^{n,1} + Z_{2}g_{2}^{n,2} + Z_{3}g_{3}^{n,3} + Z_{4}g_{4}^{n,4} + \nu_{0N_{t}}^{n,4}
$$

(5.63)

with the additive noise vector $\nu_{0N_{t}}$. The matrix $Z_{14} = [Z_{1}, Z_{2}, Z_{3}, Z_{4}]$ of dimension $(N_{T} + 1) \times 4(v_{G} + 1)$ contains the transmitted training symbols of all transmit nodes. The matrix $Z_{14}$ and all submatrices $Z_{m}$ are cyclic. The vector $g_{n,14}^{n,14}$ of length $4(v_{G} + 1)$ contains the elements of the channel impulse responses $g_{k}^{n,m}, k = 0, \ldots, v_{G}, m = 1, \ldots, 4$ between all transmit nodes and the $n$-th receive node. The elements of the channel impulse responses $g_{k}^{n,m}, k = 0, \ldots, v_{G}, m = 1, \ldots, 4$ have to be estimated. The vector of received samples $y_{0N_{t}}^{m}$ reads to:

$$
y_{0N_{t}}^{m} = Z_{14}g_{n,14}^{n,14} + \nu_{0N_{t}}^{n,4}
$$

(5.64)

The linear least square estimate for $g_{n,14}^{n,14}$ is given by [58]:

$$
\hat{g}_{n,14}^{n,14} = \left[Z_{14}^{T}Z_{14}^{n,14}\right]^{-1}Z_{14}^{T}y_{0N_{t}}^{n,4}
$$

(5.65)

and if the additive noise is not white, this estimate becomes:
The mean-squared error (MSE) of this estimation is given by:

\[
J = E \left\| \hat{g}_{n,k}^{n,4} - g_{n,k}^{n,4} \right\|^2 = 4\sigma^2_n \text{tr} \left( Z_{k+4}^T Z_{k+4} \right)^{-1} \tag{5.67}
\]

It is not hard to show that the MSE is minimized if and only if \( Z_{k+4}^T Z_{k+4} \) is diagonal \([91]\).

To construct a training sequence resulting in the minimum MSE, we should get more insight about the structure of this matrix. The matrix \( Z_{k+4}^T Z_{k+4} \) is expanded to:

\[
Z_{k+4}^T Z_{k+4} = \begin{bmatrix} Z_1^T & Z_2^T & Z_3^T & Z_4^T \end{bmatrix} \begin{bmatrix} Z_1 & Z_2 & Z_3 & Z_4 \end{bmatrix} \tag{5.68}
\]

Therefore, to obtain a diagonal matrix, we should construct training sequences so that their corresponding Toeplitz forms are orthogonal (in a conventional matrix multiplication sense).

To do this, one can construct a sequence and then look for another sequence orthogonal to this one by an exhaustive search or though a systematic construction. This process is repeated each time to find another sequence orthogonal to other ones. Although this is a straight forward approach to compose orthogonal sequences and it might not be too hard to find short sequences, as the length of sequences (or the number of sequences) increases, finding an orthogonal set of sequences becomes extremely tedious and the necessity of a better solution is inevitable.

Here, we propose to use orthogonal space-time code which is generally defined as \( T \times N \) matrix \( R \) with real entries \( r_1, -r_1, r_2, -r_2, \ldots, r_K, -r_K \) such that

\[
R^T R = c \left( r_1^2 + r_2^2 + \cdots + r_K^2 \right) I_N \tag{5.69}
\]
where \( c \) is a constant [54]. The existence of this matrix is limited to \( N=2, 4 \) and 8; there is no orthogonal space-time code for other values of \( N \) [54]. Luckily, this perfectly fits to our application with \( N=4 \). The process of generating the orthogonal space-time code is as follows.

Four arbitrary sequences \( z^1, z^2, z^3, \) and \( z^4 \) are generated. A block code is designed as:

\[
\begin{pmatrix}
  z^1 & z^2 & z^3 & z^4 \\
  z^2 & -z^1 & z^4 & -z^3 \\
  z^3 & -z^4 & -z^1 & z^2 \\
  z^4 & z^3 & -z^2 & -z^1
\end{pmatrix}
\] (5.70)

The transmit nodes start to send this block of training symbols row by row. At the receiver, each receive node discards the first \( \nu^G + 1 \) received symbols corresponding to each row of training block, and the Toeplitz form of these four sequences is constructed. It is easy to verify that

\[
y^{\text{total}}_k = \begin{bmatrix}
  y^1_k \\
  y^2_k \\
  y^3_k \\
  y^4_k
\end{bmatrix} = \begin{bmatrix}
  Z_1 & Z_2 & Z_3 & Z_4 \\
  Z_2 & -Z_1 & Z_4 & -Z_3 \\
  Z_3 & -Z_4 & -Z_1 & Z_2 \\
  Z_4 & Z_3 & -Z_2 & -Z_1
\end{bmatrix} \begin{bmatrix}
  g^{1,14} \\
  g^{2,14} \\
  g^{3,14} \\
  g^{4,14}
\end{bmatrix} + \nu
\] (5.71)

where \( y^{\text{total}}_k \) is the congregated vector from received symbols at four independent transmissions, \( Z_m \) is the Toeplitz form of \( z^m \) and \( \nu \) is the total noise vector, i.e. \( \nu = [(\nu^1)^t, (\nu^2)^t, (\nu^3)^t, (\nu^4)^t]^t \).

Constructing the training sequences by this way guarantees zero off-diagonal elements in matrix \( Z^1_{14}Z^1_{14} \). It remains to find suitable sequences \( z^n \) that have an impulse-like autocorrelation sequence [16]. Our simulation indicates that using this space-time code structure without any further effort in searching for sequences that have impulse-like autocorrelation property will substantially reduce the estimation error. Besides, the process of training symbol transmission can be long enough to achieve a proper level of estimation accuracy.
5.6 FEXT Cancellation

The *far-end crosstalk* (FEXT) loss quantifies the undesired signal coupling between adjacent pairs at the far-end of cabling. FEXT is interference produced by signals emanating from adjacent transmitters at the far end of the 10 Gigabit Ethernet link (see Fig. 5.13).

![Figure 5.13: Crosstalk signal from far-end transmitter (FEXT).](image)

The requirement for FEXT cancellation in 40GBASE-T system directly relies on the background noise level assumption and achieving the maximum possible reach length. One can determine the requirement for FEXT cancellation from the power spectral density (PSD) of the interfering signal. For example, the FEXT crosstalks and the spectral density of only one intruder of a 50m CAT-7A cable is drawn in Fig. 5.14 and 5.15. It is noted that the interfering energy from this source is very small compared to the background noise level of -146dBm/Hz. If we change the background noise level to a smaller value, then the energy of the FEXT signal must be attenuated properly to bring it down to about the same level of background noise. However, with these assumptions, the capacity will be increased and the objective of transmitting 40Gbps over this cable is achieved with a better margin. This also means that we can use a longer cable if the designer wishes to extend the reach mode to a longer distance.
Figure 5.14: Far-end crosstalk (FEXT) of 50m CAT-7A cable.

Figure 5.15: Power spectrum density of far-end crosstalk signal in a 50m CAT-7a cable.

With the current design of CAT-7A cable, only a few dB of FEXT cancellation is necessary, or it can be left uncompensated if a low cost system is our high priority. However,
there are a few remarks regarding FEXT cancellation worth to mention here. Compared to challenges of echo this sounds easy. Nonetheless, FEXT has the unfortunate characteristics of not only being formed by signals which do not originate at the same end of the cable, but also non-causal arrival times (FEXT can arrive before the signal which generated it arrives on its wire) and wide variation in time of arrival. The wide variation in arrival time substantially increases the complexity of cancellation circuitry, and when the FEXT arrives before the generating signal, even advanced FEXT cancellation techniques based on tentative decisions will not achieve the required cancellation, requiring integration of FEXT cancellation with other systems elements. Having said that, for simulation purpose, we made synchronous assumption and by introducing a variable length buffer to compensate varying arrival time of FEXT pulse, an adaptive NLMS canceller is designed to remove FEXT signal.

5.7 Simulations and Results

In this section, we present the results of echo cancellation using shortening impulse response filter and block-convolution implementation. The channel and echo impulse responses of CAT-7A copper cable are used in our simulations.

First, we demonstrate the SIRF method by the following example. The channel impulse response of a 50m CAT-7A cable is show in Fig. 5.16. This impulse response has over 200 non-zero taps. A 55-tap shortening filter is designed to squeeze this impulse response into a 60-tap window after an optimum 19 delay samples. The resulting shortened impulse response is also shown in Fig. 5.16.
The echo (return loss) of a 50m in frequency domain is shown in Fig. 5.17. Although the channel impulse response is directly affected by the cable length, its effect on echo impulse response is very minor since echo is a measure of the signal reflections occurring along a transmission line and related to impedance mismatch in a cable channel. However, different cable lengths mean different optimum signal bandwidths, and this changes the time spreading of echo signal of 40GBASE-T application over different lengths. For simulation purposes, we use the characteristics of a 50m cable. The signal bandwidth is about 1550MHz and it is sampled at Nyquist rate. The noise power spectral density is assumed -146 dBm/Hz. Other far-end (FEXT) and near-end (NEXT) crosstalks are not considered here.

Figure 5.16: Original and shortened impulse response of a 50m CAT-7A cable.
For joint channel equalization and echo cancellation, a 250-taps symbol-spaced linear equalizer is assumed. The lengths of target channel and echo impulse responses are set to 80 and 650, respectively. The optimum delay parameters are $(\Delta_b, \Delta_c) = (25, 2)$. The minimum MSE after this joint shortening by a linear equalizer is $1.33 \times 10^{-6}$. The original and shortened channel and echo impulse responses are depicted in Fig. 5.18. Obviously, further reductions in $\text{MSE}_{\text{min}}$ can be achieved by increasing $N$, $v_b$, or $v_c$.

A comparison to the proposed method is made considering both MMSE unit-tap joint shortening and MSSNR joint shortening. The unit-tap MMSE joint shortening presented in [1] was performed with the same set of parameters. The minimum achievable MSE was $2.56 \times 10^{-5}$ with optimum delay parameters $(\Delta_b, \Delta_c) = (25, 4)$. As we can see, the proposed method outperforms the unit-norm joint shortening by 13dB. Both methods have estimated very similar values for the optimum delay parameters. The MSSNR joint-shortening is also carried out using
the same \((\Delta_b, \Delta_e)\) obtained from our proposed method. The resulting MMSE was \(1.22 \times 10^{-6}\) which is only .37dB better than our proposed joint shortening.

The block-convolution implementation of adaptive echo canceller is also performed for a 50m CAT-7A cable. A block length of 64 is chosen which results in a reasonable complexity reduction and its corresponding FFT/IFFT engine is fairly simple to implement in hardware. The simulation results of echo canceller implemented by block-convolution adapted by standard LMS and normalized LMS are shown in Fig. 5.19. The achievable ERLE is 64.36dB.

Figure 5.18: Channel and echo impulse responses shortened by unit-norm joint shortening filter.
Chapter 6

Coding

6.1 Introduction

Forward error correcting codes (FEC) are key elements in establishing reliable high speed data links. In end-to-end reliable communications, e.g. TCP protocol, packet retransmission can happen due to random errors caused by interfering signals. This retransmission drops the system throughput substantially. This is not desirable especially in backbone Ethernets where time-
critical applications are running. Coded modulation introduced by Ungerboeck [94,95,96] is an efficient way to combine redundant parity check bits to information bits. The fundamental principle of trellis coded modulation, set partitioning, is briefly reviewed in this section. A block coded modulation scheme that shares the same principle with trellis coded modulation is proposed for high-rate, high-gain coding schemes. This code is constructed from a low-rate, high-gain Low Density Parity Check code (LDPC) [41] which is inclusively studied in this chapter. The architecture of parallel decoder of this code is analyzed as well. A multi-channel system is designed and developed to reduce the complexity of the system in terms of logic gate count and power consumption.

6.2 Coded Modulation

On bandwidth-limited channels, nonbinary signal alphabets such as M-PAM must be used to approach capacity [42]. On the other hand, binary codes are required to establish a reliable link between the transmitter and receiver. For moderate coding gains at moderate complexity, the two principal classes of packings are lattices and trellis codes, which are analogous to block and convolutional codes, respectively. The principles of construction of the best such of codes are well understood, and it seems likely that the best codes have been found. After adding the redundant coding bits, a proper mapping should be used to code the aggregated binary input stream to an alphabet set with higher cardinality. In [70], Massey suggested to couple the channel-encoding process with the modulation process at the transmitter. He also advised to perform joint decoding and demodulation at the receiver in order to optimize the performance of the error-correcting algorithm.

Considering an $M$-PAM modulation, Ungerboeck [95][96] realized that to achieve, e.g., 3 bits/dim, 16-PAM could be used instead of 8-PAM (Fig. 6.1) at the price of higher error
But with this expanded signal constellation, the redundant coding bits can be accommodated too. Furthermore, he set up a few simple rules to construct a signal constellation with higher Euclidean distance instead of their Hamming distance using a trellis diagram for introducing memory in the system. He developed the idea and presented a general structural modulation, known as trellis-coded modulation, to combine a conventional convolutional code with multi-dimensional multi-level signal sets.

Figure 6.1: Capacity of multi-level PAM signaling.

To summarize this, the key ideas in the invention of trellis codes were:

- use of minimum squared Euclidean distance as the design criterion;
- coding on subsets of signal sets using convolutional coding principles (e.g., trellises and the Viterbi algorithm).

### 6.3 Trellis Coded Modulation

A lattice constellation \( C(\Lambda, \mathcal{R}) = (\Lambda + t) \cap \mathcal{R} \) is a finite set of points in a lattice translate \( \Lambda + t \) that lies within a compact bounding region \( \mathcal{R} \) of \( N \)-space. An \( N \)-dimensional lattice \( \Lambda \) is
the set of all points \( \{ \mathbf{x} : \mathbf{x} = \mu_1 \mathbf{b}_1 + \cdots + \mu_N \mathbf{b}_N \} \) where \( \mathbf{x} \) is an \( m \)-dimensional vector in \( \mathbb{R}^m \), \( \mathbf{b}_1, \ldots, \mathbf{b}_N \) are \( N \) linearly independent basis vectors in \( \mathbb{R}^m \), and \( \mu_1, \ldots, \mu_N \) range through all integers. In other words, a typical large-constellation trellis code is designed as follows. One starts with a large low dimensional constellation, which in practice is almost always a lattice constellation based on a version of an \( n \)-dimensional integer lattice \( \mathbb{Z}^n \), such as M-PAM or QAM \([37][38]\), and a boundary region \( \mathcal{R} \). One can then form an \( m \)-fold Cartesian product constellation

\[
C(\mathbb{Z}^n, \mathcal{R})^m = C(\mathbb{Z}^{nm}, \mathcal{R}^m)
\]

(6.1)

which is still based on an \( mn \)-dimensional integer lattice \( \mathbb{Z}^{nm} \).

Every level of the partitioning creates a sublattice \( \Lambda' \) of the original lattice \( \Lambda \). An example of this partitioning is shown in Fig. 6.2 for a 2-dimensional lattice for 16-QAM constellation. The remaining points, if we delete \( \Lambda' \) from \( \Lambda \), are a shifted version \( \Lambda_s \) of \( \Lambda' \), called its coset. The constellation \( C(\mathbb{Z}^{nm}, \mathcal{R}^m) \) is partitioned into subsets of equal size, where the number of subsets is typically a power of two, say \( 2^b \). Initially this was done by a sequence of two-way partitions in which the minimum squared distance within subsets was maximized at each level. Subsequently, it was recognized that the resulting constellations were almost always lattice constellations \( C(\Lambda', \mathcal{R}^m) \) based on a sublattice \( \Lambda' \) of index \( \mathbb{Z}^{nm} / \Lambda' \) in \( \mathbb{Z}^{nm} \). In other words, \( \mathbb{Z}^{nm} \) is the union of \( 2^b \) cosets of \( \Lambda' \), and the subsets are the points of \( C(\mathbb{Z}^{nm}, \mathcal{R}^m) \) that lie in each such coset. The sublattice \( \Lambda' \) is usually chosen to be as dense as possible.

In four dimensions, for example, there is a chain of sub-lattices of \( \mathbb{Z}^4 \) as

\[
\mathbb{Z}^4 \supseteq D_4 \supseteq R\mathbb{Z}^4 \supseteq RD_4 \supseteq \cdots
\]

where \( D_4 \) is the 4-dimensional Barnes-Wall lattice \([25]\) and \( R \) is the \( 4 \times 4 \) matrix:
Alternatively, this is the chain of mod-2 lattices corresponding to the (4,4,1), (4,3,2), (4,2,2) and (4,1,4) binary linear block codes. This chain may alternatively be written as \( \mathbb{Z}^4 / D_4 / R\mathbb{Z}^4 / RD_4 / \cdots \). Each partition is two-way. The corresponding minimum squared distances are \( 1/2/2/4/\cdots \). Thus, a 4-D constellation \( C(\mathbb{Z}^4, R) \) with minimum squared distance 1 may be partitioned into 2 subsets of the form \( C(D^4, R) \) with minimum squared distance 2 within subsets, 8 subsets of the form \( C(2D^4, R) \) with minimum squared distance 4 within subsets, etc. Again, the bounding region \( R \) should contain an equal number of points in each subset. The fundamental coding gain \([39][25]\), \( \gamma(\Lambda) = d_{\min}^2 / V(\Lambda)^{2/N} \) relates the minimum squared Euclidean distance \( d_{\min}^2 \) to the fundamental volume per two dimensions that is directly related to the signal energy.

\[
R = \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 1 & -1
\end{bmatrix}
\] (6.2)
A trellis code encoder then operates as shown in Fig. 6.3. Some of the input data bits are encoded in a rate-$\frac{k}{k+r_G}$ 2$^r$-state binary convolutional encoder. Almost always $r_G$ is chosen to be equal 1, so the code redundancy is 1 bit per $mn$ dimensions. The encoder output sequence of $k+r_G$-tuples selects a corresponding sequence of subsets of $C(\mathbb{Z}_2^{mn}, R^m)$ (cosets of $\Lambda'$). The convolutional code and the labeling of the subsets are chosen primarily to maximize the minimum squared distance $d_{min}^2(C)$ between signal point sequences in any possible encoded subset sequence, and secondarily to minimize the maximum possible number $K_{min}(C)$ of nearest-neighbor sequences.

Figure 6.2: Set partitioning of a 16-QAM signal set into subsets with increasing minimum distance. The final partition level used by the encoder is the fourth level - that is, the subsets with two signal points each.
Finally, other input data bits select the actual signal points to be transmitted from the selected subsets. If there is any shaping, it is done at this level.

The decoder makes an error if the path it follows through its trellis does not coincide with the path taken by the encoder, for example, the decoder starts an error at time $j$ by following an incorrect path which remerges with the correct sequence at time $j + L$. In general, there are many overlapping error paths possible. If the $e_{ij}$ is the $i$-th error path departing from the correct path at time $j$, the probability of error, $P$, is the probability of the union of the individual errors $e_{ij}$, i.e.,

$$P = P\left(\bigcup_{i,j}e_{ij}\mid \text{given correct path } c\right)$$  \hspace{1cm} (6.3)

Therefore, the average error probability is given by:

$$\bar{P} = \sum_c p(c)P\left(\bigcup_{i,j}e_{ij}\mid \text{given correct path } c\right)$$  \hspace{1cm} (6.4)

Applying the union bound on this probability twice, the average probability of error can be bounded by

$$\bar{P} \leq \sum_c p(c)\sum_{e_i} P\{e_i\mid c\}$$  \hspace{1cm} (6.5)

where $e_i$ is the event that an error starts at an arbitrary but fixed time unit.
Consider a TCM code with words in AWGN and received signal $r$. Each component of $r$ is a component of the transmitted signal, plus a zero-mean Gaussian noise with variance $N_0/2$.

The probability $\mathcal{P}\{e_i|c\}$, denoted by $\mathcal{P}_{c\rightarrow e_i}$, is easily evaluated as

$$
\mathcal{P}_{c\rightarrow e_i} = Q \left( \sqrt{d_{i}^2 \frac{E_b}{N_0}} \right)
$$

(6.6)

where $E_b$ is the energy per information bit. This equation is commonly called the pairwise error probability. The upper bound of $\bar{P}$ form Eq. (6.5) now becomes

$$
\bar{P} \leq \sum_c p(c) \sum_{e_i \in \mathcal{F}} Q \left( \sqrt{d_{i}^2 \frac{E_b}{N_0}} \right) \leq \sum_{d_i \in \mathcal{D}} A_{d_i} Q \left( \sqrt{d_{i}^2 \frac{E_b}{N_0}} \right)
$$

(6.7)

where the right-hand side of this equation is obtained by counting how often each of the squared Euclidean distances $d_{i}^2$ occurs between the signals on $c$ and $e_i$ in a particular trellis code. $\mathcal{D}$ is the set of all possible such distances $d_{i}^2$, and $A_{d_i}$ is the number of times $d_{i}^2$ occurs, called multiplicity of $d_{i}^2$.

### 6.4 Low-Density Parity Check Code (LDPC)

LDPC codes significantly differ from the trellis codes and block codes. First, they are constructed in a random manner; second, they have a decoding algorithm whose complexity is linear in the block length of the code. Combined with their spectacular performance, this makes LDPC codes a compelling class of codes for many applications. In its simplest form, an LDPC code is a linear block code with a parity-check matrix that is *sparse*.

---

4 Sparse matrices have a small number of nonzero entries.
In [41], Gallager proposed constructing LDPC codes by randomly placing 1’s and 0’s in an \( \mathcal{N} \times \mathcal{N} \) parity-check matrix \( H \) subject to the constraint that each row of \( H \) had the same number \( d_c \) of 1’s and each column of \( H \) had the same number \( d_v \) of 1’s. Codes of this form are referred to as regular \((d_v,d_c)\)-LDPC codes of length \( \mathcal{N} \). In a \((d_v,d_c)\)-LDPC code each information bit is involved in \( d_c \) parity checks and each parity-check bit involves \( d_v \) information bits. This is best understood by the Tanner graph which is a graphical representation of LDPC codes by bipartite graphs (Fig. 6.4). A bipartite graph is a graph in which the nodes may be partitioned into two subsets such that there are no edges connecting nodes within a subset. In the context of LDPC codes, the two subsets of nodes are referred to as variable nodes and check nodes. There is one variable node for each of the bits in the code, and there is one check node for each of the rows of \( H \). This actually breaks the LDPC code into simple parity-check codes, each checks only a small number of code bits.

\[
\frac{\mathcal{M} d_c}{\mathcal{N} d_v} = \frac{d_c}{\mathcal{N}},
\]

which approaches zero as the block length gets large and leads to the name low-density parity-check codes. Like a regular block code, given a parity-check matrix \( H \) for the LDPC code, we can find a corresponding \( \mathcal{K} \times \mathcal{N} \) generator matrix \( G \) such that \( GH^t = 0 \). The

![Bipartite graph for a (3, 4) regular LDPC code.](image)
generator matrix can be used as an encoder according to \( v = uG \). Decoding of LDPC code has been an active research area for the past decade and many efficient algorithms have been developed. Here we consider the massage passing or belief propagation algorithm to accomplish the decoding task.

### 6.4.1 Belief Propagation Decoding

Since the \( H \) matrix is sparse, it can be represented efficiently using lists of its nonzero locations. In this notation, bits are typically indexed by \( n \) or \( n' \) (e.g., \( c_n \)) and the checks are typically indexed by \( m \) or \( m' \) (e.g., \( z_m \)). The set of bits that participate in check \( z_m \) (i.e., the nonzero elements on the \( m \)-th row of \( H \)) is denoted \( \mathcal{N}_m = \{ n : H_{mn} = 1 \} \). Thus, we can write the \( m \)-th check as

\[
 z_m = \sum_{n \in \mathcal{N}_m} c_n 
\]  

(6.8)

The set of bits that participates in check \( z_m \) except for bit \( n \) is denoted \( \mathcal{N}_{m,n} = \mathcal{N}_m \setminus n \). The notation \( |\mathcal{N}_m| \) indicates the number of elements in the set \( \mathcal{N}_m \). These sets should be considered ordered lists, with the \( i \)-th element of \( \mathcal{N}_m \) being indicated by \( \mathcal{N}_m(i) \). The set of checks in which bit \( c_n \) participates (i.e., the nonzero elements of the \( n \)-th column of \( H \)) is denoted \( \mathcal{M}_n = \{ m : H_{nm} = 1 \} \). For a regular LDPC code, \( |\mathcal{M}_n| = d_c \). Let \( \mathcal{M}_{n,m} = \mathcal{M}_n \setminus m \) be the set of checks in which bit \( c_n \) participates except for check \( m \).

The belief propagation decoding is based on decoding as a function of code constraint that we would explain it by the following example. Consider a binary-input AWGN channel (BI-AWGN) case (the derivation for binary symmetric channel is exactly the same, however we
limit ourselves to BI-AWGNC as it is the case almost in this dissertation). Consider one of the
parity check constraints, say $c_1 \oplus c_2 \oplus c_3 = 0$. The channel-output probabilities are continuous and
equal to:

$$
\mathcal{P}(r, c_k) = \begin{cases} 
\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\sqrt{E})^2} & c_k = 1 \\
\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y+\sqrt{E})^2} & c_k = 0 
\end{cases}
$$

(6.9)

From the $\mathcal{P}(r, c_k)$ probabilities, it is possible to compute the probability that the parity
constraint $\mathcal{P}(E)$ is satisfied. The bit-level constraint that these three bits always sum to zero is
useful in decoding and provides information from any of the individual bits to the other two for
decoding. For any constraint event $E$, the maximum a posteriori decoder that observes or uses
the constraint would then maximize:

$$
\max_{c_k=0,1} \mathcal{P}(c_k \mid E)
$$

(6.10)

The a posteriori probability $\mathcal{P}(c_k \mid E)$ can be written as $\mathcal{P}(c_k \mid E) = \alpha \mathcal{P}(c_k) \mathcal{P}(E \mid c_k)$ where
$\alpha$ is a constant independent of any specific value of the encoder output. In general, the constraint
could be written as $E(c_1, c_2, \ldots, c_N) = E(e) = 0$, and the set of all $e$ values that satisfies the
constraint is $S_E = \{e \mid E(e) = 0\}$. The quantity $\mathcal{P}(E \mid c_k)$ is the extrinsic probability given by

$$
\mathcal{P}_{ext}(c_k) = \mathcal{P}(E \mid c_k) = \alpha_k \sum_{e \in S_E \cap c_k \text{ is fixed}} \prod_{i=1}^{N} \mathcal{P}(r,c_k)
$$

(6.11)

where $\alpha_k$ is a constant that does not depend on $c_k$ and it is inconsequential in subsequent
maximization over $c_k$. Therefore, the a posteriori probability is:
Thus, the probability $\mathcal{P}(E)$ is:

$$\mathcal{P}(E) = \sum_{c \in S_E} \prod_{i=1}^{N} \mathcal{P}(r_i, c_k)$$  \hspace{1cm} (6.13)

where the sum is over all vectors $c$ that satisfy this constraint. Each channel output sample initiates calculation of an intrinsic $\mathcal{P}(c_k)$ in $\mathcal{P}(c_k | E) = \alpha \mathcal{P}(c_k) \mathcal{P}(E | c_k)$. The constraint manifests itself through the set $S_E$ where certain bit combinations are not allowed.

The belief propagation decoding of LDPC code starts by initializing all variable nodes and their outgoing variable messages to the value of the corresponding received bit represented as a log-likelihood ratio of the received symbol defined as:

$$\lambda(c_n | r) = \log \frac{\mathcal{P}(c_n = 1 | r)}{\mathcal{P}(c_n = 0 | r)} = \log \frac{\mathcal{P}(c_n = 1 | r, \{ r_i, i \neq n \})}{\mathcal{P}(c_n = 0 | r, \{ r_i, i \neq n \})}$$  \hspace{1cm} (6.14)

After simple algebra and manipulation, the likelihood ratio Eq. (6.14) can thus be written:

$$\lambda(c_n | r) = \log \mathcal{P}(r_n | c_n = 1) + \log \mathcal{P}(c_n = 1 | \{ r_i, i \neq n \}) - \log \mathcal{P}(r_n | c_n = 0) - \log \mathcal{P}(c_n = 0 | \{ r_i, i \neq n \})$$  \hspace{1cm} (6.15)

For a Gaussian channel the first term in Eq. (6.15) equals to $L_c r_n$ where $L_c = 2 \sqrt{\mathbb{E}_c} / \sigma^2$ is the channel reliability. We observe that the terms in the sum can be identified as

$$\lambda(c_n | r) = L_c r_n + \log \frac{\mathcal{P}(c_n = 1 | \{ r_i, i \neq n \})}{\mathcal{P}(c_n = 0 | \{ r_i, i \neq n \})}$$  \hspace{1cm} (6.16)

where the intrinsic term is determined by the explicit measurement $r_n$, affecting the bit $c$, and the extrinsic term is determined by the information provided by all the other observations and the code structure. The probabilities in the extrinsic term can be expressed in terms of the parity
checks. Let \( z_m \), denote the parity check computed using the \( m \)-th check associated with \( c_n \), except for \( c_n \). That is,

\[
z_{m,n} = \sum_{n \in \mathcal{N}_{m,n}} c_n'
\]  

(6.17)

Figure 6.5: Tree representation of bit nodes and check nodes connectivity for the first and second tiers.

If \( c_n = 1 \), then \( z_{m,n} + c_n = 0 \); that is \( z_{m,n} = 1 \) for all the checks \( m \in \mathcal{M}_n \), in which \( c_n \) participates. Similarly, if \( c_n = 0 \), then \( z_{m,n} = 0 \) for all \( m \in \mathcal{M}_n \). We can write Eq. (6.16) as:

\[
\lambda(c_n | r) = L_c r_n + \log \frac{\mathcal{P}(z_{m,n} = 1 \text{ for all } m \in \mathcal{M}_n | \{r_i, i \neq n\})}{\mathcal{P}(z_{m,n} = 0 \text{ for all } m \in \mathcal{M}_n | \{r_i, i \neq n\})}
\]  

(6.18)

We now invoke the assumption that the graph associated with the code is cycle-free. Then the set of bits associated with \( z_{m,n} \) is independent of the bits associated with \( z_{m',n} \), for \( m' \neq m \). We thus have:

\[
\lambda(c_n | r) = L_c r_n + \log \frac{\prod_{m \in \mathcal{M}_n} \mathcal{P}(z_{m,n} = 1 | \{r_i, i \neq n\})}{\prod_{m \in \mathcal{M}_n} \mathcal{P}(z_{m,n} = 0 | \{r_i, i \neq n\})}
\]  

\[
= L_c r_n + \sum_{m \in \mathcal{M}_n} \log \frac{\mathcal{P}(z_{m,n} = 1 | \{r_i, i \neq n\})}{\mathcal{P}(z_{m,n} = 0 | \{r_i, i \neq n\})}
\]  

(6.19)

Now, we define the following log likelihood ratio:

\[
\lambda(z_{m,n} | \{r_i, i \neq n\}) = \mathcal{P}(z_{m,n} = 1 | \{r_i, i \neq n\}) / \mathcal{P}(z_{m,n} = 0 | \{r_i, i \neq n\})
\]  

(6.20)
Then,

$$\lambda(c_n|r) = L_c r_n + \sum_{m \in \mathcal{M}_n} \lambda(z_{m,n} | \{r_i, i \neq n\})$$

$$= L_c r_n + \sum_{m \in \mathcal{M}_n} \lambda \left( \sum_{j \in \mathcal{N}_{m,n}} c_j | \{r_i, i \neq n\} \right)$$

(6.21)

Under the assumption that the checks in $z_{m,n}$ are conditionally independent (if there are no cycles in the graph), we invoke the \textit{tanh} rule of to write:

$$\lambda(c_n|r) = L_c r_n - 2 \sum_{m \in \mathcal{M}_n} \tanh^{-1} \left( \prod_{j \in \mathcal{N}_{m,n}} \tanh \left( -\frac{\lambda(c_j | \{r_i, i \neq n\})}{2} \right) \right)$$

(6.22)

Computation now requires knowing the $\lambda(c_j | \{r_i, i \neq n\})$, the conditional likelihoods of the bits which connect to the checks of $c_n$. How are these obtained? They are obtained the same way as $\lambda(c_n)$: that is, we remove from $c_n$ its distinguished role, and treat all variable nodes alike. However, we must be careful to deal only with the extrinsic information. Let,

$$\eta_{m,n} = -2 \tanh^{-1} \left( \prod_{j \in \mathcal{N}_{m,n}} \tanh \left( -\frac{\lambda(c_j | \{r_i, i \neq n\})}{2} \right) \right)$$

(6.23)

This can be thought of as the “message” which is passed from the check node $m$ to the bit node $n$. Then Eq. (6.22) can be written:

$$\lambda(c_n|r) = L_c r_n + \sum_{m \in \mathcal{M}_n} \eta_{m,n}$$

(6.24)

This can be thought of as a message that the variable node $c_n$ sends to its check nodes.

The problem with these expressions is that we are still left with a product and the complex tanh functions. We can remedy this as follows. First, factor $\eta_{m,n}$ into its sign and magnitude:

$$\lambda(c_j | \{r_i, i \neq n\}) = \alpha_{j,n} \beta_{j,n}$$

$$\alpha_{j,n} = \text{sign} \left( \lambda(c_j | \{r_i, i \neq n\}) \right)$$

$$\beta_{j,n} = \left| \lambda(c_j | \{r_i, i \neq n\}) \right|$$

(6.25)
So that Eq. (6.23) may be rewritten as:

\[
\tanh \left( -\frac{1}{2} \eta_{m,n} \right) = \prod_{j \in \mathcal{N}_{m,n}} \alpha_{j,n} \cdot \prod_{j \in \mathcal{N}_{m,n}} \tanh \left( \frac{1}{2} \beta_{j,n} \right)
\]  

(6.26)

We then have

\[
\eta_{m,n} = - \prod_{j \in \mathcal{N}_{m,n}} \alpha_{j,n} \cdot 2 \tanh^{-1} \left( \prod_{j \in \mathcal{N}_{m,n}} \tanh \left( \frac{1}{2} \beta_{j,n} \right) \right)
\]

\[
= - \prod_{j \in \mathcal{N}_{m,n}} \alpha_{j,n} \cdot 2 \tanh^{-1} \log \log \left( \prod_{j \in \mathcal{N}_{m,n}} \tanh \left( \frac{1}{2} \beta_{j,n} \right) \right)
\]

\[
= - \prod_{j \in \mathcal{N}_{m,n}} \alpha_{j,n} \cdot 2 \tanh^{-1} \log \left( \sum_{j \in \mathcal{N}_{m,n}} \tanh \left( \frac{1}{2} \beta_{j,n} \right) \right)
\]

\[
= \prod_{j \in \mathcal{N}_{m,n}} \alpha_{j,n} \cdot \phi \left( \sum_{j \in \mathcal{N}_{m,n}} \phi \left( \beta_{j,n} \right) \right)
\]

(6.27)

where we have defined

\[
\phi \left( x \right) = - \log \left[ \tanh \left( \frac{x}{2} \right) \right] = \log \left( \frac{e^x + 1}{e^x - 1} \right)
\]

(6.28)

and we used the fact that \( \phi^{-1} \left( x \right) = \phi \left( x \right) \) for \( x > 0 \). The function is fairly well behaved, as shown in Fig. 6.6.

![Figure 6.6: Plot of \( \phi(x) \) function](image)

The decoding algorithm is summarized as follow:
• Step 1: Initialize $\lambda_m = \frac{2}{\sigma^2} r_m$, for each bit node ($\sigma^2 = N_0 / 2$).

• Step 2: Bit nodes send $\mu_{m-n} = \lambda_m$ to each check node $n \in \mathcal{M}_{n,m}$.

• Step 3: Check nodes connected to bit node $m$ send $\sum_{m \in \mathcal{M}_n} \lambda_m$

• Step 4: Variable nodes connected to check nodes $j$ send $\eta_{m,n} = -2 \tanh^{-1} \left( \prod_{j \in \mathcal{N}_{m,n}} \tanh \left( -\lambda_j / 2 \right) \right)$.

• Step 5: When a fixed number of iterations have been completed or the estimated codeword $\hat{u}$ satisfies the syndrome constraint $H \hat{u} = 0$, stop. Otherwise, return to Step 3.

6.4.2 Reduced complexity Decoders

It should be clear from the above that the log-domain SPA algorithm has lower complexity and is more numerically stable than the probability-domain SPA algorithm. We now present decoders of lower complexity which often suffer only a little in terms of performance. The degradation is typically on the order of 0.5dB, but is a function of the code and the channel.

6.4.3 The Min-Sum Decoder

Consider the update equation (6.27) for $\eta_{m,n}$ in the log-domain decoder. Note from the shape of $\phi(x)$ that the term corresponding to the smallest $\beta_{j,n}$ in the summation dominates, so that
Thus, the min-sum algorithm is simply the log-domain SPA with Eq. (6.27) replaced by

\[
\phi \left( \sum_{j \in \mathcal{N}_{mn}} \phi \left( \beta_{j,n} \right) \right) \simeq \phi \left( \phi \left( \min_{j \in \mathcal{N}_{mn}} \beta_{j,n} \right) \right)
\]

(6.29)

Thus, the min-sum algorithm is simply the log-domain SPA with Eq. (6.27) replaced by

\[
\eta_{m,n} = \prod_{j \in \mathcal{N}_{mn}} \alpha_{j,n} \cdot \min_{j \in \mathcal{N}_{mn}} \beta_{j,n}
\]

(6.30)

It can also be shown that, in the BI-AWGNC case, the initialization \( \eta_{m,n} = 2 r_n / \sigma^2 \) maybe replaced by \( \eta_{m,n} = r_n \) when the min-sum algorithm is employed. The advantage, of course, is that knowledge of the noise power in unnecessary in this case.

### 6.5 Block Coded Modulation

We follow the structure of block coded modulation adopted for 10GBASE-T, and try to generalize the formulations. The goal is to design a high-rate high-gain coding scheme from a low-rate powerful code. Suppose that \( g_c \) denotes the minimum required coding gain. An \( [\mathcal{N}, \mathcal{K}] \) block code, either systematic or nonsystematic, operating in the waterfall region is considered here. We construct the transmit symbols as follows. A block of \( \mathcal{K} \) data bits is transformed into \( \mathcal{N} \) coded bits, and arranged into a block with \( n_r \) rows and \( n_c \) columns (If \( \mathcal{N} \neq n_r n_c \) a few dummy bits can be inserted, therefore we assume \( \mathcal{N} = n_r n_c \) hereafter). Another block of \( n_u n_c \) uncoded bits are stacked with the block of coded bits (Fig. 6.7). Then every group of \( n_u + n_c \) bits, read column wise from the constructed block, is mapped to one constellation point in a two dimensional constellation \( \mathcal{X} \), e.g., QAM constellation. The real and imaginary parts of signal points are assigned to two twisted-pairs. The signal points obtained from this constructed block
can be assigned to the other two pairs alternately, or another block can be constructed and is assigned to the remaining pairs. The results yield a 4-dimensional signal. We denote the cardinality of the QAM constellation as $|\mathcal{X}|$. The total rate of this coding scheme is

$$r = \frac{k + s.c}{n + s.c} > r_C = \frac{k}{n}$$

(6.31)

where $r_C$ is the rate of the original block code.

Figure 6.7: Construction of block coded modulation.

The signal space $\mathcal{X}$ is partitioned into a number of cosets $\Lambda$ by set partitioning rules, such that $\delta_i^2 \leq g_c \leq \delta_{i+1}^2$, where $\delta_i^2$ is the minimum intraset distance at partitioning level $l$. The coded bits are used for coset selection and uncoded bits select the constellation points in each coset. This requires $n_r \geq l + 1$ guaranteeing the overall coding gain of $g_c$. To determine $n_a$, assume that a target bit rate of $R_b$ is desired. If a total bandwidth of $W$ is available, the maximum deliverable symbol rate without intersymbol interference is $R_{sym} = 2W/(1+\alpha)$ where $\alpha$ is the roll-off factor of pulse-shaping filter. Hence, the number of bits sent by each symbol is $n_b = R_b / R_{sym}$. Then, the cardinality of the 4-D space supporting this coded system reads as:
Therefore, if we restrict ourselves to rectangular 2-dimentional lattices, \( n_r \) can be determined from \( |X| \geq 2^{n_{b}/r} \). Unfortunately, the parameter \( r_i \) in the right-hand side of Eq. (6.32) is implicitly a function of \( n_r \). This can be resolved by the following rather simple algorithm.

The algorithm starts initializing the total rate with roughly a small value. This will simply overestimate \( n_r \), to make sure that the signal constellation is large enough to accommodate all the \( 2^{n_{b}/r_i} \) signal points. Then, the actual rate is calculated according to Eq. (6.31), and an upper bound for \( n_a \) is obtained, i.e., \( n_b/(2r_i)−n_r \). If \( n_r \) falls under this constraint, then the algorithm stops, otherwise \( n_r \) is decreased one unit and the rate gets updated. This procedure continues until a value of \( n_a \) that satisfies all the constraints is obtained.

Algorithm 1.

1: initialize \( r=0.5 \);
2: set \( |X|^2 = 2^{n_b}/r \);
3: \( s=|\log_2 |X|-n| \);
4: update \( r \) according to: \( r=(k+s.c)/(n+s.c) \)
5: if \( s \leq n_b/(2r)−n \)
   exit;
else
   \( s \leftarrow s−1 \);
   update \( r \): \( r=(k+s.c)/(n+s.c) \);
end
6: goto 2

\[
\mathcal{X} = |X|^2 \geq 2^{n_{b}/r} = 2^{R_b(1+\alpha)/2Wr}
\] (6.32)
At the receiver, first the likelihood ratios of coded bits $u_k, k = 1, \ldots, n_r$ are calculated and estimations for the coded bits are obtained.

\[
\Lambda^k_n = \frac{\Pr(u_k = 1 | r)}{\Pr(u_k = -1 | r)} = \frac{\sum_{u, u'_k = 1} \exp \left\{ -\frac{1}{2\sigma^2_v} r - \mu(u) \right\}}{\sum_{u, u'_k = -1} \exp \left\{ -\frac{1}{2\sigma^2_v} r - \mu(u) \right\}}
\] (6.33)

where $k = 1, \ldots, n_r$. Once the estimation $\hat{u}_1$ is provided, the remaining uncoded bits can be estimated by minimum Euclidean distance rule in coset $\Lambda^{(\hat{u}_1)}$, i.e.

\[
\hat{u}_2 = \mu^{-1} \left\{ \arg \min_{x_n \in \Lambda^{(\hat{u}_1)}} \| r - x_n \| \right\}
\] (6.34)

where $x = \mu(u_{n+1:n_r+n_b})$ is the labeling function.

### 6.6 Architecture of a Parallel LDPC Decoder

To explore the performance and implementation issues of the decoder architecture introduced in section 6.4, we study the underlying hardware to implement the required functions of message passing between the bit and check nodes. The interconnection of the bit and check nodes is determined by the LDPC code itself where for each edge in the code graph physical nets must be instantiated to carry messages between the bit and check nodes. The decoder architecture also requires a method to load new data packets into the decoder and write out packets once they have been decoded. The data path, check node, bit node, and data input/output architecture of the parallel LDPC decoder are described in the following sections.

The data path of the parallel decoder is illustrated in Fig. 6.8. The messages from bit nodes to check nodes and the messages returning from the check nodes to the bit nodes are
carried on distinct sets of wires. The total number of message wires required to connect between the variable and check nodes is calculated as:

\[
(n \times d_v) \text{ graph edges} \times \text{bits/message} \times 2 \text{ paths}
\]  

(6.35)

To ensure correct synchronous execution of the message passing algorithm, it is necessary to insert registers into the data path to align the messages corresponding to each decoder iteration. By associating the registers with the variable nodes as shown in Fig. 6.8, the check nodes become a purely combinatorial logic block which simplifies the floor planning of the overall decoder.

![Diagram](image)

Figure 6.8: Data path between a variable node and one of its check nodes.

### 6.6.1 Check Node Architecture

Each check node performs a parity check across all variables in a row of \( H \). As shown in Fig. 6.9a, the row parity is XOR-ed with each check node input to calculate the value that all other variables in the group imply each individual variable node should take. Along with the parity determination, an implied reliability of the parity in the log-likelihood domain is computed. The reliability update is performed in the log domain according to Eq. (6.27) so that it is
multiplication- and division-free. At the output of the check node, each result is converted back into the log-likelihood domain. The architecture of the reliability update is shown in Fig. 6.9b where the hyperbolic trigonometric functions required by Eq. (6.27) have been merged with the logarithmic and exponentiation functions [12].

![Figure 6.9: Parity check architecture.](image)

### 6.6.2 Variable Node Architecture

The architecture of the variable nodes is shown in Fig. 6.10. At the packet start signal, the decoding of a new packet is commenced and the previous packets results are loaded into the output shift registers. For the first decoding iteration, the messages sent to the check nodes are the sign-magnitude representations of the log-likelihood of the received value, since for a Gaussian channel the received values are the log-likelihoods up to a scaling factor. All messages passed between the variable and check nodes are represented as a sign bit and $b$ magnitude bits. For subsequent iterations, each message entering the variable node together with the received value are converted to 2’s complement and summed. The sign of the sum represents the current
estimate of the decoded bit at each variable node. Outgoing messages are then formed as the
group sum minus the input message of each individual edge. This is the value all other connected
checks and the received value imply that each check should use for the next parity update. All
values are converted back to a sign-magnitude representation and registered, to be used by the
check nodes connected to the variable node in the next decoder iteration. In the case that the
group sum is zero or the outgoing messages sum is zero, the sign bit used is that of the received
bit, as this is the most probable value for the decoded bit.

![Variable node architecture](image)

Figure 6.10: Variable node architecture

### 6.7 Multichannel System

The main idea of the proposed multi-channel approach is based on an important
observation about the water-filling bound repeated here:

\[
\lim_{N \to \infty} \frac{\maximize_{\epsilon_n} W \sum_{n=1}^{N} \log_{2} \left(1 + \frac{\mathcal{E}_n}{\Gamma N_n f_n} \right)}{\sum_{n=1}^{N} \mathcal{E}_n} = N \mathcal{E}_s
\]  

(6.36)
Discrete multi tone (DMT) is an efficient implementation to achieve water-filling bound. At the receiver, a one-tap equalizer $W_n = 1/|H(f_n)|$ is used at each sub-band to compensate for channel attenuations. Because it is a single tap and cannot change the SNR, there is no noise enhancement due to this equalization, unlike other linear or nonlinear time-domain equalizers. When the number of sub-channels is not large enough, transmission on each sub-channel can be treated as regular single-carrier communication, but the noise enhancement in each sub-channel is less than when the entire bandwidth is used for single-carrier data transmission.

Recall the system margin maximization from chapter 4 noting the following remarks. Fig. 6.11 illustrates the system margin versus bandwidth for 40Gbps data transmission over 50m CAT-7A cable. To achieve the required system reliability, with a 6dB margin, a coding scheme with 6dB gain and an optimum bandwidth around 1500 MHz are necessary. A 6dB coding gain can be achieved by some complex multi-dimensional trellis coding or LDPC coded modulation. The coding and decoding schemes must operate at a speed proportional to bandwidth, which is 1600 MHz in this example.

If we divide this bandwidth, say to two sub-bands, with the same average transmit power, we conjecture that the sub-channel with a less severe insertion loss would need a lower coding gain to achieve the requirements. This is somewhat similar to the argument that water-filling always outperforms over uniform power distribution.

In general, we divide up the bandwidth of $W$ into $N$ parallel sub-channels where $w_n$ is the bandwidth of the $n$-th channel. Each sub-channel has its own dedicated equalizer. The capacity and system margin of the $n$-th channel are defined as

$$C_n = w_n \gamma_{\infty}(w_n, g_n) = w_n \log_2 \exp \left( \frac{1}{w_n} \int_0^{w_n} \ln \left( 1 + \frac{\text{SNR}(f + \sum_{i=1}^{n-1} w_i)}{\Gamma(g_n)} \right) df \right)$$  \hspace{1cm} (6.37)

$$\gamma_n^{m}(w_n, g_n) = 3g_n \gamma_{\infty}(w_n, 1)2 \frac{R_n}{W} \left[ Q^{-1}(P_e / N_e) \right]^{-2}$$  \hspace{1cm} (6.38)
where $R_n \leq C_n$ is the data rate of the $n$-th sub-channel.

The total bit rate of $R_b$ is distributed among these channels while we must satisfy the following objectives:

- **bandwidth constraint:** $\sum_n w_n = W$
- **rate constraint:** $\sum_n R_n = R_b$
- **each sub-channel meets the required margin, i.e.,** $\gamma_n^m \geq \gamma_c$, $n = 1, \ldots, N$.
- **coding gains constraint:** $1 \leq g_n \leq g_c$, $n = 1, \ldots, N$

Our goal is to minimize the total power. A meaningful cost function in this regard can be expressed as

$$\phi(w_n, g_n) = k_1 w_n e^{k_2 g_n}$$  \hspace{1cm} (6.39)

where we assumed exponential growth of coding and decoding scheme as the coding gain increases. We also assume that chip power and area are linearly scaled by frequency. Although we simply modeled this by a first-order approximation, more accurate and complex dependency can be incorporated into this model.

Figure 6.11: System margin of 40GBASE-T system over 50m CAT-7A cable.
Regarding this optimization problem, we should note that although the objective function is separable, the margin constraint is not. This is due to the fact that the margin of \( n \)-th channel is the margin of sub-channel in \( \sum_{k=1}^{n-1} w_k, w_n \) frequency range, which clearly depends on \( w_k, k < n \). Besides, we show that the objective function and the constraints for this optimization problem are convex; therefore, well developed algorithms to solve the nonlinear convex problem can be exploited here.

**Lemma 1:** \( \phi(w, g) \) is convex in both \( w \) and \( g \) variables.

**proof:** \( \phi(w, g) \) is an exponential function in \( g \), which is convex. Also, \( \phi(w, g) \) is an affine map\(^5\) in \( w \). Therefore, \( \phi(w, g) \) is the product of two convex functions, which leads to convexity of \( \phi(w, g) \).

**Lemma 2:** The system margin of a bandwidth-limited, strictly monotonous decreasing attenuation channel, \( \gamma^m(w, g) \), is a concave function in both \( w \) and \( g \) variables.

**proof:** From Eq. (6.38), the concavity of \( \gamma^m(w, g) \) in \( g \) is immediate. It is left to show that \( \gamma_\infty(w, 1) \) is concave in \( w \). This can be shown in different ways. Here we present a simple

\[
\begin{align*}
\text{minimize } & f_0(w, g) = \sum_n \phi(w_n, g_n) \\
\text{subject to } & \sum_n R_n = R_u \\
& \gamma_n(w_n, g_n) \geq \gamma_c, n = 1, \ldots, N \\
& 1 \leq g_n \leq g_c, n = 1, \ldots, N
\end{align*}
\]

\[(6.40)\]

\(^5\) An affine map between two vector spaces (strictly speaking, two affine spaces) consists of a linear transformation followed by a translation \( x \rightarrow Ax + b \).
algebraic proof. By checking Eq. (6.37) and noticing that logarithm is a concave function and linear fractionals preserve concavity [13], the concavity of $\gamma_{\infty}(w,1)$ becomes apparent.

These two lemmas enable us to perform conventional convex optimization algorithms to obtain the parameters for multi-channel transmission with its block diagram shown in Fig. 6.12.

![Figure 6.12: Block diagram of multi-channel system.](image)

### 6.8 Simulations and Results

As mentioned before, TCM requires doubling of constellation set size, subsequent partitioning into subsets in order to increase the distance between points. Now, assuming a symbol rate of 2.857GSym/Sec (corresponds to optimum bandwidth of $\approx 1.542$GHz and small roll-off factor of .08); the total number of bits carried by each symbol is 14. These 14 bits can be divided into two 7-bit groups and coded modulated separately by two 256-QAM, separately. Each constellation would require $2 \times 2^7 = 256$ points. The minimum number of PAM levels (M) that has as many points is 16-PAM ($16^2 \geq 256$). This basically maps a 7-bit input to an 8-bit output that is transmitted as 16-PAM, wherein 7 bits are the information bits plus one redundant bit. Adding redundancy does not increase bandwidth due to the fact that signals are now accommodated over
more PAM levels. Simulation result of this encoder using an 8-state rate-2/3 convolutional encoder with generator matrix $G=[04; 02; 11]$ is shown in Fig. 6.13. The asymptotic gain of this code is about 3.98dB.

![Figure 6.13: Simulation results of a 2-D 8-state trellis coded modulation.](image)

There are two major issues with this scheme. First, it requires two separate encoder and decoders which subsequently occupy more chip area and consume more power. Secondly, this code offers a total rate of $7/8=.875$ which is not considerably high. These issues could be resolved employing a 4-D trellis code explained as follows. Suppose a symbol rate of 2.6667GSym/Sec that corresponds to optimum bandwidth of $\approx 1.44$GHz and roll-off factor of .08, 15 data bits enter the encoder. With one extra redundant bit added by convolution encoder, there are totally 16 bits that select a point among the 4-D signal constellation. The signal constellation would require $2^{16}$ points to accommodate these information bits. The minimum number of PAM levels ($M$) that has as many points is 16-PAM ($16^4 \geq 2^{16}$). Thus, this encoder offers a code rate of $15/16=.9375$ and a lower symbol rate that can mitigate the hardware design.
Now, we present the simulation results of LDPC assisted coded modulation for 40GBASE-T system over 50m CAT-7A cable. Before that, a technique that is commonly used to analyze the performance of the belief propagation decoding is briefly overviewed. Density evolution computes, at each iteration starting 0, the message distribution along each class of edge relative to the value of variable associated with each edge. For a given channel, density evolution can determine the bit error probability approached zero as the number of iterations (or block length) goes to infinity. Of course for a well designed code, after a few iterations the probability of error is small enough to stop the process. For the Binary Erasure Channel (BEC) this density simplifies to a probability (of erasure), but for a general case density evolution describes the evolution of densities.

As a demonstration, we used a rate 0.5 (3,6)-LDPC code with 408 parity check bits and 816 block length. This code is selected from the Encyclopedia of Sparse Block Codes by D. MacKay [69]. As the performance is independent of the transmitted codeword, we are free to choose a particular codeword and to analyze the performance of the system assuming that this
codeword was sent. Fig. 6.15 illustrates the probability density of a message sent from a check node to its correspondence bit nodes when “all 0” code is transmitted. Note that after 7 iteration the density evolves and moves toward positive values (there is a $1-2c$ mapping due to BPSK modulation).

![Histograms](https://via.placeholder.com/150)

Figure 6.15: Histograms of the bit-to-check information for various decoder iterations.

Another interesting way of checking the performance of a specific LDPC code is EXIT chart which is an effective visualization of belief propagation decoding (generally, the EXIT chart is a power tool to analyze any iterative decoding method). EXIT chart represents the information exchanged between two components of an iterative decoder on a two-dimensional chart. One component is plotted with its input on the horizontal axis and its output on the vertical axis. The other component is plotted with its input on the vertical axis and its output on the horizontal axis. There is a key assumption on this representation that the messages are independent. Fig. 6.16 illustrates the information (in bits) transferred from a bit node to check node and from the check node to the bit node. Iterative decoding will be successful if and only if the two curves do not cross.
A 6dB coding gain is required for this system to achieve error probability in the order of $10^{-12}$ with 6dB system margin. An example of 6dB set partitioning required for this coding on a 64-QAM constellation is shown in Fig. 6.17.

A simulation was performed to determine the number of levels in each coordinate in the coded modulated system by varying the signal bandwidth or correspondingly the symbol rate. A roll-off factor of 8% is assumed. The results of this simulation are summarized in Table 6.1. The corresponding system margin at each symbol rate is also calculated and given in this table. This bandwidth results in a system margin greater than 6dB. Among these, the one that derives a lower
number of levels is preferred. However, as the signal bandwidth increases, the design of mixed-signal circuitry, e.g., A/D and D/A becomes more complex and challenging. Furthermore, for the sake of implementation ease, we may prefer the number of levels to be a power of 2.

For this specific example further simplifications in terms of mapping can be applied by careful labeling. As explained in section 6.5, the two coded bits $u_1$ and $u_2$ determine one of the cosets $+$, $*$, $\%$, or $\#$. The remaining 6 uncoded bits select a constellation point from the selected coset. However, if we select cosets $\%$ and $+$ when $u_1=0$ and cosets $\#$ and $*$ when $u_1=1$ (accordingly, select cosets $\%$ and $\#$ when $u_2=0$ and cosets $+$ and $*$ when $u_2=1$), the problem of this 2-D set partitioning and coset selection becomes two separate 1-D set partitioning and coset selection: $u_1$ does a 6dB set partitioning over real axis on a 16-PAM constellation and $u_2$ does a 6dB set partitioning in imaginary axis on the same 16-PAM constellation. The 6 uncoded bits are also divided into two individual groups to select signal points in the selected cosets in real and imaginary axes.

Table 6.1: Number of levels versus bandwidth.

<table>
<thead>
<tr>
<th>Bandwidth [MHz]</th>
<th>1500</th>
<th>1550</th>
<th>1600</th>
<th>1650</th>
<th>1700</th>
<th>1750</th>
<th>1800</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-PAM</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Margin [dB]</td>
<td>6.05</td>
<td>6.1</td>
<td>6.13</td>
<td>6.11</td>
<td>6.08</td>
<td>6.05</td>
<td>6.0</td>
</tr>
</tbody>
</table>

The scaled log-likelihood ratio (LLR) of coded bits as a function of received signal is plotted in Fig. 6.18 at low and high SNR values. Since this function involves many complex operations, a simplified version of this function can be used, i.e.

$$
\sigma^2_{llr}(x) \approx \begin{cases} 
-2x - 4 & -2 \leq x \leq -1 \\
2x & -1 \leq x \leq +1, \quad llr(x) = llr(x + 4) \\
-2x + 4 & +1 \leq x \leq +2
\end{cases}
$$

(6.41)
where the following approximation is made:

\[
\max^* (x, y, \ldots) = \log \left( e^x + e^y + \ldots \right) \approx \max (x, y, \ldots)
\]  

(6.42)

Results of computer simulations along with theoretical bounds are depicted on Fig. 6.19. The figure confirms achieving the 6 dB required coding gain.

![Figure 6.18: Scaled LLR values of coded bits](image1)

![Figure 6.19: Bit error probability (BER) of coded and uncoded systems](image2)
For Multichannel system, only two parallel channels are considered here for the sake of implementation ease. Moreover, we need to do some simplifications as explained in the following in order to make the problem tractable. An important question in solving this optimization problem is how to divide the total data rate and assign them to sub-channels. It makes more sense if we assign higher data rate to a sub-channel that has a less severe condition, i.e., lower insertion loss (like water-filling algorithm). Furthermore, we try to obtain a coding scheme with lower gain and less complexity for the channel with higher bandwidth while satisfying the system margin, as it was explained in section 6.7. We propose to use 4-D trellis coded modulation. This code can achieve a coding gain of about 4-5dB at low to moderate complexity.

Suppose that the bandwidth of the first channel is $w_1$. The symbol rate of this channel is $R_{s1} = 2w_1/(1+\alpha)$, where $\alpha$ is the excess bandwidth. If $r_{b1}$ represents the number of uncoded bits per 4-D symbol, and $M$ is the number of PAM level in constituent one dimension of this constellation (we assumed that the constellation is based on $Z^4$ lattice or its variants), then the following should hold for the expanded signal constellation to accommodate one extra coded bit from Viterbi encoder,

$$M^4 \geq 2^{r_{b1}+1} \quad (6.43)$$

From the past experiences, $M=16$ seems to be the maximum number of levels that could be accommodated with the current technological advances and innovations in designing analog-to-digital converters at the speeds we are interested. The quantization noise and jitter are the main limiting factors in designing mixed-signal circuits for higher number of levels. Therefore, we fix the number of PAM levels to $M=16$. Thus, the corresponding data rate for the first sub-channel becomes

$$R_{b1} = r_{b1}R_{s1} = 30w_1/(1+\alpha) \quad (6.44)$$
The data rate of the second sub-channel obviously becomes $R_{b2} = 40 \times 10^9 - R_{b1}$. Now, the set of constraints is defined and we can solve the problem. To solve this optimization problem with convex equality and inequality constraints, we consider the Interior Point method [13] [45] [103]. The variables for this optimization are $w_1$, $w_2$ and $g_1$, while $g_2 = 6$dB is fixed. Results of this optimization are presented in Table 6.2.

Table 6.2: Parameters of 2-channel implementation

<table>
<thead>
<tr>
<th></th>
<th>Channel 1</th>
<th>Channel 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_b$ (Gbps)</td>
<td>27.368</td>
<td>12.632</td>
</tr>
<tr>
<td>Bandwidth (MHz)</td>
<td>985.2</td>
<td>605</td>
</tr>
<tr>
<td>Coding gain (dB)</td>
<td>1.8</td>
<td>6</td>
</tr>
<tr>
<td>Code class</td>
<td>4-D trellis</td>
<td>LDPC</td>
</tr>
<tr>
<td>PAM levels</td>
<td>16</td>
<td>9</td>
</tr>
</tbody>
</table>

As we predicted, the channel with less severe insertion loss carries a higher data rate over a higher bandwidth. However, the coding gain for this sub-channel is reduced and this contributes to substantial power reduction.

We finish this section noting the following remarks. First, we did not consider a guard band between the two sub-channels. However, the inclusion of a guard band into this analysis will only impact the upper sub-channel which is assigned to a lower data rate and less complex system. Secondly, we could take the coding gain of the second sub-channel as another variable and perform the optimization procedure. Secondly, in the implementation of this 2-channel system, the number of ADC converters is doubled, but each one operates at a lower frequency which makes the implementation much easier, as we know that the complexity of mixed-signal circuits grows exponentially.
7.1 Introduction

High data rate and increased bandwidth of 40GBASE-T (or 100GBASE-T) impose serious challenges in designing low-cost and power efficient implementation required for both servers and the core data-center network. Clearly, lower power and higher level of integration will drive port density and cost over time. There is no question that the 40GBASE-T PHY uses a more significant fraction of the total power budget available than was the case for the previous generation of coax- or fiber-optic-based network adaptors. In this section, we intend to briefly discuss some of the issues of hardware implementation, specially for mixed-signal circuits. Since CAT-7A cables are shielded and isolate the internal signals from the external interference sources, the objective of transmitting data at the rate of 40Gbps is in fact background-noise-limited communication system. It is assumed (and it was confirmed in the process of 10GBASE-T development) that the noise from the analog circuits is much stronger than the thermal-induced noise of the shielded cables. Here we present a review of the dominant noise sources in analog circuits and provide some guidelines for implementing proper circuits and estimate the cost, power and latency of such an implementation.

7.2 Mixed-Signal Circuits

Analog-to-digital converters set serious limitations on the characteristics of the transmitted signal, and therefore they are sometimes a capacity-limiting factor for the link. It was
the case for 10GBASE-T standard, for which a preliminary survey of the ADC technology was made in order to determine the technical feasibility of the 10Gbps transmission [53].

The sampling speed of the ADC determines the limit on the exploitable bandwidth of the channel: according to the Nyquist criterion, the bandwidth of the signal that is reliably transformed by the ADC, is at most half of the sampling rate. For instance, a 3GS/s ADC can detect analog signals with a bandwidth of at most 1500 MHz. The precision of the ADC, expressed in terms of effective number of bits (ENOB), can be interpreted as a white noise limit on the channel.

As any electronic component, an ADC introduces noise between the input and its output. However, unlike any other electronic component, the ADC input voltage is analog, while the output voltage is digital. Hence, even an ideal ADC introduces some inherent noise, due to the precision error that is made when converting voltage values from a continuous analog scale to a discrete digital scale. Therefore, an unavoidable white noise results from the finite number of comparators in the ADC, comparators which are needed in order to distinguish between the discrete output quantization levels. Thus, this noise is called quantization noise. A quick calculation shows that the ratio between the input signal power and the quantization noise power at the output of an ideal ADC is [48]:

$$\text{SNR}_{\text{ideal}} (\text{dB}) = 6.02 \cdot \text{NOB} + 1.76$$  \hspace{1cm} (7.1)

where NOB is the integer number of bits necessary to code the output level. For instance, an ideal ADC with a precision of 11 bits, outputs towards a set of \(2^{11} = 2048\) discrete levels as it contains \(2^{11} - 1 = 2047\) comparators. Its SNR is then 67.98 dB.

However ADCs are real electronic components, and therefore they are affected by other unwanted noise sources. Generally, ADC performance is limited by:

- **Thermal noise** – especially for the precision of low and medium speed ADCs
• **Aperture jitter noise** – for the precision of high speed ADCs

• **Comparator ambiguity noise** – in terms of speed and precision of high-speed ADCs.

These noises are add to the quantization noise and they enter into the calculation of the output SNR of the real ADC, which by convention, is then converted back in an effective number of bits (ENOB), according to:

\[
\text{ENOB} = \frac{\text{SNR}_{\text{real}} (\text{dB}) - 1.76}{6.02} \quad (7.2)
\]

An ADC with a precision of 11 ENOB has therefore a total white noise of 67.98 dB below the level of the input signal power level. For a 10dBm signal transmitted over a bandwidth of 1500 MHz, the quantization noise would be 67.98 dB below the -81.76dBm/Hz signal average power density, at -149.74dBm/Hz.

A -140dBm/Hz quantization white noise floor has been retained for the 10GBASE-T standardization group discussions. However, in the capacity calculations, described in Chapter 3, we have considered a more optimistic hypothesis of the white noise level of -146dBm/Hz.

### 7.2.1 Quantization Noise

We can study the ADC requirements in terms of quantization noise by examining the requirements of data transmission, e.g., the error probability. For an M-ary PAM modulation, the error probability is given by [86]:

\[
p_{\text{PAM}} = 2 \left(1 - \frac{1}{M}\right) Q\left(\sqrt[3]{\frac{3\text{SNR}}{M^2}}\right) \quad (7.3)
\]

One can apply a given coding gain \(g_c\) and system margin \(\gamma_m\), to obtain:
The SNR requirement as a function of M, the quantization levels, for different coding gains and system margins at the given probability of error of $10^{-12}$ is shown in Fig. 7.1. It is noted from this figure that the minimum SNR to achieve the error probability of $10^{-12}$, for example in 16-PAM case, is 36.35dB. This corresponds to an ADC with at least 6 ENOB. To confirm this, we setup a fixed-point simulation of a 16-PAM modulator and demodulator for 40Gbps transmission ($BW \approx 1440MHz$ and $P_s = 10dBm$) and we varied the ENOB of ADC. The setup of this simple system and the result of the bit error rate (BER) simulation are depicted in Fig. 7.2. This simulation shows that at least 9 bits are required to achieve a performance close to the floating-point implementation at error probability of $10^{-8}$.

![Figure 7.1: Minimum SNR per symbol vs. bits per symbol for M-PAM transmission.](image)

However, the previous calculations show that the ADC ENOB should be at least 10.4 to bring down the quantization noise below the Gaussian background noise of -146dBm/Hz given 1440 MHz of bandwidth.

$$P_{\text{PAM}} = 2 \left(1 - \frac{1}{M}\right) Q\left(\frac{g_c}{\gamma_m} \frac{3\text{SNR}}{M^2} - 1\right)$$  \hspace{1cm} (7.4)
Fig. 7.3 presents the ADC ENOB requirement versus bandwidth assuming the Nyquist rate sampling. The admissible operating region is above the shadowed region. Interestingly, increasing the signal bandwidth downgrades the ADC ENOB. This is simply due to higher symbol rates which lower the PAM levels for a fixed data rate. Therefore, one can find the trade-offs between the ADC ENOB and sampling speeds where simply speaking the higher sampling speeds, the less constraints on the ADC ENOB. Unfortunately, designing faster ADC means higher power consumption and more sophisticated analog circuits. Definitely, these parameters are not scaled linearly with frequency and more investigations are required to find out the optimum operating point.

Figure 7.2: Block diagram of a fix-point implementation of 16-PAM modulator/demodulator.
7.2.2 Clock Jitter

Clock jitter is one of the sample-to-sample variations in the clock source, and it can limit the noise performance in a system. As the ADC’s input bandwidth increases, the demands on the clock source become even greater, because the slew rate of the incoming signal increases with frequency and generates a greater error voltage. The amount of voltage error that is generated by clock jitter, as it relates to slew rate, can be calculated using [48]:

\[ v_{err} = \text{slew rate} \times t_{jitter} \tag{7.5} \]

To see how clock jitter affects SNR as a function of analog frequency, the following equation can be used [32][33]:

\[ \text{SNR} = -20 \times \log_{10} \left( 2\pi \times f_{\text{analog}} \times j_{\text{rms}} \right) \tag{7.6} \]

A family of curves that show the effect of clock jitter on effective number of bits of an ADC across several analog frequencies is illustrated in Fig. 7.4.
It is noted from this figure that increasing the signal bandwidth puts more constraints on the uncertainty of system’s clock.

**7.2.3 ADC Requirements Imposed by Echo Canceller**

Since the echo is the strongest crosstalk among many other impairments, including internal signals or other naturally induced noises, the performance of an echo canceller can be fully realized only if the ADC device in the receive path has a sufficient number of bits to represent the smallest residual echo. The quantization noise level contributed by both an $N$ bit ADC converter and an echo canceller is about [17]:

$$Q_e = 10 \log_{10} \left( \frac{2}{3} \left( \frac{1}{2^N} \right)^2 \right)$$  (7.7)
The reachable echo cancellation level, $K_{eccan}$, is related to the maximum echo level representable by the ADC converter $E_{AD}$, the quantization noise level $Q_e$, and the excessive adaptation noise level $\varepsilon$.

\[
K_{eccan} = E_{AD} - Q_e - \varepsilon
\]

(7.8)

For known values of $K_{eccan}$, $E_{AD}$, and $\varepsilon$, solving for $N$ through $Q_e$, we have:

\[
N \geq \frac{1}{6.02} \left( K_{eccan} - E_{AD} - \varepsilon \right) - \frac{1.76}{6.02}
\]

(7.9)

For $E_{AD} = 10\log_{10} \frac{2}{3} \left( \frac{1}{2} \right)^3$ corresponding to a uniformly distributed maximum echo level representable by the ADC converter, and $\varepsilon = 3\text{dB}$, we obtain $N = 12$. In practice, the echo is more likely to have a Gaussian distribution according to the law of large numbers or the central limiting theory [80] for the long echo time span. Since any overflow at the ADC converter will affect the receiver performance at the detection point, we need to properly set up the ADC dynamic range according to the average echo power. Using a Gaussian distribution, if we set the maximum representable value to 7.44 times that of the average echo power, the probability of overflow will be $1 \times 10^{-13}$, i.e.,

\[
1 - \int_{-7.44}^{7.44} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)} dx = 1 \times 10^{-13}
\]

(7.10)

which is an order of magnitude lower than the required receiver error rate. For this dynamic range and a Gaussian distribution, the echo level becomes

\[
E_{AD} = 20\log_{10} \frac{0.5}{7.44} = -23.45\text{dB}
\]

(7.11)
and the required number of bits becomes $N = 14$. However, the echo might only have a truncated Gaussian distribution. In this case, a 12-bit ADC converter is good for a maximum to root mean power ratio of 1.8 and a 13-bit ADC is good for a ratio of 3.6. The actual distribution of the echo is to be verified by experiment. A 13-bit ADC device is necessary to achieve a 60 dB echo cancellation level. A 12-bit ADC might be able to do the job with marginal performance if all other echo canceller hardware works extremely well.

7.3 System Complexity

The overall complexity of 40GBASE-T system is mainly due to analog mix-signal circuits, e.g. ADC and DAC, and DSP algorithms which consist of channel equalization, echo cancellation, timing recovery, and finally coding and decoding. The FEXT and NEXT cancellers are not used for this system and still the requirements are met by a proper transmit power and code selection. This is mainly because of proper connector design and doubly shielded property of CAT-7A cables that makes the twisted-pairs almost isolated from one another. For the sake of comparison, we will use the 10GBASE-T parameters as base lines (Table 7.1).

<table>
<thead>
<tr>
<th>Table 7.1: 10GBASE-T FIR filters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong># of taps</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Echo $500 \times 4$</td>
</tr>
<tr>
<td>NEXT $100 \times 12$</td>
</tr>
<tr>
<td>FEXT $300 \times 12$</td>
</tr>
<tr>
<td>Total 6800</td>
</tr>
</tbody>
</table>
These numbers were obtained assuming 0.24mW/tap at 800Mbaud implemented in 90nm technology. It is reported that these cancellers account for 40% of the chip area and roughly 40% of the total power consumption. This power has linear dependency on frequency and squared voltage. Equivalent numbers can be obtained in 65nm technology, by voltage scaling (1.2v in 90nm and 1v for 65nm). More reduction is also applicable considering dimension scaling which lowers the parasitic capacitance and resistance. Assuming another 25% reduction due to dimension scaling, the total power of crosstalk cancellers becomes .85Watts. To estimate the power of these cancellers in 40GBASE-T implemented in 65n technology, we must consider the scaling due to symbol rate. The symbol rate of 40GBASE-T is about 3Gsym/sec; therefore, a scaling of 3/0.8 is required on all these figures and numbers. The results are summarized in Table 7.2.

Table 7.2: 40GBASE-T FIR filters.

<table>
<thead>
<tr>
<th></th>
<th># of taps</th>
<th>Power (Watts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echo</td>
<td>1200×4</td>
<td>2.25</td>
</tr>
<tr>
<td>NEXT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FEXT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4800</td>
<td>2.25</td>
</tr>
</tbody>
</table>

Another major contributor to the chip power, size, and latency is the encoder and decoder. We discussed that coding gain of about 6dB offers very satisfactory results in terms of error probability and system margin ($P_e \leq 10^{-12}$ and $\gamma_m \geq 6$dB). The coding gain requirements in 10GBASE-T is a lot more severe (9dB gain with 12dB set partitioning). LDPC(2048,1723) is the standard code used in this application. The intrinsic latency of this code with 8 iterations is about
350ns (30ns for encoding and 320ns for decoding). However, practical implementations reported a latency as high as 800ns. The encoder and decoder occupy about 25% of the chip’s area.

A good alternative for this code is the LDPC(1024, 833) which was extensively reviewed and analyzed in 10GBASE-T study group. This code achieves lower coding gains but offers considerably less complexity and latency. It is reported that the latency of this code is as low as 160ns. A good figure of merit in comparing LDPC codes is the number of edges connecting variable nodes and check nodes. The number of edges of LDPC(1024, 833) is 10240, while the LDPC(2048, 1723) has 12280 edges. This means that power consumption can be reduced by 17% employing this code. Roughly speaking, an extra 25% power reduction is achieved employing this code since LDPC(1024, 833) achieves its performance in 6 iterations while LDPC(2048, 1723) requires 8 iterations.

In pursuing the power and latency reduction, much simpler codes can be designed. For this purpose, we designed a simple LDPC(512,409) code. The specifications of this code and other codes are summarized in Table 7.3.

<table>
<thead>
<tr>
<th>Code</th>
<th># of edges</th>
<th># of iterations</th>
<th>Estimated power (Watts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDPC(2048,1723)</td>
<td>12280</td>
<td>8</td>
<td>5.625</td>
</tr>
<tr>
<td>LDPC(1024,833)</td>
<td>10240</td>
<td>6</td>
<td>3.517</td>
</tr>
<tr>
<td>LDPC(512,409)</td>
<td>7680</td>
<td>5</td>
<td>2.198</td>
</tr>
</tbody>
</table>

The rest of a 40GBASE-T chip will be the mixed-signal analog circuits (e.g. ADC, DAC, line drivers, PLL, and other analog circuits) which are about 35% of a 10GBASE-T chip. Two main parameters for these circuits are power and chip area. Unfortunately, to estimate these parameters we cannot use the 10GBASE-T baselines, as the specification of analog circuits are
not scalable linearly with improvements in VLSI technology. An estimated partitioning of 40GBASE-T chip area is shown in Fig. 7.5.

![Figure 7.5: Estimated partitioning of 40GBASE-T chip area.](image)

To model the latency of 40GBASE-T system, proper framing should be considered. Unfortunately, there is, generally speaking, a strong trade-off between coding and latency (if block coding is considered for 40GBASE-T). Usually, higher coding gains require longer data blocks. Several factors govern the latency and delay of LDPC decoder, including:

- Block length
- Frame structure
- Number of pipelining stages

The system latency drops down for shorter block lengths. However, shorter code can not offer higher gains and require higher speed for full parallel implementation. The clock frequency versus the block length for a fully parallel LDPC decoder is illustrated in Fig. 7.6. The corresponding latency of this decoder is also shown in Fig. 7.7.

Remarks below explain this figure. First, changing the number of iterations in the decoding process does not change the latency. This, in fact, can be explained simply. The clock frequency of the decoder is (for a fully parallel decoder):
where $R_b$ is the data rate, $i_{\text{max}}$ is the maximum number of iterations, and finally $K$ is the length of uncoded bits block. Moreover, the latency of this decoder is:

$$\tau = 2i_{\text{max}}T_{\text{clk}} = 2i_{\text{max}}f_{\text{clk}}^{-1} = 2i_{\text{max}} K / (R_b i_{\text{max}}) = 2K / R_b$$

which is independent of the number of iterations. However, a lower number of iterations results in a lower power.

Figure 7.6: Clock frequency of LDPC decoder versus block length.

Figure 7.7: Latency of LDPC decoder versus block length for 40GBASE-T system.
Second, the latency of the LDPC decoder for the 40GBASE-T system is smaller than 10GBASE-T decoder. As it is obvious from Eq. (7.13), this is due to the much higher data rate and consequently the faster clock frequency. Finally, we should emphasize here that in this analysis we only considered the latency of the decoder; the delay of the demapper or other read/write operations are assumed to be in the same order as the decoder latency.
8.1 Summary of Results

Scaling Ethernet heights, as much as 100Gbps, are approaching faster than expected. 100G Ethernet over copper could result in performance gains and cost savings for customers of corporate LAN in the near future. Currently, several 10G links are used in data centers to aggregate switch connections and links to supercomputer clusters for computer-intensive research and data processing. This dissertation addressed the technical feasibility and system design of 40 and 100GBASE-T systems over standard CAT-7A cable. The summary of the results are as follows:

1. The background noise level in CAT-7A cable turns out to be an important factor in 40 and 100GBASE-T applications as we are reaching the theoretical maximum throughput of these channels. Although these cables are much better in terms of isolation and thermal noise compared to other UTP cables, we set the background noise level to -146dBm/Hz in our simulations. This is mainly because the noise from analog front-end is dominant in these systems. The measured noise levels of cable itself were reported as low as -160dBm/Hz.

2. Simulations of CAT-7A cables revealed that data transmission at 40Gbps is limited to ~50 meters. To extend this range to 100 meters at reasonable cost, while ensuring robust and reliable performance through four connections, Category 8 cabling is the most likely choice to support the emerging IEEE 40/100GBASE-T application. With various degrees of DSP, the objective of
running 40GBASE-T over 100m of Category 8 cable can be achieved with some efforts by the silicon vendors probably in the next 2 or 3 generations of CMOS technology. Using digital signal processing techniques, the silicon vendors can cancel, or compensate for, most of the internal impairments including insertion loss and return loss. Also, based on our modeling and analysis, the maximum achievable rate over 20m cables is well above 100Gbps.

3. Although capacity analysis is useful for technical feasibility assessment of these systems, we presented two important optimization procedures to find the optimum bandwidth for a given set of constraints, e.g., probability of error, system margin, average transmit power, etc. These optimization procedures provided more insight in designing ultra high-speed systems in terms of system trade-offs.

4. In full-duplex transmission, the receiver front-end needs to optimize the level of echo cancellation. Echo cancellation imposes very harsh constraints on the speed and precision of mixed-signal circuitry. It turned out that achieving high echo cancellation level is an extremely challenging design problem in 40 and 100GBASE-T systems. Higher symbol rates (and also cable longer cables) result in larger number of filter taps in corresponding canceller. Very high-bandwidth transmission makes the design of broadband hybrid circuits extremely difficult, sometimes even impractical, to achieve fair isolation between transmitter and receiver. The result of such imperfect isolation is very long echo impulse responses which have to be cancelled digitally. Finite-length impulse response (FIR) implementation of echo canceller results in over 1000 taps. Unfortunately, long digital echo cancellers suffer from convergence issues and high quantization noise of fixed-point implementation.
5. FEXT and NEXT, the alien NEXT and alien ELFEXT, which are the crosstalk coupling into a twisted-pair from surrounding cables and other external noise sources, are negligible thanks to excellent shielding and engineering design of CAT-7A cable.

6. One of the major applications of rate optimization algorithms developed in this dissertation is code selection. The coding gain required to achieve the given system margin and probability of error (typically 6dB and $10^{-12}$, respectively) is determined by these optimization procedures. It was shown in this dissertation that a coding scheme with a coding gain greater than 6dB is promising to meet the requirements of 40GBASE-T over 50m.

Based on the investigated properties of standard twisted-pair copper channels, this research provided extensive study of communications schemes and coding to provide a reliable and high data rate link for this channel. Several methods were proposed to tackle the burdens of practical implementation issues in 40/100GBASE-T system. The summary of these efforts is presented as follows:

1. The primary objective of this research was to obtain or define the reach mode of 40/100 Gbps transmission over CAT-7A cable. This research provided a *how-far how-fast* model to identify both short and full reach-modes of these systems. This model takes into account all theoretical aspects and implementation issues to setup a reliable communication link for a required quality-of-service.

2. Two rate optimization procedures, minimizing the probability of error and maximizing the system margin, are the key elements in developing the aforementioned *how-far how-fast* model. These algorithms play a crucial rule in connecting the physical properties of twisted-pair cables, the required quality-of-service and finally the engineering implications.
3. To support a data rate of 40/100Gbps over copper wire, a very long FIR echo canceller is required to achieve a proper cancellation level. We proposed the MMSE joint channel shortening technique to preserve the implementation costs, power consumption and convergence of (possibly adaptive) FIR echo cancellers. The previously proposed unit-tap MMSE joint channel and echo impulse response shortening is revisited in this dissertation and the computational complexity of search algorithm is reduced by carefully examining the structure of underlying matrices. The unit-norm constraint is proposed for target channel impulse response in the MMSE joint shortening framework, and the corresponding algorithm is offered. The unit-norm constraint is already proved to outperform the unit-tap constraint, which was the main motivation of our work. We also proved that the unit-norm constraint on target channel impulse response is sufficient to avoid non-trivial answer for target echo impulse response. The analytical solution for MMSE joint shortening problem subjected to a single norm-constraint is formulated and the MATLAB algorithm is presented.

4. Perfect impedance matching and isolation have been a major challenge in designing connectors for cabling assembly. Expensive connectors are required to attach the twisted-pair cable to end-user (a server computer for example). By advances in silicon technology, the complex signal processing becomes cheaper to implement. It would be reasonable to move the crosstalks and impairments mitigation into the DSP, at least partially, and therefore reduce the cost by using regular and cheap connector. We applied the MIMO theory in reducing the NEXT interference. Theoretically, the MIMO NEXT cancellation enables us to incorporate joint processing and reduce any source of redundancy. This same approach can be applied to FEXT cancellation.
5. Block coded modulation is considered to achieve high coding gain while conserving high spectral efficiency. LDPC coded modulation is considered in this dissertation to obtain a high-rate and high-gain coding scheme from a powerful low-rate high-gain LDPC code. We designed a system to achieve 6dB gain with simple decoder circuit to substantially reduce the implementation cost and power consumption. We also present an algorithm to reduce the power consumption of single-carrier data transmission over twisted-pair copper cables for the future IEEE networking physical connectivity standard, 40/100GBASE-T. High-complexity coding schemes employed for multi-gigabits transmission systems for higher reliability, e.g., system margin, is one of the major sources of power dissipation. The proposed method reduces the total power consumption by properly partitioning the channel into several sub-channels and using lower complexity coding schemes for the sub-channel with less insertion loss while the margin requirement is met in the entire channel. In sub-channels with lower coding gains, trellis coded modulation could be a good candidate accomplishing this coding gain, while for sub-channels with higher coding gain our developed LDPC coded modulation is the best competitor.

6. Designing a low-cost and power efficient implementation of 40/100GBASE-T system is clearly the challenging part. Some of the issues of hardware implementation, specially for mixed-signal circuits were addressed in this dissertation. As the CAT-7A cables provide a shielded environment and the twisted-pairs are isolated almost perfectly, the performance of the system is in fact limited to that of analog interface circuits. Dynamic and static behavior of ADC and DAC circuits were discussed to prevent the degradation in the overall system performance.
8.2 Future Work

Based on the contributions and tools developed in this dissertation, future research in development of ultra high-speed communications over standard twisted-pair cables should focus more on practical aspects of these fields.

- Inexact background noise level assumption can severely affect other system parameters. Better understanding of noise sources and accurate measurements are vital in determining the exact noise floor for the upcoming VLSI technologies.

- In channel equalization, we evaluated the performance of MMSE decision-feedback equalizer and Tomlinson-Harashima precoding. MMSE Interference Cancellation (MMSE-IC) and turbo equalization are better candidates in obtaining better performance obviously at the cost of higher latency and power consumption.

- Achieving high echo return loss enhancement by digital filtering seems to be extremely ambitious for 40/100GBASE-T systems. Return loss is a measure of the signal reflections occurring along a transmission line and relate to impedance mismatch in a cabling channel. Therefore, some efforts in designing broadband matching circuit and hybrid echo cancellation could alleviate this problem to a large extent.

- In this dissertation, we showed that the multi-channel method can compensate for the shortcomings of single-carrier transmission over dispersive channels in terms of noise enhancement, and it consequently lessens the need for very complex coding scheme. However, more investigations are necessary on this method considering the guard band and practical implementation issues.
8.3 Publication

The following is a list of publications.


Appendix A

Quadratic Optimization with unit-tap constraint

This appendix presents the unit-tap constraint quadratic optimization. We are interested in the following optimization problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) = x^tAx + 2b^tx \\
\text{subject to} & \quad x_j = 1, \exists j \in \mathcal{J}
\end{align*}
\]  \tag{A.1}

where \(x \in \mathbb{R}^n\), \(A \in \mathbb{R}^{n \times n}\) is symmetric positive definite, \(b \in \mathbb{R}^n\), and finally \(\mathcal{J}\) is any subset of indices \(\{1, 2, \ldots, n\}\). The constraint \(x_j = 1\) can be replaced by \(x^te_j = 1\) where \(e_j\) is the \(j^{th}\) unit vector. The standard approach to solve this problem is Lagrange Multiplier method as follows:

\[
L(x, \lambda) = x^tAx + 2b^tx + \lambda(1 - x^te_j)
\]  \tag{A.2}

Euler optimality condition, i.e., \(\partial L / \partial x = 0\) gives:

\[
x = A^{-1}(\lambda e_j - b)
\]  \tag{A.3}

Applying the given constraint, optimum values for \(\lambda\) and \(x\) are:

\[
\lambda^* = 1 + b^tA^{-1}(\cdot, j) / A^{-1}(j, j)
\]

\[
x^* = A^{-1}\left(1 + b^tA^{-1}(\cdot, j) / A^{-1}(j, j)\right)e_j - b
\]  \tag{A.4}

After simple manipulation, the minimum value of quadratic objective function \(f_0(x^*)\) can be calculated as:
The second term, $\mathbf{b}^t \mathbf{A}^{-1} \mathbf{b}$, is independent of $j$, therefore the optimum $j$ that results in a global minimum is:

$$j^* = \arg\min_{j \in \mathcal{J}} \frac{(1 + \mathbf{b}^t \mathbf{A}^{-1}(:,j))^2}{\mathbf{A}^{-1}(j,j)}$$  (A.6)
Appendix B

Quadratic Optimization with unit-norm constraint

In this appendix, we review the following norm-constraint optimization problem.

\[
\text{minimize } f_0(x) = x^t A x + 2b^t x \\
\text{subject to } \| x \|_2 = 1
\]  \hspace{1cm} (B.1)

The Lagrangian of Eq. B.1 is

\[
L(x, \lambda) = x^t (A - \lambda I)x + 2b^t x + \lambda
\]  \hspace{1cm} (B.2)

Setting the gradient of the Lagrange function to zero results in normal equation

\[
(A - \lambda I)x = -b
\]  \hspace{1cm} (B.3)

In general, there are many pairs \((\lambda, x)\) satisfying the normal equation. Moreover, for fixed multiplier \(\lambda = \mu_j\) with the \(j\)th eigenvalue of \(A\), the unconstrained normal equation may have multiple solutions. If the multiplier \(\lambda \neq \mu_j\) for all \(j = 1, \ldots, n\), the corresponding vector \(x\) is uniquely determined by \(\lambda\) and

\[
x = -(A - \lambda I)^{-1} b
\]  \hspace{1cm} (B.4)

In this case, the multipliers required are the roots of the secular equation

\[
f(\lambda) = \| (A - \lambda I)^{-1} b \|_2 = 1
\]  \hspace{1cm} (B.5)

We are interested in the solutions \((\lambda, x)\) with respect to the optimal multiplier \(\lambda^*\). It is shown in [43] that the minimizer of problem Eq. (B.1) is given as the solution of the normal
equation with the largest $\lambda$. We would not get into many other details on the feasibility conditions and solving the normal equation as they are discussed comprehensively in [43].

Another approach to solve the problem Eq. (B.1) is as follows. The dual function of the primal problem is

$$
g(\lambda) = \inf_{\lambda} L(x, \lambda)$$

$$
\begin{cases} 
\lambda - b'(A - \lambda I)^{-1} b & (A - \lambda I) \succeq 0 \\
\infty & b \notin R(A - \lambda I)
\end{cases}
$$

Using the Schur complement [13], we can express the dual problem as

$$
\begin{align*}
\text{maximize} & \quad \gamma \\
\text{subject to} & \quad \lambda \geq 0 \\
& \quad \begin{bmatrix} A - \lambda I & b \\ b' & \lambda - \gamma \end{bmatrix} \succeq 0
\end{align*}
$$

This is a semidefinite program for which there are very efficient algorithm already developed to solve them [103]. After obtaining the optimum $\lambda^*$, one can find the optimum $x^*$ from Eq. (B.2) applying the Euler optimality condition, i.e. $\partial L / \partial x = 0$.

$$
x^* = -(A - \lambda^* I)^{-1} b
$$
Appendix C
Generalized Eigenvalue Problem

In this appendix, we review some remarks of following minimization problem:

\[
\begin{align*}
\text{minimize} & \quad w^t A w \\
\text{subject to} & \quad w^t B w = 1
\end{align*}
\]  
(C.1)

The solution to this problem is related to the generalized eigenvalue problem for two \(n \times n\) matrices \(A\) and \(B\). The eigenvalues of the pencil \(A - \lambda B\) with \(\lambda \in \mathbb{C}\) are elements of the set \(\lambda(A,B)\) defined by [47][52]:

\[
\lambda(A,B) = \{ z \in \mathbb{C} : \det(A,B) = 0 \}
\]  
(C.2)

For \(\lambda \in \lambda(A,B)\) where \(Ax = \lambda Bx, x \neq 0\), then \(x\) is referred to as an eigenvector of \(A - \lambda B\). It was shown in Chapter 5 that this problem can be reduced to ordinary eigenvalue problem if \(B\) is positive definite. In a general case, a better solution to this problem can be obtained by generalized Schur decomposition theorem [47]. This theorem states that for any \(n \times n\) matrices \(A\) and \(B\), there exists unitary matrices \(Q\) and \(Z\) such that \(Q^H AZ = T\) and \(Q^H B Z = S\) where \(T\) and \(S\) are upper triangular matrices. Moreover, if for some \(k\), \(t_{kk}\) and \(s_{kk}\) are both zero, then the remaining quotients \(t_{ii} / s_{ii} \neq 0\) can assume arbitrary values. Otherwise, \(\lambda(A,B) = \{ t_{ii} / s_{ii} : s_{ii} \neq 0 \}\). Thus, the solution to problem Eq. (C.1) is \(\lambda_{\text{min}}(A,B)\).

Another interesting solution of the problem when \(B\) is not invertible was proposed by Melsa et. al [74] where we state here. The positive semidefinite matrix \(B\) can be written as

\[
B = [U \quad N] \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} [U^t \quad N^t]
\]  
(C.3)
where $\Sigma^2$ is a diagonal matrix of the positive eigenvalues of $B$. The matrix $U$ includes the orthonormal eigenvectors associated with $\Sigma^2$, and the matrix $N$ contains an orthonormal basis of the null space of $B$. The columns of $N$ are orthogonal to the columns of $U$. Assume that $w$ is of length $n$ and that $B$ is of rank $r$ (i.e. $\dim(\Sigma^2) = r$). Every vector $w$ can be written as

$$w = U\Sigma^{-1}y + Nz \quad \text{(C.4)}$$

where $y$ is an arbitrary vector of length $t$ and $z$ is an arbitrary vector of length $n-t$. If $y'y = 1$, then using Eq. (C.3) and Eq. (C.4):

$$wBw = (U\Sigma^{-1}y + Nz)U\Sigma^2U'y(U\Sigma^{-1}y + Nz) = y'y = 1 \quad \text{(C.5)}$$

The problem is now to choose $y$ and $z$ to minimize $J = (y'\Sigma^{-1}U'y + z'N'y)A(U\Sigma^{-1}y + Nz)$ given $y'y = 1$. The minimize is performed first with respect to $z$. $J$ can be rewritten as:

$$y'\Sigma^{-1}U'A\Sigma^{-1}y + 2y'\Sigma^{-1}U'Az + z'N'ANz \quad \text{(C.6)}$$

The optimum $z$ is obtained by setting the derivative of Eq. (C.6) with respect to $z$ equal to zero. The results is: $z = -(N'AN)^{-1}N'A\Sigma^{-1}y$. Substituting this $z$ into $J$, a quadratic form of variable $y$ is obtained as:

$$(y'\Sigma^{-1}U'y - y'\Sigma^{-1}U'y(N'AN)^{-1})A(U\Sigma^{-1}y - N(N'AN)^{-1}N'A\Sigma^{-1}y) \quad \text{(C.7)}$$

that is equal to $y'C'y$ for some appropriately defined matrix $C$. Therefore, the resulting optimum $w$ is $w_{opt} = (I - N(N'AN)^{-1}N'A)U\Sigma^{-1}l_{\min}$ where $l_{\min}$ is the unit-norm eigenvector associated to minimum eigenvalue of matrix $C$. 


Bibliography


69. D. MacKay, Encyclopedia of Sparse Block Codes, online [http://www.inference.phy.cam.ac.uk/mackay/codes/data.html#l51](http://www.inference.phy.cam.ac.uk/mackay/codes/data.html#l51)


Ali Enteshari

Ali Enteshari, was born in 1976 in Kermanshah, Iran. He received B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran-Iran, both in electrical engineering in 1999 and 2001, respectively.

He joined the Center of Information and Communications Technology Research at the Pennsylvania State University in August 2006 pursuing the Ph.D. degree in electrical engineering. He started his research on ultra high-speed data transmission over standard twisted pair copper cables after his first semester at Penn State. This research supported by Nexans Inc. for leading edge wired line communication system has attracted many leading telecommunication industries and academia. His paper on this work won the best paper award at the DesignCon 2009 conference, Santa Clara, CA, which is a highly prestigious conference among the leading industries such as IBM, Intel, Cisco, Infineon and many others. This award goes to papers which make a substantial contribution to the educational mission of DesignCon, reporting on important results, methodologies, or case studies of special significance. He gave a technical talk at 2nd Signal Integrity Symposium in Harrisburg, March 2008, and one presentation at IEEE Higher Speed Study Group, Atlanta, Nov. 2007.

His research interests include broadband data transmission over copper wires, equalization, coding, signal processing and optimization. He is a member of Institute of Mathematical Statistics, review of IEEE Transactions on Communication, and reviewer of IEEE international conf. on communication and MTT-S international microwave symposium.