Multi-Carrier Digital Communications

Theory and Applications of OFDM
Second Edition

Multi-Carrier Digital Communications

Theory and Applications of OFDM
Second Edition

Ahmad R. S. Bahai
National Semiconductor
Stanford University
University of California Berkeley

Burton R. Saltzberg
Consultant on Digital Communications
Formerly Bell Laboratories

Mustafa Ergen
University of California Berkeley

Springer

eBook ISBN: 0-387-22576-5
Print ISBN: 0-387-22575-7

©2004 Springer Science + Business Media, Inc.

Print ©2004 Springer Science + Business Media, Inc.
Boston

All rights reserved

No part of this eBook may be reproduced or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without written consent from the Publisher

Created in the United States of America

Visit Springer's eBookstore at: http://www.ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

Contents

Preface XVII

Acknowledgements XXI

1 Introduction to Digital Communications **1**

1.1 Background 1

1.2 Evolution of OFDM 5

2 System Architecture **15**

2.1 Wireless Channel Fundamentals 15

2.1.1 Path Loss 16

2.1.2 Shadowing 17

2.1.3 Fading Parameters 18

2.1.4 Flat Fading 21

2.1.5 Frequency Selective Fading 22
2.1.6 Fast Fading 23
2.1.7 Slow Fading 23
2.1.8 Rayleigh Fading 23
2.1.9 Ricean Fading 25
2.1.10 Distortions for Wireless Systems 26
2.1.11 Diversity Techniques in a Fading Environment 26
2.2 Digital Communication System Fundamentals 27
2.2.1 Coding 28
2.2.2 Modulation 29
2.3 Multi-Carrier System Fundamentals 33
2.4 DFT 35
2.5 Partial FFT 41
2.6 Cyclic Extension 42
2.7 Channel Estimation 45
2.8 Modelling of OFDM for Time-Varying Random Channel 48
2.8.1 Randomly Time-Varying Channels 48
2.8.2 OFDM in Randomly Time-Varying Channels 51

3 Performance over Time-Invariant Channels 55
3.1 Time-Invariant Non-Flat Channel with Colored Noise 55
3.2 Error Probability 56
3.3 Bit Allocation 59

3.4 Bit and Power Allocation Algorithms for Fixed Bit Rate 66

4 Clipping in Multi-Carrier Systems 69

4.1 Introduction 69

4.2 Power Amplifier Non-Linearity 71

4.3 Error Probability Analysis 73

4.3.1 System Model 77

4.3.2 BER Due to Clipping 78

4.4 Performance in AWGN and Fading 86

4.5 Bandwidth Regrowth 94

5 Synchronization 99

5.1 Timing and Frequency Offset in OFDM 99

5.2 Synchronization & System Architecture 104

5.3 Timing and Frame Synchronization 105

5.4 Frequency Offset Estimation 107

5.5 Phase Noise 108

6 Channel Estimation and Equalization 117

6.1 Introduction 117

6.2 Channel Estimation 117

6.2.1 Coherent Detection 119

6.2.2 Block-Type Pilot Arrangement 122

6.2.3 Comb-Type Pilot Arrangement 126

6.2.4 Interpolation Techniques 127

6.2.5 Non-coherent Detection 130
6.2.6 Performance 130
6.2.7 Channel Estimation for MIMO-OFDM . . . 134
6.3 Equalization 137
6.3.1 Time Domain Equalization 138
6.3.2 Equalization in DMT 142
6.3.3 Delay Parameter 145
6.3.4 AR Approximation of ARMA Model 146
6.3.5 Frequency Domain Equalization 149
6.3.6 Echo Cancellation 152
6.3.7 Appendix - Joint Innovation Representation of ARMA Models 160

7 Channel Coding 167
7.1 Need for Coding 167
7.2 Block Coding in OFDM 168
7.3 Convolutional Encoding 173
7.4 Concatenated Coding 179
7.5 Trellis Coding in OFDM 180
7.6 Turbo Coding in OFDM 185

8 ADSL 189
8.1 Wired Access to High Rate Digital Services 189
8.2 Properties of the Wire-Pair Channel 190

8.3 ADSL Systems 200

9 Wireless LAN Applications 203

9.1 Introduction 203

9.1.1 Background 203

9.1.2 Pros and Cons 204

9.2 Topology 206

9.2.1 Independent BSS 206

9.2.2 Infrastructure BSS 208

9.2.3 Services 211

9.3 Architecture 213

9.4 Medium Access Control 217

9.4.1 DCF Access 221

9.4.2 Markov Model of DCF 227

9.4.3 PCF 233

9.5 Management 239

9.5.1 Synchronization 239

9.5.2 Scanning 240

9.5.3 Power Management 241

9.5.4 Security 242

9.6 IEEE 802.11 and 802.11b Physical Layer 244

9.6.1 Spread Spectrum 245

9.6.2 FHSS Physical Layer 247

9.6.3 DSSS Physical Layer 249

9.6.4 IR Physical Layer 252
9.6.5 HR/DSSS Physical Layer 252
9.6.6 RF Interference 253
9.7 IEEE 802.11a Physical Layer 254
9.7.1 OFDM Architecture 255
9.7.2 Transmitter 256
9.7.3 Receiver 269
9.7.4 Degradation Factors 274
9.8 Performance of 802.11a Transceivers 275
9.8.1 Data Rate 276
9.8.2 Phase Noise 278
9.8.3 Channel Estimation 279
9.8.4 Frequency Offset 280
9.8.5 IQ Imbalance 281
9.8.6 Quantization and Clipping Error 283
9.8.7 Power Amplifier Nonlinearity 285
9.8.8 Hard or Soft Decision Decoding 286
9.8.9 Co-channel Interference 286
9.8.10 Narrowband Interference 288
9.8.11 UWB Interference 288
9.8.12 Performance of 64QAM 289
9.9 Rate Adaptation 291
9.10 Zero IF Technology 292

9.11 IEEE 802. 11e MAC Protocol 293
9.11.1 IEEE 802.11e MAC Services 294
9.11.2 IEEE 802.11e MAC Architecture 295
9.11.3 Hybrid Coordination Function (HCF) 295
9.11.4 HCF Controlled Channel Access 299
9.11.5 Admission Control 300
9.11.6 Block Acknowledgement 301
9.11.7 Multi-rate Support 301
9.11.8 Direct Link Protocol 301
9.12 HIPERLAN/2 302
9.12.1 Protocol Architecture 304
9.12.2 Data Link Control 304
9.12.3 Convergence Layer 310
9.12.4 HIPERLAN/2 vs 802.11a 310
9.13 MMAC-HiSWAN 312
9.14 Overview of IEEE 802.11 Standards 312
9.14.1 IEEE 802.11c - Bridge Operation Procedures 312
9.14.2 IEEE 802.11d - Global Harmonization 313
9.14.3 IEEE 802.11f - Inter Access Point Protocol . 313
9.14.4 IEEE 802.11g - Higher Rate Extensions in the 2.4GHz Band 313
9.14.5 IEEE 802.11h - Spectrum Managed 802.11a . 314
9.14.6 IEEE 802.11i - MAC Enhancements for Enhanced Security 315

9.14.7 IEEE 802.1x - Port Based Network Access Control 316

9.14.8 IEEE 802.1p - QoS on the MAC Level 316

10 Digital Broadcasting 317

10.1 Broadcasting of Digital Audio Signals 317

10.2 Signal Format 320

10.3 Other Digital Broadcasting Systems 323

10.3.1 DAB in the U.S.A 323

10.4 Digital Video Broadcasting 324

11 OFDM based Multiple Access Techniques 327

11.1 Introduction 327

11.2 OFDM-FDMA 329

11.3 OFDM-TDMA 330

11.4 Multi Carrier CDMA (OFDM-CDMA) 331

11.5 OFDMA 336

11.5.1 OFDMA Architecture 337

11.5.2 Resource Allocation Regarding QoS 341

11.5.3 Resource Allocation Regarding Capacity . . 342

11.6 Flash-OFDM 343

11.7 OFDM-SDMA 346

12 Ultra WideBand Technologies 349

12.1 Impulse Radio 350

12.2 Multiband Approach 352
12.3 Multiband OFDM 352

13 IEEE 802.16 and WiMAX 357
13.1 Introduction 357
13.2 WiMAX 358
13.3 WirelessMAN OFDM 359
13.4 802.16 MAC 361
13.5 Conclusion 362

14 Future Trends 363
14.1 Comparison with Single Carrier Modulation 363
14.2 Mitigation of Clipping Effects 365
14.3 Overlapped Transforms 366
14.4 Advances in Implementation 370

Bibliography 373

List of Figures 393

List of Tables 405

Index 407

We dedicate this book
to our families...

Preface

Multi-carrier modulation, Orthogonal Frequency Division Multiplexing (OFDM) particularly, has been successfully applied to a wide variety of digital communications applications over the past several years. Although OFDM has been chosen as the physical layer standard for a diversity of important systems, the theory, algorithms, and implementation techniques remain subjects of current interest. This is clear from the high volume of papers appearing in technical journals and conferences.

Multi-carrier modulation continues to evolve rapidly. It is hoped that this book will remain a valuable summary of the technology, providing an understanding of new advances as well as the present core technology.

The Intended Audience

This book is intended to be a concise summary of the present state of the art of the theory and practice of OFDM technology. The authors believe that the time is ripe for such a treatment. Particularly based on one of the author's long experience in development of wireless systems (AB), and the other's in wireline systems (BS), we have attempted to present a unified presentation of OFDM performance and

implementation over a wide variety of channels.

It is hoped that this will prove valuable both to developers of such systems and to researchers and graduate students involved in analysis of digital communications.

In the interest of brevity, we have minimized treatment of more general communication issues. There exist many excellent texts on communication theory and technology. Only brief summaries of topics not specific to multi-carrier modulation are presented in this book where essential. As a background, we presume that the reader has a clear knowledge of basic fundamentals of digital communications.

Highlights of the Second Edition

During the past few years since the publication of the first edition of this text, the technology and application of OFDM has continued their rapid pace of advancement. As a result, it became clear to us that a new edition of the text would be highly desirable. The new edition provides an opportunity to make those corrections and clarifications whose need became apparent from continued discussions with many readers. However, the main purpose is to introduce new topics that have come to the forefront during the past few years, and to amplify the treatment of other subject matter.

Because of the particularly rapid development of wireless systems employing OFDM, we have introduced a section early in the text on wireless channel fundamentals. We have extended and modified our analysis of the effects of clipping, including simulation results that have been reported in a recent publication. These new results are restated here. A section on channel estimation has been added to the chapter on equalization. The chapter on local area networks has been greatly expanded to include the latest technology and applications. Three totally new chapters are added, on OFDM multiple access tech-

nology, on ultra wideband technology and on WiMAX (IEEE 802.16).

Organization of This Book

We begin with a historical overview of multi-carrier communications, wherein its advantages for transmission over highly dispersive channels have long been recognized, particularly before the development of equalization techniques. We then focus on the bandwidth efficient technology of OFDM, in particular the digital signal processing techniques that have made the modulation format practical. Several chapters describe and analyze the sub-systems of an OFDM implementation, such as clipping, synchronization, channel estimation, equalization, and coding. Analysis of performance over channels with various impairments is presented.

The book continues with descriptions of three very important and diverse applications of OFDM that have been standardized and are now being deployed. ADSL provides access to digital services at several Mbps over the ordinary wire-pair connection between customers and the local telephone company central office. Digital Broadcasting enables the radio reception of high quality digitized sound and video. A unique configuration that is enabled by OFDM is the simultaneous transmission of identical signals by geographically dispersed transmitters. And, the new development of wireless LANs for multi-Mbps communications is presented in detail. Each of these successful applications required the development of new fundamental technology.

Finally, the book concludes with describing the OFDM based multiple access techniques, ultra wideband technology and WiMAX.

Acknowledgements

The two authors of the first edition of this text are very pleased to include our colleague as an additional author, and gratefully acknowledge his extensive contributions in making this second edition possible.

The first edition of this text has been used for classes in University of California Berkeley, Stanford University, University of Cambridge, CEI-Europe and in other institutions. We are grateful to colleagues in the institutions where this book has been used. For the first edition, we acknowledge the extensive review and many valuable suggestions of Professor Kenji Kohiyama, our former colleagues at AT&T Bell Laboratories and colleagues at Algorex. Gail Bryson performed the very difficult task of editing and assembling this text. The continuing support of Kambiz Homayounfar was essential to its completion.

In preparing the second edition, we acknowledge Professor Pravin Varaiya for his valuable help, Dr. Haiyun Tangh for providing some graphs, National Semiconductor and IMEC for providing the MATLAB simulation tool, Sinem Coleri for proof-reading. We wish to express our appreciation to Fran Wilkinson for editing.

Last, but by no means least, we are thankful to our families for their support and patience.

Despite all our efforts to keep the text error free, for any that remain, any comments, corrections and suggestions received will be much appreciated for the future printings. We can be reached via e-mail at *bahai@stanford.edu, bsaltzberg@worldnet.att.net,* and *ergen@eecs.berkeley.edu.* We will post any corrections and comments at the Web site ***http://ofdm.eecs.berkeley.edu/*** in addition to the support materials that may be necessary to prepare a lecture or paper.

Chapter 1

Introduction to Digital Communications

1.1 Background

The physical layer of digital communications includes mapping of digital input information into a waveform for transmission over a communication channel and mapping of the received waveform into digital information that hopefully agrees with the original input since the communication channel may introduce various forms of distortion as well as noise [3].

The simplest form of such communication, as least conceptually, is Pulse Amplitude Modulation (PAM), shown in Figure 1.1. Here the transmitted waveform is of the form

$$s(t) = \sum_n a_n g(t - nT) \tag{1.1}$$

where the information to be transmitted is given by the sequence of a_ns, $1/T$ is the symbol rate, and $g(t)$ is the impulse response of the

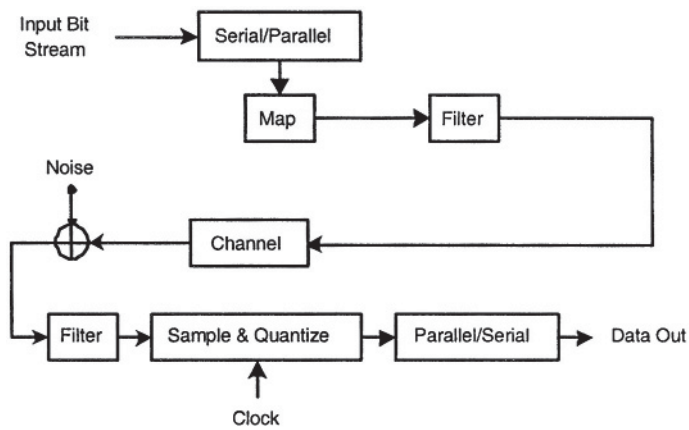

Figure 1.1: A basic PAM system

transmit filter, usually low-pass. The a_ns are chosen from an alphabet of size L, so the bit rate is $\frac{1}{T}log_2 L$. It is desirable that the alphabet be both zero mean and equally spaced. The values of a_n can be written as

$$\{-A(L-1), \ldots, -A, A, \ldots, A(L-1)\}. \tag{1.2}$$

Assuming the a_ns are equiprobable, the transmitted power is

$$A^2 \frac{L^2-1}{3T} \int_{-\infty}^{\infty} g^2(t)dt. \tag{1.3}$$

At the receiver, the signal is filtered by $r(t)$, which may include an adaptive equalizer (sampled), and the nearest permitted member of the alphabet is output. In order to avoid inter-symbol interference, it is desirable that $x(t) = g(t) * h(t) * r(t) = 0$ for all $t = kT$, k an integer $\neq 0$, where $h(t)$ is the channel impulse response. This is the Nyquist criterion, which is given in the frequency domain by:

$$\sum_m X(f + \frac{m}{T}) = const \tag{1.4}$$

The minimum bandwidth required is $1/2T$. This is met by a frequency response that is constant for $-1/2T < f < 1/2T$, whose corresponding time response is

$$x(t) = \frac{sin(\pi t/T)}{(\pi t/T)} \tag{1.5}$$

Some excess bandwidth, denoted by the roll-off factor, is desirable in order for the time response to decay more quickly. Note that $r(t)$ is not a matched filter, because it must satisfy the inter-symbol interference constraint.

If $r(t)$ has gain such that the alphabet levels of $x(0)$ are also spaced by $2A$, then errors will occur when the noise at the sampler satisfies $|n| > A$ for interior levels, or $n > A$ or $n < -A$ for the outer levels. If the noise is Gaussian with power spectral density $N(f)$ at the receiver input, then the noise variance is:

$$\sigma^2 = \int_{-\infty}^{\infty} N(f)|R(f)|^2 df \tag{1.6}$$

and the error probability per symbol is

$$P_e = \frac{2(L-1)}{L} Q(\frac{A}{\sigma}), \tag{1.7}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2} dy \tag{1.8}$$

is the normal error integral.

PAM is only suitable over channels that exist down to, but might not necessarily include, zero frequency. If zero frequency is absent, a modulation scheme that puts the signal spectrum in the desired frequency band is required. Of particular interest, both in its own right

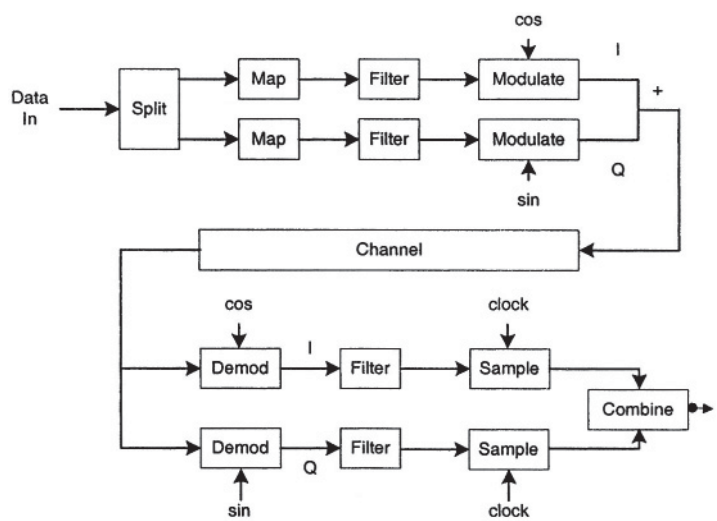

Figure 1.2: A basic QAM system

and as a component of OFDM, is Quadrature Amplitude Modulation (QAM). The simplest form of QAM, shown in Figure 1.2, may be thought of as two PAM signals, modulated by carriers at the same frequency but 90 degrees out of phase. At the receiver, demodulation by the same carriers separates the signal components. Unlike some other modulation schemes, such as FM, QAM is bandwidth efficient in that it requires the same bandwidth as a PAM signal of the same bit rate. Furthermore, the performance of QAM in noise is comparable to that of PAM. The QAM line signal is of the form

$$\sum_n a_n g(t - nT) cos\omega t - \sum_n b_n g(t - nT) sin\omega t. \qquad (1.9)$$

This line signal may also be written in the form of:

$$Re\{\sum_n c_n g(t - nT) e^{j\omega t}\}, \qquad (1.10)$$

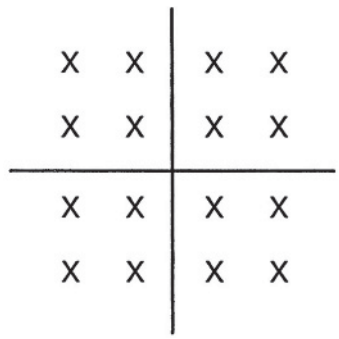

Figure 1.3: A QAM constellation

where the pair of real symbols a_n and b_n are treated as a complex symbol $c_n = a_n + jb_n$. The required bandwidth for transmitting such complex symbols is $1/T$. The complex symbol values are shown as a "constellation" in the complex plane. Figure 1.3 shows the constellation of a 16-point QAM signal, which is formed from 4-point PAM.

It is not necessary that the constellation be square. Figure 1.4 shows how input information can be mapped arbitrarily into constellation points. A constellation with a more circular boundary provides better noise performance. By grouping n successive complex symbols as a unit, we can treat such units as symbols in $2n$-dimensional space. In this case, Figure 1.4 can be extended to include a large enough serial-to-parallel converter that accommodates the total number of bits in n symbols, and a look-up table with $2n$ outputs.

1.2 Evolution of OFDM

The use of Frequency Division Multiplexing (FDM) goes back over a century, where more than one low rate signal, such as telegraph, was carried over a relatively wide bandwidth channel using a separate car-

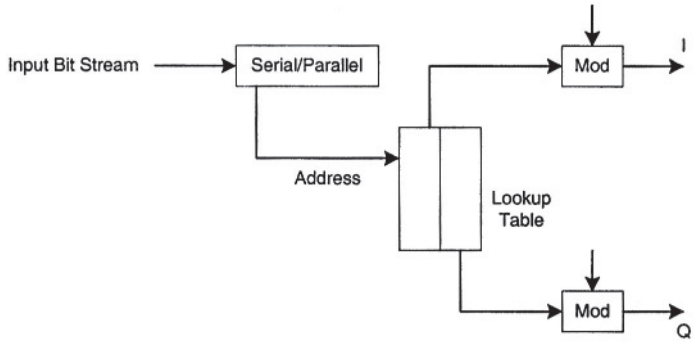

Figure 1.4: General form of QAM generation

rier frequency for each signal. To facilitate separation of the signals at the receiver, the carrier frequencies were spaced sufficiently far apart so that the signal spectra did not overlap. Empty spectral regions between the signals assured that they could be separated with readily realizable filters. The resulting spectral efficiency was therefore quite low.

Instead of carrying separate messages, the different frequency carriers can carry different bits of a single higher rate message. The source may be in such a parallel format, or a serial source can be presented to a serial-to-parallel converter whose output is fed to the multiple carriers.

Such a parallel transmission scheme can be compared with a single higher rate serial scheme using the same channel. The parallel system, if built straightforwardly as several transmitters and receivers, will certainly be more costly to implement. Each of the parallel subchannels can carry a low signalling rate, proportional to its bandwidth. The sum of these signalling rates is less than can be carried by a single serial channel of that combined bandwidth because of the unused guard space between the parallel sub-carriers. On the other hand, the

single channel will be far more susceptible to inter-symbol interference. This is because of the short duration of its signal elements and the higher distortion produced by its wider frequency band, as compared with the long duration signal elements and narrow bandwidth in sub-channels in the parallel system.

Before the development of equalization, the parallel technique was the preferred means of achieving high rates over a dispersive channel, in spite of its high cost and relative bandwidth inefficiency. An added benefit of the parallel technique is reduced susceptibility to most forms of impulse noise.

The first solution of the bandwidth efficiency problem of multi-tone transmission (not the complexity problem) was probably the *Kineplex* system. The Kineplex system was developed by Collins Radio Co. [4] for data transmission over an H.F. radio channel subject to severe multi-path fading. In that system, each of 20 tones is modulated by differential 4-PSK without filtering. The spectra are therefore of the $sin(kf)/f$ shape and strongly overlap. However, similar to modern OFDM, the tones are spaced at frequency intervals almost equal to the signalling rate and are capable of separation at the receiver.

The reception technique is shown in Figure 1.5. Each tone is detected by a pair of tuned circuits. Alternate symbols are gated to one of the tuned circuits, whose signal is held for the duration of the next symbol. The signals in the two tuned circuits are then processed to determine their phase difference, and therefore the transmitted information. The older of the two signals is then quenched to allow input of the next symbol. The key to the success of the technique is that the time response of each tuned circuit to all tones, other than the one to which it is tuned, goes through zero at the end of the gating interval, at which point that interval is equal to the reciprocal of the frequency separation between tones. The gating time is made somewhat shorter than the symbol period to reduce inter-symbol interference, but efficiency of 70% of the Nyquist rate is achieved. High performance over actual long H.F. channels was obtained, although at a high im-

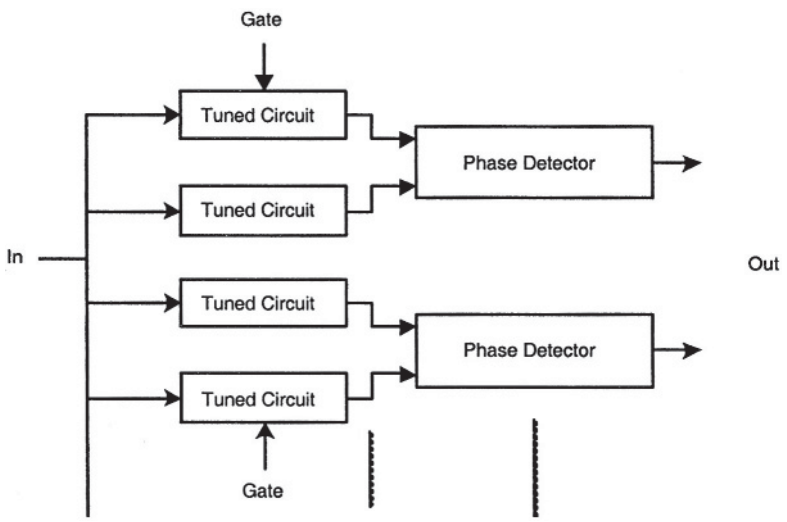

Figure 1.5: The Collins Kineplex receiver

plementation cost. Although fully transistorized, the system required two large bays of equipment.

A subsequent multi-tone system [5] was proposed using 9-point QAM constellations on each carrier, with correlation detection employed in the receiver. Carrier spacing equal to the symbol rate provides optimum spectral efficiency. Simple coding in the frequency domain is another feature of this scheme. The above techniques do provide the orthogonality needed to separate multi-tone signals spaced by the symbol rate. However the $sin(kf)/f$ spectrum of each component has some undesirable properties. Mutual overlap of a large number of sub-channel spectra is pronounced. Also, spectrum for the entire system must allow space above and below the extreme tone frequencies to accommodate the slow decay of the sub-channel spectra. For these reasons, it is desirable for each of the signal components to be bandlimited so as to overlap only the immediately adjacent sub-carriers, while remaining orthogonal to them. Criteria for meeting this objec-

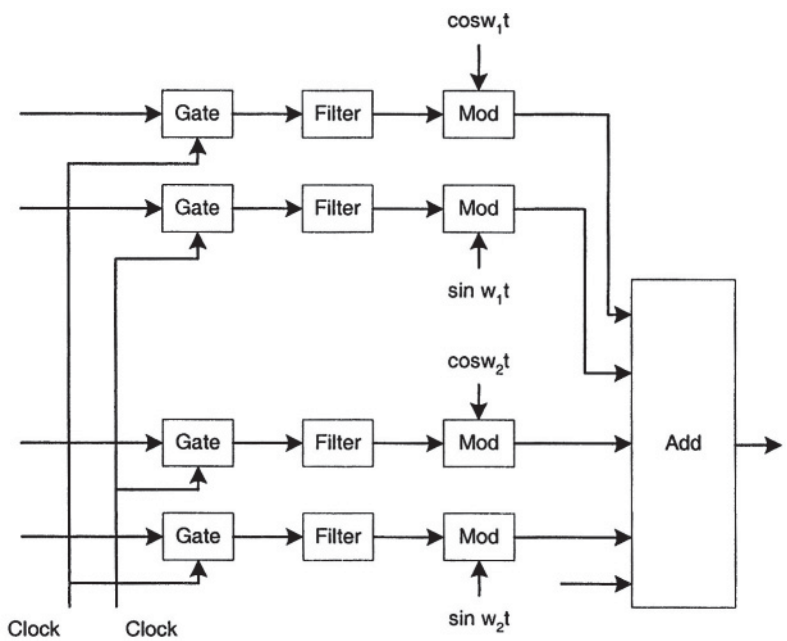

Figure 1.6: An early version of OFDM

tive is given in References [6] and [7].

In Reference [8] it was shown how bandlimited QAM can be employed in a multi-tone system with orthogonality and minimum carrier spacing (illustrated in Figure 1.6). Unlike the non-bandlimited OFDM, each carrier must carry Staggered (or Offset) QAM, that is, the input to the I and Q modulators must be offset by half a symbol period. Furthermore, adjacent carriers must be offset oppositely. It is interesting to note that Staggered QAM is identical to Vestigial Sideband (VSB) modulation. The low-pass filters $g(t)$ are such that $G^2(f)$, the combination of transmit and receive filters, is Nyquist, with the roll-off factor assumed to be less than 1.

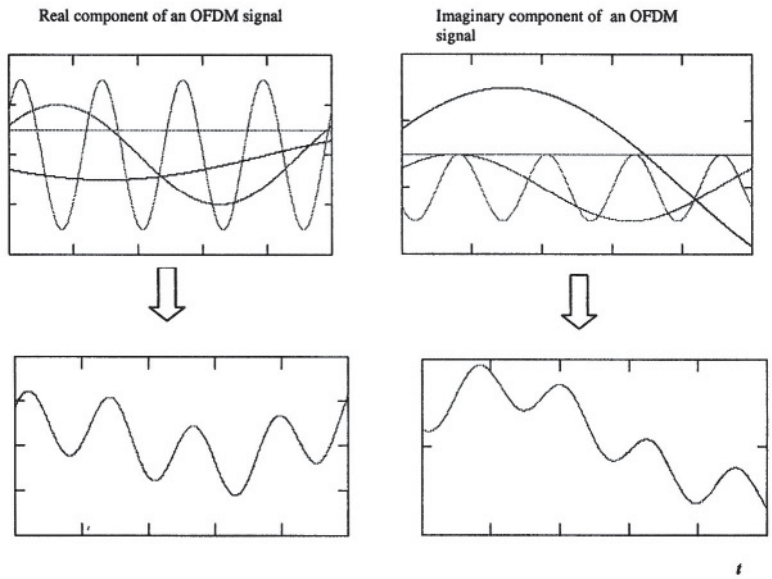

Figure 1.7: OFDM modulation concept: Real and Imaginary components of an OFDM symbol is the superposition of several harmonics modulated by data symbols

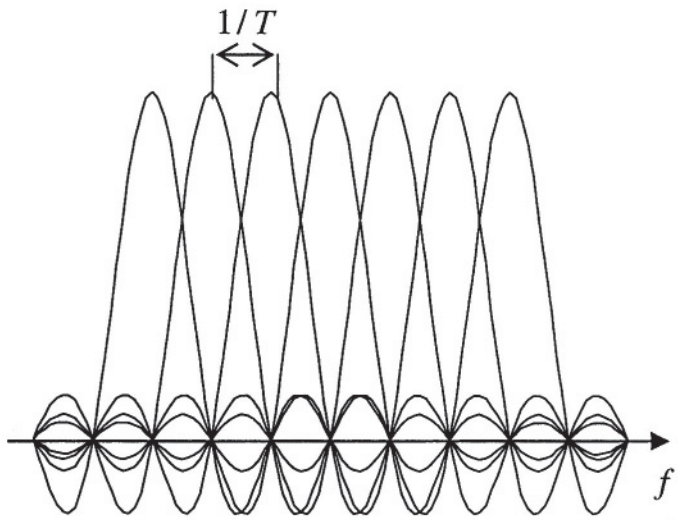

Figure 1.8: Spectrum overlap in OFDM

The major contribution to the OFDM complexity problem was the application of the Fast Fourier Transform (FFT) to the modulation and demodulation processes [9]. Fortunately, this occurred at the same time digital signal processing techniques were being introduced into the design of modems. The technique involved assembling the input information into blocks of N complex numbers, one for each sub-channel. An inverse FFT is performed on each block, and the resultant transmitted serially. At the receiver, the information is recovered by performing an FFT on the received block of signal samples. This form of OFDM is often referred to as Discrete Multi-Tone (DMT). The spectrum of the signal on the line is identical to that of N separate QAM signals, at N frequencies separated by the signalling rate. Each such QAM signal carries one of the original input complex numbers. The spectrum of each QAM signal is of the form $sin(kf)/f$, with nulls at the center of the other sub-carriers, as in the earlier OFDM systems, and as shown in Figure 1.8 and Figure 1.9.

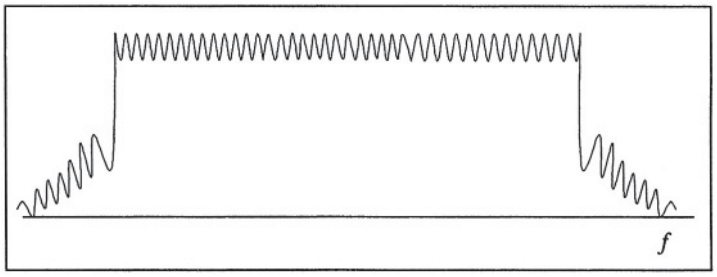

Figure 1.9: Spectrum of OFDM signal

A block diagram of a very basic DMT system is shown Figure 1.10. Several critical blocks are not shown. As described more thoroughly in Chapter 2, care must be taken to avoid overlap of consecutive transmitted blocks, a problem that is solved by the use of a cyclic prefix. Another issue is how to transmit the sequence of complex numbers from the output of the inverse FFT over the channel. The process is straightforward if the signal is to be further modulated by a modulator with I and Q inputs.

Otherwise, it is necessary to transmit real quantities. This can be accomplished by first appending the complex conjugate to the original input block. A $2N$-point inverse FFT now yields $2N$ real numbers to be transmitted per block, which is equivalent to N complex numbers.

The most significant advantage of this DMT approach is the efficiency of the FFT algorithm. An N-point FFT requires only on the order of $NlogN$ multiplications, rather than N^2 as in a straightforward computation. The efficiency is particularly good when N is a power of 2, although that is not generally necessary. Because of the use of the FFT, a DMT system typically requires fewer computations per unit time than an equivalent single channel system with equalization. An overall cost comparison between the two systems is not as clear, but the costs should be approximately equal in most cases. It should be noted that the bandlimited system of Figure 1.6 can also

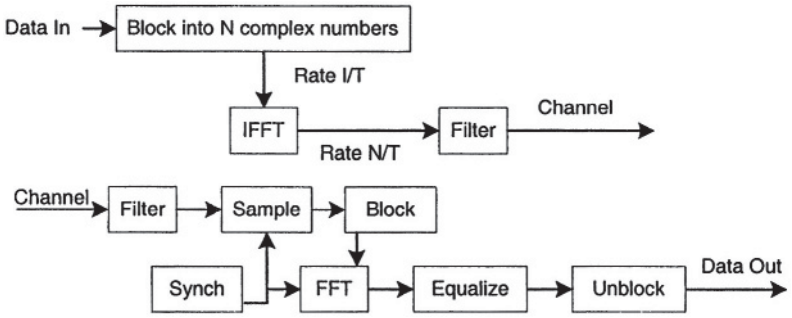

Figure 1.10: Very basic OFDM system

be implemented with FFT techniques [10], although the complexity and delay will be greater than DMT. Over the last 20 years or so, OFDM techniques and, in particular, the DMT implementation, has been used in a wide variety of applications [64]. Several OFDM voice-band modems have been introduced, but did not succeed commercially because they were not adopted by standards bodies. DMT has been adopted as the standard for the Asymmetric Digital Subscriber Line (ADSL), which provides digital communication at several Mbps from a telephone company central office to a subscriber, and a lower rate in the reverse direction, over a normal twisted pair of wires in the loop plant. OFDM has been particularly successful in numerous wireless applications, where its superior performance in multi-path environments is desirable. Wireless receivers detect signals distorted by time and frequency selective fading. OFDM in conjunction with proper coding and interleaving is a powerful technique for combating the wireless channel impairments that a typical OFDM wireless system might face, as is shown in Figure 1.11. A particularly interesting configuration, discussed in Chapter 10, is the Single Frequency Network (SFN) used for broadcasting of digital audio or video signals. Here many geographically separated transmitters broadcast identical and synchronized signals to cover a large region. The reception of such signals by a receiver is equivalent to an extreme form of multi-

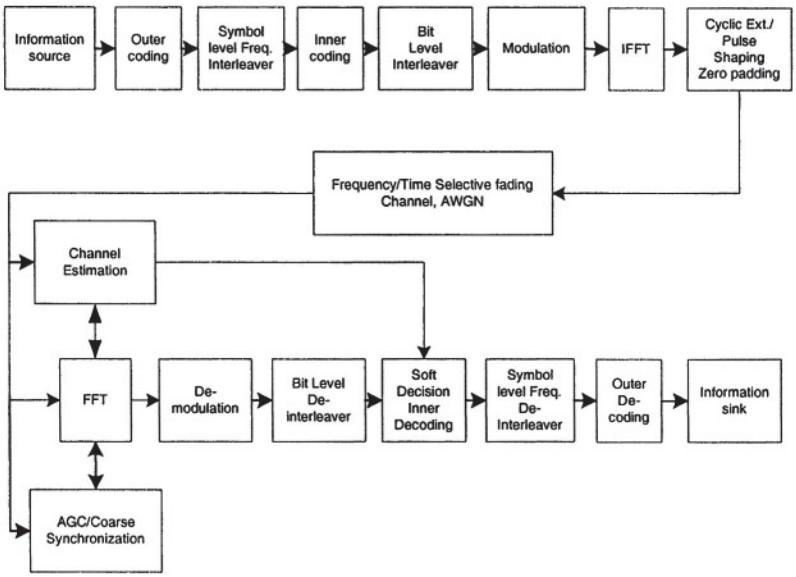

Figure 1.11: A typical wireless OFDM architecture

path. OFDM is the technology that makes this configuration viable.

Another wireless application of OFDM is in high speed local area networks (LANs). Although the absolute delay spread in this environment is low, if very high data rates, in the order of many tens of Mbps, is desired, then the delay spread may be large compared to a symbol interval. OFDM is preferable to the use of long equalizers in this application. It is expected that OFDM will be applied to many more new communications systems over the next several years.

Chapter 2

System Architecture

This chapter presents a general overview of system design for multi-carrier modulation. First, a review of wireless channels are discussed. Second, fundamentals of digital communication is cited and finally, the major system blocks of OFDM is analyzed.

2.1 Wireless Channel Fundamentals

Wireless transmission uses air or space for its transmission medium. The radio propagation is not as smooth as in wire transmission since the received signal is not only coming directly from the transmitter, but the combination of reflected, diffracted, and scattered copies of the transmitted signal. It is interesting and rewarding to examine the effects of propagation to a radio signal since consequences determine data rate, range, and reliability of the wireless system.

Reflection occurs when the signal hits a surface where partial energy is reflected and the remaining is transmitted into the surface. Reflection coefficient, the coefficient that determines the ratio of reflec-

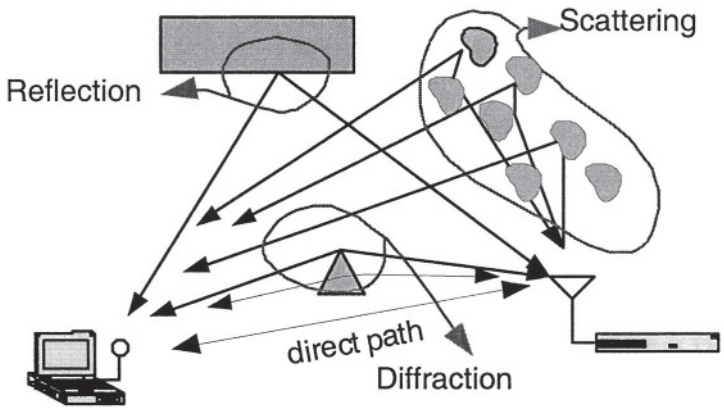

Figure 2.1: Wireless propagation

tion and transmission, depends on the material properties. Diffraction occurs when the signal is obstructed by a sharp object which derives secondary waves. Scattering occurs when the signal impinges upon rough surfaces, or small objects. Received signal is sometimes stronger than the reflected and diffracted signal since scattering spreads out the energy in all directions and consequently provides additional energy for the receiver which can receive more than one copies of the signal in multiple paths with different phases and powers. Figure 2.1 illustrates the propagation mechanisms [2].

Models play an important role in the designs of wireless communication systems. There are key modelling parameters that help to characterize a wireless channel: path loss, delay spread, coherence bandwidth, doppler spread, and coherence time.

2.1.1 Path Loss

The average received power diminishes with the distance. In free space when there is a direct path between transmitter and receiver and when

there are no secondary waves from the medium objects, the received power is inversely proportional to square of the carrier frequency and square of the distance. If P_R and P_T are the received and transmitted powers respectively, the following relation describes free space propagation [1],

$$P_R \propto \frac{G * P_T}{f^2 * d^\alpha}, \tag{2.1}$$

where f is the carrier frequency, d is the distance between transmitter and receiver, G is the power gain from the transmit and receive antennas, $\alpha = 2$ is the path loss component and *path loss* is defined as P_T/P_R. One can infer from the equation that the larger the carrier frequency the smaller the operating range.

When there is clutter in the medium, it is not easy to determine the received signal since now the direct path component is added with the multipath components at the receiver. Adopted method is to use a model that is derived from analytical and empirical methods. A good model is a model that obeys the theoretical findings and test results.

2.1.2 Shadowing

Signal power attenuates randomly with distance when the medium includes obstructions. Equation 2.1 is not enough to describe the real environment since Equation 2.1 gives the same path loss to the different receivers having the equal distance apart from the transmitter. However, if the two locations have different surroundings then the variations in the signal should be different. This behavior is called *shadowing*. Measurements indicate that the path loss is random and distributed log-normally. Modified version of Equation 2.1 for indoor path loss is called log normal shadowing and follows the formula,

$$PL(dB) = PL(d_o) + 10\alpha \log_{10}(d/d_o) + X_\sigma, \tag{2.2}$$

where d_o is the reference distance and X_σ is a zero mean Gaussian random variable with a standard deviation σ. σ and α are determined

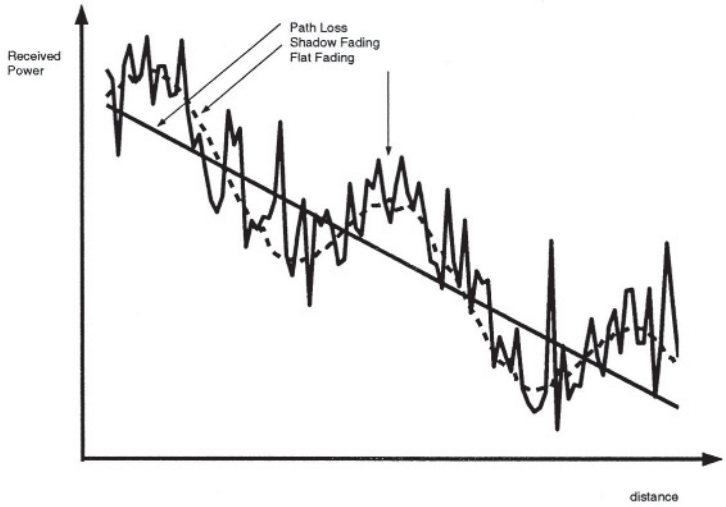

Figure 2.2: Received signal versus distance © [1]

from measured data.

A typical situation is depicted in Figure 2.2. The figure is useful to understand the difference between path loss, shadowing and flat fading, which is described in the following section.

2.1.3 Fading Parameters

One of the most intriguing aspect in wireless communication is *fading* which is present when there are multipath components. Multipath components arrive at the receiver at slightly different times. If there is movement in the system then there is also phase difference between the received components which leads to shift in the frequency. These multiple received copies apply constructive or destructive interference to the signal and create a standing wave pattern that shows rapid signal strength changes, frequency shifts or echoes.

Multipath propagation, movement and bandwidth are the factors that influence the fading. Multipath delay nature of the channel is quantified by *delay spread* and *coherence bandwidth.* The time-varying nature of the channel caused by movement is quantified by *doppler spread* and *coherence time.*

√ **Delay Spread:**

The reflected parts arrive later than the original signal to the receiver. RMS delay spread (σ_τ) is the metric to characterize this delay in terms of second order moment of the channel power profile. Typical values are on the order of microseconds in outdoor and on the order of nanoseconds in indoor radio channels [2]. Depending on the symbol duration (T_s), σ_τ plays an important role to find out how the signal is treated by the channel.

√ **Coherence Bandwidth:**

The channel power profile is coupled with channel frequency response through Fourier transform. Coherence bandwidth (B_c) is used to quantify the channel frequency response and it is inversely proportional to RMS delay spread. Coherence bandwidth is used to measure how flat the channel bandwidth is. Flatness is described as the close correlation between the two frequency components. This is important since a signal having a larger bandwidth (B_s) than B_c is severely distorted. For a 0.9 correlation $B_c \approx 1/50\sigma_\tau$ [2].

√ **Doppler Shift:**

Movement causes shift in the signal frequency. When the stations are moving, the received signal frequency is different than the original signal frequency. Doppler shift is defined as the change in the frequency.

Suppose a mobile is moving at velocity ν as in Figure 2.3. The difference in path length of received signals is $l = d\cos\theta$ where

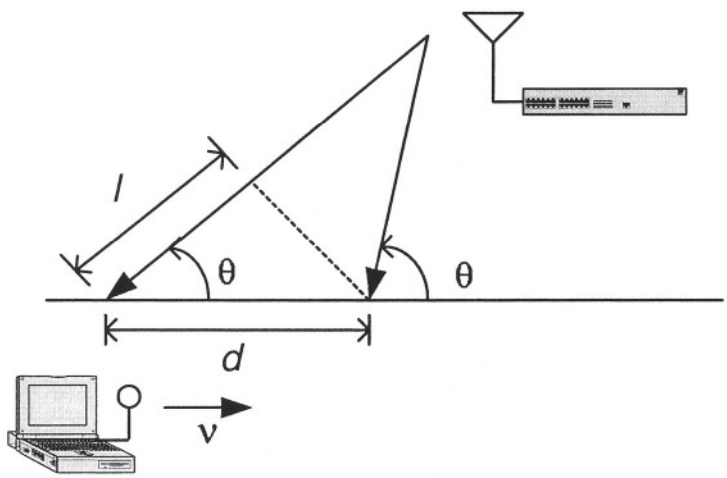

Figure 2.3: Doppler effect

$d = \nu\Delta t$. The phase change is

$$\Delta\phi = \frac{2\pi l}{\lambda} \tag{2.3}$$

and Doppler shift is given by

$$f_d = \frac{1}{2\pi}.\frac{\Delta\phi}{\Delta t} = \frac{\nu}{\lambda}.\cos\theta \tag{2.4}$$

From the Equation 2.4, one can see that if the movement is toward the signal generator, the doppler shift is positive otherwise it is negative [2].

√ **Doppler Spread:**

A channel shows a time varying nature when there is a movement in either the source or destination or even objects in the middle. Doppler spread (B_D) is the measure of maximum broadening of the spectrum due to Doppler shift. Thereby B_D is $f_m = \nu/\lambda$ where f_m is the maximum Doppler shift.

√ **Coherence Time:**

The time domain dual of Doppler spread is coherence time (T_c) where $T_c \approx 1/f_m$. Coherence time identifies the time period wherein two received signal have high amplitude correlation.

√ **Coherence Distance:**

Coherence distance gives a measure of a minimum distance between points in space for which the signal are uncorrelated. This distance is important for multiple antenna systems. It is 0.5 wavelength for wide beamwidth receive antennas and about 10 and 20 wavelengths for low-medium and high BTS antenna heights, respectively [141].

2.1.4 Flat Fading

Delay spread and coherence bandwidth are independent from Doppler spread and coherence time. The signal goes to *flat fading* if the channel bandwidth is greater than the signal bandwidth ($B_c > B_s$). Spectral shape of the signal remains but the gain changes. Narrow band signals fall into this category since their bandwidth is small as compared to the channel bandwidth. Flat fading channels bring challenges such as variation in the gain and in the frequency spectrum. Distortion in the gain may cause deep fades thereby requiring significant increase in the power in some frequencies. Destructive interferences may cause deep nulls in the signal power spectrum which is a particular problem for narrow band signals since any null in a frequency may cause lost of the signal.

There are various ways to reduce the fading distortion [1]. Diversity is one way which leverages multiple independent channels. Since the channels are independent they have lower probability to experience fades at the same time. Space diversity is one of the most efficient diversity technique compared to time or frequency. Multiple slightly

separated antennas in space are used to create independent paths to achieve uncorrelated channels.

Coding is another way which spreads the signal power among the bandwidth and lessens the power allocated in the null frequencies. As a result, loss occurs in the smaller portion of the signal power.

Another method is chopping the bandwidth into small bandwidths as in orthogonal frequency division multiplexing (OFDM). In each subcarrier small portion of the code is sent. Thus, if there is a loss in a subcarrier then it is recoverable since it corresponds to the small part of the code.

2.1.5 Frequency Selective Fading

If channel bandwidth is smaller than the signal bandwidth then the channel creates frequency selective fading. Spectral response of a radio signal will show dips due to the multipath. Channel bandwidth limits the signal bandwidth which leads to overlapping of the successive symbols in the time domain due to the convolution ($\sigma_\tau > T_s$). Overall symbol duration becomes more than the actual symbol duration. This phenomenon is called *intersymbol interference* (ISI).

There are several ways to combat ISI [1]. Equalization is one way that tries to invert the effects of the channel. In OFDM, each subcarrier undergoes flat fading since its bandwidth is less than the coherence bandwidth. Therefore ISI is eliminated within an OFDM symbol but OFDM symbols may overlap in time. A portion of the OFDM symbol, larger than the coherence bandwidth, is appended to the symbol to overcome the ISI between symbols. Equalization is simple in OFDM since each subcarrier only needs a one tap equalizer.

2.1.6 Fast Fading

Channel bandwidth may also change due to movement. Fast fading occurs if the coherence time is smaller than the symbol duration of the signal ($T_s > T_c$). The channel becomes time varying and within a symbol duration rapid changes occur in impulse response of the channel. Low data rates is vulnerable to fast fading since the symbol duration is wide in low data rates.

2.1.7 Slow Fading

If the signal bandwidth is much higher than the Doppler spread ($B_c > B_d$) then the shift in the frequency due to Doppler shift is insignificant. This is called slow fading and the channel impulse response changes slowly.

Figure 2.4 shows a table that summarizes the relationship between fading parameters and signal. As it can be seen from the figure there are two independent effects that classify the wireless channel: multipath and movement which are quantified by Delay spread and Doppler spread respectively.

2.1.8 Rayleigh Fading

Constructive and destructive nature of multipath components in flat fading channels can be approximated by Rayleigh distribution if there is no line of sight which means when there is no direct path between transmitter and receiver. The Rayleigh distribution is basically the magnitude of the sum of two equal independent orthogonal Gaussian random variables and the probability density function (pdf) is given by

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \; 0 \leq r \leq \infty \tag{2.5}$$

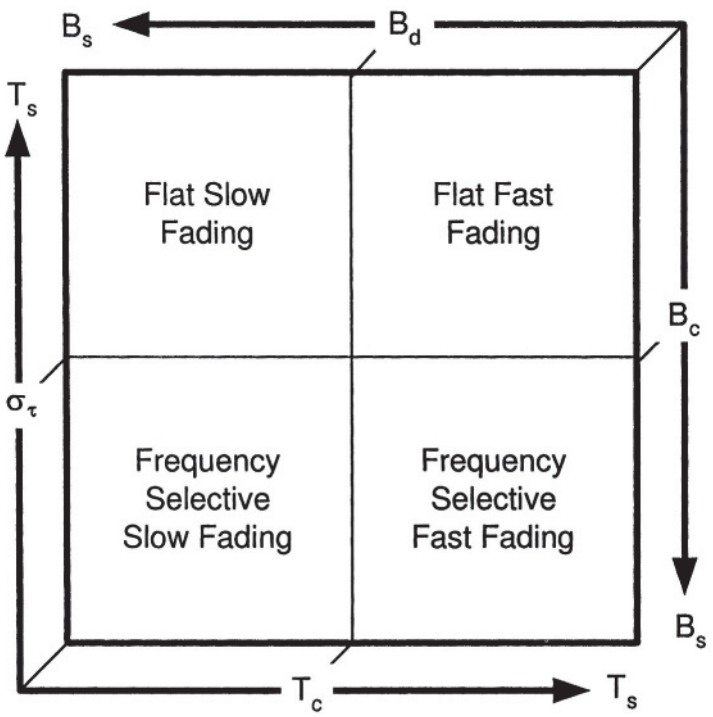

Figure 2.4: Fading illustration © [2]

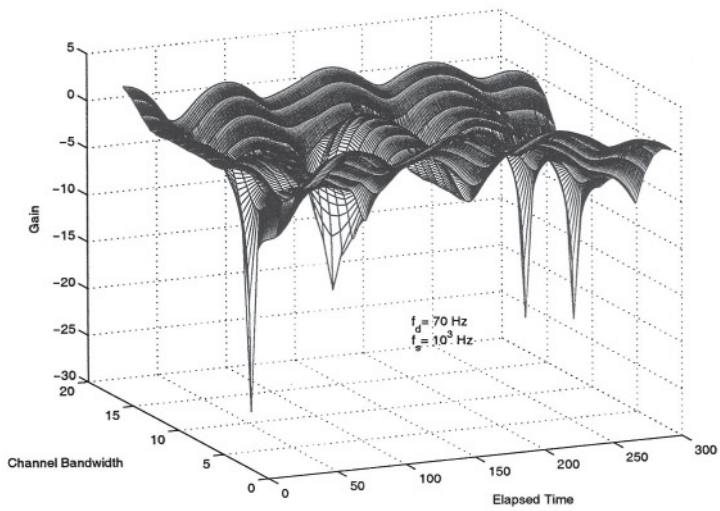

Figure 2.5: A flat fading channel where f_s is sampling frequency

where σ^2 is the time-average power of the received signal [2].

Figure 2.5 depicts a flat fading channel where the amplitude of each multipath component is created by a Rayleigh distribution.

2.1.9 Ricean Fading

When there is line of sight, direct path is normally the strongest component goes into deeper fade compared to the multipath components. This kind of signal is approximated by Ricean distribution. As the dominating component run into more fade the signal characteristic goes from Ricean to Rayleigh distribution.

The Ricean distribution is given by

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{(r^2+A^2)}{2\sigma^2}} I_o(\frac{Ar}{\sigma^2}) \; for(A \geq 0, r \geq 0) \tag{2.6}$$

where A denotes the peak amplitude of the dominant signal and $I_o(.)$ is the modified Bessel function of the first kind and zero-order [2].

2.1.10 Distortions for Wireless Systems

Wireless systems are heavily dependent on the wireless environment. Radio receivers should be able to combat the detrimental effects of the multipath channel. Typical distortions that wireless systems face are listed below.

- AWGN (receiver thermal noise)
- Receiver carrier frequency and phase offset
- Receiver timing offset
- Delay spread
- Fading
- Co-channel and adjacent channel interference
- Non-linear distortion, impulse noise

2.1.11 Diversity Techniques in a Fading Environment

Diversity techniques are important to cope with fading environment as well as coding. It tries to provide the receiver with multiple fade replicas of the same information bearing signal. If there are L independent replicas and if the p denote the outage probability then overall probability is p^L which is less than p. Diversity gain is at the highest when the diversity branches are uncorrelated. Some of the techniques are listed below and will be explained in detail in the subsequent chapters.

- √ **Frequency Diversity:** Multiple channels are separated by more than the coherence bandwidth. Using RAKE receiver, OFDM, equalization are possible techniques for frequency diversity. It can not be used over frequency-flat channel.
- √ **Time Diversity:** Channel coding, interleaving are possible techniques but it is not effective over slow fading channels.
- √ **Space Diversity:** Using multiple antennas for space diversity is efficient, coherent distance plays an important role for the selection of places of the antennas.
- √ **Angle/Direction Diversity:** Antennas with different polarizations/directions can be used for reception and transmission.

Notes

Wireless channels are studied in many books and articles, for a brief tutorial see [1] and detailed information can be found in [2].

2.2 Digital Communication System Fundamentals

The digital communication systems are composed of a transmitter, a receiver and the channel. The transmitter accepts input bits. If the initial source is analog, it is inverted to bit stream by an analog-to-digital (A/D) converter. The A/D converter samples and quantizes the continuous signal and assigns a binary number within a set. The sampling interval is determined by the *Nyquist rate* which states that in order to recover the signal, the sampling rate should be bigger than equal $2W$ where W is the bandwidth of the signal.

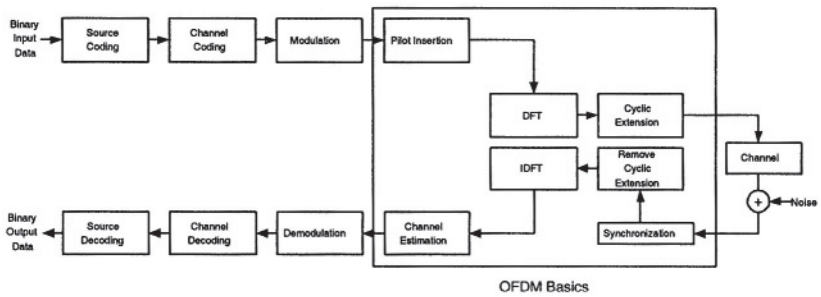

Figure 2.6: Components of a multi-carrier system

2.2.1 Coding

Source coding is used to reduce the redundancy in binary data to obtain a compressed version. An information theoretic channel is shown in Figure 2.7.

The concept of entropy defines the average information contained in a signal. The measure of entropy is the average uncertainty in the random variable. If $\log p(x_i)$ is the uncertainty $I(x_i)$ for a given discrete alphabet, then the entropy $H(X)$ is defined for an alphabet Ω as

$$H(X) = E\{I(x_i)\} = -\sum_j p(x_i) \log_2 p(x_i) \tag{2.7}$$

if M is the cardinality of the discrete alphabet set, then H is bounded by $0 \leq H(X) \leq \log_2 M$ since the maximum uncertainty is achieved if the random variable is uniformly distributed ($p(x_j) = 1/M$). The entropy of a random variable is the lower bound that can be achieved in data compression.

The expected length $L = \sum_i p(x_i) l(x_i)$ of the signal is constrained by $H \leq L$ where $l(x_i)$ is the length of code word of x_i. The basic idea of data compression is assigning short code words to the most frequent outcomes.

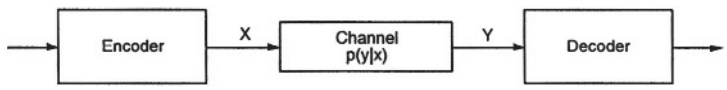

Figure 2.7: A communication channel in information theoretic view

In a communication systems, the transmitted symbols X are subject to change due to the channel impairments. As a result, the received symbols Y are not exactly the same as X. The channel capacity which is defined as the maximum achievable data rate is closely connected to the mutual information between X and Y.

Mutual information $I(X, Y)$ is a measure of the correlation between two random variables [16].

$$I(X,Y) = H(X) - H(X|Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (2.8)$$

This measure is necessary since it is the mean of information capacity C in communication. Capacity is defined as $C = max_{p(x_i)} = I(X;Y)$ and Shannon states that all rates up to C are reliable.

Channel coding is essential to introduce code words that have data rate R less than equal to C. Unlike data compression, the channel coding introduces controllable redundancy to the transmitted data. Block, convolutional, and trellis coding are described in detail and issues about concatenated and turbo coding are cited intensely in Chapter 7.

The combination of source coding $(H \leq R = L)$ and channel coding $(R \leq C)$ define the reliable communication system [16].

2.2.2 Modulation

Modulation of a signal changes binary bits into an analog waveform. Modulation can be done by changing the amplitude, phase, and fre-

quency of a sinusoidal carrier. There are several digital modulation techniques used for data transmission. We will review the ones that are basic or frequently used and appropriate for OFDM. The nature of OFDM only allows the signal to modulate in amplitude and phase. There can be coherent or non-coherent modulation techniques. Unlike non-coherent modulation, coherent modulation uses a reference phase between the transmitter and the receiver which brings accurate demodulation together with receiver complexity.

Each of these classes can be represented as M-ary modulation which is the general representation of all modulation schemes. One of M possible signals can be sent in M-ary signalling schemes for a duration T. When $M = 2$ then the modulation scheme reduces to the binary modulation of its kind [18].

Coherent Modulation

Amplitude Shift Keying (ASK), Phase Shift Keying (PSK), Quadrature-Amplitude Modulation (QAM) are appropriate coherent modulation techniques for OFDM.

1. ***M*-ary ASK**

   Amplitude Shift Keying (ASK) carry information in amplitude. In a given set of $\{A_1, \cdots, A_M\}$, the signal s_i is

   $$s_i(t) = A_i \cos \omega t \tag{2.9}$$

   The 2-ASK also known as BPSK is a famous modulation technique which is being widely used. If E is the signal energy per symbol, then the constellation points reside at $\sqrt{2E/T}$ and $-\sqrt{2E/T}$.

2. ***M*-ary PSK**

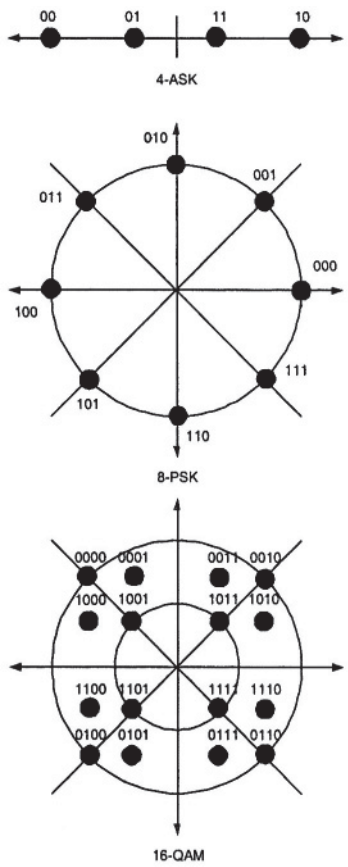

Figure 2.8: Constellation diagram

M-ary PSK shifts the phase for different constellations given a constant amplitude. There are M possible signals which has $\theta_i = 2(i-1)/M$ phases, each signal is placed on a circle and represented as

$$s_i(t) = \sqrt{\frac{2E}{T}}\cos(2\pi f_c t + \frac{2\pi}{M}(i-1)),\ i = 1, 2, \cdots, M \quad (2.10)$$

Notice that 2-PSK is again BPSK.

3. ***M*-ary QAM**

   M-ary QAM is a hybrid scheme which has amplitude modulation as well as a phase modulation. The M signals are represented as

$$s_i(t) = \sqrt{\frac{2E_0}{T}}a_i\cos(2\pi f_c t) + \sqrt{\frac{2E_0}{T}}b_i\cos(2\pi f_c t),\ 0 \le t \le T \quad (2.11)$$

   $\{a_i, b_i\}$ is an element of the L-by-L matrix:

$$\begin{bmatrix} (-L+1, L-1) & (-L+3, L-1) & \cdots & (L-1, L-1) \\ (-L+1, L-3) & (-L+3, L-3) & \cdots & (L-1, L-3) \\ \vdots & \vdots & & \vdots \\ (-L+1, -L+1) & (-L+3, -L+1) & \cdots & (L-1, -L+1) \end{bmatrix} \quad (2.12)$$

   where $L = \sqrt{M}$ [18].

   The case for $M = 4$ is same as 4-PSK which is also known as Quadrature-PSK (QPSK).

Non-coherent Modulation

IEEE 802.11 and HIPERLAN/2 uses coherent modulation, but low data rate systems such as Digital Audio Broadcasting (DAB) prefers

non-coherent modulation. Non-coherent modulation is useful when there is no knowledge of the carrier phase. It simplifies the receiver structure since there is no need for channel estimation, carrier or phase tracking.

Differential techniques are very common in non-coherent modulation. The data is encoded in phase or amplitude which is the difference between phases or amplitudes of two consecutive symbols.

Differential-PSK which is used in DAB carry the data in phase difference of two M-PSK modulated signals,

$$s_i(t) = A\cos(\omega t + \Delta\theta) \tag{2.13}$$

where $\Delta_i\theta = \theta_i - \theta_{i-1}$.

Differential detection uses two consecutive symbols and calculates $\theta_{i-1} + \Delta\theta = \theta_i$ which gives a signal that needs to be PSK demodulated.

2.3 Multi-Carrier System Fundamentals

Let $[D_0, D_1, ..., D_{N-1}]$ denote data symbols which is where $D_t = s_i(t)$ for $i \in \{1, \cdots, M\}$. Digital signal processing techniques, rather than frequency synthesizers, can be deployed to generate orthogonal sub-carriers. The DFT as a linear transformation maps the complex data symbols $[D_0, D_1, ..., D_{N-1}]$ to OFDM symbols $[d_0, d_1, ..., d_{N-1}]$ such that

$$d_k = \sum_{n=0}^{N-1} D_n e^{j2\pi n\frac{k}{N}} \tag{2.14}$$

The linear mapping can be represented in matrix form as:

$$\bar{d} = \bar{W}\bar{D}, \tag{2.15}$$

where:

$$\bar{W} = \begin{bmatrix} 1 & \cdots & 1 & 1 \\ 1 & W & \cdots & W^{N-1} \\ 1 & W^2 & \cdots & W^{2(N-1)} \\ \vdots & & & \vdots \\ 1 & W^{N-1} & & W^{N(N-1)} \end{bmatrix} \tag{2.16}$$

and,

$$W = e^{j\frac{2\pi}{N}} \tag{2.17}$$

$\bar{W}$ is a symmetric and orthogonal matrix. After FFT, a cyclic pre/postfix of lengths k_1 and k_2 will be added to each block (OFDM symbol) followed by a pulse shaping block. Proper pulse shaping has an important effect in improving the performance of OFDM systems in the presence of some channel impairments, and will be discussed in Chapter 5. The output of this block is fed into a D/A at the rate of f_s and low-pass filtered. A basic representation of the equivalent complex baseband transmitted signal is

$$x(t) = \sum_{n=0}^{N-1} \{D_n e^{j2\pi \frac{n}{N} f_s t}\} \tag{2.18}$$

for

$$-\frac{k_1}{f_s} < t < \frac{N + k_2}{f_s} \tag{2.19}$$

A more accurate representation of OFDM signal including windowing effect is

$$x(t) = \sum_{l=-\infty}^{\infty} \sum_{k=-k_1}^{N+k_2} \sum_{n=0}^{N-1} \{D_{nl} e^{j2\pi \frac{n}{N} k}\} w(t - \frac{k}{f_s} - lT) \tag{2.20}$$

D_{nl} represents the nth data symbol transmitted during the lth OFDM block, $T = (N + k_1 + k_2)/f_s$ is the OFDM block duration, and $w(t)$ is the window or pulse shaping function. The extension of the OFDM

block is equivalent to adding a cyclic pre/postfix in the discrete domain. The received signal for a time-varying random channel is

$$r(t) = \int_0^{\infty} x(t-\tau)h(t,\tau)d\tau + n(t). \tag{2.21}$$

The received signal is sampled at $t = \frac{k}{f_s}$ for $k = \{-k_1, ..., N+k_2-1\}$. With no inter-block interference, and assuming that the windowing function satisfies $w(n-1) = \delta_{nl}$, the output of the FFT block at the receiver is

$$\tilde{D}_m = \frac{1}{N}\sum_{k=0}^{N-1} r_k e^{-j2\pi m \frac{k}{2N}}, \tag{2.22}$$

where

$$r_k = \sum_{n=0}^{N-1} H_n D_n e^{j2\pi \frac{n}{2N} k} + n(k). \tag{2.23}$$

A complex number H_n is the frequency response of the time-invariant channel $h(t-\tau)$ at frequency n/T. So,

$$\tilde{D}_m = \begin{cases} H_n D_n + N(n), & n = m \\ N(n), & n \neq m \end{cases} \tag{2.24}$$

$n(t)$ is white Gaussian noise with a diagonal covariance matrix of $E(n(k)n(l)) = \delta I$. Therefore, the noise components for different subcarriers are not correlated,

$$E(\bar{n}(k)\bar{n}^*(l)) = W\delta I W^T = \delta I. \tag{2.25}$$

where $\bar{n}(k)$ is the vector of noise samples $\{ n(k), \cdots, n(k-N)\}$.

A detailed mathematical analysis of OFDM in multi-path Rayleigh fading is presented in the last section.

2.4 DFT

The key components of an OFDM system are the Inverse DFT in the transmitter and the DFT in the receiver. These components must im-

plement

$$d_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} D_k W_N^{kn},\ n = 0, \cdots, N-1 \tag{2.26}$$

and,

$$D_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} d_k W_N^{-kn},\ n = 0, \cdots, N-1 \tag{2.27}$$

respectively.

These operations perform reversible linear mappings between N complex data symbols $\{D\}$ and N complex line symbols $\{d\}$. It can be seen that the two operations are essentially identical.

The scale factor $1/\sqrt{N}$ provides symmetry between the operations and also preservation of power. Frequently, a scale factor of $1/N$ is used in one direction and unity in the other instead. In an actual implementation, this is immaterial because scaling is chosen to satisfy considerations of overflow and underflow rather than any mathematical definition.

A general N-to-N point linear transformation requires N^2 multiplications and additions. This would be true of the DFT and IDFT if each output symbol were calculated separately. However, by calculating the outputs simultaneously and taking advantage of the cyclic properties of the multipliers $e^{\pm j2\pi nk/N}$, Fast Fourier Transform (FFT) techniques reduce the number of computations to the order of $N \log N$. The FFT is most efficient when N is a power of two. Several variations of the FFT exist, with different ordering of the inputs and outputs, and different use of temporary memory. One variation, decimation in time, is shown in Figure 2.9.

Figure 2.10 shows the architecture of an OFDM system capable of using a further stage of modulation, employing both in-phase and quadrature modulators. This configuration is common in wireless communication systems for modulating baseband signals to the required IF or RF frequency band. It should be noted that the basic con-

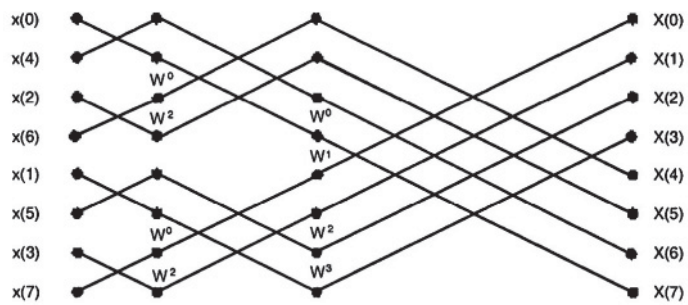

Figure 2.9: An FFT implementation (decimation in time)

figuration illustrated does not account for channel dispersion, which is almost always present. The channel dispersion problem is solved by using the cyclic prefix which will be described later.

Small sets of input bits are first assembled and mapped into complex numbers which determine the constellation points of each sub-carrier. In most wireless systems, smaller constellations are formed for each sub-carrier. In wireline systems, where the signal-to-noise ratio is higher and variable across the frequency range, the number of bits assigned to each sub-carrier may be variable. Optimization of this bit assignment is the subject of bit allocation, to be discussed in Chapter 3 and 11.

If the number of sub-carriers is not a power of two, then it is common to add symbols of zero value to the input block so that the advantage of using such a block length in the FFT is achieved. The quantity N which determines the output symbol rate is then that of the padded input block rather than the number of sub-carriers.

The analog filters at the transmitter output and the receiver input should bandlimit the respective signals to a bandwidth of $1/T$. Low pass filters could be used instead of the band-pass filters shown, placed on the other side of the modulator or demodulator. The transmit filter

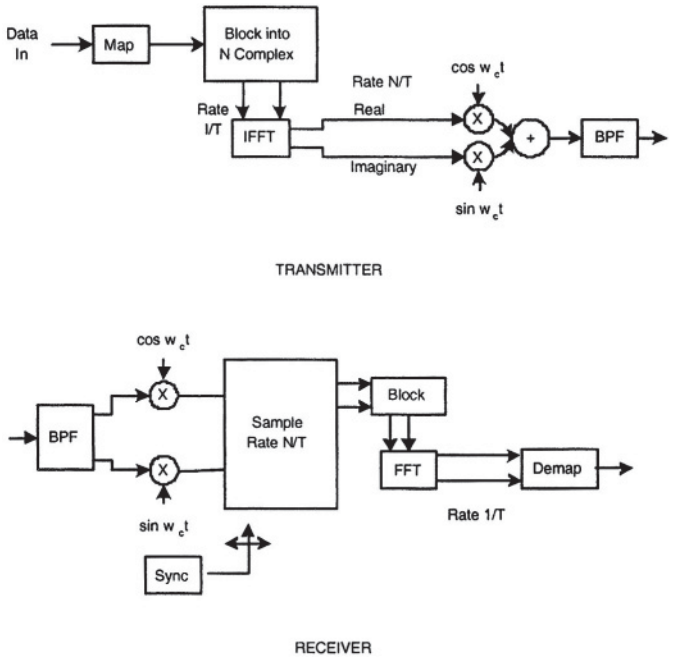

Figure 2.10: System with complex transmission

eliminates out-of-band power which may interfere with other signals. The receive filter is essential to avoid aliasing effects. The transmitted signal has a continuous spectrum whose samples at frequencies spaced $1/T$ apart agree with the mapped input data. In particular, the spectrum of each sub-carrier is the form $sinc(1/T)$, whose central value is that input value, and whose nulls occur at the central frequencies of all other sub-carriers.

The receiver operations are essentially the reverse of those in the transmitter. A critical set of functions, however, are synchronization of carrier frequency, sampling rate, and block framing. There is a minimum delay of $2T$ through the system because of the block assembly functions in the transmitter and receiver.

The FFT functions may be performed either by a general purpose DSP or special circuitry, depending primarily on the information rate to be carried. Some simplification compared with full multiplication may be possible at the transmitter, by taking advantage of small constellation sizes.

The number of operations per block of duration T is $KN \log_2 N$, where K is a small quantity. To compare this with a single carrier system, the number of operations per line symbol interval T/N is $K \log_2 N$, which is substantially below the requirement of an equalizer in a typical single carrier implementation for wireline applications. In most wireline systems it is desirable to transmit the transformed symbols d_n without any further modulation stages. In this case, it is only possible to transmit real line symbols, and not the above complex quantities. The problem is solved by augmenting the original sequence D_n by appending its complex conjugate to it, as shown in Figure 2.11. The $2N$-point IFFT of this augmented sequence is then a sequence of $2N$ real numbers, which is equivalent in bandwidth to N complex numbers.

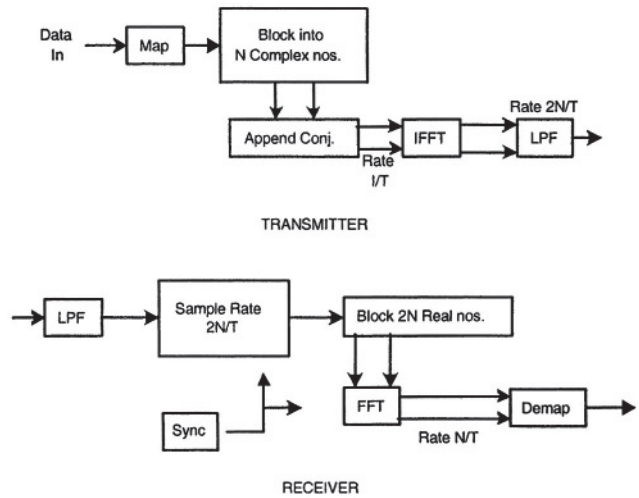

Figure 2.11: System with real transmission

The augmented sequence is formed from the original sequence as

$$D_n^{'} = \begin{cases} D_n, & n = 1, \cdots, N-1 \\ D_{2N-n}^{*}, & n = N+1, \cdots, 2N-1 \end{cases} \tag{2.28}$$

In order to maintain conjugate symmetry, it is essential that $D_0^{'}$ and $D_N^{'}$ be real. If the original D_0 is zero, as is common, then $D_0^{'}$ and $D_N^{'}$ are set to zero. Otherwise, $D_0^{'}$ may be set to $Re(D_0)$ and $D_N^{'}$ to $Im(D_0)$.

For the simple case of $D_0 = 0$, the output of the IFFT is:

$$\begin{aligned} d_m &= \sum_{n=0}^{2N-1} D_n^{'} e^{\frac{j\pi mn}{N}} \\ &= 2Re \sum_{n=0}^{N-1} D_n e^{\frac{j\pi mn}{N}} \\ &= 2\sum_{n=0}^{N-1} [A_n cos\frac{\pi mn}{N} - B_n sin\frac{\pi mn}{N}], \quad m = 0, \cdots, 2N-1, \end{aligned} \tag{2.29}$$

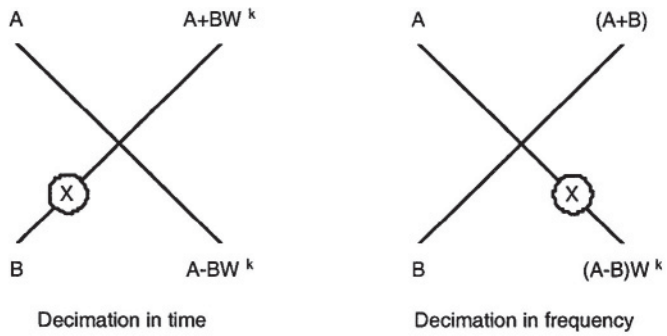

Figure 2.12: Two different techniques for FFT butterfly

where $D_n = A_n + jB_n$. Of course the scaling by a factor of two is immaterial and can be dropped. This real orthogonal transformation is fully equivalent to the complex one, and all subsequent analyzes are applicable.

2.5 Partial FFT

In some applications, the receiver makes use of a subset of transmitted carriers. For example, in a digital broadcasting system receiver [12] that decodes a group of sub-carriers (channels) or in some multirate applications, the receiver has several fall-back modes so using a repetitive structure is advantageous. One of the benefits of an OFDM system with an FFT structure is the fact that it lends itself to a repetitive structure very well. This structure is preferred compared to the required filtering complexity in other wideband systems. Two common structures are shown in Figure 2.12.

An example of partial FFT is shown in Figure 2.13. In order to detect (receive) the marked point at the output, we can restrict the

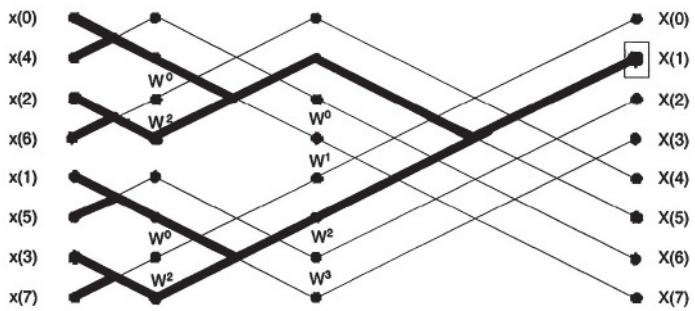

Figure 2.13: Partial FFT (DIT)

FFT calculation to the marked lines. Therefore, a significant amount of processing will be saved.

Two main differences between decimation in time (DIT) and decimation in frequency (DIF) are noted [11]. First, for DIT, the input is bit-reversed and output is in natural order, while in DIF , the reverse is true. Secondly, for DIT complex multiplication is performed before the add-subtract operation, while in DIF the order is reversed. While complexity of the two structures is similar in typical DFT, this is not the case for partial FFT [12]. The reason is that in the DIT version of partial FFT, a sign change (multiplication by 1 and -1) occurs at the first stages, but in the DIF version it occurs in later stages.

2.6 Cyclic Extension

Transmission of data in the frequency domain using an FFT, as a computationally efficient orthogonal linear transformation, results in robustness against ISI in the time domain. Unlike the Fourier Transform (FT), the DFT (or FFT) of the circular convolution of two signals is

Figure 2.14: Prefix and postfix cyclic extension

equal to the product of their DFT's (FFT).

$$\begin{aligned} FT\{d_n * h_n\} &= FT\{d_n\} \times FT\{h_n\} \\ DFT\{d_n \otimes h_n\} &= DFT\{d_n\} \times DFT\{h_n\}, \end{aligned} \tag{2.30}$$

where $*$ and $\otimes$ denote linear and circular convolution respectively.

The signal and channel, however, are linearly convolved. After adding prefix and postfix extensions to each block, the linear convolution is equivalent to a circular convolution as shown in Figure 2.14. Instead of adding a prefix and a postfix, some systems use only prefix, then by adjusting the window position at the receiver proper cyclic effect will be achieved.

Using this technique, a signal, otherwise aliased, appears infinitely periodic to the channel. Let's assume the channel response is spread over M samples, and the data block has N samples then:

$$y(n) = \sum_{m=0}^{N-1} d(m)h(n-m) \times R_N(n), \; n = 0, 1, \cdots, N+M-1 \tag{2.31}$$

where $R_N(n)$ is a rectangular window of length N. To describe the effect of distortion, we proceed with the Fourier Transform noting that convolution is linear

$$\begin{aligned} Y(k) &= Y(e^{jw}) \Big|_{\omega = k\omega_s} \\ &= \{D(e^{jw}) \times H(e^{jw})\} * \frac{sin\omega\frac{N}{2}}{sin\frac{\omega}{2}} e^{j\omega\frac{N-1}{2}} \Bigg|_{\omega = k\omega_s} \end{aligned} \tag{2.32}$$

After linear convolution of the signal and channel impulse response, the received sequence is of length $N + M - 1$. The sequence is truncated to N samples and transformed to the frequency domain, which is equivalent to convolution and truncation. However, in the case of cyclic pre/postfix extension, the linear convolution is the same as the circular convolution as long as channel spread is shorter than guard interval. After truncation, the DFT can be applied, resulting in a sequence of length N because the circular convolution of the two sequences has period of N.

Intuitively, an N-point DFT of a sequence corresponds to a Fourier series of the periodic extension of the sequence with a period of N. So, in the case of no cyclic extension we have

$$\sum_{i=-\infty}^{\infty} \sum_{m=0}^{N-1} d(m)h(n + iN - m), \tag{2.33}$$

which is equivalent to repeating a block of length $N + M - 1$ with period N. This results in aliasing or inter-symbol interference between adjacent OFDM symbols. In other words, the samples close to the boundaries of each symbol experience considerable distortion, and with longer delay spread, more samples will be affected. Using cyclic extension, the convolution changes to a circular operation. Circular convolution of two signals of length N is a sequence of length N so the inter-block interference issue is resolved.

Proper windowing of OFDM blocks, as shown later, is important to mitigating the effect of frequency offset and to control transmitted signal spectrum. However, windowing should be implemented after cyclic extension of the frame, so that the windowed frame is not cyclically extended. A solution to this problem is to extend each frame to $2N$ points at the receiver and implement a $2N$ FFT. Practically, it requires a $2N$ IFFT block at the transmitter, and $2N$ FFT at the receiver. However, by using partial FFT techniques, we can reduce the computation by calculating only the required frequency bins.

If windowing was not required, we could have simply used zero

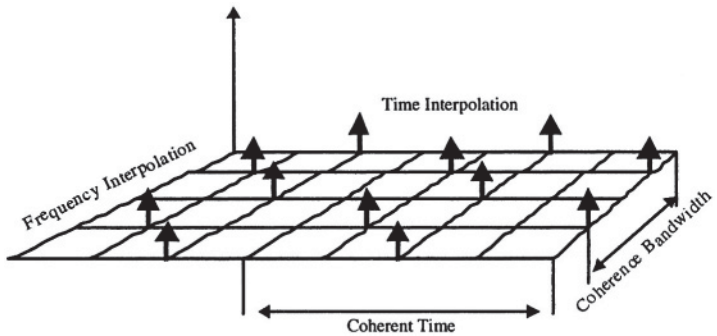

Figure 2.15: Pilot positioning in time and frequency

padded pre/postfix. Before the DFT at the receiver, we could copy the beginning and the end of frame as prefix and postfix. This creates the same effect of cyclic extension with the advantage of reducing transmit power and causing less ISI.

The relative length of cyclic extension depends on the ratio of the channel delay spread to the OFDM symbol duration.

2.7 Channel Estimation

Channel estimation inverts the effect of non-selective fading on each sub-carrier. The channel estimation in OFDM is usually coherent detection which is based on the use of pilot subcarriers in given positions of the frequency-time grid. Chapter 6 covers the channel estimation in more detail. Time varying channels are required to implement repeated pilot symbols. Nature of the channel such as coherence time and coherence bandwidth determines the pilot spacing in time and frequency. As it can be seen from Figure 2.15, the number of pilots necessary to use can be reduced by spacing the pilots less than coherence bandwidth in frequency and less than in coherence time in time

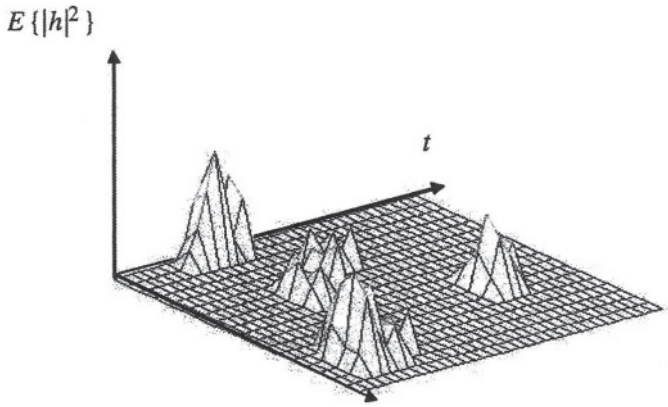

Figure 2.16: Typical impulse response of a wireless channel

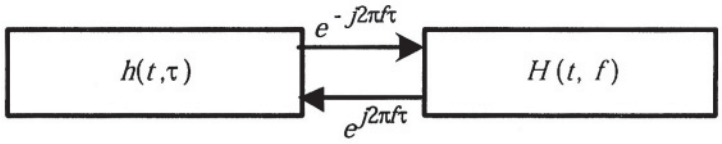

Figure 2.17: Relationship between system functions

domain.

Let

$$h_t = (h_1, h_2, \cdots, h_M)^T \tag{2.34}$$

represent the overall impulse response of the channel, including transmitter and receiver filters, with length M [15]. The N-element training sequence is represented by

$$\begin{aligned} &X = [X_{ij}], 1 \leq i \leq M+N-1, \qquad 1 \leq j \leq M \\ &X_{ij} = \begin{cases} x_{i-j+1}, & 1 \leq i-j+1 \leq N \\ 0, & elsewhere \end{cases} . \end{aligned} \tag{2.35}$$

The noise samples are

$$n = (n_1, n_2, \cdots, n_{M+N-1})^T, \tag{2.36}$$

with

$$R_n = E\{nn^*\} = LL^* \tag{2.37}$$

in which L is a lower triangular matrix, which can be calculated by the Cholesky method. The received signal is then

$$r = Xh_t + n \tag{2.38}$$

After correlation or matched filtering, the estimate of the impulse response is:

$$\hat{h}_t = X^* r = X^* X h_t + X^* n. \tag{2.39}$$

If the additive noise is white, the matched filter is the best estimator in terms of maximizing signal-to-noise ratio. However, the receiver filter colors the noise. With a whitening filter, the estimate is given in [111]. An unbiased estimator, which removes the effect of sidelobes, is

$$\hat{h}_t = (X^* R_n^{-1} X)^{-1} X^* R_n^{-1} r = h + (X^* R_n^{-1} X)^{-1} X^* R_n^{-1} n. \tag{2.40}$$

In the dual system architecture, circular convolution is replaced by multiplication. The repetition of the training sequence results in circular convolution. Therefore, Equation 2.39 is replaced by:

$$\hat{H}^\omega = X^\omega X^{\omega^*} H^\omega + X^{\omega^*} N^\omega, \tag{2.41}$$

where superscript ω shows Fourier Transform and all products are scalar. The same procedure can be applied to Equation 2.40. This procedure is also called frequency equalization.

2.8 Modelling of OFDM for Time-Varying Random Channel

2.8.1 Randomly Time-Varying Channels

Linear randomly time-varying channels can be characterized by their impulse response , which represents system response at time t due to an impulse at time $t - \tau$. So, for input $x(t)$, the output would be $y(t) = \int h(t,\tau)x(t-\tau)d\tau$.

Other kernels can be defined to relate input and output in time and frequency domains [13]. Since the impulse response is time-varying, it is expected that the system transfer function is time-varying. The transfer function is defined as

$$H(t,f) = \int e^{-j2\pi ft} h(t,\tau)d\tau, \tag{2.42}$$

$$y(t) = \int X(f)H(t,f)e^{j2\pi ft}df. \tag{2.43}$$

If the input is a single tone $e^{j2\pi mt}$ the output is:

$$\int e^{j2\pi m(t-\tau)} h(t,\tau)d\tau = e^{j2\pi mt}H(t,m), \tag{2.44}$$

which shows that the time-varying transfer function has an interpretation similar to time-invariant systems. Namely, the output of the system for a sinusoidal function is the input multiplied by the transfer function. In general, $h(t,\tau)$ is a random process and its exact statistical description requires a multi-dimensional probability distribution. A practical characterization of the channel is given by first and

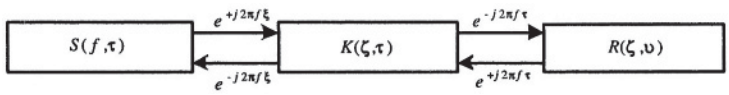

Figure 2.18: Relationship between correlation functions

second order moments of system functions. For example, many randomly time-varying physical channels are modelled as wide sense stationary non-correlated scatterers. Therefore, they are white and non-stationary with respect to delay variable τ and stationary with respect to t.

$$E\{h(t,T)h^*(t^{'},\tau^{'})\} = K(t-u,\tau)\delta(\tau-\tau^{'}), \tag{2.45}$$

$K(t,\tau)$ is called the tap gain correlation. The scattering function of the channel is defined as

$$S(f,\tau) = \int K(\psi,\tau)e^{-j2\pi f\psi}d\psi. \tag{2.46}$$

For any fixed τ, the scattering function may be regarded as the complex-valued spectrum of the system. The relationship between different system functions is shown in the Figure 2.17 and Figure 2.18.

Coherence bandwidth of the received signal (or channel) is defined as the frequency distance beyond which frequency responses of the signal are uncorrelated. The reciprocal of delay spread is an estimate for coherence bandwidth. Coherence time is the time distance beyond which the samples of received signal (or channel impulse response) are uncorrelated. Since the channel and signal are Gaussian, independence of samples is equivalent to their being non-correlated. The reciprocal of Doppler spread is an estimate of coherence time [14].

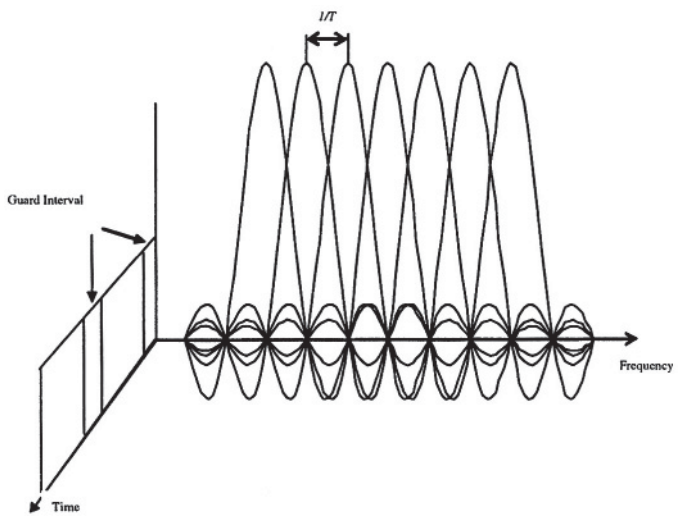

Figure 2.19: OFDM time and frequency span

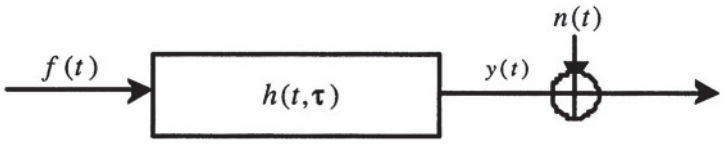

Figure 2.20: Time-varying channel

2.8.2 OFDM in Randomly Time-Varying Channels

In general, OFDM is a parallel transmission technique in which N complex symbols modulate N orthogonal waveforms $\psi_i(t)$ which maintain their orthogonality at the output of the channel. This requires that the correlation function of the channel output process, in response to different waveforms, has orthogonal eigenfunctions.

Let $h(t,\tau)$ be the Rayleigh distributed, time-varying complex impulse response of the channel, while $R(t,\tau)$ denotes the complex correlation of the output in response to a basic pulse shape $f(t)$:

$$R(t,\tau) = E\{y(t)y(t+\tau)^*\}. \tag{2.47}$$

Using Karhunen-Loeve expansion, the output of the channel $y(t)$ can be represented as

$$y(t) = \sum_i D_i y_i \varphi_i(t), \tag{2.48}$$

where the $\varphi_i(t)$s are eigenfunctions of the correlation function $R(t,\tau)$ and

$$y_i = \int y(t)\varphi_i(t)dt, \tag{2.49}$$

Using Karhunen-Loeve expansion, the output of the channel $y(t)$ can be represented as

$$y(t) = \sum_i y_i \varphi_i(t), \tag{2.50}$$

where the $\varphi_i(t)$s are eigenfunctions of the correlation function $R(t,\tau)$ and

$$y_i = \int y(t)\varphi_i(t)dt \tag{2.51}$$

The corresponding eigenvalues of the correlation function are λ_i where

$$\int R(t,\tau)\varphi_i(\tau)d\tau = \lambda_i \varphi_i(t). \tag{2.52}$$

For a Gaussian random process $y(t)$, the coefficients in Equation 2.50 are Gaussian random variables and their powers are the corresponding

eigenvalues. If the input consists of orthogonal modulating functions which are harmonics of the form

$$\psi_k(t) = u(t)e^{j2\pi f_k t}, \tag{2.53}$$

where $u(t)$ is the basic waveform such as time-limited gating function. The eigenfunctions of the received signal correlation function $R_k(t,\tau)$ are,

$$\varphi_i(t)e^{j2\pi f_i t}, \tag{2.54}$$

where again $\varphi_i(t)$ are the eigenfunctions' output correlation functions in response to basic waveform $u(t)$. Therefore, the orthogonality condition is

$$\int \varphi_i(t)\varphi_k^*(t)e^{j2\pi(f_i - f_k)}dt = 0 \tag{2.55}$$

which can be achieved by increasing $\delta f = f_{i+1} - f_i$. Analytically this means that l_2 sub-spaces spanned by different eigenfunctions are mutually orthogonal. In other words, the frequency response of the channel for different frequency bins is independent. For a pure tone signal of frequency f_k, the output of random channel is:

$$\begin{aligned} y_k(t) &= \int h(t,\tau)x(t-\tau)d\tau = \int h(t,\tau)e^{j2\pi f_k(t-\tau)}d\tau \\ &= e^{j2\pi f_k t}\int h(t,\tau)e^{-j2\pi f_k \tau}d\tau \end{aligned} \tag{2.56}$$

The transfer function of a time-varying channel is defined as

$$H(f,t) = \int e^{-j2\pi f\tau}h(t,\tau)d\tau, \tag{2.57}$$

so the output of channel can be represented as

$$y_k(t) = H(f_k,t)e^{j2\pi f_k t}, \tag{2.58}$$

and the orthogonality requirement of the output processes holds. If the filter gains $H(f_k,t)$ are orthogonal, and the tap filters are non-correlated and Gaussian, then they will be independent too. As an example, consider orthogonal functions of the form

$$\varphi_{ik}(t) = u(t-iT)e^{2j\pi f_k(t-jT)}. \tag{2.59}$$

This set of orthogonal functions is used to modulate data symbols and transmit them in a single OFDM symbol timing interval such that the orthogonality in the frequency domain is preserved. By choosing the above set of orthogonal functions, the eigenfunctions used in Karhunen-Loeve expansion will be related as shown in Equation 2.54. The optimum receiver maximizes a *posteriori* probability (conditional probability of the received signal given a particular sequence is transmitted). Usually, channel characteristics do not change during one or a few symbols. Therefore, the channel output can be represented as

$$y(t) = \sum_k H(f_k, t) e^{j2\pi f_k t}. \tag{2.60}$$

The received signal $r(t) = y(t) + n(t)$ with $n(t)$ representing additive white Gaussian noise can be expanded as

$$r(t) = \sum_i r_i \varphi_i(t), \tag{2.61}$$

where

$$\begin{aligned} r_i &= y_i + n_i \\ n_i &= \int n(t)\varphi_i(t)dt \\ y_i &= \int y(t)\varphi_i(t)dt \end{aligned} \tag{2.62}$$

Notice that the noise projection should be interpreted as a stochastic integral, because a Riemann integral is not defined for white noise. r_is are independent zero mean Gaussian random variables and their powers are

$$E|r_i|^2 = E|H_i|^2 + \bar{n_i^2}, \tag{2.63}$$

assuming that data symbols are of equal and unit power. So, the problem at hand reduces to the detection of a Gaussian random process corrupted by Gaussian noise.

Chapter 6

Channel Estimation and Equalization

6.1 Introduction

Time-variant and frequency-selective fading channels present a severe challenge to the designer of a wireless communication system. An OFDM receiver implements dual role to tackle this problenv.channel estimation or equalization. Several choices are possible for the implementation of a receiver depending on the modelling of the channel and complexity invested in each task. The nature of the OFDM enables powerful estimation and equalization techniques.

6.2 Channel Estimation

In OFDM link, modulated bits are disturbed during the transmission through the channel since the channel introduces amplitude and phase shifts due to frequency selective and time-varying nature of the radio

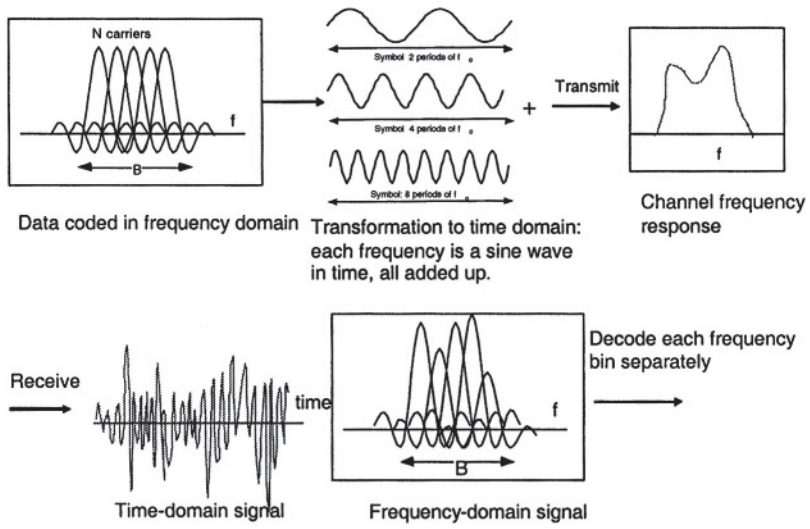

Figure 6.1: Observation of the distortion by channel [87]

channel. In order for the receiver to acquire the original bits, it needs to take into account these unknown changes. Figure 6.1 illustrates a typical OFDM transmission and the effects of the channel. The receiver applies either coherent detection or non-coherent detection to recover the original bits. Coherent detection uses reference values that are transmitted along with data bits. The receiver can tune into those values and estimate the channel only at the locations where the reference values are located. The entire channel can be estimated by using several interpolation techniques. Non-coherent detection on the other hand, does not use any reference values but uses differential modulation where the information is transmitted in difference of the two successive symbols. The receiver uses two adjacent symbols in time or two adjacent subcarriers in frequency to compare one with another to acquire the transmitted symbol [52].

In this section, we will review channel estimation techniques. We

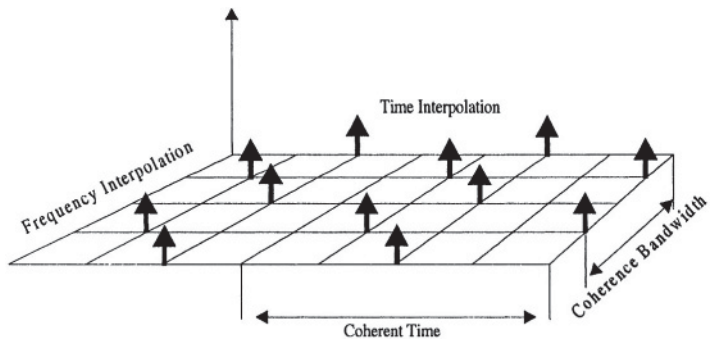

Figure 6.2: Pilot positioning in time and frequency

start with coherent and non-coherent detection wherein we describe the block type and comb type channel estimation and continue with the interpolation techniques. Finally, channel estimation in MIMO-OFDM system is discussed.

6.2.1 Coherent Detection

The coherent detection can be performed by either inserting pilot tones into all of the subcarriers of OFDM symbols with a period in time or inserting pilot tones into each OFDM symbol with a period in frequency. In the case of time-varying channels, the pilot signal should be repeated frequently. The spacing between pilot signals in time and frequency depends on coherence time and bandwidth of the channel. One can reduce the pilot signal overhead by using a pilot signal with a maximum distance of less than the coherence time and coherence bandwidths. Then, by using time and frequency interpolation, the impulse response and frequency response of the channel can be calculated. Figure 6.2 illustrates the pilot arrangement in time-frequency plane.

If B_d is Doppler spread and σ_τ is delay spread, a suitable choice

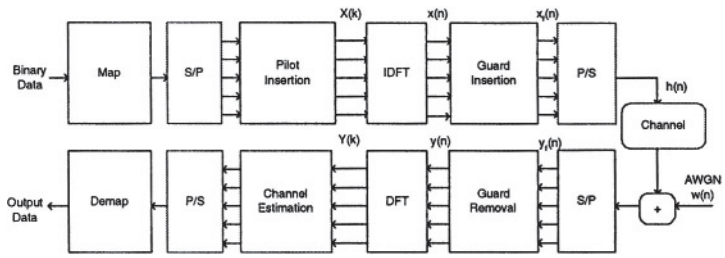

Figure 6.3: Baseband OFDM system

for pilot spacing in time N_p^t, and in frequency N_p^f, is as follows:

$$N_p^t \approx \frac{1}{B_d T_{symb}} \quad N_p^f \approx \frac{1}{\triangle f \sigma_\tau} \tag{6.1}$$

where $\triangle f$ is subcarrier bandwidth, T_{symb} is symbol time.

Various types exist in the pilot arrangement depending on the channel condition. Those types can be classified into two at higher level. The first one is the block type pilot arrangement for slow fading channels. All subcarriers are used once and the estimated channel is used for the coming symbols since the channel changes very slowly for slow fading channels, and the channel is highly correlated for the consecutive symbols. The second approach is the comb type pilot arrangement, which interleaves the pilots over time and frequency. Comb-type estimation uses several interpolation techniques to estimate the entire channel. Flat fading channels can be estimated by comb-type pilot arrangement and pilot frequency increases if the channel is frequency selective fading.

The OFDM system based on pilot channel estimation is given in Figure 6.3. The binary information is first grouped and mapped according to the modulation in "signal mapper". After inserting pilots either to all sub-carriers with a specific period or uniformly between the information data sequence, IDFT block is used to transform the data sequence of length N $\{X(k)\}$ into time domain signal $\{x(n)\}$

with the following equation:

$$\begin{aligned} x(n) = & \quad IDFT\{X(k)\} \qquad n = 0, 1, 2, \ldots, N-1 \\ = & \quad \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi kn}{N}} \end{aligned} \tag{6.2}$$

where N is the DFT length. Following IDFT block, guard time, which is chosen to be larger than the expected delay spread, is inserted to prevent inter-symbol interference. This guard time includes the cyclically extended part of OFDM symbol in order to eliminate inter-carrier interference (ICI). The resultant OFDM symbol is given as follows:

$$x_f(n) = \begin{cases} x(N+n), & n = -N_g, -N_g+1, \ldots, -1 \\ x(n), & n = 0, 1, \ldots, N-1 \end{cases} \tag{6.3}$$

where N_g is the length of the guard interval. The transmitted signal $x_f(n)$ will pass through the frequency selective time varying fading channel with additive noise. The received signal is given by:

$$y_f(n) = x_f(n) \bigotimes h(n) + w(n) \tag{6.4}$$

where $w(n)$ is Additive White Gaussian Noise (AWGN) and $h(n)$ is the channel impulse response.The channel response h can be represented by [56]:

$$h(n) = \sum_{i=0}^{r-1} h_i e^{j\frac{2\pi}{N} f_{D_i} T n} \delta(\sigma_\tau - \tau_i) \; 0 \leq n \leq N-1 \tag{6.5}$$

where r is the total number of propagation paths, h_i is the complex impulse response of the ith path, f_{Di} is the ith path Doppler frequency shift, σ_τ is delay spread index, T is the sample period and τ_i is the ith path delay normalized by the sampling time. At the receiver, after passing to discrete domain through *A/D* and low pass filter, guard time is removed:

$$\begin{aligned} & y_f(n) & & -N_g \leq n \leq N-1 \\ & y(n) = y_f(n+N_g) & & n = 0, 1, \ldots N-1 \end{aligned} \tag{6.6}$$

Then $y(n)$ is sent to DFT block for the following operation:

$$\begin{aligned} Y(k) &= DFT\{y(n)\} \qquad k = 0, 1, 2 \ldots N-1 \\ &= \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} y(n) e^{-j\frac{2\pi kn}{N}} \end{aligned} \tag{6.7}$$

Assuming there is no ISI, [59] shows the relation of the resulting $Y(k)$ to $H(k) = DFT\{h(n)\}$, $I(k)$ (ICI because of Doppler frequency), and $W(k) = DFT\{w(n)\}$, with the following equation [55]:

$$Y(k) = X(k)H(k) + I(k) + W(k) \; k = 0, 1 \ldots N-1 \tag{6.8}$$

where

$$H(k) = \sum_{i=0}^{r-1} h_i e^{j\pi f_{D_i} T} \frac{sin(\pi f_{D_i} T)}{\pi f_{D_i} T} e^{-j\frac{2\pi\tau_i}{N}k},$$

$$I(k) = \sum_{i=0}^{r-1} \sum_{\substack{K=0, \\ K \neq k}}^{N-1} \frac{h_i X(K)}{N} \frac{1 - e^{j2\pi(f_{D_i}T - k + K)}}{1 - e^{j\frac{2\pi}{N}(f_{D_i}T - k + K)}} e^{-j\frac{2\pi\tau_i}{N}K},$$

Following DFT block, the pilot signals are extracted and the estimated channel $H_e(k)$ for the data sub-channels is obtained in channel estimation block. Then the transmitted data is estimated by:

$$X_e = \frac{Y(k)}{H_e(k)} \qquad k = 0, 1, \ldots N-1 \tag{6.9}$$

Then the binary information data is obtained back in "signal demapper" block.

6.2.2 Block-Type Pilot Arrangement

Block type pilot channel estimation has been developed under the assumption of slow fading channels. The estimation of the channel for this block-type pilot arrangement can be based on Least Square (LS) or Minimum Mean-Square (MMSE). The MMSE estimate has been shown to give 10-15 dB gain in signal-to-noise ratio (SNR) for the

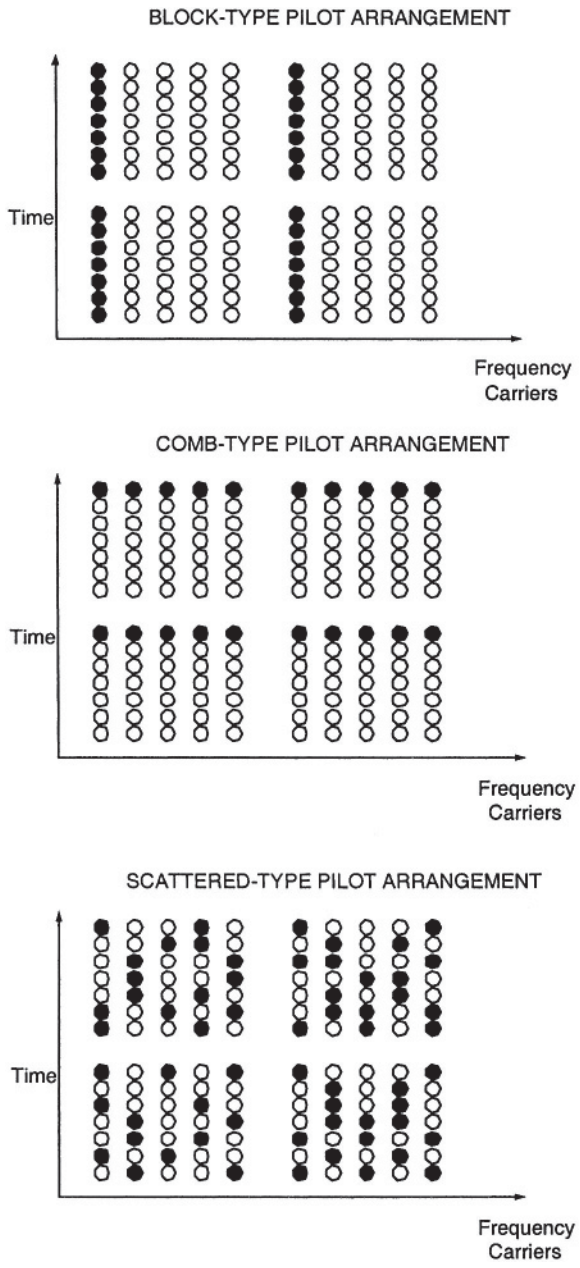

Figure 6.4: Pilot arrangement

same mean square error of channel estimation over LS estimate [53]. In [54], a low-rank approximation is applied to linear MMSE by using the frequency correlation of the channel to eliminate the major drawback of MMSE, which is complexity.

In block-type pilot based channel estimation, OFDM channel estimation symbols are transmitted periodically, in which all sub-carriers are used as pilots as seen in Figure 6.4. If the channel is constant during the block, there will be no channel estimation error since the pilots are sent at all carriers.

If inter symbol interference is eliminated by the guard interval, the Equation (6.8) is written in matrix notation as

$$Y = XFh + W \tag{6.10}$$

where

$$\begin{aligned} X &= diag\{X(0), X(1), \ldots, X(N-1)\} \\ Y &= [Y(0)Y(1)\ldots Y(N-1)]^T \\ W &= [W(0)W(1)\ldots W(N-1)]^T \\ H &= [H(0)H(1)\ldots H(N-1)]^T = DFT_N\{h\} \\ F &= \begin{bmatrix} W_N^{00} & \ldots & W_N^{0(N-1)} \\ \vdots & \ddots & \vdots \\ W_N^{(N-1)0} & \ldots & W_N^{(N-1)(N-1)} \end{bmatrix} \\ W_N^{nk} &= \frac{1}{N} e^{-j2\pi \frac{n}{N} k} \end{aligned} \tag{6.11}$$

If the time domain channel vector h is Gaussian and uncorrelated with the channel noise W, the frequency domain MMSE estimate of h is given by [54]:

$$H_{MMSE} = FR_{hY}R_{YY}^{-1}Y \tag{6.12}$$

where

$$R_{hY} = E\{hY\} = R_{hh}F^H X^H$$

$$R_{YY} = E\{YY\} = XFR_{hh}F^H X^H + \sigma^2 I_N \tag{6.13}$$

are the cross covariance matrix between h and Y and the auto covariance matrix of Y. R_{hh} is the auto-covariance matrix of h and σ^2 represents the noise variance $E\{|W(k)|^2\}$.

The LS estimate on the other hand minimizes the difference between the received value and estimated value. Thus, LS minimizes $(Y - XFh)^H(Y - XFh)$.

The estimated channel in LS is represented by

$$H_{LS} = X^{-1}Y, \tag{6.14}$$

Decision-feedback channel estimation

When the channel is slow fading, the channel estimation inside the block can be updated using the decision feedback equalizer at each sub-carrier. Decision feedback equalizer for the kth sub-carrier can be described as follows:

- The channel response at the kth sub-carrier estimated from the previous symbol $\{H_e(k)\}$ is used to find the estimated transmitted signal $\{X_e(k)\}$.

$$X_e(k) = \frac{Y(k)}{H_e(k)} \qquad k = 0, 1, \ldots N-1 \tag{6.15}$$

- $\{X_e(k)\}$ is mapped to the binary data through "signal demapper" and then obtained back as $\{X_{pure}(k)\}$ through "signal mapper".

- The estimated channel $H_e(k)$ is updated by:

$$H_e(k) = \frac{Y(k)}{X_{pure}(k)} \qquad k = 0, 1, \ldots N-1 \tag{6.16}$$

Since the decision feedback equalizer has to assume that the decisions are correct, the fast fading channel may cause the complete loss of estimated channel parameters. Therefore, as the channel fading becomes faster, there happens to be a compromise between the estimation error due to the interpolation and the error due to loss of channel tracking. For fast fading channels the comb-type based channel estimation performs better.

6.2.3 Comb-Type Pilot Arrangement

The comb-type pilot channel estimation, has been introduced to satisfy the need for equalizing when the channel changes even in one OFDM block. The comb-type pilot channel estimation consists of algorithms that estimate the channel at pilot frequencies and that interpolate the channel.

The estimation of the channel at the pilot frequencies for comb-type based channel estimation can be based on LS, MMSE or Least Mean-Square (LMS). MMSE has been shown to perform much better than LS. In [55], the complexity of MMSE is reduced by deriving an optimal low-rank estimator with singular-value decomposition.

The interpolation of the channel for comb-type based channel estimation can depend on linear interpolation, second order interpolation, low-pass interpolation, spline cubic interpolation, and time domain interpolation. In [55], second-order interpolation has been shown to perform better than the linear interpolation. In [56], time-domain interpolation has been proven to give a lower bit-error rate (BER) compared to linear interpolation.

In comb-type pilot based channel estimation, the N_p pilot signals are uniformly inserted into $X(k)$ according to the following equation:

$$\begin{aligned} X(k) &= X(mL+l) \\ &= \begin{cases} x_p(m), & l=0 \\ inf.data & l=1,\ldots L-1 \end{cases} \end{aligned} \tag{6.17}$$

where $L = (\text{number of carriers}/N_p)$ and $x_p(m)$ is the mth pilot carrier value. We define $\{Hp(k), \quad k = 0, 1, \ldots Np\}$ as the frequency response of the channel at pilot sub-carriers. The estimate of the channel at pilot sub-carriers based on LS estimation is given by:

$$H_e = \frac{Y_p}{X_p}, \qquad k = 0, 1, ...N_p - 1 \tag{6.18}$$

where $Y_p(k)$ and $X_p(k)$ are output and input at the kth pilot sub-carrier respectively.

Since LS estimate is susceptible to noise and ICI, MMSE is proposed while compromising complexity. Since MMSE includes the matrix inversion at each iteration, the simplified linear MMSE estimator is suggested in [57]. In this simplified version, the inverse only needs to be calculated once. In [55], the complexity is further reduced with a low-rank approximation by using singular value decomposition.

6.2.4 Interpolation Techniques

In comb-type pilot based channel estimation, an efficient interpolation technique is necessary in order to estimate channel in time and frequency. The pilot symbols can be interleaved within a symbol or between the symbols. The estimated pilot values should be connected in order to get an estimated value for the data values. The following techniques are written for interpolation within the symbol but it can also apply to interpolation between the symbols.

Linear Interpolation:

The linear interpolation method is shown to perform better than the piecewise-constant interpolation in [58]. The channel estimation at the data-carrier k, $mL < k < (m+1)L$, using linear interpolation is

given by:

$$\begin{aligned} H_e(k) &= H_e(mL+l) \qquad 0 \le l < L \\ &= (H_p(m+1) - H_p(m))\tfrac{l}{L} + H_p(m) \end{aligned} \tag{6.19}$$

Second-order Gaussian Interpolation:

The second-order Gaussian interpolation is shown to fit better than linear interpolation [55]. The channel estimated by second-order interpolation is given by:

$$\begin{aligned} H_e(k) &= H_e(mL+l) \\ &= c_1 H_p(m-1) + c_0 H_p(m) + c_{-1} H_p(m+1) \end{aligned}$$

$$where \quad \begin{cases} c_1 = \frac{\alpha(\alpha-1)}{2}, \\ c_0 = -(\alpha-1)(\alpha+1),\ \alpha = \frac{l}{N} \\ c_{-1} = \frac{\alpha(\alpha+1)}{2} \end{cases} \tag{6.20}$$

Low-pass Interpolation

The low-pass interpolation is performed by inserting zeros into the original sequence and then applying a lowpass FIR filter that allows the original data to pass through unchanged and interpolates between them such that the mean-square error between the interpolated points and their ideal values is minimized.

Spline cubic Interpolation:

The spline cubic interpolation produces a smooth and continuous polynomial fitted to given data points. It finds a curve that is in an equation of three degrees or less that connects a set of data points. In each point weights are attached and a flexible strip is shaped around the weights.

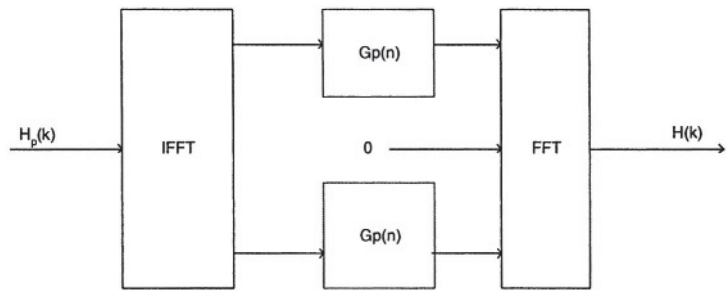

Figure 6.5: Time domain interpolation

Time domain Interpolation:

The time domain interpolation is a high-resolution interpolation based on zero-padding and DFT/IDFT [59]. After obtaining the estimated channel $\{Hp(k),\, k = 0, 1, Np - 1\}$, it is first converted to time domain by IDFT:

$$G(n) = \sum_{k=0}^{N_p-1} H_p e^{j\frac{2\pi kn}{N_p}}, \qquad n = 0, 1, \ldots N_p - 1 \tag{6.21}$$

Then, by using the basic multi-rate signal processing properties [60], the signal is interpolated by transforming the N_p points into N points with the following method:

$$M = \frac{N_p}{2} + 1$$

$$G_N = \begin{cases} G_p, & 0 \leq n < M - 2 \\ 0, & \frac{N_p}{2} \leq N - M \\ G_p(n - N + 2M - 1), & -M \leq n - N < -1 \end{cases} \tag{6.22}$$

The estimate of the channel at all frequencies is obtained by:

$$H(k) = \sum_{n=0}^{N-1} G_N(n) e^{-j\frac{2\pi}{N}nk},\ 0 \leq k \leq N - 1 \tag{6.23}$$

The pilot symbols do not need to be periodic, they can be arranged according to the channel condition dynamically with additional overhead. the number of pilots and their positions can be changed dynamically. This type estimation is called scattered estimation as shown in Figure 6.4.

6.2.5 Non-coherent Detection

Non-coherent detection uses differential modulation technique wherein the information to be transmitted is contained in the phase difference ψ_k of two successive modulation symbols x_k and x_{k-1} as in M-DPSK and DQPSK. and the received phase differences φ_k of the corresponding received symbols y_k and y_{k-1} [65].

$$\psi_k = \arg(\frac{x_k}{x_{k-1}}), \quad \varphi_k = \arg(\frac{y_k}{y_{k-1}}) \tag{6.24}$$

In non-coherent detection, there is no requirement for the channel estimation and consequently, there is no overhead introduced by the pilot symbols. Coherent detection, on the other hand, performs better than non-coherent detection since coherent detection is not affected by phase changes since training values are used to estimate the reference phases and amplitudes[66].

6.2.6 Performance

The performance results are from a system with the parameters indicated in Table 6.2.6. Perfect synchronization and no ISI is assumed. The HE channel model of the ATTC (Advanced Television Technology Center) and the Grande Alliance DTV laboratory's ensemble E model is used. The impulse response of the Rayleigh channel is given by:

Parameters	Specifications
FFT Size	1024
Number of Active Carriers(N)	128
Pilot Ratio	1/8
Guard Interval	256
Guard Type	Cyclic Extension
Sample Rate	44.1KHz
Bandwidth	17.5KHz
Signal Constellation	16QAM
Channel Model	Rayleigh Fading
Coding Rate	1/1

Table 6.1: Simulation parameters for channel estimation

$$\begin{aligned} h(n) &= \alpha(n) + 0.3162\alpha(n-2) \\ &+ 0.1995\alpha(n-17) + 0.1296\alpha(n-36) \\ &+ 0.1\alpha(n-75) + 0.1\alpha(n-137) \end{aligned} \tag{6.25}$$

In block type estimation, each block consists of a fixed number of symbols, which is 30 in the simulation. Pilots are sent in all the sub-carriers of the first symbol of each block and channel estimation is performed by using LS estimation. The channel is estimated at the beginning of the block and used for all the following symbols of the block.

In decision feedback equalizer, in addition to the block type estimation, the successive symbols after the pilots are used to calibrate the estimated values of the pilot symbols. In this way, channel can be tracked through the channel variations.

Values is estimated in either LS or Least Mean Square (LMS) method. The LMS estimator uses one tap LMS adaptive filter at each

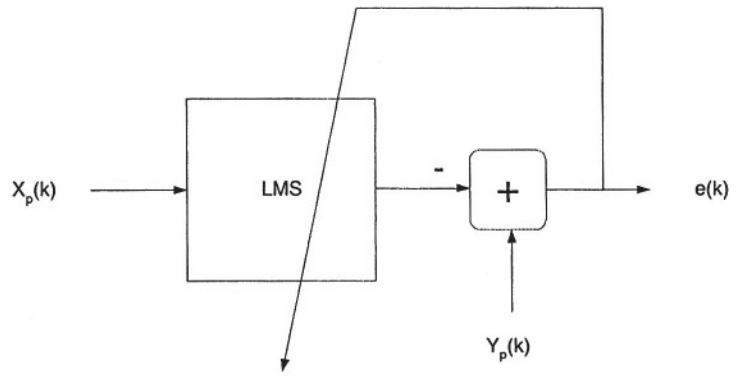

Figure 6.6: LMS scheme

pilot frequency. The first value is found directly through LS and the following values are calculated based on the previous estimation and the current channel output as shown in Figure 6.6.

All of the possible interpolation techniques (linear interpolation, second order interpolation, low-pass interpolation, spline cubic interpolation, and time domain interpolation) are applied to LS estimation result to investigate the interpolation effects and linear interpolation are applied to LMS estimation results to compare with the LS overall estimation results.

The legends "linear, second-order, low-pass, spline, time domain" denote interpolation schemes of comb-type channel estimation with the LS estimate at the pilot frequencies, "block type" shows the block type pilot arrangement with LS estimate at the pilot frequencies, "decision feedback" means the block type pilot arrangement with LS estimate at the pilot frequencies and with decision feedback, and "LMS" is for the linear interpolation scheme for comb-type channel estimation with LMS estimate at the pilot frequencies.

Figures 6.7 gives the BER performance for Rayleigh fading channel, with static channel response given in equation (6.25).

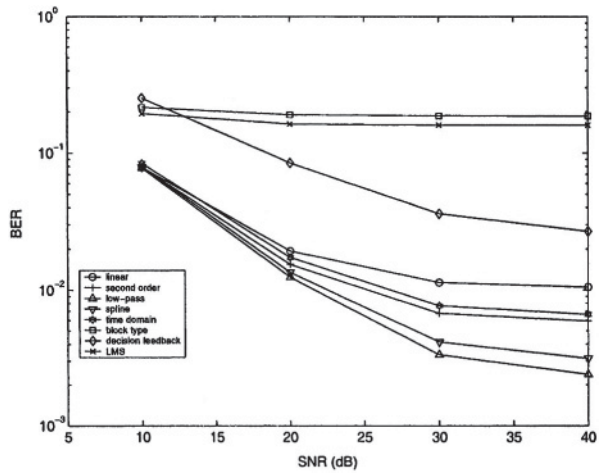

Figure 6.7: 16QAM modulation with Rayleigh fading (Doppler Freq. 70Hz)

These results show that the block-type and decision feedback estimation BER values are 10-15dB higher than that of the comb-type estimation type. This is because the channel transfer function changes so fast that there are even changes for adjacent OFDM symbols.

The comb-type channel estimation with low pass interpolation achieves the best performance among all the estimation techniques for 16QAM modulation. The performance among the comb-type channel estimation techniques usually range from the best to the worst as follows: low-pass, spline, time-domain, second-order and linear. The result was expected since the low-pass interpolation used in simulation does the interpolation such that the mean-square error rate between the interpolated points and their ideal values is minimized.

Figure 6.8 shows the performance of channel estimation methods for 16QAM modulation in Rayleigh fading channel for different Doppler frequencies. The general behavior of the plots is that BER increases as the Doppler spread increases. The reason is because of

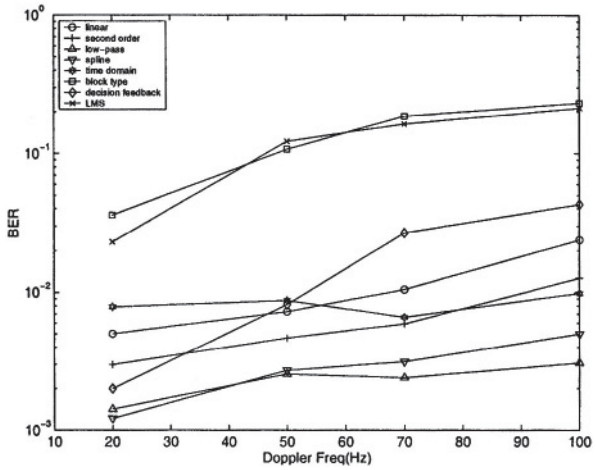

Figure 6.8: 16QAM modulation with Rayleigh fading (SNR 40dB)

the existence of severe ICI caused by Doppler shifts. Another observation from this plot is that the decision feedback block type channel estimation performs better than the comb-type based channel estimation for low Doppler frequencies except low-pass and spline interpolation. One can observe that time-domain interpolation performance improves compared to other interpolation techniques as Doppler frequency increases [52].

6.2.7 Channel Estimation for MIMO-OFDM

Multiple transmit and receive antennas are a strong choice to improve capacity in OFDM systems. MIMO systems leverage the transmitter and receiver diversity. Channel estimation is an important issue for the MIMO-OFDM systems since computational complexity increases rapidly with the increase in the number of antennas. Any estimation error becomes crucial since the system becomes more sensitive.

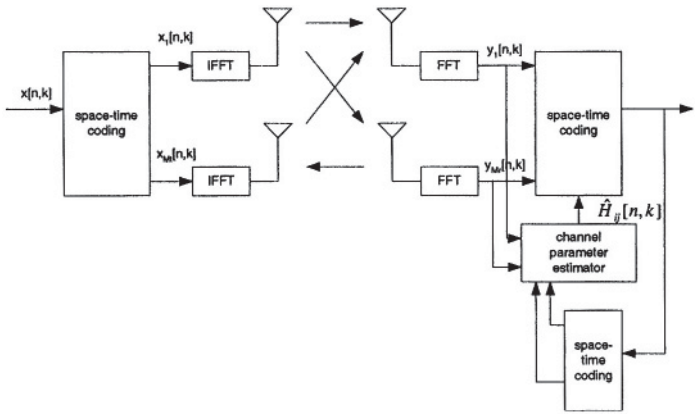

Figure 6.9: MIMO-OFDM architecture

In the estimation process of the channel in any transmitter receiver pair, the signals transmitted by other transmitters act as an interference. The challenge can be overcome by inserting pilot tones orthogonal in time, meaning whenever a pilot tone is inserted in a subcarrier, all other transmitters do not send anything in that subcarrier [73]. This method is suitable to the cases when the symbol time is larger than the coherence time. If the symbol time is smaller than the coherence time, then the correlation can be leveraged to improve the channel estimation. Assume $H_n(t)$ is the estimate of the channel at time t at the nth pilot tone. The $H_n(t+1)$ can be estimated given $H_n(t)$ and $H_n(t+1)$ by MMSE or LS estimation [62].

For a MIMO-OFDM system with M_t transmit and M_r receive antennas, there are $M_t \times M_r$ channels to be estimated. All transmitters send their signals simultaneously and the received signal is a superposition of the transmitted signals that are distorted by the channel. The received signal at the kth subcarrier of the nth block from the jth

receive antenna described in [75], [73] is expressed as:

$$y_j[n,k] = \sum_{i=1}^{M_t} H_{ij}[n,k]x_i[n,k] + w_j[n,k], \tag{6.26}$$

for $j = 1, \ldots, M_r$ and $k = 0, \ldots, K-1$, where $x_i[n,k]$ is the symbol transmitted from the ith transmit antenna at the kth subcarrier of the nth block, $H_{ij}[n,k]$ is the channel's frequency response at the kth subcarrier of the nth block corresponding to the ith transmit and the jth receive antenna, and, $w_j[n,k]$ is additive (complex) Gaussian noise that is assumed to be independent and identically distributed (i.i.d.) with zero-mean and variance σ.

The received symbol in each antenna is independent from one another. Following the notation in [75] and omitting the receiver antenna subscript j the estimated channel is as follows:

Define

$$\begin{aligned} y[n] &\triangleq (y[n,0], y[n,1], \cdots, y[n,K-1])^T, \\ \hat{D}_i &\triangleq diag\{\hat{x}_i[n,0], \cdots, \hat{x}_i[n,K-1]\}, \\ w[n] &\triangleq (w[n.0], w[n,l], \cdots, w[n,K-1])^T, \end{aligned}$$

where $\hat{x}_i[n,k]$'s are detected symbols. Let rank-χ approximation of the channel response be:

$$(H_i[n,0], \cdots, H_i[n,K-1])^T = Uh_i[n], \tag{6.27}$$

where

$$h_i[n] \triangleq (h_i[n,0], \cdots, h_i[n,\chi-1])^T,$$

and U is the rank-χ transform matrix which is

$$U = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_K & \cdots & W_K^{\chi-1} \\ \vdots & \cdots & \ddots & \vdots \\ 1 & W_K^{(K-1)} & \cdots & W_K^{(\chi-1)(K-1)} \end{pmatrix},$$

where $W_K = e^{-j\frac{2\pi}{K}}$ and equation (6.26) can be written as

$$x[n] = \sum_{i=1}^{M_t} \hat{D}_i[n] U h_i[n] + w[n]. \tag{6.28}$$

Let

$$\hat{h}[n] \triangleq (\hat{h}_1^T[n], \hat{h}_2^T[n], \cdots, \hat{h}_{M_t}^T[n]),$$

from [73], the LS estimate is given as

$$\hat{h}[n] = (T^H[n]T[n])^{-1} T^H[n] y[n] \tag{6.29}$$

where

$$T[n] = (\hat{D}_1[n]U, \hat{D}_2[n]U, \cdots, \hat{D}_{M_t}[n]U). \tag{6.30}$$

The matrix inversion introduces complexity to the system. There are proposals that use training sequences which optimize and simplify the channel estimation during the training period [74].

6.3 Equalization

One of the advantages of a multi-carrier system is its robustness against inter-symbol interference. The longer duration of OFDM symbols provides higher immunity against delay spread and ISI. As long as channel dispersion is not longer than the OFDM symbol guard interval, system performance does not degrade due to ISI and use of the time domain equalization is not usually mandated.

However, in case of higher data rates and channels with extensive time dispersion, an equalizer is unavoidable. Albeit, the structure of the equalizer is different from that of single carrier systems. The purpose of equalization is not the complete removal, but the restriction of inter-symbol interference to a tolerable extent.

In frequency domain equalization, the absence of inter-channel interference (ICI) is used to compensate for channel complex gain

at each sub-carrier frequency. Using time-frequency duality, complex gain compensation after FFT is equivalent to a convolution of a FIR filter in time domain (residual equalization).

In this section, we review main differences of equalization in single and multi-carrier and follow up with analytical and practical aspects of time domain equalization. Then, frequency domain equalization and echo cancellation are discussed. Using duality principle, we present a unified approach to time and frequency domain equalization.

6.3.1 Time Domain Equalization

A brief review of decision feedback equalizer structure and performance should help the reader to readily understand the analysis and design of equalizers for multi-carrier.

Decision Feedback Equalizers

In a decision feedback equalizer, the forward filter whitens the noise and produces a response with post-cursor ISI only, and the feedback filter then cancels that post-cursor ISI. The optimum forward filter for a zero-forcing DFE can be considered as cascade of a matched filter followed by a linear equalizer and a causal whitening filter whose transfer function can be found by spectral factorization of the channel power spectrum. The forward filter, therefore, first minimizes ISI, then noise power at the input of slicer, reintroducing causal ISI. Notice that the separation and ordering has only analytic significance [79]. In practice, all forward filters are usually lumped together in one adaptive filter.

In an MSE decision feedback equalizer, the optimum forward filter minimizes the mean-square residual ISI and noise power. In both MSE-DFE and ZF-DFE we intend to minimize the error term at the input of the slicer, which in the absence of error propagation includes

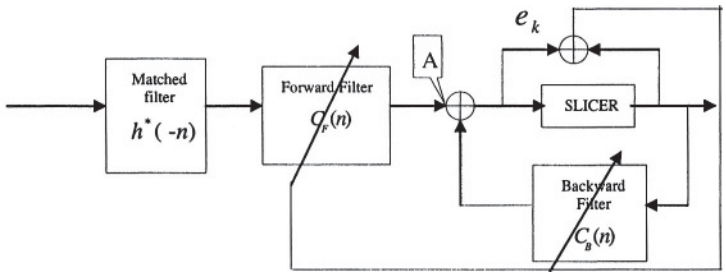

Figure 6.10: General decision feedback equalizer

residual ISI and noise. Let $h(n)$ represent the channel impulse response, hence $h^*(-n)$ is the impulse response of the matched filter. Forward filter $c_F(n)$ performs ISI suppression and noise whitening. The feedback filter $c_B(n)$ eliminates the interference of previous symbols, hence $1+c_B(n)$ has to be a monic and causal filter. Therefore, its output should have post-cursor inter-symbol interference only, namely

$$h(n) * h^*(-n) * c_F(n) = 1 + c_B(n) \tag{6.31}$$

On the other hand the variance of the noise, at the output of the matched filter is

$$\sigma^2 \int_{-\pi/2}^{\pi/2} S_h(e^{j\omega t}) d\omega \tag{6.32}$$

where S_h is power spectrum of the channel impulse response. In order to minimize noise power at the input of the slicer, $c_F(n)$ should whiten the noise. Rewriting Equation 6.31 in the frequency domain, the forward filter transfer function is

$$C_F(z) = \frac{1 + C_B(z)}{H(z)H^{*-1}(z^{-1})}. \tag{6.33}$$

The forward filter may apply further linear transformation to the input signal to meet some optimality criteria such as minimizing square

error or peak distortion. To whiten a stochastic process with spectrum of $\sigma^2 S_h(z)$ it should satisfy the following equation:

$$\sigma^2 S_h(z) \frac{1 + C_B(z)}{H(z)H^{*^{-1}}(z^{-1})} \frac{1 + C_B^*(z^{*^{-1}})}{H^*(z^{*^{-1}})H(z)} = \sigma_{DFE}^2 \tag{6.34}$$

Since

$$S_h(z) = H(z)H^{*^{-1}}(z^{-1}) \tag{6.35}$$

from Equation 6.34 we have:

$$S_h(z) = (1 + C_B(z))(1 + C_B^*(z^{*^{-1}})). \tag{6.36}$$

The same procedure [79] can be followed for the MSE equalizer to obtain

$$S_h'(z) = (1 + C_B(z))(1 + C_B^*(z^{*^{-1}})). \tag{6.37}$$

In the following, we discuss some important issues about the DFE with finite tap FIR feedback and forward filters as a preface for future discussions.

The input to the feedback filter is noise free, therefore the feedback filter eliminates the post-cursor ISI both in ZF-DFE and MSE-DFE. The error term used for training is memoryless and defined as:

$$e_k = x_k - \hat{x}_k. \tag{6.38}$$

where x_k is transmitted symbol, which is the same as the output of the slicer when there is no error and $\hat{x}_k$ is the output of the equalizer. Minimizing $|e_k|^2$ requires that the input signal be orthogonal to the error term. Since the input to the feedback filter is noise free, we design the forward filter to minimize the error power which leads to the following equations:

$$E(\varepsilon_k y_{k-i}^*) = E(x_k - \sum_j c_{F_j} y_{k-j} - \sum_{j'} c_{B_{j'}} x_{k-j'}) y_{k-i}^* = 0 \tag{6.39}$$

where $-\infty < i < \infty$ and:

$$y_k = \sum_{n=0} p_n x_{k-n} + \eta_k. \tag{6.40}$$

and

$$p_i = \sum_{n=0}^{l-i} h_n h_{n+i}^* \quad i = 0, \cdots, l$$

$$p_{-i} = p_i^*.$$

where l is channel spread in symbols.

Therefore,

$$\sum c_{F_i} E\{\sum_n p_n x_{k-n-j} \sum_{n'} p_{n'}^* x_{k-n'-i}^*\} + \sum_j c_{F_j} E\{\eta_{k-j}\eta_k^*\}$$

$$= \sum_{n=0}^{N} p_n E(x_k x_{k-n-i})$$

$$\Rightarrow \sum_j c_{F_j} \sum_n p_n p_{n+j-i}^* + N_o \sum_j c_{F_j} p_j = p_{-i}^*. \tag{6.41}$$

η_k is a colored Gaussian noise with the following auto-correlation function

$$E(\eta_i \eta_j^*) = \begin{cases} N_o p_{i-j}, & |i-j| \le l \\ 0, & otherwise \end{cases} \tag{6.42}$$

The solution to Equation 6.39 is

$$C_F = \Gamma^{-1}\theta, \tag{6.43}$$

where

$$\Gamma_{ij} = E\{y_{k+i} y_{k+j}^*\},\ 0 \le i, j < N. \tag{6.44}$$

N is the number of forward taps and

$$\theta_i = E\{x_k y_{k+i}\},\ 0 \le i < N. \tag{6.45}$$

Coefficients of the backward filter can be found by

$$C_B = P^T C_F, \tag{6.46}$$

where

$$C_F = \begin{pmatrix} c_{F_0} \\ c_{F_{-1}} \\ \\ c_{F_{-N}} \end{pmatrix} \quad C_B = \begin{pmatrix} c_{B_1} \\ c_{B_2} \\ \\ c_{B_M} \end{pmatrix} \quad P = \begin{pmatrix} p_0 & p_1 & \cdots & p_M \\ p_1 & p_2 & & p_{M+1} \\ & & & \\ p_N & & & p_{N+M} \end{pmatrix} \tag{6.47}$$

The forward filter c_F is anti-causal which means that the main tap is assumed to be the closest one to the slicer. The forward taps should remove the entire post-cursor. Practically, the matched filter is usually absorbed in the forward equalizer[1] then replaces is Equations 6.42 and 6.46. Notice that the feedback filter is designed to minimize the post-cursor ISI at the input of the slicer. In practice, c_B and c_F are adjusted adaptively.

6.3.2 Equalization in DMT

Multi-carrier systems are robust against inter-symbol interference as long as orthogonality of adjacent symbols are preserved in the frequency domain. In other words, the symbol duration of the OFDM signal is extended to beyond T. However, if channel impulse response spread is more than the prefix extension, inter-symbol interference can degrade the performance.

There are several fundamental differences between time domain equalization requirements of OFDM channels and classic decision feedback equalizers:

a) An equalizer of an OFDM system does not need to cancel the ISI entirely, but to limit its length. Unlike a single carrier system in which equalization minimizes ISI, we only need to reduce it to a time span less than the length of guard interval.

b) The channel impulse response is modelled as an ARMA system. Usually, the channel impulse response is too long. Therefore, a model of the form

$$\frac{A(z)}{1+B(z)} \tag{6.48}$$

is appropriate.

[1]We can interpret this as replacing the matched filter with a whitening matched filter.

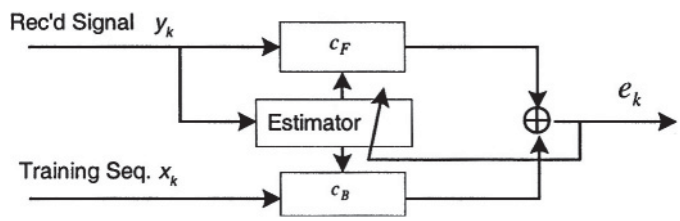

Figure 6.11: Equalizer configuration in training mode

c) The equalizer does not work as a DFE in data mode. Unless channel characteristics change rapidly, equalizer taps are adjusted during the training sequence and the same tap values are preserved during data mode.

d) The feedback filter is not in the loop and need not to be monic or causal. In single carrier systems, the slicer output is fed to the feedback filter to remove the post-cursor effect. However, truncating the impulse response does not require a monic feedback filter. A typical structure is shown in Figure 6.11

e) The error is not memoryless. The error term used in typical DFE is derived from the scalar input and output of the slicer. In a DMT equalizer the error is derived from the output of an impulse shortening filter and reference filter.

f) The location of the impulse response window has a big impact on the performance of equalizer.

The overall structure of the equalizer in training mode is shown in the following diagram:

In steady state mode, the equalizer will include only c_F . The error term is

$$e_k = c_F * Y_k - c_B * X_k \tag{6.49}$$

where X_k and Y_k are vectors of training and received samples respectively. In general, we can use adaptive techniques to find near optimum equalizer coefficients. A large number of taps prevents us from using RLS type of algorithms. On the other hand, a wide spread of eigenvalues of the input signal covariance matrix can slow down convergence of the equalizer. A different technique for the equalizer tap adjustment is based on the linear prediction theory as explained below.

During acquisition, the frame is repeated and can be used for equalizer training. In this approach, we use second order statistics of the channel to produce a one shot estimate of equalizer coefficients. In order to use the predictor format of linear regression techniques, without loss of generality, assume c_F is monic. So, the predictor format is:

$$\hat{y}(n|n-1) = \varphi^T(n)C \tag{6.50}$$

where

$$\begin{aligned} \varphi(n) \quad &= [-y(n-1) \\ &\quad -y(n-2) \\ &\quad \vdots \\ &\quad -y(t-N) \\ &\quad x(n-1) \\ &\quad x(n-2) \\ &\quad \vdots \\ &\quad x(n-M)] \end{aligned} \tag{6.51}$$

and,

$$C = [\ C_F \quad C_B\ ]^T. \tag{6.52}$$

The least square solution of linear regression is:

$$\hat{C} = \Gamma^{-1}\theta \tag{6.53}$$

with:

$$\Gamma = \begin{pmatrix} A_{N\times N} & B_{N\times M} \\ B^*_{M\times N} & I_{M\times N} \end{pmatrix}, \tag{6.54}$$

where

$$A_{ij} = R_{yy}(i-j) = \frac{1}{N}\sum_{n=1}^{N} y(n-i)y(n-j), \tag{6.55}$$

and

$$B_{ij} = R_{xy}(i-j) = \frac{1}{N}\sum_{n=1}^{N} x(n-i)y(n-j), \tag{6.56}$$

and

$$\theta = [\ D_N \quad E_M\ ]^T \tag{6.57}$$

with

$$D_i = R_{yy}[i],\ E_i = R_{yx}[i], \tag{6.58}$$

where R_{yy} and R_{xy} are auto-correlation and cross correlation functions. We can use channel impulse response to calculate coefficients directly

$$A_{ij} = \sum_n h_n h^*_{n+j-i} \tag{6.59}$$

$$B_{ij} = h^*_{-i+j}. \tag{6.60}$$

The estimation of channel impulse response can help in choosing the optimum window location. As explained previously, c_B is not necessarily causal, therefore, delay parameter plays an important role in the performance of the equalizer. Other important parameters are the lengths of the two filters c_F and c_B.

6.3.3 Delay Parameter

Unlike classic decision feedback equalizers, an ARMA equalizer can have a feedback filter with non-causal transfer function i.e. the general error term is:

$$c_F * Y_k - c_B * X_{k-d} \tag{6.61}$$

the delay unit d has significant effect on overall performance. The number of taps M of filter c_B is approximately the same as the prefix length and number of taps for the forward filter $N \leq M$. The delay unit should be chosen properly to capture the most significant window of the impulse response. Brute force trial and error is one option. A more intelligent technique uses the estimate of the impulse response to pick a proper starting point where the energy of the impulse response is above a threshold. The threshold should be the ratio of the tap power to the overall window power. Since the length of the window function and the forward equalizer should be as short as possible, the proper choice of starting point is of significant importance. In other words, in an ARMA model, the number of zeroes are not assigned properly.

6.3.4 AR Approximation of ARMA Model

The Direct solution of Equation 6.53 requires a non-Toeplitz matrix inversion and is numerically sensitive and intensive. One technique to alleviate the numerical difficulty of least square solution is by using techniques for AR approximation of the channel response ARMA model, which results in a numerically attractive Toeplitz correlation matrix. A description of embedding techniques is presented in the Appendix. In the following, the procedure for this analysis is briefly described. By using embedding techniques, as explained in the Appendix, the ARMA model of Equation 6.48 is approximated by a two-channel AR model.

If the covariance matrix is available the problem can be reduced to the solution of a normal (Yule-Walker) equation. The multi-channel version of the Levinson algorithm is a computationally efficient solution for the two-channel AR model at hand. As a reference, the recursive technique is presented here:

Initialization:

$$\begin{aligned} C_1^f[1] &= -R^{-1}[0]R[1] \\ C_1^b[1] &= -R^{-1}[0]R[-1] \end{aligned} \tag{6.62}$$

where

$$C_i^f[j] = \begin{bmatrix} -a_i^j & b_i^j \\ 0 & 0 \end{bmatrix}$$

is the jth forward prediction coefficient of two-channel model at i^{th} iteration and $C_i^b[j]$ is the jth backward prediction coefficient at i^{th} iteration and

$$R[i] = \begin{bmatrix} R_{yy}[i] & R_{yx}[i] \\ R_{xy}[i] & R_{xx}[i] \end{bmatrix}.$$

If the input signal x is white then

$$R[i] = \begin{bmatrix} R_{yy}[i] & R_{yx}[i] \\ 0 & 0 \end{bmatrix}$$

with

$$R[0] = \begin{bmatrix} R_{yy}[i] & R_{yx}[i] \\ R_{xy}[i] & I \end{bmatrix}$$

Recursion:

$$\begin{aligned} C_i^f[i] &= K_i^f \\ C_k^f[i] &= C_{k-1}^f + K_k^f[i]C_{k-1}^b[k-i] \\ C_i^b[i] &= K_i^b \\ C_k^b[i] &= C_{k-1}^b + K_k^b[i]C_{k-1}^f[k-i] \\ C_k^b[i] &= C_{k-1}^b + K_k^b[i]C_{k-1}^f[k-i] \end{aligned} \tag{6.63}$$

where the reflection coefficients are,

$$\begin{aligned} K_k^f &= -\Delta_k^{f^T} \textstyle\sum_{k-1}^{b^{-T}} \\ K_k^b &= -\Delta_k^{f^T} \textstyle\sum_{k-1}^{b^{-T}} \\ \Delta_k^f &= \textstyle\sum_{i=0}^{k-1} R[k-i]C_{k-1}^{f^T}[i]. \end{aligned} \tag{6.64}$$

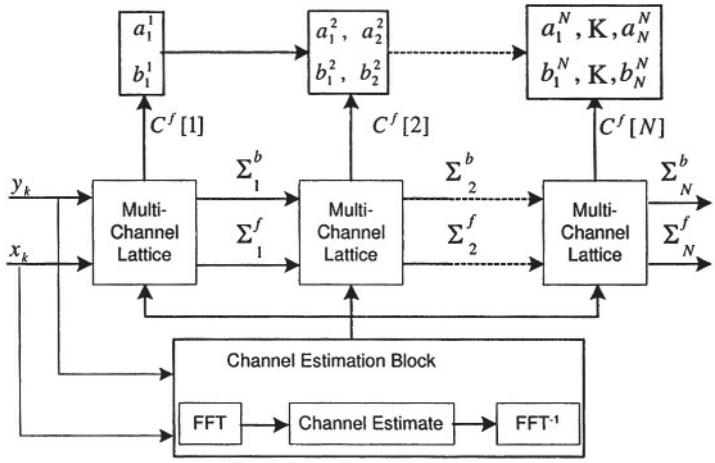

Figure 6.12: Equalization using order update

Figure 6.12 shows the structure of an equalizer using a lattice filter. This structure represents the case of equal number of poles and zeros. The extension to unequal number of poles and zeroes is straightforward and explained in [79].

The above approach is convenient for batch or frame based processing where the entire frame is buffered and processed for equalization. Instead of inverting the large matrix of Equation 6.43 for example, this structure can be used in the training mode of an ADSL receiver. This technique can be implemented iteratively in time and order when the covariance is not known in advance. The time/order updating algorithm increases the order while updating the coefficients of the equalizer. In some cases, we can preserve the order and continue updating the parameters using adaptive techniques. This technique is useful when the delay profile of the channel does not change significantly but some slow adaptation improves the performance.

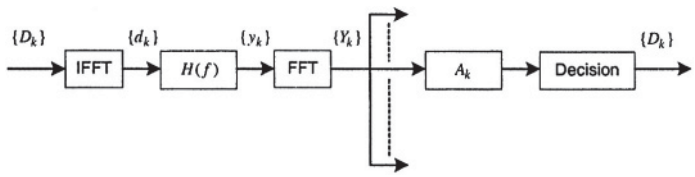

Figure 6.13: An OFDM system with frequency domain equalization

6.3.5 Frequency Domain Equalization

Once it is assured that orthogonality of the sub-carriers is maintained, possibly through the use of the cyclic prefix and time domain equalization as previously discussed, then the final frequency domain equalization of an OFDM signal is an extremely simple process. This is certainly one of the key advantages of OFDM.

After demodulation, the sub-carriers will be subjected to different losses and phase shifts, but there will be no interaction among them. Frequency domain equalization therefore consists solely of separate adjustments of sub-carrier gain and phase, or equivalently of adjusting the individual decision regions. In the case where the constellations consist of equal amplitude points, as in PSK, this equalization becomes even simpler in that only phase needs to be corrected for each sub-carrier, because amplitude has no effect on decisions.

A simplified picture of the place of frequency domain equalization is shown in Figure 6.13, where the equalizer consists of the set of complex multipliers, $\{A\}$, one for each sub-carrier.

Here the linear channel transfer function $H(f)$ includes the channel, the transmit and receive filters, and any time domain equalization if present. $H(f)$ is assumed bandlimited to less than N/T for a complex channel. Cyclic extension is not shown although it is almost certain to be present. The following analysis assumes that both amplitude and phase need to be corrected, and that equalization consists

of multiplying each demodulated component by a quantity such that a fixed set of decision regions may be used.

The signal presented to the demodulator is

$$y(t) = \sum_{n=0}^{N-1} d_n h(t - n\frac{T}{N}). \tag{6.65}$$

The kth output of the demodulator is then

$$Y_k(f) = D_k H_k, \; k = 0, 1, \cdots, N-1, \tag{6.66}$$

where the H_k are samples of $H(f)$

$$H_k = h(\frac{k}{T}). \tag{6.67}$$

Thus each output is equal to its associated input data symbol multiplied by a complex quantity which differs among the outputs, but are uncoupled. Equalization at its simplest then consists of setting the multipliers to $1/H_k$ for each non-zero channel.

The above approach is optimum in every sense under high signal-to-noise conditions. It also produces minimum probability of error at any noise level, and is an unbiased estimator of the input data D_k. However if the criterion to be optimized is minimum mean-square error particularly, then the optimum multipliers are modified to

$$A_k = \frac{1}{H_k} \frac{1}{1 + \frac{\sigma_k^2}{|D_k H_k|^2}} \tag{6.68}$$

where σ_k^2 is the noise power in the demodulated sub-channel. However, this value produces a biased estimator and does not minimize error probability.

As a practical implementation issue, for variable amplitude constellations, it is frequently desirable to have a fixed grid of decision

regions. The A_ks can then be scaled in amplitude such that the separation of constellation points is constant.

Since a shift τ in timing phase is equivalent to a phase shift

$$H_k = e^{j2\pi \frac{k}{T}\tau} \tag{6.69}$$

frequency domain equalization readily corrects for such timing shift.

In principle frequency domain equalization could be employed when orthogonality is lost due to interference among OFDM symbols. In this case, rather than a simple multiplier per sub-channel, a matrix multiplication would be required. This approach is bound to require more computational load than the combination of time and frequency domain equalizers.

During system initialization, the any time domain equalizer must be adjusted before frequency domain equalization is performed. Then any periodic test signal with full frequency content, such as a repeated segment of a PN sequence without cyclic prefix, may be used to adapt the frequency domain equalizer. The demodulator outputs should be averaged over many periods to minimize the effects of noise, and the reciprocal of these measured values used to set the multipliers. During steady state data operation adaptation may be performed using a single variable LMS algorithm as shown in Figure 6.14. This LMS algorithm will adapt to the values given by Equation 6.68.

An interesting interpretation of time domain and frequency domain equalizer can be obtained by studying their role in channel distortion compensation. As discussed earlier, a typical channel model for DMT and channels with long impulse response is an ARMA model of the form in Equation 6.48. Time domain equalization shortens the impulse response to a tolerable level for an OFDM system. Mathematically, it's equivalent to compensating the AR part of the channel impulse response $1/(1 + B(z))$. So after successful time domain equalization, the equivalent impulse response of the channel is reduced to a FIR filter of short duration $A(z)$. Since it does not violate orthogonal-

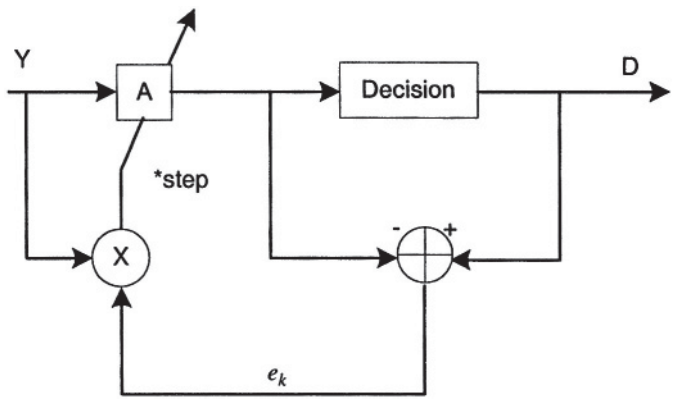

Figure 6.14: LMS adaptation of a frequency domain equalizer multiplier

ity of sub-carriers, we can remove its effect after the FFT by frequency domain equalization. The above process is shown in Figure 6.15.

6.3.6 Echo Cancellation

In wireline applications, it is usually highly desirable to communicate full duplex (simultaneous transmission in both directions) over a single wire-pair. A classic sub-system known as the hybrid provides partial separation, but is not sufficient for high performance digital applications.

Two possible techniques are frequency division and time division duplexing. In the former, different frequency bands are used in each direction, with analog filtering to provide separation. In the latter, alternating time slots are used to transmit alternately in each direction. For symmetric communication, that is equal rates for the two directions, both of these techniques require at least double the bandwidth

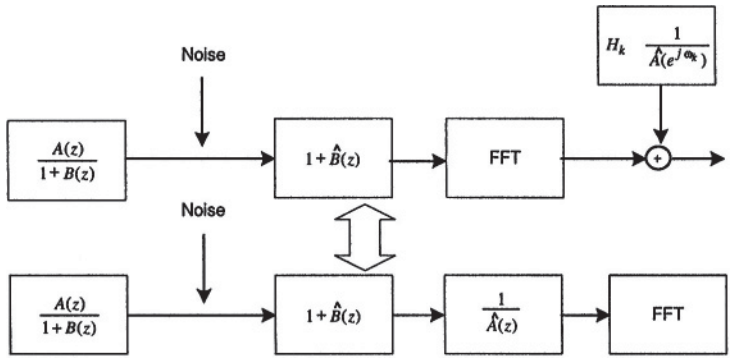

Figure 6.15: Time and frequency domain equalization

as unidirectional transmission, which is a very severe performance penalty.

Far preferable is the technique of echo cancellation, where at each end the leakage of transmitted signal into the co-located receiver is cancelled by creating and subtracting a replica of that leakage. The precision required is quite high, as illustrated in the example of a bad case wireline system shown in Figure 6.16

In this example, the loss of the line is $40\ dB$ while the hybrid provides only $15\ dB$ of loss. The leakage of the transmitted signal, or "echo", is therefore $25\ dB$ higher than the desired received signal. One consequence is that the level which the front end of the receiver, in particular the A/D converter, must accommodate is dominated by the echo rather than the received signal. Of more interest here is the high degree of echo cancellation required. Assume, as in the figure, that a signal-to noise ratio of $30\ dB$ is required, and that uncancelled echo can be treated as noise. If we use the reasonable design rule that the uncancelled echo should be at least $10\ dB$ below the line noise, thereby increasing the noise by less than $0.4\ dB$, the level of the uncancelled echo must be below -80 dB. Because the hybrid provides only $15\ dB$ of reduction, the echo canceller must provide an additional

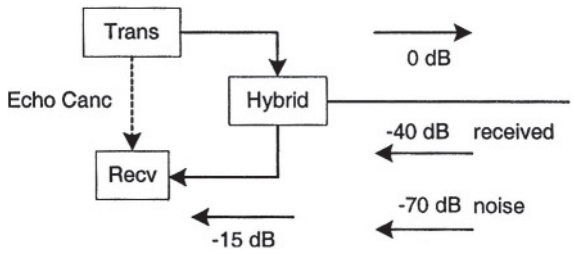

Figure 6.16: An example system illustrating echo cancellation requirement

65 dB of rejection. This requirement is far more severe than that of equalization, so that a much greater arithmetic precision and number of taps is required.

In principle, the same form of echo canceller used in single carrier systems could be used in OFDM. This is shown in Figure 6.17.

Residual echo is minimized when the tap coefficients of the echo canceller are equal to the corresponding samples of the echo path response $p_k = p(kT)$, where the sampled echo is

$$r_n = r(nT) = \sum_k d_k p_{n-k} \tag{6.70}$$

Adaptation is similar to that of equalization. Residual error results from echo samples beyond the time span of the canceller, and also from finite precision in signal sampling and in the coefficients. Unlike the case of single carrier equalization, the inputs to the canceller are samples of signals quantized by an *A/D* converter as opposed to exact data symbols. In one variation [80], adaptation is performed in the frequency domain, followed by an IFFT and time domain cancellation.

In any variation, time domain cancellation requires extensive computation, reducing a prime advantage of OFDM. Another possibility is

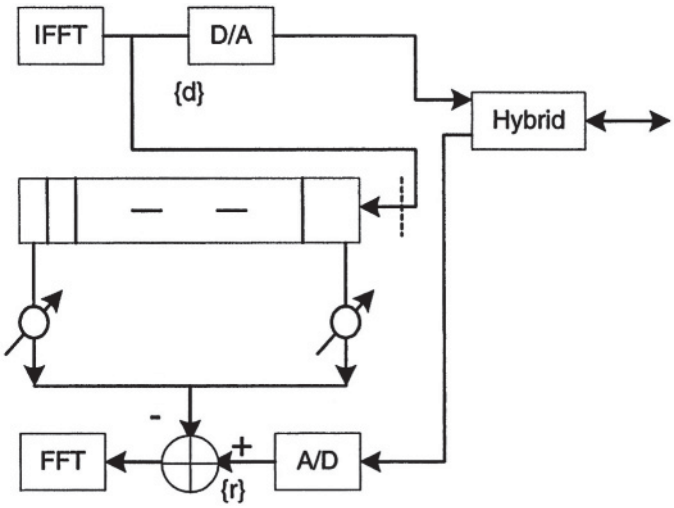

Figure 6.17: A basic echo canceller

frequency domain cancellation. If the echo of the transmitted OFDM signal maintains orthogonality, then echo cancellation may be performed very easily after demodulation in the receiver. An advantage of this approach is the use of exact data values as inputs to the canceller rather than quantized signal samples. Furthermore, orthogonality leads to simple adaptation. The canceller consists of one complex multiplication per sub-carrier, as in the frequency domain equalizer. In order for this technique to be valid, there must be no inter-symbol interference among OFDM signals. Otherwise a matrix multiplication is required rather than a single multiply per sub-carrier [81]. In another approach described in Reference [82], the time domain equalization is modified so as to shorten both the channel response and the echo response to a time span less than the duration of the cyclic prefix. The resulting equalizer and echo canceller can then both be performed quite simply in the frequency domain as shown in Figure 6.18.

Techniques have been proposed for echo cancellation in OFDM

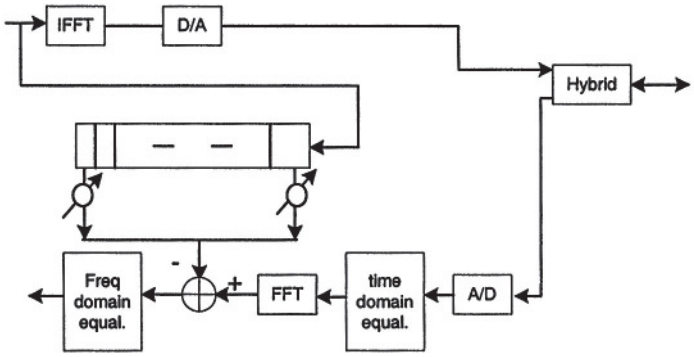

Figure 6.18: Frequency domain echo cancellation

that combine time and frequency domain components in ways to reduce total required computation [82], [84]. Shown in Figure 6.19 is the approach of Reference [82] for a symmetric system.

The time domain canceller is a short one if the echo response is not much longer than the duration of the cyclic prefix. Its functions are to undo the effect of cyclic extension, and to cancel the overlap from the previous symbol. The frequency domain canceller can now deal with orthogonal symbols as before.

In asymmetric applications such as ADSL, the penalty for using frequency division duplexing is not as great because the bandwidth used in one direction is relatively small. However for optimum performance, the use of overlapping bands and echo cancellation is employed. Shown in Figure 6.20 and Figure 6.21 are modifications of the previous structure [82] when the same OFDM symbol rate but different numbers of sub-carriers are used in each direction, with overlap of the frequencies of some of the sub-carriers.

Figure 6.11 illustrates a case in which the transmit rate is lower than the receive rate, as in the remote terminal of ADSL. Because the transmit spectrum is smaller than the receive spectrum, it is replicated

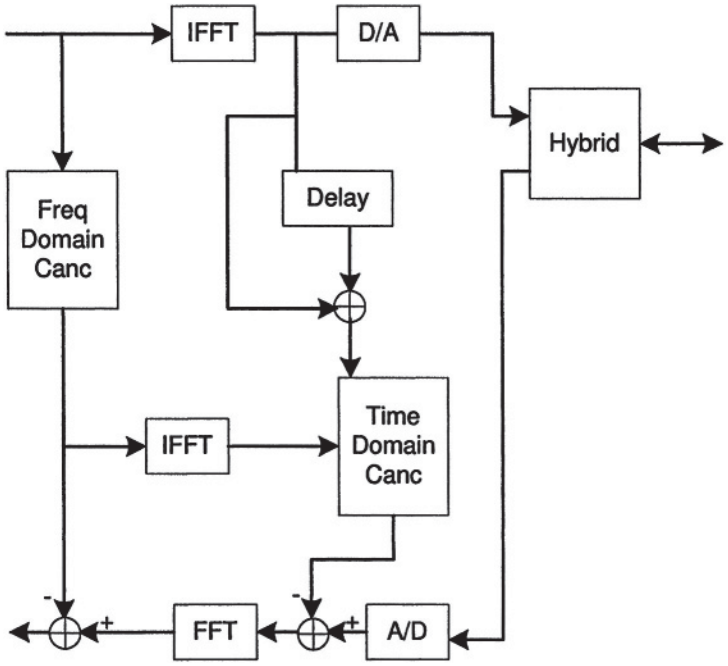

Figure 6.19: Combination echo cancellation for symmetric transmission

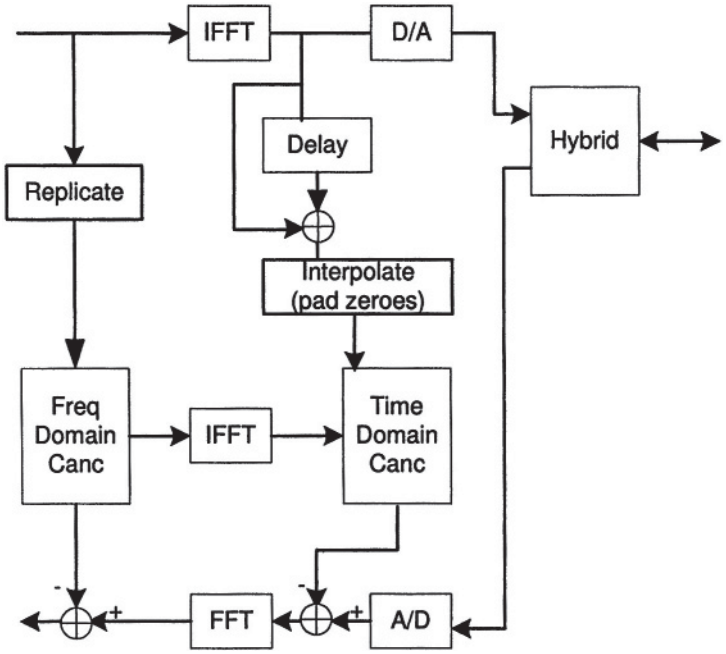

Figure 6.20: Echo cancellation where the transmit rate is lower than receive rate

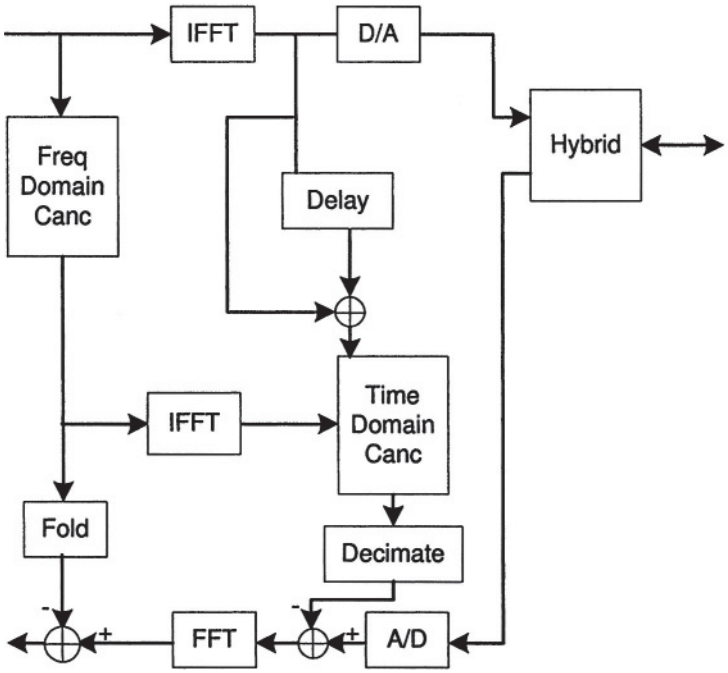

Figure 6.21: Echo cancellation where the receive rate is lower than transmit rate

so that the number of sub-carriers is increased to equal that of the received signal. For the time domain portion of the echo canceller, the transmit signal is interpolated to the higher rate by inserting the appropriate number of zeroes between the time samples.

The converse case, in which the transmit rate is higher than the receive rate, as in the central office terminal of an ADSL link, is shown Figure 6.21. The size of the input to the frequency domain canceller is reduced by folding subsets of the OFDM symbol on top of each other and adding them to produce a number of sub-carriers equal to the number received. The time domain sampling rate is reduced by computing the short echo at further spaced points.

Initial adaptation, for any echo cancellation structure, may be performed at the same time and with the same training signals as the equalization at the other end. Having the remote transmitter silent improves this initial adaptation.

6.3.7 Appendix - Joint Innovation Representation of ARMA Models

Some of the least squares estimation algorithms developed for AR models such as ladder form realization, can be extended to ARMA models by using embedding technique. Namely, an ARMA model can be embedded into a two channel AR model of joint input/output process [85]. The same approach can be used to model an equalizer for an ARMA channel by extending the results of the equalizer design for AR channels. Unlike AR models, the covariance matrix of an ARMA model is not Toeplitz and that results in some numerical computational difficulties. Therefore, by using embedding techniques we can transform an ARMA model to a multi-channel AR model with a block-Toeplitz covariance matrix.

First, we present a geometric interpretation of prediction techniques for ARMA processes which is the basis for understanding Levinson Algorithm.

In linear prediction theory, we predict x_n by $\hat{x}_n^m$, a linear combination of previous m samples $\{x_{n-1}, x_{n-2}, \cdots, x_{n-m}\}$ as a mth order linear predictor, such that an error criteria like

$$E\{|x_n - \hat{x}_n^m|^2\} \tag{6.71}$$

is minimized. The concept of an innovation process is a critical one in prediction and estimation. It represents the amount of new information obtained by observing a new sample x_n which was not available in previous observations:

$$v_n = x_n - x_{n|n-1} \tag{6.72}$$

Estimation of sample x_n using the previous m samples leads to an mth order forward linear predictor

$$\hat{x}^m_{n|n-1} = -\sum_{k=1}^{m} c[k]x_{n-k} \tag{6.73}$$

while using future m samples leads to a backward m^{th} order linear predictor

$$\hat{x}^m_{n|n-1} = -\sum_{k=1}^{m} c[k]x_{n+k} \tag{6.74}$$

Error terms for backward and forward predictors represent an innovation process.These error terms are defined as:

$$\begin{aligned} e^m_f[n] &= x_n - \textstyle\sum_{k=1}^{m-1} c[k]x_{n-k} \\ e^m_b[n] &= x_{n-m} - \textstyle\sum_{k=1}^{m-1} c[k]x_{n-m+k}. \end{aligned} \tag{6.75}$$

A recursive implementation of linear predictors utilizes Gram-Schmidt orthogonalization to transform samples into orthogonal random variables [86]

By interpretation of the expression in Equation 6.71 as a metric of the Hilbert space spanned by process x, a useful geometrical representation for linear prediction can be offered as shown in Figure 6.22. Next we briefly review the Levinson algorithm which is the origin of important concepts such as lattice filters and equalizers. Originally, the Levinson algorithm was proposed as an efficient numerical technique for inverting Toeplitz matrices.

Its order-recursive nature leads to important concepts in estimation and equalization. Minimizing the error criteria 6.71 for the predictor of order results in the normal equation

$$\Phi_m c_m = \phi_m \tag{6.76}$$

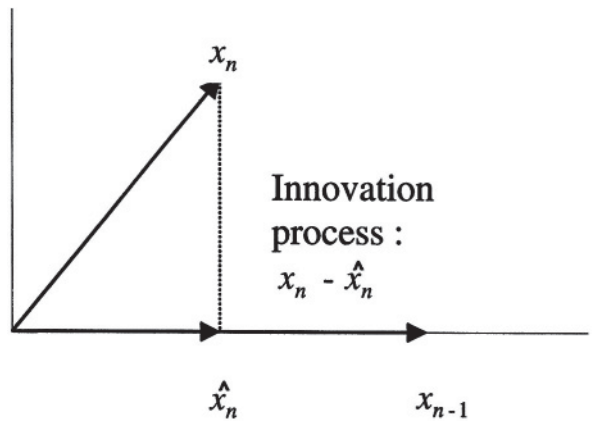

Figure 6.22: Geometric interpretation of linear predictors

where Φ_m is Toeplitz matrix of correlations $\Phi_{kl} = \phi(k-l)$ with

$$\phi(n) = E\{x_l x_{l+n}\} \tag{6.77}$$

The Levinson algorithm calculates the n coefficients of nth order linear predictor using $n-1$ coefficients of $(n-1)^{th}$ order linear predictor. As a reference we provide the scalar Levinson equations:

Initialization:

$$c_1[1] = \frac{-\phi[1]}{\phi[0]}, \; e_0 = \phi[0] \tag{6.78}$$

Recursion:

$$\begin{aligned} c_k[k] &= -\frac{\phi[k]+\sum_{l=1}^{k-1} c_{k-l}[l]\phi[k-l]}{\varepsilon_{k-1}} \\ c_k[i] &= c_{k-1}[i] + c_k[k]c^*_{k-1}[k-i] \\ e_k &= (1-|c_k[k]|^2)e_{k-1} \end{aligned} \tag{6.79}$$

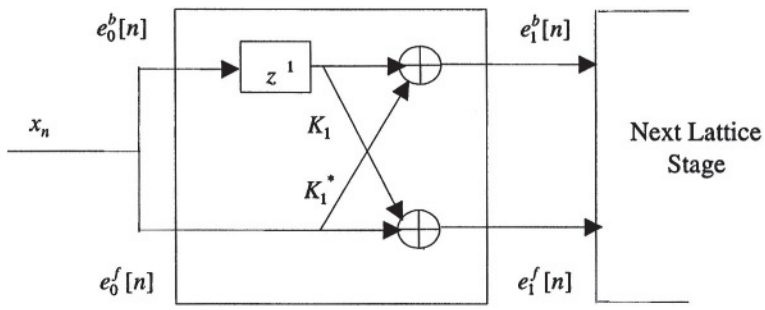

Figure 6.23: Lattice filter representation of predictors

where $c_k[i]$ represents i^{th} coefficient of the kth order predictor. In the above equations, we did not explicitly distinguish between backward and forward coefficients because optimal backward coefficients are reversed in time and complex conjugate of optimal forward predictor coefficients.

When the exact order of the predictor is unknown, we can monitor the error term e_k after each iteration. An extensive treatment of this issue is discussed in model order estimation theory.

The above order recursive technique is the basis of lattice representation of prediction filter as shown in Figure 6.23.

In general, an ARMA model can be represented as:

$$y_k + b_1 y_{k-1} + \cdots + b_N y_{k-N} = a_0 + a_1 x_k + \cdots + a_N x_{k-N} \quad (6.80)$$

where x_k is the input process and y_k represents the output process. The innovation process of y_k which represents additional information obtained from observation of sample y_k is:

$$e_{k,N} \equiv y_k - \hat{y}_{k|k-1} = \begin{bmatrix} B_N & -A_N \end{bmatrix} \begin{bmatrix} Y_N \\ X_N \end{bmatrix} \quad (6.81)$$

where

$$\begin{aligned} X_N^T &= [x_k, \cdots, x_{k-N}] \\ Y_N^T &= [y_k, \cdots, y_{k-N}] \\ A_N^T &= [a_0, \cdots, a_N] \\ B_N^T &= [1, b_1, \cdots, b_N] \end{aligned}$$

The problem of linear prediction is: given a set of observed samples of $\{y_{n-1}, y_{n-2}, \cdots, y_{n-m}, x_{n-1}, \cdots, x_{n-m}\}$ an estimate of unobserved sample y_n is required, usually without covariance information. The prediction filter is chosen to minimize the power of prediction error which means designing a filter of the form

$$H_N(z) = \frac{a_0 z^N + a_1 z^{N-1} + \cdots + a_N}{z^N + b_1 z^{N-1} + \cdots + b_N} \tag{6.82}$$

such that

$$y(z) = H_N(z) e_{k,N}(z) \tag{6.83}$$

where $e_{k,N}$ is as close to a white noise process as possible.

One approach to linear prediction of ARMA models is to introduce an innovation process for input x_k, then embed it into a two-channel AR process. The general approach for embedding is discussed extensively in the literature. Here we discuss the embedding technique for the special case where x_k is white. By introducing the augmented process $w_k = [\ y_k \quad x_k\ ]$, the new prediction error is

$$e_{k,N} = C_N w_k, \tag{6.84}$$

where

$$e_{k,n} = \begin{bmatrix} e_N^x \\ e_N^y \end{bmatrix} = \begin{bmatrix} a_0 x_k \\ x_k \end{bmatrix} \tag{6.85}$$

and

$$C_N = \left[\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \begin{bmatrix} b_1 & -a_1 \\ 0 & 0 \end{bmatrix}, \cdots, \begin{bmatrix} b_N & -a_N \\ 0 & 0 \end{bmatrix} \right] \tag{6.86}$$

Geometric interpretation of the ARMA model using the embedded technique is straightforward and useful. In the ARMA model the

space of observations is spanned by augmented process $[x, y]$. Then the joint backward and forward innovations can be interpreted as orthogonal projections on the sub-spaces spanned by future and past observation respectively.

The Levinson recursion algorithm can be applied to the augmented (multi-channel) system as well. However, unlike the scalar case, backward and forward predictor coefficients are not time-reversed and complex conjugate versions of each other.