OFDM Baseband Receiver Design for Wireless Communications

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0001

OFDM Baseband Receiver Design for Wireless Communications

Tzi-Dar Chiueh
National Taiwan University, Taiwan
Pei-Yun Tsai
National Central University, Taiwan

John Wiley and Sons (Asia) Pte Ltd

Petitioner Samsung Ex-1026, 0002

Copyright © 2007 John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop, # 02-01, Singapore 129809, tel: 65-64632400

Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as expressly permitted by law, without either the prior written permission of the Publisher, or authorization through payment of the appropriate photocopy fee to the Copyright Clearance Center. Requests for permission should be addressed to the Publisher, fax: 65-64646912, email:enquiry@wiley.com.sg

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book. All trademarks referred to in the text of this publication are the property of their respective owners.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, UK

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data

Chiueh, Tzi-Dar, 1960-
OFDM Baseband receiver design for wireless communications / Tzi-Dar Chiueh, Pei-Yun Tsai.
p. cm.
Includes index.
ISBN 978-0-470-82234-0 (cloth)
1. Wireless communication systems. I. Tsai, Pei-Yun. II. Title.
TK5103.2.C4657 2007
621. 384–dc22 2007020124

ISBN 978-0-470-82234-0 (HB)

Typeset in 10/12pt Times by Thomson Digital Noida, India
Printed and bound in Singapore by Markono Print Media Pte Ltd
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

To my Dad Chin-Mu, my wife Jill, my daughter Joanne, and my son Kevin.

–Tzi-Dar Chiueh

To my families for their constant encouragement and support.

–Pei-Yun Tsai

Petitioner Samsung Ex-1026, 0004

Contents

Preface **xi**

About the Authors **xv**

Acknowledgements **xvii**

1. Introduction **1**

1.1 Wireless Communication Systems 1
1.1.1 Digital Broadcasting Systems 1
1.1.2 Mobile Cellular Systems 3
1.1.3 Wireless Network Systems 5
Bibliography 8

2. Digital Modulation **9**

2.1 Single-Carrier Modulation 9
2.1.1 Power Spectral Densities of Modulation Signals 11
2.1.2 PSK, QAM and ASK 11
2.1.3 CPFSK and MSK 14
2.1.4 Pulse Shaping and Windowing 15
2.2 Multi-Carrier Modulation 17
2.2.1 Orthogonal Frequency-Division Multiplexing 19
2.2.2 OFDM-Related Issues 19
2.2.3 OFDM Transceiver Architecture 24
2.2.4 OFDM System Examples 26
Bibliography 28

3. Multiple Access and Error-Correcting Codes **31**

3.1 Multiple Access 31
3.1.1 Frequency-Division Multiple Access (FDMA) 31
3.1.2 Time-Division Multiple Access (TDMA) 31
3.1.3 Code-Division Multiple Access (CDMA) 33
3.1.4 Carrier Sense Multiple Access (CSMA) 34
3.2 Spread Spectrum and CDMA 35
3.2.1 PN Codes 35
3.2.2 Direct-Sequence Spread Spectrum 38
3.2.3 Frequency-Hopping Spread Spectrum 40
3.3 Error-Correcting Codes 41
3.3.1 Block Codes 42
3.3.2 Reed-Solomon Codes 45
3.3.3 Convolutional Codes 48
3.3.4 Low-Density Parity-Check Codes 53
Bibliography 56

4. Signal Propagation and Channel Model **59**

4.1 Introduction 59
4.2 Wireless Channel Propagation 59
4.2.1 Path Loss and Shadowing 60
4.2.2 Multipath Fading 61
4.2.3 Multipath Channel Parameters 62
4.3 Front-End Electronics Effects 68
4.3.1 Carrier Frequency Offset 68
4.3.2 Sampling Clock Offset 69
4.3.3 Phase Noise 70
4.3.4 IQ Imbalance and DC Offset 70
4.3.5 Power Amplifier Nonlinearity 73
4.4 Channel Model 74
4.4.1 Model for Front-End Impairments 75
4.4.2 Multipath Rayleigh Fader Model 77
4.4.3 Channel Models Used in Standards 78
Bibliography 84

5. Synchronization **85**

5.1 Introduction 85
5.2 Synchronization Issues 86
5.2.1 Synchronization Errors 86
5.2.2 Effects of Synchronization Errors 86
5.2.3 Consideration for Estimation and Compensation 90
5.3 Detection/Estimation of Synchronization Errors 91
5.3.1 Symbol Timing Detection 91
5.3.2 Carrier Frequency Offset Estimation 100

Petitioner Samsung Ex-1026, 0006

5.3.3 Residual CFO and SCO Estimation 104
5.3.4 Carrier Phase Estimation 107
5.4 Recovery of Synchronization Errors 107
5.4.1 Carrier Frequency Offset Compensation 108
5.4.2 Sampling Clock Offset Compensation 110
Bibliography 113

6. Channel Estimation and Equalization 115

6.1 Introduction 115
6.2 Pilot Pattern 116
6.3 Pilot-Based Channel Estimation 118
6.3.1 Channel Estimation by Block-Type Pilot Symbols 122
6.3.2 Channel Estimation by Comb-Type Pilot Symbols 125
6.3.3 Channel Estimation by Grid-Type Pilot Symbols 132
6.4 Adaptive Channel Estimation 137
6.5 Equalization 138
6.5.1 One-Tap Equalizer 138
6.5.2 Multiple-Tap Equalizer 138
Bibliography 140

7. MIMO Techniques 143

7.1 Introduction 143
7.2 MIMO Basics 145
7.2.1 Capacity 145
7.2.2 Diversity 146
7.3 MIMO–OFDM 147
7.3.1 MIMO Pilot Pattern 149
7.3.2 MIMO–OFDM Synchronization 151
7.3.3 MIMO–OFDM Channel Estimation 155
7.4 MIMO Encoding and Detection 158
7.4.1 Space Block Codes 158
7.4.2 Spatial Multiplexing 161
7.4.3 Spatial De-correlation 167
Bibliography 170

8. From Algorithm to Bit-True Design 171

8.1 Design Flow Overview 171
8.2 Effect of Additive Impairment Sources 174
8.3 Analog-to-Digital Conversion 176
8.3.1 ADC Distortions 176
8.3.2 Signal Probability Distributions 178
8.3.3 Dynamic Range and Precision Setting 180
8.4 Finite Precision Effect in Digital Baseband Processing 181
8.4.1 Fixed-Point Data Format 182
8.4.2 Fixed-Point Error Model 182

Petitioner Samsung Ex-1026, 0007

8.4.3 Finite Precision Effect in FIR Filters 187
8.4.4 Finite Precision Effect in IIR Filters 188
8.5 Conversion from Floating-Point Design to Bit-True Design 189
8.5.1 Metrics for Performance Evaluation 190
8.5.2 Interpolative Design Flow 190
8.5.3 Simulation-Based Approaches 192
8.5.4 Analytical Approaches 192
Bibliography 194

9. Circuit Techniques 195

9.1 Introduction 195
9.2 FFT 195
9.2.1 FFT Algorithms 196
9.2.2 Architecture 201
9.2.3 Comparison 206
9.3 Delay Buffer 208
9.3.1 SRAM/Register File-Based Delay Buffer 209
9.3.2 Pointer-Based Delay Buffer 211
9.3.3 Gated Clock Strategy 212
9.3.4 Comparison 214
9.4 Circuits for Rectangular-to-Polar Conversion 216
9.4.1 Arctangent Function 216
9.4.2 Magnitude Function 221
9.4.3 Comparison 225
9.5 Circuits for Polar-to-Rectangular Conversion 225
9.5.1 Trigonometric Approximation 228
9.5.2 Polynomial Approximation 229
9.5.3 Comparison 230
Bibliography 231

10. System Examples 233

10.1 MC-CDMA Downlink Receiver IC 233
10.1.1 System Description 234
10.1.2 Transmitter and Receiver Design 234
10.1.3 Circuit Design 238
10.1.4 Experimental Results 241
10.2 MIMO–OFDM Cognitive Radio Receiver IC 243
10.2.1 System Overview 243
10.2.2 Architecture and Circuit Design 245
10.2.3 Experimental Results 250
Bibliography 252

Index 255

Petitioner Samsung Ex-1026, 0008

Preface

Orthogonal frequency-division multiplexing (OFDM) has become the favorite modulation technology for wireless communication systems. To address the needs of OFDM receiver design, we have developed this book based on course materials for a class in digital communication IC design. This book is ideal for advanced undergraduate and post-graduate students from either VLSI design or signal processing backgrounds. For engineers working on algorithms or hardware of wireless communications systems, this book provides a comprehensive understanding of the state-of-the-art OFDM design technology and will be a valuable reference.

The topics in this book include theories, algorithms, architectures and circuits of OFDM wireless communication systems. One special feature of this book lies in last three chapters, from which our readers can learn how to develop signal processing algorithms oriented from hardware implementation and how to design ICs for wireless OFDM systems. These techniques are lively illustrated through two design examples dealing with two OFDM systems that currently attract much attention–MC-CDMA for future cellular communications and MIMO–OFDM for next-generation WLAN with cognitive radio capability.

This book is organized into three parts. The first part of the book aims at reviewing background knowledge that includes fundamentals of modulation and communication, signal propagation and channel modeling. In this part, ideas behind formulas, rather than mathematical derivations, are emphasized and several examples are provided to allow easy comprehension of the concepts. In the second part, in-depth treatment of two essential signal processing tasks–synchronization and channel estimation–is offered. Then, MIMO (multiple-input multiple-output) techniques with their application to OFDM systems are also delineated. This part of the book aims to present the readers with modern signal-processing algorithms in OFDM baseband receivers. The third part of the book talks about hardware design, from design methodology to design of essential blocks. Finally, the book ends with a couple of examples that cover the latest OFDM receiver IC developments. The following gives a more detailed description of the content in each chapter.

Petitioner Samsung Ex-1026, 0009

Chapter 1 reviews several important wireless communication standards, including digital broadcasting systems, mobile cellular systems and wireless data network systems. Without exception, OFDM is the modulation scheme of choice for all standards, exemplifying the importance of OFDM technology in wireless communications.

Chapter 2 discusses digital modulation techniques, including both single-carrier modulation and multi-carrier modulation. The introduction to conventional single-carrier modulation techniques serves as the basis for explaining the multi-carrier OFDM modulation. Basic OFDM processing operations, such as discrete Fourier transform (DFT)/inverse discrete Fourier transform (IDFT), guard interval insertion, guard-band reservation and spectrum shaping, are addressed. The phenomenon of high peak-to-average power ratios in OFDM modulation signals is also illustrated. Then, several standards are described to exemplify OFDM system parameters, such as the fast Fourier transform (FFT) size, guard-interval ratio and guard band ratio with regard to Doppler frequency and channel maximum excess delay.

Multiple access schemes, meaning to support a number of users in the same communication link, are discussed in Chapter 3. In addition, spread spectrum techniques, from which CDMA is derived, are illustrated. In that section, several important codes popularly used in CDMA and spread spectrum systems are also presented. Error-correcting codes, indispensable in digital communication systems, are introduced subsequently. Several prevailing error-correcting codes and their decoding strategies are covered. They include convolutional codes, Reed-Solomon codes and low-density parity check (LDPC) codes.

Wireless receiver design is impossible without a thorough understanding of the impairments to signals during propagation. Chapter 4 discusses propagation mechanisms, fading phenomena and other non-ideal effects in the channel and transceiver front-ends. Passing through a wireless channel, communication signals suffer from path loss and a shading effect, which prominently weaken received signal strength. In addition, delay spread, Doppler spread and angle spread in the signal are possible and they produce signal replicas with different arrival times, distorted spectra and incident angles. Front-end electronic non-ideality must also be taken into consideration when designing wireless receivers. Oscillator mismatch as well as relative motion between the transmitter and the receiver causes carrier frequency offset and sampling clock offset. Unmatched branches in the up-/down-conversion path may result in I–Q imbalance and DC offset. A power amplifier with limited linear region is another source of amplitude and phase distortion. In Chapter 4, details about all of the above will be presented.

Synchronization is one of the critical issues in all communication systems, wired and wireless alike. Algorithms for synchronizing the phase and frequency of the carrier signal as well as the sampling clock signal in OFDM receivers are the main topic of Chapter 5. The chapter starts with descriptions of carrier frequency offset, carrier phase error, sampling clock offset and symbol timing offset and their impacts on the received OFDM signals. For each synchronization error, several estimation algorithms with performance comparison are presented. Then time-domain and frequency-domain compensation approaches are introduced. Their pros and cons are also given to help the designers to make knowledgeable and appropriate decisions for their designs.

Chapter 6 concentrates on the channel estimation tasks in OFDM receivers. To perform channel estimation, a receiver often relies on some reference signals, such as the preamble and the pilot signals. As a result, channel estimation algorithms are categorized according to the available reference signal pattern. Channel statistics and characteristics of channel

Petitioner Samsung Ex-1026, 0010

power-delay profiles can also be exploited to obtain better estimation results. Although one prominent advantage of OFDM lies in its simple yet effective one-tap equalization, this chapter will discuss multiple-tap equalizers in OFDM receivers, as they are needed more and more due to inter-carrier interference caused by mobile channels.

As multiple-input multiple-output (MIMO) techniques continue to show promising results in enhancing communication performance in regard to transmission efficiency and quality of service, MIMO has become a very important topic. In Chapter 7, the multiple antenna configurations are first discussed, and then MIMO–OFDM systems are described. New pilot patterns and modified synchronization and channel estimation for MIMO–OFDM systems are presented. Then, the chapter goes on to discuss several MIMO techniques, such as space–time block codes, spatial multiplexing, spatial decorrelation/beam-forming and their detection methods.

Chapter 8 presents the hardware design methodology for communication receiver design. Systematic approaches to map the system-level and the algorithm-level design to the architecture-level and the circuit-level description are first delineated. The effects of finite precision and clipping in analog-to-digital converters (ADC) and limited word-lengths of data path signals are discussed. Specifically, word-length optimization for hardware functional blocks is illustrated. The propagation of quantization errors and the change in signal dynamic range after several common fixed-point arithmetic operations in communication circuits are illustrated. The chapter ends with the introduction of techniques for converting a design in floating-point arithmetic to a corresponding fixed-point design.

Chapter 9 illustrates architectures and circuits that are widely used in OFDM systems, including fast Fourier transform (FFT) processors, delay buffers, circuits for rectangular-to-polar conversion and polar-to-rectangular conversion. A couple of hardware-oriented FFT algorithms are first introduced, followed by several FFT architectures. Pipelined architectures can perform FFT at sample rate, though consuming more hardware resources. On the other hand, memory-based architectures are area-efficient, but may require higher clock rate and complicated control in memory addressing. A delay buffer can be efficiently implemented in shift registers or SRAMs, depending on its length. The chapter also presents several circuits for rectangular-to-polar conversion, which are needed when the phase or magnitude of a complex value is desired. Furthermore, circuits for polar-to-rectangular conversion, needed to generate sinusoidal waveforms, are also introduced at the end of this chapter.

Finally, in the tenth chapter of the book, two OFDM receiver designs are reported. First, a downlink MC-CDMA baseband receiver for future mobile cellular communications is introduced in detail. Then, the design of a cognitive-radio receiver using MIMO–OFDM technology is given. With these two examples, the readers can comprehend firsthand how the algorithms and circuits introduced in the book can be applied in real-life designs.

T. D. Chiueh and P. Y. Tsai
Taipei, Taiwan

Petitioner Samsung Ex-1026, 0011

About the Authors

The authors and their groups at the National Taiwan University, Taipei, Taiwan, and the National Central University, Taoyuan, Taiwan, have been doing research in wireless communication baseband IC design for more than a decade, focusing especially on OFDM (orthogonal frequency-division multiplexing) systems recently. The research results have been published in important international journals and conferences, and are recognized by domestic and international awards.

Tzi-Dar Chiueh received his Ph.D. in electrical engineering from the California Institute of Technology in 1989 and he is now a Professor of Electrical Engineering at the National Taiwan University (NTU). Since August 2004, he has also served as the Director of the Graduate Institute of Electronics Engineering at the same university. He has held visiting positions at ETH Zurich Switzerland and at the State University of New York at Stony Brook. Professor Chiueh has received the Acer Longtern Award 11 times and the MXIC Golden Silicon Award in 2002 and 2005. His teaching efforts were recognized four times by the Teaching Excellence Award from the NTU. Professor Chiueh was the recipient of the Distinguished Research Achievements Award from the National Science Council, Taiwan, in 2004, and was awarded the Himax Chair Professorship at the NTU in 2006. He is the author of more than 140 technical papers, many of which are on algorithm, architecture and integrated circuits for baseband communication systems.

Pei-Yun Tsai received her Ph.D. in electrical engineering from the National Taiwan University in 2005 and she is now an Assistant Professor of Electrical Engineering at the National Central University, Taoyuan, Taiwan. Professor Tsai has received the Acer Longtern Award, MXIC Golden Silicon Award and 1st Asian Solid-State Circuit Conference Student Design Contest Outstanding Award in 2005. Her research interests include signal-processing algorithms and architectures for baseband communication systems.

Acknowledgements

First of all, the authors would like to thank the contribution of Dr Ming-Luen Liu from Mediatek, Inc., Hsinchu, Taiwan, for writing an important part of this book: Chapter 8, and the contribution of Mr. Po-An Chen for the second example in Chapter 10. We would also like to thank the reviewers: Professor Yuan-Hao Huang and Professor Hsin-Pin Ma of the National Tsing Hua University, Hsinchu, Taiwan; Chun-Hao Liao, Yen-Shuo Chang, Jun-Wei Lin, Jin-Hao Yu, Chun-Hao Liu, Yu-Yen Chen, Chien-Yi Wang, Jing-Yeu Yang, all of the Graduate Institute of Electronics Engineering, National Taiwan University. Their valuable comments have greatly improved the content of this book.

Tzi-Dar Chiueh also wishes to thank all former and current students of his MicroSystem Research Laboratory (MSRL) at the National Taiwan University for their tremendous research work. Last, but not least, the constant encouragement from Professor Liang-Gee Chen from the NTU for the development of this book's materials is appreciated.

Petitioner Samsung Ex-1026, 0013

1

Introduction

All wireless communication standards, existing and under development, adopt or consider adopting orthogonal frequency-division multiplexing (OFDM) as the modulation technique. It is clear that OFDM has become the definitive modulation scheme in current and future wireless communication systems.

1.1 Wireless Communication Systems

Pursuance for better ways of living has been instrumental in advancing human civilization. Communication services available at any time and place free people from the limitation of being attached to fixed devices. Nowadays, thanks to the remarkable progress in wireless technology, affordable wireless communication service has become a reality. Mobile phones hook people up whenever and wherever they want. Digital audio and video broadcasting offers consumers high-resolution, better-quality and even interactive programmes. The devices are now thin, light, small and inexpensive. Furthermore, smart mobile phones capable of multimedia and broadband internet access are showing up on the shelves.

Several projects studying wireless networks with different extents of coverage are under way. They will enable wireless access to internet backbone everywhere, either indoors or outdoors and in rural or metropolitan areas. In the following, their evolution and future developments will be introduced. The essential role that the orthogonal frequency-division multiplexing (OFDM) technique plays in wireless communication systems will also become very clear.

1.1.1 Digital Broadcasting Systems

In the modern world, most people fill the need for information and entertainment through audio and video broadcasting. The inauguration of AM radio can be traced back to the early twentieth century, whilst analog TV programmes were first broadcast before the Second World War. Around the middle of twentieth century, FM radio programmes became available. These technologies, based on analog communication, brought news, music, drama, movies and much more into our daily lives. To provide more and better programmes, digital broadcasting techniques, such as digital audio broadcasting (DAB) and digital video broadcasting (DVB), began to replace the analog broadcasting technologies in the past several years.

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0014

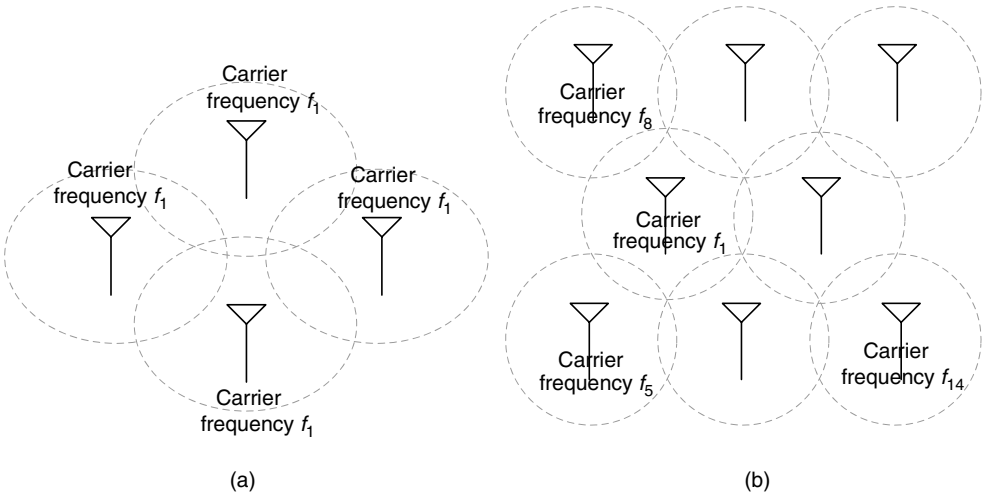


Figure 1.1. (a) Single-frequency network and (b) multi-frequency network

Digital Audio Broadcasting (DAB)

DAB is among the first standards that use the OFDM technique. The DAB project started in mid-1980 [1]. Based on OFDM, DAB has one distinct benefit: a single-frequency network (SFN). In a single frequency broadcasting network, one carrier frequency can be used for all transmitters to broadcast the same radio programme in the entire country without suffering from co-channel interference. On the other hand, in the FM system, only one out of approximately 15 possible frequencies can be used, resulting in a very inefficient frequency re-use factor of 15. A single-frequency network and a multi-frequency network are illustrated in Figure 1.1.

In the DAB system, it is not necessary to search for radio stations as is necessary with AM/FM radios. The programmes of all radio stations are integrated in so-called multiplexes. Multiplexes save on the maintenance cost of individual radio stations. In addition, variable bandwidths can be assigned to each programme, fulfilling their respective demands for sound quality. Music radio multiplexes can transmit at a rate up to the highest-quality 192 Kbps, while mono talk and news programmes may use only 80 Kbps. Futhermore, the DAB system features better mobile reception quality thanks to the OFDM technique.

Digital Video Broadcasting (DVB)

DVB is the European standard for digital television broadcasting [2]. The DVB standards include DVB-S for satellites, DVB-C for cables, DVB-T for terrestrial transmission and DVB-H for low-power handheld terminals. Among them, DVB-T and DVB-H utilize OFDM as the modulation scheme. DVB-T receivers started shipping in late-1990 and now digital DVB-T programmes are available in many countries. As the DAB system, DVB-T/H technology also supports countrywide single-frequency networks. In addition, DVB-T/H standards offer several modes of operation that are tailored for large-scale SFN and high-mobility reception.

Petitioner Samsung Ex-1026, 0015

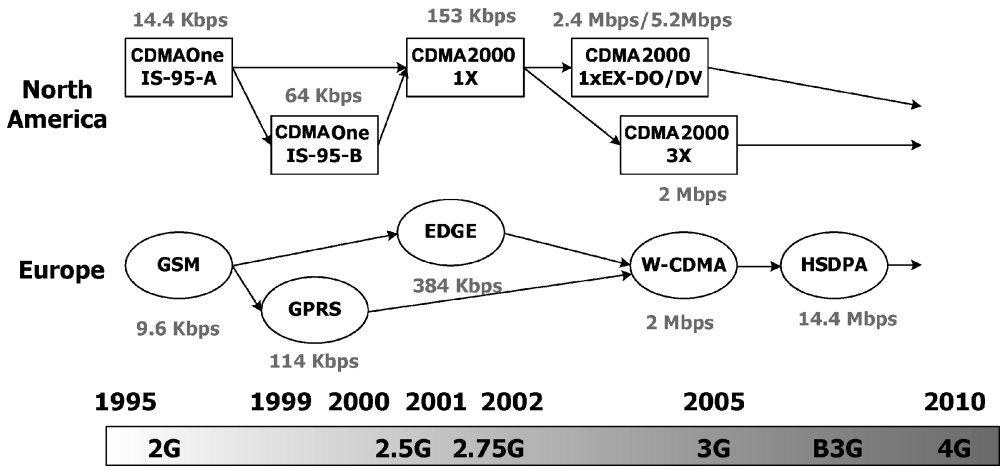


Figure 1.2. Evolution of major mobile cellular communication systems

The basic digital stream in DVB-T is the MPEG-2 transport stream that contains one or more programme streams. Each stream multiplexes compressed video, audio and data signals. The DVB-T standard can support a data rate of MPEG-2 high-definition TV (HDTV), which is up to 31 Mbps. In DVB-H, high-speed IP services as an enhancement of mobile telecommunication networks are offered. Moreover, DVB standard has allowed for integration with bi-direction data connections through other access technology, thus enabling interactive applications between the viewers and the TV stations.

1.1.2 Mobile Cellular Systems

Mobile phones are now a necessity to several billions of people in the world. Their functionalities range from voice service to picture, video and broadband data services. Figure 1.2 shows the migration from the second-generation (2G) to the third-generation (3G), and then onward to the fourth-generation (4G) mobile cellular communication systems. In 2G, the GSM system is used as the European standard and CDMAOne IS-95 is adopted in North America. Both of them offer digital voice services at around 10 Kbps. Afterwards, General Packet Radio Service (GPRS) and Enhanced Data rate for Global Evolution (EDGE) systems provide transmission rates of up to several hundreds of Kbps as an enhancement of the GSM standard. Similarly, CDMA2000 1X upgraded the data transmission to 300 Kbps in North America.

Currently, 3G standards provide data services with a data rate of up to 2 Mbps to accommodate multimedia applications. Two main-stream 3G standards are CDMA2000 3X and wideband-CDMA (W-CDMA). The enhanced version of W-CDMA has been standardized as High Speed Downlink Packet Access (HSDPA), which is regarded as 3.5G and can achieve about a 10-Mbps transmission rate. The third-Generation Partnership Project (3GPP) long-term evolution (LTE) has started to plan possible solutions to future mobile communication technology. The main features include [3]:

Petitioner Samsung Ex-1026, 0016

- spectral efficiency up to 10 b/s/Hz;
- provision of a flexible radio resource management to enlarge cell coverage and improve system efficiency;
- supporting internet protocol version 6 (IPv6) multimedia services with low power consumption and high performance; and
- supporting mobility up to 250 Km/hr.

In order to satisfy high spectral efficiency, low power consumption and excellent performance requirements, advanced techniques are necessary in any future 4G system.

Modulation

In the downlink 3GPP-LTE evolved universal terrestrial radio access (E-UTRA) project, OFDM is considered as the modulation scheme [4]. OFDM has the distinct advantage that it can combat frequency-selective fading channels, which is quite a challenge for receivers of wideband systems. Additionally, OFDM can achieve efficient spectrum utilization, flexible subcarrier allocation and adaptable subcarrier modulation [3].

MIMO

Multiple antennas can be used at the transmitter and at the receiver of a communication system. Such systems are called multiple input and multiple output (MIMO) systems. MIMO systems may be implemented in several different ways and can be categorized into three types. The first type of MIMO system provides spatial diversity and enhances power efficiency. It includes space–time/frequency block code (STBC/SFBC), space–time trellis code (STTC) and delay diversity systems. The second type of MIMO system implements spatial multiplexing to increase its transmission rate. Independent data streams are transmitted over a group of antennas. At the receiver, signals from several antennas are detected and the transmitted information recovered. In the third type of MIMO system, some capacity gain can be achieved over non-MIMO systems by pre-processing the signals to be transmitted according to the channel characteristics and then decoding the received signals accordingly.

Link Adaptation

Link adaptation algorithms, composed of adaptive modulation and coding (AMC), are also regarded as one prominent technique for future communication systems. Its basic concept is to adapt transmission parameters according to channel conditions. Modulation schemes and rates for forward-error-correction codes are the fundamental adaptable parameters. Other parameters, such as power levels, signal bandwidth and spreading factor in spread spectrum and CDMA systems, are also settings that can be adjusted [5].

Radio Resource Management

Flexible radio resource management (RRM) policies become indispensable in future wireless systems as users with various multimedia applications require different quality of service (QoS). Three major topics include scheduling, power control and interference

Petitioner Samsung Ex-1026, 0017

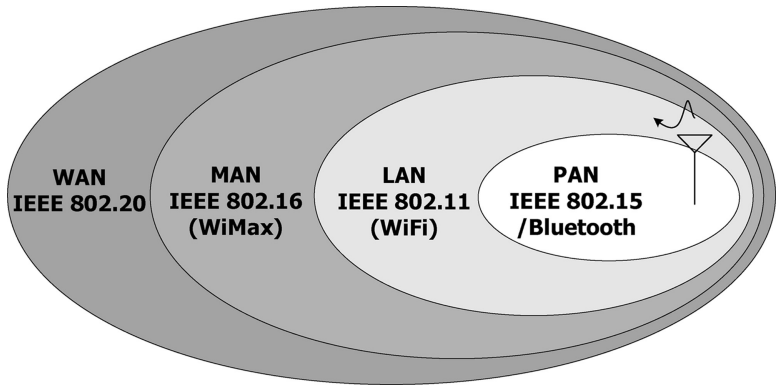


Figure 1.3. Illustration of several IEEE wireless network standards

mitigation. Scheduling is very crucial in that a large number of different applications need to be supported. Priority-based management of different queues must be supported to satisfy all sorts of QoS requirements and cope with a variety of traffic flows [5]. Power-control algorithms, on the other hand, are designed to minimize overall power consumption. The benefits brought about by power control are lower interference level and longer battery life. Finally, the interference mitigation methods include interference randomization, interference cancellation and interference avoidance.

1.1.3 Wireless Network Systems

Bluetooth and IEEE 802.11 wireless local area network are two famous wireless networks. Actually, the Institute of Electrical and Electronics Engineers (IEEE) has already defined several wireless data network standards, from small-area to large-area, as depicted in Figure 1.3. The smallest one is the wireless personal area network (PAN), which covers only several meters around a user. Operating in a bigger environment than wireless PAN, the IEEE 802.11 wireless local area network (LAN) is by far the most successful and prevalent wireless computer network standard. In wireless LAN, short-distance communications within several tens of meters and up to 100 meters are provided. The metropolitan area network (MAN) extends its coverage to several kilometers–the range of typical cells in urban areas. The wide-area network (WAN) is the standard with the largest coverage and it supports communications over up to tens of kilometers, including hilly terrains and rural areas. With all these networks, uninterrupted internet access can be made available whenever and wherever the users desire.

Personal Area Network (PAN)

The IEEE 802.15 working group is responsible for the standardization of wireless PAN [6]. Portable and mobile infotainment products such as cameras, personal digital assistants (PDAs) and handsets can benefit greatly from incorporating the function of wireless PAN connection. Several projects are coordinated by the IEEE 802.15 working group. IEEE 802.15.1 was developed based on the Bluetooth standard. In the enhanced data rate (EDR)

Petitioner Samsung Ex-1026, 0018

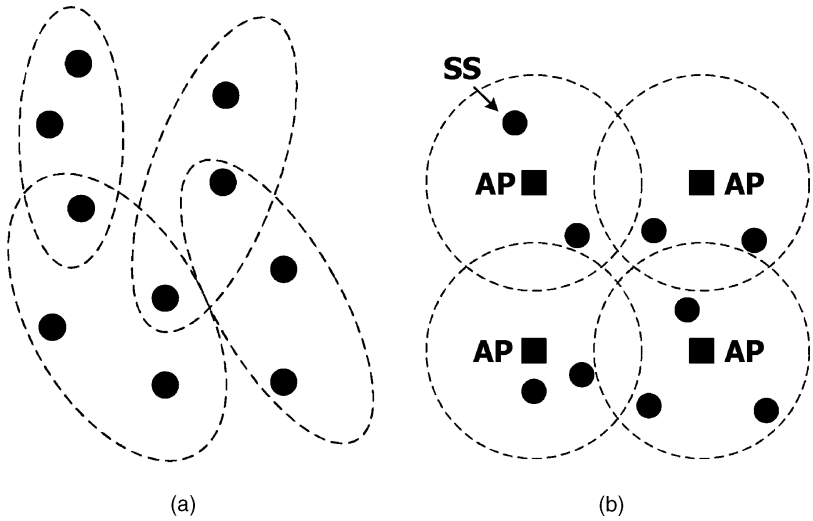


Figure 1.4. (a) Scatter network and (b) cellular network

standard of Bluetooth 2.0, scatter ad hoc connections (shown in Figure 1.4(a)) with a peak data rate of 2.178 Mbps is achieved. The frequency band used is the industrial, scientific and medical (ISM) band at 2.4 GHz.

The IEEE 802.15.3 task group works on high-rate, low-cost and low-power solutions. The standard was released in 2003. It adopts ad hoc peer-to-peer networking and supports data rates from 11 to 55 Mbps. In 2002, a project authorization request (PAR) initiated the development of a high-data-rate ultra wide-band (UWB) standard as the IEEE 802.15.3a standard, which was regarded as an enhanced amendment for high-speed multimedia and imaging applications to the IEEE 802.15.3 standard. UWB communications are defined as systems whose emitted signal bandwidths exceed 500 MHz or 25% of the carrier frequency. Two proposals were presented: multi-band orthogonal frequency-division multiplexing (MB-OFDM) and direct-sequence UWB (DS-UWB). Formed in 2005, the IEEE 802.15.3c group endeavoured to develop alternative physical layer solutions exploiting millimeter waves–the band around 57–64 GHz, to be specific. This standard is geared towards short-range applications that require very high data rates of up to 2 Gbps, such as real-time high-definition video streaming.

On the other hand, the IEEE 802.15.4 standard aims to provide a wireless solution with a low data rate but low power consumption and longer battery life. The target applications include house automation, remote control and toy interaction. This standard operates in the ISM radio bands: 868 MHz in Europe, 915 MHz in the USA and 2.4 GHz in most countries. Data rates of 250, 40 and 20 Kbps are supported with very low-complexity devices to allow years of operation.

Local Area Network (LAN)

The working group of IEEE 802.11, also known as WiFi, defines a series of wireless LAN standards [7]. Unlike the scatter ad hoc network of wireless PAN, the 802.11 wireless LAN

Petitioner Samsung Ex-1026, 0019

adopts cellular radio architecture using base stations, called access points (AP), to control the traffic to/from the subscriber station (SS) within their respective cells, as shown in Figure 1.4(b). The access points are usually connected to a wireline backbone to set up links to the internet.

The first IEEE 802.11 standard was released in 1997 using either frequency hopping spread spectrum (2.4 GHz), direct sequence spread spectrum (2.4 GHz) or infrared (IR) as the transmission technology. The supported data rates were 1 and 2 Mbps. Two years later, IEEE 802.11b, which uses a complementary code keying (CCK) modulation scheme, was ratified as an amendment. It extends the transmission rate to 5.5 and 11 Mbps. With a data rate that was five times higher than the previous generation, IEEE 802.11b products suddenly became very popular in the market. Simultaneously, in 1999, another OFDM wireless LAN standard (IEEE 802.11a) was proposed and it increased the maximum data rate to 54 Mbps. Because the 2.4 GHz ISM band is very crowded, IEEE 802.11a uses another band at around 5 GHz with a low level of interference. Unfortunately, a higher carrier frequency incurs more penetration loss and also increases the cost of radio-frequency components. As a result, IEEE 802.11g was approved in 2003 to transmit at 2.4 GHz using the same OFDM technique as in IEEE 802.11a and yet achieving a data rate of up to 54 Mbps. In addition, IEEE 802.11g is backward compatible to IEEE 802.11b. It has so many conveniences and advantages that IEEE 802.11g or dual-band (2.4/5 GHz), tri-mode (11a/b/g) products are now very well received in the market.

In 2004, a new task group (IEEE 802.11n) was formed to increase the wireless LAN data rate further. A very aggressive spectral efficiency higher than 15 bps/Hz is proposed and it needs to offer interoperability with existing 802.11a/b/g networks. Wireless technologies including OFDM modulation and MIMO techniques with up to four antennas are adopted. Other measures put forward in the 802.11n proposals are higher code rate, low-density parity check code (LDPC), 20/40MHz channelization and reduction in guard interval overhead.

Metropolitan Area Network (MAN)

The 802.16 is the IEEE standard for a wireless MAN, sometimes also dubbed WiMAX [8]. It specifies an air interface for fixed and broadband wireless access systems and aims to provide a solution to the so-called 'last-mile' internet connection problem. In the countryside, deployment of wired digital subscriber loop (DSL), cable or optical fibre can be very expensive. On the contrary, with the wireless IEEE 802.16 networks, residents in rural areas can connect to the internet effortlessly.

Originally, the 802.16 and 16c defined a single-carrier system operating at frequencies ranging from 10 to 66 GHz. Later, the 802.16a defined several modes, such as single-carrier, OFDM and orthogonal frequency-division multiple access (OFDMA), in licensed and unlicensed bands from 2 to 11 GHz. The 802.16-2004, originally known as the 802.16d, includes the standards defined in the 802.16/16c and 16a. One year later, the 802.16e-2005 proposed a revision with more enhanced mobility than the 802.16d and it was thus called mobile WiMAX. The major revision is a scalable OFDM scheme in the OFDMA mode to restrict the Doppler effect regardless of the bandwidth used. In addition, the highest carrier frequency was reduced from 11 to 6 GHz. The new standard also incorporated several MIMO techniques to enhance its performance in terms of coverage, frequency re-use and bandwidth efficiency.

Petitioner Samsung Ex-1026, 0020

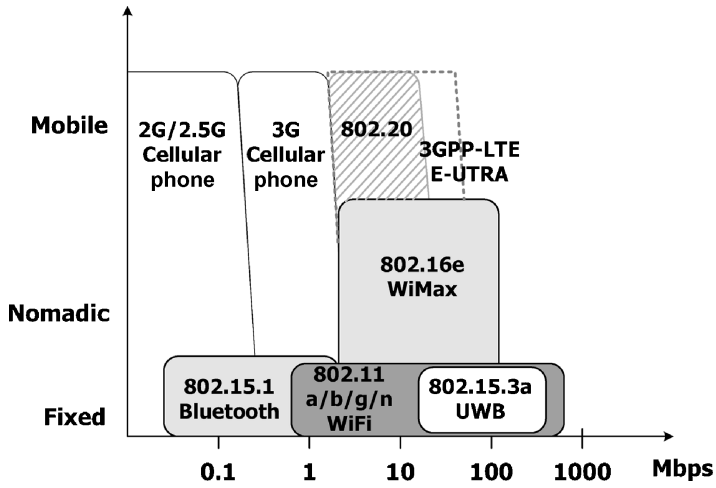


Figure 1.5. Mobility versus transmission rate of several wireless communication standards

Wide Area Network (WAN)

The wireless network with maximum coverage is the WAN, also called mobile broadband wireless access (MBWA) [9]. The working group was established in 2002 with a scope to provide IP services with full mobility of up to 250 Km/hr in cells with a radius of tens of kilometers. 802.22 networks operate with a carrier frequency below 3.5 GHz. In the draft, a combination of the OFDM scheme and MIMO techniques was considered.

In Figure 1.5, the mobility and data rate of several wireless data communication network standards are illustrated. The IEEE 802.11n wireless LAN provides the highest transmission rate but can only be used in fixed reception. On the other hand, IEEE 802.20 and 3GPP-LTE E-UTRA support the highest mobility of 250 Km/hr with a data rate possibly approaching 100 Mbps. Note that all these advanced standards have one feature in common, namely they all use OFDM. This illustrates the fact that OFDM is and will be the modulation technology of choice in wireless communications.

Bibliography

[1] *Digital Broadcast*, available online at www.worlddab.org/.
[2] *Digital Video Broadcasting*, available online at www.dvb.org/.
[3] P. Zhang, X. Tao, J. Zhang, Y. Wang, L. Li and Y. Wang, 'A vision from the future: Beyond 3G TDD,' *IEEE Commun. Mag.*, vol. 43, Jan. 2005, pp. 38–44.
[4] 3GPP, 'Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7),' 3GPP TR 25.814 V7.1.0, Sep. 2006.
[5] R. Fantacci, F. Chiti, D. Marabissi, G. Mennuti, S. Morosi and D. Tarchi, 'Perspectives for present and future CDMA-based communications systems,' *IEEE Commun. Mag.*, vol. 2, Feb. 2005, pp. 95–100.
[6] *IEEE 802.15 Working Group for WPAN*, available online at www.ieee802.org/15/.
[7] *IEEE 802.11 Wireless Local Area Networks*, available online at www.ieee802.org/11/.
[8] *IEEE 802.16 Working Group on Broadband Wireless Access Standards*, available online at www.ieee802.org/16/.
[9] *IEEE 802.20 Mobile Broadband Wireless Access (MBWA)*, available online at www.ieee802.org/20/.
[10] H. Yang, 'A road to future broadband wireless access: MIMO-OFDM-based air interface,' *IEEE Commun. Mag.*, vol. 1, Jan. 2005, pp. 53–60.

Petitioner Samsung Ex-1026, 0021

2

Digital Modulation

Digital modulation is the process of representing binary information using segments of different sinusoidal waveforms. The parameters that can be adjusted in a sinusoidal wave are its amplitude, frequency and phase.

2.1 Single-Carrier Modulation

Single-carrier modulation techniques use only one sinusoidal wave at all times, while in the multi-carrier modulation techniques, several sinusoidal waves are transmitted simultaneously. Basic single-carrier modulation techniques modify only one of the three parameters–amplitude, frequency and phase–of the sinusoidal wave according to the binary information to be transmitted. These techniques are called *amplitude shift keying* (ASK), *frequency shift keying* (FSK) and *phase shift keying* (PSK), respectively, according to the parameter of the sinusoidal wave that is modified. The basic time unit in digital modulation techniques is a *symbol*, which is composed of a segment of the sinusoidal waveform. If there are only two possible different symbols in a digital modulation, then it is called a *binary modulation*. Figure 2.1 depicts sample waveforms of binary ASK, binary FSK and binary PSK.

Note that ASK and PSK are linear modulation, whose symbol waveforms are sinusoidal waveforms of the same frequency. A more compact and clear representation of all possible symbols in such modulation is drawn by the phasor representations of all the possible symbols on a phasor plane. This representation is called the *signal constellation* of a digital modulation. The signal constellations of the binary ASK (BASK) and binary PSK (BPSK) are illustrated in Figure 2.2(a) and (b). FSK, however, does not have a signal constellation, as its symbol waveforms use sinusoidal waves of different frequencies.

With the signal constellation representation, higher-order modulation techniques that have a large number of possible symbol waveforms can be clearly described. For instance, M-ary PSK has M possible symbol waveforms with different phases, and carries $log_2 M$ bits per symbol. Figure 2.2(c) and (d) show the signal constellation of *quaternary PSK* (QPSK) and *8PSK*.

More advanced digital modulation techniques change more than one parameter in a segment of sinusoidal wave. For instance, if both the amplitude and the phase of the

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0022

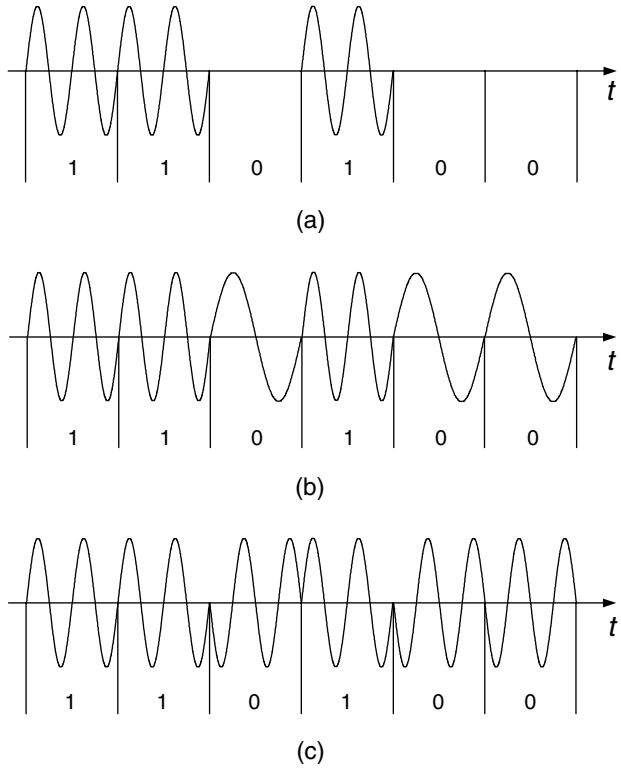


Figure 2.1. Waveforms of three basic digital modulation techniques: (a) ASK, (b) FSK and (c) PSK

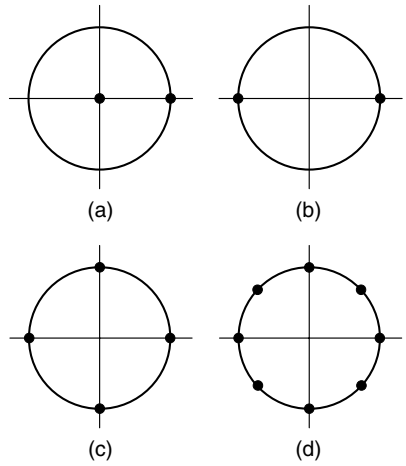


Figure 2.2. Signal constellations of some basic single-carrier digital modulation techniques: (a) binary ASK, (b) binary PSK, (c) QPSK and (d) 8PSK

Petitioner Samsung Ex-1026, 0023

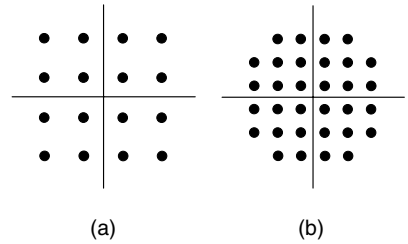


Figure 2.3. Signal constellations of two quadrature amplitude modulation techniques: (a) 16-QAM and (b) 32-QAM

sinusoidal wave are changed, then it is called *quadrature amplitude modulation* (QAM). This is because both levels of the in-phase and quadrature-phase components are modified. There are two types of QAMs: square-type and cross-type. Figure 2.3 shows the signal constellations of 16-QAM (square-type) and 32-QAM (cross-type).

A special kind of FSK maintains the phase continuity of the modulated waveform at symbol boundary and is called *continuous-phase frequency shift keying* (CPFSK). One of such modulation that has the minimum frequency spacing between the frequencies of two possible symbol waveforms is called *minimum shift keying* (MSK). It has the distinct feature that its sidelobe fall-off is sharper than that of other digital modulation techniques, and it has been adopted in several wireless communication systems.

2.1.1 Power Spectral Densities of Modulation Signals

Digital modulation techniques not only have to be robust against all sorts of impairments in the channel and in the transceiver itself, but they also have to be spectrally efficient. A digital modulation system designer must consider the limitation on the signal bandwidth when selecting the digital modulation technique. In the following, the signal power spectral densities of the digital modulation techniques presented in the previous section will be examined.

2.1.2 PSK, QAM and ASK

PSK and QAM are both linear modulation using sinusoidal waves of the same frequency. In general, their baseband signals can all be represented as

$$x(t) = \sum_{k=-\infty}^{\infty} (x_I(k) + jx_Q(k))\Pi(t/T - k), \tag{2.1}$$

where $\Pi(\cdot)$ is the rectangular pulse that is unity when its argument is between 0 and 1 and is zero otherwise; and $(x_I(k), x_Q(k))$ is the k-th baseband symbol. For the PSK signal, $(x_I(k), x_Q(k)) = (\cos((2k+1)\pi/M), \sin((2k+1)\pi/M)), k = 0, 1, \ldots, M-1$; for the square QAM signal, $(x_I(k), x_Q(k))$ are the coordinates in the signal constellation.

Petitioner Samsung Ex-1026, 0024

The auto-correlation function of $x(t)$ is first computed,

$$\begin{aligned}
&\Phi_{xx}(t, t+\tau) \\
&= E[x^*(t)x(t+\tau)] \\
&= E\left[\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}[(x_I(k) - jx_Q(k))(x_I(l) + jx_Q(l))] \cdot \Pi(t/T - k) \cdot \Pi((t+\tau)/T - l)\right].
\end{aligned} \tag{2.2}$$

Without loss of generality, let $\tau = rT + \delta T$, where r is an integer and $0 \le \delta < 1$. Then note that

$$\Pi(t/T - k) \cdot \Pi((t+\tau)/T - l) = 0 \qquad \text{for } l - k > r + 1 \quad \text{or} \quad l - k < r.$$

Therefore, Equation (2.2) becomes

$$\begin{aligned}
&\Phi_{xx}(t, t+\tau) \\
&= \sum_{k=-\infty}^{\infty}\{E[(x_I(k) - jx_Q(k))(x_I(k+r) + jx_Q(k+r))]\Pi(t/T - k) \cdot \Pi((t+\delta T)/T - k) \\
&\quad + E[(x_I(k) - jx_Q(k))(x_I(k+r+1) + jx_Q(k+r+1))]\Pi(t/T - k) \cdot \Pi((t+\delta T)/T - k - 1)\}.
\end{aligned}$$

Since $x(t)$ is cyclo-stationary, one can get its auto-correlation function by averaging over a period T and

$$\begin{aligned}
&\overline{\Phi_{xx}}(\tau) \\
&= \frac{1}{T}\int_0^T \Phi_{xx}(t, t+\tau)dt \\
&= \frac{1}{T}\left\{E[(x_I(0) - jx_Q(0))(x_I(r) + jx_Q(r))]\int_0^T \Pi(t/T) \cdot \Pi((t+\delta T)/T)dt\right. \\
&\quad \left. + E[(x_I(0) - jx_Q(0))(x_I(r+1) + jx_Q(r+1))]\int_0^T \Pi(t/T) \cdot \Pi((t+\delta T)/T - 1)dt\right\} \\
&= E[(x_I(0) - jx_Q(0))(x_I(r) + jx_Q(r))] \cdot (1 - \delta) \\
&\quad + E[(x_I(0) - jx_Q(0))(x_I(r+1) + jx_Q(r+1))] \cdot \delta.
\end{aligned} \tag{2.3}$$

The baseband symbols are assumed independent and they are also all zero-mean random variables (see Figure 2.2(b), (c) and (d) and Figure 2.3), so

$$\overline{\Phi_{xx}}(\tau) = \begin{cases} E[x_I(0)^2 + x_Q(0)^2](1 - \delta) & r = 0 \\ E[x_I(0)^2 + x_Q(0)^2]\delta & r = -1 \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

Petitioner Samsung Ex-1026, 0025

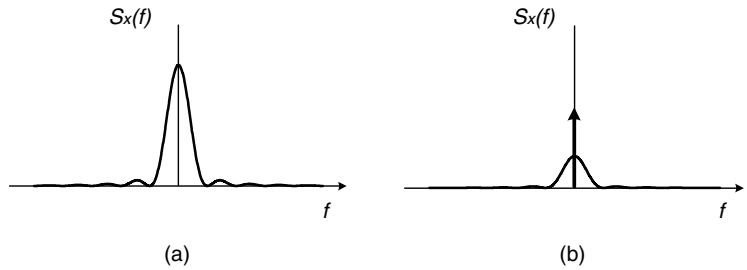


Figure 2.4. Power spectral density functions of (a) PSK and QAM signals and (b) ASK signal

Let the variance of the baseband symbols be σ^2, then, without loss of generality, one has

$$\sigma^2 = E[x_I(0)^2 + x_Q(0)^2],$$

and

$$\overline{\Phi_{xx}}(\tau) = \begin{cases} \sigma^2(1-\tau/T) & 0 < \tau < T \\ \sigma^2(1+\tau/T) & -T < \tau < 0. \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

Taking the Fourier transform of the above auto-correlation function, it is clear that the power spectral density functions of the PSK and QAM signals are identical and are given by [1]

$$S_x(f) = \sigma^2 T \left(\frac{\sin \pi f T}{\pi f T} \right)^2. \tag{2.6}$$

Figure 2.4(a) depicts such a power spectral density function. The power spectral density of the binary ASK modulation signal can then be computed similarly. The baseband waveform of binary ASK modulation can be regarded as a superposition of a constant 1/2 plus a bipolar waveform with two levels: 1/2 and $-1/2$. Note that the second waveform is exactly that of a BPSK baseband signal except for a proportional constant. The auto-correlation function of the binary ASK signal turns out to be the sum of a constant plus one quarter of the auto-correlation function of the BPSK signal. Consequently, the power spectral density of the ASK modulation signal is an impulse function centerd at DC, plus a scaled power spectral density function of the PSK signal, as shown in Figure 2.4(b). Moreover, the power is divided equally between the discrete impulse function and the continuous PSK-like power spectral density.

The passband modulation signal is generated by a quadrature modulator using the following formula:

$$\begin{aligned} y(t) &= \mathrm{Re}\{x(t)e^{j2\pi f_c t}\}. \\ &= \frac{1}{2}\{x(t)e^{j2\pi f_c t} + x^*(t)e^{-j2\pi f_c t}\}. \end{aligned} \tag{2.7}$$

Petitioner Samsung Ex-1026, 0026

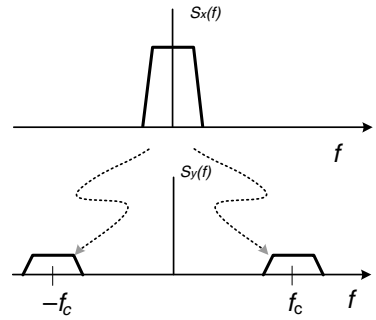


Figure 2.5. Relation between passband and baseband power spectral density functions

Hence, the passband power spectral density function of the ASK, PSK and QAM signals is given by

$$S_y(f) = \frac{1}{4}[S_x(f - f_c) + S_x(-f - f_c)], \tag{2.8}$$

namely making two copies of the baseband power spectral density function; moving one copy to frequency f_c and the other copy to frequency $-f_c$; then multiplying both copies by one quarter, as shown in Figure 2.5

2.1.3 CPFSK and MSK

The derivation of the power spectral density function of CPFSK is quite complicated. A complete treatment of this subject can be found in Chapter 4 of [2] and only binary CPFSK will be discussed here. The two possible symbol waveforms of binary CPFSK are given by

$$y(t) = \cos(2\pi f_c t + 2\pi i(h/2T)t + \phi(kT)) \qquad t \in (kT, kT + T), \tag{2.9}$$

where i is either $+1$ or -1; f_c is the center frequency; T is the symbol period; h is the modulation index that specifies the distance between the two frequencies; and $\phi(\cdot)$ is the phase that makes the phase of the waveform continuous at the symbol boundary.

As the modulation index increases, the two frequencies grow farther apart and the power spectral density becomes less smooth and more bi-modal. The minimum modulation index without losing the orthogonality between the two sinusoidal waves is 0.5. MSK is the CPFSK that has a modulation index of 0.5. The power spectral density of the MSK signal is given by [2]

$$S_y(f) = \frac{16T}{\pi^2}\left(\frac{\cos 2\pi fT}{1 - 16f^2T^2}\right)^2. \tag{2.10}$$

Petitioner Samsung Ex-1026, 0027

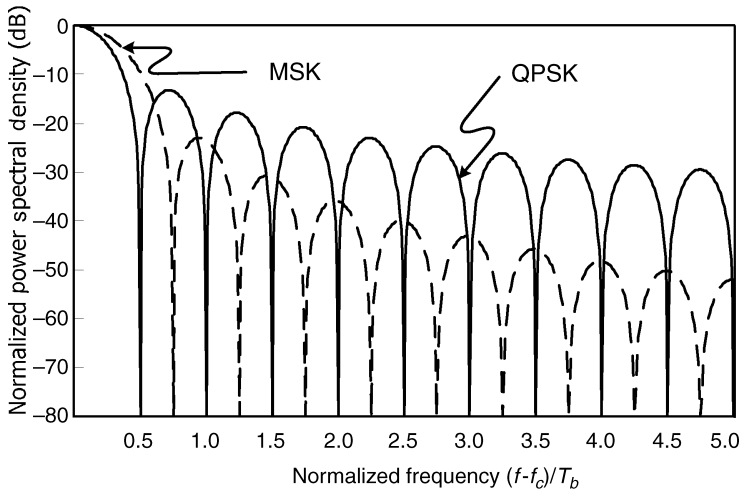


Figure 2.6. Power spectral density functions of QPSK and MSK signals

Figure 2.6 illustrates the comparison between the power spectral density functions of the QPSK signal and the MSK signal. The frequency is normalized with respect to the bit duration (T_b) and $T = 2T_b$ for QPSK as it transmits two bits per symbol. Note that the main lobe of the MSK signal power spectral density function is 50% wider than that of the QPSK signal, but its sidelobe falls off more sharply than that of the QPSK signal.

2.1.4 Pulse Shaping and Windowing

Symbol waveforms are concatenated to form the signal to be up-converted in frequency and transmitted by the transmitter. As shown in the previous section, such a signal can have significant sidelobe so that the signals in the adjacent frequency band can suffer severe interference from such 'spill-over.' To prevent such *adjacent-channel interference* (ACI), most wireless communication standards set a mask on the spectrum of the transmitted signal. In order to meet that spectrum mask, the transmitter usually applies some form of pulse-shaping filtering or time-domain windowing.

Two famous pulse-shaping filters are the *raised-cosine filter* and the *Gaussian filter*. Both of them are low-pass filters, and their impulse responses are given by

$$h_{RCF}(t) = \frac{\sin(\pi t/T)}{\pi t/T} \cdot \frac{\cos(\pi \beta t/T)}{1 - 4\beta^2 t^2/T^2}, \tag{2.11}$$

and

$$h_{Gaussian}(t) = \sqrt{\frac{2\pi}{\ln 2}}(BT)e^{-\frac{2\pi^2}{\ln 2}(BT)^2 t^2}, \tag{2.12}$$

Petitioner Samsung Ex-1026, 0028

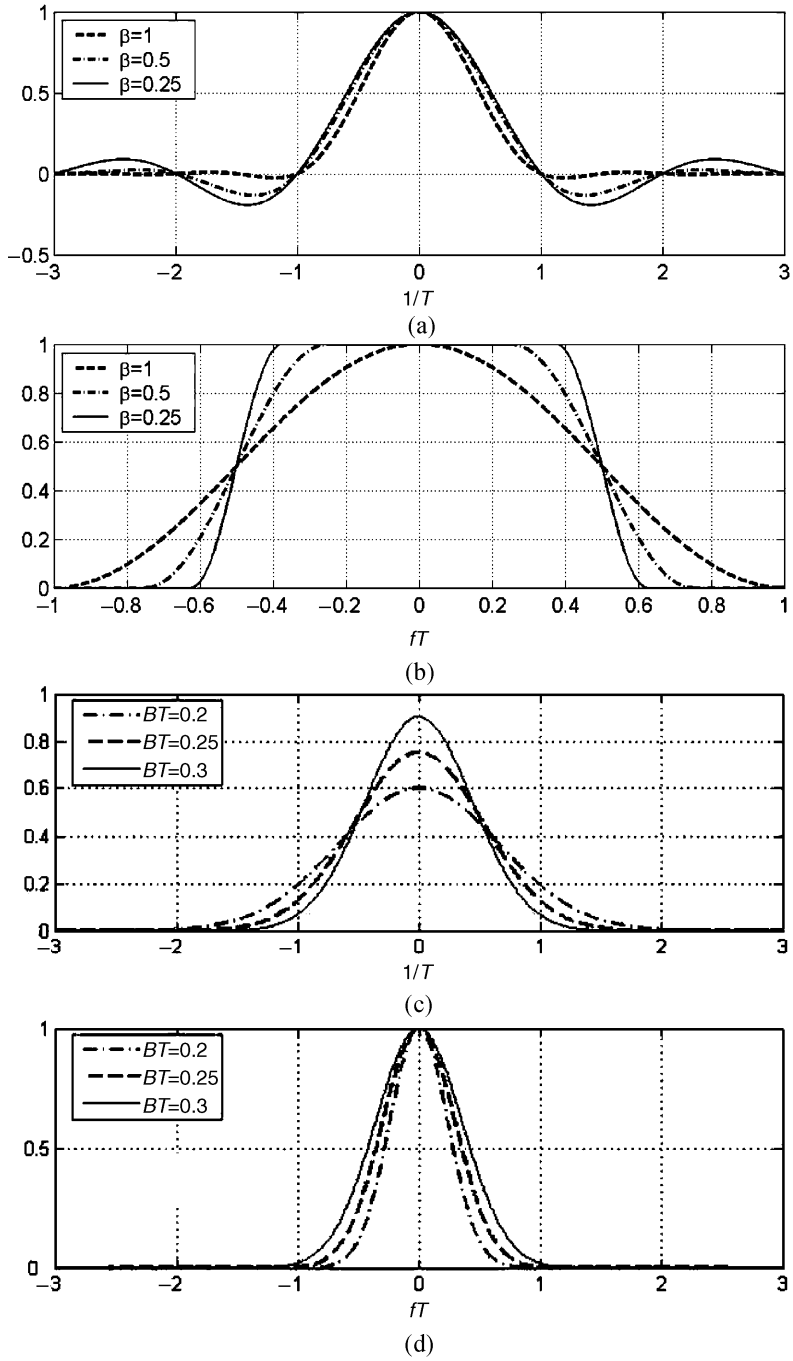


Figure 2.7. Pulse-shaping filters: (a) impulse response of the raised-cosine filter, (b) frequency response of the raised-cosine filter, (c) impulse response of the Gaussian filter and (d) frequency response of the Gaussian filter

Petitioner Samsung Ex-1026, 0029

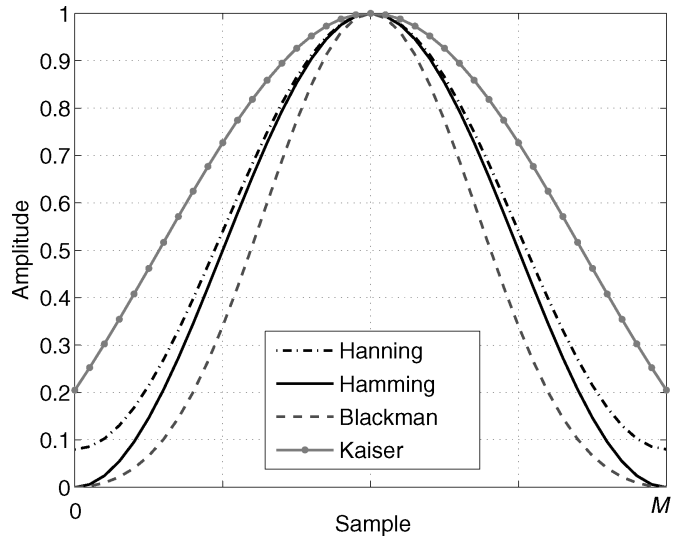


Figure 2.8. Windowing functions

where β is the roll-off factor, taking values in the range $0 \leq \beta \leq 1$; and B is the filter's 3-dB bandwidth.

Their frequency responses are

$$H_{RCF}(f) = \begin{cases} T & 0 \leq |f| \leq \dfrac{1-\beta}{2T} \\ \dfrac{T}{2}\left\{1 + \cos\left[\dfrac{\pi T}{\beta}\left(|f| - \dfrac{1-\beta}{2T}\right)\right]\right\} & \dfrac{1-\beta}{2T} \leq |f| \leq \dfrac{1+\beta}{2T} \\ 0 & |f| > \dfrac{1+\beta}{2T} \end{cases} \tag{2.13}$$

$$H_{Gaussian}(f) = e^{-\frac{\ln 2}{2}\left(\frac{f}{BT}\right)^2}. \tag{2.14}$$

Figure 2.7 depicts the impulse responses and frequency responses of the raised-cosine filter and the Guassian filter.

Windowing can also limit out-of-band signal energy by smoothing the time-domain signal waveform at the symbol boundary where segments of very different waveforms adjoin. This can prevent abrupt changes in the time-domain waveform so that the spectrum will have weaker high-frequency components and thus can meet the mask set by the standard. There exist many famous windowing functions in the signal processing literature. They include Hamming, Hanning, Blackman and Kaiser [3]. Figure 2.8 illustrates these windows with $(M + 1)$ samples. The shape parameter of the Kaiser window is set to 3.

2.2 Multi-Carrier Modulation

Wireless signals to be transmitted over the air usually suffer frequency-selective fading, namely different frequency components are faded quite differently by the channel.

Petitioner Samsung Ex-1026, 0030

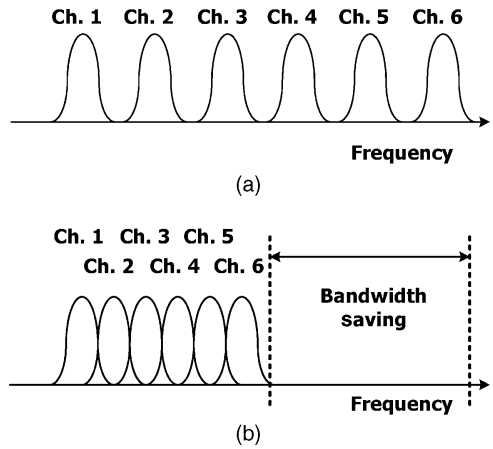


Figure 2.9. (a) Conventional non-overlapping multi-carrier modulation; (b) overlapping multi-carrier modulation

Researchers have studied this problem extensively and proposed many solutions. In conventional single-carrier systems, complex equalization schemes are adopted to combat frequency-selective fading. The ideal equalizer has a frequency response that is the exact inverse of that of the channel. This usually entails an infinite number of equalizer taps. What is worse, noises inflicted on the signal can be enhanced through the equalizer when a deep fade occurs. As a result, even with the best equalizer, a deep fade can still result in communication link failure in single-carrier systems. The first proposal to use parallel data transmission to combat frequency-selective fading channels was published around 1967 [4]. In that system, only a small number of sub-channels use carriers that fall within each deep-faded frequency band. With the help of error-correcting codes, data along those corrupted sub-channels can be recovered. Thus, error-correcting codes are indispensable in all multi-carrier systems.

In early parallel transmission systems, a few non-overlapping sub-channels share the whole frequency band, as shown in Figure 2.9(a). Independent data are modulated on different sub-channels and then these sub-channels are frequency-multiplexed. The purpose of the non-overlap is to eliminate the possible interference among adjacent sub-channels, also known as *inter-carrier interference* (ICI). Note that the guard band between two adjacent sub-channels constitutes a waste of spectrum. In the mid-1960s, spectral efficiency was improved by overlapping the sub-channels, as shown in Figure 2.9(b), which saves up to 50% of the spectrum used. Toward this end, *Orthogonal Frequency-Division Multiplexing* (OFDM) was developed. OFDM is not only a frequency multiplexing technique that mandates orthogonality among sub-channel signals, but also a special case of multi-carrier modulation. Consequently, OFDM can be regarded as either a multiplexing technique or as a modulation scheme.

In the following, the OFDM principle and its mathematical expression are first introduced. Then, the design of OFDM system parameters, such as guard interval ratio and FFT size, will be discussed. Finally, several recent communication standards adopting the OFDM modulation scheme will be addressed.

Petitioner Samsung Ex-1026, 0031

2.2.1 Orthogonal Frequency-Division Multiplexing

Conventionally, the multi-carrier transmitter consists of a set of modulators, each with different carrier frequencies. The transmitter then combines the modulator outputs and generates the transmitted signal. Suppose that the N data to be transmitted are X_k, $k = 0, 1, \ldots, N-1$, where X_k is a complex number in a given constellation, such as QPSK or QAM. Also suppose that the kth carrier frequency for X_k is f_k. Then, the complex-valued multi-carrier transmitter output is given by

$$x(t) = \sum_{k=0}^{N-1} X_k e^{j2\pi f_k t}.$$

Modern communication systems often implement their transmitters and receivers digitally whenever they can. A digital transmitter will generate its output in a sampled-data fashion. By letting $t = nT_s$, where T_s is the sample interval, the digital multi-carrier transmitter output is now

$$x(nT_s) = \sum_{k=0}^{N-1} X_k e^{j2\pi f_k nT_s}.$$

Furthermore, if the carrier frequencies are uniformly spaced in the frequency domain by a frequency spacing of f_S, i.e. $f_k = kf_S, k = 0, 1, \ldots, N-1$, then

$$x(nT_s) = \sum_{k=0}^{N-1} X_k e^{j2\pi k f_S nT_s}.$$

Let $f_S = 1/(NT_s)$–the minimum separation to keep orthogonality among signals on different modulators–then the OFDM signal is given by

$$x_n = x(nT_s) = \sum_{k=0}^{N-1} X_k e^{j2\pi nk/N}.$$

These carriers are called subcarriers and usually there is one more modulation to translate all these subcarriers to a higher frequency band. Except for a multiplying constant $(1/N)$, the above formula is the equation of an N-point *inverse discrete Fourier transform* (IDFT). If N is a power of two, then there exist many fast and efficient algorithms and architectures for implementing such an IDFT operation. It is such efficient digital realization of the OFDM transmitter that makes the OFDM technology a feasible solution to advanced communication systems. Without specific time-domain windowing on the OFDM symbol (x_n) to shape its waveform, the OFDM subcarriers have *sinc*-shaped spectra, as shown in Figure 2.10.

2.2.2 OFDM-Related Issues

Guard Interval

In wireless channels, a receiver may receive several delayed replicas of the transmitted signal, which is known as the multipath effect (for details, see Chapter 4). Figure 2.11(a)

Petitioner Samsung Ex-1026, 0032

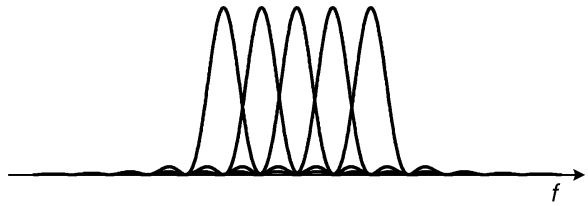


Figure 2.10. OFDM signal power spectral density

shows a scenario in which there are two copies of the received waveform–one on time and the other delayed by some time. Inter-symbol interference (ISI) is induced because the tail part of symbol 1 will interfere with the processing of symbol 2. To eliminate ISI, a guard interval of N_g samples is usually inserted at the beginning of each OFDM symbol, as depicted in Figure 2.11(b). The length of the guard interval is made longer than the delay spread of the wireless channel. As a result, the degree of delay spread in the operating environments must be obtained beforehand. Note that the guard interval actually wastes transmission resources, hence the ratio of the guard interval length to the effective OFDM symbol duration is usually kept below 1/4.

During the guard interval, the transmitter can send null waveform. This scheme is called *zero padding* (ZP) transmission and is illustrated in Figure 2.11(b). A ZP–OFDM system has lower transmission power and a simpler transmitter structure. Unfortunately, the ZP–OFDM scheme introduces ICI, as the orthogonality among subcarriers is destroyed when multiple copies of the time-shifted ZP–OFDM waveform are received. To remove ICI, *cyclic prefixing* (CP) transmission is preferred. The cyclic prefix is an exact copy of a segment of the OFDM symbol located toward the symbol end. Figure 2.11(c) illustrates a typical CP–OFDM signal. To generate the CP signal, an additional buffer is required in an OFDM transmitter. In the following treatments, CP–OFDM is assumed unless otherwise specified.

Null Subcarriers

To prevent significant leakage to adjacent bands, OFDM systems usually do not transmit any data on the subcarriers near the two edges of the assigned band. These unused subcarriers are known as guard subcarriers or virtual subcarriers. The collection of all the unused subcarriers is called the *guard band*. As the OFDM signal power spectrum has quite high sidelobes, reservation of the guard band helps to reduce the out-of-band emission and thus eases the requirements on transmitter front-end filters. Nevertheless, adoption of guard band wastes some assigned bandwidth and decreases spectral efficiency of the OFDM system. In addition to guard bands, some subcarriers around DC frequency (subcarrier index 0) may also be made null in order to evade the large yet unwanted DC and low-frequency components generated by the receiver front-end.

Spectrum Shaping

Wireless communication signals are regulated by spectrum masks which define the allowable maximum in-band and out-of-band signal power spectrum. Figure 2.12 shows

Petitioner Samsung Ex-1026, 0033

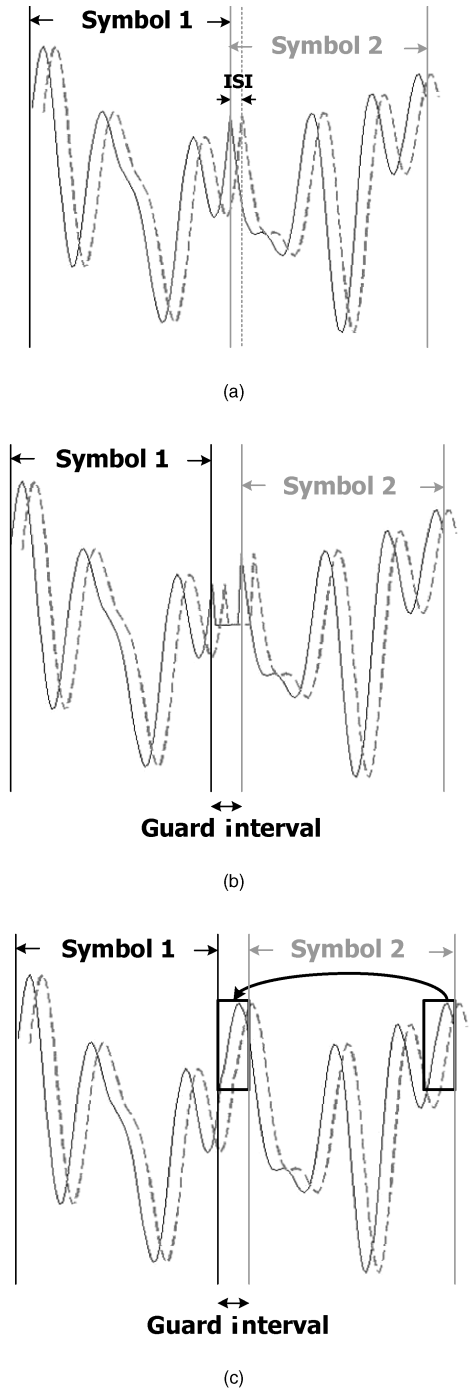


Figure 2.11. (a) Illustration of ISI due to mutlipath delay; (b) zero-padding guard interval to avoid ISI; (c) guard interval with cyclic prefix to eliminate ISI and ICI

Petitioner Samsung Ex-1026, 0034

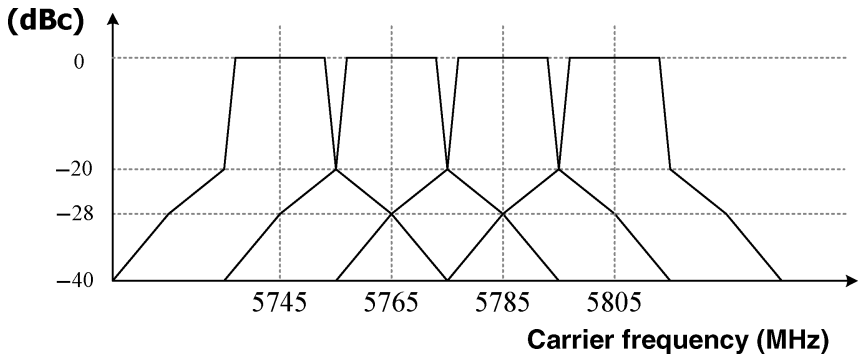


Figure 2.12. Spectrum mask for wireless LAN signal in the U-NII band

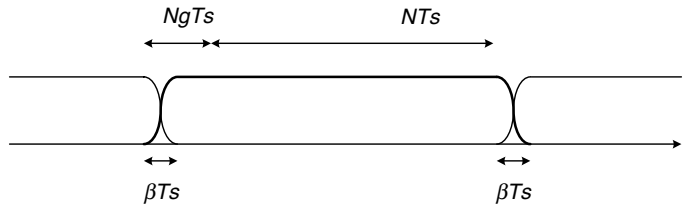


Figure 2.13. Time-domain windowing of the OFDM signals

the signal spectrum mask for the IEEE 802.11a wireless LAN system. For each channel, a 20-MHz bandwidth is allocated. The rectangular time window of the DFT operation and waveform discontinuity at the boundary of adjacent OFDM symbols make the signal spectrum sidelobe fall off more slowly than that specified by the spectrum mask. As such, windowing the time-domain OFDM waveform is often applied to achieve a signal spectrum that complies with the mask.

Of all the popular windows, the raised-cosine window is commonly adopted, and it takes the form of

$$w(t) = \begin{cases} 0.5 + 0.5\cos(\pi + \pi(t + N_g T_s)/(\beta T_s)) & -N_g T_s \le t \le (-N_g + \beta)T_s \\ 1 & (-N_g + \beta)T_s \le t \le NT_s, \\ 0.5 + 0.5\cos((t - NT_s)\pi/(\beta T_s)) & NT_s \le t \le (N + \beta)T_s \end{cases} \tag{2.15}$$

where βT_s is the length of the roll-off region and N and N_g are the number of OFDM symbol samples and the number of guard interval samples, respectively. The OFDM symbol is first cyclically extended at the end by β samples, and then the extended $(N + N_g + \beta)$-sample symbol waveform is multiplied by the window. The next symbol overlaps with the current one by β samples, as indicated in Figure 2.13. Note that the effective guard interval is shortened somewhat by β samples, yet the waveform within $[(-N_g + \beta)T_s, NT_s]$ remains unchanged through the windowing operation.

Petitioner Samsung Ex-1026, 0035

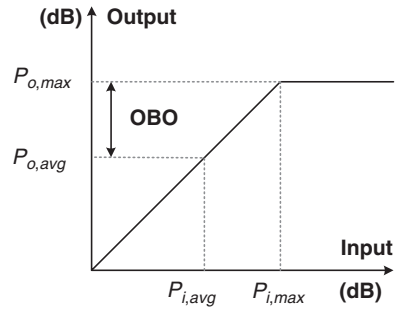


Figure 2.14. Output back-off point of power amplifier

In addition to time-domain windowing, low-pass filtering of the baseband OFDM signal helps to suppress transition-band and out-of-band emission. As the impulse response of this filter is convolved with the OFDM signal, the equivalent channel impulse response seen by the receiver is broadened.

Peak-to-Average Power Ratio

The *peak-to-average power ratio* (PAPR), defined as the ratio of the peak power to the average power, has been one weakness for OFDM communication systems. The PAPR formula is given by

$$\text{PAPR} = \frac{\max|x(t)|^2}{E\{|x(t)|^2\}}. \tag{2.16}$$

In the extreme case, in which all the subcarriers are coherently and equally summed up, the time-domain OFDM signal can have a PAPR of about N. For example, the PAPR of a 256-subcarrier OFDM system can be as high as 256, or equivalently 24 dB.

Such a high PAPR demands high dynamic range in the ensuing amplifier, especially the power amplifier (PA) in the transmitter. If not biased properly, the PA easily enters into saturation, causing nonlinear amplification of large-magnitude signals. To accommodate such large-dynamic-range signals linearly, the PA must work at an operating point, $P_{o,avg}$, that is quite inefficient in terms of power consumption. Namely, a large output back-off (OBO), shown in Figure 2.14, must be implemented. The output back-off is defined as the output saturation power to the average output power of a PA,

$$\text{OBO} = 10\ \log_{10}\frac{P_{o,max}}{P_{o,avg}} \quad (\text{dB}). \tag{2.17}$$

Petitioner Samsung Ex-1026, 0036

To reduce the PAPR, many approaches have been proposed. Clipping and windowing the peak signals exceeding some threshold are one of the possible solutions. However, they may introduce in-band distortion and out-of-band radiation [5]. Some suggested using coding techniques [6,7]. Depending on the input data, the signals to be modulated on all the subcarriers are chosen from a set of codewords that corresponds to waveforms with a lower PAPR. The drawback of the coding techniques is the overhead in transmission efficiency. Scrambling codes can also be adopted to destroy signal regularity, which can incur a high PAPR. In yet another method, the transmitted signals are carefully adjusted by convex optimization to minimize PAPR under the constraints of allowable constellation error and out-of-band energy [8].

2.2.3 OFDM Transceiver Architecture

A general OFDM transmitter integrates several functions, including inverse DFT processing, guard interval insertion and spectrum shaping. In a receiver, besides DFT processing and guard interval removal, additional efforts are required to handle the channel-fading effect and synchronization issues between the transmitter and the receiver.

Without loss of generality, consider only one transmitted symbol. The $N + N_g$ transmitted samples are denoted as

$$\mathbf{x} = [\overbrace{x_{N-N_g} \ \cdots \ x_{N-1}}^{\text{guard interval samples}} \ x_0 \ x_1 \ \cdots \ x_{N-1}]^T, \tag{2.18}$$

where N_g is the number of guard interval samples. Also assume that the time-invariant channel impulse response with R taps is given by

$$\mathbf{h}^T = [h_0 \ h_1 \ \cdots \ h_{R-1}] \tag{2.19}$$

and that $R \leq N_g$. Then, the received signals $\mathbf{z} = [z_0 \ z_1 \ \cdots \ z_{N-1}]^T$ after removal of the cyclic prefix can be expressed as

$$\mathbf{z} = \begin{bmatrix} \overbrace{0 \ \cdots \ 0}^{N_g-R+1} & h_{R-1} & h_{R-2} & \cdots & h_0 & \overbrace{0 \qquad\qquad\qquad\qquad \cdots \ 0}^{N-1} \\ 0 \ \cdots & 0 & h_{R-1} & h_{R-2} & \cdots \ h_0 & \qquad\qquad\qquad\qquad \cdots \ 0 \\ & & & & \cdots & \\ 0 & & & & \cdots & 0 \ h_{R-1} \ h_{R-2} \ \cdots \ h_0 \end{bmatrix} \mathbf{x} + \mathbf{v}, \tag{2.20}$$

where $\mathbf{v}$ is the Gaussian noise and the channel matrix is of the dimension $N \times (N + N_g)$. As opposed to linear convolution of the transmitted signals with the channel impulse response in the single carrier systems, in OFDM systems, the channel impulse response is circularly

Petitioner Samsung Ex-1026, 0037

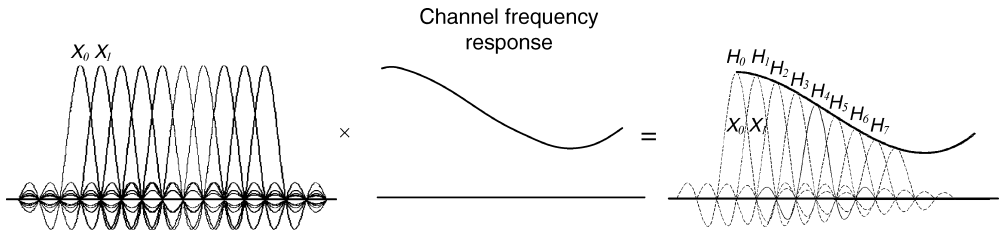


Figure 2.15. Channel frequency response and channel fading effect on the OFDM subcarriers

convolved with the transmitted signals owing to the cyclic prefix in OFDM signals. The received signals can be expressed as

$$\mathbf{z} = \begin{bmatrix} h_0 & 0 & \cdots & & & 0 & h_{R-1} & h_{R-2} & \cdots & h_1 \\ h_1 & h_0 & 0 & & & \cdots & 0 & h_{R-1} & \cdots & h_2 \\ & & & & & \cdots & & & & \\ h_{R-1} & h_{R-2} & \cdots & h_0 & 0 & & & & \cdots & 0 \\ 0 & h_{R-1} & \cdots & h_1 & h_0 & & & & \cdots & 0 \\ & & & & & \cdots & & & & \\ 0 & \cdots & & & 0 & h_{R-1} & & & \cdots & h_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_{N-1} \end{bmatrix} + \mathbf{v},$$
$$= \tilde{\mathbf{Q}}\tilde{\mathbf{x}} + \mathbf{v}, \tag{2.21}$$

where $\tilde{\mathbf{x}}$ consists of the last N elements in $\mathbf{x}$. Note that the circulant matrix $\tilde{\mathbf{Q}}$ can be diagonalized by the DFT and IDFT matrices, yielding

$$\tilde{\mathbf{Q}} = \mathbf{F}^{-1}\mathbf{H}\mathbf{F}, \tag{2.22}$$

where $\mathbf{F}$ and $\mathbf{F}^{-1}$ are the DFT and IDFT matrices, respectively. The matrix $\mathbf{H}$ is a diagonal matrix,

$$\mathbf{H} = \begin{bmatrix} H_0 & 0 & \cdots & 0 \\ 0 & H_1 & \cdots & 0 \\ & \vdots & & \\ 0 & \cdots & 0 & H_{N-1} \end{bmatrix}, \tag{2.23}$$

where each diagonal element corresponds to the frequency-domain channel response at the corresponding subcarrier, as shown in Figure 2.15. Consequently, a low-complexity one-tap equalizer with a coefficient equal to the inverse of the frequency-domain channel response at that subcarrier can effectively mitigate the channel fading effect. Figure 2.16 depicts the architecture of the transmitter and receiver of a typical OFDM communication system.

Petitioner Samsung Ex-1026, 0038

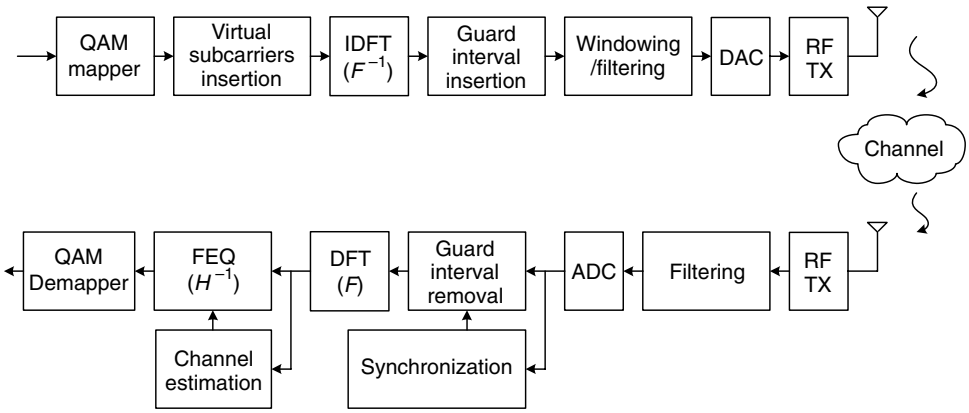


Figure 2.16. Typical OFDM transceiver architecture

2.2.4 OFDM System Examples

Subcarrier spacing in OFDM systems must be carefully designed so that each subcarrier suffers only flat fading. Usually, the extent of channel coherence determines the maximum subcarrier spacing. On the other hand, due to parallel transmission on numerous subcarriers, the OFDM symbol period is made longer than that of the single-carrier symbols. Unfortunately, time-selective fading compromises the long OFDM symbols. In the case in which the channel response varies significantly within an OFDM symbol, ICI will arise, degrading the system performance. The OFDM symbol period is thus upper bounded by the length of the interval that the channel stays stationary. As the symbol period is the inverse of the subcarrier spacing, the minimum subcarrier spacing is thus determined by how fast the channel is changing. Chapter 4 will discuss in more detail the characterization of a channel in terms of its coherence in time domain and in frequency domain.

Given the signal bandwidth and the subcarrier spacing, the FFT size can then be determined. The guard interval depends on the channel maximum excess delay, and can be adjusted according to the operating environment. The guard band ratio considers the sidelobe attenuation to fit in the spectrum mask. The data transmission rate can be calculated from the symbol period, the number of subcarriers and the adopted constellation size, taking into account the overhead of guard interval and guard band. Table 2.1 lists some essential parameters of several mature and future OFDM systems.

The digital audio broadcasting (DAB) system provides four operation modes for different transmission bands using different FFT sizes, yet the same 1.5-MHz bandwidth [9]. The higher the carrier frequency, the wider the subcarrier spacing and the shorter the symbol period so that it can resist time-selective fading. The same guard interval ratio is used for all four modes. Hence, the mode with the longest period is suitable for operating in the long-delay-spread channels, namely large-scale cells. A quarter of subcarriers are reserved as guard band. Considering the loss in the guard interval and the guard band, the maximum data rate that can be achieved is 1.8 Mbps, with $\pi/4$-DQPSK modulation and an error-correcting code with 2/3 code rate.

Petitioner Samsung Ex-1026, 0039

Table 2.1. System parameters of several wireless communication standards using the OFDM technology

	DAB				DVB-T		802.11a/g	802.16e-OFDM	3GPP-LTE EUTRA					
Carrier Frequency (GHz)	≤0.375	≤1.5	≤1.5	≤3	0.4 ~ 0.8		2.5, 5.8	2-11	2					
Sample Frequency (MHz)	2				6.8, 8, 9.14		20	8/7*B, 7/6*B	1.92	3.84	7.68	15.36	23.04	30.72
Bandwidth (MHz)	1.5				6, 7, 8		20	1.5~28	1.25	2.5	5	10	15	20
FFT Size	2048	1024	512	256	8192	2048	64	256	128	256	512	1024	1536	2048
Used Subcarriers	1536	768	384	192	6817	1705	52	200	76	151	301	601	901	1201
Guardband Ratio	0.25				0.1678	0.1675	0.1875	0.2185	0.41					
Subcarrier Spacing (KHz)	1	2	4	8	1.116(1)	4.464(1)	312.5	125(2)	15					
FFT Period (μs)	1000	500	250	125	896(1)	224(1)	3.2	8(2)	66.7					
Guard Interval (μs)	246	123	62	31	224, 112, 56, 28(1)	56, 28, 14, 7(1)	0.8	2, 1, 0.5, 0.25(2)	4.67, 16.67					
Guard Interval Ratio	1/4				1/4, 1/8, 1/16, 1/32		1/4	1/4, 1/8, 1/16, 1/32	9/128, 1/4					
Constellation	DQPSK				QPSK, 16-QAM, 64-QAM		BPSK, QPSK, 16-QAM, 64-QAM	BPSK, QPSK, 16-QAM, 64-QAM	QPSK, 16-QAM, 64-QAM					
Maximum Data Rate (bps)	1.8M				31.67M(1)		54M	104.7M(2)	> 100M(3)					

(1): For a 8-MHz channel (2): For a 28-MHz channel (3): For a 20-MHz channel

Petitioner Samsung Ex-1026, 0040

The digital video broadcasting-terrestrial (DVB-T) system delivers digital video signals in the frequency band from 400 to 800 MHz [10]. Each DVB-T signal can be accommodated in a channel with either 6,7 or 8-MHz bandwidth. DVB-T supports 2K FFT and 8K FFT. In the 2K mode, 1705 out of 2048 subcarriers are used to carry the data and pilot signals. As the symbol period is shorter, this mode is designed to tackle the scenarios with fast-changing channel conditions. On the other hand, the 8K mode has a longer guard interval, making it capable of coping with multipath channels with a long delay spread. Of all the 8192 subcarriers in the 8K-mode DVB-T signal, 6817 are used for data and pilot transmission. With the high-order 64-QAM constellation, the highest data rate of the 8-MHz DVB-T system is 31.67 Mbps with a code rate of 7/8.

The IEEE 802.11a/g wireless LAN can operate in ISM band around 2.4–2.5 GHz and in U-NII band around 5.8 GHz [11]. A total of 64 subcarriers divide the whole 20-MHz bandwidth. This high-speed wireless LAN standard allows stationary or nomadic reception in indoor environments. In the system, the guard band ratio and the guard interval ratio are 0.1875 and 0.25, respectively, and the maximum information data rate is 54 Mbps.

The wireless MAN, IEEE 802.16e-2005, contains three modes: single-carrier, OFDM and OFDMA [12]. In the OFDM mode, an FFT size of 256 is used for various bandwidths ranging from 1.5 to 28 MHz and the carrier frequency is from 2 to 11 GHz. The sampling frequency can be 8/7 or 7/6 times the bandwidth and the guard interval ratio varies from 1/32 to 1/4. Protected by a code with rate 3/4, the highest data rate is 104.7 Mbps, with the smallest guard interval overhead and the widest channel bandwidth.

The 3GPP-long-term evolution (3GPP-LTE) evolved universal terrestrial radio access (E-UTRA) project, intending to improve the UMT mobile phone standard to meet the so-called fourth-generation (4G) requirements, adopts OFDMA with various FFT sizes in the downlink transmission [13]. Its bandwidth varies from 1.25 to 20 MHz to provide flexible allocation. The subcarrier spacing is fixed at 15 KHz, which results in possible FFT sizes ranging from 128 to 2048. The optimal cell size is 5 Km so that 16.67 μs is reserved for the guard interval. The target downlink data rate is 100 Mbps in the 20-MHz bandwidth.

Bibliography

[1] A. B. Carlson, P. B. Crilly and J. C. Rutledge, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 4th ed., Boston, MA: McGraw-Hill, 2002.

[2] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA: McGraw-Hill, 2001.

[3] A. V. Oppenheim and R. W. Schafer, *Digital-Time Signal Processing*. Eaglewood Cliffs, NJ: Prentice Hall, 1989.

[4] B. R. Salzberg,'Performance of an efficient parallel data transmission system', *IEEE Trans. Commun.*, vol. 15, Dec. 1967, pp. 805–813.

[5] R. O'Neill and L. B. Lopes, 'Envelope variations and spectral splatter in clipped multicarrier signals', in *Proc. of IEEE Symposium on Personal Indoor and Mobile Radio Communications*, Sep. 1995, pp. 71–75.

[6] A. E. Jones, T. A. Wilkinson and S. K. Barton, 'Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes', *IEE Electronics Letters*, pp. 2098–2099, Dec. 1994.

[7] J. A. Davis and J. Jedwab, 'Peak-to-mean power control and error correction for OFDM transmission using Golay sequences and Reed-Muller codes', *IEE Electronics Letters*, Feb. 1997, pp. 267–268.

[8] A. Aggarwal and T. H. Meng, 'Minimizing the peak-to-average power ratio of OFDM signals via convex optimization', in *Proc. of IEEE Global Telecommunications Conference*, Dec. 2003, pp. 2385–2389.

[9] ETSI, 'Digital Audio Broadcasting (DAB), guidelines and rules for implementation and operation, part 1: system outline', ETSI Technical Report, TR 101 496-1 v1.1.1, Nov. 2000.

Petitioner Samsung Ex-1026, 0041

[10] ETSI, 'Digital Video Broadcasting (DVB), framing structure, channel coding and modulation for digital terrestrial television', ETSI European Standard, EN 300 744 v1.2.1, Jul. 1999.
[11] IEEE, 'Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 GHz Band', IEEE Standard 802.11a-1999, Sep. 1999.
[12] IEEE, 'Draft IEEE standard for local and metropolitan area networks, part 16: Air Interface for fixed and mobile broadband wireless access systems, amendment for physical and medium access control layers for combined fixed and mobile operation in licensed bands', IEEE P802.16e/D9, Jun. 2005.
[13] 3GPP, 'Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7)', 3GPP TR 25.814 V7.1.0, Sep. 2006.

Petitioner Samsung Ex-1026, 0042

4

Signal Propagation and Channel Model

Wireless communication signals experience many adverse effects as they travel through the transmitter/receiver electronics, antennas and the radio-frequency channel. Understanding these effects and modeling them accurately will lay a solid foundation to wireless receiver design.

4.1 Introduction

Wireless communication systems offer many advantages, such as mobility, easy access and installation; however, they also suffer more limitations than wireline transmission systems, such as limited capacity, spectrum shortage and service quality uncertainties. In wireless communication systems, signals are transmitted over the air in the radio-frequency (RF) band. There are numerous non-ideal factors that affect the quality of the received signals and, thus, the reliability of wireless communication. As a result, these wireless channel effects place some fundamental limitation on the capability of wireless communication systems.

Wireless radio channels are extremely dynamic and time-varying due to channel variations and user movement. Making matters worse is the fact that wireless transmission is usually off and on owing to the inconsistent nature of the wireless channels. Hence, most analysis and estimation of the wireless channels must be re-started every so often. In addition to the effects of the wireless channels, the transmission signal also suffers distortions caused by RF front-end circuits. To construct an equivalent baseband channel model for a wireless communication system, signal propagation characteristics as well as many circuit impairments must be taken into consideration. Baseband channel modeling is very crucial to the design of wireless baseband receivers, as any inaccuracy inevitably leads to either deficient receiver designs or costly receivers that are over-designed.

4.2 Wireless Channel Propagation

In a wireless communication system, due to numerous obstacles in the propagation environment, RF signals usually travel along several different paths, arising from reflection,

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0043

scattering and diffraction. The collective effect of the above mechanisms is random and complicated. Thus, usually, these different propagation mechanisms are combined and categorized as *path loss*, *shadowing* and *multipath fading*.

4.2.1 Path Loss and Shadowing

Radio signals propagating through the free space attenuate at a rate that is inversely proportional to the squared distance between the transmitter and the receiver:

$$P_r(d) \propto \left(\frac{\lambda}{4\pi d}\right)^2, \tag{4.1}$$

where $P_r(d)$ is the received power at distance d from the transmitter and λ is the wavelength of the carrier signal. The above scenario assumes no obstruction between the transmitter and receiver and is called *line of sight* (LOS). Numerous measurement experiments indicate that such LOS channel is not the norm in most wireless communication systems and, in non-LOS (NLOS) situations, the attenuation will be more severe.

Path loss represents the degree of signal power attenuation as the distance between the transmitter and receiver increases. A simplified log-distance path loss model [1] is given by

$$L = 10n \log\left(\frac{d}{d_{ref}}\right) + L_{ref}, \tag{4.2}$$

where n is the path loss exponent; d is the separation between the transmitter and receiver; and L_{ref} is the path loss value in free space for the reference distance d_{ref}. According to the terrain and obstacles, the path loss exponent n varies from 2 (free space) to 6 (severe obstruction). Note that in order to obtain the exact path loss at a certain location, it is necessary to have L_{ref}. Usually, the reference path loss is either calculated using the free-space formula or averaged over measurements at distance d_{ref} from the transmitter.

The previous formula of path loss fails to consider the fact that two locations with identical distance from the transmitter may experience quite different signal attenuation due to obstacle positions and surrounding environments. Measurements after measurements show that the actual signal loss at a distance d is random with a log-normal distribution [1]. Shadowing describes such a random effect and it is superposed on the path loss model introduced before. Shadowing varies faster with distance than path loss and, in a range of hundreds of meters, the signal strength variation can be 20 dB. Including the log-normal-distributed shadowing effect, the total loss then is given by

$$L = 10n \log\left(\frac{d}{d_{ref}}\right) + L_{ref} + X, \tag{4.3}$$

where X represents the shadowing effect after transforming to dB, and is a normal-distributed (Gaussian-distributed) random variable. The standard deviation of the shadowing variable can be higher than 10 dB in some cellular and indoor environments [1,2].

Petitioner Samsung Ex-1026, 0044

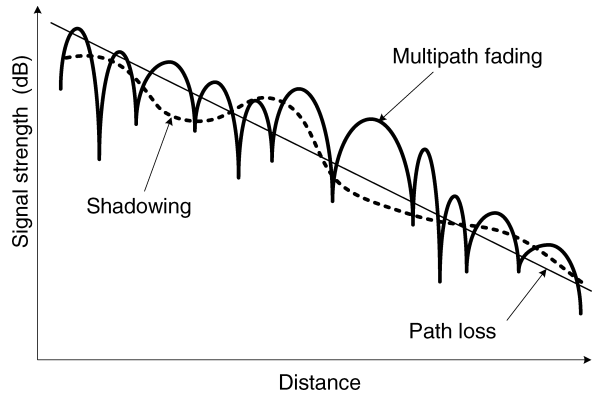


Figure 4.1. Illustration of path loss, shadowing and multipath fading

4.2.2 Multipath Fading

In addition to large-scale channel attenuation effects, namely path loss and shadowing, there exist rapid channel fluctuations within a small region. For instance, signal strength variation can reach up to 40 dB in one half-wavelength (3 cm in a system with 5-GHz carrier). Such fading is caused by the combined effect of different versions of the transmitted signal that arrive at the receiver along different paths. It is thus called multipath fading. As they arrive at the receiver, the different versions of the transmitted signal are constructively and destructively combined in amplitude and phase. Since the signals interfere with one another on the scale of a fraction of a wavelength, the total received signal power can vary within a small region on the order of the wavelength of the carrier signal. Figure 4.1 depicts an illustration of these three effects that cause fluctuation in the strength of signals that have traversed the channel. Note that the figure is not drawn to scale, as the multipath fading actually fluctuates within a much smaller region than shown in the figure.

To describe multipath fading, one can deterministically specify the delay, phase shift and attenuation of each path. This approach, however, inevitably uses too many resources, since the number of paths required to yield a reasonably accurate time-varying channel can be prohibitively large. A statistical fading model assumes that the received signal actually consists of an infinite number of independent signals that impinge on the receiver from different directions (and thus phase shifts), albeit with approximately the same degree of attenuation. From the central limit theorem, the baseband received signal will experience a complex channel gain that has independent Gaussian-distributed real and imaginary parts with zero mean and equal variance. Consequently, the amplitude of the complex baseband channel gain is Rayleigh-distributed [1–3]. The Rayleigh fading model describes the case in which all paths from the transmitter to the receiver are obstructed to a certain degree. There are cases in which there exists a line-of-sight (LOS) path between the transmitter and the receiver, when the stronger signal component along the LOS path must be considered in addition to all other non-LOS components. In such LOS cases, the amplitude of the baseband channel gain becomes Rician-distributed [1–3].

Petitioner Samsung Ex-1026, 0045

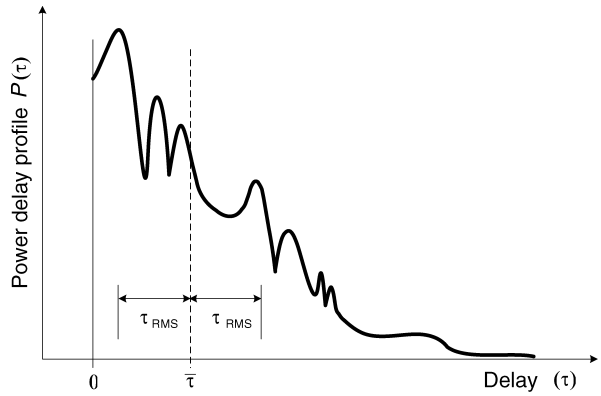


Figure 4.2. Typical multipath channel power delay profile

4.2.3 Multipath Channel Parameters

In a time-varying multipath channel, the signal will be subject to several types of signal dispersion: delay dispersion, spectrum dispersion and dispersion in arrival angle. In the following, several parameters that help to characterize the extent of these types of dispersion will be introduced.

Delay Spread: Time Dispersion

In the multipath channel, the arrival times of the different versions of the transmitted signal along distinctive paths will be spread in time. This phenomenon is called *delay spread.* To characterize the extent of channel delay spread, one needs to first find the *power delay profile* (PDP) of a channel, which describes the time distribution of the received signal power when an impulse waveform is transmitted through the channel under consideration. Assume that the baseband complex impulse response of a multipath (possibly time-varying) channel is given by

$$h(\tau;t) = \sum_r \beta_r(t) e^{j\theta_r(t)} \delta(\tau - \tau_r), \tag{4.4}$$

where r is the path index; $\beta_r(t)$ is the path gain; $\theta_r(t)$ is the phase shift; τ_r is the time delay of the r-th path; and $\delta(.)$ denotes the Dirac delta function. Then, the power delay profile $P(\tau;t)$ takes the form of

$$P(\tau;t) \equiv |h(\tau;t)|^2 = \sum_r \beta_r^2(t) \delta(\tau - \tau_r). \tag{4.5}$$

Figure 4.2 illustrates the power delay profile of a typical multipath channel.

Petitioner Samsung Ex-1026, 0046

The *RMS (root mean square) delay spread* (τ_{RMS}) is a measure of the amount of signal dispersion in time and is an important indicator for the wireless channel. For simplicity, assume that the channel is stationary, then the mean excess delay $\bar{\tau}$ is defined as

$$\bar{\tau} = \int_{-\infty}^{\infty} \tau P_{\text{norm}}(\tau) d\tau, \tag{4.6}$$

where $P_{\text{norm}}(\tau)$ is the normalized PDP, and

$$P_{\text{norm}}(\tau) = \frac{P(\tau)}{\int_{-\infty}^{\infty} P(\tau) d\tau}. \tag{4.7}$$

Then, the RMS delay spread τ_{RMS} is defined as

$$\tau_{\text{RMS}} = \sqrt{\int_{-\infty}^{\infty} (\tau - \bar{\tau})^2 P_{\text{norm}}(\tau) d\tau}. \tag{4.8}$$

Note that the RMS delay spread, though commonly used to indicate the degree of channel spreading in time, may sometimes be misleading. Two channels with identical RMS delay spread may act quite differently on the transmitted signal. It is therefore good practice to use many instances of the channel with a RMS delay spread in simulation and average the performance results in all the cases.

Taking the Fourier transform of the channel power delay profile yields the auto-correlation of the channel frequency response w.r.t. frequency difference Δf, $\Phi_{HH}(\Delta f)$ [3]. If the channel power delay profile is an impulse function, then the auto-correlation function will become unity over all possible Δf. This implies that with the single-impulse (single-path) channel, the channel frequency responses at any two different frequencies are identical (100% correlated) or, equivalently, the channel is all-pass.

On the other hand, if the delay spread is relatively large, the auto-correlation function falls off more rapidly. Define the *coherence bandwidth*, B_c, of a channel as the frequency displacement (Δf) at which $|\Phi_{HH}(\Delta f)|$ becomes less than a threshold. In other words, the channel has rather different characteristics at two frequencies with a difference larger than the coherence bandwidth. The exact formula of the coherence bandwidth depends inversely on the RMS delay spread and linearly on a constant that, in turn, depends on the threshold and the shape of $|\Phi_{HH}(\Delta f)|$:

$$B_c = \frac{1}{K \tau_{\text{RMS}}}, \tag{4.9}$$

where K ranges from 5 to 50 for a threshold from 0.5 to 0.9 [3].

According to the relation between the coherence bandwidth and the signal bandwidth, a wireless communication system can be categorized as under *flat fading* or under *frequency-selective fading*. Flat fading refers to the case in which the bandwidth of the transmitted signal is much smaller than the coherence bandwidth, making the channel frequency response rather coherent (flat) inside the signal band. On the other hand, in the frequency-selective fading case, the signal bandwidth is larger than the channel coherence

Petitioner Samsung Ex-1026, 0047

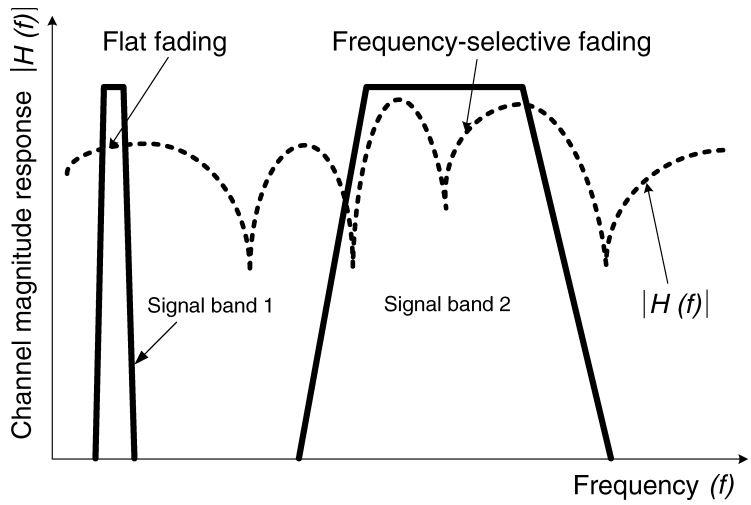


Figure 4.3. Channel frequency response and signal bands for flat fading and frequency-selective fading

bandwidth. Consequently, different frequency components of the signal suffer different amplitude attenuations and phase shifts, making it harder to equalize the received signal for accurate detection. Figure 4.3 illustrates the scenarios of flat fading and frequency-selective fading.

Doppler Spread: Frequency Dispersion

Wireless communication frees the users from being attached to a device and gives them the freedom of moving around during communication. When either the transmitter, the receiver or the scatterers is in motion, the received signal will be dispersed in frequency as a result of the Doppler effect. For instance, if a pure tone is transmitted from a fixed transmitter, then a moving receiver will receive a signal whose spectrum has been shifted by a finite frequency. The frequency shift is related to the speed of the motion and the angle between the signal arrival direction and the direction of motion. Figure 4.4 depicts a typical scenario, in which a receiver is moving with a constant speed V; the Doppler frequency shift f_D between the signals received at point A and point B is

$$f_D = \frac{\Delta\phi}{2\pi\Delta t} = \frac{V\cos\theta}{\lambda}, \tag{4.10}$$

where

$$\Delta\phi = \frac{2\pi\Delta l \cdot \cos\theta}{\lambda} = \frac{2\pi V \cdot \Delta t \cdot \cos\theta}{\lambda} \tag{4.11}$$

Petitioner Samsung Ex-1026, 0048

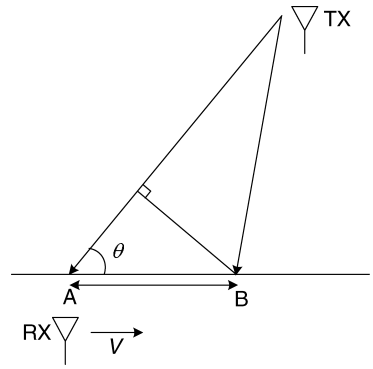


Figure 4.4. Illustration of a signal reception under the Doppler effect

is the phase shift between the two locations; l is the distance between point A and point B; and λ is the wavelength of the carrier signal.

The above scenario considers only one signal path. In practice, there can be several paths between the transmitter and the receiver. Then, the received signal power spectral density will be spread over a bandwidth limited by the *maximum Doppler frequency*, $f_m = V/\lambda$. To characterize the extent of such frequency spreading of the received signal, *Doppler spread* of a channel can be defined. At first, transmit a pure tone with frequency f_c and measure the received signal power spectrum. This spectrum is called the *Doppler spectrum* and can be denoted as $S_D(f)$. Normally, the spectrum will be spread across $[f_c - f_m, f_c + f_m]$. The mean frequency of this spectrum is calculated according to

$$\bar{f} = \int_{-\infty}^{\infty} f S_{\text{norm}}(f) df, \tag{4.12}$$

where

$$S_{\text{norm}}(f) = \frac{S_D(f)}{\int_{-\infty}^{\infty} S_D(f) df}.$$

Then, the RMS Doppler spread B_{RMS} is defined as

$$B_{\text{RMS}} = \sqrt{\int_{-\infty}^{\infty} (f - \bar{f})^2 S_{\text{norm}}(f) df}. \tag{4.13}$$

The Doppler spread is obviously bounded by the maximum Doppler frequency. Its exact value depends on the shape of the Doppler spectrum.

A classical Doppler spectrum is derived by assuming that the local scatterers are uniformly distributed and the receiving antenna is omnidirectional. Then, all scattered signals arrive with the same amplitude but their phases and arrival angles are both uniformly

Petitioner Samsung Ex-1026, 0049

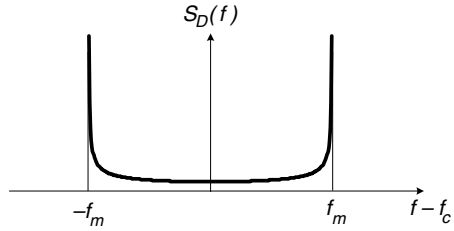


Figure 4.5. The classical Doppler spectrum

distributed. In this case, the combined received signal power around f will be contributed by the scattered components arriving from an angle at around

$$\theta = \cos^{-1}\left(\frac{f - f_c}{f_m}\right). \tag{4.14}$$

Now, let the two derivations signal power be equivalent; then, one has

$$S_D(f)|df| = S_D(\theta)|d\theta|, \tag{4.15}$$

where $S_D(f)$ is the power spectral density of the received signal with respect to frequency, while $S_D(\theta)$ is the received signal power spectral density with respect to the arrival angle. Note that $S_D(\theta)$ is a constant due to the above uniform-distribution assumption. Consequently, $S_D(f)$ can be derived by differentiating Equation (4.14) with respect to f and

$$\begin{aligned} S_D(f) &\propto \left|\frac{d\theta}{df}\right| \\ &= \left|\frac{1}{\sqrt{1-(f-f_c)^2/f_m^2}}\right| \left|\frac{d((f-f_c)/f_m)}{df}\right| \\ &= \frac{1}{\sqrt{f_m^2-(f-f_c)^2}}, \quad f_c - f_m < f < f_c + f_m. \end{aligned} \tag{4.16}$$

Figure 4.5 plots the classical Doppler spectrum with center frequency f_c and Doppler frequency f_m.

To characterize the time variation of the channel as opposed to the signal spectrum spreading, one can compute the auto-correlation function of the channel response at a certain frequency w.r.t. some time difference Δt, $\Phi_{HH}(\Delta t)$. The coherence time T_c is defined as the size of the time interval when the magnitude of $\Phi_{HH}(\Delta t)$ stays above a certain threshold. In other words, within a time interval shorter than the coherence time, the channel can be assumed time-invariant. Obviously, slower relative motion introduces smaller f_m, and thus a

Petitioner Samsung Ex-1026, 0050

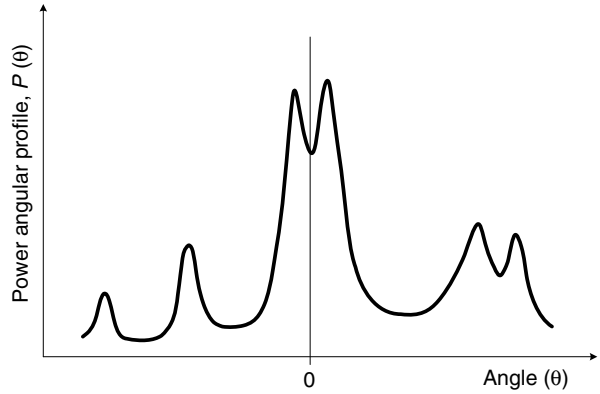


Figure 4.6. A typical signal power angular profile

slower varying channel and a larger coherence time. A rule of thumb between the coherence time and the maximum Doppler frequency is given by [1]

$$T_c \approx \frac{0.423}{f_m}. \tag{4.17}$$

Depending on how fast the channel is varying with respect to the signal symbol period, a channel can also be categorized as *slow fading* or *fast fading*. In the slow fading channel, the channel impulse response remains unchanged in one symbol time (T_{sym}), i.e. $T_{\text{sym}} \ll T_c$. On the other hand, when $T_{\text{sym}} > T_c$, the signal within one symbol receives channel effects that are changing quite quickly and, in this case, the channel is under fast fading. It should be noted that the slow or fast fading channel is related to only the coherence time of the channel and not the actual delay or the delay spread of the channel.

Angle Spread: Space Dispersion

Through a multipath wireless channel, the transmitted signals along different paths not only experience different levels of attenuation and phase shift, but also arrive at the receiver from different directions, namely different angle-of-arrivals (AOA). When there is only one LOS path, the signal arrives from a single angle. As there are more and more paths, the range of AOA gets broader. Figure 4.6 shows a typical power spectrum with respect to the AOA. Similarly, at the transmitter end, one can also define angle-of-departure (AOD). The *angle spread* of AOA and that of AOD are both important parameters that describe the extent of signal spreading in spatial angle.

To compute the angle spread of the AOA, one first finds the mean AOA from the angle power spectrum, $P(\theta)$, according to

$$\bar{\theta} = \int_{-\infty}^{\infty} \theta P_{\text{norm}}(\theta) d\theta, \tag{4.18}$$

Petitioner Samsung Ex-1026, 0051

where $P_{\text{norm}}(\theta)$ is the normalized angle power spectrum,

$$P_{\text{norm}}(\theta) = \frac{P(\theta)}{\int_{-\infty}^{\infty} P(\theta)d\theta}. \tag{4.19}$$

Then, the RMS angle spread θ_{RMS} is defined as

$$\theta_{\text{RMS}} = \sqrt{\int_{-\infty}^{\infty} (\theta - \bar{\theta})^2 P_{\text{norm}}(\theta)d\theta}. \tag{4.20}$$

Since the power of the received signal fluctuates as the receiver antenna rotates, angle spreading results in selective fading due to antenna orientation. In addition, the characteristics of a multipath channel also depend on the location of the transmitter or the receiver. Consider a fixed transmitter; as the receiver moves, the channel response will also change accordingly. As in the previous two situations, one can compute or measure the auto-correlation function of the channel response at a certain time and frequency w.r.t. space displacement Δr, $\Phi_{HH}(\Delta r)$. Then, the degree of space selective fading can be characterized by the coherence distance D_c, which is defined as the spatial displacement when the magnitude of the auto-correlation function remains higher than a threshold. In other words, the coherence distance of a multipath channel is the minimum distance between two nearby receiver locations that will receive signals affected quite differently by the channel.

4.3 Front-End Electronics Effects

In baseband receiver design, the designer has to consider not only the degradation caused by the imperfect channel and noise, but also the non-idealities resulting from the RF and analog parts in the transmitter and the receiver. These non-idealities include carrier frequency offset, sampling clock offset, phase noise, IQ imbalance, DC offset and power amplifier (PA) non-linearity.

4.3.1 Carrier Frequency Offset

Carrier frequency offset (CFO) occurs when the local oscillator signal for down conversion in the receiver does not synchronize with the carrier signal contained in the received signal. This phenomenon can be attributed to two factors: frequency mismatch in the transmitter and the receiver oscillators, and the Doppler effect as the transmitter and/or the receiver is moving. When this occurs, the received signal will be shifted in frequency, as shown in Figure 4.7.

For an OFDM system, the orthogonality among subcarriers is maintained only if the receiver uses a local oscillation signal that is synchronous with the carrier signal contained in the received signal. Otherwise, mismatch in carrier frequency can result in inter-carrier interference (ICI). Practically, the oscillators in the transmitter and the receiver can never be oscillating at identical frequency. Hence, carrier frequency offset always exists, even if there is no Doppler effect.

In a standard-compliant communication system, such as the IEEE 802.11a wireless LAN, the oscillator precision tolerance is specified to be less than ±20 ppm, so that the CFO is in

Petitioner Samsung Ex-1026, 0052

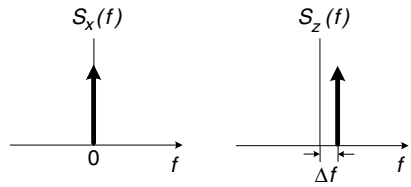


Figure 4.7. The received baseband signal spectrum $(S_z(f))$ is shifted by the CFO (Δf) with respect to the transmitted signal $(S_x(f))$

the range from −40 ppm to +40 ppm. For example, if the TX oscillator runs at a frequency that is 20 ppm above the nominal frequency and if the RX oscillator is running at 20 ppm below, then the received baseband signal will have a CFO of 40 ppm. With a carrier frequency of 5.2 GHz in this standard, the CFO is up to ±208 KHz. In addition, if the transmitter and/or the receiver is moving, the Doppler effect adds some hundreds of Hz in frequency spreading. Compared with the CFO resulting from the oscillator mismatch, the Doppler effect in this case is relatively minor.

4.3.2 Sampling Clock Offset

Sampling clock offset (SCO) is quite similar to the CFO, as they both originate from oscillator mismatch and the Doppler effect. When oscillators with mismatched frequencies are used to drive the sampling clocks of the DAC in the transmitter and the ADC in the receiver, SCO may occur. Figure 4.8(a) illustrates an example in which the sampling clock mismatch causes the received waveform to be sampled at time instants that are progressively skewing.

Motion between the transmitter and the receiver effectively makes the signal waveform to contract or expand in time. Therefore, even without sampling clock mismatch, the sampled waveform at the receiver can still be suffering errors in sampling times, as in the previous case. Figure 4.8(b) shows a case in which the Doppler effect causes the received waveform to expand and thus sample-time error occurs, even though the ADC and DAC are synchronously clocked.

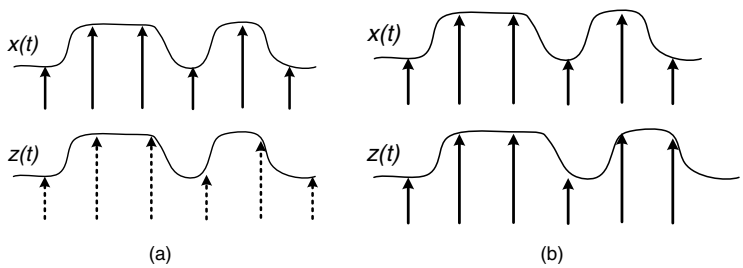


Figure 4.8. (a) Sampling error occurs when the baseband transmitted signal $x(t)$ and the baseband received signal $z(t)$ are sampled at different clock rates. (b) Baseband received signal $z(t)$ is expanded due to the Doppler effect, resulting in sampling error, even without clock rate mismatch

Petitioner Samsung Ex-1026, 0053

4.3.3 Phase Noise

Phase noise in an oscillation signal is related to its *jitter*. The former is usually described in the frequency domain as the noise spectrum centerd at the oscillation frequency, while the latter is represented in the time domain to measure how precise the oscillation periodicity is. Phase noise occurs because oscillators cannot generate pure sinusoidal waves with impulsive spectra. Real-life oscillation signal spectra have a vestigial sideband around the oscillation frequency.

The phase noise θ_n of a free-running oscillator output signal can be modeled as a discrete-time Wiener-Leỳy process [4] and

$$\theta_n(t_n) = \theta_n(t_{n-1}) + \psi(t_n), \tag{4.21}$$

where $\psi(t_n)$ is the phase increment at time t_n. Note that $\psi(t_n)$ is an i.i.d. Gaussian-distributed random variable with variance

$$\sigma_\psi^2 = 2\pi B T_s, \tag{4.22}$$

where B represents the two-sided 3-dB bandwidth, and T_s is the sampling interval. The power spectral density of the local oscillation (LO) signal with such phase noise, $e^{j2\pi f_c t + j\theta_n(t)}$, is given by [5]

$$\frac{2}{\pi B}\frac{1}{\left[1+\left(\frac{2(f-f_c)}{B}\right)^2\right]}. \tag{4.23}$$

When such an LO signal is used in the receiver down-converting mixer, the resulting baseband signal spectrum will be the passband signal spectrum convolved with the LO spectrum with non-zero bandwidth. As a result, the frequency components in the original signal will now be 'blended' with their neighbouring components. OFDM receivers are very sensitive to this spectral spreading, as it may introduce common phase error on all subcarriers as well as ICI among subcarriers [6]. In practical OFDM receivers, the LO signal is generated by a frequency synthesizer based on a phase-locked loop (PLL). In the PLL, the oscillation frequency of the controlled oscillator is usually locked to a free-running oscillator, such as a crystal oscillator. In this case, [7] offers a typical spectrum for the phase-noise-inflicted LO signal.

4.3.4 IQ Imbalance and DC Offset

Direct conversion, also known as zero-IF or homodyne, receiver structure translates the RF (passband) signal directly from the carrier frequency (f_c) to DC (baseband) using one stage of mixing. The traditional heterodyne receiver structure needs an intermediate frequency (IF) stage between the RF and baseband signals. The direct conversion receiver structure has several advantages: no IF stage, no image-rejection filter and easy integration due to low component count. However, a direction conversion RF front-end suffers from two major drawbacks, namely IQ imbalance and DC offset [8].

Petitioner Samsung Ex-1026, 0054

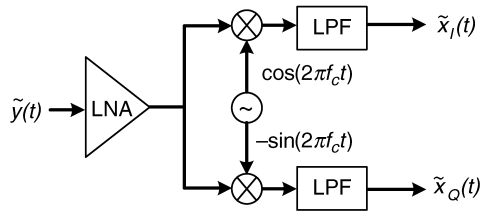


Figure 4.9. Architecture of the direct conversion receiver

IQ Imbalance

A direct conversion receiver uses two quadrature sinusoidal signals to perform the so-called quadrature down-conversion. This process requires shifting the local oscillator (LO) by 90 degrees to produce the quadrature sinusoidal component. Figure 4.9 depicts the structure of such a direct conversion receiver. When there exist mismatches between the gain and phase of the two sinusoidal signals and/or mismatch along the two branches of down-conversion mixers, gain amplifier and low-pass filters, the quadrature baseband signals will be corrupted.

Suppose the received passband signal is identical to the transmitted signal and is given as

$$y(t) = \mathrm{Re}\{x(t)e^{j2\pi f_c t}\} = x_I(t)\cos(2\pi f_c t) - x_Q(t)\sin(2\pi f_c t), \tag{4.24}$$

where $x(t) = x_I(t) + jx_Q(t)$ is the transmitted baseband signal. Assume that the gain error is $20\log\frac{1+\alpha}{1-\alpha}$ dB and the phase error is ϕ degrees. Then we can model such imbalance using mismatched local oscillator output signals

$$\begin{aligned} &2(1+\alpha)\cos(2\pi f_c t - \phi/2) \\ &-2(1-\alpha)\sin(2\pi f_c t + \phi/2). \end{aligned} \tag{4.25}$$

Multiplying the passband signal by the two LO signals and passing through a pair of low-pass filters, one has the demodulated baseband signals as

$$\begin{aligned} \tilde{x}_I(t) &= (1+\alpha)\left[x_I(t)\cos\left(\frac{\phi}{2}\right) - x_Q(t)\sin\left(\frac{\phi}{2}\right)\right] \\ \tilde{x}_Q(t) &= (1-\alpha)\left[x_Q(t)\cos\left(\frac{\phi}{2}\right) - x_I(t)\sin\left(\frac{\phi}{2}\right)\right]. \end{aligned} \tag{4.26}$$

The above equation clearly indicates that IQ imbalance causes interference between the I and Q signals. Figure 4.10 depicts an example of how the IQ imbalance can distort the QPSK baseband signals.

To analyze IQ imbalance in the frequency domain, Equation (4.26) can be rewritten as

$$\begin{aligned} \tilde{x}(t) &= \tilde{x}_I(t) + j\tilde{x}_Q(t) \\ &= \left[\cos\left(\frac{\phi}{2}\right) + j\alpha\sin\left(\frac{\phi}{2}\right)\right]x(t) + \left[\alpha\cos\left(\frac{\phi}{2}\right) - j\sin\left(\frac{\phi}{2}\right)\right]x^*(t) \\ &= Ax(t) + Bx^*(t), \end{aligned} \tag{4.27}$$

Petitioner Samsung Ex-1026, 0055

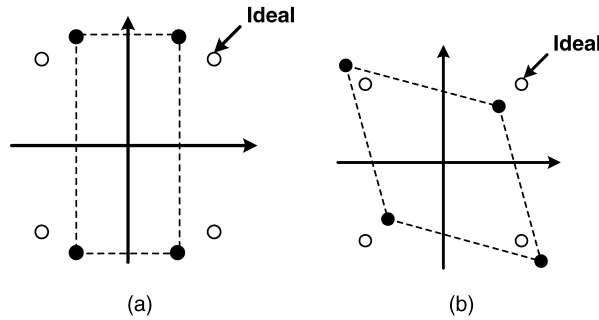


Figure 4.10. IQ imbalance effect on QPSK signals: (a) gain error in LO signals, (b) phase error in LO signals

where * denotes complex conjugate. In an OFDM system, the baseband signal consists of several subcarriers; complex conjugating the baseband signal of the k-th subcarrier carrying data X_k is identical to carrying X_k^* on the $(-k)$-th subcarrier

$$\left((X_{k,I} + jX_{k,Q})e^{j2\pi kf_s t}\right)^* = (X_{k,I} - jX_{k,Q})e^{-j2\pi kf_s t} = X_k^* e^{j2\pi(-k)f_s t}.$$

Equivalently, the received baseband OFDM signal under the IQ imbalance effect is given by

$$\tilde{X}_k = AX_k + BX_{-k}^*. \tag{4.28}$$

In conclusion, besides a complex gain imposed on current subcarrier data X_k, IQ imbalance also introduces ICI from the mirror subcarrier. The ICI term makes OFDM receivers very sensitive to the IQ imbalance effect. To solve this problem, the designer can request stringent specification of the matching of the two branches in the front-end or compensate the imbalance in the baseband receiver [9].

DC Offset

DC offset arises from self-mixing and nonlinearity in the front-end [8,10]. Components with even-order nonlinearity will generate a large DC value at its output when there is a strong in-band input. For example, an amplifier has a nonlinear output response as

$$\alpha_1 x(t) + \alpha_2 x^2(t).$$

Assume that there is a single-tone interferer, $A\cos(2\pi ft)$, within the passband of this amplifier; then, the amplifier output is

$$(1/2)\alpha_2 A^2 + \alpha_1 A\cos(2\pi ft) + (1/2)\alpha_2 A^2 \cos(4\pi ft). \tag{4.29}$$

Note that the first term is a DC component related to the interferer.

Petitioner Samsung Ex-1026, 0056

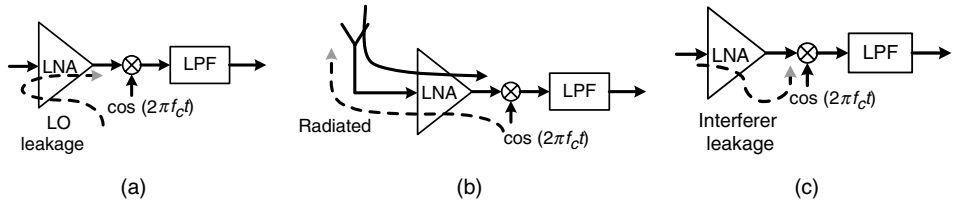


Figure 4.11. DC offset generation mechanisms: (a) LO leakage, (b) LO re-radiation and (c) interference leakage

The self-mixing is due to the finite isolation between the LO and RF input ports of a mixer, which is typical of silicon-based ICs. Self-mixing has three possible mechanisms: (1) LO leakage (static DC), (2) LO re-radiation (dynamic DC), and (3) strong in-band interferers. The LO signal can leak with sufficiently high power to the input port of the mixer or even that of the low-noise amplifier (LNA). The leaked LO signal can then come back into the mixer RF input port and mix with itself, thereby generating a static DC component at the mixer output, as depicted in Figure 4.11(a). The LO signal also can leak to the antenna and be re-radiated; the re-radiated LO signal can reflect off obstructions and be recaptured by the front-end, resulting in a time-varying or dynamic DC offset, as shown in Figure 4.11(b). A strong nearby interferer, such as another user's leaked LO, can also find a path to the mixer LO input port and mix with itself (interferer self-mixing) to produce a dynamic DC offset. Figure 4.11(c) illustrates such a phenomenon.

The DC offset can be much stronger than the desired signal, such as more than 40 dB in a GSM system. When this happens, the DC offset will saturate all the following stages and make the receiver fail. Popular schemes for handling DC offset include AC-coupling, offset estimation/cancelling and many other methods that use adaptive techniques [8].

4.3.5 Power Amplifier Nonlinearity

The power amplifier (PA) in the transmitter is often one of the most power-consuming blocks in a wireless communication device. The efficiency of the power amplifier very likely determines the system power consumption level and, thus, the battery life of the communication device. The PA efficiency is measured in terms of the output power delivered to the load divided by the power delivered from the power supply.

No different from other amplifiers, PAs exhibit nonlinear behavior and saturating gain at large input levels. Their transfer characteristics may exhibit a certain degree of nonlinearity, depending on the specifications. In Section 2.2.2, a piecewise linear PA transfer characteristic has been used to explain the output back-off in PA operation. When taking power efficiency into account, PAs usually operate near the saturation region, which is highly nonlinear. Nonlinear amplification distorts transmission signal, and OFDM signals in particular are very sensitive to this effect. Hence, PA linearity becomes an important issue in wireless communication system design.

The linearity of a PA is often characterized by two input-power levels: P_{1dB} and IIP3. As shown in Figure 4.12, the 1-dB compression point (P_{1dB}) is the input power level at

Petitioner Samsung Ex-1026, 0057

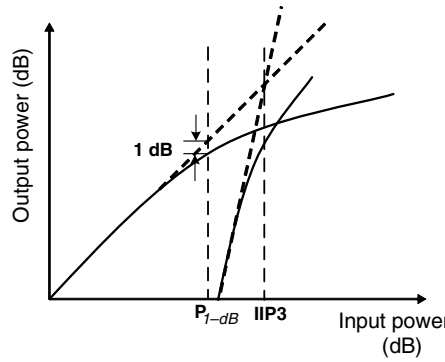


Figure 4.12. Illustration of PA nonlinearity

which the actual output power is exactly 1 dB lower than the ideal output power of a perfectly linear amplifier. The input referred third-order intercept point (IIP3) refers to the input power level at which two output power versus input power characteristics intercept. One line indicates the extrapolated first-harmonic output power versus the input power and the other plots the extrapolated third-harmonic output power versus the input power.

Constellation Distortion

The nonlinear characteristics of power amplifiers introduce two types of distortion: AM/AM (amplitude modulation/amplitude modulation) and AM/PM (amplitude modulation/phase modulation) effects. They result in output signal amplitude and phase modulation when the input signal envelope fluctuates. This can bring about severe signal constellation distortion, especially when non-constant-envelope modulation schemes are used, which is very often the case in spectral-efficient communication systems.

Inter-Modulation Distortion

Using a single tone as the input of a nonlinear PA, the output waveform will contain the original sine wave as well as its harmonics. These harmonic terms can be eliminated by filtering. However, when more than one tone is present, the distortion may be at the same frequency as the input signal. These distortions are known as *inter-modulation distortion* (IMD), which cannot be eliminated by filtering.

4.4 Channel Model

Considering all the above effects and the multipath fading channel, an equivalent channel model for baseband transceiver design can be constructed. Figure 4.13 illustrates the block diagram of one such baseband channel model. The model includes all the effects discussed above, namely PA nonlinearity, multipath fading, AWGN, CFO, SCO, phase noise, IQ imbalance and DC offset.

Petitioner Samsung Ex-1026, 0058

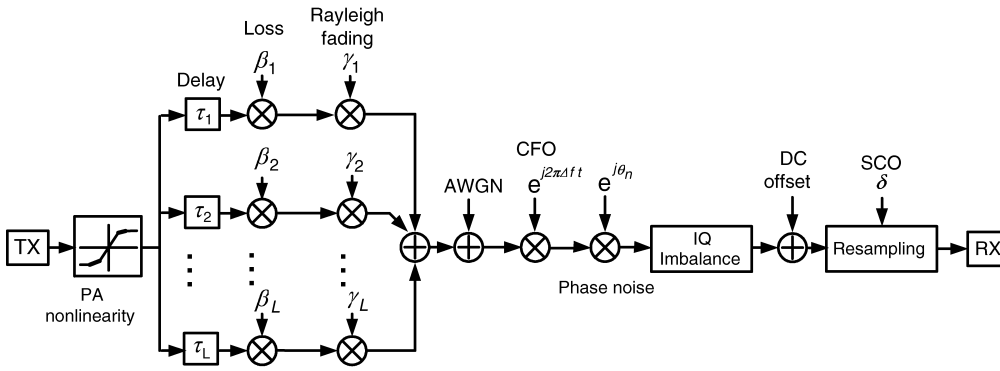


Figure 4.13. Equivalent baseband channel model

4.4.1 Model for Front-End Impairments

Power Amplifier Nonlinearity

A very popular nonlinear power amplifier model is proposed by Saleh [11]. It is a nonlinear memoryless transformation between the complex envelopes of the single-tone input and the single-tone output signals. The amplified signal can be expressed as

$$\tilde{y}(t) = y(t)G(|y(t)|), \tag{4.30}$$

where the amplifier gain is given by

$$G(|y(t)|) = \frac{A(|y(t)|)e^{j\Phi(|y(t)|)}}{|y(t)|}. \tag{4.31}$$

Obviously, the AM/AM and AM/PM effects are now modeled as two functions: $A(\cdot)$ and $\Phi(\cdot)$. For a typical traveling-wave-tube amplifier, they are modeled as

$$A(r) = \frac{\nu_a r}{1 + \eta_a r^2}, \tag{4.32}$$

$$\Phi(r) = \frac{\nu_\phi r^2}{1 + \eta_\phi r^2}, \tag{4.33}$$

where ν_a is the small-signal gain. In [5], a plausible choice of the parameters is given and $\nu_a = 1, \eta_a = 0.25, \nu_\phi = \pi/12$ and $\eta_\phi = 0.25$.

Petitioner Samsung Ex-1026, 0059

For a solid-state PA, the two functions become [5]

$$A(r) = \frac{v_a r}{[1 + (v_a r/A_0)^{2p}]^{1/2p}}, \tag{4.34}$$

$$\Phi(r) = v_\phi \left(\frac{v_a r}{A_0}\right)^4, \tag{4.35}$$

where p controls the transition from the linear region to the saturation region.

Carrier Frequency Offset

In an equivalent baseband channel model, the CFO can be simulated by multiplying $\exp(j2\pi\Delta f t)$ to model the phase rotation, where Δf is the CFO. The formula can be implemented by a phase accumulator with input equal to Δf and its output connected to a phase-to-complex-sinusoid converter. Then, a complex multiplier rotates the baseband signal using the generated complex sinusoidal waveform.

Phase Noise

The Wiener phase noise effect can be simulated by a phase accumulator whose input is a Gaussian random number with a predefined variance. The phase accumulator output can be combined with the phase of the CFO model mentioned above. Then, both the phase noise model and the CFO model can share one set of complex sinusoid generation and complex multiplication.

IQ Imbalance

IQ imbalance can be simulated by computing the gain and phase imbalance and applying them to the baseband signal by means of several real multipliers and adders.

DC offset

The DC offset can be simulated by inflicting a time-invariant DC offset to the baseband signal using complex addition. The dynamic DC offset is usually harder to model, as it is quite random. One can use a slowly varying sinusoidal waveform and/or some randomly on–off rectangular waveform to model dynamic DC offset.

Sampling Clock Offset

To simulate the SCO in the baseband equivalent channel, one can use interpolation to generate samples at times that are not at exact multiples of the sampling interval. An ideal interpolator is a digital filter whose impulse response is a *sinc* function. However, it has infinite taps and is non-causal. Therefore, a low-complexity piecewise parabolic interpolator (Farrow structure) [12] is often adopted to model the SCO effect in the baseband equivalent channel. Note that though the Farrow structure is simple and easy to implement, it can cause severe magnitude degradation, especially for high-frequency components of the signal. As such, a fractional-delay all-pass filter, which has a very flat frequency response, can be adopted instead [13].

Petitioner Samsung Ex-1026, 0060

Additive Noise

Electronic and thermal noises are inherent in wireless channels and analog circuits in the RF front-end. Flicker noise and shot noise are also major sources of noises in electronic devices. For simplicity, however, in a baseband channel model, all noises are combined and modeled as an additive complex Gaussian random process. The real part and imaginary part are independent and identically distributed with zero mean and variance $\sigma_N^2/2$.

4.4.2 Multipath Rayleigh Fader Model

Stationary Multipath Model

In a stationary multipath model, the delays, magnitudes and phase shifts of all the paths are randomly decided but are constant throughout. A well known power delay profile was proposed by Saleh and Valenzuela [14]. In their model, signals along different paths arrive in clusters and the first arrival time of each cluster is modeled as a Poisson process with rate Λ. Within each cluster, subsequent individual paths arrive according to another Possion process with rate λ. Typically, $\lambda \gg \Lambda$; hence, each cluster contains many paths. According to the model, the conditional probability density functions of the l-th cluster and the r-th path of the l-th cluster are given, respectively, by

$$p(T_l|T_{l-1}) = \Lambda e^{-\Lambda(T_l - T_{l-1})}, l = 1, 2, \ldots, \tag{4.36}$$

and

$$p(\tau_r^{(l)}|\tau_{r-1}^{(l)}) = \lambda e^{-\lambda(\tau_r^{(l)} - \tau_{r-1}^{(l)})}, l, r = 1, 2, \ldots, \tag{4.37}$$

where T_l is the arrival time of the l-th cluster and $\tau_r^{(l)}$ is the arrival time of the r-th path, measured from the beginning of the l-th cluster.

Given an instance of all arrival path delays, the corresponding channel impulse response takes the form of

$$h(\tau) = \sum_{l=0}^{\infty}\sum_{r=0}^{\infty} h_{l,r}\delta(\tau - T_l - \tau_r^{(l)}) = \sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \beta_{r,l} e^{j\theta_{r,l}}\delta(\tau - T_l - \tau_r^{(l)}). \tag{4.38}$$

The phase shift of the r-th arrival path in the l-th cluster, $\theta_{r,l}$, is a uniform random variable over $[0, 2\pi)$ and the magnitude of the gain of that path is $\beta_{r,l}$, which is a Rayleigh-distributed random variable, whose mean-squared value is described by the double-exponential decay formula given by

$$\overline{\beta_{r,l}^2} = \beta^2 e^{-T_l/\Gamma_1} e^{-\tau_{r,l}/\Gamma_2}, \tag{4.39}$$

where β^2 is the average power of the zeroth path of the zeroth cluster, and Γ_1 and Γ_2 are power decay time constants for the clusters and paths, respectively. In principle, the number of clusters and the number paths in one cluster can both be infinite. However, practically,

Petitioner Samsung Ex-1026, 0061

they should be chosen appropriately, according to the effective delay spread of the channel being modeled.

Non-Stationary Fading Model

To model the time-varying characteristic of a non-stationary channel, it is necessary to include a time-varying random process for each path gain, shown as γ_r in Figure 4.13. A fading model that assumes many uniformly distributed scattering objects around the receiver was proposed in 1974 by Jakes [15]. Under such an assumption, a fading waveform can be modeled as a sum of sinusoids. When a pure tone is passed through this fading model, the auto-correlation of the received signal becomes a Bessel function in $f_m\tau$, where f_m is the maximum Doppler frequency and τ is the time difference. Taking the Fourier transform of this auto-correlation function, an approximation to the classical U-shaped Doppler spectrum of Equation (4.16) is obtained.

However, the original model suffers from non-zero cross-correlation between the multiple fading waveforms that it generates. Hence, in [15], a modified model with zero cross-correlation among generated fading waveforms was proposed. With N a power of two, the formula for the r-th fading waveform generated by this model takes the form of

$$\gamma_r(t) = \sqrt{\frac{2}{N_0}} \sum_{l=1}^{N_0} A_{r,l}(\cos\phi_l + j\sin\phi_l)\cos(2\pi f_l t + \theta_{r,l}), \tag{4.40}$$

where $N_0 = N/4$; $\phi_l = l\pi/N_0$; $f_l = f_m \cos\alpha_l$; $\alpha_l = 2\pi(l-0.5)/N$; $\theta_{r,l}$ is a random phase shift uniformly distributed in $[0, 2\pi)$; and $A_{r,l}$ denotes the l-th element of the r-th Walsh-Hadamard codeword with length N_0. Given that $\log(N_0) = s+1$, the length-N_0 Walsh-Hadamard code is an orthogonal code given by

$$A_{r,l} = (-1)^{r_s l_s + r_{s-1} l_{s-1} + \cdots + r_0 l_0},$$

where

$$r = r_s 2^s + r_{s-1} 2^{s-1} + \cdots + r_0,$$
$$l = l_s 2^s + l_{s-1} 2^{s-1} + \cdots + l_0,$$

and $r_s, r_{s-1}, \cdots, r_0, l_s, l_{s-1}, \cdots, l_0$ are binary numbers that equal either 1 or 0. Note that the Walsh-Hadamard codewords are orthogonal to one another:

$$\frac{1}{N_0} \sum_{l=0}^{N_0-1} A^*_{r,l} A_{j,l} = \begin{cases} 1, & r = l, \\ 0, & r \neq l. \end{cases} \tag{4.41}$$

4.4.3 Channel Models Used in Standards

Generally speaking, channel power delay profiles that can be used to evaluate and compare receiver quality are usually specified in wireless communication standard documents. For

Petitioner Samsung Ex-1026, 0062

example, in cellular communications, the GSM standard and the third-generation partner project (3GPP) both specify several different modes of operations and their associative channel models [16,17]. In broadcasting, DVB-T standard also specifies two different channel models for the developers to test and evaluate receiver designs [18]. The original 802.11 wireless LAN standard used the Saleh and Valenzuela cluster model and generated channels of different RMS delay spreads for simulation. In the recent IEEE 802.11n task group, the cluster model is extended to MIMO transmission scenarios [19]. In the following, details about some of these channel models are presented.

3GPP

The 3GPP provides several models to facilitate computer simulation of 3G receivers under development [17]. The channel impulse response is assumed to be

$$h(\tau; t) = \sum_r h_r(t)\delta(\tau - \tau_r). \tag{4.42}$$

The path gain $h_r(t)$ is the product of the pass loss β_r and the fading waveform $\gamma_r(t)$, which has a Rayleigh-distributed magnitude and the classical Doppler spectrum specified in Equation (4.16). If a direct wave exists, which corresponds to a non-fading path, $\gamma_r(t)$ becomes a sinusoidal wave with the specified Doppler frequency. The default mobile speeds for the different scenarios suggested by 3GPP are listed in Table 4.1. The models for the typical urban, rural area and hilly terrain scenarios are listed in Tables 4.2, 4.3 and 4.4, respectively [17].

802.11n Wireless LAN

The IEEE 802.11n is the next-generation high-throughput wireless LAN using the multiple-antenna technology. In [19], channel models for single-antenna RX-TX (SISO) as well as multiple-antenna (MIMO) are provided. There are a total of six channel models: A, B, C, D, E and F. In the models, several clusters are given and, for each path, not only the excess delay and the power level, but also angle-of-arrival (AOA)/angle-of-departure (AOD) and angle spreads (AS) at RX/TX are specified. Three channel models from [19] are listed in Tables 4.5, 4.6 and 4.7. They correspond to models with delay spreads of 30, 50 and 100 ns, respectively.

Table 4.1. Default mobile speeds for the channel models

Channel model	Mobile speed
Typical urban	3 km/hr
	50 km/hr
	120 km/hr
Rural area	120 km/hr
	250 km/hr
Hilly terrain	120 km/hr

Petitioner Samsung Ex-1026, 0063

Table 4.2. Channel model for typical urban area [17]

Tap	Delay (μs)	Average relative power (dB)	Doppler spectrum
1	0	−5.7	**Classical**
2	0.217	−7.6	**Classical**
3	0.512	−10.1	**Classical**
4	0.514	−10.2	**Classical**
5	0.517	−10.2	**Classical**
6	0.674	−11.5	**Classical**
7	0.882	−13.4	**Classical**
8	1.230	−16.3	**Classical**
9	1.287	−16.9	**Classical**
10	1.311	−17.1	**Classical**
11	1.349	−17.4	**Classical**
12	1.533	−19.0	**Classical**
13	1.535	−19.0	**Classical**
14	1.622	−19.8	**Classical**
15	1.818	−21.5	**Classical**
16	1.836	−21.6	**Classical**
17	1.884	−22.1	**Classical**
18	1.943	−22.6	**Classical**
19	2.048	−23.5	**Classical**
20	2.140	−24.3	**Classical**

IEEE 802.11n systems mainly operate in indoor environments, where the transmitter and the receiver are usually fixed while people may move in between. Therefore, in [19], a new Doppler spectrum is specified and

$$S_D(f) \propto \frac{1}{1 + A(f/B_{\mathrm{RMS}})^2}, \tag{4.43}$$

where B_{RMS} is the Doppler spread and A is a constant. This Doppler spectrum for an indoor channel is 'bell'-shaped and is quite different from the classical U-shaped spectrum. Typical

Table 4.3. Channel model for rural area [17]

Tap	Delay (μs)	Average relative power (dB)	Doppler spectrum
1	0	−5.2	**Direct path** $f_D = 0.7f_m$
2	0.042	−6.4	**Classical**
3	0.101	−8.4	**Classical**
4	0.129	−9.3	**Classical**
5	0.149	−10.0	**Classical**
6	0.245	−13.1	**Classical**
7	0.312	−15.3	**Classical**
8	0.410	−18.5	**Classical**
9	0.469	−20.4	**Classical**
10	0.528	−22.4	**Classical**

Petitioner Samsung Ex-1026, 0064

Table 4.4. Channel model for hilly terrain area [17]

Tap	Delay (μs)	Average relative power (dB)	Doppler spectrum
1	0	−3.6	**Classical**
2	0.356	−8.9	**Classical**
3	0.441	−10.2	**Classical**
4	0.528	−11.5	**Classical**
5	0.546	−11.8	**Classical**
6	0.609	−12.7	**Classical**
7	0.625	−13.0	**Classical**
8	0.842	−16.2	**Classical**
9	0.916	−17.3	**Classical**
10	0.941	−17.7	**Classical**
11	15.000	−17.6	**Classical**
12	16.172	−22.7	**Classical**
13	16.492	−24.1	**Classical**
14	16.876	−25.8	**Classical**
15	16.882	−25.8	**Classical**
16	16.978	−26.2	**Classical**
17	17.615	−29.0	**Classical**
18	17.827	−29.9	**Classical**
19	17.849	−30.0	**Classical**
20	18.016	−30.7	**Classical**

Table 4.5. Channel model C for IEEE 802.11n [19]. AOA, AOD and AS are all in degrees

		Cluster 1					Cluster 2				
Tap	τ **(ns)**	**Power (dB)**	**AOA**	**AS (RX)**	**AOD**	**AS (TX)**	**Power (dB)**	**AOA**	**AS (RX)**	**AOD**	**AS (TX)**
1	0	0	290.3	24.6	13.5	24.7					
2	10	−2.1	290.3	24.6	13.5	24.7					
3	20	−4.3	290.3	24.6	13.5	24.7					
4	30	−6.5	290.3	24.6	13.5	24.7					
5	40	−8.6	290.3	24.6	13.5	24.7					
6	50	−10.8	290.3	24.6	13.5	24.7					
7	60	−13.0	290.3	24.6	13.5	24.7	−5.0	332.3	22.4	56.4	22.5
8	70	−15.2	290.3	24.6	13.5	24.7	−7.2	332.3	22.4	56.4	22.5
9	80	−17.3	290.3	24.6	13.5	24.7	−9.3	332.3	22.4	56.4	22.5
10	90	−19.5	290.3	24.6	13.5	24.7	−11.5	332.3	22.4	56.4	22.5
11	110						−13.7	332.3	22.4	56.4	22.5
12	140						−15.8	332.3	22.4	56.4	22.5
13	170						−18.0	332.3	22.4	56.4	22.5
14	200						−20.2	332.3	22.4	56.4	22.5

Petitioner Samsung Ex-1026, 0065

Table 4.6. Channel model D for IEEE 802.11n [19]. AOA, AOD and AS are all in degrees

		Cluster 1					Cluster 2					Cluster 3				
Tap	τ **(ns)**	**Power (dB)**	**AOA**	**AS (RX)**	**AOD**	**AS (TX)**	**Power (dB)**	**AOA**	**AS (RX)**	**AOD**	**AS (TX)**	**Power (dB)**	**AOA**	**AS (RX)**	**AOD**	**AS (TX)**
1	0	0	158.9	27.7	332.1	27.4										
2	10	−0.9	158.9	27.7	332.1	27.4										
3	20	−1.7	158.9	27.7	332.1	27.4										
4	30	−2.6	158.9	27.7	332.1	27.4										
5	40	−3.5	158.9	27.7	332.1	27.4										
6	50	−4.3	158.9	27.7	332.1	27.4										
7	60	−5.2	158.9	27.7	332.1	27.4										
8	70	−6.1	158.9	27.7	332.1	27.4										
9	80	−6.9	158.9	27.7	332.1	27.4										
10	90	−7.8	158.9	27.7	332.1	27.4										
11	110	−9.0	158.9	27.7	332.1	27.4	−6.6	320.2	31.4	49.3	32.1					
12	140	−11.1	158.9	27.7	332.1	27.4	−9.5	320.2	31.4	49.3	32.1					
13	170	−13.7	158.9	27.7	332.1	27.4	−12.1	320.2	31.4	49.3	32.1					
14	200	−16.3	158.9	27.7	332.1	27.4	−14.7	320.2	31.4	49.3	32.1					
15	240	−19.3	158.9	27.7	332.1	27.4	−17.4	320.2	31.4	49.3	32.1	−18.8	276.1	37.4	275.9	36.8
16	290	−23.2	158.9	27.7	332.1	27.4	−21.9	320.2	31.4	49.3	32.1	−23.2	276.1	37.4	275.9	36.8
17	340						−25.5	320.2	31.4	49.3	32.1	−25.2	276.1	37.4	275.9	36.8
18	390											−26.7	276.1	37.4	275.9	36.8

Petitioner Samsung Ex-1026, 0066

Table 4.7. Channel model E for IEEE 802.11n [19]. AOA, AOD and AS are all in degrees

		Cluster 1					Cluster 2					Cluster 3					Cluster 4				
Tap	τ (ns)	Power (dB)	AOA	AS (RX)	AOD	AS (TX)	Power (dB)	AOA	AS (RX)	AOD	AS (TX)	Power (dB)	AOA	AS (RX)	AOD	AS (TX)	Power (dB)	AOA	AS (RX)	AOD	AS (TX)
1	0	−2.6	163.7	35.8	105.6	36.1															
2	10	−3.0	163.7	35.8	105.6	36.1															
3	20	−3.5	163.7	35.8	105.6	36.1															
4	30	−3.9	163.7	35.8	105.6	36.1															
5	50	−4.5	163.7	35.8	105.6	36.1	−1.8	251.8	41.6	293.1	42.5										
6	80	−5.6	163.7	35.8	105.6	36.1	−3.2	251.8	41.6	293.1	42.5										
7	110	−6.9	163.7	35.8	105.6	36.1	−4.5	251.8	41.6	293.1	42.5										
8	140	−8.2	163.7	35.8	105.6	36.1	−5.8	251.8	41.6	293.1	42.5										
9	180	−9.8	163.7	35.8	105.6	36.1	−7.1	251.8	41.6	293.1	42.5	−7.9	80.0	37.4	61.9	38.0					
10	230	−11.7	163.7	35.8	105.6	36.1	−9.9	251.8	41.6	293.1	42.5	−9.6	80.0	37.4	61.9	38.0					
11	280	−13.9	163.7	35.8	105.6	36.1	−10.3	251.8	41.6	293.1	42.5	−14.2	80.0	37.4	61.9	38.0					
12	330	−16.1	163.7	35.8	105.6	36.1	−14.3	251.8	41.6	293.1	42.5	−13.8	80.0	37.4	61.9	38.0					
13	380	−18.3	163.7	35.8	105.6	36.1	−14.7	251.8	41.6	293.1	42.5	−18.6	80.0	37.4	61.9	38.0					
14	430	−20.5	163.7	35.8	105.6	36.1	−18.7	251.8	41.6	293.1	42.5	−18.1	80.0	37.4	61.9	38.0					
15	490	−22.9	163.7	35.8	105.6	36.1	−19.9	251.8	41.6	293.1	42.5	−22.8	80.0	37.4	61.9	38.0	−20.6	182.0	40.3	275.7	38.7
16	560						−22.4	251.8	41.6	293.1	42.5						−20.5	182.0	40.3	275.7	38.7
17	640																−20.7	182.0	40.3	275.7	38.7
18	730																−24.6	182.0	40.3	275.7	38.7

Petitioner Samsung Ex-1026, 0067

Doppler spread is given at about 6 Hz for a 5.25-GHz center frequency and 3 Hz for a 2.4-GHz center frequency.

For each tap in the profile of a specific channel model, the extension from SISO channel to MIMO channel is quite elaborate and involves several steps, as follows [19–21]:

- generate independent fading waveform;
- calculate a matrix specifying the correlation properties of all pairs of TX and RX antennas using the AS, AOA and AOD; and
- construct the correlated fading waveforms for all pairs of TX/RX antennas.

Bibliography

[1] T. S. Rappaport, *Wireless Communications*, *Principles and Practices*, 2nd ed. Eaglewood Cliffs, NJ: Prentice Hall, 2002.
[2] S. Sempei, *Applications of Digital Wireless Technologies to Global Wireless Communications.* Eaglewood Cliffs, NJ: Prentice Hall, 1997.
[3] A. Goldsmith, *Wireless Communications.* Cambridge, UK: Cambridge University Press, 2005.
[4] T. Pollet, M. Van Bladel and M. Moeneclaey, 'BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise', *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, Feb./Mar./Apr. 1995, pp. 191–193.
[5] A. Costa and S. Pupolin, 'M-QAM-OFDM system performance in the presence of a nonlinear amplifier and phase noise', *IEEE Trans. Commun.*, vol. 50, no. 3, Mar. 2002, pp. 462–472.
[6] S. Wu and Y. Bar-Ness, 'OFDM systems in the presence of phase noise: Consequences and solutions', *Trans. Commun.*, vol. 52, no. 11, Nov. 2004, pp. 1988–1996.
[7] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications.* Boston, MA: Artech House Publishers, 2000.
[8] B. Razavi, 'Design considerations for direct-conversion receivers', *IEEE Trans. Circuits Syst. II*, vol. 44, no. 6, Jun. 1997, pp. 428–435.
[9] M. Valkama, M. Renfors and V. Koivunen, 'Advanced methods for I/Q imbalance compensation in communication receivers', *IEEE Trans. Signal Processing*, vol. 49, no. 10, Oct. 2001, pp. 2335–2344.
[10] R. Svitek and S. Raman,'DC offsets in direct-conversion receivers: Characterization and implications', *IEEE Microwave*, vol. 6, no. 3, Sep. 2005, pp. 76–81.
[11] A. A. M. Saleh, 'Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers', *IEEE Trans. Commun.*, vol. COM-29, no. 11, Nov. 1981, pp. 1715–1720.
[12] C. W. Farrow, 'A continuously variable digital delay element', in *Proc. of the IEEE International Symposium on Circuits and Systems*, vol. 3, Jun. 1988, pp. 2641–2645.
[13] C. C. Fu, T. P. Wang, K. C. Chang, C. H. Liao and T. D. Chiueh, 'A real-time digital baseband channel emulation system for OFDM Communications', in *Proc. of IEEE Asia Pacific Conference on Circuits and Systems*, Dec. 2006, pp. 984–987.
[14] A. A. M. Saleh and R. A. Valenzuela, 'A statistical model for indoor multipath propagation', *IEEE J. of Select. Areas Commun.*, vol. 5, no. 2, Feb. 1987, pp. 128–137.
[15] P. Dent, G. E. Bottomley and T. Croft, 'Jakes fading model revisited', *IEE Electronics Letters*, vol. 29, no. 13, Jun. 1993, pp. 1162–1163.
[16] ETSI, 'Digital cellular telecommunicaiton systems (phase 2+): Radio transmission and reception', ETSI TS 100 910, V5.12.0, 2001.
[17] 3GPP, 'Deployment aspects', TR 25.943, v2.1.0, Jun. 2001.
[18] ETSI, 'Digital Video Broadcasting (DVB): framing structure, channel coding and modulation for digital terrestrial television', ETSI EN 300 744 V1.5.1, 2004.
[19] V. Erceg, L. Schumacher, P. Kyntsi, et al., 'TGn channel models', IEEE 802.11-03/940r4, May 2004.
[20] L. Schumacher, K. I. Pedersen and P. E. Mogensen, 'From antenna spacings to theoretical capacities: guidelines for simulating MIMO systems', in *Proc. IEEE International Symposium on Personal*, *Indoor and Mobile Radio Communications Conference*, vol. 2, Sep. 2002, pp. 587–592.
[21] L. Schumacher, 'WLAN MIMO Channel Matlab program', available online at www.info.fundp.ac.be/∼lsc/Research/IEEE_80211_HTSG_CMSC/distribution_terms.html.

Petitioner Samsung Ex-1026, 0068

5

Synchronization

Synchronization is a major task in wireless communication receivers. Normally, the phase and frequency of the carrier signal as well as those of the digital clock signal must be retrieved and maintained at all times to ensure reliable receiver operation.

5.1 Introduction

The synchronization issue is inevitable in all signal transmission systems. In digital transmission, though the bit streams are inherently discrete-time signals, all physical media, such as radio channels or transmission lines, are continuous-time in nature. In wireless communications, most physical transmission media are inefficient in transmitting baseband signals. Consequently, the digital baseband transmitted signal has to be converted to a continuous-time waveform and then modulated by a higher-frequency carrier signal. After the modulated signal passes through the physical media, several inverse processing procedures, including sampling/digitization and demodulation, are applied. Figure 5.1 illustrates a simplified diagram of a wireless communication system.

In wireless communication receivers, *coherent demodulation* needs to make use of a *local oscillation* (LO) signal that has exactly the same carrier frequency and phase as the carrier signal contained in the received signal. Furthermore, accurate sampling clock frequency and phase allow the demodulator to recover the transmitted digital data more effectively. Unfortunately, the receiver is unsynchronized with the transmitter most of the time and thus does not have matching timing reference from which the carrier signal and the sampling clock signal can be regenerated. As is often the case, the receiver derives its LO and clock signals from a controlled oscillator. As such, oscillator mismatch causes carrier frequency/phase error and clock frequency/phase error. In reality, the controlled oscillator not only cannot maintain a stable frequency/phase in its output signal, but also suffers from time-varying *phase noise*. Even with perfect oscillator matching, the unknown propagation delay between the transmitter and the receiver introduces additional phase offset in the LO and the clock signals. Besides, the Doppler effect due to relative motion between the transmitter and the receiver imposes additional frequency shift on the received

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0069

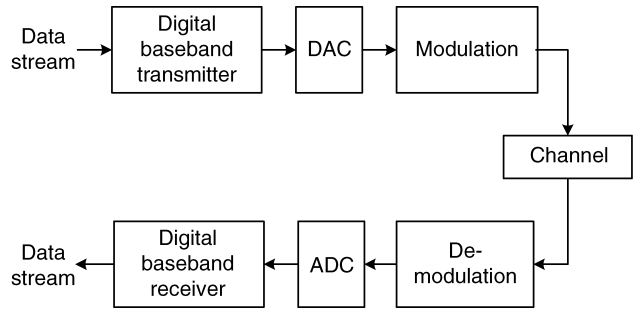


Figure 5.1. Block diagram of a simplified wireless communication system

signal. All these unavoidable impairments undermine the demodulator performance and can even render the demodulator useless if they are not properly tackled. OFDM communication systems, based on the orthogonality among subcarrier signals, are more vulnerable to these synchronization errors than conventional single-carrier communication systems.

This chapter will discuss synchronization errors and explain their respective effects. System performance degradation due to synchronization errors will also be discussed. In the past, numerous algorithms that tried solving the synchronization issues in OFDM systems have been developed. These algorithms will be covered according to the type of synchronization errors that they solve. Finally, several effective schemes for time-domain and frequency-domain synchronization error compensation will also be treated.

5.2 Synchronization Issues

5.2.1 Synchronization Errors

Figure 5.2 illustrates possible synchronization errors in OFDM baseband signals. For clarity, noise, fading and interference are not considered:

- *Carrier frequency offset* (CFO), Δf, causes the received complex baseband signal to rotate at a frequency of Δf.
- *Carrier phase error*, $\phi(t)$, introduces an additional phase rotation term in the received complex baseband signal.
- *Sampling clock offset* (SCO), δ, results in sampling the received continuous-time waveform at an interval of $(1+\delta)T_s$ instead of the ideal T_s.
- *Symbol timing offset*, T_d, refers to the error in the symbol boundary at the receiver from the actual boundary in the received waveform.

5.2.2 Effects of Synchronization Errors

OFDM systems have efficient spectrum utilization due to the overlapping spectra of subcarrier signals. However, this is only true when the orthogonality among subcarriers is

Petitioner Samsung Ex-1026, 0070

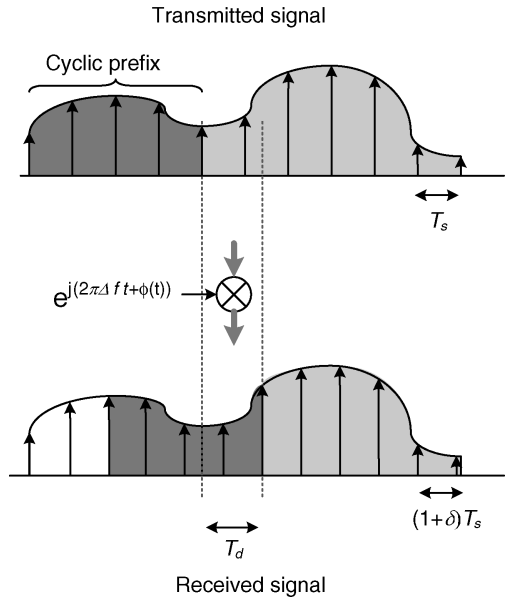


Figure 5.2. Synchronization errors in OFDM baseband signals

preserved, namely, when there is perfect carrier and clock synchronization. If this is not the case, system performance will be degraded due to *inter-symbol interference* (ISI) and *inter-carrier interference* (ICI). In the following, synchronization errors and their effects on the received frequency-domain signals will be discussed in detail.

An OFDM baseband symbol is generated by modulating N complex data using the *inverse discrete Fourier transform* (IDFT) on N subcarriers. The subcarrier spacing is $1/T$, where T is the duration of N time-domain samples. Note that in order to combat ISI, a cyclic prefix of N_g samples is inserted at the beginning of each symbol. Consequently, each OFDM symbol has $N + N_g$ samples. The n-th time-domain sample of the i-th transmitted symbol can be expressed as

$$x(i(N+N_g)T_s + N_gT_s + nT_s) = x_{i,n} = \frac{1}{N}\sum_{k=-N/2+1}^{N/2} X_{i,k}e^{j2\pi nk/N} \qquad (5.1)$$
$$n = -N_g, \cdots, N-1,$$

where $x(t)$ is the transmitted signal, and $X_{i,k}$ is the k-th complex-value frequency-domain signal of the i-th symbol.

Assume that the received signal has been corrupted by a time-varying multipath fading channel and the impulse response at time t is given by

$$h(t,\tau) = \sum_r h_r(t)\cdot\delta(\tau - \tau_r(t)), \qquad (5.2)$$

Petitioner Samsung Ex-1026, 0071

where τ denotes the time delay variable; the gain and delay of the r-th path are denoted by $h_r(t)$ and $\tau_r(t)$, respectively. By convolving the transmitted signal with the channel impulse response and adding the channel noise, $v(\tau)$, the received signal is given by

$$z(\tau) = \sum_r h_r(t) \cdot x(\tau - \tau_r(t)) + v(\tau). \tag{5.3}$$

Further assume that the channel is almost stationary, namely, $h_r(t) = h_r,\ \tau_r(t) = \tau_r$. The frequency-domain channel distortion, H_k, becomes

$$H_k = \sum_r h_r e^{-j2\pi \frac{\tau_r k}{T_s N}}. \tag{5.4}$$

Carrier Frequency Offset (CFO)

Given a carrier frequency offset, Δf, the received continuous-time signal will be rotated by a constant frequency and is in the form of

$$z_{i,n} = z(t) e^{j2\pi\Delta f t}|_{t=i(N+N_g)T_s+N_gT_s+nT_s}. \tag{5.5}$$

The carrier frequency offset can be first normalized with respect to subcarrier spacing $(f_S = 1/(NT_s))$ and then decomposed into the integral component (ϵ_I) and fractional component (ϵ_f): $\Delta f = (\epsilon_I + \epsilon_f) f_S$ and $-0.5 \le \epsilon_f < 0.5$. The received frequency-domain signal then becomes

$$\begin{aligned} Z_{i,k} = {} & X_{i,k-\epsilon_I} H_{k-\epsilon_I} \frac{\sin(\pi\epsilon_f)}{N \sin(\frac{\pi\epsilon_f}{N})} e^{j2\pi \frac{i(N+N_g)+N_g}{N}(\epsilon_I+\epsilon_f)} e^{j\pi\frac{N-1}{N}\epsilon_f} \\ & + \sum_{l=-N/2+1, l\neq k-\epsilon_I}^{N/2} X_{i,l} H_l \frac{\sin(\pi(\epsilon_I+\epsilon_f+l-k))}{N \sin(\frac{\pi(\epsilon_I+\epsilon_f+l-k)}{N})} e^{j2\pi \frac{i(N+N_g)+N_g}{N}(\epsilon_I+\epsilon_f)} e^{j\pi\frac{N-1}{N}(\epsilon_I+\epsilon_f+l-k)} \\ & + V_{i,k}. \end{aligned} \tag{5.6}$$

The second term of the above equation denotes the ICI, namely signals from other subcarriers that interfere with the desired subcarrier signal. Also note that $V_{i,k}$ is the channel noise component in the k-th subcarrier of the i-th symbol. The fractional carrier frequency offset, ϵ_f, results in attenuation in magnitude, phase shift and ICI, while the integer carrier frequency offset, ϵ_I, causes index shift as well as phase shift in the received frequency-domain signals. Note that the phase shift is identical in every subcarrier and is also proportional to the symbol index i.

Carrier Phase Offset

The carrier phase offset contains an initial constant phase offset between the transmitter and the receiver as well as the phase noise from oscillators and other transceiver components.

Petitioner Samsung Ex-1026, 0072

Usually, the phase noise in a receiver can be modelled as a Wiener-Lévy process [1] with a power spectral density in the form of

$$S(f) = \frac{2}{\pi B} \frac{1}{\left[1 + \left(\frac{2f}{B}\right)^2\right]}, \tag{5.7}$$

where B defines the two-sided 3-dB bandwidth. The received frequency-domain signal with a carrier phase offset, $e^{j(\phi_{i,n}+\phi_0)}$, is given by

$$\begin{aligned} Z_{i,k} = {} & X_{i,k} H_k e^{j\phi_0} \left(\frac{1}{N} \sum_{n=0}^{N-1} e^{j\phi_{i,n}} \right) \\ & + \sum_{l=-N/2+1, l\neq k}^{N/2} X_{i,l} H_l e^{j\phi_0} \left(\frac{1}{N} \sum_{n=0}^{N-1} e^{j\phi_{i,n}} e^{j2\pi n(k-l)/N} \right), \\ & + V_{i,k}, \end{aligned} \tag{5.8}$$

where $\phi_{i,n}$ is the phase noise sampled at the n-th sample in the i-th symbol and ϕ_0 is the constant phase offset at the beginning of i-th symbol. The carrier phase offset has two effects: ICI and extra phase shift. The latter is due to the constant phase offset and the averaged phase noise. Note that as the 3-dB bandwidth of the phase noise becomes small compared with the symbol rate, $1/T$, the signal suffers less ICI. In other words, if the phase noise is relatively constant within an OFDM symbol, there is very little ICI.

Sampling Clock Offset (SCO)

If the received continuous-time signal is sampled at an interval of $(1+\delta)T_s$ rather than the ideal T_s, then the n-th received sample of the i-th symbol takes the form of

$$z_{i,n} = z(t)|_{t=i(N+N_g)(1+\delta)T_s + N_g(1+\delta)T_s + n(1+\delta)T_s}, \quad n = -N_g, \cdots, N-1. \tag{5.9}$$

Assume that there is no ISI contamination inside the DFT window; then the k-th frequency-domain received signal of the i-th symbol is given by

$$\begin{aligned} Z_{i,k} = {} & X_{i,k} H_k \frac{\sin(\pi\delta k)}{N \sin(\frac{\pi\delta k}{N})} e^{j2\pi \frac{i(N+N_g)+N_g}{N}\delta k} e^{j\pi \frac{N-1}{N}\delta k} \\ & + \sum_{l=-N/2+1, l\neq k}^{N/2} X_{i,l} H_l \frac{\sin(\pi((1+\delta)l-k))}{N \sin(\frac{\pi((1+\delta)l-k)}{N})} e^{j2\pi \frac{i(N+N_g)+N_g}{N}\delta l} e^{j\pi \frac{N-1}{N}[(1+\delta)l-k]} \\ & + V_{i,k}. \end{aligned} \tag{5.10}$$

From Equation (5.10), it is clear that the sampling clock offset, δ, causes magnitude attenuation and phase shift in the transmitted signal. Note that the phase shift has constant increment proportional to k and δ as the symbol index i increases. Besides, inter-carrier interference (the second term in the above formula) is also introduced.

Petitioner Samsung Ex-1026, 0073

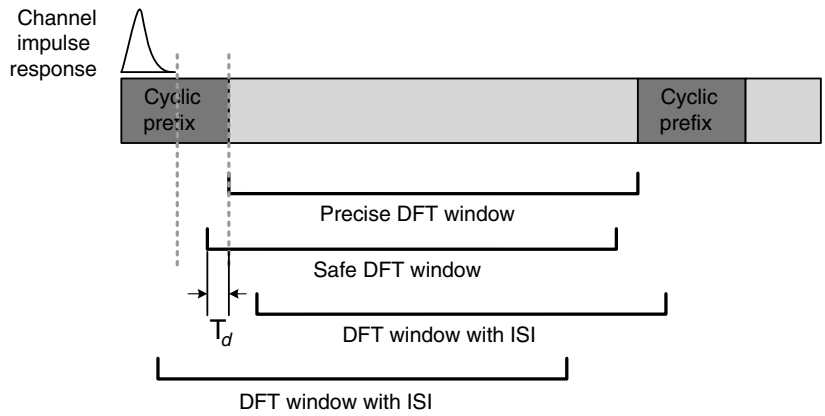


Figure 5.3. Scenarios of DFT window location

Symbol Timing Offset

Assume that the channel maximum excess delay τ_{max} is shorter than the guard interval; the DFT window adopted in the receiver can have several scenarios, as shown in Figure 5.3. When the DFT window is too early (by T_d) and the signal in the window is not contaminated by the previous symbol, the transformed received signal is still free of ISI. In this case, additional phase shift is introduced in the transformed signal and

$$Z_k = X_k H_k e^{-j2\pi T_d k/(NT_s)} + V_k, \tag{5.11}$$

where Z_k, X_k and V_k are the received frequency-domain signal, the transmitted frequency-domain signal and the noise signal of the k-th subcarrier, respectively. Note that without loss of generality, one can consider only one symbol; hence, the subscript i is dropped.

On the other hand, when the DFT window leads or lags by a large degree, ISI will be introduced and both the magnitude and phase of Z_k will be distorted [2]:

$$Z_k = X_k H_k \frac{N - |T_d|/T_s}{N} e^{-j2\pi T_d k/(NT_s)} + \text{ISI} + V_k. \tag{5.12}$$

Note that there exists a slight decrease in the transformed signal magnitude. This is because of the N samples collected for DFT operation; only $N - |T_d|/T_s$ samples are from the desired symbol. The contribution of the other samples is expressed in the ISI term.

5.2.3 Consideration for Estimation and Compensation

In OFDM receivers, signal processing can be applied 'in the time domain' or 'in the frequency domain'. For receiver synchronization, the issue is where to estimate and where to compensate. In the time domain or in the frequency domain? For this question, the receiver designer needs to consider transmission types, system resources, latency, performance and other factors.

The transmission types can be categorized as packet-based or frame-based. In the packet-based systems, such as the IEEE 802.11a/g wireless LAN, user data are organized into packets, and the maximum length of one packet is limited so that channel impairments and

Petitioner Samsung Ex-1026, 0074

synchronization parameters remain almost stationary within one packet. Each packet starts with some preamble signals to facilitate receiver synchronization. What follows is the header that contains important packet information, such as modulation type, code rate, packet length, etc. After the header, the user data then compose the rest of the packet. With such a packet structure in the 802.11a/g wireless LAN, the receiver must be ready for signal detection right after reception of the preamble signals. This makes it necessary for the synchronization block to response promptly, namely to estimate and compensate for synchronization errors in the received signal as quickly as the signal is received. Because the DFT block requires many cycles to compute the frequency-domain signal, most synchronization tasks of the 802.11a/g receiver are usually conducted in the time domain. Moreover, the preamble signals are periodic repetitions and have a good auto-correlation property, making time-domain signal processing preferable.

In the frame-based OFDM systems like digital audio broadcasting (DAB) and digital video broadcasting-terrestrial (DVB-T), data are transmitted continuously. The receiver can spend longer on synchronization before the users notice anything. However, since the receiver may be in operation for quite a while, tracking mechanisms with adaptive compensation for time-varying channel effects and impairments are necessary. Therefore, for the receivers of frame-based OFDM systems, the designers have more latitude in choosing where to estimate the synchronization errors and where to compensate them.

Although the aforementioned synchronization errors mainly incur phase shift in the received frequency-domain signals, ISI and ICI are two sources of interference not to be taken lightly. To mitigate the ISI, it is necessary to adjust the DFT window to within the ISI-free region. Likewise, time-domain CFO/SCO compensation is mandatory when ICI is to be suppressed. Finally, frequency-domain estimation algorithms often have limitation in their capability and one must ensure that the synchronization errors are within the frequency-domain estimators' limits. Usually, this is achieved through time-domain estimation/compensation of coarse synchronization errors.

In OFDM receivers, both the CFO and SCO estimation/compensation blocks are phase-locked loops. Loop delay can play an important role in the convergence performance of a phase-locked loop. In OFDM receivers, frequency-domain synchronization estimation experiences delay of about two symbols–one symbol duration for FFT operation and another one for bit-reverse reordering of the transformed signals [3]. Hence, a synchronization loop that adopts time-domain compensation based on frequency-domain estimation results can expect no less than two symbols of loop delay, making it harder for the loop to converge.

5.3 Detection/Estimation of Synchronization Errors

In this section, various time-domain and frequency-domain algorithms for estimating synchronization errors mentioned in the previous section will be described.

5.3.1 Symbol Timing Detection

As soon as the receiver starts up, it must search for the presence of OFDM symbols in the received signal. Usually, symbol detection exploits some form of repetition in the received OFDM signal. This can be the cyclic prefix inserted in the guard interval or the preambles consisting of identical periods, such as the preambles of IEEE 802.11a/g and the IEEE 802.16e-2005 OFDM mode signals, as shown in Figure 5.4. In the figure, 'CP' refers to

Petitioner Samsung Ex-1026, 0075

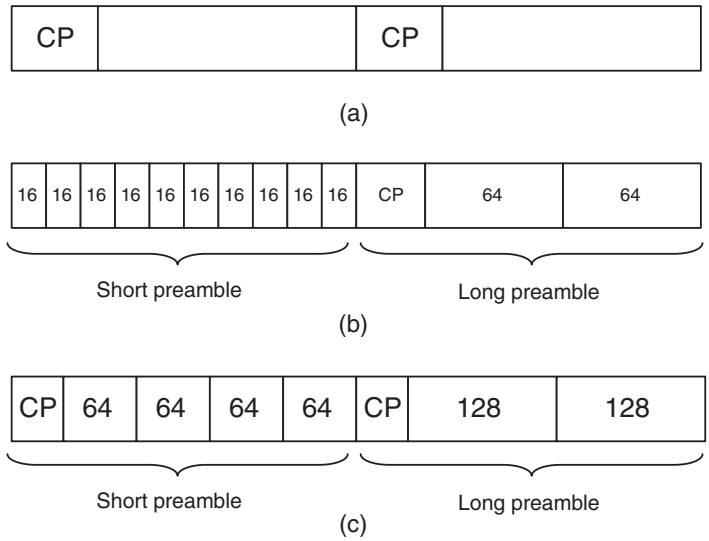


Figure 5.4. Time-domain periodic repetitions in (a) general OFDM symbols, (b) preambles of IEEE 802.11a and (c) preambles of IEEE 802.16e-2005 OFDM mode

the cyclic prefix and the number inside the block indicates the number of samples in one period.

Taking advantage of the repetition in OFDM signals, numerous symbol (timing) detection schemes have been proposed. However, due to CFO, SCO and other channel impairments, these schemes often achieve only coarse symbol timing detection. Additional timing refinement algorithms must be implemented to improve timing accuracy in the receiver and avoid ISI completely. In other words, usually, there are two phases of timing synchronization in OFDM receivers: initial coarse symbol timing detection and ensuing fine symbol timing detection.

Coarse Symbol Timing Detection

Denote the received time-domain baseband samples under consideration, $z_m, m = 0, 1, \cdots$. Let the repetition interval length be denoted as R and the separation between two adjacent intervals be defined as L.

Delay and Correlate

The *delay and correlate* (DC) algorithm for the symbol timing detection algorithm is quite straightforward. It searches for repetition in the received signal using a correlator and a maximum searcher. The time index of the sample with the maximum auto-correlation of the received signal is found by [4]:

$$\begin{aligned} \Phi_{\mathrm{DC}}(m) &= |\sum_{r=0}^{R-1} z_{m-r} z_{m-r-L}^{*}|, \\ \hat{m}_{\mathrm{DC}} &= \arg\max_{m} \Phi_{\mathrm{DC}}(m). \end{aligned} \tag{5.13}$$

Petitioner Samsung Ex-1026, 0076

This DC algorithm is simple in complexity, but there are a couple of drawbacks. First of all, the peak magnitude of the $\Phi_{\mathrm{DC}}(m)$ in different symbols may fluctuate because the power of z_m varies with time. Secondly, when the correlation window moves away from the repeated periods, the magnitude of the correlator output may not fall off as expected, since the correlation between two unrelated segments of noisy OFDM signal is random, especially when the SNR is low. In this case, low correlation length (R) can lead to a large error in the detected symbol boundary.

ML Metric

A *maximum likelihood (ML) timing synchronization algorithm* for OFDM receivers is proposed in [5]. Based on the assumption that the received signals are uncorrelated except for the replicas, this method evaluates

$$\begin{aligned} \Phi_{\mathrm{ML}}(m) &= 2\left|\sum_{r=0}^{R-1} z_{m-r} z_{m-r-L}^{*}\right| - \rho \sum_{r=0}^{R-1} \left(|z_{m-r}|^2 + |z_{m-r-L}|^2\right), \\ \hat{m}_{\mathrm{ML}} &= \arg \max_{m} \Phi_{\mathrm{ML}}(m), \end{aligned} \tag{5.14}$$

where

$$\rho \triangleq \frac{1}{1 + (\sigma_n^2/\sigma_s^2)} \tag{5.15}$$

and σ_s^2/σ_n^2 is the signal-to-noise ratio (SNR). The complexity of the ML algorithm is quite high because of the hardware for estimating ρ. Besides, error in SNR estimation often renders it less reliable than other methods.

MMSE Metric

A *minimum mean squared error* (MMSE) criterion for joint symbol and carrier frequency synchronization that exploits the periodicity of the signal was proposed in [6]. The metric for searching symbol boundary is given by

$$\begin{aligned} \Phi_{\mathrm{MMSE}}(m) &= \sum_{r=0}^{R-1} |z_{m-r}|^2 + \sum_{r=0}^{R-1} |z_{m-r-L}|^2 - 2\left|\sum_{r=0}^{R-1} z_{m-r} z_{m-r-L}^{*}\right|, \\ \hat{m}_{\mathrm{MMSE}} &= \arg \min_{m} \Phi_{\mathrm{MMSE}}(m). \end{aligned} \tag{5.16}$$

Note that the MMSE metric is equivalent to a special case of the ML metric with $\rho = 1$. In the ML and MMSE symbol timing detection algorithms, consideration of the signal power, $|z_{m-r}|^2$ and $|z_{m-r-L}|^2$, indeed improves the timing uncertainty related to the received signal power fluctuation in the delayed correlation algorithm. Moreover, it has been shown that both ML and MMSE algorithms outperform the delay and correlate algorithm in symbol timing detection [7].

Normalized Metrics

Another group of symbol timing detection algorithms adopt a power-normalized metric. The metric was first proposed by Schmidl and Cox [8]. This method uses a preamble that consists

Petitioner Samsung Ex-1026, 0077

of two identical segments, each of which has $N/2$ samples. The symbol timing detection metric takes the form of

$$\Phi_{\mathrm{S}}(m) = \frac{\left|\sum_{r=0}^{N/2-1} z_{m-r} \cdot z_{m-r-N/2}^{*}\right|^2}{\left(\sum_{r=0}^{N/2} |z_{m-r}|^2\right)^2}. \tag{5.17}$$

The maximum of $\Phi_{\mathrm{S}}(m)$ indicates the end of the preamble.

Minn et al. [9] proposed a more general preamble structure consisting of U identical segments with different polarities. The sign of each segment is assigned with a view to achieving the steepest possible roll-off correlation. For $U = 4$, a typical preamble has the following form,

$$[-A \;\; A \;\; -A \;\; -A], \tag{5.18}$$

where A is the preamble segment with $N/4$ samples. The timing detection metric for this preamble is given by

$$\Phi_{\mathrm{M}}(m) = \frac{4}{3}\frac{\left|\sum_{u=0}^{2} s(u)s(u+1)\sum_{r=0}^{N/4-1} z_{m-r-uN/4} \cdot z_{m-r-(u+1)N/4}^{*}\right|^2}{\left(\sum_{u=0}^{2}\sum_{r=0}^{N/4-1} |z_{m-r-uN/4}|^2\right)^2}, \tag{5.19}$$

where $s(u)$ is the sign of the u-th preamble segment.

The main drawback of the delay-and-correlation, ML and MMSE algorithms is that when the preamble has more than two identical segments (e.g. the short preambles in IEEE 802.11a/g and IEEE 802.16e-2005), there will be a plateau or a wide basin in the correlator output waveforms. The same phenomenon can also be observed in the Schmidl's algorithm. The width of the plateau or the basin is approximately (depending on the SNR and the thresholds that decide the two edges) equal to the length of the extra cyclic prefix/segments minus the length of the channel maximum excess delay. Figure 5.5 illustrates this effect for the three algorithms under an AWGN channel with zero channel excess delay. Theoretically, the plateau/basin indicates the ISI-free region for the DFT window. In reality, however, noises contained in the received signal may cause the maximum/minimum to drift away from the optimal point. Consequently, a moving average with a length longer than the plateau/basin helps to filter out the noise component and improve the detection accuracy. Another remedy is to use a longer period in the preamble and thus a longer window of correlation. This effectively enhances the signal-to-noise ratio of the metrics and thus can also make the timing detection more robust.

Figure 5.6 depicts waveforms of the Schmidl's algorithm and the Minn's algorithm [10]. The simulation was conducted under the perfect channel, namely no noise and channel distortion, so as to illustrate the characteristics of the two normalized metrics. The OFDM symbol has 1024 samples plus a guard interval of 128 samples. As is clear from the figure, the Schmidl's algorithm produces a plateau due to the extra cyclic extension of the guard interval, while the Minn's algorithm indeed does away with the plateau and has a sharp roll-off in its metric.

Petitioner Samsung Ex-1026, 0078

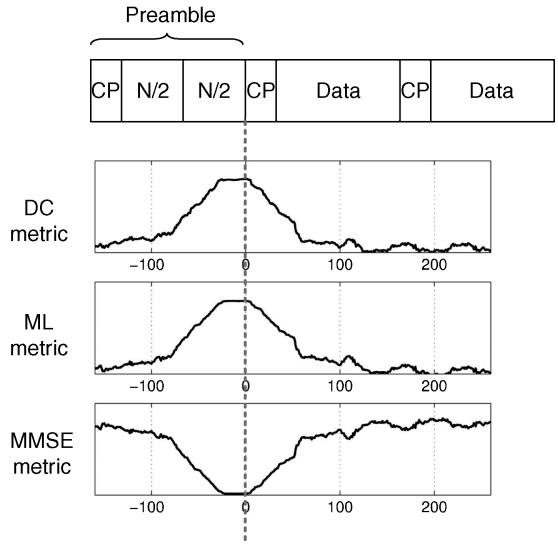


Figure 5.5. Waveforms in three symbol timing detection algorithms

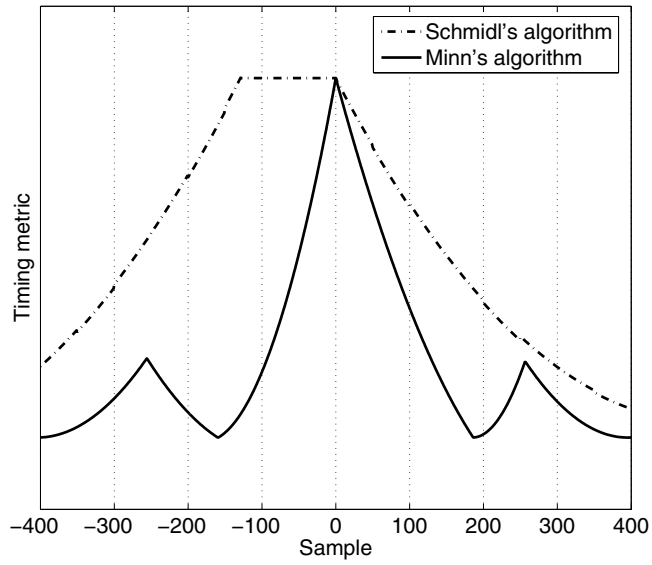


Figure 5.6. Waveforms of the metrics used in Schmidl's and Minn's symbol timing estimators

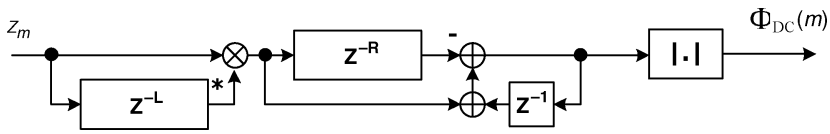


Figure 5.7. Diagram of the circuit that computes the DC metric

It is worth mentioning that the windowed correlation of all the above metrics can be reformulated in a recursive form. For example, the delay and correlation output at sample $m+1$ can be obtained from the result at sample m by

$$\sum_{r=0}^{R-1} z_{m+1-r} z^*_{m+1-r-L} = \left(\sum_{r=0}^{R-1} z_{m-r} z^*_{m-r-L} \right) + \left(z_{m+1} z^*_{m+1-L} - z_{m+1-R} z^*_{m+1-R-L} \right). \tag{5.20}$$

Note that only one complex multiplication, one addition and one subtraction are needed to compute a new output. As an example of this hardware reduction, Figure 5.7 depicts the architecture that computes the DC metric.

Fine Symbol Timing Detection

Usually, coarse symbol timing detection algorithms acquire only rough timing information, and there may still exist a large timing error. A further refinement step that finds more accurate symbol timing is therefore necessary. In the frame-based OFDM systems, averaging over a number of symbols makes a viable solution, since there is large slack in the acquisition time budget in this type of system and, as the averaging interval is extended, the acquired timing information is more robust.

For the packet-based systems, unfortunately, an ISI-free DFT window must be derived as soon as possible in order to proceed with tasks such as channel estimation and packet header detection. If the fine symbol timing cannot be obtained in time, delay lines are needed to buffer the received signal.

The mechanism for fine symbol timing detection mainly relies on matching the time-domain received waveform with the preamble waveform to obtain the channel impulse response (CIR) and then the optimal symbol timing. Another timing detection algorithm obtains the channel impulse response from the estimated channel frequency response through inverse DFT. However, this algorithm has long latency due to the need for the extra inverse DFT operation.

Cross-Correlation

Instead of correlating the noisy received waveform with a delayed version of the noisy received waveform, as in the coarse symbol timing detection algorithm, the receiver can correlate the received noisy signal with the 'clean' preamble waveform using a *matched filter* [11]. The optimal timing can be obtained by first computing the cross-correlation function

$$\Phi_{zp}(m) = \sum_{q=0}^{Q-1} z_{m+q} \cdot p^*_q, \tag{5.21}$$

Petitioner Samsung Ex-1026, 0080

where Q is the length of the preamble and $q = 0, \cdots, Q-1$, denote preamble samples. Then find the location at which its maximum magnitude occurs:

$$\hat{m}_{\mathrm{MAX}} = \arg \max_{m} |\Phi_{zp}(m)|. \tag{5.22}$$

In practice, OFDM systems usually use a PN-code-like preamble in the transmitter; hence, the preamble matched filter output also provides an estimate of the CIR [12], from which the optimal symbol timing can be derived. Let the periodic auto-correlation function of a preamble, p_q, be defined as

$$\Phi_{pp}(n) = \left| \sum_{q=0}^{Q-1} p_q \cdot p^*_{[q-n]_Q} \right|, \tag{5.23}$$

where $[\cdot]_Q$ denotes the modulo-Q operation. As an example, Figure 5.8 shows the periodic auto-correlation function of the long preamble in the IEEE 802.16e-2005 OFDM mode system. Note that with only one-sample shift in the time index between the two waveforms ($n = \pm 1$), the attenuation is obvious, which is similar to that of the impulse function. Therefore, the output waveform of the matched filter for such a preamble represents an estimate of the channel impulse response.

Example:

Figure 5.9 depicts the simulated output magnitude of the matched filter when receiving a preamble signal. The channel used in the simulation has three taps, each two samples apart and 3-dB attenuation in power. Note that the simulated magnitude waveform indeed resembles the three-tap channel profile used in the simulation.

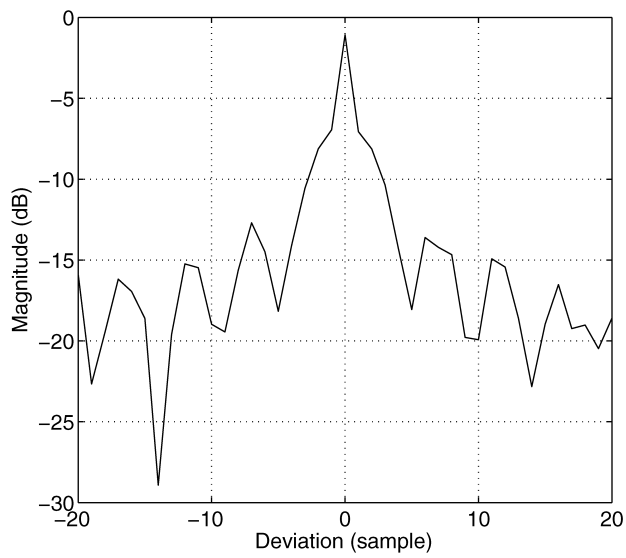


Figure 5.8. Impulse-like auto-correlation function of the long preamble in the IEEE 802.16e-2005 OFDM mode

Petitioner Samsung Ex-1026, 0081

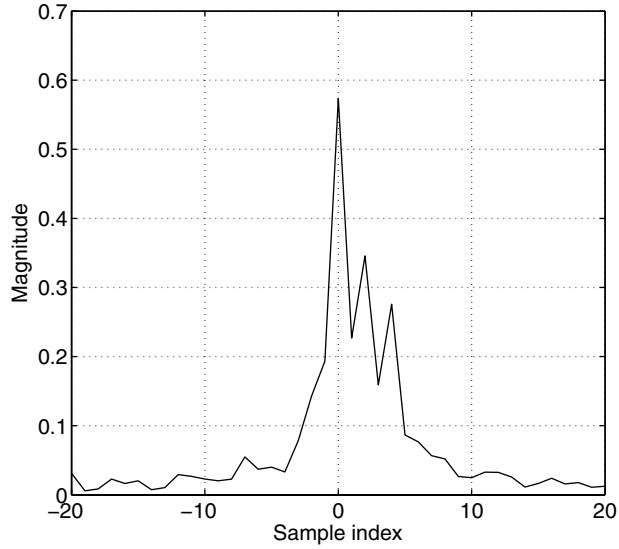


Figure 5.9. Resolved channel profile from a matched filter output

Another algorithm selects the earliest sample with magnitude greater than some threshold, usually some percentage of the peak magnitude [13]. For instance,

$$\hat{m}_{\mathrm{TH}} = \min\{m \big| \ |\Phi_{zp}(m)| > \gamma |\Phi_{zp}(\hat{m}_{\mathrm{MAX}})|\}, \tag{5.24}$$

where $0 < \gamma < 1$ and $|\Phi_{zp}(\hat{m}_{\mathrm{MAX}})|$ is the peak magnitude of Φ_{zp}. The threshold (γ) must be chosen judiciously; it must be small enough to reject the ISI and large enough to avoid false alarm. In practice, using the first multipath component rather than later paths as the start of the DFT window often reduces the probability of ISI contamination.

The architecture of a matched filter, which calculates one cross-correlation output sample every clock cycle, is illustrated in Figure 5.10(a). This design is a transposed finite-impulse-response (FIR) filter structure. A direct FIR filter structure moves all the delay elements to the top branch and has a longer latency. Note that in both architectures, there are as many complex multipliers as there are taps. The complexity can become prohibitive as Q gets large. In order to reduce the complexity, quantizing both the input signals and the coefficients to shorter word-length can help to strike a balance between accuracy and hardware complexity [14]. Another low-complexity implementation is the *correlator bank*, as shown in Figure 5.10(b). Only one complex multiplier and accumulator (MAC) is required for each correlation lag. As a result, the needed hardware complexity is proportional to the search range.

Frequency Response Estimate

Another method of estimating the CIR is from the frequency-domain channel frequency response [15]. If $\mathbf{F}^{-1}$ denotes the $N \times N$ IDFT matrix, then the time-domain channel impulse response $\hat{\mathbf{h}} = [\hat{h}_{-(N/2)+1}\ \hat{h}_{-(N/2)+2} \cdots \hat{h}_{N/2}]^T$ can be obtained by

$$\hat{\mathbf{h}} = \mathbf{F}^{-1} \cdot \mathbf{X}^{-1} \cdot \mathbf{z}, \tag{5.25}$$

Petitioner Samsung Ex-1026, 0082

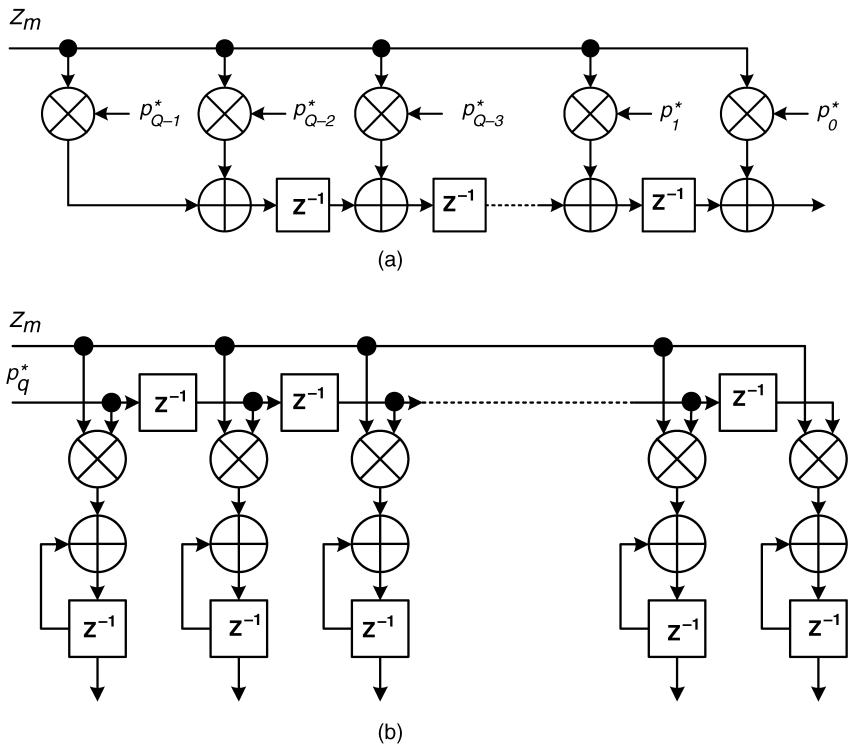


Figure 5.10. (a) Architecture of the transposed matched filter and (b) architecture of the correlator bank

where $\mathbf{z} = [Z_{-(N/2)+1}\, Z_{-(N/2)+2} \cdots Z_{N/2}]^T$ is the received frequency-domain subcarrier data and $\mathbf{X}$ is a diagonal matrix whose i-th diagonal element is the known transmitted signal at the i-th subcarrier. Then, starting from $\hat{h}_0$ onwards, once $\hat{h}_m$ exceeds a threshold, the index m is regarded as the symbol timing. Note that this method introduces extra latency due to the inverse Fourier transformation block.

Frequency-Domain Phase Shift

Symbol timing offset results in phase shift of the frequency-domain signals, as indicated in Equations (5.11) and (5.12). Hence, frequency-domain signal phases can provide information for fine symbol timing detection. Unfortunately, the frequency-domain signal phase is also influenced by the channel phase. To get rid of the channel effect during timing offset estimation, one can compute the phase difference between adjacent subcarriers, given that adjacent subcarriers usually suffer similar channel fading and that the common channel phase is eliminated. The symbol timing offset estimation is then obtained as

$$\hat{m}_{\mathrm{PS}} = \frac{N}{2\pi} \angle \left(\sum_k (Z_{k+1} Z_k^*)(X_{k+1} X_k^*)^* \right), \tag{5.26}$$

where $\angle$ denotes the phase of the complex number; Z_k and X_k are the received frequency-domain signal and the transmitted frequency-domain signal of the k-th subcarrier, respectively.

Petitioner Samsung Ex-1026, 0083

All cross-correlation-based algorithms and frequency-domain based algorithms cannot handle the signal with large residual CFO [13]. Thus, carrier frequency offset must be compensated in advance. However, as will become evident in the following, the fractional carrier frequency offset, ϵ_f, can be jointly estimated when the coarse symbol timing is detected. Hence, it can be compensated prior to the fine symbol timing detection. As for the integer carrier frequency offset (ϵ_I), several sets of matched filter/correlator banks, each using coefficients modulated by different values of integer CFO, can jointly detect the integer CFO and the fine symbol timing. For the frequency-response-estimate and the frequency-domain-phase-shift methods, the integer carrier frequency offset causes a frequency shift of subcarriers and renders them ineffective. Hence, compensation of the integer CFO must be performed before the fine symbol timing operation.

5.3.2 Carrier Frequency Offset Estimation

Usually, the CFO is normalized with respect to the subcarrier spacing $f_S = 1/(NT_s)$, where T_s is the sample period. The integer part of the normalized value is called integer CFO, while the fractional part is called fractional CFO.

Fractional CFO Estimation

Maximum Likelihood Estimation

An estimate of the carrier frequency offset, if within a certain limit, can be obtained simultaneously when the coarse symbol timing is acquired by the algorithms mentioned earlier. The *maximum likelihood CFO estimator* is given by

$$\widehat{\Delta f} = \frac{1}{2\pi L T_s} \angle \left(\sum_{r=0}^{R-1} z_{m-r} z_{m-r-L}^{*} \right). \tag{5.27}$$

Note that the phase can only be resolved in $[-\pi, \pi]$, and the above formula estimates only the part of the CFO that is within $[-1/(2LT_s), 1/(2LT_s)]$ Hz. If $L = N$, then $\widehat{\Delta f} = \hat{\epsilon}_f f_S$, the part of the CFO that is within plus and minus half subcarrier spacing, a.k.a. fractional CFO. In the case in which $L > 1/(\Delta f T_s)$, frequency ambiguity occurs, and the total CFO must be resolved by additional integer CFO estimation.

Example:

In the DVB-T system, there is no particular training symbol for symbol timing detection, and the CP associated with each OFDM symbol is adopted for repetition detection. Therefore, the distance L between two segments is equal to N and the range of estimated CFO is within $[-0.5f_S, 0.5f_S]$. In the IEEE 802.16e-2005 OFDM mode, the short preamble has four repetitions, each having $N/4$ samples, and thus $-2f_S \leq \hat{\Delta f} \leq 2f_S$.

BLUE

If the preamble has U identical repetitions, where $U > 2$, then another *best linear unbiased estimator* (BLUE) exploiting the correlation of the repeated segments is possible [16].

Petitioner Samsung Ex-1026, 0084

Assume that there are R samples in a segment, so, in total, $UR = N$ samples are available. The BLUE estimation algorithm starts with computing several linear auto-correlation functions with uR samples of delay:

$$\Phi_{\mathrm{BLUE}}(u) = \frac{1}{N - uR} \sum_{m=uR}^{N-1} z_m z_{m-uR}^*, \quad 0 \le u \le K. \tag{5.28}$$

Then, the phase differences between all pairs of two auto-correlation functions with delay difference of R are computed

$$\varphi(u) = [\angle\{\Phi_{\mathrm{BLUE}}(u)\} - \angle\{\Phi_{\mathrm{BLUE}}(u-1)\}]_{2\pi} \quad 1 \le u \le K, \tag{5.29}$$

where $[\cdot]_{2\pi}$ denotes a modulo-2π operation and K is a design parameter less than U. Note that each $\varphi(u)$ represents an estimate of the CFO, scaled by a constant. The smaller the constant u, the better accuracy it achieves. To gain an effective CFO estimate, the BLUE estimator uses a weighted average of all $\varphi(u)$ and computes

$$\widehat{\Delta f}/f_S = \frac{U}{2\pi} \sum_{u=1}^{K} w_u \varphi(u), \tag{5.30}$$

where

$$w_u = 3\frac{(U-u)(U-u+1) - K(U-K)}{K(4K^2 - 6UK + 3U^2 - 1)}, \tag{5.31}$$

The optimal value for K achieving the minimal variance of $\widehat{\Delta f}$ is $U/2$. The range of estimated carrier frequency offset is $-Uf_S/2 \le \widehat{\Delta f} \le Uf_S/2$.

With some modification, this estimator can also be applied to preambles consisting of several repeated segments with specific sign changes. With proper acquired symbol timing, the received U segments of the preamble are multiplied by their respective signs, and then the same method as the BLUE can be applied [9].

Integer CFO Estimation

In the IEEE 802.16e-2005 OFDM mode standard, the oscillator deviation is within ± 8 ppm. With the highest possible carrier frequency of 10.68 GHz, the maximum carrier frequency offset is about ± 171 KHz when the transmitter LO and the receiver LO both have largest yet opposite-sign frequency deviations, which is also equivalent to ± 11 subcarrier spacing (f_S). In the 6-MHz DVB-T system, assuming that the oscillator deviation is within ± 20 ppm and the carrier frequency is around 800 MHz, the maximum carrier frequency offset can be up to ± 38 subcarrier spacing (f_S) in the 8K transmission mode. From the previous discussion, it is clear that the estimated carrier frequency offset obtained simultaneously in the coarse symbol boundary detection has ambiguity in frequency. In the following, the algorithms for resolving such frequency ambiguity in the estimated carrier frequency offset will be presented.

Petitioner Samsung Ex-1026, 0085

Time-Domain Correlation

In the 802.16e-2005 OFDM mode, the initial estimated carrier frequency offset is within $[-2f_S, 2f_S]$. Besides this estimation, additional frequency offset of $\pm 12f_s$, $\pm 8f_s$ or $\pm 4f_s$ is possible given a CFO range of $\pm 11 f_S$. In order to estimate this additional integer carrier frequency offset, a matched filter matching the fractional-CFO-compensated received signal against the modulated long preamble waveforms can be used [17]. The coefficients of the matched filter are the complex conjugate of the long preamble and they are modulated by a sinusoidal wave whose frequency is a possible integer CFO mentioned above. The output of the matched filter will have a maximum peak value if its coefficients are modulated by the carrier with the correct integer CFO. It is possible to deploy one such matched filter for each possible integer CFO. In this case, seven matched filters are needed. In [17], the authors proposed to use only one set of matched filter hardware that handles different integer CFOs sequentially. In addition, as suggested previously in the symbol timing detection subsection, the coefficients of the matched filter can be quantized to -1, 0, 1 to reduce hardware complexity.

Example:

Assume that in an 802.16e-2005 OFDM mode system, the received signal suffers a carrier frequency offset of $4.2\, f_S$. As soon as the coarse symbol timing is detected, the fractional carrier frequency offset of about $0.2\, f_S$ can be estimated and compensated. Then, the compensated signal is passed to the matched filter with several sets of integer-CFO-modulated long preamble as coefficients. Figure 5.11 depicts the matched filter output waveforms for three cases of different integer CFO. It is clear that only the output waveform with coefficients modulated by the carrier frequency offset of $+4\, f_S$ has a distinctive peak, indicating that the integer CFO is $+4\, f_S$.

Frequency-Domain Auto-Correlation

Besides the time-domain matched filter with modulated preamble method, there exist a few integer CFO detection algorithms that resort to frequency-domain correlation. Let $Z_{i,k}$ be the received frequency-domain signal at the k-th subcarrier of the i-th symbol.

In the DVB-T signals, there are no special training symbols or preambles; hence, continual pilot subcarriers are utilized for integer CFO synchronization. As the integer carrier frequency offset causes frequency shift of the received frequency-domain signals, the following correlation function between the signal values on a set of subcarriers of two consecutive symbols can be computed to detect the continual pilot subcarriers [18–20]:

$$\Phi_{\mathrm{FDAC}}(g) = \sum_{j=0}^{J-1} Z_{i,\alpha_j+g} Z^*_{i-1,\alpha_j+g}, \quad g = 0, \pm 1, \pm 2, \cdots, \tag{5.32}$$

where the indices of J pilot subcarriers are denoted by α_j, $j = 0, 1, \cdots, J-1$. Note that if the k-th subcarrier carries random data, then the product inside the summation of the above formula is zero mean. On the other hand, if the k-th subcarrier carries a time-invariant continual pilot value and if the channel is quasi-stationary, the product will have an average

Petitioner Samsung Ex-1026, 0086

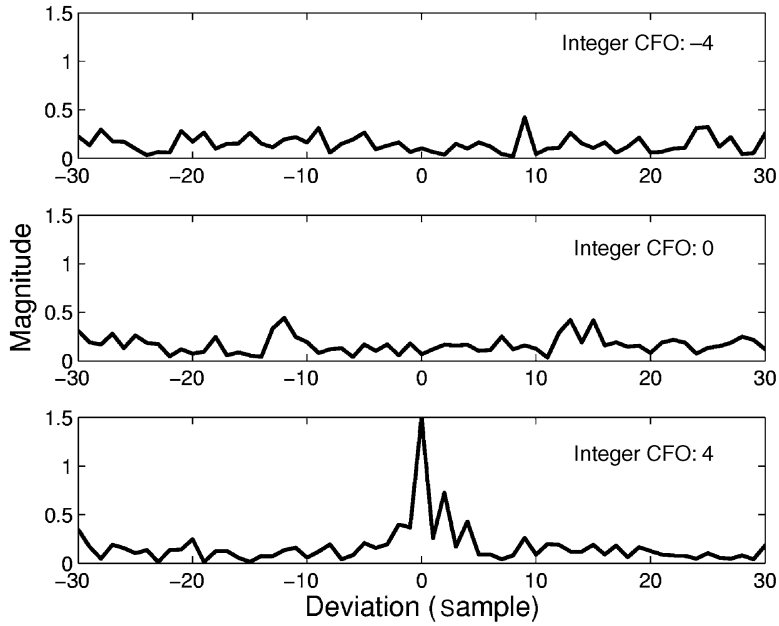


Figure 5.11. Matched filter output waveforms for three different values of integer CFO when the received signal suffers an integer carrier frequency offset of $+4f_S$

magnitude equivalent to that of the squared pilot signal. The integer CFO can be found by locating the subcarrier index offset g with the largest $|\Phi_{\text{FDAC}}(g)|$:

$$\hat{g} = \arg \max_{g} |\Phi_{\text{FDAC}}(g)|. \tag{5.33}$$

The above scheme works quite well because the channel phase can be cancelled when the channel is quasi-stationary. However, under fast-fading channels, the performance of this algorithm will be much degraded. Furthermore, the magnitude of the $\Phi_{\text{FDAC}}(g)$ becomes attenuated if many of the subcarriers being considered are deeply faded. Hence, enough pilot subcarriers are necessary to guarantee low probability of simultaneous fading on pilot subcarriers. To implement this method, one complex multiplier and several accumulators are needed. The computation complexity of the frequency-domain auto-correlation method is thus lower than that of the aforementioned time-domain matched filter approach.

Frequency-Domain Cross-Correlation

In the OFDM systems, usually the frequency-domain pilot power is made larger than that of the data subcarriers. The receiver can utilize such a property and find the set of subcarriers with maximum overall power:

$$\hat{g} = \arg \max_{g} \sum_{j=0}^{J-1} |Z_{i,\alpha_j+g}|^2. \tag{5.34}$$

Petitioner Samsung Ex-1026, 0087

Although this algorithm uses only one symbol and thus can resist the impact of fast-fading channels, frequency-domain channel response can influence its performance. Therefore, its validity also depends on a large number of pilot subcarriers.

Frequency-Domain PN-Correlator

A low-complexity integer CFO detection algorithm adopting a differential PN-code encoding on the pilot subcarriers is proposed in [21]. Let the length of the PN code be M and denote it as $c_m, m = 0, 1, \cdots, M-1$. Without loss of generality, let $X_{\alpha_0} = 1$ and set

$$X_{\alpha_{j+1}} = c_{[j]_M} X_{\alpha_j}, \quad j = 0, 1, \cdots, J-2, \tag{5.35}$$

where $[\cdot]_M$ denotes modulo M. The algorithm calculates

$$\Phi_{\mathrm{PN}}(g) = \left| \sum_{j=0}^{J-2} Z_{i,\alpha_j+g} Z^*_{i,\alpha_{j+1}+g} c_{[j]_M} \right|, \tag{5.36}$$

and the integer carrier frequency offset is given by

$$\hat{g} = \arg \max_g \ \Phi_{\mathrm{PN}}(g). \tag{5.37}$$

In contrast to the previous frequency-domain auto-correlation (FDAC) algorithm, this approach can resist fast-fading channels to a much larger degree, since it considers only one symbol. However, coherent fading between adjacent pilot subcarriers is implied in the algorithm. If this is not true, then the effectiveness of the algorithm can be compromised. As in the algorithms mentioned above, more pilot subcarriers make the probability of accurate detection higher.

5.3.3 Residual CFO and SCO Estimation

Although the CFO in the received signal has been estimated and compensated in the receiver, still some residual CFO may exist. Besides, the CFO contained in the received signal may very well be time-varying and, thus, it needs to be continuously tracked. The received signal also suffers from sampling clock offset (SCO), which may cause a gradual drift of the safe DFT window in addition to extra phase shift in the received frequency-domain signals. In the frame-based OFDM systems, both the residual CFO tracking and the SCO tracking are inevitable, since the receiver may operate for a long while. In the packet-based OFDM systems, however, the influences of these two offsets depend on the packet length and the magnitude of the offsets.

The sampling clock offset may not be easily estimated from the time-domain signal. However, it can be examined through the phase shift of frequency-domain pilot signals. The residual CFO can also be estimated in a similar way. In many OFDM wireless communication standards, such as DVB-T, IEEE 802.11a/g and IEEE 802.16e-2005 OFDM mode, dedicated pilot subcarriers are allocated to facilitate receiver synchronization.

As Equation (5.6) indicates, the phase shifts in the received frequency-domain data caused by the CFO are identical at all subcarriers provided that the ICI is ignored. On the other hand, the SCO causes phase shifts that are proportional to respective subcarrier indices. Figure 5.12 shows the simulated effect on the phase of the frequency-domain data of two adjacent OFDM

Petitioner Samsung Ex-1026, 0088

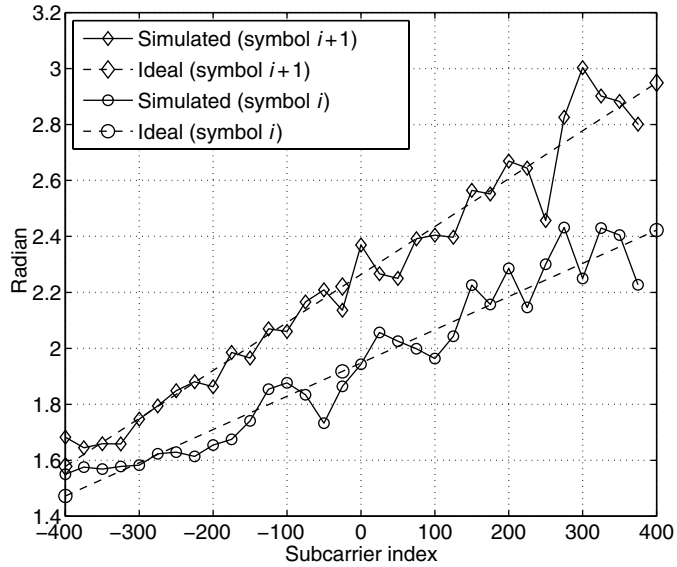


Figure 5.12. Phases of subcarrier data in two OFDM symbols with carrier frequency offset equal to 0.05 subcarrier spacing, $\delta = -100$ ppm and Gaussian noise

symbols when the symbols are distorted by noise, CFO and SCO. The CFO is 0.05 f_S and the SCO ratio, δ, is -100 ppm. The received data contain ICI and noise, and therefore the phases deviate from the two ideal straight lines. Conventionally, the SCO can be estimated by computing a slope from the plot of measured pilot subcarrier phase differences versus pilot subcarrier indices [22]. Recently, joint estimation of CFO and SCO has also been studied extensively. In the following, two joint estimation algorithms will be discussed.

Joint LLS Estimator

In [23], the *linear least squares* (LLS) algorithm is applied to the plot of pilot subcarrier phase differences versus pilot subcarrier indices in order to obtain the intercept and the slope of the best-fit line, which are related to the CFO and SCO, respectively. Assume that the pilot subcarriers are symmetrical to the DC subcarrier. Define the phase difference between two OFDM symbols at the pilot subcarrier α_j as θ_j:

$$\theta_j = \angle(Z_{i,\alpha_j} Z^*_{i-1,\alpha_j}),\ j = 0, 1, \cdots, J-1. \tag{5.38}$$

The LLS algorithm for the residual carrier frequency offset and sampling clock offset is given by

$$\hat{\epsilon}_f = \frac{\sum_{j=0}^{J-1} \theta_j}{\left(2\pi \frac{N+N_g}{N}\right) J} \tag{5.39}$$

Petitioner Samsung Ex-1026, 0089

and

$$\hat{\delta} = \frac{\left(\sum_{j=0}^{J-1} \theta_j \alpha_j\right)}{\left(2\pi \frac{N+N_g}{N}\right)\left(\sum_{j=0}^{J-1} \alpha_j^2\right)}. \tag{5.40}$$

Such an estimation algorithm that is based on the phase difference across two OFDM symbols can remove the common channel fading terms in slow-fade scenarios. Consequently, estimation schemes such as this can be applied before channel estimation and equalization.

Joint WLS Estimator

Though the joint LLS estimation algorithm provides accurate estimation results in the AWGN channel, diverse channel responses on the pilot subcarriers can render its estimation useless. For instance, phases of several deeply faded pilot subcarriers, when entered into the estimation of the joint LLS formulas, can contribute a large error in the estimation results. On the other hand, the phases of those subcarriers with little fading are naturally more reliable. Consequently, weighting the subcarrier data is advantageous, and data of deeply-faded subcarriers should be assigned smaller weights to minimize their adverse effect on estimation accuracy.

The *weighted least squares* (WLS) algorithm for joint estimation of CFO and SCO [24] estimates the two offsets as

$$\hat{\epsilon}_f = \frac{\left(\sum_{j=0}^{J-1} w_j \alpha_j^2\right)\left(\sum_{j=0}^{J-1} w_j \theta_j\right) - \left(\sum_{j=0}^{J-1} w_j \alpha_j\right)\left(\sum_{j=0}^{J-1} w_j \theta_j \alpha_j\right)}{\left(2\pi \frac{N+N_g}{N}\right)\left[\left(\sum_{j=0}^{J-1} w_j\right)\left(\sum_{j=0}^{J-1} w_j \alpha_j^2\right) - \left(\sum_{j=0}^{J-1} w_j \alpha_j\right)^2\right]} \tag{5.41}$$

and

$$\hat{\delta} = \frac{\left(\sum_{j=0}^{J-1} w_j\right)\left(\sum_{j=0}^{J-1} w_j \theta_j \alpha_j\right) - \left(\sum_{j=0}^{J-1} w_j \alpha_j\right)\left(\sum_{j=0}^{J-1} w_j \theta_j\right)}{\left(2\pi \frac{N+N_g}{N}\right)\left[\left(\sum_{j=0}^{J-1} w_j\right)\left(\sum_{j=0}^{J-1} w_j \alpha_j^2\right) - \left(\sum_{j=0}^{J-1} w_j \alpha_j\right)^2\right]}. \tag{5.42}$$

The weight w_j should be inversely proportional to the variance of phase error, which depends on noise, ICI and the complex channel gain. Usually, the residual synchronization error is so small that the ICI term can be neglected and w_j depends only on the channel gain of the pilot subcarrier:

$$w_j \propto |\hat{H}_{\alpha_j}|^2. \tag{5.43}$$

Petitioner Samsung Ex-1026, 0090

In many wireless communication receivers, the CFO and SCO are derived from the same oscillator source, such as receivers for the IEEE 802.11a/g WLAN standard. The joint estimation algorithm can then be reduced to estimate one oscillator offset, δ. The normalized CFO is $\epsilon = \delta \times f_c T$, where f_c is the carrier frequency and $T = 1/f_S$ is the symbol time. Furthermore, if the pilot subcarriers are uniformly distributed with a spacing of D subcarriers, the phase difference θ_j takes the form of

$$\begin{aligned}\theta_j &= 2\pi \frac{N+N_g}{N}(\alpha_j \delta + \epsilon) + e_j \\ &= 2\pi \frac{N+N_g}{N}(jD + f_c T)\delta + e_j,\end{aligned} \tag{5.45}$$

where e_j comes from ICI and AWGN. Then, the WLS estimate for δ is reduced to

$$\hat{\delta} = \frac{\sum_j w_j \theta_j (jD + f_c T)}{2\pi \frac{N+N_g}{N} \sum_j w_j (jD + f_c T)^2}. \tag{5.45}$$

5.3.4 Carrier Phase Estimation

The carrier phase offset contains the constant phase difference between the received signal and the receiver oscillator as well as the random phase noise caused by transceiver RF/analog electronics. The constant phase difference is harmless and can be absorbed in the channel response, which will be dealt with in the equalizer. Unfortunately, the phase noise can result in unpredictable phase shift in the received frequency-domain signals. Pilot subcarriers usually help to estimate the variation of the carrier phase in the frequency-domain signals:

$$\hat{\phi}_i = \angle\left(\sum_j \frac{Z_{i,\alpha_j}}{\hat{H}_{\alpha_j} X_{\alpha_j}}\right), \tag{5.46}$$

where $\hat{\phi}_i$ is the estimated *common phase error* (CPE) of the i-th symbol, and $\hat{H}_{\alpha_j}$ and X_{α_j} are the channel estimate and the transmitted data at the subcarrier α_j, respectively.

The carrier phase estimation is important in the packet-based OFDM systems because in these systems, the channel estimates are acquired in the beginning of a packet. Hence, by examining the phase of the received pilot subcarrier signals, the phase-noise-induced common phase error term can be obtained and then properly compensated. Carrier phase estimation is also essential in the frame-based OFDM systems when the channel response has to be interpolated along the time direction, such as channel estimates for data subcarriers from those of the scattered pilots in the DVB-T system. Removing the common phase error term diminishes noises contained in the estimated channel gains and, thus, more accurate channel interpolation results can be achieved.

5.4 Recovery of Synchronization Errors

As the synchronization errors are estimated, the receiver will process the received signal to remove or mitigate the effects from these errors. Clearly, orthogonality among subcarriers

Petitioner Samsung Ex-1026, 0091

must be restored first, otherwise the ICI will be incurred and the receiver performance will be deteriorated. Recently, the trend in communication transceiver implementation is to adopt fully digital baseband processors. This means that free-running oscillators are used to generate the carrier frequency as well as the sampling clock that drives the analog-to-digital converters (ADC) and the baseband processor. The CFO/SCO, caused by oscillator mismatch and the Doppler effect, is recovered by digital signal processing techniques instead of adjusting the local oscillation (LO) frequency in the demodulator and the clock generator. The following discussions will focus on the digital solutions of compensating the OFDM synchronization errors, both in the time domain and in the frequency domain.

5.4.1 Carrier Frequency Offset Compensation

Equation (5.6) indicates that ICI arises due to fractional CFO, ϵ_f. In [25], analytic analysis showed that in the AWGN channel and when the number of subcarriers is large, the SNR degradation due to the fractional CFO, D_{SNR}, is given by

$$D_{\text{SNR}} \approx \frac{10}{3\ln 10}(\pi\epsilon_f)^2 \frac{E_s}{N_0} \quad \text{dB}. \tag{5.47}$$

In order to suppress the ICI and thereby reduce SNR degradation, the residual carrier frequency offset must be sufficiently small. For example, when using the 64-QAM constellation, it is better to keep the residual CFO below 0.01 f_S to ensure that $D_{\text{SNR}} < 0.3$ dB for moderate signal-to-noise ratio (E_s/N_0). On the other hand, when QPSK is used, the residual CFO can be up to 0.03 f_S.

Time-Domain Derotator

To compensate for the CFO and limit the residual CFO, a time-domain derotator is commonly used. The derotator is simply a complex multiplier, which rotates the complex-valued input by a phase. The phase for de-rotation is controlled by a *numerical-controlled oscillator* (NCO) and it is fed to the multiplier in the form of cosine/sine values of the phase. Ideally, the NCO should be running at a frequency that is the negative of the CFO contained in the received signal so as to remove the CFO completely. In practice, this is never true, since the CFO is not a constant and is buried in the signal with noises/interferences, so the true CFO is not easy to come by all the time. Usually, a phase-locked loop is adopted in the receiver for estimating and compensating the CFO. Through the feedback loop, the residual error can be maintained within a certain level and the receiver remains synchronized with the carrier.

Figure 5.13 depicts one baseband receiver with such a configuration. The frequency-domain CFO estimator generates CFO estimates continuously, which may be contaminated by noises and interferences. Then, a loop filter is used to filter out unwanted components. Then, the filtered signal for frequency control is passed into the NCO, which outputs the digital sinusoidal waveform to the complex multiplier. Note that the DFT and the bit-reversal blocks induce long latency in the loop. In [26], the effect of extra loop delay is analyzed. The analysis pointed out that the stability region of feasible filter coefficients becomes smaller and that the loop bandwidth may be decreased. This is crucial in the acquisition mode when a fast response is required. However, when it is applied in the tracking mode, the

Petitioner Samsung Ex-1026, 0092

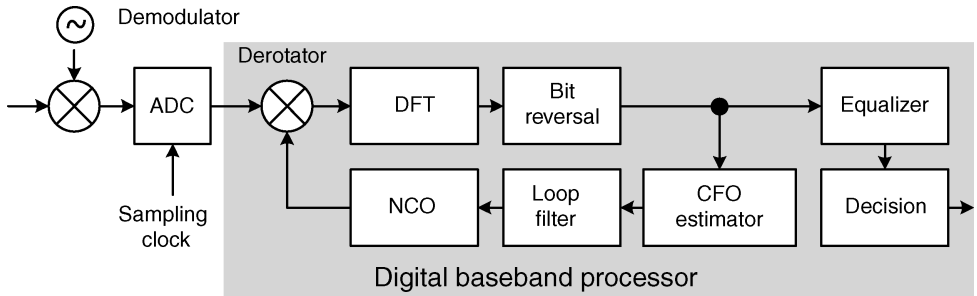


Figure 5.13. OFDM baseband receiver architecture with a CFO synchronization phase-locked loop that uses time-domain compensation

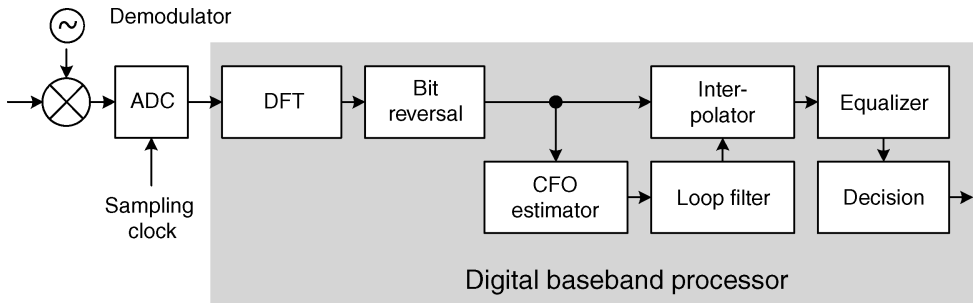


Figure 5.14. OFDM baseband receiver architecture using a frequency-domain interpolator to compensate the carrier frequency offset

requirement of smaller jitter agrees with the narrower loop bandwidth. The above demonstrates a good example, showing that, in most cases, trade-offs must be considered in designing the architecture of an OFDM baseband receiver.

Frequency-Domain Interpolator

To avoid the long delay in the CFO frequency-domain estimation/time-domain compensation loop, the receiver can compensate the CFO in the frequency domain. In this case, a phase rotator for compensating the frequency-domain receiver signals is inadequate, since severe ICI may have been induced. As such, it is necessary to adopt a frequency-domain interpolator [27], which interpolates among the received frequency-domain signals to get signals at the exact frequencies and thereby mitigates the ICI. Such a receiver structure is illustrated in Figure 5.14. The ICI terms are proportional to the coefficients of the *sinc* function if rectangular window is applied to the time-domain-received signal before the DFT operation. Consequently, near-by subcarriers affect the desired subcarrier to a greater extent. The farther away a subcarrier is from the target subcarriers, the smaller the corresponding induced interference. Interpolators with more taps can help to eliminate the ICI more thoroughly and reconstruct the subcarrier in question. Note, however, that an interpolator with many taps requires much higher complexity than the time-domain de-rotator.

Petitioner Samsung Ex-1026, 0093

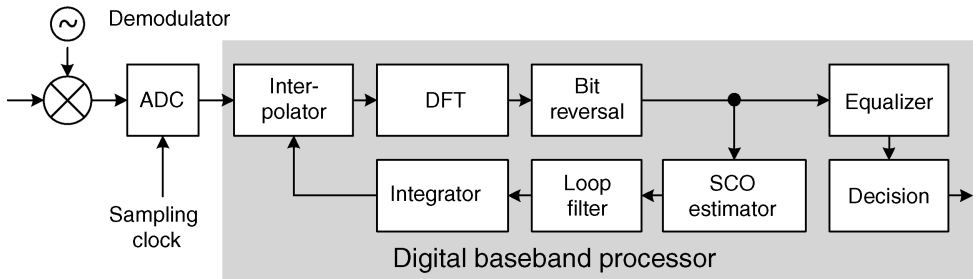


Figure 5.15. OFDM baseband receiver architecture with time-domain sampling clock offset compensation using an interpolator

5.4.2 Sampling Clock Offset Compensation

ICI is also incurred by the sampling clock offset, as shown in Equation (5.10). To be precise, signal attenuation and the magnitude of ICI are proportional to δ and the subcarrier index k. In [28], the SNR degradation in the k-th subcarrier, $D_{\mathrm{SNR},k}$, due to the SCO is analyzed and is given by

$$D_{\mathrm{SNR},k} \approx 10\log_{10}\left(1+\frac{1}{3}(\pi\delta \mathrm{k})^2\frac{\mathrm{E_s}}{\mathrm{N_0}}\right) \quad \mathrm{dB}. \tag{5.48}$$

The formula is derived under the assumptions that the number of subcarriers is large; the ICI is the dominant factor; and the channel is AWGN. Evidently, substantial degradation in SNR occurs at the subcarrier with larger index k.

Time-Domain Interpolator

In the OFDM communication systems with a large number of subcarriers such as the DVB-T system with 8K subcarriers, time-domain SCO compensation should be adopted. Otherwise, signal distortion in the high-frequency subcarriers will become unbearable. To this end, an interpolator capable of calculating samples with fractional delay is a possible solution. The block diagram of such an OFDM baseband receiver is shown in Figure 5.15.

The ideal interpolator to implement fractional delay uses the *sinc* function as its coefficients:

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{5.49}$$

Unfortunately, the *sinc* function is infinitely long and noncausal, making it infeasible in real-time implementation. Nevertheless, it elicits other interpolator solutions to the fractional delay problem. Among the various types of interpolators, the polynomial-based interpolators are more advantageous [29]. The polynomial-based interpolators usually have good frequency characteristics. In addition, their coefficients are easily computed online, making them amenable to real-time implementation.

Petitioner Samsung Ex-1026, 0094

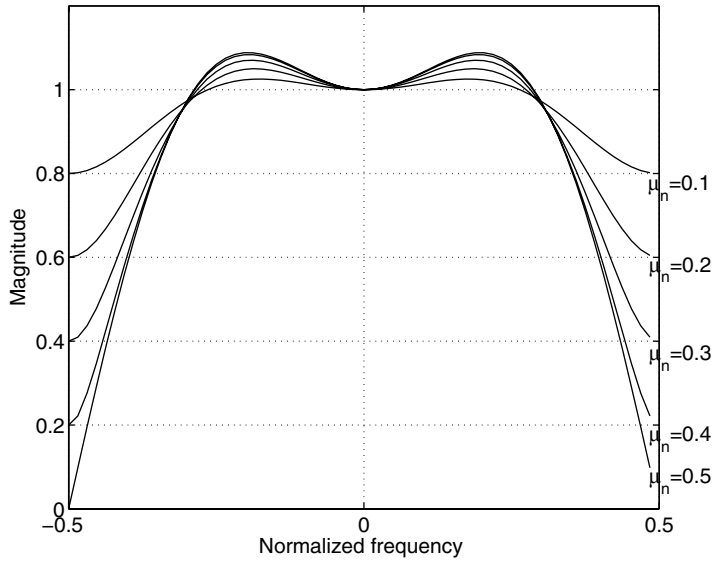


Figure 5.16. Frequency-domain magnitude response of the piecewise-parabolic interpolator with different fractional delays (μ_n)

Among different polynomial orders, the second-order piecewise-parabolic interpolator has adequate complexity as well as moderate magnitude distortion when compared to the first-order linear interpolator and the third-order cubic interpolator. Assume that a sampling clock offset ratio δ is detected, and that the interpolator needs to re-sample the signal (z_m) with a sample interval of 1-δ. The interpolator output (re-sampled signal), q_n, is formulated as

$$\begin{aligned} q_n = {} & z_{m_n+2} \cdot (-0.5\mu_n + 0.5\mu_n^2) + z_{m_n+1} \cdot (1.5\mu_n - 0.5\mu_n^2) \\ & + z_{m_n} \cdot (1 - 0.5\mu_n - 0.5\mu_n^2) + z_{m_n-1} \cdot (-0.5\mu_n + 0.5\mu_n^2), \end{aligned} \tag{5.50}$$

where $m_n = \lfloor n(1-\delta) \rfloor$ and $\mu_n = n(1-\delta) - m_n$, $0 \le \mu_n < 1$.

The magnitude responses of the above interpolator for different fractional delays are plotted in Figure 5.16. The piecewise-parabolic interpolator introduces magnitude shaping in the frequency domain, which must be handled by the equalizer. Also note that the magnitude distortion depends on the fractional delay, μ_n. Therefore, the equalizer coefficients need to be adjusted according to the time-varying interpolator magnitude frequency response, even though the channel is stationary. Finally, the subcarriers in the high-frequency band are subject to severe attenuation and, hence, usually, the signal needs to be over-sampled.

All-Pass Fractional-Delay Filter

A finite-order all-pass digital filter with a near-linear phase response that corresponds to a fractional-sample delay can be the ideal solution to SCO compensation. In general, the infinite impulse response (IIR) filter structure can achieve unit magnitude response (all-pass

Petitioner Samsung Ex-1026, 0095

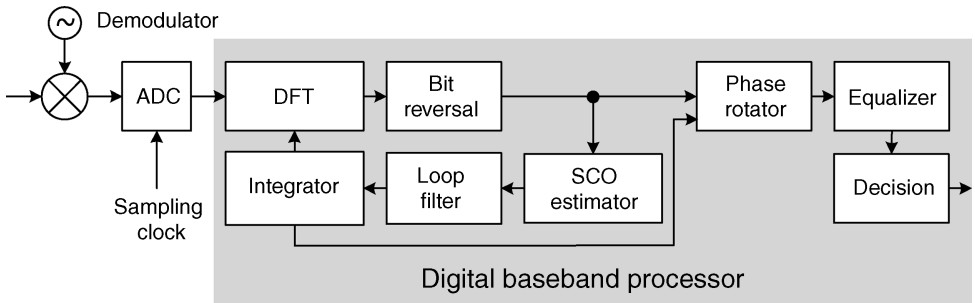


Figure 5.17. OFDM baseband receiver architecture that uses the frequency-domain phase rotator for SCO compensation

criterion) more easily. The transfer function of one such class of all-pass digital filters is given by

$$H(z) = \frac{z^{-N}A(z^{-1})}{A(z)} = \frac{z^{-N}(a_0 + a_1 z + \cdots + a_N z^N)}{a_0 + a_1 z^{-1} + \cdots + a_N z^{-N}}, \tag{5.51}$$

where N is the order of the digital filter; the numerator polynomial is a mirrored version of the denominator $A(z)$; and the coefficients are assumed to be real-valued. To implement a fractional-delay filter, it is imperative to make the phase response of the filter as linear as possible. Besides, since this type of filter has the IIR structure, their stability must be carefully examined. An all-pass fractional-delay filter with maximally flat group delay at zero frequency has been proposed [30]. The coefficients of such a filter that corresponds to a fractional delay (d) take the form of

$$a_k = (-1)^k \binom{N}{k} \frac{(d)_k}{(N+d+1)_k}, \quad k = 0, 1, 2 \ldots, N, \tag{5.52}$$

where d is in the range $[-0.5, 0.5]$, and $(d)_k$ is the k-term product of $d, (d+1), \cdots$, and $(d+k-1)$. In SCO compensation, the fractional delay is time-varying and thus the coefficients must be computed constantly, which may require considerable hardware complexity [31]. Fortunately, the fractional delay is usually expressed as a fixed-point number and thus coefficients can be stored in tables addressed by d. Such an implementation can greatly reduce the complexity of the SCO compensation block [32].

Frequency-Domain Rotator

In the case in which the OFDM communication system has few subcarriers or when the SCO can be controlled within a very small range, the ICI term can be ignored. Hence, only the phase shift of the received frequency-domain signal, which is proportional to the subcarrier index, needs to be corrected. A frequency-domain phase rotator is inserted before the equalizer to reduce the loop latency as well as to avoid the magnitude distortion. The OFDM receiver that uses frequency-domain phase rotator for SCO compensation is depicted in

Petitioner Samsung Ex-1026, 0096

Figure 5.17. In this receiver, the sampling clock offset estimate is first filtered and then integrated to generate the fractional delay. Subsequently, the fractional delay is used to correct the phase shift of the subcarrier data. Meanwhile, the DFT window is adjusted when the fractional delay overflows (> 1) or underflows (< 0).

Bibliography

[1] L. Tomba, 'On the effect of Wiener phase noise in OFDM systems', *IEEE Trans. Commun.*, vol. 46, May 1998, pp. 580–583.

[2] M. Speth, F. Classen and H. Meyr, 'Frame synchronization of OFDM systems in frequency selective fading channels', in *Proc. IEEE Vehicular Technology Conf.*, Phoenix Arizona, USA, May 1997, pp. 1807–1811.

[3] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989.

[4] T. Keller and L. Hanzo, 'Orthogonal frequency division multiplex synchronization techniques for wirelss local area networks', in *Proc. IEEE Int. Symp. on Personal, Indoor, and Mobile Radio Communications*, Taipei, Taiwan, Oct. 1996, pp. 963–967.

[5] M. Sandell, J. J. van de Beek and P. O. Borjesson, 'Timing and frequency synchronization in OFDM systems using the cyclic prefix', in *Proc. of the Int. Symp. on Synchronization*, Essen, Germany, 1995, pp. 16–19.

[6] P. R. Chevillat, D. Maiwald and G. Ungerboeck, 'Rapid training of a voiceband data-modem receiver employing an equalizer with frational-T spaced coefficients', *IEEE Trans. Commun.*, vol. 35, no. 9, pp. 869–876, Sep. 1987.

[7] S. H. Muller-Weinfurtner, 'On the optimality of metric for coarse frame synchronization in OFDM: A comparison', in *Proc. IEEE Int. Symp. on Personal, Indoor, and Mobile Radio Communications*, Boston, MA, Sep. 1998, pp. 533–537.

[8] T. M. Schmidl and D. C. Cox, 'Robust frequency and timing synchronization for OFDM', *IEEE Trans. Commun.*, vol. 45, Dec. 1997, pp. 1613–1621.

[9] H. Minn, V. K. Bhargava and K. B. Letaief, 'A robust timing and frequency synchronization for OFDM systems', *IEEE Trans. Wireless Commun.*, vol.2, Jul. 2003, pp. 822–839.

[10] B. Park, H. Cheon, C. Kang and D. Hong, 'A novel timing estimation method for OFDM Systems', *IEEE Communications Letters*, vol. 7, no. 5, May 2003.

[11] F. Tufvesson, O. Edfors and M. Faulkner, 'Time and frequency synchronization for OFDM using PN-sequence preambles', in *Proc. IEEE Vehicular Technology Conference*, Sep. 1999, vol. 4, pp. 2203–2207.

[12] M. L. Liou and T. D. Chiueh, 'A low-power digital matched filter for direct-sequence spread spectrum signal acquisition', *IEEE Journal of Solid-State Circuits*, vol. 36, no. 6, June 2001, pp. 933–943.

[13] A. Fort, J.-W. Weijers, V. Derudder, W. Eberle and A. Bourdoux, 'A performance and complexity comparison of auto-correlation and cross-correlation for OFDM burst synchronization', in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 2, Apr. 2003, pp. II-341–344.

[14] R. van Nee and R. Prasad, *OFDM Wireless Mutlimedia Communications.* Uitgave: Artech House Publishers, 2000.

[15] B. Yang, K. B. Letaief, R. S. Chen and Z. Cao, 'Timing recovery for OFDM transmission', *IEEE J. Selected Area in Comm.*, vol. 18, Nov. 2000, pp. 2278–2291.

[16] M. Morelli and U. Mengali, 'An improved frequency offset estimator for OFDM applications', *IEEE Commun. Lett.*, vol. 3, Mar. 1999, pp. 75–77.

[17] Y. C. Lei, I. W. Lai and T. D. Chiueh, 'Design of a Baseband Receiver for IEEE 802.16 OFDM Mode Subscriber Station', in *Proc. IEEE International Symp. on Circuits and Systems*, Kos, Greece, May 2006.

[18] D. S. Han, J. H. Seo and J. J. Kim, 'Fast carrier frequency offset compensation in OFDM systems', *IEEE Trans. Consumer Electron.*, vol. 47, Aug. 2001, pp. 364–369.

[19] T. M. Schmidl and D. C. Cox, 'Blind frequency and timing synchronization for OFDM', *IEEE Trans. Commun.*, vol. 45, Dec. 1997, pp. 1613–1621.

[20] B. Ai, J. Ge, Y. Wang, S. Yang, P. Liu and G. Liu, 'Frequency offset estimation for OFDM in wireless communications', *IEEE Trans. Consumer Electron.*, vol. 50, Feb. 2004, pp. 73–77.

[21] Y. H. Kim, Iickho Song, Seokho Yoon and S. R. Park, 'An efficient frequency offset estimator for OFDM systems and its performance characteristics', *IEEE Trans. Veh. Technol.*, vol. 50, Sep. 2001, pp. 1307–1312.

Petitioner Samsung Ex-1026, 0097

[22] M. Speth, S. Fechtel, G. Fock and H. Meyr, 'Optimum receiver design for OFDM-based broadband transmission-part II: A case study', *IEEE Trans. Commum.*, vol. 49, Apr. 2001, pp. 571–578.
[23] S. Y. Liu and J. W. Chong, 'A study of joint tracking algorithms of carrier frequency offset and sampling clock offset for OFDM-based WLANs', in *Proc. IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions*, Chengdu City, China, Jun. 2002, pp. 109–113.
[24] P. Y. Tsai, H. Y. Kang and T. D. Chiueh, 'Joint weighted least squares estimation of carrier frequency offset and timing offset for OFDM systems over multipath fading channel', *IEEE Trans. Vehicular Technol.*, vol. 54, Jan. 2005, pp. 211–224.
[25] T. Pollet, M. van Bladel and M. Moeneclaey, 'BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise', *IEEE Trans. Commun.*, vol. 43, Apr. 1995, pp. 191–193.
[26] J. W. M. Bergmans, 'Effect of loop delay on stabilit of discrete-time PLL', *IEEE Trans. Circuits Syst. I*, vol. 42, Apr. 1995, pp. 229–231.
[27] M. Luise, M. Marselli and R. Reggiannini, 'Low-complexity blind carrier frequency recovery for OFDM signals over frequency-selective radio channels', *IEEE Trans. Commun.*, vol. 50, Jul. 2002, pp. 1182–1188.
[28] T. Pollet, P. Spruyt and M. Moeneclaey, 'The BER performance of OFDM systems using non-synchronized sampling', in *Proc. IEEE Global Telecommunications Conference*, San Francisco, CA, Nov. 1994, pp. 253–257.
[29] L. Erup, F. M. Gardner and R. A. Harris, 'Interpolation in digital modems-part II: Implementation and performance', *IEEE Trans. Commun.*, vol. 41, Jun. 1993, pp. 998–1008.
[30] T. I. Laakso, V. Valimaki, M. Karjalainen and U. K. Laine, 'Splitting the unit delay', *IEEE Signal Processing Mag.*, vol. 13, Jan. 1996, pp. 30–60.
[31] J.S. Park, B. K. Kim, J. G. Chung and K. K. Parhi, 'High-speed tunable fractional-delay allpass filter structure', in *Proc. of IEEE International Symposium on Circuits and Systems*, May 25–28, 2003.
[32] T. P. Wang and T. D. Chiueh, 'A low-complexity fractional delay all-pass filter design for time-domain interpolation', in *Proc. International Symposium on VLSI Design, Automation, and Test*, HsinChu, Taiwan, April, 2007.

Petitioner Samsung Ex-1026, 0098

6

Channel Estimation and Equalization

Channel estimation provides information about distortion of the transmission signal when it propagates through the channel. This information is then used by equalizers so that the fading effect and/or co-channel interference can be removed and the original transmitted signal can be restored.

6.1 Introduction

Channel estimation plays an important role in a communication receiver. In order to mitigate hostile channel effects on the received signal, precise channel estimation is required to provide information for further processing of the received signal. Channel estimators can be categorized as *non-data-aided* or *data-aided*. Non-data-aided or *blind* channel estimators estimate channel response by statistics of the received signals. No specialized reference (training) signals are needed and the transmission efficiency is retained for systems using such channel estimation schemes. However, without precise knowledge of the transmitted signals, a large number of data must be collected in order to obtain reliable estimation. On the other hand, the data-aided channel estimators require known reference (training) signals to be transmitted. Rapid and accurate channel estimation can be achieved by comparing the received and transmitted reference signals. A sufficient number of such reference signals must be inserted according to the degree of channel variation, namely coherence time and coherence bandwidth of the channel under estimation.

OFDM has been known to be quite spectral efficient over frequency-selective fading channels. By dividing a frequency-selective-faded signal band into a large number of narrow-band flat-fading subchannels, high-rate transmission is then achieved by using a compact constellation on each subcarrier. To obtain precise channel estimation for equalizing each subcarrier, most OFDM-based communication standards, such as DVB-T, IEEE 802.11a/g and IEEE 802.16e, provide some forms of reference signals, namely preamble or pilot signals. As a result, this chapter will focus mainly on the data-aided channel estimation algorithms for OFDM communications.

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0099

Once channel estimates at data subcarriers are derived, the receiver performs equalization to compensate for signal distortion. A simple one-tap equalizer is often employed in OFDM systems to deal with flat-faded signals on each subcarrier. However, there are times at which the multipath channel varies so fast that the channel state cannot be regarded as unchanged within one symbol period. In such cases, interference among subcarriers, also known as *inter-carrier interference* (ICI), is induced and must be eliminated in the receiver.

In this chapter, several popular pilot (reference signal) arrangements in OFDM systems will be first introduced. Then, channel estimation algorithms based on different pilot patterns will be addressed. Adaptive channel estimation algorithms that can track the state of the channel when it varies with time will also be discussed. Then, channel equalization schemes as well as ICI cancellation for OFDM signals that travel through highly time-selective fading channels will be covered.

6.2 Pilot Pattern

In OFDM systems, usually dedicated pilot subcarriers are interlaced with data subcarriers. Several possible pilot patterns are depicted in Figure 6.1. The full circles denote the pilot subcarriers while open circles designate data subcarriers. A pilot pattern can be described by a 2×2 sampling matrix $\mathbf{Y} = [\mathbf{y}_1 \;\; \mathbf{y}_2]$ [1]. Two spanning vectors, $\mathbf{y}_1 = [y_{11} \;\; y_{21}]^T$ and $\mathbf{y}_2 = [y_{12} \;\; y_{22}]^T$, generate the doubly periodic sub-lattice of all pilots in the time–frequency lattice. Five pilot patterns are commonly seen in OFDM systems:

- One pilot symbol full of pilot subcarriers for every y^B symbol, also known as the block type:

$$\mathbf{Y}_{\mathrm{B}} = \begin{bmatrix} y^B & 0 \\ 0 & 1 \end{bmatrix}, \tag{6.1}$$

  and $y^B = 7$ in Figure 6.1(a).
- Pilot subcarriers at some predefined locations y^C subcarriers apart for all symbols, also known as the comb type:

$$\mathbf{Y}_{\mathrm{C}} = \begin{bmatrix} 1 & 0 \\ 0 & y^C \end{bmatrix}, \tag{6.2}$$

  and $y^C = 5$ in Figure 6.1(b).
- Pilot subcarriers with fixed spacing in both time and frequency [2]. The periods in time and frequency are y_1^R and y_2^R, respectively:

$$\mathbf{Y}_{\mathrm{R}} = \begin{bmatrix} y_1^R & 0 \\ 0 & y_2^R \end{bmatrix}, \tag{6.3}$$

  and $y_1^R = 4, y_2^R = 4$ in Figure 6.1(c).
- Pilot subcarriers with regular shift in time and frequency [3]:

$$\mathbf{Y}_{\mathrm{P}} = \begin{bmatrix} y_1^P & 1 \\ 0 & y_2^P \end{bmatrix}, \tag{6.4}$$

  and $y_1^P = 4, y_2^P = 3$ in Figure 6.1(d).

Petitioner Samsung Ex-1026, 0100

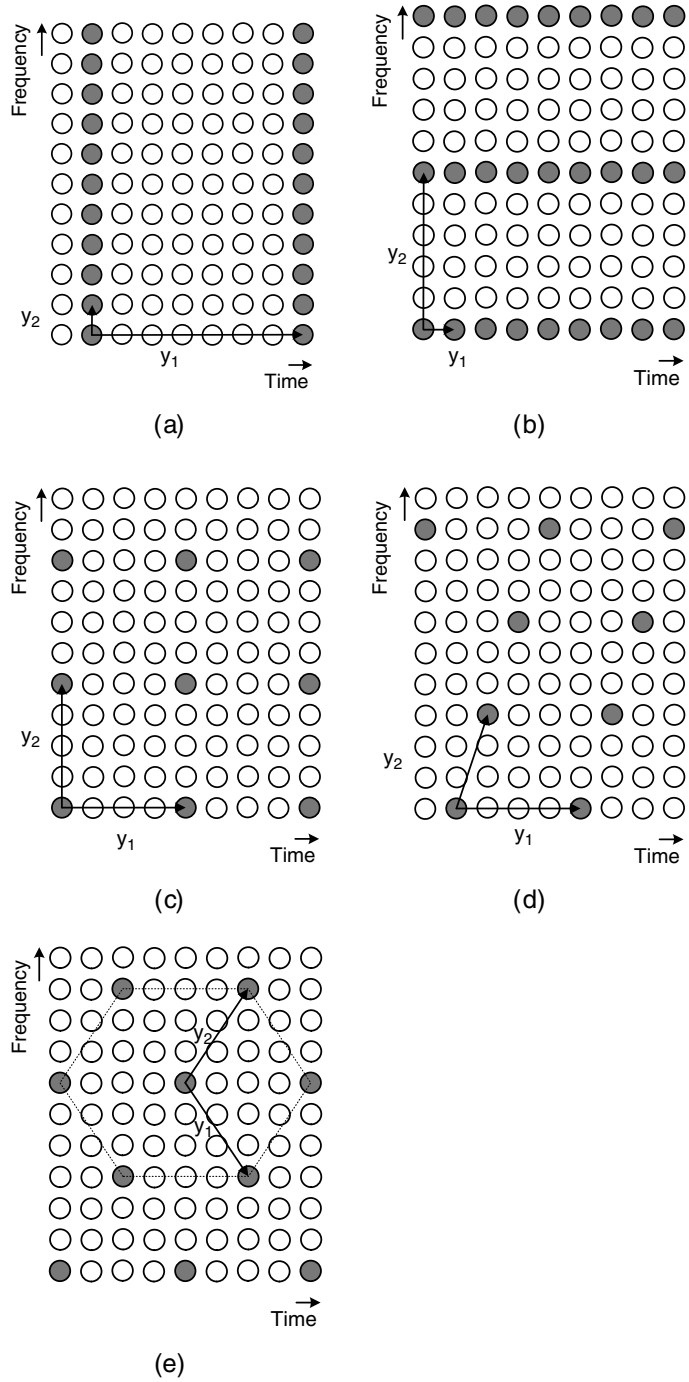


Figure 6.1. Illustrations of several common pilot patterns: (a) block type, (b) comb type, (c) rectangular grid, (d) parallelogram-shaped grid and (e) hexagonal grid

Petitioner Samsung Ex-1026, 0101

- Pilot subcarriers with hexagonal distribution in the time–frequency lattice [1]:

$$\mathbf{Y}_{\mathrm{H}} = \begin{bmatrix} y_1^H & y_1^H \\ -y_2^H & y_2^H \end{bmatrix}, \tag{6.5}$$

and $y_1^H = 2, y_2^H = 3$ in Figure 6.1(e).

In the block-type pilot arrangement, one specific symbol full of pilot subcarriers is transmitted periodically, such as the preamble symbols in the IEEE 802.16e-2005 OFDM mode standard. A receiver can obtain estimation for channel gains at all subcarriers from this preamble symbol and then apply estimated channel gains to equalize the data symbols that follow. These pilot symbols must appear at a frequency tens of times higher than the Doppler frequency in order to ensure the validity of the channel estimates. In other words, the interval between two consecutive pilot symbols must be significantly shorter than the channel coherence time. Consequently, block-type pilot pattern is suitable for systems operating under slow-fading channels.

For the comb-type arrangement, a number of subcarriers are reserved for pilot signals, which are transmitted continuously. Channel estimation can then be performed uninterruptedly based on these pilot subcarriers in every symbol. The spacing of pilot subcarriers must be less than the coherence bandwidth of the channel. The receiver can thus compute the channel estimates for non-pilot subcarriers through the estimated frequency-domain channel responses at the pilot subcarriers. It is widely accepted that effective and accurate channel estimation in fast-fading channels must rely on the comb-type pilot arrangement [4].

For the remaining three pilot patterns, pilot subcarriers provide sub-sampling of the two-dimensional channel responses. Since pilot subcarriers are distributed in the shapes of rectangles, parallelograms or hexagons in the time–frequency lattice, the sampling theorem must be obeyed in both dimensions to avoid the aliasing effect. These scattered pilot arrangements reduce the pilot density and thus improve spectral efficiency. The pilot density is inversely proportional to the area of the quadrilateral spanned by the two vectors of the sampling matrix and is given by

$$|\det([\mathbf{y}_1 \;\; \mathbf{y}_2])|^{-1}, \tag{6.6}$$

where $\det(\cdot)$ is the determinant of a matrix. According to [5], if the two-dimensional channel spectrum is circularly bandlimited, the most efficient pilot pattern is a hexagonal grid, since it requires 13.4% fewer samples than the rectangular grid. Although the choice of the pilot pattern relates to a two-dimensional sampling problem, it may very well be treated as two one-dimensional sampling problems, since radio channels have uncorrelated characteristics in time and frequency [6].

6.3 Pilot-Based Channel Estimation

Figure 6.2 shows a typical block diagram of a baseband OFDM system based on pilot-aided channel estimation. The inverse DFT block transforms frequency-domain data, X_k, on the

Petitioner Samsung Ex-1026, 0102

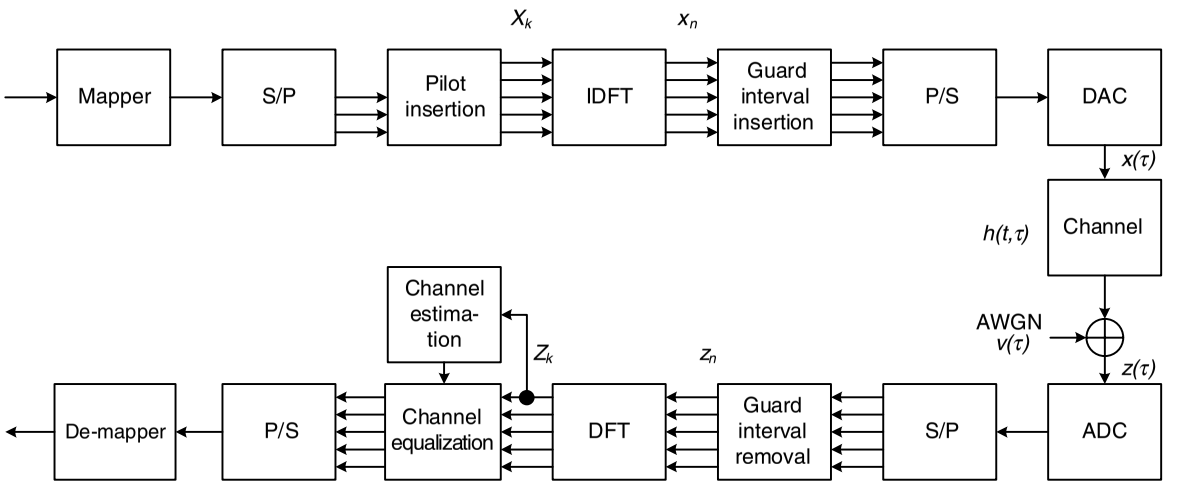


Figure 6.2. Block diagram of a typical pilot-aided OFDM system

Petitioner Samsung Ex-1026, 0103

k-th subcarrier into time-domain samples x_n as

$$x_n = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} X_k e^{j2\pi nk/N}, \quad n = -N_g, \cdots, 0, \cdots, N-1, \tag{6.7}$$

where N is the number of total subcarriers and N_g is the number of guard interval samples reserved for coping with time-domain dispersion of the channel.

Assume that the impulse response of the multipath fading channel is given by

$$h(t,\tau) = \sum_r h_r(t) \cdot \delta(\tau - \tau_r), \tag{6.8}$$

where the gain and delay of the r-th path are denoted by $h_r(t)$ and τ_r, respectively. The path gains, $h_r(t)$, are *wide-sense stationary* (WSS) narrow-band complex Gaussian processes and are mutually independent. The received signal, which has been corrupted by the multipath fading channel and contaminated by the additive white Gaussian noise ($v(\tau)$), takes the form of

$$z(\tau) = \sum_r h_r(t) \cdot x(\tau - \tau_r) + v(\tau), \tag{6.9}$$

where $x(\tau)$ is the continuous-time representation of the transmitted discrete-time signal, x_n. After converting the received continuous-time signal back to a discrete-time signal, z_n, the receiver then removes the cyclic prefix and transforms z_n to the frequency domain:

$$Z_k = \sum_{n=0}^{N-1} z_n \cdot e^{-j2\pi nk/N}, \quad k = -N/2, \cdots, N/2-1. \tag{6.10}$$

Assume that the duration of the cyclic prefix is long enough so that there is no *inter-symbol interference* (ISI). Furthermore, assume that the channel is almost stationary within one symbol period, namely $h_r(t) = h_r$; then, the frequency-domain channel response at subcarrier k, H_k, is given by

$$H_k = \sum_r h_r e^{-j2\pi \frac{\tau_r k}{T_s N}}, \tag{6.11}$$

where T_s is the sample interval. Denote V_k as the frequency-domain counterpart of $v(\tau)$; the frequency-domain received baseband data Z_k then becomes

$$Z_k = X_k H_k + V_k. \tag{6.12}$$

Due to the time-domain sampling effect, the equivalent discrete-time *channel impulse response* (CIR), g_n for $n = 0, 1, \cdots, N-1$, is the summation of discrete-time circular *sinc* waveforms–one per arrival path and weighted by the path gain h_r. Each circular *sinc*

Petitioner Samsung Ex-1026, 0104

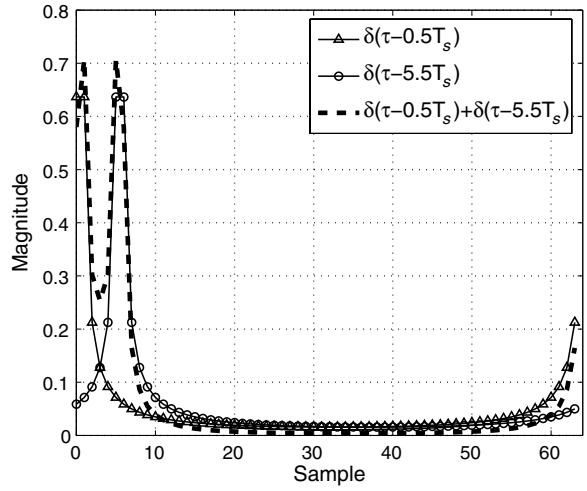


Figure 6.3. Energy leakage in discrete-time channel impulse magnitude response

waveform is coupled with rotating phasor due to the fractional time difference between the arrival time of that path and the sampling time instant:

$$\begin{aligned} g_n &= \frac{1}{N}\sum_{k=-N/2}^{N/2-1} H_k e^{j2\pi\frac{nk}{N}} \\ &= \frac{1}{N}\sum_{k=-N/2}^{N/2-1}\sum_{r} h_r e^{-j2\pi\frac{\tau_r k}{T_s N}} e^{j2\pi\frac{nk}{N}} \\ &= \frac{1}{N}\sum_{r} h_r e^{-j\pi(\frac{\tau_r}{T_s}-n)/N} \frac{\sin\left(\pi\left(\frac{\tau_r}{T_s}-n\right)\right)}{\sin\left(\frac{\pi}{N}\left(\frac{\tau_r}{T_s}-n\right)\right)}. \end{aligned} \tag{6.13}$$

If the delay τ_r does not coincide with a sample point, which is usually the case, *energy leakage* will occur [7]. Figure 6.3 depicts the discrete-time channel impulse magnitude response of a two-path channel, $\delta(\tau - 0.5T_s) + \delta(\tau - 5.5T_s)$. Note that the FFT size is set to 64. It is clear that the signal energy has spread to all other taps in g_n. However, most of the energy is still concentrated around the original pulse locations. The received time-domain signals z_n can be viewed as the circular convolution of the transmitted sequence (x_n) and the discrete-time channel impulse response (g_n) plus noise.

Define the DFT matrix, **F**, as

$$\mathbf{F} = \begin{bmatrix} 1 & e^{-j2\pi(-N/2)\cdot 1/N} & \dots & e^{-j2\pi(-N/2)\cdot(N-1)/N} \\ & \vdots & & \vdots \\ 1 & e^{-j2\pi(N/2-1)\cdot 1/N} & \dots & e^{-j2\pi(N/2-1)\cdot(N-1)/N} \end{bmatrix}. \tag{6.14}$$

Petitioner Samsung Ex-1026, 0105

After rewriting Equation (6.12) in matrix-vector notation, the received signal vector takes the form of

$$\mathbf{z} = \mathbf{XFg} + \mathbf{v} = \mathbf{Xh} + \mathbf{v}, \tag{6.15}$$

where

$$\begin{aligned}
\mathbf{z} &= [Z_{-N/2} \;\; Z_{-N/2+1} \;\; \cdots \;\; Z_{N/2-1}]^T, \\
\mathbf{g} &= [g_0 \;\; g_1 \;\; \cdots \;\; g_{N-1}]^T, \\
\mathbf{v} &= [V_{-N/2} \;\; V_{-N/2+1} \;\; \cdots \;\; V_{N/2-1}]^T, \\
\mathbf{h} &= [H_{-N/2} \;\; H_{-N/2+1} \;\; \cdots \;\; H_{N/2-1}]^T, \\
\mathbf{X} &= \mathrm{diag}[X_{-N/2} \;\; X_{-N/2+1} \;\; \cdots \;\; X_{N/2-1}],
\end{aligned} \tag{6.16}$$

and diag$[\cdot]$ constructs a matrix using the arguments as the diagonal elements.

Now, suppose that the channel responses at all data subcarriers are estimated using pilot subcarriers and are denoted as $\hat{H}_k, k = -N/2, \cdots, 0, \cdots, N/2-1$; the received data can then be equalized by

$$\hat{X}_k = \frac{Z_k}{\hat{H}_k}, \quad k = -N/2, \cdots, 0, \cdots, N/2-1. \tag{6.17}$$

Previously, the pilot patterns have been categorized into the block type, the comb type and the grid type (rectangle, parallelogram and hexagon). In the following, channel estimation algorithms for these three types of pilot patterns will be introduced.

6.3.1 Channel Estimation by Block-Type Pilot Symbols

To facilitate discussing channel estimators based upon the block-type pilot pattern, only the pilot symbol is examined in the received signals. Consequently, X_k is known to the receiver for all k.

MMSE Estimator

Assume that the equivalent time-domain channel impulse response, $\mathbf{g}$, is a random vector with Gaussian distribution and is uncorrelated with the noise, $\mathbf{v}$. Assume that $\mathbf{v}$ has a covariance matrix $\sigma_v^2\mathbf{I}_N$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Then, the *minimum mean-squared error* (MMSE) estimator that minimizes $E\{(\hat{\mathbf{g}} - \mathbf{g})^H(\hat{\mathbf{g}} - \mathbf{g})\}$ takes the form of [7]

$$\hat{\mathbf{g}}_{\mathrm{MMSE}} = \mathbf{R}_{\mathbf{gz}}\mathbf{R}_{\mathbf{zz}}^{-1}\mathbf{z}, \tag{6.18}$$

where

$$\mathbf{R}_{\mathbf{gz}} = E\{\mathbf{gz}^H\} = \mathbf{R}_{\mathbf{gg}}\mathbf{F}^H\mathbf{X}^H \tag{6.19}$$

is the cross-covariance matrix between $\mathbf{g}$ and $\mathbf{z}$;

$$\mathbf{R}_{\mathbf{zz}} = E\{\mathbf{zz}^H\} = \mathbf{XFR}_{\mathbf{gg}}\mathbf{F}^H\mathbf{X}^H + \sigma_v^2\mathbf{I}_N \tag{6.20}$$

Petitioner Samsung Ex-1026, 0106

is the auto-covariance matrix of $\mathbf{z}$; $\mathbf{R_{gg}}$ is the auto-covariance matrix of $\mathbf{g}$, assumed to be known in advance. As a result, the MMSE frequency-domain channel response $\hat{\mathbf{h}}_{\text{MMSE}} = [\hat{H}_{-N/2} \ \cdots \ \hat{H}_{N/2-2} \ \hat{H}_{N/2-1}]^T$ is given by

$$\hat{\mathbf{h}}_{\text{MMSE}} = \mathbf{F}\hat{\mathbf{g}}_{\text{MMSE}}. \tag{6.21}$$

Low-Rank Approximation

The MMSE channel estimator can be written more explicitly by substituting Equations (6.15) and (6.18) into Equation (6.21):

$$\hat{\mathbf{h}}_{\text{MMSE}} = \mathbf{R_{hh}}(\mathbf{R_{hh}} + \sigma_v^2(\mathbf{X}^H\mathbf{X})^{-1})^{-1}(\mathbf{X}^{-1}\mathbf{z}), \tag{6.22}$$

where $\mathbf{R_{hh}} = E\{\mathbf{hh}^H\}$ is the auto-covariance matrix of $\mathbf{h}$. In MMSE channel estimation, matrix inversion is required for each symbol. Low-rank approximation is proposed to reduce the complexity of the MMSE estimator [8]. In this method, $(\mathbf{X}^H\mathbf{X})^{-1}$ is replaced by its expectation, $E\{(\mathbf{X}^H\mathbf{X})^{-1}\}$. Then, since $\mathbf{R_{hh}}$ is a square and Hermitian matrix, rank reduction is achieved through *eigen decomposition*, which decomposes the channel auto-covariance matrix into

$$\mathbf{R_{hh}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H. \tag{6.23}$$

Note that $\mathbf{U}$ is an unitary matrix consisting of eigenvectors. The diagonal matrix $\boldsymbol{\Lambda}$ is equal to $\text{diag}[\lambda_0 \ \lambda_1 \ \lambda_2 \ \cdots \ \lambda_{N-1}]$, where $\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1}$, and λ_n is its eigenvalue. The rank-p MMSE channel estimation is [8]

$$\hat{\mathbf{h}}_{\text{ED}} = \mathbf{U}\boldsymbol{\Delta}_p\mathbf{U}^H\mathbf{X}^{-1}\mathbf{z}. \tag{6.24}$$

The n-th entry of the diagonal matrix $\boldsymbol{\Delta}_p$ is

$$\rho_n = \begin{cases} \dfrac{\lambda_n}{\lambda_n + (\alpha/SNR)} & n = 0, 1, \cdots, p-1 \\ 0 & n = p, p+1, \cdots, N-1, \end{cases} \tag{6.25}$$

where

$$\alpha = E\{|X_k|^2\}E\{|1/X_k|^2\} \tag{6.26}$$

is a constant determined by the constellation. For example, $\alpha = 17/9$ for 16-point QAM modulation.

It has been shown that the minimum rank p in the channel estimation problems approximately corresponds to N_g [8]. It can be interpreted that the eigenvalue λ_n is the channel power of the n-th transform coefficient. Hence, after about N_g taps, the effective channel power almost vanishes and thus can be neglected. The block diagram of this low-rank channel estimator is depicted in Figure 6.4.

Petitioner Samsung Ex-1026, 0107

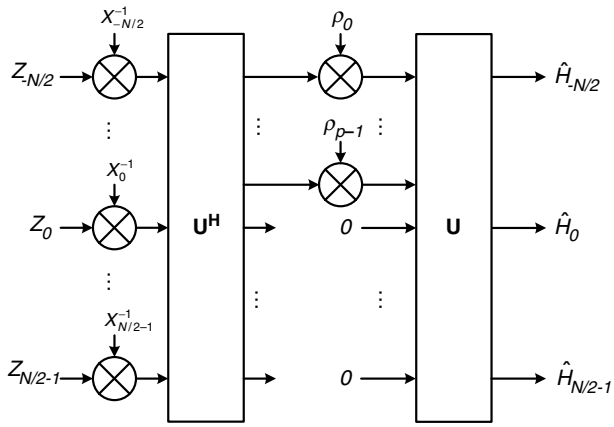


Figure 6.4. Block diagram of the low-rank channel estimator by eigen decomposition

Least-Squares Estimator

The *least-squares* (LS) channel estimator is also a *maximum likelihood* (ML) estimator, which assumes that the time-domain channel impulse response is deterministic and tries to find $\hat{\mathbf{g}}_{LS}$ which minimizes $(\mathbf{z} - \mathbf{XFg})^H(\mathbf{z} - \mathbf{XFg})$. The LS solution is then given by [7]

$$\hat{\mathbf{h}}_{\mathrm{LS}} = \mathbf{F}(\mathbf{F}^H\mathbf{X}^H\mathbf{XF})^{-1}\mathbf{F}^H\mathbf{X}^H\mathbf{z}. \tag{6.27}$$

Note that in block-type pilot symbols, matrices $\mathbf{F}$ and $\mathbf{X}$ are invertible square matrices; hence,

$$\hat{\mathbf{h}}_{\mathrm{LS}} = \mathbf{F}(\mathbf{F}^{-1}\mathbf{X}^{-1}(\mathbf{X}^H)^{-1}(\mathbf{F}^H)^{-1})\mathbf{F}^H\mathbf{X}^H\mathbf{z} = \mathbf{X}^{-1}\mathbf{z}. \tag{6.28}$$

Reduced-Order ML Estimator

A reduced-order ML estimator that exploits the finite length of the channel impulse response is proposed in [9]. It assumes that the maximum excess delay is shorter than the length of cyclic prefix (N_g), so the LS channel estimation is just the frequency-domain counterpart of the finite-length time-domain channel impulse response estimation $\mathbf{q}$ plus noise $\mathbf{u}$, and

$$\hat{\mathbf{h}}_{\mathrm{LS}} = \mathbf{F}\begin{bmatrix}\mathbf{q}\\ 0\end{bmatrix} + \mathbf{u}. \tag{6.29}$$

Retaining only the first N_g time-domain channel impulse response samples and nullifying the remaining noisy samples can enhance the estimation quality. So, defining $\mathbf{F}_l$ as the matrix

Petitioner Samsung Ex-1026, 0108

containing the first N_g columns in $\mathbf{F}$ and $\mathbf{F}_l^\dagger = (\mathbf{F}_l^H \mathbf{F}_l)^{-1} \mathbf{F}_l^H$ as the pseudo-inverse of $\mathbf{F}_l$, then the reduced-order ML estimate is given by

$$\hat{\mathbf{h}}_{\mathrm{ML}} = \mathbf{F}_l \mathbf{F}_l^\dagger \hat{\mathbf{h}}_{\mathrm{LS}}. \tag{6.30}$$

It can be interpreted as transforming the frequency-domain LS estimate to the time domain, eliminating the noisy samples in the tail of the waveform, and then transforming back to the frequency domain.

Discussions

The channel response (either time-domain or frequency-domain) is inherently a random process. The MMSE and the low-rank approximation estimators regard the channel response as a stationary random vector. Consequently, the statistics about this random vector, such as the covariance matrix and the signal-to-noise ratio (SNR), can help the estimation. However, it also implies that knowledge about the channel covariance matrix and the operating SNR necessitates additional time and complexity to come by. On the other hand, the channel response is regarded as a deterministic yet unknown vector in the LS and reduced-order ML channel estimation. This assumption basically uses a snap shot to represent the slow-varying random process.

With the knowledge of extra statistical information, the MMSE estimator can outperform the LS estimator. Nevertheless, in high-SNR scenarios, the MMSE estimator boils down to the LS estimator. Both the low-rank approximation and the reduced-order ML estimator eliminate the noisy part of the estimated channel impulse response. Hence, the reduced-order ML estimator gets better estimation quality than the LS estimator. However, the low-rank approximation loses statistical information contained in the noise subspace and performs poorer than the original MMSE estimator.

6.3.2 Channel Estimation by Comb-Type Pilot Symbols

In OFDM systems with comb-type pilots, the MMSE criterion and the ML criterion mentioned in the previous section can still be used [10, 11]. However, the ML estimator and the low-rank approach can be derived if and only if the number of pilot subcarriers is greater than the number of channel taps or the length of the guard interval. The MMSE estimator, on the other hand, can be applied without such a constraint. In [10] and [11], interested readers can obtain similar derivations as Equations (6.18), (6.24) and (6.30) for the comb-type pilot-aided channel estimation.

In addition to these approaches, a variety of comb-type pilot-aided channel estimation schemes have been proposed. Among them, two classes of methods are very popular, namely time-domain windowing and frequency-domain interpolation. In the following, some definitions are first introduced before further discussion on this topic. As shown in Figure 6.5, there are a total of N subcarriers. Among these N subcarriers, only $N_u + 1$ subcarriers are used for transmission, and the others are reserved as guard bands on both ends of the signal spectrum. There are $M_p + 1$ non-zero pilot subcarriers for channel estimation inserted in the $N_u + 1$ subcarriers, where $M_p = N_u/D$. Note that D is an integer and that $M = N/D$.

Petitioner Samsung Ex-1026, 0109

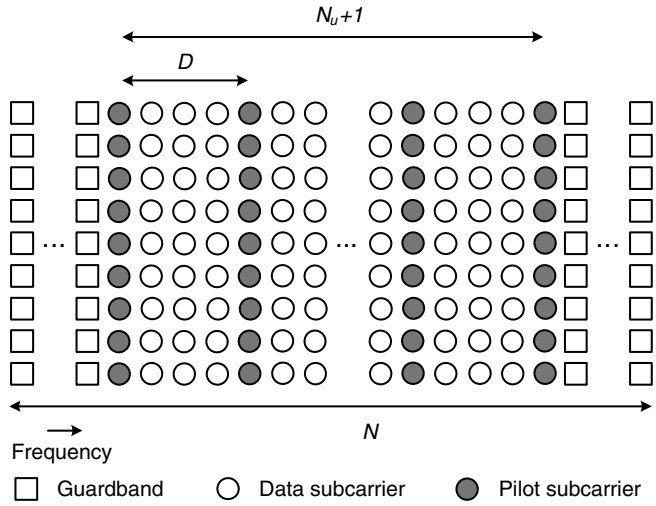


Figure 6.5. Comb-type pilot subcarriers allocation

In the time-domain windowing algorithms, time-domain channel impulse response (CIR) is reconstructed by first inverse Fourier transforming the frequency-domain channel response at only the pilot subcarriers. In this case, the number of pilot subcarriers M_p must be greater than the maximum excess delay to avoid aliasing, namely $M > \tau_{\mathrm{MAX}}/T_s$ [12]. Henceforth, different windowing techniques are applied to the contaminated time-domain channel impulse response in order to diminish the noise level as well as the aliasing effect. Then, zeros are padded to form a total of N samples. Consequently, the frequency-domain channel responses are derived by transforming the 'cleaned-up' time-domain CIR back to the frequency domain. The frequency-domain interpolation algorithms up-sample the channel frequency responses at pilot subcarriers and then apply various interpolators to estimate the channel responses at data subcarriers.

Mathematically, the two classes of channel estimation algorithms can be shown to be equivalent. From the received frequency-domain signals at the pilot subcarriers, the channel responses at the pilot subcarriers can be given by

$$\tilde{H}_{mD} = Z_{mD}/X_{mD}, \quad m = -M_p/2, \cdots, M_p/2. \tag{6.31}$$

Applying the M-points IDFT to these channel responses, one then has the noisy reconstructed time-domain channel impulse response, $\tilde{q}_n$, as

$$\tilde{q}_n = \frac{1}{M} \sum_{m=-M_p/2}^{M_p/2} \tilde{H}_{mD} e^{j2\pi mn/M} = q_n + \nu_n, \qquad n = 0, 1, \cdots, M-1, \tag{6.32}$$

where ν_n denotes the noise term. Note that the reconstructed CIR, $\tilde{q}_n$, is not equivalent to the original CIR, g_n. This is because the guard bands cause low-pass filtering of g_n and the

Petitioner Samsung Ex-1026, 0110

sub-sampling in the frequency-domain pilot subcarriers results in folding of the original channel impulse response with a period of M samples. As such, the reconstructed channel impulse response is

$$q_n = g_n \otimes \frac{\sin(\pi(N_u+1)n/N)}{N\sin(\pi n/N)} \otimes \sum_{i=0}^{D-1} \delta(n - iM). \tag{6.33}$$

Assume that a time-domain window is applied to $\tilde{q}_n$ and the window is given by

$$\mathbf{w} = [w_{-M/2+1+d} \quad w_{-M/2+2+d} \quad \cdots \quad w_{M/2+d}]^T, \tag{6.34}$$

where d denotes the left boundary of the window. Then, the frequency-domain channel estimates can be derived from Fourier transforming the windowed channel impulse response as

$$\begin{aligned}\hat{H}_k &= \sum_{n=-M/2+1+d}^{M/2+d} \tilde{q}_n w_n e^{-j2\pi nk/N} \\ &= \frac{1}{M} \sum_{m=-M_p/2+1}^{M_p/2} \tilde{H}_{mD} \sum_{n=-M/2+1+d}^{M/2+d} w_n e^{-j2\pi n(k-mD)/N} \\ &= \sum_{m=-M_p/2+1}^{M_p/2} \tilde{H}_{mD} W_{k-mD}.\end{aligned} \tag{6.35}$$

The equation above can be interpreted as interpolation in the frequency domain using $\tilde{H}_{mD}$ as the base points and the interpolation coefficients

$$W_l = \frac{1}{M} \sum_{n=-M/2+1+d}^{M/2+d} w_n e^{-j2\pi nl/N}. \tag{6.36}$$

Similarly, in a frequency-domain interpolation algorithm, a corresponding time-domain window function can be derived. For a set of J-tap interpolation coefficients, W_l, the windowing function takes the form of

$$w_n = \frac{1}{D} \sum_{l=-JD/2+1}^{JD/2} W_l e^{j2\pi nl/N}. \tag{6.37}$$

Rectangular Windowing

In the windowing method proposed in [13], the time-domain channel impulse response samples below a certain threshold are cut off. In the same vein, the window in [14] only keeps the more significant samples. In [15], however, the window covers M samples. All the

Petitioner Samsung Ex-1026, 0111

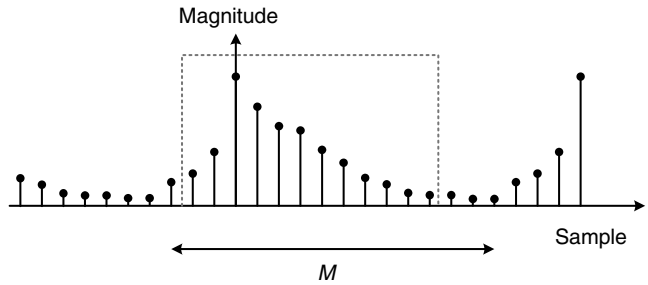


Figure 6.6. Time-domain rectangular window

above approaches can be regarded as applying a rectangular window to the reconstructed time-domain channel impulse response as shown in Figure 6.6 except that the position and the length of the window are different.

In low-SNR scenarios, the windows that depend on a threshold and the windows that determine selected samples according to the sample magnitude can reduce the estimation error level. On the contrary, in high-SNR environments, gathering more time-domain samples as in [15] can avoid neglecting samples with smaller magnitude and thus will have better estimation quality.

MMSE Windowing

In [12], an MMSE window that minimizes the mean-squared error ($E\{(\hat{\mathbf{h}}-\mathbf{h})^H(\hat{\mathbf{h}}-\mathbf{h})\}$) is proposed. Assume that the frequency-domain channel responses at the pilot subcarriers are

$$\tilde{\mathbf{h}}_p = \left[\tilde{H}_{-\frac{M_p}{2}D} \ \cdots \ \tilde{H}_{\frac{M_p}{2}D}\right]^T \tag{6.38}$$

and the estimated channel responses are represented as

$$\hat{\mathbf{h}} = [\hat{H}_{-\frac{N_u}{2}} \ \hat{H}_{-\frac{N_u}{2}+1} \ \cdots \ \hat{H}_{\frac{N_u}{2}}]^T. \tag{6.39}$$

If the weighting vector $\mathbf{w}$ is applied to the time-domain reconstructed CIR, then $\hat{\mathbf{h}}$ is given by

$$\hat{\mathbf{h}} = \mathbf{F}_d \cdot \mathrm{diag}(\mathbf{w}) \cdot \mathbf{G}_d \tilde{\mathbf{h}}_p, \tag{6.40}$$

where the function $\mathrm{diag}(\cdot)$ generates a matrix whose diagonal terms are the entries in its argument. Matrices $\mathbf{G}_d$ and $\mathbf{F}_d$ are the $M \times (M_p+1)$ IDFT matrix and the $(N_u+1) \times M$ DFT matrix, respectively. They take the form of

$$\mathbf{G}_d = \frac{1}{M}\begin{bmatrix} e^{j2\pi(-\frac{M}{2}+d+1)(-\frac{M_p}{2})/N} & \cdots & e^{j2\pi(-\frac{M}{2}+d+1)(\frac{M_p}{2})/N} \\ \vdots & & \vdots \\ e^{j2\pi(\frac{M}{2}+d)(-\frac{M_p}{2})/N} & \cdots & e^{j2\pi(\frac{M}{2}+d)(\frac{M_p}{2})/N} \end{bmatrix}, \tag{6.41}$$

Petitioner Samsung Ex-1026, 0112

and

$$\mathbf{F}_d = \begin{bmatrix} e^{-j2\pi(-\frac{N_u}{2})(-\frac{M}{2}+d+1)/N} & \dots & e^{-j2\pi(-\frac{N_u}{2})(\frac{M}{2}+d)/N} \\ \vdots & & \vdots \\ e^{-j2\pi\frac{N_u}{2}(-\frac{M}{2}+d+1)/N} & \dots & e^{-j2\pi\frac{N_u}{2}(\frac{M}{2}+d)/N} \end{bmatrix}. \tag{6.42}$$

Note that the parameter d controls the time-domain samples for windowing.

The MMSE window is derived as [12]

$$\begin{aligned} \mathbf{w}_{\text{MMSE}} &= \left(E\{\text{diag}(\mathbf{G}_d\tilde{\mathbf{h}}_p)^H\mathbf{F}_d^H\mathbf{F}_d \cdot \text{diag}(\mathbf{G}_d\tilde{\mathbf{h}}_p)\}\right)^{-1} \cdot E\{\text{diag}(\mathbf{G}_d\tilde{\mathbf{h}}_p)^H\mathbf{F}_d^H\mathbf{h}\} \\ &= \left[(\mathbf{G}_d\mathbf{R}_{\tilde{\mathbf{h}}_p\tilde{\mathbf{h}}_p}\mathbf{G}_d^H)^* \odot (\mathbf{F}_d^H\mathbf{F}_d)\right]^{-1} \cdot \text{diag}^{-1}\left(\mathbf{F}_d^H\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_p}\mathbf{G}_d^H\right), \end{aligned} \tag{6.43}$$

where $\odot$ indicates a component-wise product, and the function $\text{diag}^{-1}(\cdot)$ generates a vector with entries from the diagonal terms of its argument matrix.

Polynomial-Based Interpolation

Typically, in the OFDM system, the pilot subcarriers over-sample the channel frequency response by at least twofold to avoid aliasing in the time domain CIR. In other words, $M \geq 2\tau_{\text{MAX}}/T_s$ [16]. The channel responses at non-pilot subcarriers are then interpolated from the responses at the pilot subcarriers. Polynomial-based interpolators have been popular in the frequency-domain interpolation algorithms due to their low implementation complexity. Linear interpolation has been proposed to estimate the frequency-domain channel responses at data subcarriers [17]. For the k-th subcarrier to be interpolated, let $k/D = m + \mu$, where $0 \leq \mu < 1$, and $m = \lfloor \frac{k}{D} \rfloor$, the largest integer smaller than k/D. Then, the linear interpolation method obtains the channel response at the k-th subcarrier as [17]

$$\hat{H}_k = \hat{H}_{D(m+\mu)} = (1-\mu)\tilde{H}_{mD} + \mu\tilde{H}_{(m+1)D}. \tag{6.44}$$

The estimation quality can be improved by using higher-order polynomials. However, the implementation grows more complicated as the order increases. A piecewise second-order polynomial interpolation is adopted in [11], and

$$\begin{aligned} \hat{H}_k &= \hat{H}_{D(m+\mu)} \\ &= C_0\tilde{H}_{mD} + C_{-1}\tilde{H}_{(m+1)D} + C_{-2}\tilde{H}_{(m+2)D}, \end{aligned} \tag{6.45}$$

where

$$\begin{cases} C_0 = (1-\mu)(2-\mu)/2 \\ C_{-1} = \mu(2-\mu) \\ C_{-2} = -\mu(1-\mu)/2 \end{cases}. \tag{6.46}$$

Petitioner Samsung Ex-1026, 0113

Other high-order polynomial-based interpolators, such as the piecewise parabolic interpolator and the cubic interpolator [18], take in four base points for interpolation:

$$\begin{aligned}\hat{H}_k &= \hat{H}_{D(m+\mu)} \\ &= C_1\tilde{H}_{(m-1)D} + C_0\tilde{H}_{mD} + C_{-1}\tilde{H}_{(m+1)D} + C_{-2}\tilde{H}_{(m+2)D}.\end{aligned} \tag{6.47}$$

In the piecewise parabolic interpolator, the coefficients are given by

$$\begin{cases} C_1 = -\alpha\mu + \alpha\mu^2 \\ C_0 = 1 + (\alpha - 1)\mu - \alpha\mu^2 \\ C_{-1} = (\alpha + 1)\mu - \alpha\mu^2 \\ C_{-2} = -\alpha\mu + \alpha\mu^2 \end{cases}. \tag{6.48}$$

Usually, α is set to 0.5 to provide better interpolation quality. On the other hand, the coefficients of the cubic interpolator are

$$\begin{cases} C_1 = -\frac{1}{3}\mu + \frac{1}{2}\mu^2 - \frac{1}{6}\mu^3 \\ C_0 = 1 - \frac{1}{2}\mu - \mu^2 + \frac{1}{2}\mu^3 \\ C_{-1} = \mu + \frac{1}{2}\mu^2 - \frac{1}{2}\mu^3 \\ C_{-2} = -\frac{1}{6}\mu + \frac{1}{6}\mu^3 \end{cases}. \tag{6.49}$$

With four base points and smoother curves formed by the interpolation coefficients, the piecewise parabolic and cubic interpolators have relatively flat main lobes in the corresponding time-domain windows when compared with the linear interpolator. In addition, the high-order polynomial-based interpolators have relatively lower sidelobes and, thus, diminishing noisy samples in the estimated channel response.

Shifted Raised-Cosine Interpolation

In an OFDM receiver, uncertainty in timing often leads to poor demodulation performance. Therefore, channel estimation needs to carefully consider the timing of the channel impulse response. Suppose that the timing acquisition is accurate and the obtained symbol boundary is aligned to the origin, then the first impulse of the CIR will lie at the origin and the rest of the CIR appears inside the guard interval $[0, N_g T_s]$. In reality, however, due to the energy leakage, there exist pre-cursor as well as post-cursor in the reconstructed channel impulse response. A good time-domain window must preserve the major portion of the reconstructed channel impulse response and, at the same time, reject the aliased and noisy components. It is clear from Figure 6.7 that the time-domain window must be shifted to the right instead of centring at the origin. Note that shifting the window in the time domain is equivalent to rotating the phase of the interpolation coefficients in the frequency domain.

The window must be flat over the interval where the CIR is strong so that no distortion will be introduced [19]. At the two ends of the window, the weighting should be smaller in order to suppress noises and the aliasing effect. Finally, smoother weighting in the time domain entails faster fall-off in the frequency-domain interpolation coefficients, and thus

Petitioner Samsung Ex-1026, 0114

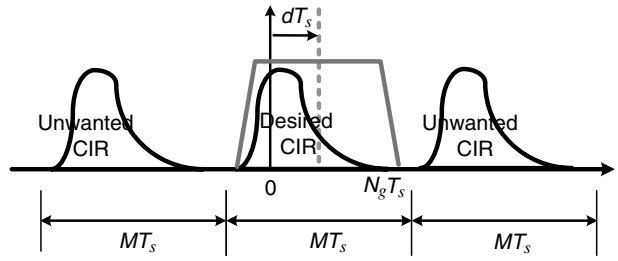


Figure 6.7. Periodic CIR in the time domain

fewer of them are needed. In light of the above considerations, a special interpolator has been designed [20], in which raised-cosine function is used as the frequency-domain interpolation coefficients,

$$W_{l,\mathrm{RC}} = \frac{\sin(\pi Ml/N)}{\pi Ml/N} \cdot \frac{\cos(\pi\beta Ml/N)}{1 - 4\beta^2 (Ml/N)^2} \cdot e^{-j2\pi dl/N}. \tag{6.50}$$

The parameter β is the roll-off factor, which decides the excess width of the window's main lobe. The parameter d controls the position of the window shifted to the right so as to include as much CIR energy as possible.

Discussions

In general, the MMSE windowing algorithm achieves the best estimation performance among all methods. One drawback is, of course, that the receiver must have all the channel statistics available. The MMSE window has larger weights for the strong channel taps and smaller weights for the weak channel taps. On the other hand, the rectangular windowing algorithm simply discards all the weak channel taps and thus cannot obtain precise channel estimates.

The corresponding time-domain windows of the polynomial-based interpolators and the raised-cosine interpolator are depicted in Figure 6.8. Since the windows for polynomial-based interpolators have rather significant side lobes and are centred at the origin, they can be only used in those cases in which the pilot subcarriers sample the channel frequency response at a high enough rate. To wit, for the channel estimation methods using the polynomial-based interpolators, the number of pilot subcarriers must be much larger than the normalized maximum excess delay τ_{MAX}/T_s, otherwise there will be time-domain aliasing. The shifted raised-cosine interpolator, on the other hand, has a window with a flat top and no side lobe, so its window can be free of waveform distortion and aliasing. In addition, it provides the capability of shifting the window position to allow for maximum coverage of the CIR samples and high-fidelity channel estimation. All of the above features make the shifted raised-cosine interpolation channel estimator attractive, especially when there are only a limited number of pilot subcarriers.

Petitioner Samsung Ex-1026, 0115

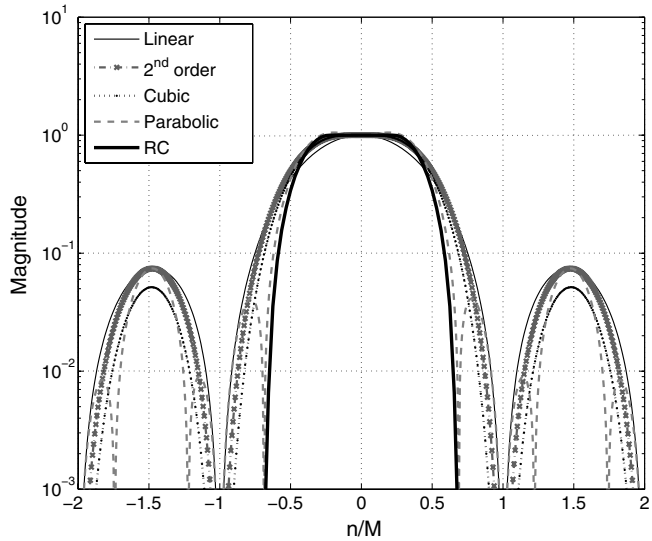


Figure 6.8. Time-domain window of polynomial-based interpolators and raised-cosine interpolator

6.3.3 Channel Estimation by Grid-Type Pilot Symbols

The rectangular, parallelogram-shaped and hexagonal pilot configurations sample the channel frequency response in both time domain and frequency domain. Such arrangements minimize the overheads spent on pilot subcarriers in channels that experience both time-selective fading and frequency-selective fading. According to the sampling theorem, any band-limited one-dimensional deterministic signal can be represented by samples taken at the Nyquist rate. The theorem holds also for multi-dimensional signals.

The two-dimensional correlation function of the response of a time- and frequency-selective channel must be derived first, before the introduction of the two-dimensional channel estimation. Assume that the time-domain channel impulse response is

$$h(t,\tau) = \sum_r h_r(t) \cdot \delta(\tau - \tau_r). \tag{6.51}$$

The time-varying path gain, $h_r(t)$, is a wide-sense stationary complex Guassian process with an average power σ_r^2 and mutually independent. Without loss of generality, the total path power is normalized, namely $\sum_r \sigma_r^2 = 1$. The time-varying channel frequency response at time t is

$$H(t,f) = \int_{-\infty}^{\infty} h(t,\tau) e^{-j2\pi f\tau} d\tau = \sum_r h_r(t) e^{-j2\pi f \tau_r}. \tag{6.52}$$

The two-dimensional correlation function of the channel frequency response becomes [6]

$$\begin{aligned} \Psi_H(\Delta t, \Delta f) &= E\{H(t+\Delta t, f+\Delta f) H^*(t,f)\} \\ &= \psi_t(\Delta t)\psi_f(\Delta f), \end{aligned} \tag{6.53}$$

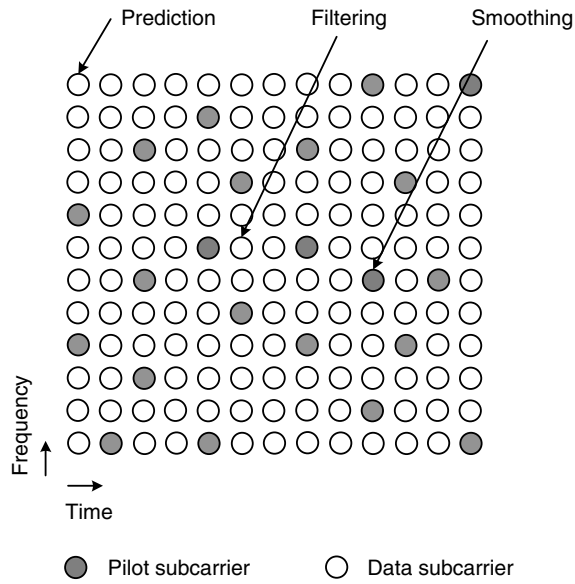


Figure 6.9. Arbitrary sampling grid used in a two-dimensional MMSE interpolator for smoothing, filtering and prediction

with

$$\psi_f(\Delta f) = \sum_r \sigma_r^2 e^{-j2\pi\Delta f \tau_r}. \tag{6.54}$$

and

$$\psi_t(\Delta t) = J_0(2\pi f_D \Delta t), \tag{6.55}$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind and f_D is the Doppler frequency.

From Equation (6.53), it can be clear that the two-dimensional correlation function of the channel frequency response $\Psi_H(\Delta t, \Delta f)$ is separable and $\psi_t(\Delta t)$ depends on the Doppler frequency while $\psi_f(\Delta f)$ is related to the channel delay spread. This separability property will be very important for the channel estimator design in OFDM systems with grid-type pilot configuration.

Two-Dimensional MMSE Interpolation

The time-varying channel frequency response is a wide-sense stationary two-dimensional random process. Since the process is observed under noise, it is impossible to have a perfect reconstruction. The two-dimensional (2D) MMSE interpolator is the optimum linear filter/smoother/predictor in terms of the minimum mean-squared error criterion. For the grid shape as shown in Figure 6.9, the 2D MMSE interpolator can smooth the noise in the channel estimates for the pilot subcarrier; filter estimates on pilots; generate the channel response for

Petitioner Samsung Ex-1026, 0117

data subcarriers inside the pilot grid; and predict the channel response for data subcarriers outside the pilot grid [16].

Let $H_{i,k}$ denote the channel response at the k-th subcarrier of the i-th symbol. Further denote the time–frequency indices of the p-th pilot subcarrier (i_p, k_p). Define the actual channel frequency response at pilot subcarriers as

$$\mathbf{h}_p = [H_{i_0,k_0} \; H_{i_1,k_1} \; \cdots H_{i_{M_p-1},k_{M_p-1}}]^T. \tag{6.56}$$

At the pilot subcarriers, the channel estimates can be represented as

$$\tilde{\mathbf{h}}_p = \mathbf{h}_p + \mathbf{u}, \tag{6.57}$$

where

$$\tilde{\mathbf{h}}_p = [\tilde{H}_{i_0,k_0} \; \tilde{H}_{i_1,k_1} \; \cdots \; \tilde{H}_{i_{M_p-1},k_{M_p-1}}]^T, \tag{6.58}$$

and $\mathbf{u}$ is the measurement noise.

The M_p-tap 2D MMSE interpolator tries to estimate the channel response for the k-th subcarrier of the i-th symbol by the following equation

$$\hat{H}_{i,k} = \sum_{p=0}^{M_p-1} W(i,k,i_p,k_p)\tilde{H}_{i_p,k_p}. \tag{6.59}$$

The coefficients of the 2D MMSE interpolator, $W(i,k,i_p,k_p)$, minimize the mean-squared error $E\{|H_{i,k} - \hat{H}_{i,k}|^2\}$ for all i and k. If the above equations for all i and k are stacked, the vector equation is given by

$$\hat{\mathbf{h}}_{\text{MMSE}} = \mathbf{W}_{\text{MMSE}}\tilde{\mathbf{h}}_p. \tag{6.60}$$

The solution of the 2D MMSE interpolator, $\mathbf{W}_{\text{MMSE}}$, takes the form of

$$\mathbf{W}_{\text{MMSE}} = \mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_p}\left(\mathbf{R}_{\mathbf{hh}} + \sigma_u^2\mathbf{I}_{M_p}\right)^{-1}, \tag{6.61}$$

where

$$\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_p} = E\{\mathbf{h}\tilde{\mathbf{h}}_p^H\}, \tag{6.62}$$

$$\mathbf{R}_{\mathbf{hh}} = E\{\mathbf{h}\mathbf{h}^H\}, \tag{6.63}$$

σ_u^2 is the noise variance of $\mathbf{u}$ and $\mathbf{I}_{M_p}$ is the $M_p \times M_p$ identity matrix.

Although the 2D MMSE interpolator provides the optimal channel estimates, it has high complexity. According to the separability property, the correlation functions of the channel frequency responses along the time axis and the frequency axis are independent. The performance of two one-dimensional MMSE interpolators is compared with one two-dimensional MMSE interpolator in [16] and the simulation results show that they are quite similar. As such, channel estimation based on grid-type pilot subcarrier responses is often treated as two one-dimensional problems.

Petitioner Samsung Ex-1026, 0118

Double One-Dimensional Interpolations

Even though channel estimation using two one-dimensional MMSE interpolators can approach the performance of the two-dimensional MMSE interpolator, high complexity still prevents the use of MMSE interpolators in practical OFDM receivers. Hence, simpler interpolators, such as polynomial interpolation or spline interpolation, are often used. In OFDM systems with grid-type pilot patterns, applying two one-dimensional simple interpolators is often the case [21–25].

In [22], a DFT-based approach, a low-pass filter and a least-squares filter are compared for interpolating channel frequency responses in the time direction. The DFT-based approach first transforms the channel estimates at pilot subcarriers from several OFDM symbols into the transform domain which corresponds to the Doppler spectrum. By appropriate zero-padding in the transform domain and inverse transformation, the up-sampled channel frequency estimates along the time axis are obtained. The low-pass filter has a flat central region but slow roll-off in the transform domain. In contrast, the least-squares filter, generally designed by least-squares fitting through a polynomial of a given degree, has a rippled central region but a steep slope. It is shown that the DFT-based approach is favored only for dense pilots. The least-squares filter outperforms the low-pass filter in cases with high mobility.

In [23], frequency estimation for digital video broadcasting-terrestrial (DVB-T) system, which features a parallelogram-shaped pilot pattern, is investigated. Interpolation for channel frequency responses along the time axis uses either sample-and-hold, linear interpolation or *sinc* interpolation, while linear interpolation, *sinc* interpolation or regularized least-squares interpolation is studied for the frequency-axis interpolation. Clearly, in time-varying channels, the sample-and-hold approach has poor performance. However, given the pilot density and the possible Doppler frequency in the DVB-T systems, the linear interpolator has similar performance as the *sinc* interpolator for time-axis interpolation. In the frequency direction, the *sinc* interpolator is slightly better than the linear interpolator and the regularized least-squares interpolator.

Two-Dimensional Regression

Figure 6.10 shows the magnitude of a sample channel frequency response along the time and frequency axes under mobility of 120 km/hr and 2.14-μs maximum excess delay. The OFDM subcarrier spacing is 5.625 KHz and 1024 subcarriers. The sampling frequency is 5.76 MHz. Smooth variation of this response appears along the time axis and the frequency axis. With pilot subcarrier sampling the channel frequency response, a two-dimensional regression algorithm is proposed in [26] to fit the time-varying channel frequency response of all data subcarriers by a quadrature surface

$$f(x,y) = ax^2 + bxy + cy^2 + dx + ey + f. \tag{6.64}$$

The coefficients (a,b,c,d,e,f) are determined so as to minimize the error between the quadrature surface and the channel frequency response at pilot subcarriers:

$$\mathcal{E} = \sum_{p=0}^{M_p-1} |\tilde{H}_{i_p,k_p} - f(i_p,k_p)|^2 = \sum_{p=0}^{M_p-1} |\tilde{H}_{i_p,k_p} - \mathbf{p}^T \mathbf{s}_{i_p,k_p}|^2, \tag{6.65}$$

Petitioner Samsung Ex-1026, 0119

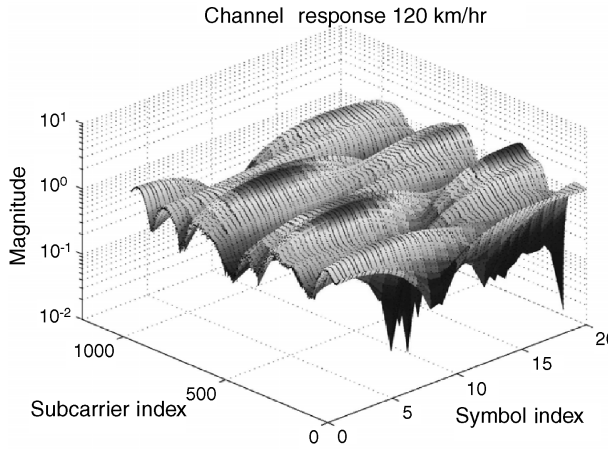


Figure 6.10. Time-varying channel frequency response

where

$$\mathbf{p} = [a \;\; b \;\; c \;\; d \;\; e \;\; f]^T, \tag{6.66}$$

and

$$\mathbf{s}_{i_p,k_p} = [i_p^2 \;\; i_p k_p \;\; k_p^2 \;\; i_p \;\; k_p \;\; 1]^T. \tag{6.67}$$

The optimal coefficient $\mathbf{p}$ can be obtained by setting the partial derivatives of the error with respect to all coefficients to be zero ($\frac{\partial \mathcal{E}}{\partial \mathbf{p}} = 0$) so that

$$\sum_{p=0}^{M_p-1} \tilde{H}^*_{i_p,k_p} \mathbf{s}_{i_p,k_p} = \sum_{p=0}^{M_p-1} \mathbf{s}_{i_p,k_p} \mathbf{s}^T_{i_p,k_p} \mathbf{p}^*. \tag{6.68}$$

Therefore, the optimal coefficients are given by

$$\mathbf{p} = \left(\sum_{p=0}^{M_p-1} \mathbf{s}_{i_p,k_p} \mathbf{s}^T_{i_p,k_p} \right)^{-1} \cdot \left(\sum_{p=0}^{M_p-1} \tilde{H}_{i_p,k_p} \mathbf{s}_{i_p,k_p} \right), \tag{6.69}$$

and the estimated channel responses are

$$\hat{H}_{i,k} = \mathbf{p}^T \mathbf{s}_{i,k}. \tag{6.70}$$

With the grid-type pilot pattern, the matrix inversion in the 2D regression algorithm is fixed and can be pre-computed. The estimation error mainly depends on the density of pilot

Petitioner Samsung Ex-1026, 0120

subcarriers. If the channel frequency response suffers deep fading in the time axis or frequency axis within the observed time–frequency block, the 2D regression cannot generate a proper fit.

Discussions

As usual, the 2D MMSE interpolator achieves optimal estimation in terms of minimum mean squared estimation errors at the cost of huge complexity. Although the double one-dimensional interpolation approach is much simpler, it somewhat sacrifices the system performance. What is worse, interpolation in the time direction requires a lot of temporary storage to buffer the data. It is noteworthy to know that the 2D MMSE interpolator can be applied to arbitrary pilot patterns and the mean-squared error of its channel estimates only depends on the pilot density, but not the pilot pattern. However, the same cannot be said about the double one-dimensional interpolation method because its performance is affected by both the pilot pattern and the pilot density [25]. The 2D-regression algorithm can be adopted for arbitrary pilot patterns, too. Its performance is also closely related to the pilot density.

6.4 Adaptive Channel Estimation

In order to find the optimal coefficients of the 2D MMSE interpolator, the receiver has to possess precise information about the channel statistics. In practice, however, this information is not readily available. Instead of using the *a priori* information of channel statistics explicitly, there are several possible approaches, known as the *adaptive filters*, that can follow the channel statistics and self-adjust their coefficients to track the channel characteristics.

In [27,28], two-dimensional adaptive channel estimation algorithms, namely normalized least-mean-square (NLMS) and recursive least-squares (RLS) adaptive filters, are proposed to make full use of the correlation of doubly-selective (time-selective and frequency-selective) fading channels. Both the adaptive filters involve three basic iterative steps: filtering to perform channel estimation; computing estimation error; updating the coefficients of the filter [27,28]. Even though both adaptive algorithms can track channel variation, they still require a few symbols to converge. Thus, they are only suitable for those cases in which the normalized Doppler frequency (with respect to the symbol rate) is below 0.01.

In addition to NLMS and RLS algorithms, a Kalman-filtering approach can also track the dynamics of time-varying channels [29]. This approach also provides a recursive solution in terms of state-space concepts to the minimum mean-squared estimation error problem of nonstationary signals. The state is the minimal requirement of the data that uniquely describe the dynamic behavior of the system. Each updated estimate of the state is calculated from the previous estimate and the current observation. Thus, the storage of the entire past observed data is unnecessary.

In summary, the RLS algorithm has fast rate of convergence compared with the NLMS algorithm, while the NLMS is simple and can be commonly adopted in hardware implementation. Both the RLS algorithm and the Kalman-filtering approach are computation intensive, since they need to explicitly evaluate the correlation of the input or the error signals. However, the RLS algorithm is a special case of the Kalman filtering approach for a stationary state-space model. Hence, the Kalman filtering approach can have better tracking properties in fast-fading channels with the dynamic state-space model.

Petitioner Samsung Ex-1026, 0121

6.5 Equalization

6.5.1 One-Tap Equalizer

OFDM systems are favored over single-carrier modulations in that the simple one-tap *frequency-domain equalizer* (FDE) can equalize OFDM signals that go through frequency-selective fading channels. In channels whose impulse responses remain constant within one OFDM symbol period, the received signal at each subcarrier takes the form of

$$Z_{i,k} = H_{i,k}X_{i,k} + V_{i,k}. \tag{6.71}$$

One-tap equalizers restore the transmitted signal by

$$\hat{X}_{i,k} = G_{i,k}Z_{i,k}, \tag{6.72}$$

where $G_{i,k}$ is the equalizer coefficient at the k-th subcarrier during the i-th symbol.

Regardless of noise, the zero-forcing equalizer simply uses the inverse of the channel response ($G_{i,k} = H_{i,k}^{-1}$) and forces the frequency-selective-faded signals back to flat faded ones. However, it may result in noise enhancement in the subcarriers that suffer deep fading.

The MMSE equalizer, which tries to minimize $E\{|\hat{X}_{i,k} - X_{i,k}|^2\}$, takes the noise component into account and equalizes the signal by

$$G_{i,k} = \frac{H_{i,k}^*}{|H_{i,k}|^2 + \dfrac{1}{SNR}}. \tag{6.73}$$

This equalizer has the advantage that the noise enhancement problem in low-SNR cases is gone. Also when SNR is high enough, it is clear that the MMSE equalizer approaches the zero-forcing equalizer.

As in the channel estimation problems, adaptive algorithms can adjust the equalizer coefficients to minimize $E\{|\hat{X}_{i,k} - X_{i,k}|^2\}$ without prior channel information. The equalized signal $\hat{X}_{i,k}$ is compared with a reference signal to obtain an error signal, $e_{i,k}$. Then, the equalizer coefficients are adjusted according to the error signal:

$$G_{(i+1),k} = G_{i,k} - g_{i,k}e_{i,k}, \tag{6.74}$$

where $g_{i,k}$ is the gain factor, depending on the NLMS/RLS algorithm. The reference signal can be known pilot data in the training mode or the slicer output in the decision-directed mode.

6.5.2 Multiple-Tap Equalizer

In fast-fading channels, channel response not only changes from the previous symbol to the current symbol, but also varies within one symbol period. The fast channel variation within one symbol period brings about inter-carrier interference, which further deteriorates the system performance. One-tap equalizers cannot cope with such situations. As such,

Petitioner Samsung Ex-1026, 0122

multiple-tap equalizers which may cancel inter-carrier interference from adjacent subcarriers are required.

Assume that the time-varying discrete channel impulse response at the n-th sample in the i-th symbol of finite length R, with $R \leq N_g$ given by $[q_0(i,n) \;\; q_1(i,n) \;\; \cdots \;\; q_{R-1}(i,n)]$. The received signal after convolving the transmitted signal with the channel impulse response becomes

$$z_{i,n} = \sum_{r=0}^{R-1} q_r(i,n) x_{i,[n-r]_N} + v_{i,n}, \tag{6.75}$$

where $[\cdot]_N$ denotes the modulo-N operation, and $v_{i,n}$ is Gaussian noise. Then, the frequency-domain data can be obtained by taking the discrete Fourier transform of $z_{i,n}$:

$$\begin{aligned} Z_{i,k} &= \sum_{n=0}^{N-1} z_{i,n} e^{-j2\pi nk/N} \\ &= \sum_{n=0}^{N-1} \left(\sum_{r=0}^{R-1} q_r(i,n) \cdot \left(\frac{1}{N} \sum_{k'=-N/2+1}^{N/2} X_{i,k'} e^{j2\pi(n-r)k'/N} \right) \right) e^{-j2\pi nk/N} + V_{i,k} \\ &= \sum_{k'=-N/2+1}^{N/2} X_{i,k'} \cdot \left(\sum_{r=0}^{R-1} \left(\frac{1}{N} \sum_{n=0}^{N-1} q_r(i,n) e^{-j2\pi n(k-k')/N} \right) e^{-j2\pi rk'/N} \right) + V_{i,k} \\ &= H_{i,k,k} X_{i,k} + \sum_{k' \neq k} H_{i,k,k'} X_{i,k'} + V_{i,k}, \end{aligned} \tag{6.76}$$

where $V_{i,k}$ is the equivalent frequency-domain noise term of $v_{i,n}$, and $H_{i,k,k'}$ represents the interference gain from subcarrier k' to the subcarrier k due to the fast-fading channel. The first term in Equation (6.76) contains the desired signal and the second term is the inter-carrier interferences from other subcarriers. Note that

$$H_{i,k,k'} = \sum_{r=0}^{R-1} \left(\frac{1}{N} \sum_{n=0}^{N-1} q_r(i,n) e^{-j2\pi n(k-k')/N} \right) e^{-j2\pi rk'/N}. \tag{6.77}$$

If $k' = k$, then

$$H_{i,k,k} = \sum_{r=0}^{R-1} \left(\frac{1}{N} \sum_{n=0}^{N-1} q_r(i,n) \right) e^{-j2\pi rk/N}, \tag{6.78}$$

which can be interpreted as the transformation of the time-averaged channel impulse response. Moreover, if the channel is time-invariant over one symbol period, then $H_{i,k,k}$ is equivalent to $H_{i,k}$, as in the case of stationary channels, and $H_{i,k,k'} = 0$ for $k \neq k'$. It means the inter-carrier interference terms all vanish and the orthogonality among subcarriers is preserved. Otherwise, $H_{i,k,k'}$ must be processed in order to get rid of the inter-carrier interferences.

Petitioner Samsung Ex-1026, 0123

In [30], it is assumed that variation of each channel path gain is linear within one OFDM symbol when the normalized Doppler frequency with respect to one subcarrier spacing is smaller than 0.1. The time domain channel impulse response is first obtained by pilot symbols. Then, time-varying channel impulse response at each sample is calculated by linear interpolation, and $H_{i,k,k'}$ can be calculated according to Equation (6.77). Under the assumption of linear variation, only s adjacent subcarriers contribute significant interferences, namely $H_{i,k,k'} = 0$ for $|k' - k| > s/2$. Given the estimation of all interference gains, the signal can be restored and ICI component cancelled by

$$\hat{X}_{i,k} = \sum_{|k'-k| \leq s/2} \mathbf{G}_{i,k,k'} Z_{i,k'}, \tag{6.79}$$

where $\mathbf{G}_{i,k,k'}$ can be obtained from all $\mathbf{H}_{i,k,k'}$ through matrix inversion.

Bibliography

[1] M. J. Fernandez-Getino Garcia, S. Zazo and J. M. Paez-Borrallo, 'Pilot pattern for channel estimation in OFDM', *IEE Electronics Letters*, vol. 36, no. 12, Jun. 2000, pp. 1049–1050.

[2] T. Mueller, K. Brueninghaus and H. Rohling, 'Performance of coherent OFDM-CDMA for broadband mobile communications', *Kluwer Wireless Personal Commun.*, vol. 2, no. 4, 1996, pp. 295–305.

[3] ETSI, 'Digital Video Broadcasting (DVB): Implementation guidelines for DVB terrestrial services; transmission aspects', ETSI Technical Report TR 100.190 v1.1.1, Dec. 1997.

[4] S. Coleri, M. Ergen, A. Puri and A. Bahai, 'Channel estimation techniques based on pilot arrangement in OFDM systems', *IEEE Trans. Broadcast.*, vol. 48, no. 3, Sep. 2002, pp. 223–229.

[5] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Dignal Processing*. Eaglewood Cliffs, NJ: Prentice Hall, 1984.

[6] Y. Li, L. J. Cimini and N. R. Sollenberger, 'Robust channel estimation for OFDM systems with rapid dispersive fading channels', *IEEE Trans. Commun.*, vol. 46, no. 7, Jul. 1998, pp. 902–915.

[7] J. -J van de Beek, O. Edfors, M. Sandell, S. K. Wilson and P. O. Borjesson, 'On channel estimation in OFDM systems', in *Proc. of IEEE Vehicular Technology Conference*, vol. 2, Jul. 1995, pp. 815–819.

[8] O. Edfors, M. Sandell, J.-J van de Beek, S. K. Wilson and P. O. Borjesson, 'OFDM channel estimation by singular value decomposition', *IEEE Trans. Commmun.*, vol. 46, no. 7, Jul. 1998, pp. 931–939.

[9] L. Deneire, P. Vandenameele, L. van der Perre, B. Guselinckx and M. Engels, 'A low-complexity ML channel estimator for OFDM', *IEEE Trans. Commun.*, vol. 51, no. 2, Feb. 2003, pp. 135–140.

[10] M. Morelli and U. Mengali, 'A comparion of pilot-aided channel estimation methods for OFDM systems', *IEEE Trans. Signal Processing*, vol. 49, no. 12, Dec. 2001.

[11] M. H. Hsieh and C. H. Wei, 'Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels', *IEEE Trans. Consumer Electron.*, vol. 44, no. 1, Feb. 1998, pp. 217–225.

[12] B. Yang, Z. Cao and K. B. Letaief, 'Analysis of low-complexity windowed DFT-based MMSE channel estimator for OFDM systems', *IEEE Trans. Commun.*, vol. 49, no. 11, Nov. 2001, pp. 1977–1987.

[13] Y. Zhao and A. Huang, 'A novel channel estimation method for OFDM mobile communication systems based on pilot signals and transform domain processing', in *Proc. of IEEE Vehicular Technology Conference*, vol. 3, May 1997, pp. 2089–2093.

[14] H. Minn and V. K. Bhargava, 'An investigation into time-domain approach for OFDM channel estimation', *IEEE Trans. Broadcast.*, vol. 46, Dec. 2000, pp. 240–248.

[15] M. J. Fernandez-Getino Garcia, J. M. Paez-Borrallo and S. Zazo, 'DFT-based channel estimation in 2D-pilot-symbol-aided OFDM wireless systems', in *Proc. of IEEE Vehicular Technology Conference*, vol. 2 , 2001, pp. 810–814.

[16] P. Hoeher, S. Kaiser and P. Robertson, 'Two dimensional pilot-symbol-aided channel estimation by Wiener filtering', in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1997, pp. 1845–1848.

Petitioner Samsung Ex-1026, 0124

[17] J. Rinne and M. Renfors, 'Pilot spacing in orthogonal frequency division multiplexing systems on practical channels', *IEEE Trans. Consumer Electron.*, vol. 42, no. 4, Nov. 1996, pp. 959–962.
[18] L. Erup, F. M. Gardner and R. A. Harris, 'Interpolation in digital modems – part II: Implementation and performance', *IEEE Trans. Commun.*, vol. 41, no. 6, Jun. 1993, pp. 998–1008.
[19] H. Tang, K. Y. Lau and R. W. Brodersen, 'Interpolation-based maximum likelihood channel estimation using OFDM pilot symbols', in *Proc. of IEEE Global Telecommunications Conference*, vol. 2, 2002, pp. 1860–1864.
[20] P. Y. Tsai and T. D. Chiueh, 'Frequency-domain interpolation-based channel estimation in pilot-aided OFDM systems', in *Proc. of IEEE Vehicular Technology Conference*, vol. 1, May 2004, pp. 420–424.
[21] C. E. Toso, P. Combelles, J. Golbrun, *et al.*, '0.5-μm CMOS circuits for demodulation and decoding of an OFDM-based digital TV signal conforming to the European DVB-T standard', *IEEE J. Solid State Circuits*, vol. 33, no. 11, Nov. 1998, pp. 1781–1792.
[22] A. A. Hutter, R. Hasholzner and J. S. Hammerschmidt, 'Channel estimation for mobile OFDM systems', in *Proc. of IEEE Vehicular Technology Conference*, vol. 1, Sep. 1999, pp. 305–309.
[23] S. H. Chen, W. H. He, H. S. Chen and Y. Lee, 'Mode detection, Synchronization and channel estimation for DVB-T OFDM', in *Proc. of IEEE Global Telecommunications Conference*, vol. 5, pp. 2416–2420.
[24] J. W. Choi and Y. H. Lee, 'Design of 2-D channel estimation filters for OFDM systems', in *Proc. of IEEE 2005 International Conference on Communication*, vol. 4, May 2005, pp. 2568–2572.
[25] J. W. Choi and Y. H. Lee, 'Optimal pilot pattern for channel estimation in OFDM systems', *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, Sept. 2005, pp. 2083–2088.
[26] M. X. Chang and Y. T. Su, '2D regression channel estimation for equalizing OFDM signals', in *Proc. of IEEE Vehicular Technology Conference*, vol. 1, May 2000, pp. 240–244.
[27] X. Hou, S. Li, D. Li, C. Yin and G. Yue, 'On two-dimensional adaptive channel estimation in OFDM systems', in *Proc. of IEEE Vehicular Technology Conference*, vol. 1, Sep. 2004, pp. 498–502.
[28] X. Hou, S. Li, C. Yin and G. Yue, 'Two-dimensional recursive least square adaptive channel estimation for OFDM systems', in *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing*, vol. 1, Sep. 2005, pp. 232–236.
[29] W. Chen and R. Zhang, 'Kalman-filter channel estimator for OFDM systems in time and frequency-selective fading environment', in *Proc. of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, May 2004, pp. 377–380.
[30] W. G. Jeon, K. H. Chang and Y. S. Cho, 'An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels', *IEEE Trans. Commun.*, vol. 47, no. 1, Jan. 1999, pp. 27–32.

Petitioner Samsung Ex-1026, 0125

7

MIMO Techniques

Multiple antenna configuration at the transmitter and/or the receiver can greatly enhance the capacity and performance of wireless communication systems. The *multiple-input multiple-output* (MIMO) technology, with multiple antennas at the transmitter and/or the receiver, has been integrated into many OFDM systems.

7.1 Introduction

According to the number of transmit (TX) antennas and the number of receive (RX) antennas, wireless systems can be classified as single-input single-output (SISO), single-input multiple-output (SIMO), multiple-input single-output (MISO) and multiple-input multiple-output (MIMO) systems, in which the input and output are with respect to the channel between the transmitter and the receiver, as shown in Figure 7.1. The advantages of employing multiple antennas and related signal processing include [1,2]:

- **Array gain** As multiple copies of the signals are received at a receiver with more than one antenna, the signals can be combined coherently to achieve gain in effective SNR. Such gain is usually called *array gain*. Combining methods such as equal-gain combining (EGC) and maximal ratio combining (MRC) are very popular [3]. In an SIMO system, the average SNR increase is proportional to the number of receive antennas. In the case of multiple transmit antennas, however, array gain can also be obtained, provided that spatial pre-coding based on the channel information is implemented. With this pre-coding, the multiple copies of transmitted signals supposedly will arrive at the single receiving antenna coherently.
- **Diversity gain** In a SISO system without signal redundancy, deeply faded signals are beyond detection. On the other hand, in wireless systems with multiple TX antennas and/or multiple RX antennas, signals can be transmitted/received with diversity so as to combat channel fading. Receive diversity in MIMO systems refers to the combination of independently faded signals from different receive antennas so that the processed signal suffers less fading than that of the receiver with only one antenna. Similarly, using coding that transmits redundant information from multiple TX antennas, transmit diversity can be achieved. Well known spatial coding techniques include space–time trellis codes (STTC) [4], space–time block codes (STBC) [5,6], space–frequency block codes (SFBC), and space–time–frequency block codes (STFBC) [7].

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0126

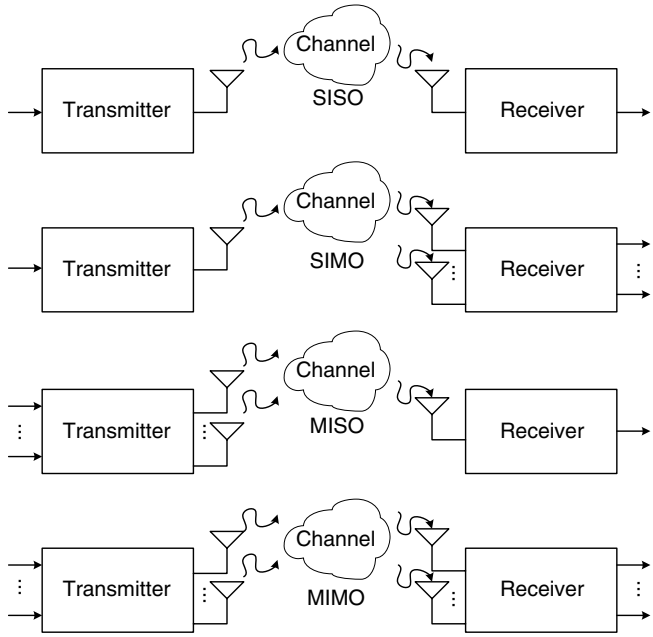


Figure 7.1. Illustration of transmitters/receivers with different antenna configurations

- **Capacity gain** MIMO technology brings one very important enhancement to wireless communications: gain in transmission rate. By multiplexing the transmitted data streams among different antennas, namely *spatial multiplexing*, an increase in data rate can be attained. This rate increase is proportional to the minimum of the number of TX antennas and the number of RX antennas. If either the transmitter or the receiver has a single antenna, then there exists no obvious capacity gain. Hence, spatial multiplexing is mainly applied to MIMO systems. Note that several data streams are simultaneously transmitted over the air and received at the RX antennas. At the receiver, these signals need to be processed to recover the information contained in the individual data streams.
- **Beamforming** When combining the received signals from multiple antennas, it is possible to create strong differentiation in gains for signals that arrive from different angles. The beam-forming technique [8] has traditionally been applied in the transmitter or the receiver to control the directionality of the transmit/receive antenna pattern. With proper knowledge of the channel and accordingly setting the combining coefficients, a beam-forming receiver can increase the antenna gain along the direction of the intended transmitter while at the same time suppressing the interferences from other directions.

Lured by these advantages, researchers have intensively studied the use of multiple antennas and, thus, this field has grown rapidly, both in theory and in implementation during the past few years [9]. The enhancement in spectral efficiency has resulted in the adoption of MIMO technology in several wireless standards. In this chapter, a brief introduction of channel capacity under different transmit/receive antenna configurations will be given first. Then, diversity gain attained by special signal arrangements in the spatial domain will be

Petitioner Samsung Ex-1026, 0127

illustrated. Incorporation of the MIMO techniques into wireless OFDM systems will then be discussed. Moreover, this chapter will also cover how SISO OFDM transceivers should be extended to enjoy MIMO advantages, especially in terms of synchronization and channel estimation. Finally, several MIMO encoding and detection schemes will be presented.

7.2 MIMO Basics

7.2.1 Capacity

A fundamental theory of communication channel capacity (C) was first proposed by Claude Shannon for additive white Gaussian channels in 1948. This theorem stipulates that the maximum error-free data rate that a channel can convey is given by

$$C = \log_2(1+\rho) \ \text{ bps/Hz}, \tag{7.1}$$

where ρ is the signal-to-noise ratio (SNR). Since then, Shannon bound represents an upper limit in spectral efficiency.

In stochastic channels, the ergodic channel capacity is used instead. This capacity uses the ensemble average capacity over distribution of channel gains. In the SISO system with a random complex channel gain, h, the ergodic channel capacity takes the form of [10]

$$C = E\{\log_2(1+\rho|h|^2)\} \ \text{ bps/Hz}. \tag{7.2}$$

Without loss of generality, the capacity of MIMO channels with P transmit antennas and Q receive antennas is first considered. Assume that the channel response from the p-th transmit antenna to the q-th receive antenna is denoted by $h^{(q,p)}$. The channel matrix is then given by

$$\mathbf{H} = \begin{bmatrix} h^{(0,0)} & \cdots & h^{(0,P-1)} \\ & \ddots & \\ h^{(Q-1,0)} & \cdots & h^{(Q-1,P-1)} \end{bmatrix}. \tag{7.3}$$

For the case without prior channel state information at the transmitter but with perfect channel knowledge at the receiver, the ergodic capacity of the MIMO channel takes the form of [1]

$$C = E\left\{\log_2\left[\det\left(\mathbf{I}_{\mathrm{Q}} + \frac{\rho}{\mathrm{P}}\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)\right]\right\} \ \text{ bps/Hz}, \tag{7.4}$$

where $\mathbf{I}_Q$ is the $Q \times Q$ identity matrix. By using eigen-decomposition, $\mathbf{H}\mathbf{H}^H$ can be written as

$$\mathbf{H}\mathbf{H}^H = \mathbf{U}\Lambda\mathbf{U}^H, \tag{7.5}$$

where $\mathbf{U}$ is a unitary matrix containing eigenvectors and the diagonal matrix Λ has R non-zero eigenvalues λ_r for $0 \le r < R$. After some derivation, Equation (7.4) can be formulated as [1]

$$C = E\left\{\sum_{r=0}^{R-1}\log_2\left(1+\frac{\rho}{P}\lambda_r\right)\right\} \ \text{ bps/Hz}. \tag{7.6}$$

Petitioner Samsung Ex-1026, 0128

Clearly, the above equation implies that the capacity of a MIMO channel is the sum of R parallel SISO sub-channels.

For a SIMO channel, namely $P = 1$, the channel vector is denoted by $\mathbf{h}_{\text{SIMO}} = [h^{(0,0)}\ h^{(1,0)}\ \cdots\ h^{(Q-1,0)}]^T$. After eigen-decomposition, there is only one non-zero eigenvalue λ_0, which is equal to $||\mathbf{h}_{\text{SIMO}}||^2$, where $||\cdot||$ is the L_2 norm of the vector that is equal to the square root of the sum of the squared elements in the vector. The ergodic capacity of a SIMO channel becomes [1]

$$C = E\{\log_2(1 + \rho||\mathbf{h}_{\text{SIMO}}||^2)\} \quad \text{bps/Hz}. \tag{7.7}$$

From the above equation, adding more receive antennas increases the $||\mathbf{h}_{\text{SIMO}}||^2$ linearly and the ergodic capacity logarithmically. On the other hand, consider the ergodic capacity of a MISO channel with $\mathbf{h}_{\text{MISO}} = [h^{(0,0)}\ h^{(0,1)}\ \cdots\ h^{(0,P-1)}]$. Similarly, with one eigenvalue $\lambda_0 = ||\mathbf{h}_{\text{MISO}}||^2$, the ergodic capacity in this case is given by [1]

$$C = E\left\{\log_2\left(1 + \frac{\rho}{P}||\mathbf{h}_{\text{MISO}}||^2\right)\right\} \quad \text{bps/Hz}. \tag{7.8}$$

No capacity gain is achieved, even with more transmit antennas, as, on average, the increase in $||\mathbf{h}_{\text{SIMO}}||^2$ and P terms in the equation tend to cancel each other out.

When the channel is known to the transmitter, the maximum MIMO capacity can be achieved via the water-pouring optimization [1]. Assume that γ_r represents the optimal transmit energy for the r-th SISO sub-channel from the water-pouring algorithm and $\sum_{r=0}^{R-1} \gamma_r = P$. The ergodic capacity is then given by [1]

$$C = E\left\{\sum_{r=0}^{R-1} \log_2\left(1 + \gamma_r \frac{\rho}{P}\lambda_r\right)\right\} \quad \text{bps/Hz}. \tag{7.9}$$

In summary, the ergodic capacity scales linearly with the number of antennas for the spatially uncorrelated MIMO channels. In addition, the ergodic SIMO capacity is slightly greater than the ergodic MISO capacity, even if the number of receive antennas in a SIMO system is equal to the number of transmit antennas in a MISO system. It is because in the MISO system, the transmitter, lacking the channel state information, is incapable of exploiting transmit array gain. Furthermore, from Equations (7.6) and (7.9), the ergodic capacity of the MIMO channels will decrease as the rank R of the channel matrix is reduced, which happens when the rows or the columns of $\mathbf{H}$ get more and more correlated.

7.2.2 Diversity

Diversity refers to the schemes that introduce more than one reception of the signals related to a single piece of information. Different receptions can be contributed by signals transmitted from different antennas, over different frequencies, at different time slots/symbols, or along different electromagnetic (EM) wave polarizations. Usually, each signal suffers different degrees of fading so that the transmitted information is unresolved only

Petitioner Samsung Ex-1026, 0129

when all the different copies of the signals are severely faded. Consequently, diversity has been adopted widely to combat fading and co-channel interference. There are four types of diversity techniques in wireless communications:

- **Time diversity** Information is transmitted redundantly in the temporal domain. Time diversity is attained when the redundancy information is separated in the time domain by longer than the channel coherence time.
- **Frequency diversity** Similar to the time diversity, the frequency diversity technique transmits redundant signals about a piece of information on carriers that are spaced by frequency separations wider than the coherent bandwidth of the channel.
- **Space diversity** Multiple antennas separated in the spatial domain are used to offer space diversity provided that rich-scattering environments create uncorrelated multipath fading in different spatial streams. In downlink, transmit diversity is preferred due to the size limitation of mobile units, while receive diversity is much more feasible in uplink.
- **Polarization diversity** EM waves with different polarizations have different propagation, reflection and scattering characteristics. Transmitting and receiving multiple copies of the signal using antennas with different polarizations also helps to enhance the performance of a wireless link.

It is intuitive that with diversity, the error probability scales exponentially with the order of diversity, assuming that different versions of the signal suffer independent fading. Hence, the diversity gain (G_d) can be found as the ratio between the slopes of probability curves at the high-SNR region for the diversity reception case and the no-diversity reception case, when both error probability curves are plotted as a function of SNR on a log–log scale [4]. In other words, for a wireless link with diversity gain G_d, the error probability at the high-SNR region satisfies

$$\log\left(\frac{P_e(\rho_1)}{P_e(\rho_2)}\right) \propto (-G_d)\log\left(\frac{\rho_1}{\rho_2}\right), \tag{7.10}$$

where ρ_1 and ρ_2 are SNRs for two channel conditions. Figure 7.2 depicts error probability curves for several different cases with and without diversity gain.

7.3 MIMO–OFDM

OFDM has the advantage of converting a wideband frequency-selective fading channel into numerous narrow-band flat-fading sub-channels. Hence, channel fading can be mitigated by a simple one-tap equalizer. On the other hand, MIMO techniques bring about a significant performance boost for wireless communications under flat-fading channels. As a result, by combining MIMO and OFDM, new high-speed wireless communication systems enjoy the benefits from both technologies.

Figure 7.3 depicts a typical MIMO–OFDM system architecture with P transmit antennas and Q receive antennas. User data are first encoded by a channel encoder block, which may consist of one or more channel encoders. Afterwards, the MIMO encoder parses the data from the channel encoder block into several spatial streams. Recall that the signals transmitted by the OFDM transmitters can be represented on a symbol–subcarrier grid in

Petitioner Samsung Ex-1026, 0130

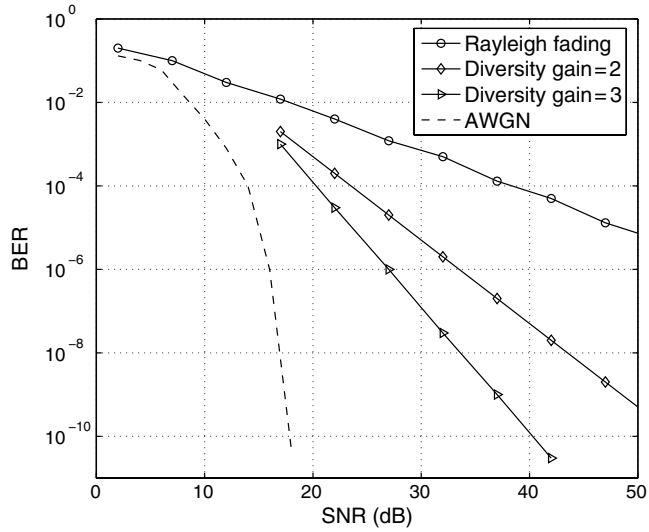


Figure 7.2. Performance improvement with diversity gain

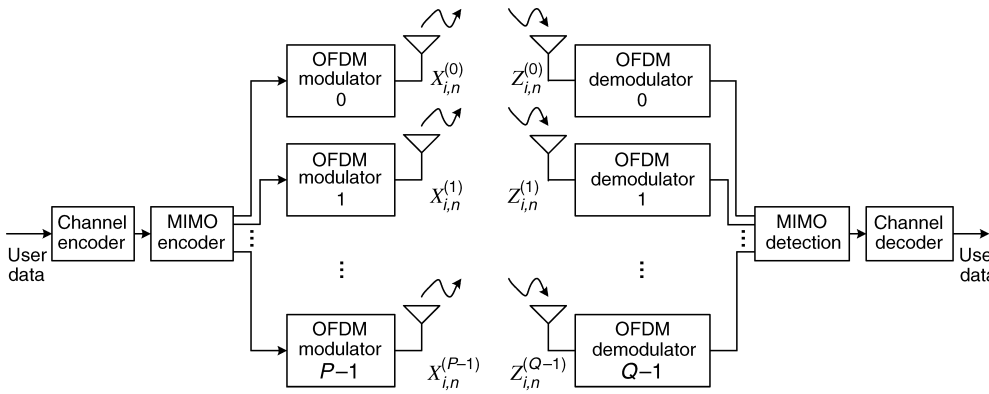


Figure 7.3. A MIMO–OFDM system with P transmit antennas and Q receive antennas

the time–frequency plane. Now, in the MIMO–OFDM systems, a third spatial dimension is introduced, and, thus, the MIMO encoder output signals can be formatted in either space–time blocks, space–frequency blocks or space–time–frequency blocks. Signals at different symbols and subcarriers but with the same space coordinate are fed into a single OFDM modulator, connected to its own transmit antenna.

Let $x_{i,n}^{(p)}$ and $z_{i,n}^{(q)}$ denote the n-th baseband sample of the i-th symbol at the p-th transmit antenna and at the q-th receive antenna, respectively. After OFDM demodulation, a MIMO detector is required to restore the original information that is transmitted. At this stage, the main objective is to eliminate or mitigate the effect of inter-antenna interference. Following the MIMO detector, a channel decoder then corrects most of the erroneous bits to achieve a better system error rate.

Petitioner Samsung Ex-1026, 0131

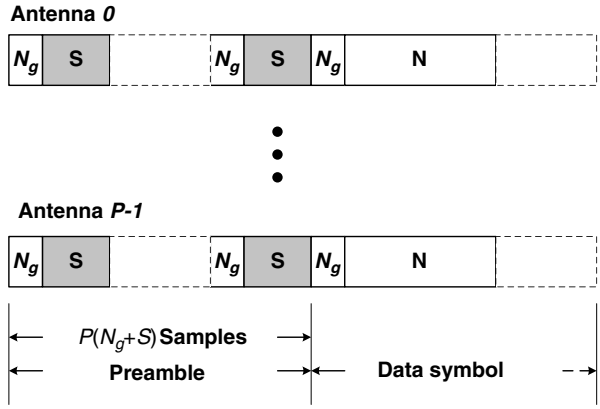


Figure 7.4. Preamble design for MIMO–OFDM systems

In such a MIMO–OFDM configuration, the considerations for inserting the training/reference signals, namely pilot signals/preambles, are totally different. Thus, synchronization and channel estimation tasks in a MIMO–OFDM receiver must be revisited. Instead of allocating pilot signals efficiently, the research on pilot patterns in MIMO–OFDM systems mainly focuses on estimating channel frequency response at all data subcarriers for each transmit–receive antenna pair. In the following, pilot insertion in MIMO–OFDM systems is first examined, and then related modifications on synchronization and channel estimation in MIMO–OFDM receivers are addressed.

7.3.1 MIMO Pilot Pattern

As in conventional SISO–OFDM systems, preamble structure plays a crucial role in synchronization and channel equalization in MIMO–OFDM systems. In [11], P preamble segments, each having S samples, are designed. Note that $S = N/U$, where U is an integer that divides N so as to shorten the preamble interval and speed up initial synchronization. Figure 7.4 depicts a preamble structure with P segments, each with N_g guard interval samples. An S-sample segment is generated by using non-zero pilot subcarriers spaced U subcarriers apart, transforming to the time domain by N-point IDFT, and keeping only the first S samples. Note that one preamble segment is U times shorter than one OFDM symbol. Denote the l-th frequency-domain pilot signal in the i-th preamble segment transmitted by antenna p as $X_i^{(p)}(l)$. The preamble structure is given by

$$\mathbf{\Gamma}_{\mathrm{PR}}(l) = \begin{bmatrix} X_0^{(0)}(l) & X_0^{(1)}(l) & \cdots & X_0^{(P-1)}(l) \\ X_1^{(0)}(l) & X_1^{(1)}(l) & \cdots & X_1^{(P-1)}(l) \\ & & \cdots & \\ X_{P-1}^{(0)}(l) & X_{P-1}^{(1)}(l) & \cdots & X_{P-1}^{(P-1)}(l) \end{bmatrix}. \tag{7.11}$$

Comb-type pilot patterns have also been proposed for MIMO–OFDM systems [12]. In this scheme, pilot signals across P symbols can be collected to form a space–time pilot block. An

Petitioner Samsung Ex-1026, 0132

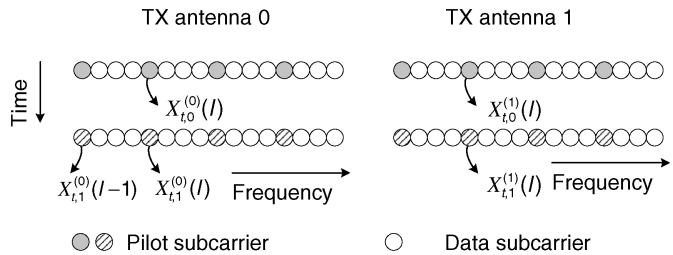


Figure 7.5. Space–time pilot insertion in MIMO–OFDM systems

example with $P = 2$ is illustrated in Figure 7.5. Note that one space–time pilot block is made up of P full OFDM symbols and the channel is assumed stationary in that period. Consequently, this pilot configuration is not suitable for fast-fading channels. Define $X_{t,i}^{(p)}(l)$ as the pilot signal at the l-th pilot subcarrier in the i-th symbol from the p-th antenna and the space–time pilot signal block for pilot subcarrier l takes the form of

$$\mathbf{\Gamma}_{\mathrm{ST}}(l) = \begin{bmatrix} X_{t,0}^{(0)}(l) & X_{t,0}^{(1)}(l) & \cdots & X_{t,0}^{(P-1)}(l) \\ X_{t,1}^{(0)}(l) & X_{t,1}^{(1)}(l) & \cdots & X_{t,1}^{(P-1)}(l) \\ & & \cdots & \\ X_{t,P-1}^{(0)}(l) & X_{t,P-1}^{(1)}(l) & \cdots & X_{t,P-1}^{(P-1)}(l) \end{bmatrix}. \tag{7.12}$$

The third pilot-insertion scheme is proposed in particular for fast-fading channels [13]. This scheme inserts pilot subcarriers in a space–frequency manner rather than among multiple OFDM symbols, as in the space–time pilot-insertion scheme. Assume that the channel impulse response has a maximum excess delay of M samples. Then, M equally spaced pilot subcarriers are adequate to estimate the channel impulse response correctly. A pilot-subcarrier set includes M subcarriers that are equally spaced with a spacing of D subcarriers, where $D = N/M$. Since the lowest-frequency pilot subcarrier can be located at subcarrier $0, 1, 2, \cdots$, or $D - 1$, there are D distinct pilot-subcarrier sets. In SISO–OFDM systems, any one of these D sets can be used for channel estimation. In MIMO–OFDM systems, however, at least P pilot-subcarrier sets are required to estimate the channel responses associated with P transmit antennas. The space–frequency pilot signal matrix is defined as

$$\mathbf{\Gamma}_{\mathrm{SF}}(l) = \begin{bmatrix} X_{f,0}^{(0)}(l) & X_{f,0}^{(1)}(l) & \cdots & X_{f,0}^{(P-1)}(l) \\ X_{f,1}^{(0)}(l) & X_{f,1}^{(1)}(l) & \cdots & X_{f,1}^{(P-1)}(l) \\ & & \cdots & \\ X_{f,P-1}^{(0)}(l) & X_{f,P-1}^{(1)}(l) & \cdots & X_{f,P-1}^{(P-1)}(l) \end{bmatrix}, \tag{7.13}$$

where $X_{f,i}^{(p)}(l)$ represents the l-th pilot subcarrier in the i-th set transmitted by the p-th antenna. The pilot insertion configuration with $M = D = 4$ and $P = 2$ is shown in Figure 7.6. Note that the pilot subcarrier density is higher than that of the space–time pilot-insertion scheme depicted in Figure 7.5. This is because the channel is assumed in fast

Petitioner Samsung Ex-1026, 0133

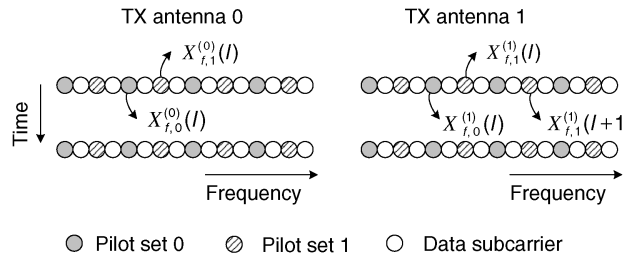


Figure 7.6. Space–frequency pilot insertion in MIMO–OFDM systems

fading and, thus, more pilot resources are needed to estimate the channel response in every symbol.

The three square matrices $\mathbf{\Gamma}_{\mathrm{PR}}(l)$, $\mathbf{\Gamma}_{\mathrm{ST}}(l)$ and $\mathbf{\Gamma}_{\mathrm{SF}}(l)$ for the three schemes are often designed as unitary matrices. The reason will be clear later, when MIMO channel estimation is discussed. For instance, in the case of real-valued pilot signals, matrices $\mathbf{\Gamma}_{\mathrm{PR}}(l)$, $\mathbf{\Gamma}_{\mathrm{ST}}(l)$ and $\mathbf{\Gamma}_{\mathrm{SF}}(l)$ of size 2×2 can take the form of

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{7.14}$$

If four transmit antennas are used, the following matrix from the space–time block codes [6] can be used:

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{bmatrix}. \tag{7.15}$$

For complex-valued pilot signals, the 2×2 unitary matrix can be chosen as [12]

$$\mathbf{S}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}. \tag{7.16}$$

Extending from the above matrix, one can derive the $P \times P$ unitary matrix for P equal to a power of two using

$$\mathbf{S}_4 = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{S}_2 & \mathbf{S}_2 \\ \mathbf{S}_2 & -\mathbf{S}_2 \end{bmatrix} \tag{7.17}$$

recursively. For those cases in which P is not a power of two, the $P \times P$ DFT matrix with entries proportional to $e^{-j2\pi pi/P}$ can be used [13].

7.3.2 MIMO–OFDM Synchronization

Since pilot signals of different transmit antennas have a certain relationship (Equations (7.14)–(7.17)), MIMO–OFDM systems can adopt the synchronization techniques in SISO–OFDM systems with only slight modification. For details about OFDM synchronization algorithms, the readers are referred to Chapter 5 of this book.

Petitioner Samsung Ex-1026, 0134

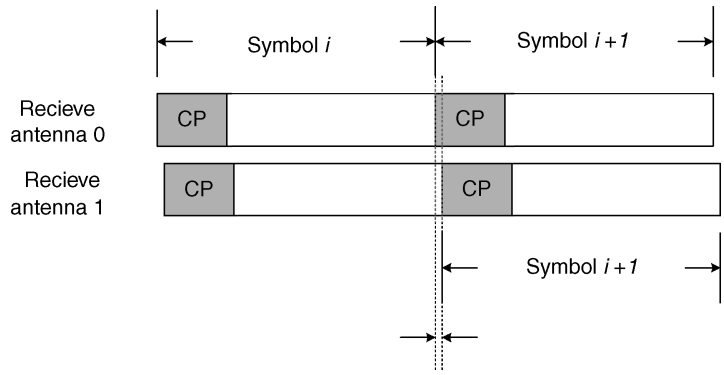


Figure 7.7. Symbol timings of different receive branches in MIMO–OFDM systems

Symbol Timing Detection

In [11,14,15], the delay and correlate algorithm and the normalized metric are adopted for the coarse symbol timing detection in MIMO–OFDM receivers. All of them still take advantage of the cyclic prefix in OFDM signals. The signals may arrive at the receive antennas at different times, causing slight misalignment of the symbol timings among the receive branches, as shown in Figure 7.7. As the misalignment is usually very small for normal antenna separation, one single symbol timing for all receive branches is usually assumed. The delay and correlate metric in the q-th receive branch takes the form of [11]

$$\begin{aligned} \Phi_{\mathrm{DC}}^{(q)}(m) &= |\sum_{r=0}^{R-1} z_{m-r}^{(q)} (z_{m-r-L}^{(q)})^*|, \\ \hat{m}^{(q)} &= \arg\max_m \ \Phi_{\mathrm{DC}}^{(q)}(m), \end{aligned} \tag{7.18}$$

where R denotes the correlation length; L indicates the separation between two adjacent replicas/repetitions; and $z_m^{(q)}$ is the received time-domain samples in the q-th receive branch. The acquired symbol timings from all receive branches corresponding to the peak values of $\Phi_{\mathrm{DC}}^{(q)}(m)$ are then averaged to obtain the estimated symbol timing.

A normalized metric for coarse symbol timing detection in MIMO–OFDM systems is proposed in [14,15]. This metric inherently gives more weight to the receive branches with stronger signal levels, and

$$\begin{aligned} \Phi_{\mathrm{NM}}(m) &= \frac{|\sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} z_{m-r}^{(q)} \cdot (z_{m-r-L}^{(q)})^*|^2}{(\sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} |z_{m-r}^{(q)}|^2)^2}, \\ \hat{m} &= \arg\max_m \ \Phi_{\mathrm{NM}}(m). \end{aligned} \tag{7.19}$$

As to the fine symbol timing detection in MIMO–OFDM receivers, both the time-domain cross-correlation [11,14] and the frequency–response–estimate algorithm can be employed [15]. Extending directly from the time-domain cross-correlation algorithm in SISO–OFDM

Petitioner Samsung Ex-1026, 0135

systems, the fine symbol timing detector for MIMO receivers requires P cross-correlations at each of the Q receive branches:

$$\hat{m}_{\mathrm{CC}}^{(q)} = \arg\max_{m} \sum_{p=0}^{P-1} \left| \sum_{r=0}^{S-1} z_{m+r}^{(q)} \cdot (s_r^{(p)})^* \right|^2, \tag{7.20}$$

where S is the length of preamble segment $s_r^{(p)}$ from transmit antenna p. When the preambles from P transmit antennas are related, as in Equations (7.14)–(7.17), then matching with any preamble is sufficient. The refined timing can be derived by finding the time index corresponding to the peak of the sum of all the correlator outputs or averaging the optimal time indices obtained in all receive branches.

In the frequency–response–estimate algorithm for MIMO systems, a superimposed time-domain MIMO channel response $\hat{g}_m^{(q)}$ is first derived in each receive branch. To avoid ISI, the fine symbol timing in each receive path is selected as the earliest arrival path with magnitude greater than some threshold of the peak in the obtained channel response:

$$\hat{m}_{\mathrm{TH}}^{(q)} = \min\left\{ m \;\middle|\; |\hat{g}_m^{(q)}| > \gamma |\hat{g}_{\mathrm{PM}}^{(q)}| \right\}, \tag{7.21}$$

where γ is the threshold ratio and $\hat{g}_{\mathrm{PM}}^{(q)}$ denotes the peak magnitude in $\hat{g}_m^{(q)}$. Note that this fine symbol timing detection must be done after the carrier frequency offset (including integer CFO and fractional CFO) has been compensated. The final symbol timing can be similarly derived from all $\hat{m}_{\mathrm{TH}}^{(q)}$ or $\hat{g}_m^{(q)}$.

Carrier Frequency Offset Estimation

In MIMO–OFDM systems, the transmit antennas are often co-located, and so are the receive antennas. Hence, it is valid to assume that only one oscillator is referenced in either the transmitter side or the receiver side. As a result, a single carrier frequency offset is to be estimated for the multiple receive branches. The maximum likelihood estimation for the fractional carrier frequency offset is quite popular in MIMO–OFDM systems [11,14,15].

Another fractional carrier frequency offset estimation algorithm for MIMO–OFDM systems applies different weights to the receive signals according to the respective degrees of channel fading [16]. The preamble is designed so that each transmit antenna uses non-overlapping subcarriers to facilitate separation of signals from different transmit antennas. At each receive branch, the cross-correlation between the received signal and the known preamble is examined. The magnitude of the cross-correlation output reflects the channel fading between the corresponding transmit and receive antenna pair. Based on the channel fading information, weights are applied to the received signals to emphasize those with stronger channel gains and, at the same time, suppress those that are deeply faded. Then, the carrier frequency offset is estimated based on the phase of delay correlation output of weighted signals.

For integer carrier frequency offset, frequency-domain cross-correlation [11] and frequency-domain PN-correlation [15] can be used with slight modification. First, the received signals must be compensated by the estimated fractional carrier frequency offset. Then, the compensated signals are transformed into the frequency domain. The

Petitioner Samsung Ex-1026, 0136

frequency-domain cross-correlation algorithm for one specific receive antenna is similar to that in the SISO case:

$$\hat{\epsilon}_I^{(q)} = \arg\max_{\epsilon_I} \ |\sum_l Z_{\alpha_l+\epsilon_I}^{(q)} X_{\alpha_l}^{(p)*}|, \tag{7.22}$$

where $X_k^{(p)}$ and $Z_k^{(q)}$ denote the transmitted and received signals at the k-th subcarrier from transmit antenna p to receive antena q, and α_l denotes the l-th pilot subcarrier. Because only one correct estimate of the integer frequency offset is allowed, inconsistent estimates from receive branches invalidate themselves. In this case, the cross-correlation output of only one receive antenna (preferably the one with strongest reception) is used.

Note that Equations (7.14)–(7.17) do not specify the relationship between neighboring pilot subcarriers. Differential encoding using a PN-sequence on adjacent pilot subcarriers can further be applied. The frequency-domain pilot-subcarrier signals of all receive branches are first summed:

$$\bar{Z}_{\alpha_l+\epsilon_I} = \sum_{q=0}^{Q-1} Z_{\alpha_l+\epsilon_I}^{(q)}. \tag{7.23}$$

The phase differences between adjacent pilot subcarriers are matched to the known PN-sequence to detect the possible frequency shift caused by the integer carrier frequency offset. Given that the pilot signals are encoded with a PN-sequence c_k of length M (see Equation (5.35)), the integer carrier frequency offset can be estimated by

$$\begin{aligned} \Phi_{\mathrm{PN}}(\epsilon_I) &= \frac{|\sum_l \bar{Z}_{\alpha_l+\epsilon_I} \bar{Z}^*_{\alpha_{l+1}+\epsilon_I} c_{[l]_M}|}{\sum_l |\bar{Z}_{\alpha_l+\epsilon_I}|^2 \cdot \sum_l |\bar{Z}_{\alpha_{l+1}+\epsilon_I}|^2}, \\ \hat{\epsilon}_I &= \arg\max_{\epsilon_I} \ \Phi_{\mathrm{PN}}(\epsilon_I), \end{aligned} \tag{7.24}$$

where $[\cdot]_M$ is modulo-M operation. Note that in this method, similar fading on two neighboring pilot subcarriers is assumed. In other words, the channel coherent bandwidth must be wider than the frequency separation between two neighboring pilot subcarriers.

Residual Carrier Frequency Offset and Sampling Clock Offset Estimation

In [17], a maximum likelihood estimation algorithm for CFO and SCO tracking in the MIMO–OFDM systems is proposed. In this method, the received frequency-domain signals are assumed to suffer from interferences and noises modelled as Gaussian distributions. Given the MIMO channel responses matrix at the l-th pilot subcarrier as

$$\mathbf{H}_{\alpha_l} = \begin{bmatrix} H_{\alpha_l}^{(0,0)} & \cdots & H_{\alpha_l}^{(0,P-1)} \\ & \ddots & \\ H_{\alpha_l}^{(Q-1,0)} & \cdots & H_{\alpha_l}^{(Q-1,P-1)} \end{bmatrix}, \tag{7.25}$$

and known pilot signals from all transmit antennas,

$$\mathbf{x}_{\alpha_l} = [X_{\alpha_l}^{(0)} \ X_{\alpha_l}^{(1)} \ \cdots \ X_{\alpha_l}^{(P-1)}]^T, \tag{7.26}$$

Petitioner Samsung Ex-1026, 0137

the maximum likelihood estimate of the CFO takes the form of

$$\hat{\epsilon}_f = \frac{-\arg\left(\sum_{l=0}^{L-1} \mathbf{z}_{\alpha_l}^H \mathbf{H}_{\alpha_l} \mathbf{x}_{\alpha_l}\right)}{2\pi \frac{N+N_g}{N}}, \tag{7.27}$$

where $\mathbf{z}_{\alpha_l} = [Z_{\alpha_l}^{(0)}\ Z_{\alpha_l}^{(1)}\ \cdots\ Z_{\alpha_l}^{(Q-1)}]^T$ denotes the received signals at subcarrier α_l in Q receive branches. As there is no close-form solution for SCO maximum likelihood estimation, the estimated SCO is approximated by

$$\hat{\delta} \approx \frac{-1}{2\pi \frac{N+N_g}{N}} \sum_{j=0}^{L-1} \sum_{l=0, l\neq j}^{L-1} \frac{\arg\left((\mathbf{z}_{\alpha_l}^H \mathbf{H}_{\alpha_l} \mathbf{x}_{\alpha_l}) \left(\mathbf{z}_{\alpha_j}^H \mathbf{H}_{\alpha_j} \mathbf{x}_{\alpha_j} \right)^H \right)}{\alpha_l - \alpha_j}, \tag{7.28}$$

where L is the number of pilot subcarriers. If the space–time pilot insertion block is repeated every P OFDM symbols, then the pilot phase differences across P symbols can be computed and the LLS and WLS tracking algorithms in Section 5.3.3 can estimate the CFO and SCO in the MIMO–OFDM systems. Of course, the two algorithms are effective only when the channel does not change significantly during any P–OFDM–symbol period.

7.3.3 MIMO–OFDM Channel Estimation

Channel estimation becomes more complicated for MIMO–OFDM systems since a channel matrix of the dimension $Q \times P$ is to be estimated. The estimation is usually performed in conjunction with either space–time pilot insertion or space–frequency pilot insertion.

Space-Time Pilot Insertion

For the systems with space–time pilot insertion, the channel is assumed stationary for the P-symbol duration; in other words,

$$H_{i,\alpha_l}^{(q,p)} = H_{i+1,\alpha_l}^{(q,p)} = \cdots = H_{i+P-1,\alpha_l}^{(q,p)}, \qquad \forall\, q, p, \alpha_l, \tag{7.29}$$

where $H_{i,\alpha_l}^{(q,p)}$ represents the channel frequency response at the l-th pilot subcarrier in symbol i from antenna p to antenna q. Let the pilot signals at subcarrier α_l transmitted by the p-th antenna be $\mathbf{x}_{\mathrm{ST},\alpha_l}^{(p)} = [X_{i,\alpha_l}^{(p)}\ X_{i+1,\alpha_l}^{(p)}\ \cdots\ X_{i+P-1,\alpha_l}^{(p)}]^T$, then the $P \times P$ pilot matrix $\mathbf{D}_{\alpha_l}$ is given by

$$\mathbf{D}_{\alpha_l} = [\mathbf{x}_{\mathrm{ST},\alpha_l}^{(0)}\ \mathbf{x}_{\mathrm{ST},\alpha_l}^{(1)}\ \cdots\ \mathbf{x}_{\mathrm{ST},\alpha_l}^{(P-1)}]. \tag{7.30}$$

A $P \times 1$ vector consisting of the received pilot signals at subcarrier α_l for P symbols in the q-th receive branch can be formulated as $\mathbf{z}_{\mathrm{ST},\alpha_l}^{(q)} = [Z_{i,\alpha_l}^{(q)}\ Z_{i+1,\alpha_l}^{(q)}\ \cdots\ Z_{i+P-1,\alpha_l}^{(q)}]^T$. Denote

$$\mathbf{h}_{\mathrm{ST},\alpha_l}^{(q)} = [H_{i,\alpha_l}^{(q,0)}\ H_{i,\alpha_l}^{(q,1)}\ \cdots\ H_{i,\alpha_l}^{(q,P-1)}]^T. \tag{7.31}$$

The received signal matrix becomes

$$\mathbf{z}_{\mathrm{ST},\alpha_l}^{(q)} = \mathbf{D}_{\alpha_l} \cdot \mathbf{h}_{\mathrm{ST},\alpha_l}^{(q)} + \mathbf{v}_{\mathrm{ST},\alpha_l}^{(q)}, \tag{7.32}$$

Petitioner Samsung Ex-1026, 0138

where $\mathbf{v}_{\mathrm{ST},\alpha_l}^{(q)}$ is the noise vector and

$$\mathbf{v}_{\mathrm{ST},\alpha_l}^{(q)} = [V_{i,\alpha_l}^{(q)}\ V_{i+1,\alpha_l}^{(q)}\ \cdots\ V_{i+P-1,\alpha_l}^{(q)}]^T. \tag{7.33}$$

Hence, a straightforward frequency-domain channel estimation can be derived from Equation (7.32) if the matrix $\mathbf{D}_{\alpha l}$ is invertible and

$$\begin{aligned}\hat{\mathbf{h}}_{\mathrm{ST},\alpha_l}^{(q)} &= \mathbf{D}_{\alpha_l}^{-1}\mathbf{z}_{\mathrm{ST},\alpha_l}^{(q)} \\ &= \mathbf{h}_{\mathrm{ST},\alpha_l}^{(q)} + \mathbf{D}_{\alpha_l}^{-1}\mathbf{v}_{\mathrm{ST},\alpha_l}^{(q)}.\end{aligned} \tag{7.34}$$

Provided that the total power of the pilot signals transmitted by all antennas is constrained by $\sum_{p=0}^{P-1} |X_{i,\alpha_l}^{(p)}|^2 = P$, then the MSE of the above channel estimator is given by

$$\begin{aligned}&\frac{1}{P}E\{||\hat{\mathbf{h}}_{\mathrm{ST},\alpha_l}^{(q)} - \mathbf{h}_{\mathrm{ST},\alpha_l}^{(q)}||^2\} \\ &= \frac{1}{P}E\{(\mathbf{D}_{\alpha_l}^{-1}\mathbf{v}_{\mathrm{ST},\alpha_l}^{(q)})^H(\mathbf{D}_{\alpha_l}^{-1}\mathbf{v}_{\mathrm{ST},\alpha_l}^{(q)})\} \\ &= \frac{\sigma_v^2}{P}tr\{(\mathbf{D}_{\alpha_l}\mathbf{D}_{\alpha_l}^H)^{-1}\} \\ &\geq \sigma_v^2,\end{aligned} \tag{7.35}$$

where σ_v^2 is the variance of the noise term $V_{i,\alpha_l}^{(q)}$ and $tr\{\cdot\}$ is the trace of a matrix. The equality holds only when $\mathbf{D}_{\alpha_l}$ is a unitary matrix [12]. In other words, the pilot matrix $\mathbf{D}_{\alpha_l}$ should be a unitary matrix so as to minimize the MSE in channel estimation.

Space–Frequency Pilot Insertion

In space–frequency pilot insertion for MIMO–OFDM systems under channels of M-sample maximum excess delay, P distinct pilot-subcarrier sets need to be allocated [13]. Note that $D = N/M$ and $P \leq D$. Define

$$\mathbf{g}^{(q,p)} = [g_0^{(q,p)}\ g_1^{(q,p)}\ g_{M-1}^{(q,p)}]^T, \tag{7.36}$$

where $g_r^{(q,p)}$ is the r-th tap in the channel impulse response associated with transmit antenna p and receive antenna q. Let $\mathbf{F}_M$ be the $M \times M$ matrix whose (m, r)-th element is $e^{-j2\pi mrD/N}$, then the channel frequency response of subcarrier $mD + l$ from transmit antenna p to receive antenna q can be expressed as

$$H_{mD+l}^{(q,p)} = (\mathbf{F}_M)_m\mathbf{R}_l\mathbf{g}^{(q,p)}, \tag{7.37}$$

where $\mathbf{R}_l$ is an $M \times M$ diagonal matrix whose r-th diagonal entry is $e^{-j2\pi lr/N}$ and $(\mathbf{F}_M)_m$ is the m-th row of matrix $\mathbf{F}_M$. The time index i in the subscript is dropped for clarity.

Let the transmitted signals at the subcarriers of the l-th pilot-subcarrier set from antenna p be formulated as a diagonal matrix

$$\mathbf{O}_{\beta_l}^{(p)} = \mathrm{diag}\,\{[X_{\beta_l}^{(p)}\ X_{D+\beta_l}^{(p)}\ \cdots\ X_{(M-1)D+\beta_l}^{(p)}]\}. \tag{7.38}$$

Petitioner Samsung Ex-1026, 0139

The received signal vector for the pilot-subcarrier set β_l takes the form of

$$\mathbf{z}_{\mathrm{SF},\beta_l}^{(q)} = \mathbf{O}_{\beta_l}^{(0)}\mathbf{F}_M\mathbf{R}_{\beta_l}\mathbf{g}^{(q,0)} + \cdots + \mathbf{O}_{\beta_l}^{(P-1)}\mathbf{F}_M\mathbf{R}_{\beta_l}\mathbf{g}^{(q,P-1)} + \mathbf{v}_{\mathrm{SF},\beta_l}^{(q)}, \tag{7.39}$$

where

$$\mathbf{z}_{\mathrm{SF},\beta_l}^{(q)} = [Z_{\beta_l}^{(q)}\ Z_{D+\beta_l}^{(q)}\ \cdots\ Z_{(M-1)D+\beta_l}^{(q)}]^T \tag{7.40}$$

and

$$\mathbf{v}_{\mathrm{SF},\beta_l}^{(q)} = [V_{\beta_l}^{(q)}\ V_{D+\beta_l}^{(q)}\ \cdots\ V_{(M-1)D+\beta_l}^{(q)}]^T. \tag{7.41}$$

Assume that a total of P pilot-subcarrier sets are used and denote these pilot sets as β_0, $\beta_1, \cdots, \beta_{P-1}$, where $0 \le \beta_0 < \beta_1 < \cdots < \beta_{P-1} < D$. Stacking the received signals of these P pilot-subcarrier sets yields

$$\begin{aligned}
\mathbf{z}_{\mathrm{SF}}^{(q)} &= \begin{bmatrix} \mathbf{z}_{\mathrm{SF},\beta_0}^{(q)} \\ \mathbf{z}_{\mathrm{SF},\beta_1}^{(q)} \\ \vdots \\ \mathbf{z}_{\mathrm{SF},\beta_{P-1}}^{(q)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{O}_{\beta_0}^{(0)}\mathbf{F}_M\mathbf{R}_{\beta_0} & \mathbf{O}_{\beta_0}^{(1)}\mathbf{F}_M\mathbf{R}_{\beta_0} & \cdots & \mathbf{O}_{\beta_0}^{(P-1)}\mathbf{F}_M\mathbf{R}_{\beta_0} \\ \mathbf{O}_{\beta_1}^{(0)}\mathbf{F}_M\mathbf{R}_{\beta_1} & \mathbf{O}_{\beta_1}^{(1)}\mathbf{F}_M\mathbf{R}_{\beta_1} & \cdots & \mathbf{O}_{\beta_1}^{(P-1)}\mathbf{F}_M\mathbf{R}_{\beta_1} \\ & \vdots & & \\ \mathbf{O}_{\beta_{P-1}}^{(0)}\mathbf{F}_M\mathbf{R}_{\beta_{P-1}} & \mathbf{O}_{\beta_{P-1}}^{(1)}\mathbf{F}_M\mathbf{R}_{\beta_{P-1}} & \cdots & \mathbf{O}_{\beta_{P-1}}^{(P-1)}\mathbf{F}_M\mathbf{R}_{\beta_{P-1}} \end{bmatrix} \begin{bmatrix} \mathbf{g}^{(q,0)} \\ \mathbf{g}^{(q,1)} \\ \vdots \\ \mathbf{g}^{(q,P-1)} \end{bmatrix} \\
&+ \begin{bmatrix} \mathbf{v}_{\mathrm{SF},\beta_0}^{(q)} \\ \mathbf{v}_{\mathrm{SF},\beta_1}^{(q)} \\ \vdots \\ \mathbf{v}_{\mathrm{SF},\beta_{P-1}}^{(q)} \end{bmatrix} \\
&= \mathbf{S}\mathbf{g}^{(q)} + \mathbf{v}_{\mathrm{SF}}^{(q)}.
\end{aligned} \tag{7.42}$$

Accordingly, the least-squares estimate of the channel impulse response $\mathbf{g}^{(q)}$ can be given by

$$\hat{\mathbf{g}}^{(q)} = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\mathbf{z}_{\mathrm{SF}}^{(q)}. \tag{7.43}$$

Using a special unitary space–frequency pilot matrix such as Equation (7.13), the complicated matrix inversion $(\mathbf{S}^H\mathbf{S})^{-1}$ can be avoided. Then, the estimated channel frequency responses can be given by

$$\hat{\mathbf{h}}_{\mathrm{SF}}^{(q)} = (\mathbf{I}_P \otimes \mathbf{F}_{N\times M})\mathbf{S}^H\mathbf{z}_{\mathrm{SF}}^{(q)}, \tag{7.44}$$

Petitioner Samsung Ex-1026, 0140

where $\otimes$ denotes the Kronecker product, $\mathbf{F}_{N\times M}$ is the $N \times M$ matrix whose (n, m) element is $e^{-j2\pi nm/N}$ and

$$\hat{\mathbf{h}}_{\mathrm{SF}}^{(q)} = \begin{bmatrix} \hat{H}_0^{(q,0)} \\ \vdots \\ \hat{H}_{N-1}^{(q,0)} \\ \hat{H}_0^{(q,1)} \\ \vdots \\ \hat{H}_{N-1}^{(q,1)} \\ \vdots \\ \hat{H}_{N-1}^{(q,P-1)} \end{bmatrix}. \tag{7.45}$$

7.4 MIMO Encoding and Detection

MIMO techniques can be categorized into three main types. The first type uses coding techniques to provide spatial diversity, such as space block codes; the second type uses spatial multiplexing to achieve capacity gain; and the third type exploits the knowledge of the channel and de-correlates the channel matrix so as to minimize interferences among antennas. In the following, the encoding strategies at the transmitter and related MIMO detection algorithms at the receiver for these three types of techniques will be presented.

7.4.1 Space Block Codes

Encoding

A simple space–time block code was proposed by Alamouti [5], and this code uses orthogonal codewords to achieve full diversity between two transmit antennas without the knowledge of channel state information. The data stream is encoded in 2×2 space–time blocks with redundancy across time slots and transmit antennas, as shown in Figure 7.8. Data in a space–time block are represented as

$$\mathbf{S}_{2,1} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}, \tag{7.46}$$

where the entries in rows and columns denote signals at time slots and transmit antennas, respectively. It is apparent that two symbols are transmitted in two time slots and thus a code rate of 1 is achieved. Note that since the two columns are orthogonal, the maximum likelihood decoding is reduced to simple linear processing at the receiver.

In [6], several orthogonal space–time codes designed for more transmit antennas were presented. The full-rate and delay-optimal designs with real entries of the form $\pm s_1, \cdots, \pm s_p$ were shown for the two, four and eight transmit antennas. For the 2×2 code, the space–time block code matrix is given by

$$\mathbf{G}_{2,1} = \begin{bmatrix} s_1 & s_2 \\ -s_2 & s_1 \end{bmatrix}. \tag{7.47}$$

Petitioner Samsung Ex-1026, 0141

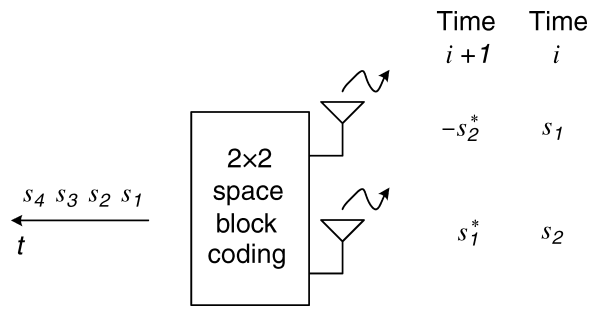


Figure 7.8. Signal transmission of the 2×2 space–time block code

The 4×4 design and the 8×8 design take the form of

$$\mathbf{G}_{4,1} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ -s_2 & s_1 & -s_4 & s_3 \\ -s_3 & s_4 & s_1 & -s_2 \\ -s_4 & -s_3 & s_2 & s_1 \end{bmatrix} \tag{7.48}$$

and

$$\mathbf{G}_{8,1} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ -s_2 & s_1 & s_4 & -s_3 & s_6 & -s_5 & -s_8 & s_7 \\ -s_3 & -s_4 & s_1 & s_2 & s_7 & s_8 & -s_5 & -s_6 \\ -s_4 & s_3 & -s_2 & s_1 & s_8 & -s_7 & s_6 & -s_5 \\ -s_5 & -s_6 & -s_7 & -s_8 & s_1 & s_2 & s_3 & s_4 \\ -s_6 & s_5 & -s_8 & s_7 & -s_2 & s_1 & -s_4 & s_3 \\ -s_7 & s_8 & s_5 & -s_6 & -s_3 & s_4 & s_1 & -s_2 \\ -s_8 & -s_7 & s_6 & s_5 & -s_4 & -s_3 & s_2 & s_1 \end{bmatrix}, \tag{7.49}$$

respectively.

For complex symbols such as the PSK and QAM constellations, the full-rate orthogonal design with entries taking the form of $\pm s_1, \pm s_2, \cdots \pm s_p$ and $\pm s_1^*, \pm s_2^*, \cdots, \pm s_p^*$ has been shown to exist only for $n = 2$ [6], namely the Alamouti code in Equation (7.46). For transmission using three and four transmit antennas, orthogonal codes exist with reduced code rates. The rate 1/2 codes with three and four transmit antennas are given by

$$\mathbf{S}_{3,1/2} = \begin{bmatrix} s_1 & s_2 & s_3 \\ -s_2 & s_1 & -s_4 \\ -s_3 & s_4 & s_1 \\ -s_4 & -s_3 & s_2 \\ s_1^* & s_2^* & s_3^* \\ -s_2^* & s_1^* & -s_4^* \\ -s_3^* & s_4^* & s_1^* \\ -s_4^* & -s_3^* & s_2^* \end{bmatrix}, \tag{7.50}$$

Petitioner Samsung Ex-1026, 0142

and

$$\mathbf{S}_{4,1/2} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ -s_2 & s_1 & -s_4 & s_3 \\ -s_3 & s_4 & s_1 & -s_2 \\ -s_4 & -s_3 & s_2 & s_1 \\ s_1^* & s_2^* & s_3^* & s_4^* \\ -s_2^* & s_1^* & -s_4^* & s_3^* \\ -s_3^* & s_4^* & s_1^* & -s_2^* \\ -s_4^* & -s_3^* & s_2^* & s_1^* \end{bmatrix}, \tag{7.51}$$

respectively. The code rate can be increased to $3/4$ by using the signals with varying magnitudes. Examples of these orthogonal codes include

$$\mathbf{S}_{3,3/4} = \begin{bmatrix} s_1 & s_2 & \frac{s_3}{\sqrt{2}} \\ -s_2^* & s_1^* & \frac{s_3}{\sqrt{2}} \\ \frac{s_3^*}{\sqrt{2}} & \frac{s_3^*}{\sqrt{2}} & \frac{-s_1-s_1^*+s_2-s_2^*}{2} \\ \frac{s_3^*}{\sqrt{2}} & -\frac{s_3^*}{\sqrt{2}} & \frac{s_1-s_1^*+s_2+s_2^*}{2} \end{bmatrix} \tag{7.52}$$

and

$$\mathbf{S}_{4,3/4} = \begin{bmatrix} s_1 & s_2 & \frac{s_3}{\sqrt{2}} & \frac{s_3}{\sqrt{2}} \\ -s_2^* & s_1^* & \frac{s_3}{\sqrt{2}} & -\frac{s_3}{\sqrt{2}} \\ \frac{s_3^*}{\sqrt{2}} & \frac{s_3^*}{\sqrt{2}} & \frac{-s_1-s_1^*+s_2-s_2^*}{2} & \frac{s_1-s_1^*-s_2-s_2^*}{2} \\ \frac{s_3^*}{\sqrt{2}} & -\frac{s_3^*}{\sqrt{2}} & \frac{s_1-s_1^*+s_2+s_2^*}{2} & -\frac{s_1+s_1^*+s_2-s_2^*}{2} \end{bmatrix}. \tag{7.53}$$

Detection

To detect the transmitted symbols in space–time block codes, all the $Q \times P$ channel responses must be stationary within one space–time code block. Let the two frequency-domain signals at subcarrier k of two consecutive OFDM symbols be coded using the Alamouti code and transmitted from two antennas. The received signals at that subcarrier of two corresponding OFDM symbols at the q-th receive antenna become

$$\begin{aligned} Z_{0,k}^{(q)} &= H_k^{(q,0)} s_1 + H_k^{(q,1)} s_2 + V_{0,k}^{(q)} \\ Z_{1,k}^{(q)} &= -H_k^{(q,0)} s_2^* + H_k^{(q,1)} s_1^* + V_{1,k}^{(q)}, \end{aligned} \tag{7.54}$$

where $H_k^{(q,p)}$ is the channel frequency response from antenna p to antenna q, and s_p is the transmitted frequency-domain data from antenna $p-1$ during the zeroth OFDM symbol, and $Z_{i,k}^{(q)}$ and $V_{i,k}^{(q)}$ are the received signal and the noise of the subcarrier in the i-th OFDM symbol

Petitioner Samsung Ex-1026, 0143

at the q-th receive antenna, respectively. The maximum-likelihood symbol detector then searches all possible values of (s_1, s_2) to find the minimium of the metric

$$\sum_{q=0}^{Q-1}(|Z_{0,k}^{(q)} - H_k^{(q,0)}s_1 - H_k^{(q,1)}s_2|^2 + |Z_{1,k}^{(q)} + H_k^{(q,0)}s_2^* - H_k^{(q,1)}s_1^*|^2). \tag{7.55}$$

Due to the orthogonality between the codewords, the cross-terms associated with the product $s_1 s_2$ are cancelled. Thus, detection of $\hat{s}_1$ and $\hat{s}_2$ is achieved by minimizing the following two metrics separately:

$$\left|\left[\sum_{q=0}^{Q-1}(Z_{0,k}^{(q)}H_k^{(q,0)*} + Z_{1,k}^{(q)*}H_k^{(q,1)})\right] - s_1\right|^2 + \left(-1 + \sum_{q=0}^{Q-1}\sum_{p=0}^{1}|H_k^{(q,p)}|^2\right)|s_1|^2 \tag{7.56}$$

and

$$\left|\left[\sum_{q=0}^{Q-1}(Z_{0,k}^{(q)}H_k^{(q,1)*} - Z_{1,k}^{(q)*}H_k^{(q,0)})\right] - s_2\right|^2 + \left(-1 + \sum_{q=0}^{Q-1}\sum_{p=0}^{1}|H_k^{(q,p)}|^2\right)|s_2|^2. \tag{7.57}$$

The linear decoding approach can also be applied to space–time block codes with three and four transmit antennas [6].

7.4.2 Spatial Multiplexing

In spatial multiplexing MIMO systems, the input data stream is first parsed into several sub-streams, known as layers, and then each sub-stream is transmitted by the transmit antennas. As a result, the streams are multiplexed in the spatial domain, hence the name of the technique.

Encoding

In 1996, G. J. Foschini at Bell Lab. first proposed a diagonal layered space–time scheme, also known as the *Diagonal Bell Labs Layered Space-Time* (D-BLAST) scheme [18]. Given P transmit antennas, the original data stream is divided into P sub-streams, and data in each sub-stream is processed by a channel encoder and mapped into complex symbols. To share the resources in the space and time dimensions, each sub-stream is periodically cycled through the set of transmit antennas, as shown in Figure 7.9. Hence, each layer can take advantage of spatial diversity and will not be in deep fade all the time.

Based on D-BLAST, the *vertical-BLAST* (*V-BLAST*) scheme, was subsequently developed in 1998. V-BLAST can be regarded as a simplified version of D-BALST with reduced decoding complexity [10]. Rather than allocating each sub-stream diagonally across space and time, the data stream is split into sub-streams by a de-multiplexer, as shown in Figure 7.10. Symbols from the channel encoder are arranged vertically in the space–time coordinate; thus, this technique is named V-BLAST. Correlation between different layers due to the multiplexing makes joint detection of all layers inevitable.

Petitioner Samsung Ex-1026, 0144

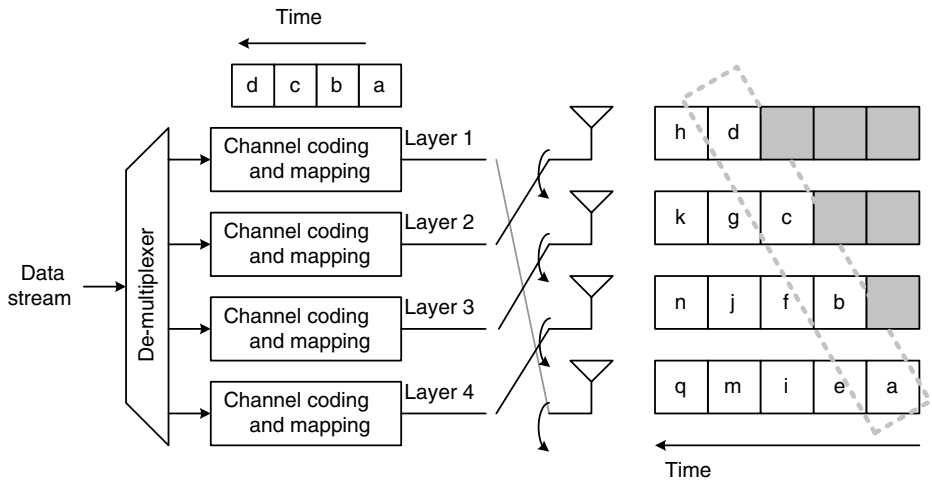


Figure 7.9. D-BLAST spatial multiplexing transmission

In addition to V-BLAST, the *horizontal-BLAST* (H-BLAST) scheme, shown in Figure 7.11, can also achieve spatial multiplexing. In H-BLAST, each sub-stream is processed by a channel encoder and fed to a transmit antenna, and thus spatial diversity is lost.

Detection

Given the $P \times 1$ transmitted signals $\mathbf{x} = [X^{(0)}\ X^{(1)}\ \cdots\ X^{(P-1)}]^T$ and the $Q \times P$ channel matrix $\mathbf{H}$ as in Equation (7.3), the received signals can be expressed as

$$\mathbf{z} = \begin{bmatrix} Z^{(0)} \\ Z^{(1)} \\ \vdots \\ Z^{(Q-1)} \end{bmatrix} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{7.58}$$

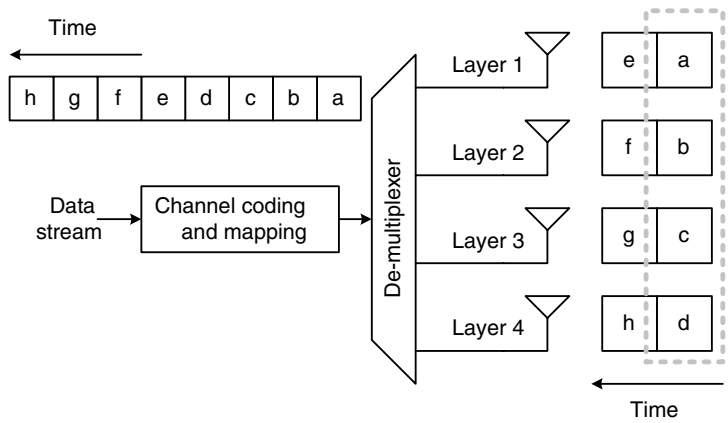


Figure 7.10. V-BLAST spatial multiplexing transmission

Petitioner Samsung Ex-1026, 0145

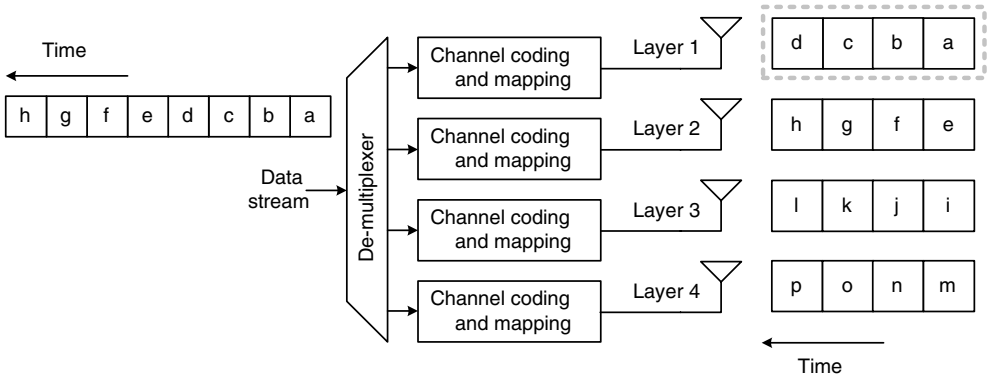


Figure 7.11. H-BLAST spatial multiplexing transmission

where $\mathbf{v}$ represents the noise vector. Since there are P unknowns and Q equations, these unknowns can be solved if $Q \geq P$.

Zero-Forcing (ZF)

The *zero-forcing* (ZF) detector simply derives the transmitted signals by solving Equation (7.58) regardless of the noise,

$$\hat{\mathbf{x}} = \mathbf{G}_{\mathrm{ZF}}\mathbf{z}, \tag{7.59}$$

where $\mathbf{G}_{\mathrm{ZF}}$ is the pseudo-inverse of the matrix $\mathbf{H}$ and is given by

$$\mathbf{G}_{\mathrm{ZF}} = \mathbf{H}^{\dagger} = (\mathbf{H}^{H}\mathbf{H})^{-1}\mathbf{H}^{H}. \tag{7.60}$$

Despite its simplicity, this approach suffers from the noise enhancement problem.

Minimum Mean-Squared Error (MMSE)

The *minimum mean-squared error* (MMSE) algorithm detects the transmitted signal through minimizing $E\{(\hat{\mathbf{x}} - \mathbf{x})^{H}(\hat{\mathbf{x}} - \mathbf{x})\}$. The detector calculates the transmitted signal vector by

$$\hat{\mathbf{x}} = \mathbf{G}_{\mathrm{MMSE}}\mathbf{z} \tag{7.61}$$

with

$$\mathbf{G}_{\mathrm{MMSE}} = \mathbf{R}_{\mathbf{xz}}\mathbf{R}_{\mathbf{zz}}^{-1}, \tag{7.62}$$

where the cross-correlation matrix $\mathbf{R}_{\mathbf{xz}}$ can be reduced and

$$\mathbf{R}_{\mathbf{xz}} = E\{\mathbf{x}\mathbf{z}^{H}\} = \mathbf{R}_{\mathbf{xx}}\mathbf{H}^{H} = \sigma_x^2\mathbf{H}^{H}. \tag{7.63}$$

In addition, the auto-correlation matrix can also re-formulated as

$$\mathbf{R}_{\mathbf{zz}} = E\{\mathbf{z}\mathbf{z}^{H}\} = [\sigma_x^2(\mathbf{H}\mathbf{H}^{H}) + \sigma_v^2\mathbf{I}_Q], \tag{7.64}$$

Petitioner Samsung Ex-1026, 0146

where σ_x^2 and σ_v^2 are the signal power and noise power, respectively. Hence, the MMSE matrix takes the form of

$$\mathbf{G}_{\text{MMSE}} = \mathbf{H}^H \left[\mathbf{H}\mathbf{H}^H + \frac{\mathbf{I}_Q}{\rho} \right]^{-1} = \left[\mathbf{H}^H \mathbf{H} + \frac{\mathbf{I}_P}{\rho} \right]^{-1} \mathbf{H}^H, \tag{7.65}$$

where $\rho = \sigma_x^2/\sigma_v^2$ is the signal-to-noise ratio. Evidently, with the SNR information, the MMSE detector avoids amplifying the noise and thus performs better than the ZF detector. It is worth noting that when the SNR is high, the MMSE detector approaches the ZF detector. On the other hand, the MMSE detector becomes a matched filter when the SNR is very low.

Ordered Successive Interference Cancellation (OSIC)

Both the ZF and the MMSE algorithms are linear detectors. There also exist nonlinear algorithms that can solve the detection problem for the data that have been spatially multiplexed. The well known *ordered successive interference cancellation* (OSIC) method was proposed when the V-BLAST prototype was announced [19]. The concept behind the OSIC algorithm is to detect the strongest undetected signal at present, then eliminate the interference caused by the resolved signal before the next decision, then do the next decision, and so on. The procedure goes on iteratively until all the transmitted signals are obtained. The OSIC process with iteration index i is described below in more detail:

- Initialization:

$$\begin{aligned} \mathbf{H}(0) &= \mathbf{H}, \\ \mathbf{z}(0) &= \mathbf{z}. \end{aligned} \tag{7.66}$$

- Ordering:
  Find the most reliable signal,

$$\begin{aligned} \mathbf{G}(i) &= \mathbf{H}(i)^\dagger = (\mathbf{H}(i)^H \mathbf{H}(i))^{-1} \mathbf{H}(i)^H, \\ \alpha_i &= \arg\min_p \|(\mathbf{G}(i))_p\|^2, \end{aligned} \tag{7.67}$$

  where $(\mathbf{G}(i))_p$ denotes the p-th row of the matrix $\mathbf{G}(i)$ and the row with the minimum norm corresponds to the transmitted signal with the highest SNR.
- Detection:
  Detect the transmitted signal identified in the previous step and then make a decision,

$$\hat{X}^{(\alpha_i)} = Q[(\mathbf{G}(i))_{\alpha_i} \mathbf{z}(i)], \tag{7.68}$$

  where $Q(\cdot)$ quantizes its argument to the nearest constellation point.
- Interference cancellation:
  Remove the interference from $\hat{X}^{(\alpha_i)}$,

$$\mathbf{z}(i+1) = \mathbf{z}(i) - \mathbf{h}_{\alpha_i} \hat{X}^{(\alpha_i)}, \tag{7.69}$$

  where $\mathbf{h}_{\alpha_i}$ is the (α_i)-th column of $\mathbf{H}$.

Petitioner Samsung Ex-1026, 0147

- Interference nulling:
  Null the effect in the channel matrix $\mathbf{H}$ corresponding to $\hat{X}^{(\alpha_i)}$,

$$\mathbf{H}(i+1) = \text{null}[\mathbf{H}(i)]^{\alpha_i}, \tag{7.70}$$

  where the null$[\cdot]^p$ operator fills the p-th column of the enclosed matrix with the all-zero column vector.
- Recursion from Equation (7.67).

The ordered set $\{\alpha_0, \alpha_1, \cdots, \alpha_{P-1}\}$ then specifies the detected sequence of the transmitted signals. In addition, matrix $\mathbf{G}(i)$ can be replaced by the MMSE solution of $\mathbf{H}(i)$ to obtain better performance.

Example:

Assume that the 2×2 channel matrix is given by

$$\mathbf{H} = \begin{bmatrix} h^{(0,0)} & h^{(0,1)} \\ h^{(1,0)} & h^{(1,1)} \end{bmatrix}.$$

The OSIC scheme checks rows of matrix $\mathbf{G}$ (equivalently columns of matrix $\mathbf{H}$) and if

$$|h^{(0,0)}|^2 + |h^{(1,0)}|^2 < |h^{(0,1)}|^2 + |h^{(1,1)}|^2,$$

then the detected signal can be computed by

$$\tilde{X}^{(1)} = \frac{1}{\det(\mathbf{H})}\left[-h^{(1,0)} \;\; h^{(0,0)}\right]\left[Z^{(0)} \;\; Z^{(1)}\right]^T, \quad \hat{X}^{(1)} = Q\left[\tilde{X}^{(1)}\right],$$

and

$$\tilde{X}^{(0)} = \frac{1}{|h^{(0,0)}|^2 + |h^{(1,0)}|^2}[h^{(0,0)*} \;\; h^{(1,0)*}]\left([Z^{(0)} \;\; Z^{(1)}]^T - \hat{X}^{(1)}[h^{(0,1)} \;\; h^{(1,1)}]^T\right),$$
$$\hat{X}^{(0)} = Q\left[\tilde{X}^{(0)}\right].$$

Otherwise, if

$$|h^{(0,0)}|^2 + |h^{(1,0)}|^2 \geq |h^{(0,1)}|^2 + |h^{(1,1)}|^2,$$

then

$$\tilde{X}^{(0)} = \frac{1}{\det(\mathbf{H})}\left[-h^{(0,1)} \;\; h^{(1,1)}\right]\left[Z^{(1)} \;\; Z^{(0)}\right]^T, \quad \hat{X}^{(0)} = Q\left[\tilde{X}^{(0)}\right],$$

and

$$\tilde{X}^{(1)} = \frac{1}{|h^{(1,1)}|^2 + |h^{(0,1)}|^2}[h^{(1,1)*} \;\; h^{(0,1)*}]\left([Z^{(1)} \;\; Z^{(0)}]^T - \hat{X}^{(0)}[h^{(1,0)} \;\; h^{(0,0)}]^T\right),$$
$$\hat{X}^{(1)} = Q[\tilde{X}^{(1)}].$$

Petitioner Samsung Ex-1026, 0148

Sphere Decoding (SD)

The *maximum likelihood* (ML) solution to the MIMO detection problem finds the P symbols satisfying:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{x \in A} ||\mathbf{z} - \mathbf{H}\mathbf{x}||^2, \tag{7.71}$$

where A is the set of all possible P-symbol combinations, and each symbol is a point from the constellation used, such as QPSK, 16-QAM or 64-QAM. The ML detector must search all possible combinations and, thus, the complexity grows exponentially with P. In light of this huge complexity, sphere decoding was proposed to reduce the search complexity in an ML detector for MIMO detection [20].

Given the zero-forcing solution

$$\hat{\mathbf{x}} = \mathbf{H}^{\dagger}\mathbf{z} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{z}, \tag{7.72}$$

compute

$$\begin{aligned}
||\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}})||^2 &= (\mathbf{x} - \hat{\mathbf{x}})^H\mathbf{H}^H\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}) \\
&= \mathbf{x}^H\mathbf{H}^H\mathbf{H}\mathbf{x} - \mathbf{z}^H\mathbf{H}^{\dagger H}(\mathbf{H}^H\mathbf{H})\mathbf{x} - \mathbf{x}^H(\mathbf{H}^H\mathbf{H})\mathbf{H}^{\dagger}\mathbf{z} + \mathbf{z}^H\mathbf{H}^{\dagger H}(\mathbf{H}^H\mathbf{H})\mathbf{H}^{\dagger}\mathbf{z} \\
&= \mathbf{x}^H\mathbf{H}^H\mathbf{H}\mathbf{x} - \mathbf{z}^H\mathbf{H}\mathbf{x} - \mathbf{x}^H\mathbf{H}^H\mathbf{z} + \mathbf{z}^H\mathbf{H}^H(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{z} \\
&= \mathbf{x}^H\mathbf{H}^H\mathbf{H}\mathbf{x} - \mathbf{z}^H\mathbf{H}\mathbf{x} - \mathbf{x}^H\mathbf{H}^H\mathbf{z} + \mathbf{z}^H\mathbf{z} - (\mathbf{z}^H\mathbf{z} - \mathbf{z}^H\mathbf{H}^H(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{z}) \\
&= ||\mathbf{z} - \mathbf{H}\mathbf{x}||^2 - \mathbf{z}^H(\mathbf{I} - \mathbf{H}^H(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H)\mathbf{z}.
\end{aligned} \tag{7.73}$$

The above equation can be rearranged and one has

$$||\mathbf{z} - \mathbf{H}\mathbf{x}||^2 = ||\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}})||^2 + \mathbf{z}^H(\mathbf{I} - \mathbf{H}^H(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H)\mathbf{z}. \tag{7.74}$$

Since the second term in the above equation is constant for all possible $\mathbf{x}$, the ML solution becomes

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{x \in A} ||\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}})||^2. \tag{7.75}$$

To avoid an exhaustive search, the sphere decoder, as its name implies, only searches for those constellation points lying within a P-dimensional hyperspace sphere centered at $\hat{\mathbf{x}}$. This, of course, only works when the radius d of the sphere is large enough to enclose the ML solution:

$$||\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}})||^2 < d^2. \tag{7.76}$$

Furthermore, by QR decomposition, the channel matrix $\mathbf{H}$ is decomposed into $\mathbf{H} = \mathbf{U}\mathbf{R}$, where $\mathbf{R}$ is an upper-triangle matrix and $\mathbf{U}$ is a unitary matrix so that $\mathbf{H}^H\mathbf{H} = \mathbf{R}^H\mathbf{R}$. Let the (i,j) element in $\mathbf{R}$ be denoted as $r_{i,j}$; then, the cost function to be minimized can be

Petitioner Samsung Ex-1026, 0149

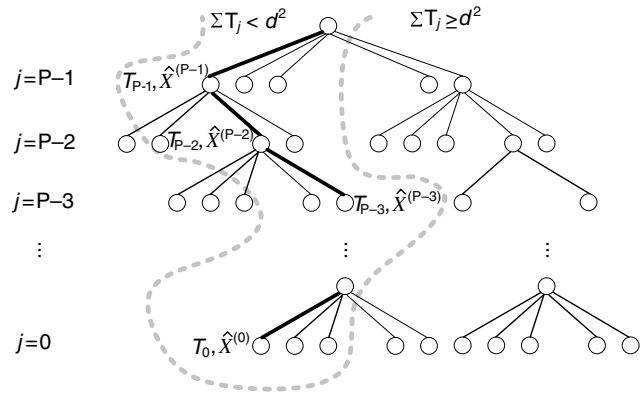


Figure 7.12. Tree-search architecture in sphere decoder

expressed as

$$(\mathbf{x}-\hat{\mathbf{x}})^H\mathbf{H}^H\mathbf{H}(\mathbf{x}-\hat{\mathbf{x}}) = |\sum_{i=0}^{P-1}(X^{(i)}-\hat{X}^{(i)})r_{0,i}|^2 + |\sum_{i=1}^{P-1}(X^{(i)}-\hat{X}^{(i)})r_{1,i}|^2 + \cdots$$
$$+ |\sum_{i=P-1}^{P-1}(X^{(i)}-\hat{X}^{(i)})r_{P-1,i}|^2 = T_0 + T_1 + \cdots + T_{P-1}, \qquad (7.77)$$

where $T_j = |\sum_{i=j}^{P-1}(X^{(i)}-\hat{X}^{(i)})r_{j,i}|^2$. Note that the cost function is now expressed as the sum of P terms.

The sphere decoder starts from variable $X^{(P-1)}$ and discards all those $X^{(P-1)}$ with $T_{P-1} \geq d^2$. With those remaining $X^{(P-1)}$, the sphere decoder proceeds to test all possible $X^{(P-2)}$ and again discards all those $(X^{(P-1)}, X^{(P-2)})$ with $T_{P-1} + T_{P-2} \geq d^2$. The process continues until the final variable $X^{(0)}$ is reached. The whole sphere decoding process can be regarded as a P-level tree search, as shown in Figure 7.12. In level p, only descendants of the parent nodes with $\sum_{j=p+1}^{P-1} T_j < d^2$ are considered. As soon as the partial sum $\sum_{j=p}^{P-1} T_j$ exceeds d^2, the node and all its descendants are removed from the search space. When all viable leaf nodes are visited, the ML solution is the leaf node with the minimum cost function. In addition to the above sphere decoding scheme, there are several modified techniques that can further improve the search efficiency [21].

It has been shown in [22] that OSIC with MMSE matrix can out-perform the original OSIC using the pseudo-inverse ZF matrix. Moreover, both nonlinear OSIC schemes are better than the linear zero-forcing detector and the MMSE detector. On the other hand, the ML solution has the optimal detection performance, albeit with very high complexity. With careful design of the radius and search strategy, sphere decoding can approach the ML performance with considerably lower complexity.

7.4.3 Spatial De-correlation

Encoding

All the aforementioned MIMO techniques do not have the luxury of known channel state information (CSI) at the transmitter. CSI at the transmitter can be derived either by the

Petitioner Samsung Ex-1026, 0150

reciprocity property of the wireless link or through a feedback channel from the receiver. With CSI, the transmitter can apply spatial pre-coding that effectively de-correlates the cross-coupled channel by diagonalizing the channel matrix and creating parallel sub-channels. When this happens, the spatial streams are de-coupled from one another.

The *singular value decomposition* (SVD) procedure is adopted to convert the channel matrix into the product of a diagonal matrix and two unitary matrices. The channel matrix $\mathbf{H}$ can be decomposed as

$$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H, \tag{7.78}$$

where $\Sigma = \text{diag}(\sigma_0, \sigma_1, \cdots, \sigma_{R-1})$ contains channel singular values $\sigma_0 \geq \sigma_1 \geq \cdots \sigma_{R-1} \geq 0$; $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, satisfying $\mathbf{U}^H\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^H\mathbf{V} = \mathbf{I}$. The de-correlating pre-coding encodes the data by $\mathbf{V}$,

$$\mathbf{x} = \mathbf{V}\mathbf{s}, \tag{7.79}$$

where $\mathbf{s} = [s_0\ s_1\ \cdots\ s_{R-1}]^T$ denotes the R spatial streams, as shown in Figure 7.13(a).

Decoding

The received signals with spatial pre-coding can be expressed as

$$\begin{aligned} \mathbf{z} &= \mathbf{H}\mathbf{x} + \mathbf{v} = \mathbf{U}\Sigma\mathbf{V}^H\mathbf{V}\mathbf{s} + \mathbf{v} \\ &= \mathbf{U}\Sigma\mathbf{s} + \mathbf{v}. \end{aligned} \tag{7.80}$$

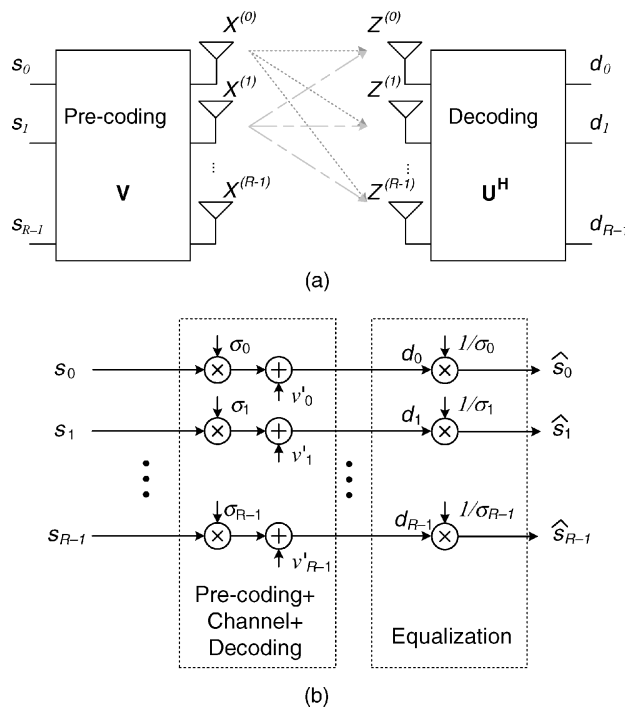


Figure 7.13. (a) Pre-processing and post-processing in SVD and (b) compensation for parallel orthogonal sub-channels

Petitioner Samsung Ex-1026, 0151

Then, the received signal is pre-multiplied by matrix $\mathbf{U}^H$ (see Figure 7.13(a)),

$$\begin{aligned}\mathbf{d} &= \mathbf{U}^H\mathbf{z} \\ &= \Sigma\mathbf{s} + \mathbf{U}^H\mathbf{v} \\ &= \Sigma\mathbf{s} + \mathbf{v}'.\end{aligned} \tag{7.81}$$

It is clear that R SISO parallel sub-channels are effectively formed and the spatial streams in $\mathbf{s}$ can be obtained by simple equalization:

$$\hat{s}_r = \sigma_r^{-1} d_r, \quad r = 0, 1, \cdots, R-1, \tag{7.82}$$

as shown in Figure 7.13(b).

Example:

Assume a 2×2 channel matrix

$$\mathbf{H} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}.$$

The column vectors in unitary matrix $\mathbf{V}$ for its SVD can be found as the eigenvectors of the matrix

$$\mathbf{H}^H\mathbf{H} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

and the singular values σ_0 and σ_1 are the square roots of the eigenvalues of $\mathbf{H}^H\mathbf{H}$. The eigenvalues of $\mathbf{H}^H\mathbf{H}$ are

$$\lambda_0 = 8, \quad \lambda_1 = 2,$$

and the corresponding orthonormal eigenvectors are

$$\mathbf{v}_0 = [1/\sqrt{2} \;\; 1/\sqrt{2}]^T, \quad \mathbf{v}_1 = [-1/\sqrt{2} \;\; 1/\sqrt{2}]^T.$$

Hence,

$$\mathbf{V} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad \Sigma = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}.$$

And the matrix $\mathbf{U}$ can be obtained by

$$\mathbf{U} = \mathbf{H}\mathbf{V}\Sigma^{-1} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/(2\sqrt{2}) & 0 \\ 0 & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Petitioner Samsung Ex-1026, 0152

Bibliography

[1] A. Paulraj, R. Nabar and D. Gore, *Introduction to space–time wireless communications*. Cambridge, UK: Cambridge University Press, 2003.

[2] H. Yang, 'A road to future broadband wireless access: MIMO–OFDM-based air interface', *IEEE Commun. Mag.*, vol. 1, Jan. 2005, pp. 53–60.

[3] T. Eng, K. Ning and L. B. Milstein, 'Comparison of diversity combining techniques for Rayleigh-fading channels', *IEEE Trans. Commun.*, vol. 44, Sep. 1996, pp. 1117–1129.

[4] V. Tarokh, N. Seshadri and A. R. Calderbank, 'Space–time codes for high data rate wireless communication: Performance criterion and code construction', *IEEE Trans. Inform. Theory*, vol. 44, Mar. 1998, pp. 744–765.

[5] S. Alamouti, 'A simple transmit diversity technique for wireless communications', *IEEE J. Select. Areas Commum.*, vol. 16, Oct. 1998, pp. 1451–1458.

[6] V. Tarokh, H. Jafarkhani and A. R. Calderbank, 'Space–time block codes from orthogonal design', *IEEE Trans. Inform. Theory*, vol. 45, no. 5, July 1999, pp. 1456–1467.

[7] G. Bauch, 'Space–time block codes versus space–frequency block codes', in *Proc. of IEEE Vehicular Technology Conference*, vol. 1, Apr. 2003, pp. 567–571.

[8] L. C. Godara, 'Applications of antenna arrays to mobile communications, part II: Beam-forming and direction-of-arrival considerations', *Proc. of IEEE*, vol. 85, Aug. 1997, pp. 1193–1245.

[9] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith and A. Naguib, 'From theory to practice: an overview of MIMO space–time coded wireless systems', *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, Apr. 2003, pp. 281–302.

[10] G. J. Foschini and M. J. Gans, 'On limits of wireless communications in a fading environment when using multiple antennas', *Wireless Personal Commun.*, vol. 6, Mar. 1998, pp. 311–335.

[11] G. L. Stuber, J. Barry, S. McLaughlin, *et al.*, 'Broadband MIMO–OFDM wireless communications', *Proc. IEEE*, vol. 92, Feb. 2004, pp. 271–293.

[12] D. Wang, G. Zhu and Z. Hu, 'Optimal pilots in frequency domain for channel estimation in MIMO–OFDM systems in mobile wireless channels', in *Proc. of IEEE Vehicular Technology Conference*, May 2004, vol. 2, pp. 608–612.

[13] Z. Wu, J. He and G. Gu, 'Design of optimal pilot-tones for channel estimation in MIMO–OFDM systems', in *Proc. of IEEE Wireless Communications and Networking Conference*, March 2005, vol. 1, pp. 12–17.

[14] A. van Zelst and T. C. W. Schenk, 'Implementation of a MIMO OFDM-based wireless LAN systems', *IEEE Trans. Signal Processing*, vol. 52, no. 2, Feb. 2004, pp. 483–494.

[15] E. Zhou, X Zhang, H. Zhao and W. Wang, 'Synchronization algorithms for MIMO OFDM systems', in *Proc. of IEEE Wireless and Communications and Networking Conference*, Mar. 2005, pp. 18–22.

[16] P. Priotti, 'Frequency synchronization of MIMO OFDM systems with frequency selective weighting', in *Proc. of IEEE Vehicular Technology Conference*, vol. 2, May 2004, pp. 1114–1118.

[17] C. Oberli and B. Daneshrad, 'Maximum likelihood tracking algorithms for MIMO–OFDM', in *Proc. of 2004 IEEE International Conference on Communications*, vol. 4, Jun. 2004, pp. 2468–2472.

[18] G. J. Foschini, 'Layered space–time architecture for wireless communication in a fading environment when using multiple antennas', *Bell Laboratories Technical Journal*, vol. 1, no. 2, Autumn 1996, pp. 41–59.

[19] P. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, 'V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel', in *Proc. of URSI International Symposium on Signals, Systems, and Electronics*, 1998, pp. 295–300.

[20] B. M. Hochwald and S. ten Brink, 'Achieving near-capacity on a multi-antenna channel', *IEEE Trans. Commun.*, vol. 53, Mar. 2003, pp. 389–399.

[21] H. S. Kim, J. Lee and S. C. Park, 'Complexity evaluation for MIMO sphere decoder with various tree-searching algorithms', in *Proc. of International Conference on Communication Technology*, Guilin, China, Nov. 2006, pp. 1–4.

[22] H. Artes, D. Seethaler and F. Hlawatsch, 'Efficient detection algorithms for MIMO channels: A geometrical approach to approximate ML detection', *IEEE Trans. Signal Processing*, vol. 51, no. 11, Nov. 2003, pp. 2808–2820.

Petitioner Samsung Ex-1026, 0153

8

From Algorithm to Bit-True Design

– by Ming-Luen Liou and Tzi-Dar Chiueh

In modern receivers, signal processing algorithms are usually executed in fixed-point arithmetic to save hardware costs as well as power consumption. Designers need to make compromises among cost, performance and hardware constraints of building blocks to obtain acceptable designs meeting the overall system performance specifications.

8.1 Design Flow Overview

System design is a series of mapping processes. As Figure 8.1 depicts, during these processes, a design starts from a clear system specification, and then descends gradually from higher levels to lower levels with more details. Checked points are specified between successive design phases to make sure that design consistency is always maintained and target specifications are met. If an intermediate design fails to meet the target functional/performance specifications or the system behaviors are not consistent with those from higher-level models, designers should revise the design for a better solution.

During the initial design phase, system algorithm is developed by constructing a floating-point functional model that meets the system specifications in both performance and functionalities. For receiver algorithm exploration, it is also necessary to analyze and model the channel effects, non-ideal characteristics and impairments. Even with functional models for all modules, the design is still inadequate because hardware-related information such as computation latency, throughput, precision and hardware complexity are still unclear. Owing to their high flexibility in system description and high design efficiency, programming languages such as C/C++ and software packages such as MATLAB are often used by the designers to construct system-level functional models. The choice of language also depends on the final implementation, which may be in hardware, software or both. Generally speaking, when the algorithm is quite sophisticated or subjected to change, and

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0154

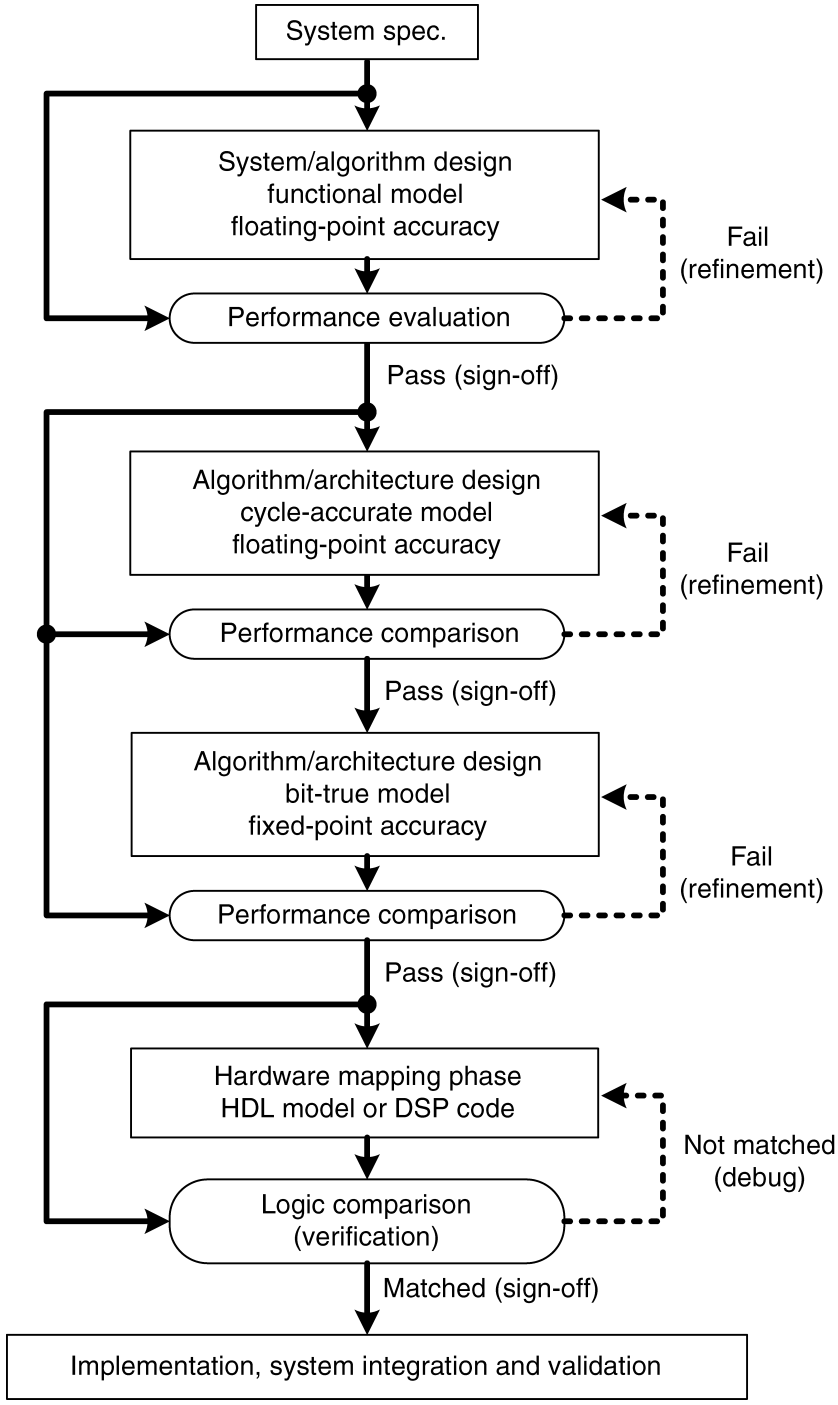


Figure 8.1. System design flow illustration

Petitioner Samsung Ex-1026, 0155

the processing delay/latency is not critical, software realization is more favorable, as system flexibility can be preserved.

When the functional model design is complete and the model behaves accurately, with its performance meeting the specifications, implementation-related information is annotated to the model. In the next design phase, the impact of signal processing latency and throughput upon the system behavior will be analyzed. For example, processing latency of modules in terms of clock cycles of synchronous systems must be estimated. After taking into account these delays, the designers may find that the system behavior has changed and target specifications are no longer met. When this happens, the designers need to evaluate approaches to reducing processing latency or even return to the previous design phase and adjust the system algorithm. At the end of this phase, the design is now said to be *cycle-accurate*.

After introducing the cycle-based processing delay, the design takes another step in hardware realization: converting from floating-point processing to fixed-point arithmetic.[1] Similarly, the design in terms of fixed-point arithmetic (so-called *bit-true design*) must be checked against the previous design to limit the performance degradation (due to the fixed-point arithmetic quantization effect) within an acceptable level. More details of the transformation will be illustrated in the following sections.

In the next phase, the task depends on how the design is to be implemented. The fixed-point arithmetic system model can be translated to the corresponding assembly code for firmware/DSP-based implementation, or translated to hardware description language (HDL) for dedicated IC or FPGA implementations. To implement the design in hardware, pipeline stages are inserted to reduce critical-path delay and registers and multiplexers are allocated to facilitate more resource sharing. Many techniques for signal processing hardware design can be found in other textbooks, such as [1]. Sometimes, these techniques may change processing latency, and thus may alter the system behavior. As always, the designers should ensure that the hardware has comparable behavior with the cycle-based fixed-point design. When it is difficult to maintain behavior consistency between the hardware model and the fixed-point arithmetic model, the designers can adjust the higher-level model to match the lower-level hardware design.

The system integration and verification task is launched once the hardware design is complete. For example, in ASIC design flow, the designers assemble functional blocks and hard/soft IPs, and perform a series of hardware synthesis, placement/routing and timing/power/testability analyzes that guarantee the functional correctness and manufacturing yield of the final design [2]. On the other hand, the designers may translate the hardware design to FPGAs, which provide fast prototyping for emulation and verification before actual silicon implementation. With this fast and flexible prototyping, the designers can evaluate system performance and reliability in more detail, and hopefully identify system level flaws/bugs before tape-out. In the case of wireless baseband receiver design, this process also helps to evaluate how well the adopted baseband channel model approximates the actual channel effect and the front-end electronics impairments.

Signal processing algorithms can be implemented in either floating-point or fixed-point arithmetic [3]. There are fewer constraints on the signal dynamic range and precision when using floating-point arithmetic, while the hardware cost is significantly higher than that of

[1]Comparatively speaking, the control of arithmetic precision is more flexible than processing latency/throughput, since the latter is usually constrained by the signal processing algorithm or the system architecture. Consequently, the timing model should be considered first to shorten the design cycle.

Petitioner Samsung Ex-1026, 0156

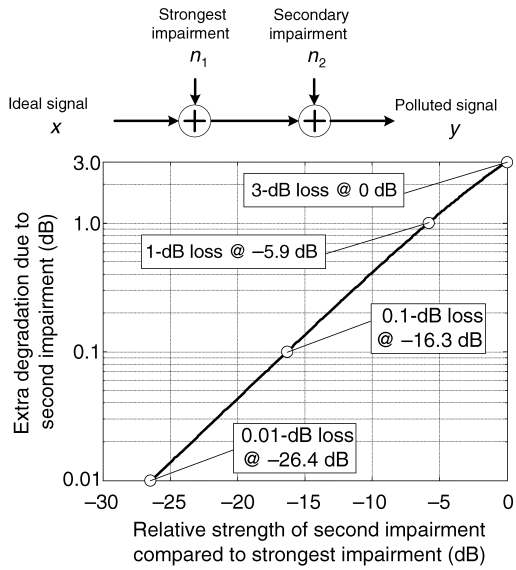


Figure 8.2. SINR degradation from a secondary impairment

the corresponding design using fixed-point arithmetic. In most cases, algorithms are first designed with high precision. After that, they will be compiled to fixed-point arithmetic on dedicated hardware (ASIC) or embedded processors to fit the implementation constraints such as hardware cost and/or power budget. This precision loss and non-trivial computation latency inevitably introduce errors that result in system characteristic change or even performance loss [3,4]. Hence, an important issue is to evaluate how the signal dynamic range and precision affect the performance of the system under development.

Throughout this chapter, the focus will be on how to convert a cycle-accurate floating-point arithmetic design to a corresponding bit-true design based on fixed-point arithmetic. The major issues include quantization effect by ADCs and its effect on the ensuing digital arithmetic operations in the receiver, also known as the *finite precision effect*. This effect includes ADC quantization/clipping effects, dynamic range setting, fixed-point arithmetic and the limit-cycle effect. In addition to describing these phenomena, the rest of this chapter will also cover fundamentals of the fixed-point design methodology in detail.

8.2 Effect of Additive Impairment Sources

In the algorithm design phase, the designers should consider signal-quality degradation caused by various impairment sources. These impairments usually appear in different forms, and their analysis is relatively simple if they can all be modelled as additive noise sources. As Figure 8.2 depicts, the ideal signal x is inflicted with two impairment sources, n_1 and n_2, where n_1 is stronger than n_2. The *signal-to-interference-and-noise-ratio* (SINR) of the polluted signal (y) is dominated by n_1, and the performance degradation caused by n_2 is determined by the relative power ratio between n_2 and n_1. For instance, if the power of n_2 is about 5.9 dB lower than that of n_1, then n_2 will cause approximately 1 dB extra performance

Petitioner Samsung Ex-1026, 0157

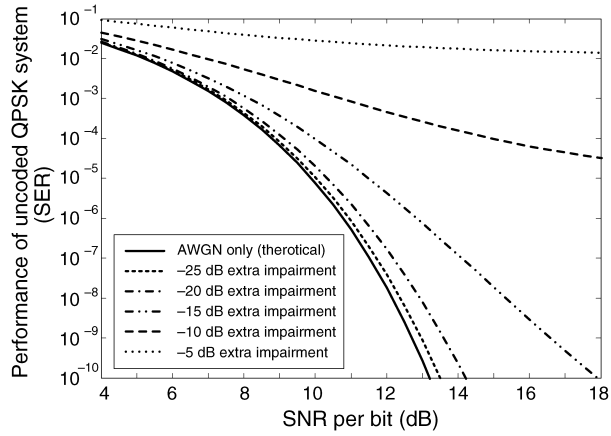


Figure 8.3. Symbol error rate degradation due to impairment with different power levels

degradation, while, if it is about 16.3 or 26.4 dB lower, there is only 0.1 or 0.01 dB extra degradation, respectively.

Figure 8.3 shows another example in which the error rate performance of a typical uncoded QPSK demodulator is considered. In addition to the thermal noise, an extra noise source with constant noise power is added. The solid curve shows the theoretical case in which there is only thermal noise. When adding the secondary noise, the symbol error rate (SER) curve moves away from the solid curve, and the loss in SNR increases with the secondary noise strength. Unlike the ideal SER that falls off as the SNR increases, the SER curve of the demodulator with an extra noise source has a lower bound (error floor).

The above two examples reveal how noisy/interfering effects may incur loss in SNR and set a limit in the system performance. Hence, these undesirable effects must be kept small relative to the noise power of the receiver operating range. In general, these secondary impairments may be from interference sources, analog front-end, estimation errors in synchronization and gain control loops, and the finite precision effect. The parameter estimator accuracy is usually determined by the estimation algorithms, while the finite precision effect can be reduced simply by increasing the signal word-lengths. To limit the SNR loss from the finite precision effect, the designers can perform simulations to derive a critical SNR at which the receiver meets the target specification. Based on the SNR and a predefined implementation loss constraint, one may derive an upper bound on the extra noise/impairment power.

Example:

Assume that an ideal demodulator works properly at SNR $= 20$ dB, and the implementation loss from ADC and the baseband receiver has to be controlled to under 1 dB. Thus, the total impairment from these modules should be 5.9 dB lower than the thermal noise, or equivalently $20 + 1 + 5.9 = 26.9$ dB lower than the signal. Similarly, if the implementation loss is limited to 0.1 dB, then the total impairment becomes $20 + 0.1 + 16.3 = 34.4$ dB lower than the signal. Compared with the 1-dB implementation loss case, this case requires 7.5-dB lower impairment power.

Petitioner Samsung Ex-1026, 0158

8.3 Analog-to-Digital Conversion

In a digital communication receiver, there is an important mixed-signal block: the analog-to-digital converter (ADC). The ADC acts as an interface between the analog and digital subsystems, and converts the continuous-time received signal $z(t)$ to the discrete-time received samples $z[n]$ for the baseband receiver. To assure that the baseband algorithms work properly, the receiver may control the dynamic range of $z(t)$ through a gain–control feedback loop.

8.3.1 ADC Distortions

The distortion of an ADC can be divided into two parts: *quantization noise* and *clipping noise*. An N-bit ADC divides its full-scale A into 2^N equal intervals, and converts a time sample $z(nT_s)$ residing in one interval to a corresponding N-bit digital code $z[n]$. The difference between $z(nT_s)$ and the middle point of the interval is called the *quantization error*. Ideally, the linear model of the quantization process is an additive uniformly distributed random noise in $(-0.5\Delta, +0.5\Delta)$, where $\Delta = A/2^N$ is usually called the least-significant bit (LSB) of the ADC. The precision of an ADC can be characterized by the *effective number of bits* (ENOB), which is defined as [5]:

$$\text{ENOB} = \left(10\log\frac{\text{average signal power}}{\text{noise and distortion power}} - 1.76\right)/6.02 \text{ bits}. \tag{8.1}$$

In an ADC, the input samples with values exceeding the maximum level are usually represented by the maximum level, which is known as the clipping level. The difference between $z(nT_S)$ and the clipping level is called the *clipping error*. The linear model of the clipping error is an additive impulsive noise source. Both quantization noise and clipping noise can be approximated as a white noise source when the number of quantization levels as well as the signal swing is large. In addition to the precision and full scale of an ADC, the range of the signal to be quantized also determines how often clipping occurs. An important metric of signal waveforms is crest factor (CF), which is defined as

$$\text{CF} = 20\log\frac{\text{peak signal amplitude}}{\text{RMS signal amplitude}}. \tag{8.2}$$

In quantising waveforms with high CF, larger headroom must be reserved to keep large samples from clipping. Moreover, for systems that have high SNR requirement (e.g. high-order modulation receivers), clipping probability must be kept low.

Consider the case in which the ADC input samples are Gaussian distributed with standard deviation σ, as shown in Figure 8.4, and set the clipping level at 2σ, namely CF of the ADC output is 6 dB. The samples within the interval $(-2\sigma, 2\sigma)$ contribute quantization errors, whereas the samples outside the interval contribute clipping errors. Obviously, the sample distribution strongly affects the power ratio of quantization error and clipping error. A proper ADC dynamic range setting should suppress the total power of the two error terms as much as possible. As Figure 8.5 depicts, clipping noise decreases as the input is attenuated so that

Petitioner Samsung Ex-1026, 0159

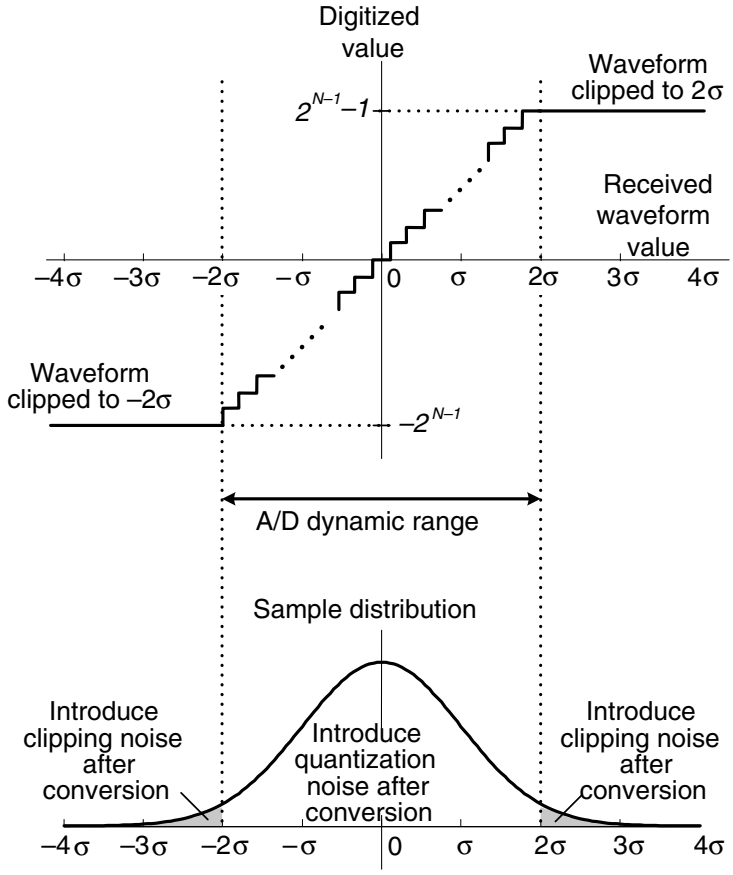


Figure 8.4. Dynamic range mapping of the ADC input signal

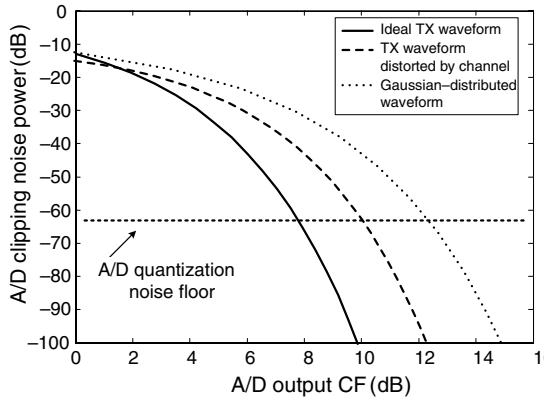


Figure 8.5. Relationship between ADC clipping noise and the CF setting

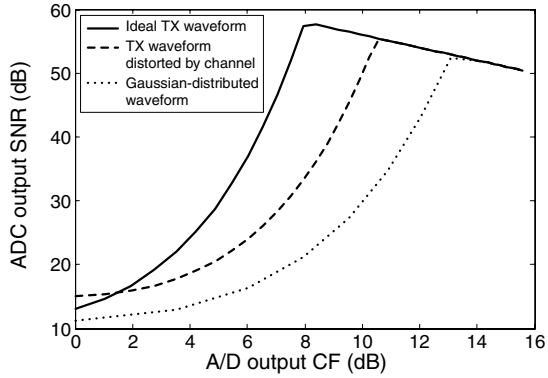


Figure 8.6. Relationship between ADC output SNR and the CF setting

the ADC output signal CF increases. As is easily seen from the figure, for the same level of clipping noise, TX samples are more 'concentrated' in values and thus have a lower CF than the Gaussian signal. For an ADC with 10-bit ENOB and Gaussian-distributed random input samples, if the design goal is to make clipping noise power close to the quantization noise floor, the ADC output CF should be set to about 12 dB. In this case, input signal samples ranging from -4σ to 4σ can be digitized without clipping.

Higher CF at ADC output can be attained by attenuating the signal level before entering the ADC. As Figure 8.6 depicts, the optimum CF setting that corresponds to the highest signal-to-conversion error ratio depends on the signal distribution. When the ADC output CF exceeds the optimum setting, a decrease in the clipping error cannot compensate for the performance loss due to signal power level reduction, thus making the ADC output SNR decrease. On the other hand, as the ADC output CF gets smaller and smaller than the optimum setting, the increase in average signal power is outstripped by the increasing clipping error, and again the ADC output SNR deteriorates.

8.3.2 Signal Probability Distributions

Signals can have very different amplitude probability distributions. From the above discussions, it is clear that the optimal CF setting of the ADC output depends on the amplitude distributions of the signals to be quantized. Figure 8.7 depicts the distributions of three typical signal sources. Samples of narrowband constant amplitude modulations (CAM) are sinusoidal-like, and their sample distributions are of the U-shaped distribution similar to the Doppler spectrum in Chapter 4. Samples of thermal noise and mixed signal sources have Gaussian distribution, which has much wider dynamic range (typically up to eight times the standard deviation σ) and hence much higher peak-to-average power ratio (PAPR). In comparison with these two signal distributions, the amplitude distribution of a typical single carrier modulation signal (such as a QAM signal) varies between the bath-tube and the Gaussian shape. Also, samples of a typical single-carrier modulation (SCM) signal have a PAPR between that of the CAM signals and that of Gaussian-distributed signals. Sample distributions of OFDM signals have higher PAPR and their amplitude distributions are approximately Gaussian. Therefore,

Petitioner Samsung Ex-1026, 0161

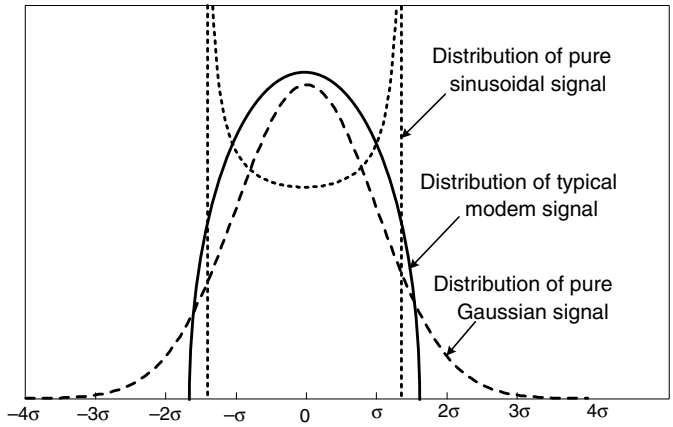


Figure 8.7. Sample distributions of various types of signal sources

the OFDM signals require a relatively large CF setting at the receiver ADC. Furthermore, in the situations in which transmitted signals undergo serious multipath fading channel with rich scattering, the received signal samples will also become Gaussian distributed.

Some signal distributions with multiple peaks are known as *multimodal distributions*. Distributions of some signals in baseband receivers are multimodal, such as the equalizer output signal. As an example, consider received signals that are 2-PAM modulated with signal levels a and $-a$. Each received sample is modelled as the sum of a transmitted signal and a Gaussian-distributed noise with variance σ^2 determined by the channel. Hence, the probability distribution of the received signal is given by the convolution of a Gaussian distribution with a two-impulse distribution that has peaks at a and $-a$, as shown in Figure 8.8. According to the characteristics of the Gaussian distribution, the range of the 2-PAM received samples can be approximated as $(-a - 4\sigma, a + 4\sigma)$.

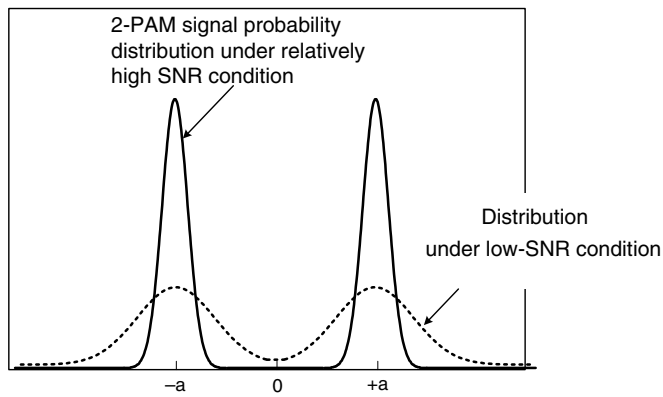


Figure 8.8. Probability distributions of two multimodal signals

Petitioner Samsung Ex-1026, 0162

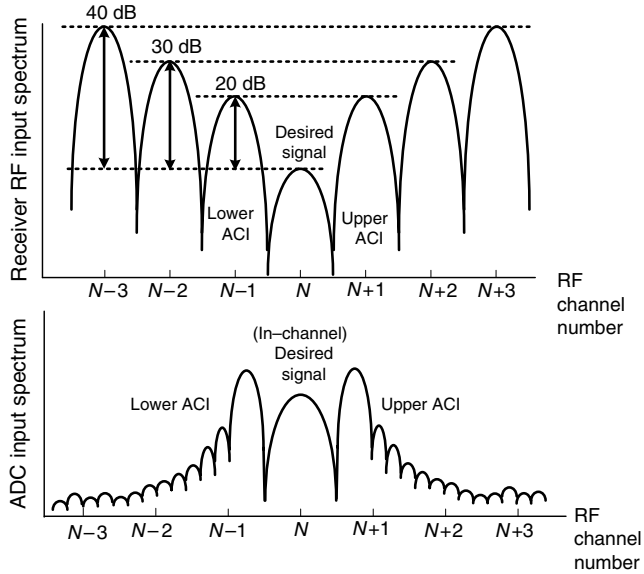


Figure 8.9. ACI conditions at the receiver front-end and ADC input

8.3.3 Dynamic Range and Precision Setting

In the early design phase of a wireless communication receiver, several decisions about the ADC must be made. The choice of conversion rate and number of bits of the ADC can have a significant influence on the power consumption, sensitivity and error rate performance of the receiver that is being developed. When determining the ADC precision specification and the dynamic range setting, several factors need to be taken into account.

(1) Input SNR for Proper Operation

The minimum input SNR at the receiver to achieve acceptable performance should be determined first. Based on this SNR requirement, the designers can estimate the relative strength between the thermal noise and the quantization/clipping noises and set an appropriate A/D precision accordingly.

(2) Signal Probability Distribution and PAPR

Probability distribution of the received signals can be characterized by statistical measures such as mean and variance. These measures and the PAPR of the signal are also major factors in the full scale of the ADC, as mentioned previously.

(3) Adjacent-Channel Interference (ACI)

In the wireless environment, the near–far problem makes adjacent-channel signal from a near transmitter overwhelm the desired signal from a transmitter much farther away. As shown in Figure 8.9, the power level of adjacent-channel signals can be tens of dB higher than the desired signal. In practice, the analog front-end and the baseband processor together

Petitioner Samsung Ex-1026, 0163

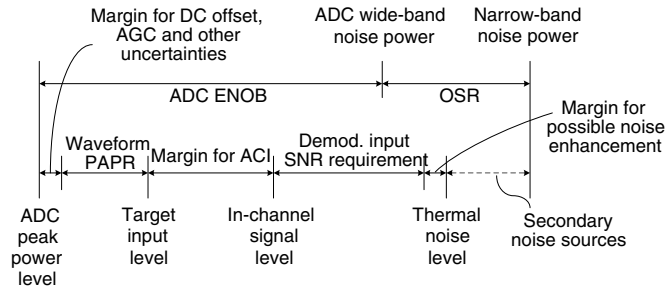


Figure 8.10. Dynamic range budget for ADC precision requirement estimation

complete the channel selection. After the front-end channel selection, strong ACI, though attenuated, still makes up a significant portion of the signal at the ADC input. In this case, the swing of the desired signal is small, and the quantization/clipping errors become quite significant. Therefore, higher ADC precision is required to limit the quantization/clipping errors to relatively smaller than the desired signal.

(4) Automatic Gain Control (AGC)

In most wireless receivers, the AGC block can track the radio signal power and set the gain accordingly to maintain the signal power level at the ADC input. However, the AGC is not perfect and the signal power level may change by too much, causing not enough headroom or too high a signal-to-quantization noise ratio. Hence, the designers must reserve some margin in the ADC dynamic range to allow for variations in the input signal power level.

(5) DC Offset

DC offset often occurs in a direct-conversion receiver structure. If it is not contained, the ADC clipping probability will increase. In most cases, this component is blocked by AC coupling at the ADC input. However, in the receivers using DC coupling, this ADC saturation due to DC offset must be considered and a suitable margin in the dynamic range is required.

Consider the dynamic range budget shown in Figure 8.10. To set an appropriate A/D specification, the designers analyze the strengths of the signal, interference and noise terms according to the items described above. The ADC precision requirement, in terms of ENOB, is normally set high enough so that the narrow-band (in-channel) noise power contributed by ADC is smaller than the front-end thermal noise. On the other hand, increasing the ADC sampling frequency provides a higher over-sampling ratio (OSR), which is an alternative choice for reducing the narrow-band noise power.

8.4 Finite Precision Effect in Digital Baseband Processing

Digital baseband processing mainly involves arithmetic functions such as addition/subtraction, multiplication and table look-up. In addition, FIR filters and IIR filters are also very useful. In the following, fixed-point data formats and error models of these functions will be investigated.

Petitioner Samsung Ex-1026, 0164

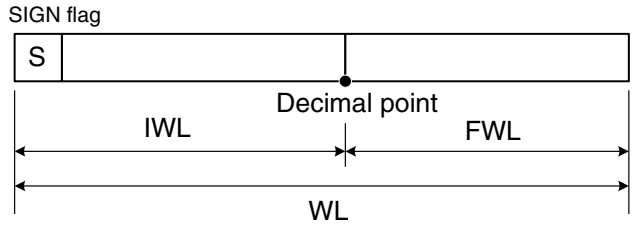


Figure 8.11. Word-length specification of a fixed-point number

8.4.1 Fixed-Point Data Format

A fixed-point number can be specified in a three-tuple [6], (WL, IWL, SIGN), in which WL and IWL denote the total word-length and the word-length for the integer part, respectively. The SIGN flag indicates whether a 2's complement number or an unsigned number is represented. A more elaborate representation uses a four-tuple that also specifies the quantization operation (truncation or rounding) [7]. Figure 8.11 shows the graphical illustration of the fixed-point format. Note that the word-length for the fraction part, FWL, is given by WL − IWL.

Although the above three-tuple specification of fixed-point number format is quite clear and simple, negative FWL is possible and can cause confusion. Hence, an alternative format specification will be adopted. In this specification, no decimal point is considered and only the locations of the most-significant bit (MSB) and the least-significant bit (LSB) in the number line are indicated:

$$\Xi(\mathrm{MSB}, \mathrm{LSB}). \tag{8.3}$$

The Ξ character can be U, which denotes the unsigned binary numbers, or S, which denotes the 2's complement numbers. MSB and LSB are both integers that can be either positive or negative. For the unsigned format, the fixed-point number range is $[0, 2^{\mathrm{MSB}+1} - 2^{\mathrm{LSB}}]$, and the minimum difference between the two nearest fixed-point numbers is 2^{LSB}. For the signed format, the most negative number is -2^{MSB}, and the most positive number is $2^{\mathrm{MSB}} - 2^{\mathrm{LSB}}$. The minimum difference is the same as the unsigned format.

8.4.2 Fixed-Point Error Model

In a fixed-point system, several different errors may be introduced during signal processing [3,4]. Three major types of errors and their effects are presented here.

(1) Truncation Error (Round-Off Error)

Truncation error is introduced when the fraction part of a fixed-point number is truncated. Generally speaking, truncation is inevitable, since signal-processing operations such as multiplications or filtering increase the output signal word-length. To keep the word-length from growing indefinitely, several LSBs in the fraction part can be discarded.

Petitioner Samsung Ex-1026, 0165

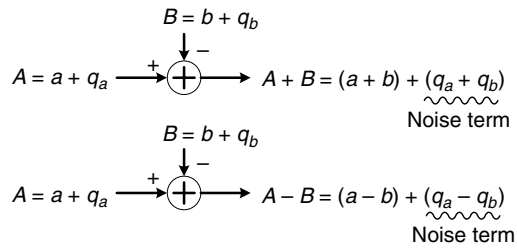


Figure 8.12. Propagation error model of a fixed-point adder/subtractor

(2) Clipping Error

Clipping error happens when the output of an arithmetic operation exceeds the range of the fixed-point representation. With proper range estimation and sufficient IWL setting, this kind of error can be avoided. However, in some cases, reserving sufficient IWL may lead to an over-conservative design with low hardware efficiency. Therefore, it is recommended to adopt clipping protection after some arithmetic operations to avoid overflow signals propagating further and crashing the whole design.

(3) Propagation Error

Error terms generated in arithmetic operations will propagate along signal/control paths of the design. Figure 8.12 illustrates the propagation error generated by a fixed-point adder/subtractor. Let A and B denote the two fixed-point variables; a and b denote the original floating-point signals; and q_a and q_b denote the truncation errors generated by earlier arithmetic operations. The output of the adder, $A + B$, can be represented as $(a + b) + (q_a + q_b)$, where $(a + b)$ is the signal term and $(q_a + q_b)$ is the propagation error term. The above formula can also be applied to the subtraction operation.

Example:

Consider the examples shown in Figure 8.13. Assume that the operands A and B are both in the format $S(+5, -5)$, the error terms q_a and q_b are of equal variance and uncorrelated to each other. After addition, the dynamic range of the result signal term increases by one bit, and the error power is doubled. In the second example, assume that the format of operands A and B are $S(+4, -4)$ and $S(+2, -7)$, respectively. The variance of q_a is much higher than that of q_b. In this case, it is deduced that the dynamic range of the sum grows by one bit from that of A, while the error term is also dominated by q_a. Hence, the precision of the adder output is close to that of the more significant operand A.

Figure 8.14 illustrates the propagation error generated by a fixed-point multiplier. Let A, B and C denote fixed-point variables; a and b denote the original signals; and q_a and q_b denote the truncation errors generated by earlier operations. The constant multiplier output, $A \times C$, can be represented as $a \times C + q_a \times C$, where $a \times C$ is the signal term and $q_a \times C$ is the

Petitioner Samsung Ex-1026, 0166

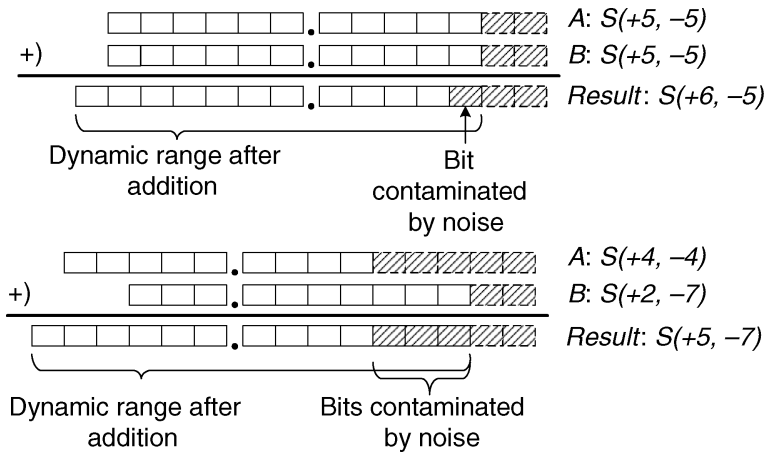


Figure 8.13. Binary addition of two fixed-point numbers

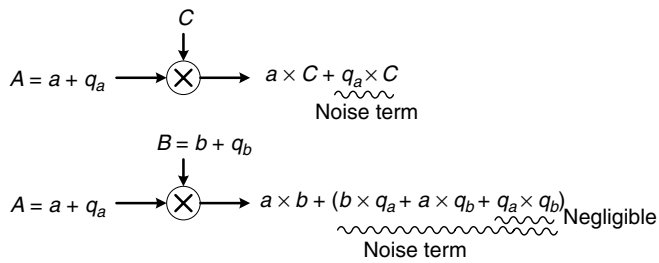


Figure 8.14. Propagation error model of fixed-point multipliers

propagation error term. On the other hand, the variable multiplier output, $A \times B$, can be approximated by $a \times b + b \times q_a + a \times q_b$, where $a \times b$ is the signal term and $b \times q_a + a \times q_b$ is the propagation error term.

Example:

Consider the example shown in Figure 8.15. Assume that the operands A and B are both in the format of $S(+2, -3)$, and that the error terms q_a and q_b are of equal variance and uncorrelated to each other. After multiplication, the dynamic range of the product increases by three bits, and the error term may grow by three bits as well. In the second example, assume that the format of operands A and B are $S(+2, -3)$ and $S(+2, -6)$, respectively. A and B are of equal variance, and the variance of q_a is much higher than that of q_b. In this case, the error term of the product is dominated by $b \times q_a$ and the signal-to-quantization noise ratio (SQNR) of the multiplier output will be close to that of operand A.

Petitioner Samsung Ex-1026, 0167

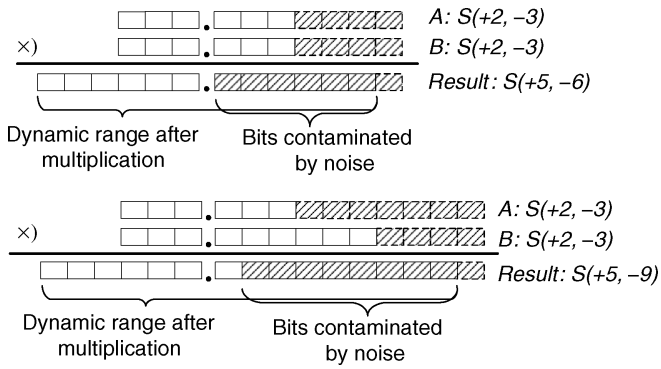


Figure 8.15. Binary multiplication of two fixed-point numbers

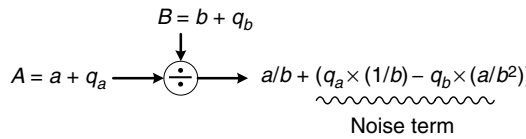


Figure 8.16. Propagation error model of a fixed-point divider

Figure 8.16 illustrates the propagation error generated by a fixed-point divider. Let A and B denote two fixed-point numbers; a and b denote the original floating-point value; and q_a and q_b denote the truncation errors generated by earlier operations. The divider output, A/B, can be approximated by $a/b + q_a/b - q_b \times a/b^2$, where a/b is the signal term and $q_a/b - q_b \times a/b^2$ is the propagation error term.

Special mathematical functions such as trigonometric and hyperbolic functions can be derived analytically or by iterative methods. If the precision requirement of the function is low, a general and practical implementation uses a look-up table (LUT). Figure 8.17 illustrates the propagation error generated by a fixed-point LUT. Assume that the mathematical function, $f(a)$, is realized by an LUT. For an input a, which has been quantized as $A = a + q_a$, there exist two independent error terms: $q_a \times f'(a)$ and q_{LUT}. The former term $q_a \times f'(a)$ depends on the LUT input word-length (address granularity) and the latter term q_{LUT} depends on the LUT output word-length. Since the LUT hardware cost scales exponentially with the input word-length and linearly with the output word-length, a more aggressive setting can be used for the input word-length.

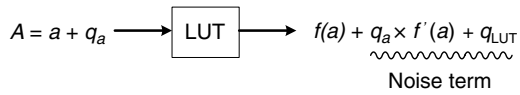


Figure 8.17. Propagation error model of a special mathematic function

Petitioner Samsung Ex-1026, 0168

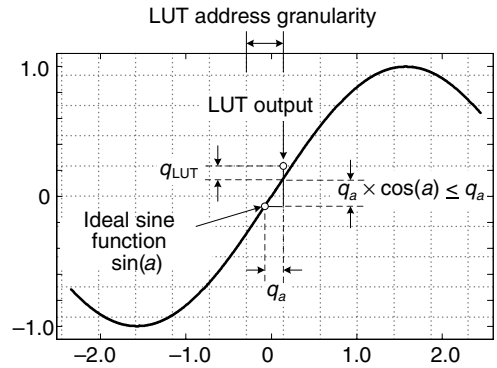


Figure 8.18. Error terms derivation in sine-function look-up table

Example:

Figure 8.18 shows an example of sine table generation. Since the absolute value of the first-order derivative of the sine function, $f'(a)$, is no greater than 1, the error of LUT is about

$$|E_{\text{PROP}}| \approx |q_a f'(a) + q_{\text{LUT}}| \leq |q_a| + |q_{\text{LUT}}|. \tag{8.4}$$

Thus, when the input is around 0, the quantization error is maximum, which is half the sum of the input LSB (Δ_{I}) and the output LSB (Δ_{O}). Similarly, for a cosine table, the maximum error occurs when the input is around $\pi/2$. If the design goal is to limit the error $|E_{\text{PROP}}|$ to under 10^{-3}, then the input word-length is minimized by letting $E_{\text{PROP}} = q_a = 10^{-3}$ and Δ_{I} becomes 2×10^{-3}. In this case,

$$\text{Input word-length} \geq \log_2(2\pi/(2 \times 10^{-3})) \approx 11.62. \tag{8.5}$$

Thus, the input word length is set to 12 bits (4096 equal levels in the range from 0 to 2π), and Δ_{O} is given by

$$2|q_{\text{LUT}}| < 2 \times 10^{-3} - (2\pi/4096) = 4.66 \times 10^{-4}. \tag{8.6}$$

Reserving two more bits for the sign and integer part of the output, then

$$\text{Output word-length} \geq 2 + \log_2\left|\frac{1}{2q_{\text{LUT}}}\right| = 13.06. \tag{8.7}$$

Finally, the output word length is set to 14 bits ($S(+1, -12)$).

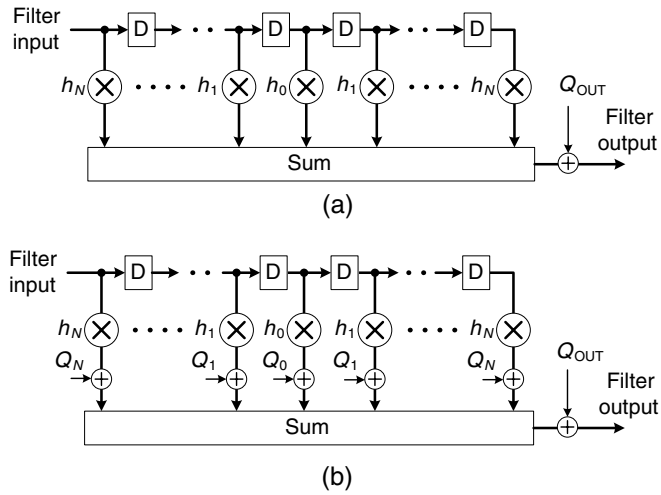


Figure 8.19. Truncation error of a direct-form FIR filter: (a) truncation only at the filter output and (b) truncation at each product term

8.4.3 Finite Precision Effect in FIR Filters

FIR filters are widespread in many signal processing systems. They are also popular in digital baseband receivers owing to their structural simplicity/regularity, unconditional stability and low sensitivity to filter coefficient errors. Besides being used for anti-aliasing, pulse shaping, interference rejection and parameter smoothing, FIR filters can also be used for sample rate conversion and equalization. For a causal N-tap FIR filter, the output sample can be represented as

$$y[n] = h_0 x[n] + h_1 x[n-1] + \cdots + h_{N-1} x[n-N+1], \tag{8.8}$$

where $h_0, h_1, \ldots, h_{N-1}$ are the filter coefficients and $x[n], x[n-1], .., x[n-N+1]$ are the time-domain input samples at time index n to $n-N+1$. If the filter coefficients are symmetric or anti-symmetric about time $(N-1)/2$, then the filter has a linear phase response and all frequency components have the same delay when passing through this filter.

Consider a $(2N+1)$-tap direct-form linear-phase FIR filter with symmetric coefficients $h_0, h_1, \ldots, h_N$. In application-specific integrated circuit (ASIC) implementation, the precisions of internal arithmetic operations in a filter are not constrained, while in processor-based implementation, the precision of each multiplication output is constrained. In the former case, the designer must estimate the propagation error at the FIR filter output and then determine the output word-length with a truncation error relatively smaller than the propagation error. With this arrangement, an extra error term, Q_{OUT}, at the filter output can be used to model the filter output truncation error, as shown in Figure 8.19(a). In the latter case, in which the processor-based implementation is concerned, each multiplication introduces an extra truncation noise, as shown in Figure 8.19(b). By the central limit theorem, the overall truncation noise can be modelled as a Gaussian noise, with variance given by the sum of individual variances.

Petitioner Samsung Ex-1026, 0170

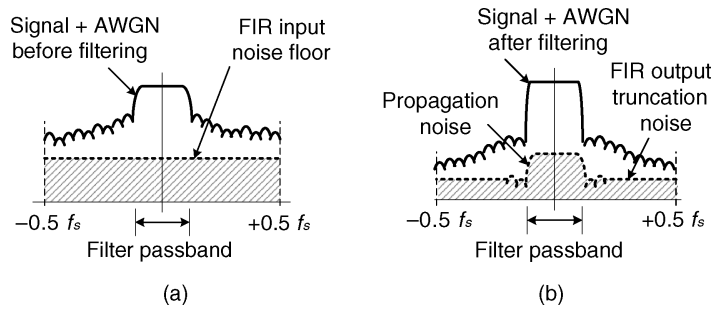


Figure 8.20. Illustration of the noise floor changes due to filtering (a) noise before filtering and (b) noise after filtering

The filter output error can be approximated by the noise filtering process illustrated in Figure 8.20. Assume that the filter input contains a white noise process with variance Q_{IN}, and the filter performs word-length truncation only after the final summation. Then, the noise component at the filter output is given by

$$Q_{\text{TOTAL}} = Q_{\text{OUT}} + Q_{\text{IN}} \times (h_0^2 + 2h_1^2 + 2h_2^2 + \cdots + 2h_N^2). \tag{8.9}$$

Note that the filtering operation suppresses propagation error components inside the filter stopband; thus, the propagation error is no longer white after the filtering operation.

It should be noted that transforming the filter coefficients into a fixed-point format does not introduce datapath quantization noise. Since the filter coefficients are usually fixed, coefficient quantization only changes the filter frequency response, such as increasing passband ripples, reducing stopband attenuation and changing the locations of filter notches. In particular, narrow-band band-pass filters and band-reject filters have high sensitivity to filter coefficients; hence, their response must be double checked after coefficient quantization.

8.4.4 Finite Precision Effect in IIR Filters

IIR filtering is another basic signal processing operation. The time domain input–output relationship of an IIR filter takes the form of

$$y[n] = a_1 y[n-1] + \cdots + a_{M-1} y[n-M+1] + b_0 x[n] + b_1 x[n-1] + \cdots + b_{N-1} x[n-N+1], \tag{8.10}$$

where $a_1 \ldots a_{M-1}$ are recursive-part coefficients; $b_0 \ldots b_{N-1}$ are non-recursive-part coefficients; $x[n] \ldots x[n-N+1]$ are the input samples; and $y[n] \ldots y[n-M+1]$ are previous output samples. IIR filters can achieve very high-frequency selectivity with much smaller filter order than the FIR filters. In digital baseband receivers, IIR filters are widely applied to parameter estimation, tone suppression (e.g. DC removal) and channel selection function.[2] On the other hand, IIR filter design is

[2]It should be noted that IIR filtering is not widely applied to baseband datapaths due to its non-constant group-delay nature. A non-constant group delay property usually introduces an ISI that increases the loading of time-domain equalizers or consumes the guard-interval margin in an OFDM receiver.

Petitioner Samsung Ex-1026, 0171

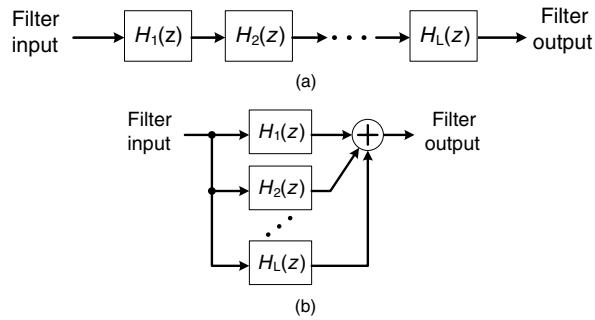


Figure 8.21. Practical IIR filter implementation: (a) cascade form and (b) parallel (poly-phase) form, where $H_1(z), H_2(z), \ldots, H_L(z)$ are first-order or second-order filters

more challenging than FIR filter design, due to its much higher sensitivity to coefficient quantization and arithmetic precision. To mitigate the vulnerability, high-order IIR filters are often decomposed to a cascade form or a parallel (poly-phase) form, as shown in Figure 8.21.

8.5 Conversion from Floating-Point Design to Bit-True Design

When developing the floating-point design, there are several performance evaluation metrics. The first step in the conversion to bit-true design is to replace the floating-point input by quantized input and check the metrics to evaluate performance, as illustrated in Figure 8.22. Generally speaking, this 'quantized-input' behavior model includes an optional scalar, a quantizer model and the original floating-point model. The optional scalar is to adjust the magnitude of the input signal to fit the dynamic range of the ensuing circuit. The quantizer model decides the precision and the clipping (overflow) behavior of the signal. The quantization can be carried out by flooring (discarding less-significant bits) or rounding operation, while the clipping can be realized by saturation or wrapping (discarding overflow bits). This quantized-input model is useful for determining the word-length of the ADC in the baseband receiver. The procedure can be applied recursively to get more signals inside the receiver represented in the fixed-point format and, at the end of this process, the corresponding bit-true design is obtained.

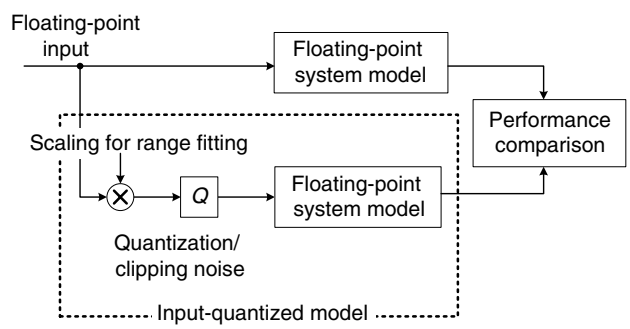


Figure 8.22. Input-quantized model and its performance evaluation

Petitioner Samsung Ex-1026, 0172

8.5.1 Metrics for Performance Evaluation

There are many performance metrics that can qualify a baseband receiver design, including (1) waveform difference, (2) signal-to-noise ratio, (3) error rate and (4) estimation accuracy. Waveform difference can be used to measure how precise a quantized design is when compared to its floating-point counterpart. The design can be as small as a simple module or as large as a whole baseband receiver. The metric is especially useful during the development of an individual module, when the whole system model is not available. On the other hand, when the whole receiver design is available, the output signal-to-noise ratio (SNR) of the receiver is an effective measure of its communication performance.

To shed more light on the quality of the receiver, coded or uncoded error rates are often evaluated and compared to specifications from baseline designs in the standard documents or other competitors. The drawback in this metric is that to reach the level of a meaningful coded error rate, millions or tens of millions of samples must be simulated for just one outcome. In terms of simulation complexity, this metric is much more expensive than the others. Parameter estimation error and variance can also work as excellent indicators of system performance. They include carrier frequency offset accuracy, timing offset accuracy, channel estimation error and others. In most real-time receivers, the elapsed time from algorithm start to estimation completion, sometimes called acquisition time, is also an important system characteristic frequently used for design qualification.

The metrics described above are all important from the system perspective. Sometimes, they are sensitive to the signal word-length, while other times they are not. Referring to the first example depicted in Figure 8.23, note that the first metric changes smoothly around the n-bit word-length and approaches the floating-point performance in higher word-length settings. This metric, such as the output SNR of a module, is only moderately sensitive to the signal word-length. Based on this trend, the designer can set the signal word-length one or two bits greater than the word-length at which the metric starts to saturate.

The second performance metric, as shown at the bottom of Figure 8.23, has a higher sensitivity to changes in the word-length. Note that the metric drops rapidly as soon as the signal word-length passes a threshold, called the knee point. Although the second metric seems to work as a clear design guideline, it may not tell the whole story. For example, the error-rate metric curve may have such a sharp fall-off only because that error correction coding has been included in the evaluation. Consequently, it is recommended that one should employ more than one performance metric to confirm word-length settings.

8.5.2 Interpolative Design Flow

In principle, one can derive the relationship between a performance metric and the word-length of any variable and then determine the word-length of each variable accordingly. However, this is not feasible, since a baseband receiver may contain thousands of variables. In all likelihood, only a small portion of these variables are critical to the system performance. On the other hand, the previous discussions indicate that the noise power of an arithmetic operator can be computed from the noise powers of its inputs and the property of that operation itself. Hence, it seems that proper word-lengths for all signals can be derived in a hierarchical manner. Indeed, a hierarchical design scheme, known as an interpolative design approach, has been proposed

Petitioner Samsung Ex-1026, 0173

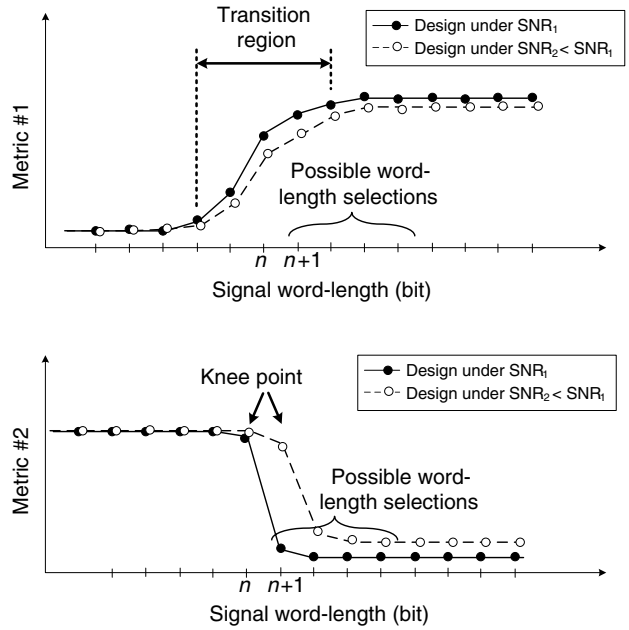


Figure 8.23. Performance characteristics used for signal word-length selection

and adopted in an interactive word-length optimization system called the Fixed-point pRogrammIng DesiGn Environment (FRIDGE) [8,9].

Consider the bit-true design scheme shown in Figure 8.24. Critical signals, such as module I/O signals, the signals whose word-lengths are specified and the signals whose word-lengths greatly affect system performance and/or hardware complexity, will be designed first. In

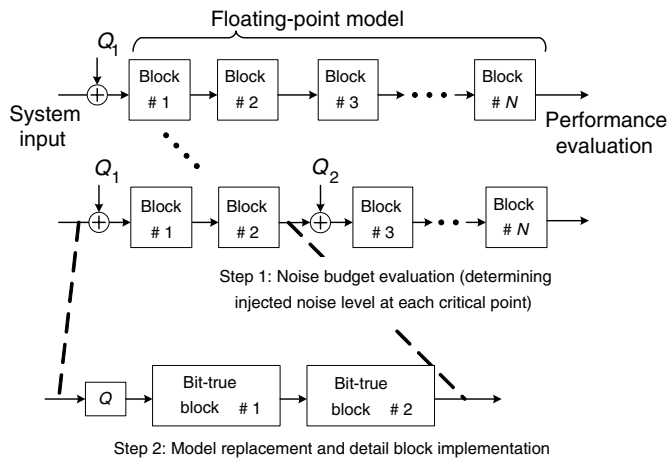


Figure 8.24. Interpolative approach for fixed-point design conversion

order of their importance, the scheme determines the noise power of each critical signal. With all critical signals' word-lengths determined, one can then associate these word-lengths with their corresponding signals. This extra information valuable to bit-true design is called *local annotation*. After all critical signals are locally annotated, the designer should perform a detail simulation to ensure that the partially bit-true design meets the target specifications.

In the next stage, the remaining signals (e.g. module internal variables) can then be transformed into fixed-point representations. This design stage is called *interpolation*, suggesting that the word-lengths of non-annotated variables will be interpolated from the local annotations of the critical signals around them. The main operation, called *format propagation*, derives the word-lengths of signals along the signal flow graph of the datapath. The FRIDGE system uses a type of interval-arithmetic (IA) to analyze the signal dynamic range, and then determines the integer word-length of each variable [8,9]. The range analysis is carried out at the program compile time to minimize the overhead of overflow casting during the simulation phase. In order to determine the fractional word-length of each variable, this system performs bit-true simulations. Moreover, to meet some hardware constraints, such as the operand word-length of a processor, the designer can assign a word-length upper bound for all variables. Finally, the system will generate a fully annotated bit-true design with specifications about the word-lengths of all signals.

The signal word-length can also be manually derived instead of resorting to an automated system. In this case, the designer should always keep the sum of noise powers (introduced by word-length truncation) under the predetermined noise budget. To reduce the design effort, it is recommended to minimize word-length truncation for internal variables and leave most of the noise budget for module I/O signals.

8.5.3 Simulation-Based Approaches

To convert signals into fixed-point format, there are two things that must be determined: integer word-length (by range analysis) and fractional word-length (by precision analysis). Both tasks can be performed by either analytical- or simulation-based approaches [8,10–12]. In simulation-based approaches, the bit-true design process is regarded as an optimization problem. A possible search-based design flow is sketched in Figure 8.25 [12]. Simply put, the designers first determine a set of test samples, called a *testbench*, and try them on the floating-point design to get a baseline performance for comparison. The he/she performs a heuristic search in the design space to find out a word-length setting that minimizes the predefined cost function (such as the hardware cost or power consumption) while simultaneously meeting the performance requirement. The performance requirement, in this case, is to keep the performance degradation of the bit-true design to within a certain threshold.

Since the design space is extremely large, a full-search strategy is not feasible. Almost always, efficient sequential searching strategies are adopted to reduce the searching time. In [10–12,13], several different sequential searching algorithms with various search times and searching qualities are proposed.

8.5.4 Analytical Approaches

In analytical approaches, a large testbench is not needed. These approaches start with evaluating the strength of the input signals and truncation noises, tracing their propagation

Petitioner Samsung Ex-1026, 0175

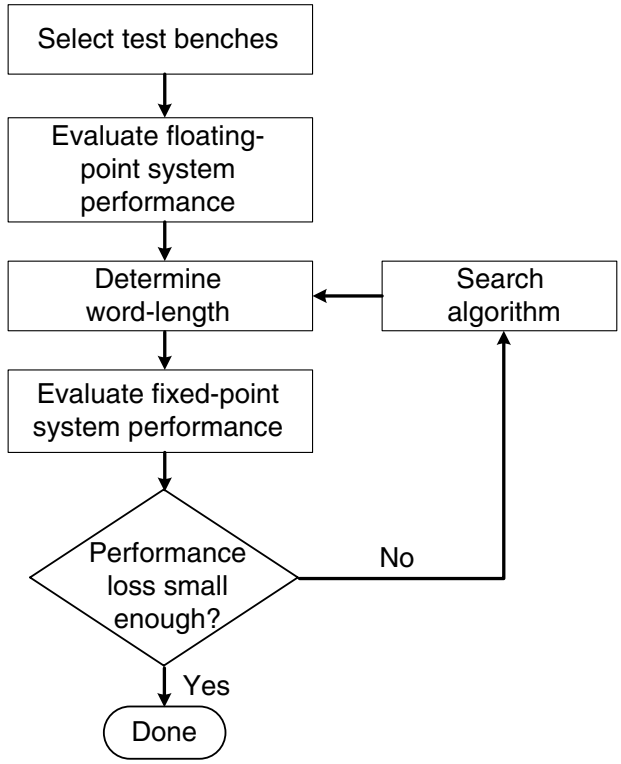


Figure 8.25. Illustration of a search-based approach for fixed-point design conversion

(propagate the source statistics along the datapath), and then determining signal word-length accordingly. It has been proposed that interval arithmetic (IA) [14,15] and affine arithmetic (AA) [16] be used in range analysis.

IA defines a set of range propagation rules for datapath operators. Following these rules, one may estimate the signal dynamic range along the datapath. However, these methods assume no data dependency, and thus the signal dynamic range estimation tends to be too pessimistic after several range propagation steps. AA is a refinement on IA and it preserves the correlation information between operands during arithmetic operations. When the variables are combined, uncertainty may be cancelled due to such correlation, and the signal dynamic range may even shrink instead of continuously growing in the IA case. Consequently, the AA approaches often yield a better estimation of signal dynamic range and noise level.

For precision analysis, there also exist several proposals, such as perturbation method and automatic differentiation [17–19]. Both of these two methods provide implementation that can evaluate noise power propagation along the datapath. Actually, many solutions to bit-true design employ both analytical approaches and simulation-based approaches. Very likely, an analytical approach may be used in range analysis, whereas extensive simulation is run to determine the signal precision setting.

Petitioner Samsung Ex-1026, 0176

Bibliography

[1] Keshab K. Parhi, *VLSI Digital Signal Processing Systems: Design and Implementation*. New York: John Wiley & Son, 1999.

[2] M. Keating and P. Bricaud, *Reuse Methodology Manual: For System-on-a-Chip Designs*, 3rd ed. New York: Kluwer Academic Publishers, 2002.

[3] D. Menard, R. Rocher, P. Scalart and O. Sentieys, 'Automatic SQNR determination in non-linear and non-recursive fixed-point systems', in *Proc. of Design, Automation and Test in Europe Conference and Exhibition*, Sep. 2004, pp. 1349–1352.

[4] A. V. Oppenheim and C. J. Weinstein, 'Effects of finite register length in digital filtering and the fast Fourier transform', *Proc. IEEE*, vol. 60, Aug. 1972, pp. 957–976.

[5] A. Moscovici, *High-Speed A/D Converters: Understanding Data Converters through SPICE*. New York: Kluwer Academic Publishers, 2001.

[6] W. Sung and K.-I. Kum, 'Simulation-based word-length optimization method for fixed-point digital signal processing systems', *IEEE Trans. Signal Processing*, Dec. 1995, pp. 3087–3090.

[7] N. Doi, T. Horiyama, M. Nakanishi and S. Kimura, 'Minimization of fractional wordlength on fixed-point conversion for high-level synthesis', in *Proc. of Asia and South Pacific Design Automation Conference*, Jan. 2004, pp. 80–85.

[8] M. Coors, H. Keding, O. Luthje and H. Meyr, 'Design and DSP implementation of fixed-point systems', *EURSASIP J. Applied Signal Processing*, vol. 2002, no. 9, pp. 908–925.

[9] H. Keding, M. Willems, M. Coors and H. Meyr, 'FRIDGE: A fixed-point design and simulation environment', in *Proc. Design Automation and Test in Europe*, 1998, pp. 429–435.

[10] H. Choi and W. P. Burleson, 'Search-based wordlength optimization for VLSI/DSP synthesis', in *Workshop on VLSI Signal Processing*, VII, Oct. 1994, pp. 198–207.

[11] K. Han, I. Eo, K. Kim and H. Cho, 'Numerical word-length optimization for CDMA demodulator', in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 4, May 2001, pp. 290–293.

[12] M. A. Cantin, Y. Savaria and P. Lavoie, 'A comparison of automatic word length optimization procedures', in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 2, May 2002, pp. 612–615.

[13] H. H. Hoos and T. Stutzle, *Stochastic Local Search: Foundations and Applications*. San Fransisco, CA: Morgan Kaufmann, 2005.

[14] R. Cmar, L. Rijnders, P. Schaumont, S. Vernalde and I. Bolsens, 'A methodology and design environment for DSP ASIC fixed point refinement', in *Proc. of Design, Automation and Test in Europe Conference and Exhibition*, Mar. 1999, pp. 271–276.

[15] J. A. Lopez, C. Carreras, G. Caffarena and O. Nieto-Taladriz, 'Fasts characterization of the noise bounds derived from coefficient and signal quantization', in *Proc. of IEEE International Symposium on Circuits and Systems*, vol. 4, May 2003, pp. 309–312.

[16] C. F. Fang, R. Rutenbar and T. Chen, 'Fast, accurate static analysis for fixed-point finite-precision effects in DSP designs', in *Proc. IEEE/ACM International Conference on Computer Aided Design*, Nov. 2003, pp. 275–282.

[17] D. Lee, A. Abdul Gaffar, R. C. C. Cheung, O. Mencer, W. Luk and G. A. Constantinides, 'Accuracy guaranteed bit-width optimization', *IEEE Trans. Computer-Aided Design*, vol. 25, Oct. 2006, pp. 1990–2000,

[18] G. A. Constantinides, 'Perturbation analysis for word-length optimization', in *Proc. of IEEE Symposium on Field-Programmable Custom Computing Machines*, Apr. 2003, pp. 81–90.

[19] A. A. Gaffer, O. Mencer, W. Luk and P. Y. K. Cheung, 'Unifying bit width optimization for fixed point and floating point designs by using automatic differentiation', in *Proc. of IEEE Symposium on Field-Programmable Custom Computing Machines*, 2004, pp. 79–88.

Petitioner Samsung Ex-1026, 0177

9

Circuit Techniques

Many types of signal-processing circuits with diverse functionalities are needed in an OFDM baseband receiver IC. Low power and low complexity are the major concerns in the design of these circuits.

9.1 Introduction

Portable wireless communication devices have experienced explosive growth recently. Consumers now demand longer and longer battery operation time in those portable devices. As a result, low-power and low-complexity circuit design for wireless communication transceivers has become crucial. In most orthogonal frequency-division multiplexing (OFDM) systems, several functional units are indispensable and play an essential role. They include fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) modules, delay lines (buffers), polar/rectangular coordinate conversion functions, numerical controlled oscillators (NCO), etc.

In OFDM systems, FFT and IFFT modules transform signals between the time domain and the frequency domain and occupy a large portion of the circuit area as well as the power consumption. Delay lines (buffers) also contribute a significant portion of area and power in the OFDM baseband receiver. Usually, they come in different memory depths and are used for data scheduling and temporary storage. OFDM synchronization often relies on the phase or magnitude of received complex signals; thus, arctangent and magnitude functions, which convert data from rectangular coordinates to polar coordinates, are also commonly seen in OFDM receivers. On the other hand, conversion from polar coordinates to rectangular coordinates is also needed for generating sinusoidal signals. In light of their importance, in the rest of this chapter, circuit design techniques of these functional units will be discussed and examined.

9.2 FFT

Recent advances in semiconductor processing technology has enabled dedicated FFT processors in applications such as telecommunications, specifically in OFDM communication systems. Thanks to efficient and low-power VLSI implementation of FFT processors,

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0178

OFDM has become a popular modulation technique for current and future wireless communication. Conventionally, direct implementation of N-point discrete Fourier transform (DFT) requires a complexity that is $O(N^2)$. The Cooley–Tukey fast Fourier transform algorithm achieves a huge complexity saving over direct implementation and requires only $O(N \log N)$ computations. Even with the reduction in complexity, circuit implementation of the Cooley–Tukey FFT algorithm is still both computation and communication intensive.

In the following, FFT algorithms are first introduced and then the FFT hardware architectures will be described. Comparisons and analysis of these methods will be made. Finally, their pros and cons and applications that they are most suitable for will be discussed.

9.2.1 FFT Algorithms

The N-point DFT is formulated as

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{nk}, \tag{9.1}$$

where $k = 0, 1, 2, \cdots, N-1$ and W_N is given by

$$W_N = e^{-j2\pi/N}. \tag{9.2}$$

Direct implementation of the above formula requires $O(N^2)$ arithmetic complexity. Cooley and Tukey proposed an FFT algorithm that decomposes the N-point DFT into recursive 2-point DFT operations, known as the radix-2 FFT. Subsequently, higher-radix FFT algorithms using larger-size DFT are suggested. Usually, FFT algorithms can be categorized into two types: decimation-in-time FFT and decimation-in-frequency FFT. In decimation-in-time FFT algorithms, the time-domain samples are not processed sequentially, while, in decimation-in-frequency FFT algorithms, the frequency-domain samples are generated in a non-sequential order, with increasing bit-reversed addresses. In the following, both decimation-in-time and decimation-in-frequency FFT will be treated and FFT algorithms of different radixes will also be introduced.

Radix-2 Algorithm

The basic concept underlying the radix-2 FFT algorithm is the use of symmetry between W_N^{nk} and $W_N^{nk+N/2}$.

Decimation-in-Time FFT

In decimation-in-time FFT, time-domain signals $x[n]$, $n = 0, 1, \cdots, N-1$, are first partitioned into even-numbered samples and odd-numbered samples. Then,

$$X[k] = \sum_{m=0}^{N/2-1} x[2m] W_N^{2mk} + \sum_{m=0}^{N/2-1} x[2m+1] W_N^{(2m+1)k}, \quad k = 0, 1, \cdots, N-1. \tag{9.3}$$

Petitioner Samsung Ex-1026, 0179

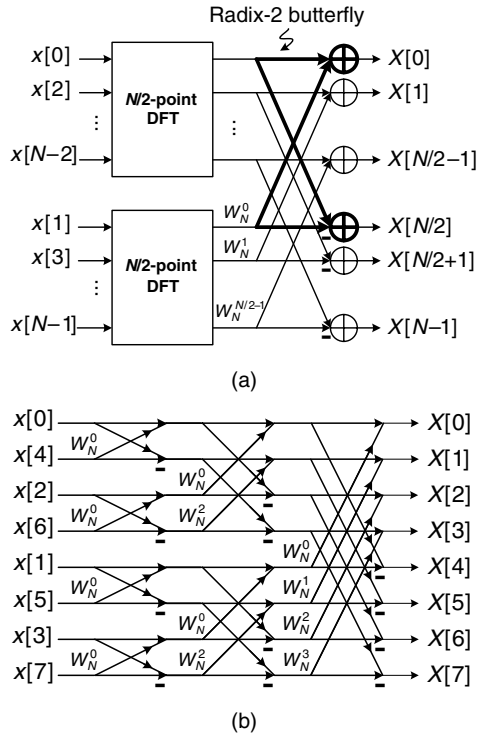


Figure 9.1. (a) Signal flow graph of radix-2 decimation-in-time decomposition and (b) its application in 8-point DFT computation

Because $W_N^{2mk} = W_{N/2}^{mk}$, for the first half of $X[k]$, $0 \le k < N/2$, Equation (9.3) becomes

$$X[k] = \sum_{m=0}^{N/2-1} x[2m]W_{N/2}^{mk} + W_N^k \sum_{m=0}^{N/2-1} x[2m+1]W_{N/2}^{mk}, \quad k = 0, 1, \cdots, N/2-1. \tag{9.4}$$

Owing to the symmetry that $W_N^{k+\frac{N}{2}} = -W_N^k$, the second half can be given by

$$\begin{aligned} X\left[k+\frac{N}{2}\right] &= \sum_{m=0}^{N/2-1} x[2m]W_{N/2}^{m(k+\frac{N}{2})} + W_N^{k+\frac{N}{2}} \sum_{m=0}^{N/2-1} x[2m+1]W_{N/2}^{m(k+\frac{N}{2})} \\ &= \sum_{m=0}^{N/2-1} x[2m]W_{N/2}^{mk} - W_N^k \sum_{m=0}^{N/2-1} x[2m+1]W_{N/2}^{mk}, \quad k = 0, 1, \cdots N/2-1. \end{aligned} \tag{9.5}$$

Hence, the N-point DFT can be regarded as a combination of $N/2$-point DFT of the even-numbered time-domain samples $x[2m]$ and $N/2$-point DFT of the odd-numbered time-domain sample $x[2m+1]$ multiplied by W_N^k, as shown in Figure 9.1(a). In addition, the basic arithmetic module called the'butterfly' operation, with one addition and one subtraction, is indicated. Similarly, one $(N/2)$-point DFT can be further decomposed into two $(N/4)$-point

Petitioner Samsung Ex-1026, 0180

DFTs and so on, until 2-point DFTs. If $N = 2^v$, the decomposition procedure can be applied v times. Figure 9.1(b) shows the signal flow graph of an 8-point radix-2 decimation-in-time FFT algorithm. It is worth noting that as its name indicates, the time-domain samples are not arranged in normal order, but in a bit-reversed-addressing order.

Decimation-in-Frequency FFT

For the decimation-in-frequency FFT, consider only even-numbered frequency-domain samples; then

$$X[2m] = \sum_{n=0}^{N-1} x[n] W_N^{n(2m)}, \quad m = 0, 1, 2, \cdots, N/2 - 1. \tag{9.6}$$

By partitioning the sum into two halves, the above equation takes the form of

$$X[2m] = \sum_{n=0}^{N/2-1} x[n] W_N^{n(2m)} + \sum_{n=N/2}^{N-1} x[n] W_N^{n(2m)} = \sum_{n=0}^{N/2-1} x[n] W_N^{n(2m)} + \sum_{n=0}^{N/2-1} x[n+N/2] W_N^{(n+\frac{N}{2})(2m)}. \tag{9.7}$$

Because $W_N^{(n+\frac{N}{2})(2m)} = W_N^{2mn} = W_{N/2}^{mn}$, Equation (9.7) becomes

$$X[2m] = \sum_{n=0}^{N/2-1} (x[n] + x[n+N/2]) W_{N/2}^{nm}, \quad m = 0, 1, 2, \cdots, N/2 - 1. \tag{9.8}$$

From Equation (9.8), it is seen that all $X[2m]$ can be computed from a $(N/2)$-point DFT of the sequence $x[n] + x[n+N/2]$. Similarly, the odd-numbered frequency-domain samples are given by

$$\begin{aligned} X[2m+1] &= \sum_{n=0}^{N/2-1} x[n] W_N^{n(2m+1)} + \sum_{n=0}^{N/2-1} x[n+N/2] W_N^{(n+\frac{N}{2})(2m+1)} \\ &= \sum_{n=0}^{N/2-1} (x[n] - x[n+N/2]) W_N^{2nm+n} \\ &= \sum_{n=0}^{N/2-1} (x[n] - x[n+N/2]) W_N^{n} W_{N/2}^{nm}, \quad m = 0, 1, \cdots, N/2 - 1. \end{aligned} \tag{9.9}$$

Figure 9.2(a) shows the *decimation-in-frequency* signal flow graph that implements Equations (9.8) and (9.9).

Similarly, if N is a power of 2, this concept can be applied recursively up to 2-point DFTs. Figure 9.2(b) illustrates the signal flow graph of one such example, with $N = 2^3$. Note that the frequency-domain samples are in a bit-reverse order.

It is clear that the decimation-in-time the decimation-in-frequency algorithms, and have the same arithmetic complexity. There are $v = \log_2 N$ stages and each stage has N complex multiplications and N complex additions. Consequently, the radix-2 FFT algorithm has $O(N \log_2 N)$ computational complexity. Since the complexity is identical in the two types of

Petitioner Samsung Ex-1026, 0181

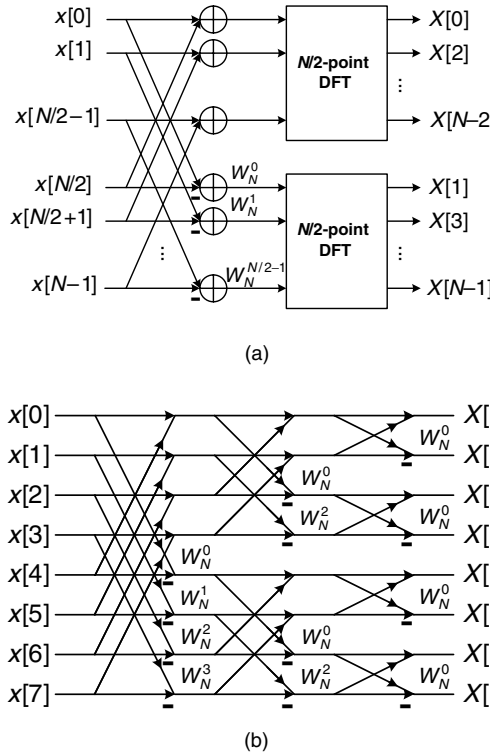


Figure 9.2. (a) Signal flow graph of radix-2 decimation-in-frequency decomposition and (b) its application in 8-point DFT computation

FFT algorithms, only the high-radix decimation-in-frequency FFT algorithms will be illustrated in the following.

Radix-4 Algorithm

Four-way symmetry of $W_N^{nk}(W_N^{nk+N/4} = -W_N^{nk+3N/4} = -jW_N^{nk})$ has been utilized to minimize the number of complex multiplications in the radix-4 algorithm. For $m = 0, 1, 2, \cdots, N/4 - 1$, its formulas can be derived as

$$
\begin{aligned}
X[4m] &= \sum_{n=0}^{N/4-1} \{x[n] + x[n+N/4] + x[n+N/2] + x[n+3N/4]\} \cdot W_{N/4}^{nm}, \\
X[4m+1] &= \sum_{n=0}^{N/4-1} \{x[n] - jx[n+N/4] - x[n+N/2] + jx[n+3N/4]\} \cdot W_N^{n} W_{N/4}^{nm}, \\
X[4m+2] &= \sum_{n=0}^{N/4-1} \{x[n] - x[n+N/4] + x[n+N/2] - x[n+3N/4]\} \cdot W_N^{2n} W_{N/4}^{nm}, \\
X[4m+3] &= \sum_{n=0}^{N/4-1} \{x[n] + jx[n+N/4] - x[n+N/2] - jx[n+3N/4]\} \cdot W_N^{3n} W_{N/4}^{nm}.
\end{aligned}
\tag{9.10}
$$

Petitioner Samsung Ex-1026, 0182

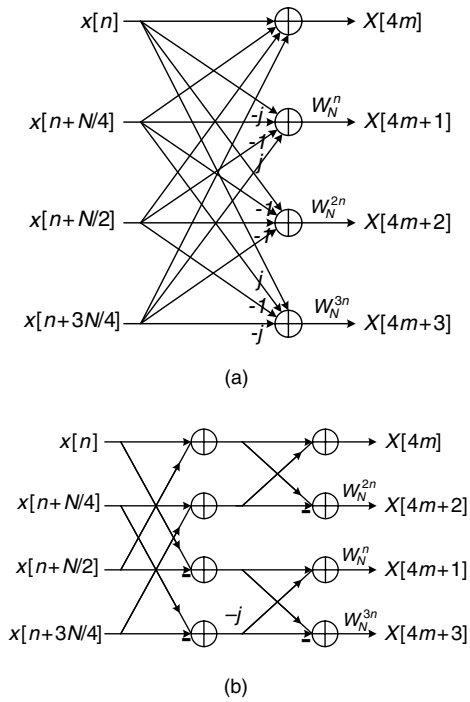


Figure 9.3. (a) Radix-4 butterfly and (b) radix-2^2 butterfly

Figure 9.3(a) depicts the signal flow graph of a radix-4 FFT butterfly operation. The radix-4 butterfly can be further decomposed and implemented by cascading two radix-2 stages, known as the radix-2^2 algorithm. The signal flow graph of such an algorithm is shown in Figure 9.3(b).

Radix-8 Algorithm

FFT algorithms using higher radix can be designed by decomposition of the frequency-domain samples into more groups at the cost of more complicated control. Radix-8 FFT algorithm simplifies multiplication with several special terms (*twiddle factors*), $W_N^{N/8}$, $W_N^{3N/8}$, $W_N^{5N/8}$ and $W_N^{7N/8}$, by

$$\begin{aligned}(a+jb)W_N^{N/8} &= -(a+jb)W_N^{5N/8} = \frac{\sqrt{2}}{2}[(a+b)+j(b-a)] \\ (a+jb)W_N^{3N/8} &= -(a+jb)W_N^{7N/8} = \frac{\sqrt{2}}{2}[(b-a)-j(a+b)].\end{aligned} \tag{9.11}$$

Note that these complex multiplications can be realized by two real-constant multiplications and two additions [1]. In addition, the constant multiplication can be replaced by shift-and-add operations. A radix-8 butterfly can be also realized by cascading three radix-2 stages,

Petitioner Samsung Ex-1026, 0183

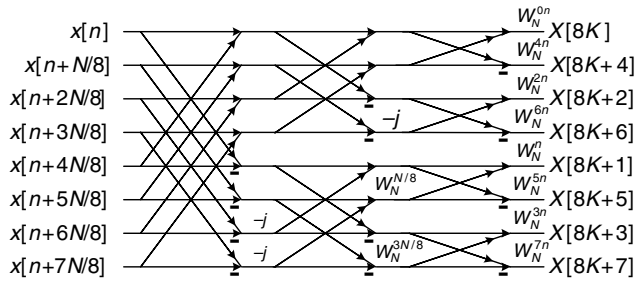


Figure 9.4. Radix-2^3 butterfly

which is called the radix-2^3 algorithm. The signal flow graph of the radix-2^3 FFT algorithm is shown in Figure 9.4. Note that the regularity of three radix-2 butterfly stages makes the algorithm amenable to pipelined implementation.

In summary, higher-radix FFT algorithms require fewer twiddle factor multiplications $(W_N^n, W_N^{2n}, \cdots)$. Also, much saving in the read-only memory (ROM) for storing the twiddle factors can be achieved. In consequence, FFT hardware solutions usually adopt radix-4 and radix-8 algorithms.

9.2.2 Architecture

Various FFT architectures have been proposed, such as memory-based architecture [2], pipelined architecture [1], cache memory architecture [3] and array architecture [4]. Among them, memory-based architecture and pipelined architecture are widely adopted, and they will be introduced in the following. The decimation-in-frequency FFT algorithms are adopted to illustrate these architectures. The same hardware design concept can be applied to the decimation-in-time FFT algorithms.

Memory-Based FFT Architecture

The *memory-based FFT architecture* (shown in Figure 9.5) consists of a main processing element and several memory blocks for rate matching, temporary storage and bit-reverse

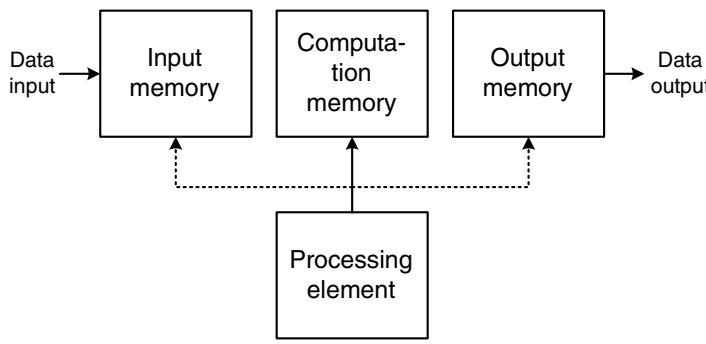


Figure 9.5. Memory-based FFT architecture

Petitioner Samsung Ex-1026, 0184

reordering. Every input to be transformed by the memory-based FFT processor goes through three phases: input buffering, computation and output reordering. Because the input data rate may not be equal to the operating frequency of the FFT processor, data are first stored in the input buffer. Once N samples are collected, the input buffer becomes the memory for the computation phase, accessed by the processor. Meanwhile, another memory block serves as the input buffer that stores yet more subsequent input data. The arithmetic processor may take quite a few cycles to complete one N-point FFT operation. Thus, the intermediate data are stored in the computation memory. As the transformation completes, the computation memory then serves as the output buffer to allow for sequence reordering, since the output data of the FFT operation are in a bit-reverse order.

Memory usage of the memory-based architecture can be reduced to $2N$ complex-valued words if the input and output buffers are shared and some in-place FFT algorithms achieving conflict-free addressing are adopted [2]. The in-place strategy works in the way that outputs of the processor are stored in the same memory addresses where the corresponding inputs are located. In [2], two main memory modules, each having four banks to accommodate the 4-input/output radix-4/2 butterfly unit, are allocated as shown in Figure 9.6(a). The radix-4/2 butterfly unit consists of two radix-2 stages and can compute either one radix-2^2 or one radix-2 butterfly operation, as shown in Figure 9.6(b).

An example of a 32-point FFT is illustrated in Figure 9.7, in which three stages of operations are performed. The first two stages are radix-4 butterfly operations, while the last stage uses a radix-2 butterfly. The data interchange mechanism is activated during the first and the third stages. The memory bank index is denoted by '**B**', while '**A**' indicates the memory address. Note that there are two different addressing modes. This memory-based architecture works as follows:

- data are written into the places where they are read from;
- the bold lines indicate the first butterfly operation in each iteration;
- in each butterfly operation, the processor accesses four data, one from each of the four banks in the current memory module;
- once the FFT is complete, the current memory module becomes the output buffer to generate FFT outputs in normal order;
- the same memory module also receives newly-arrived input data;
- in the next FFT period, the other memory module is accessed using the alternative addressing scheme, as shown in Figure 9.7(b).

In memory-based architecture, an N-point radix-r FFT computation requires $(N/r)\log_r N$ memory access and each access requires r words for read and write. Since only one processor handles the arithmetic operation, the operation clock is driven to $\log_r N/r$ times the data sample frequency.

Pipelined Architectures

Pipeline FFT architectures have the advantage of parallelism and pipelining. So these architectures are usually very fast, but they can be less flexible and require more hardware complexity.

Petitioner Samsung Ex-1026, 0185

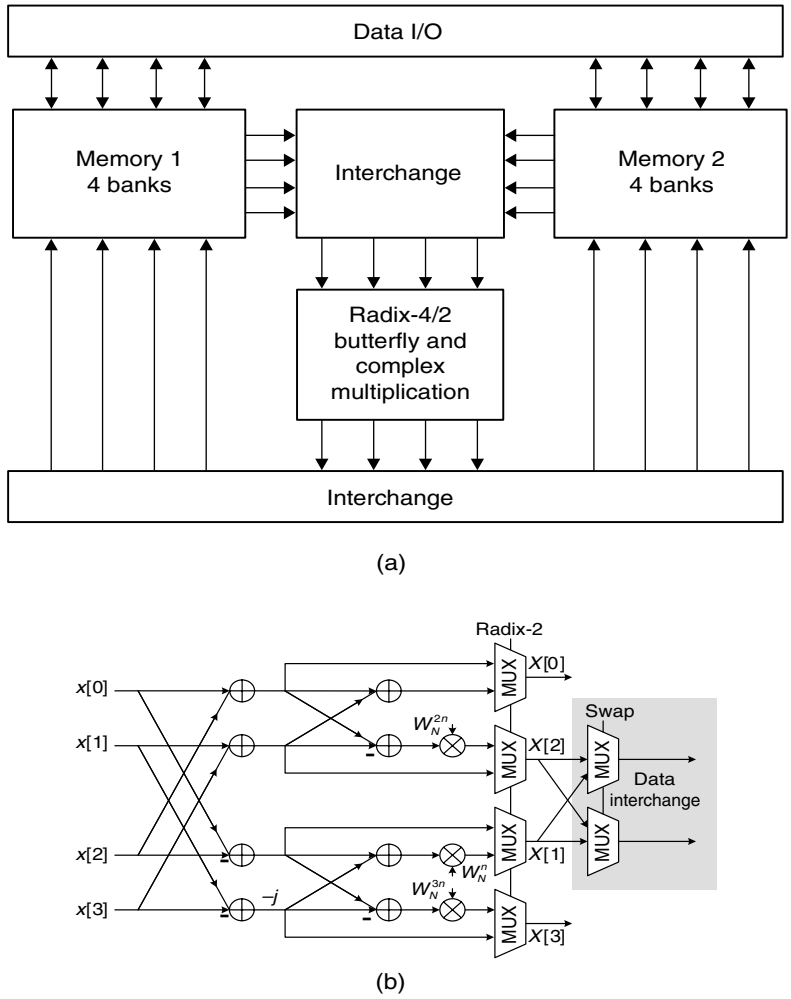


Figure 9.6. (a) Conflict-free addressing mixed-radix FFT processor and (b) radix-4/2 butterfly unit

Multi-Path Delay Commutator

The radix-2 *multi-path delay commutator* (MDC) [5] is a pipelined implementation of the radix-2 FFT algorithm. Figure 9.8 shows such an architecture with N equal to 16. The input sequence is divided into two parallel data streams by a commutator and then, with proper delay of one of the two streams, butterfly operation and twiddle factor multiplication are executed. In total, $(\log_2 N - 1)$ complex multipliers, $\log_2 N$ radix-2 butterfly unit and $(3N/2 - 2)$ words for delay are required. With a proper input buffering scheme, all processing elements (butterfly units and multipliers) can work at 100% utilization.

Petitioner Samsung Ex-1026, 0186

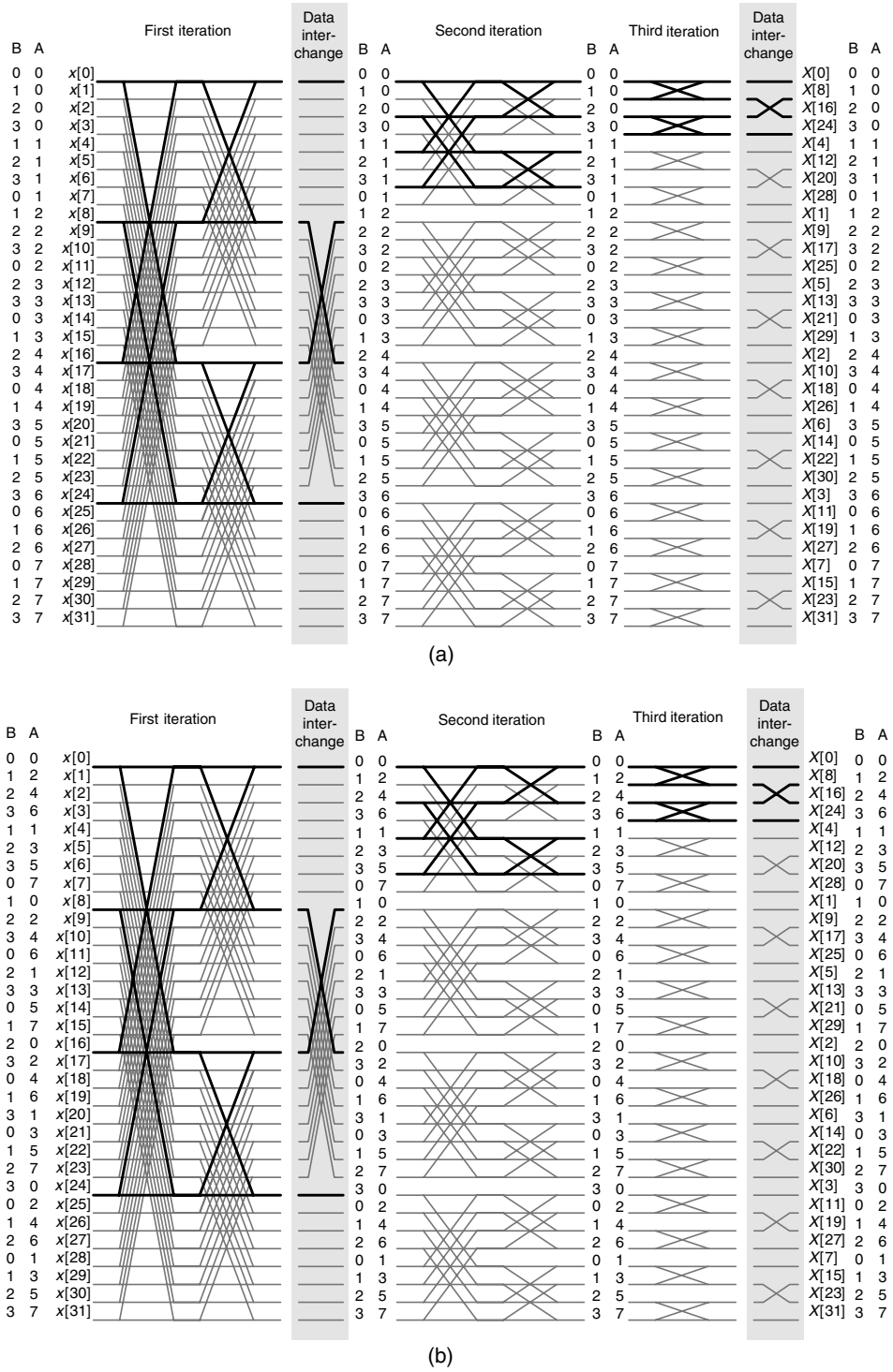


Figure 9.7. Signal flow graph of a 32-point FFT for the memory-based architecture: (a) normal addressing mode and (b) alternative addressing mode

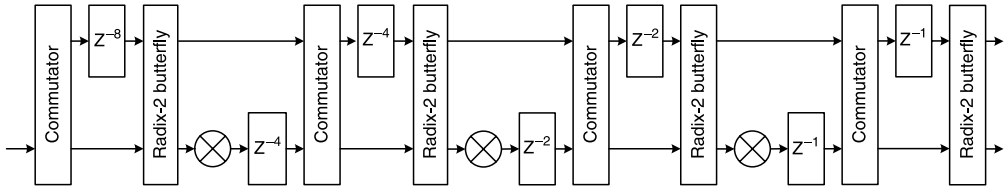


Figure 9.8. Radix-2 multi-path delay commutator FFT architecture ($N = 16$)

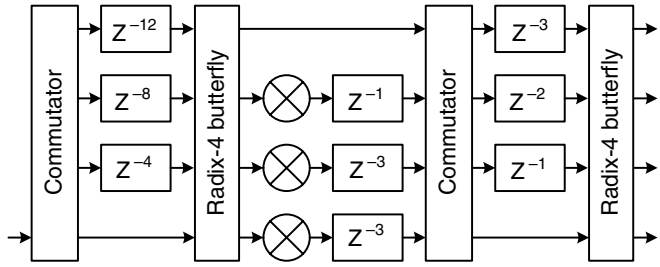


Figure 9.9. Radix-4 multi-path delay commutator FFT architecture ($N = 16$)

Radix-4 multi-path delay commutator FFT architecture (see Figure 9.9) can be implemented in a similar way as the radix-2 multi-path delay commutator FFT architecture, except that four parallel data streams are processed simultaneously and the butterfly units must handle radix-4 butterfly operation [6]. It needs $3(\log_4 N - 1)$ complex multipliers, $\log_4 N$ radix-4 butterfly units and $(5N/2 - 4)$ words of memory.

Single-Path Delay Feedback

The radix-2 *single-path delay feedback* (SDF) architecture (shown in Figure 9.10) utilizes the delay elements more efficiently by sharing the same storage between the butterfly outputs and inputs [7]. A single data stream goes through the multiplier in every stage. This architecture has the same number of butterfly units and multipliers as those in the radix-2 multi-path delay commutator FFT architecture, albeit only $N - 1$ delay elements. Note that the butterfly units and multipliers work at 50% utilization, since, half of the time, they are

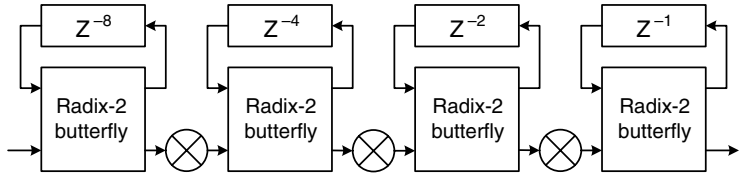


Figure 9.10. Radix-2 single-path delay feedback ($N = 16$)

Petitioner Samsung Ex-1026, 0188

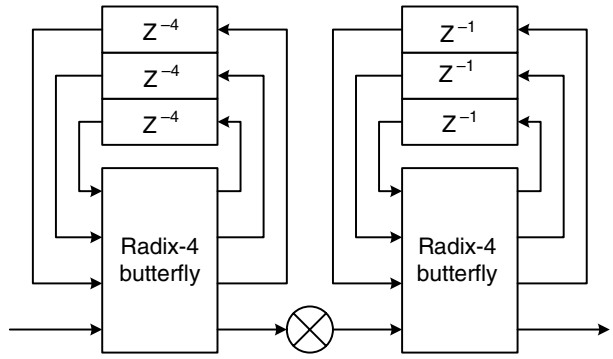


Figure 9.11. Radix-4 single-path delay feedback ($N = 16$)

bypassed. Figure 9.11 shows the single-path delay feedback FFT architecture employing the radix-4 algorithm [8]. In terms of complexity, the radix-4 single-path delay feedback FFT architecture requires $\log_4 N - 1$ complex multipliers, $\log_4 N$ radix-4 butterfly units and $N - 1$ memory words in delay lines.

An FFT processor with variable length can be useful in OFDM receivers that need to demodulate OFDM signals with different FFT sizes. For instance, the digital video broadcasting-handheld (DVB-H) standard stipulates three different FFT sizes: 2048, 4096 and 8192. One such variable-length FFT processor using the pipelined single-path delay feedback architecture was proposed in [9]. In order to accommodate different FFT sizes up to 2048, two radix-2 stages and three radix-2^3 stages are integrated as shown in Figure 9.12(a). The radix-2^3 stage is made up of three processing elements (PE1, PE2 and PE3) and a complex multiplier. The three different processing elements contain a basic butterfly unit, together with some combinational logic to deal with the symmetry of twiddle factors in higher-radix algorithms, depicted in Figure 9.12(b).

9.2.3 Comparison

The hardware requirements of the multi-path delay commutator (MDC) architectures and single-path delay feedback (SDF) architectures are compared in Table 9.1. From the table, it can be seen that due to efficient use of delay buffers, single-path delay feedback architectures have the minimum memory requirement of $N - 1$ words. In the single-path delay feedback architectures, the higher-radix algorithms reduce the number of complex multipliers and enhance hardware utilization. Note that the existence of a complex multiplier entails an accompanying twiddle-factor ROM. The twiddle-factor ROM for the first multiplier stores twiddle factors with a phase spacing of $2\pi/N$. In the later stages, the phase spacing gets larger and larger, and is equal to the radix times the phase spacing in the previous stage. This means that the size of the twiddle-factor ROM decreases rapidly in higher-radix architectures.

On the other hand, the higher-radix butterfly unit needs complicated adders with more inputs if it is not implemented by cascading several radix-2 butterfly units. In the multi-path delay commutator architecture, although the higher-radix algorithm helps to reduce the operational

Petitioner Samsung Ex-1026, 0189

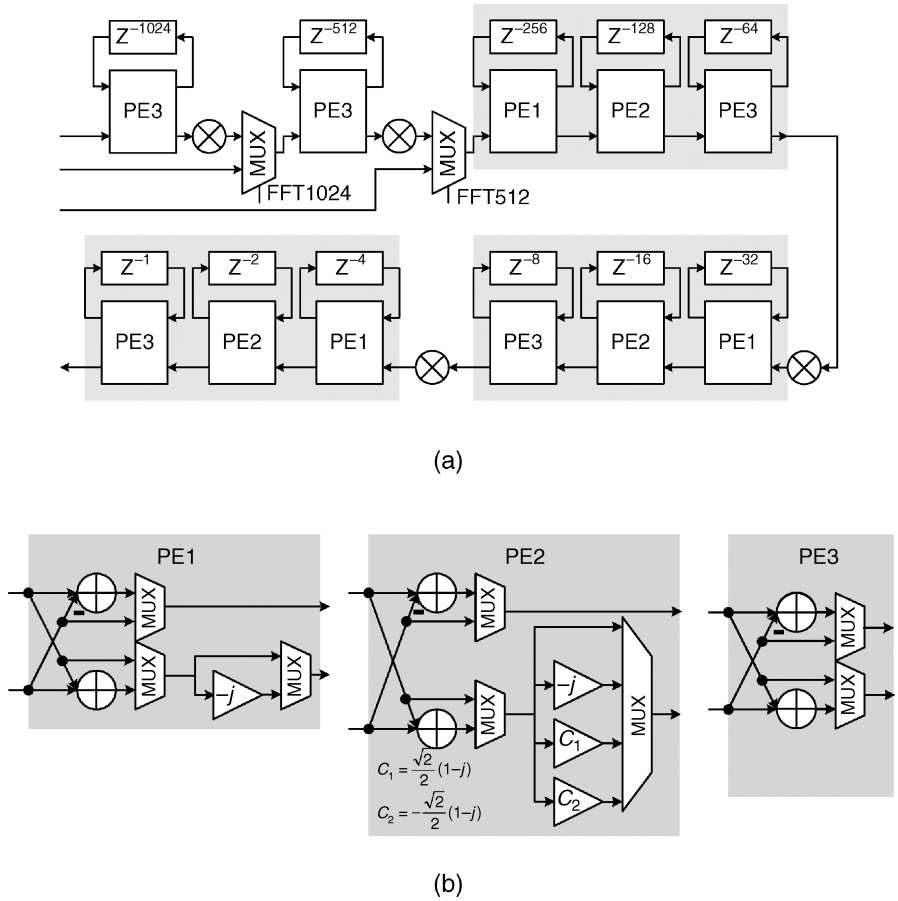


Figure 9.12. (a) Variable-length FFT processor and (b) its processing elements

clock frequency, the penalty is the increase in scheduling buffers for arranging the data so that 100% utilization can be achieved. As a result, higher-radix MDC FFT architectures are seldom used. The higher-radix single-path delay feedback FFT architectures, on the other hand, are preferred because they require less memory and fewer complex multipliers.

Table 9.1. Comparison of pipelined FFT architecture

	R2-SDF	R4-SDF	$R2^3$-SDF	R2-MDC	R4-MDC
Delay buffer	$N-1$	$N-1$	$N-1$	$3N/2-2$	$5N/2-4$
Complex adder	$2\log_2 N$	$8\log_4 N$	$2\log_2 N$	$2\log_2 N$	$8\log_4 N$
Adder utilization	50%	25%	50%	100%	100%
Complex multiplier	$\log_2 N-1$	$\log_4 N-1$	$\log_8 N-1$	$\log_2 N-1$	$3\log_4 N-1$
Multiplier utilization	50%	75%	87.5%	100%	100%
Schedule buffer	–	–	–	N	$3N$
Clock rate	1	1	1	0.5	0.25

Petitioner Samsung Ex-1026, 0190

Table 9.2. Comparison of memory-based FFT architecture and pipelined SDF FFT architecture

	Memory-based architecture	Single-path delay feedback architecture
Algorithm	Radix-r	Radix-r
Storage requirement	N	$N-1$
Memory banks (dual-port)	r	$\log_2 N$
Memory access times	$2N\log_r N$	$2N\log_2 N$
Complex multipliers	$r-1$	$\log_r N-1$
Complex adders	$2r$	$2\log_2 N$
Clock frequency	$\log_r N/r$	1
Clock cycle	$N\log_r N/r$	N

Table 9.2 compares the hardware implementation of memory-based architecture and pipelined single-path delay feedback FFT architecture. Assume that the radix-r algorithm is adopted in these two architectures. In memory-based architecture, a central memory block with size of N is partitioned into r memory banks to facilitate simultaneous access of r samples to complete the radix-r butterfly operation. In contrast, the pipelined single-path delay feedback architecture with cascaded radix-2 butterflies distributes its memory into $log_2 N$ banks with a total size of $N-1$. Note that in the memory-based architecture, random addressing is necessary to achieve conflict-free memory access. However, in the pipelined architecture, only sequential access is required and, thus, efficient implementation methods of sequential delay buffers can be adopted to reduce power consumption.

In order to accomplish an N-point FFT in the memory-based architecture, each memory bank is accessed for $\frac{N}{r}\log_r N$ times–about $2 \times r \times \frac{N}{r} log_r N$ memory accesses in total. Similarly, in the pipelined single-path delay feedback architecture, one N-point FFT demands $2 \times N \log_2 N$ memory accesses. In the memory-based architecture, $r-1$ complex multipliers and $2r$ complex adders are needed in the central processor. To achieve the same FFT operation speed as the pipelined architecture, the memory-based FFT processor needs to drive the processor clock frequency to $\frac{\log_r N}{r}$ times the input sample frequency. Consequently, memory-based architecture can be a good choice when circuit complexity rather than speed/power is of the major concern.

9.3 Delay Buffer

From the previous section, it is clear that delay buffers of different lengths are needed in the pipelined FFT architecture. In addition, delay buffers are used to store time-domain and frequency-domain data samples in synchronization and equalization blocks of all OFDM receivers. It is estimated that more than 30% of the area in OFDM baseband receivers is dedicated to delay buffers. Shift registers consisting of D-type flip-flops (DFFs) in cascade conveniently implement short-length delay buffers. However, at every clock edge, all the data move forward in a lock-step fashion and, on average, half of the DFFs change their states. In addition, each DFF in shift registers is composed of two latches–quite high complexity for one-bit storage. Therefore, a shift register is neither power economical nor area efficient, especially for long delay buffers.

Petitioner Samsung Ex-1026, 0191

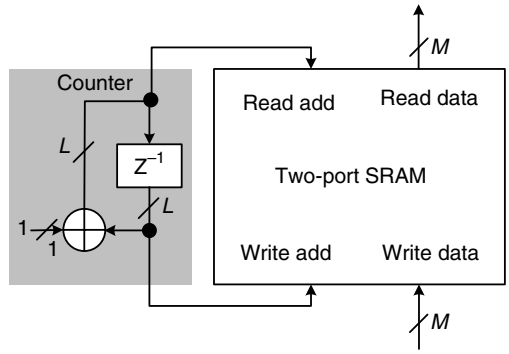


Figure 9.13. Two-port SRAM-based delay buffer. There are $2^L - 1$ stages of delay with M-bit word in each stage

Static random access memory (SRAM) and register file are alternative solutions to delay buffers and they have the advantage of compact bit cell and low power consumption. In an SRAM configured as a delay buffer, only two words are accessed (one read and one write) in every clock cycle and, thus, it consumes much less power than shift registers. A register file is structurally similar to SRAM, except that its length is limited. Despite some overhead for output sensing and address decoding circuits, SRAM is popularly adopted in long delay buffers, whereas a register file is suitable for moderate-length delay buffers. Since the delay buffer is accessed sequentially rather than randomly, special addressing techniques can be designed to further reduce the hardware complexity and power consumption. In the following, several approaches to implement a low-power delay buffer will be introduced.

9.3.1 SRAM/Register File-Based Delay Buffer

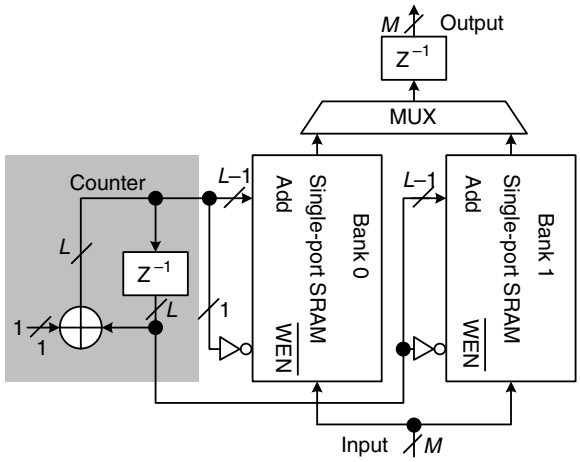
Two-Port SRAM/Register File

The two-port SRAM and register file, both with one read port and one write port, are most suitable for delay buffer implementation. Given the sequential access nature, the output address and the input address differ by one, namely the output address is $m + 1$ if the write address is m. Of course, in the next clock cycle, both addresses advance by one and the address wraps around to zero when it reaches the maximum. A simple counter can provide addresses for the two ports. Note that with an N-word SRAM/register, only $N - 1$ delay stages can be implemented. The block diagram of a two-port SRAM-based delay buffer is shown in Figure 9.13, where M denotes the data word-length.

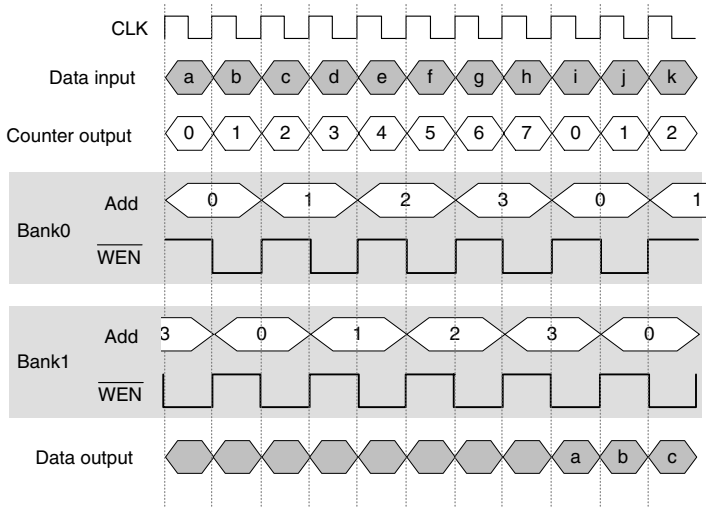
Single-Port SRAM/Register File

Although a two-port SRAM/register file can be configured as a delay buffer in a straightforward way, they sometimes are wasteful in area and power. Alternatively, two half-size ($N/2$-word) single-port SRAMs/register files can be adopted to construct a delay buffer. Two single-port SRAM modules are configured in a ping-pong fashion so that in every clock cycle, one data word is read out from one SRAM while the new data word is written into the other SRAM. In the next clock cycle, the two SRAM modules' roles are reversed.

Petitioner Samsung Ex-1026, 0192



(a)



(b)

Figure 9.14. (a) Single-port SRAM-based delay buffer and (b) its timing diagram

The main reason for such a design is to save on both area and power. For example, in a 0.18-μm CMOS technology, a two-port SRAM of size 1024 × 16 consumes 29.1 mW at a 100-MHz operating frequency and nominal supply voltage and it occupies about 0.4 mm^2. In contrast, two single-port 512 × 16 SRAM modules consume only 20.2 mW and have a size of 0.2 mm^2 – a 30% saving in power and a 50% saving in area.

As shown in Figure 9.14(a), in order to perform the alternative read and write access, the complement of the least-significant bit (LSB) of the address is used as the low-active 'write enable' control ($\overline{WEN}$). A read operation will precede a write operation on the same memory

Petitioner Samsung Ex-1026, 0193

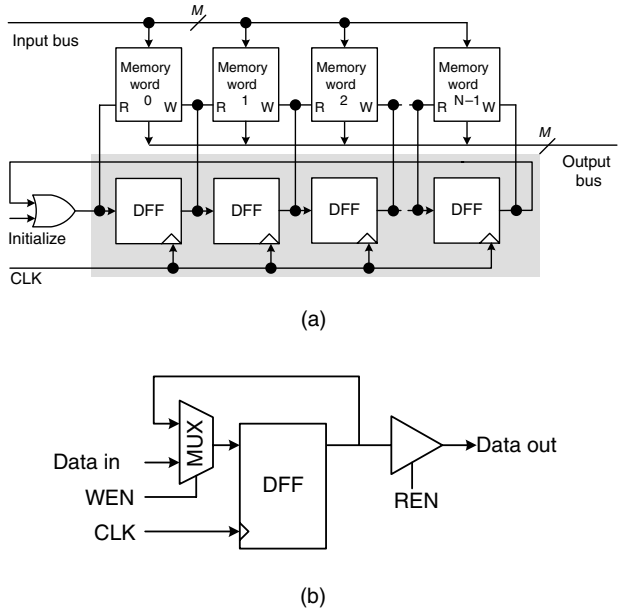


Figure 9.15. (a) Pointer-based delay buffer and (b) its memory cell

word by one clock cycle, namely the most out-of-date word in the previous cycle is overwritten by the latest incoming word. With this addressing scheme, one SRAM stores even-numbered data words and the other SRAM stores odd-numbered data words. The timing diagram of such an arrangement with $L = 3$ is depicted in Figure 9.14(b).

9.3.2 Pointer-Based Delay Buffer

To further simplify the circuit complexity, one can use a ring counter with only one active cell that points to two words to be accessed in every cycle. This method, known as the pointer-based scheme [10], is shown in Figure 9.15(a). The ring counter is made up of cascading one-bit shift registers. It is initialized with one '1' (the active cell) and the '1' propagates through the ring counter as the clock pulses arrive. One DFF in the ring counter points to two consecutive memory words for respective read and write operations. Compared with the shift-register-based delay buffer, the pointer-based delay buffer propagates only one '1' and keeps most of the DFF inactive. Besides, the ring counter is only one bit wide, whereas the shift register delay buffer shifts M-bit words. Consequently, the pointer-based delay buffer consumes much less power.

The memory cell in the pointer-based delay buffer can be either a standard SRAM cell or a DFF. If the SRAM cell is employed, then only the address decoder is replaced by the above ring counter. On the other hand, if a DFF is adopted as the memory cell, a multiplexer is required in the input path to the DFF for write control, as shown in Figure 9.15(b). A tri-state buffer is inserted in the output path to selectively drive the data to the output bus. This approach is suitable only for short delay buffers because in this scheme, the I/O control of the memory cells is more complicated than SRAM.

Petitioner Samsung Ex-1026, 0194

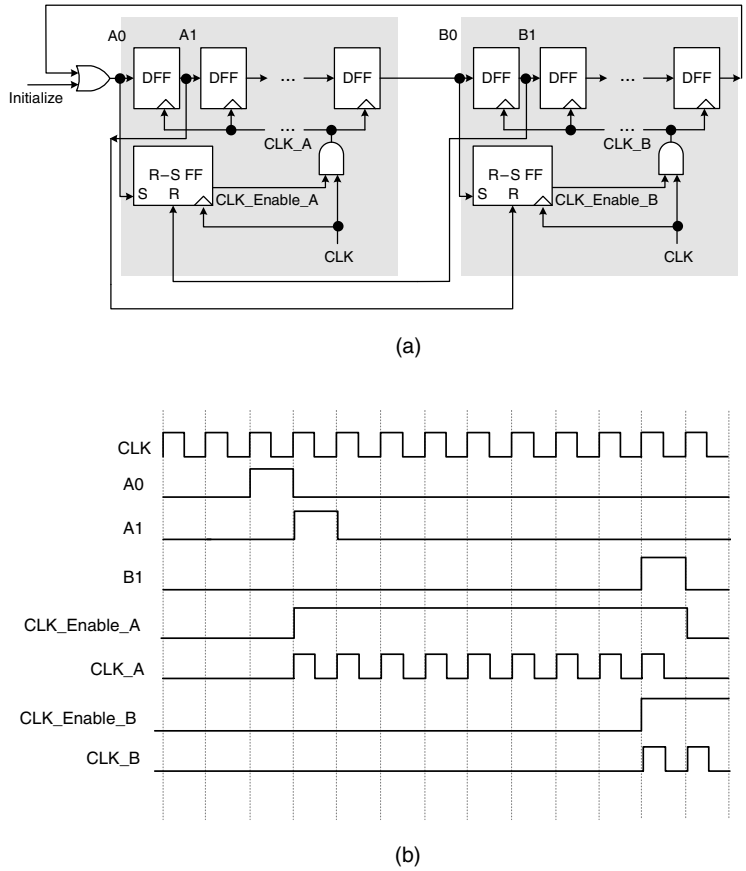


Figure 9.16. (a) Ring counter with clock gating by R–S flip-flop and (b) its timing diagram

9.3.3 Gated Clock Strategy

R-S Flip-Flop Gating

To further reduce power consumption, it is worth noting that most of the DFFs in the ring counter remain at '0' most of the time. The gated clock technique can be applied to the DFFs to save more power [11]. In [11], the ring counter is partitioned into several blocks, each having eight DFFs. Then, each block computes its own 'gate' signal, which then generates the gated clock for that block. As shown in Figure 9.16(a), when '1' arrives at the input of the first DFF in a block, the output of the R–S flip-flop is set to '1' at the next clock edge. Thus, the incoming '1' is captured by the first DFF of this block and it will continue to propagate inside that block. On the other hand, when '1' has successfully gone through the first DFF in the next block, the clock signal in the current block can henceforth be disabled. The timing diagram of this gate-clock ring counter is shown in Figure 9.16(b).

Petitioner Samsung Ex-1026, 0195

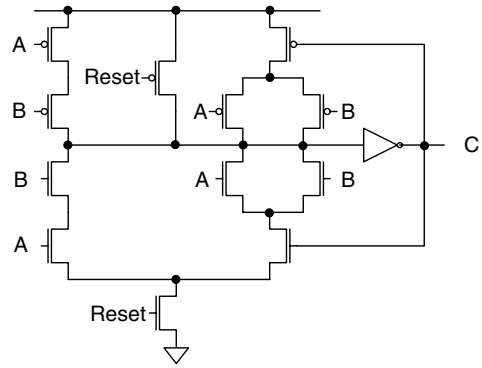


Figure 9.17. Circuit diagram of the C-element

C-Element Gating with Gated Driver Tree

Although some power is saved by gating the clock signal to all but one block, still some clock power is spent on driving the R–S flip-flops–one in every block. In [12], the R–S flip-flop is replaced by a C-element. Besides, the operating frequency is reduced to half by using the double-edge triggered (DET) flip-flop [13]. The C-element is a cell usually used in asynchronous circuits for handshaking, as shown in Figure 9.17 [14]. The logic function of a C-element is

$$C = AB + AC_{pre} + BC_{pre}, \tag{9.12}$$

where A and B are the two inputs and C and C_{pre} are the current and previous outputs. If $A = B$, then the output C is assigned to A. Otherwise, if $A \neq B$, C remains unchanged. The major advantage of the C-element is that its output is free of glitches, which is essential for a clock gating signal. Since the DFFs are replaced by DET flip-flops to run the ring counter at half speed, the gating on–off condition needs to be revised. The block diagram of the C-element gating strategy and its timing diagram are shown in Figure 9.18. When the input of the last DET flip-flop in the previous block has a transition from '0' to '1', the clock signal in the current block is enabled. When the output of the first DET flip-flop in the next block rises from '0' to '1', both inputs of the C-element go to '0' and the clock is turned off in the current block.

Besides using the C-element to gate the clock to DET flip-flops in a block, in [12], the authors propose to apply gating to the driver tree network that delivers the global clock signal to all blocks. Since, at any time, at most, two blocks need the global clock signal, so only those drivers along the path from the clock source to the blocks that need to be driven by the global clock are activated, as shown in Figure 9.19. The 'gate' signal ($\mathrm{CKE}_{i,j}$) for those drivers can utilize the same clock gating signals of their driving blocks. Thus, the driver tree 'gate' signal should be asserted when the active cell (whose output is '1') in the ring counter is one of its descendants in the quaternary driver tree. Given M blocks, each having D DET flip-flops, instead of activating all

$$\frac{M}{4} + \frac{M}{16} + \cdots = \frac{M}{3} \tag{9.13}$$

drivers, no more than $(1 + 2/D)\log_4 M$ drivers are activated.

Petitioner Samsung Ex-1026, 0196

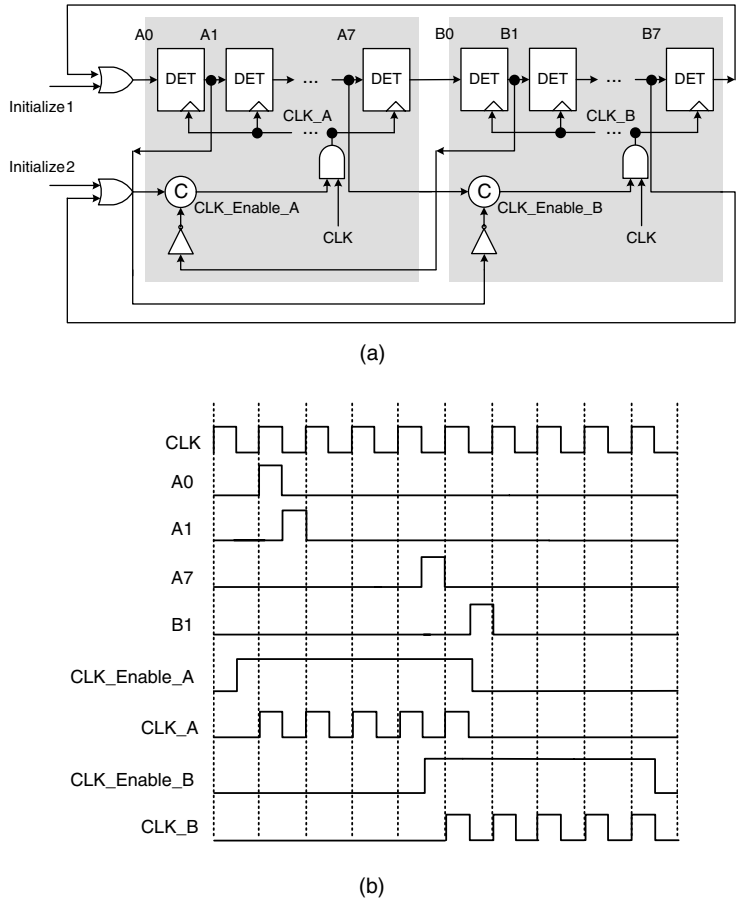


Figure 9.18. (a) Ring counter with clock gating by C-element and (b) its timing diagram

9.3.4 Comparison

Finally, the power consumption of delay buffers implemented by SRAM and a register file in a 0.18-μm CMOS technology is compared in Figure 9.20(a). The word-length is set to 16 bits and the operating frequency is 100 MHz. The results indicate that the register file consumes less power than the SRAM. But a register file may not be available for a length longer than 2048 and such is the case for the technology considered. A single-port register file is favored for a delay buffer longer than 256; otherwise, a two-port register file may be preferred.

These four delay buffer implementations are compared in terms of area and the results are given in Figure 9.20(b). For a buffer length of up to 1024, all but the two-port SRAM have fairly comparable area.

Three different length-1024 ring counter structures are simulated in a 0.18-μm CMOS technology with 1.8-V supply voltage and running at 50 MHz. In the simulation, eight DFFs are grouped in one block. The power consumption results of the three ring counters are

Petitioner Samsung Ex-1026, 0197

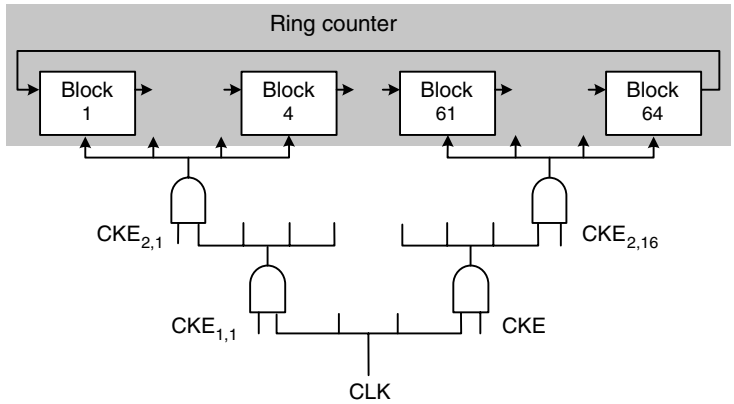


Figure 9.19. Clock driver tree and gating signal

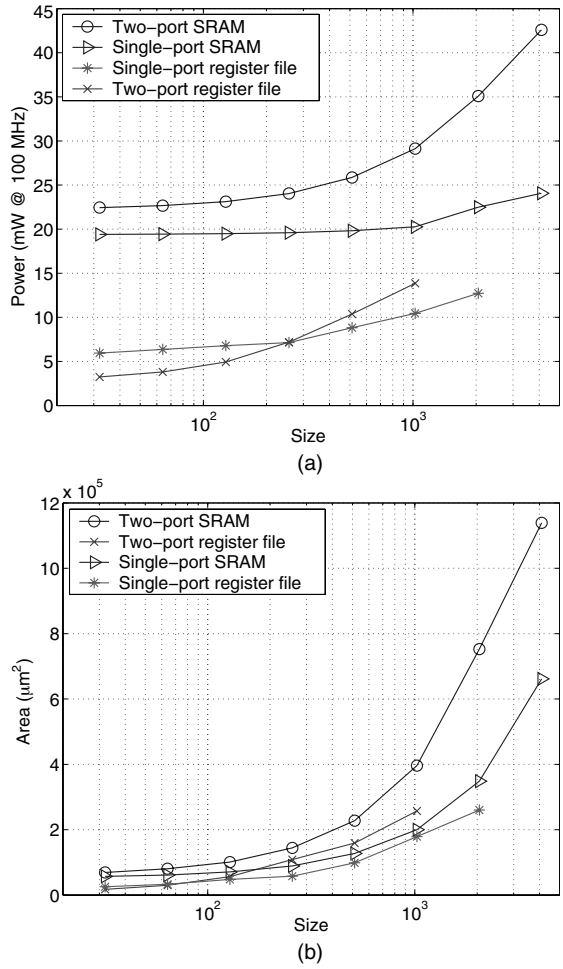


Figure 9.20. (a) Simulated power and (b) simulated area of several delay buffers using different approaches

Table 9.3. Power consumption of three ring counters

Architecture	Power (μW)
Conventional ring counter (Figure 9.15(a))	2127
R–S flip-flop gating ring counter (Figure 9.16(a))	433
C-element gating ring counter (Figure 9.18(a))	30

shown in Table 9.3. It is obvious that the clock gated by the C-elements indeed saves a significant portion of power in the ring counter.

9.4 Circuits for Rectangular-to-Polar Conversion

Baseband signal processing handles mostly complex signals. Sometimes, the magnitude or phase of a complex signal is needed for further processing. For example, in the delay correlator that detects OFDM symbol timing, the peak magnitude of the complex correlator output decides the symbol boundary, as mentioned in Chapter 4. Another example is that the phase of the delay correlator output can be used to estimate the fractional carrier frequency offset. Therefore, in this section, circuits that transform a signal from rectangular coordinate representation to polar coordinate format will be introduced.

9.4.1 Arctangent Function

Rational Approximation

The phase of a complex number $Z = X + jY$ is defined as

$$\theta = \tan^{-1}(Y/X). \tag{9.14}$$

Conventionally, a look-up table implemented by read-only memory (ROM) storing all the phases is used. Although a look-up table is simple, it still requires large area to achieve certain accuracy. A ROM-less rational approximation for the arctangent function is proposed in [15]:

$$\tan^{-1}(Y/X) \approx \frac{Y/X}{1 + 0.28125(Y/X)^2} \quad \text{(rad)}, \tag{9.15}$$

where $-1 \leq Y/X \leq 1$ and $X \geq 0$, which means that θ lies in the range $-45°$ to $+45°$. The maximum error by this rational approximation, as shown in Figure 9.21, is less than 0.3°.

The range of θ can then be extended by utilizing the symmetry of arctangent function. If $|Y| > |X|$ and $Y \geq 0$, then the arctangent approximation is derived by

$$\begin{aligned} \tan^{-1}\left(\frac{Y}{X}\right) &= \frac{\pi}{2} - \tan^{-1}\left(\frac{X}{Y}\right) \\ &\approx \frac{\pi}{2} - \frac{X/Y}{1 + 0.28125(X/Y)^2} \quad \text{(rad)}. \end{aligned} \tag{9.16}$$

Petitioner Samsung Ex-1026, 0199

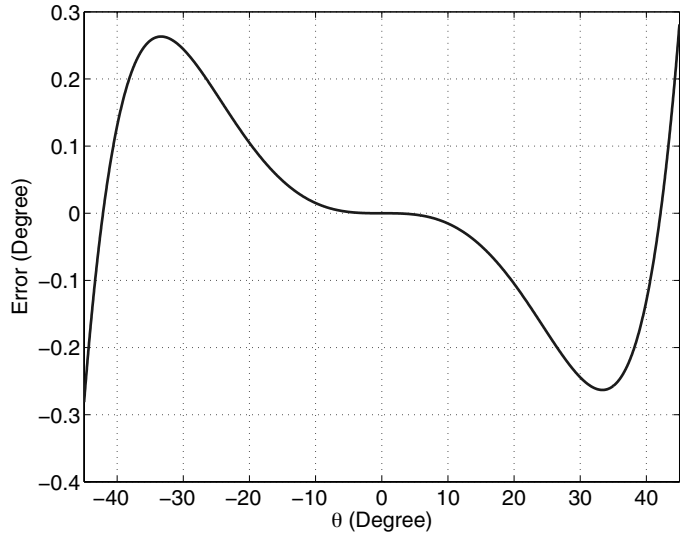


Figure 9.21. Error of rational approximation

Similarly, if $|Y| > |X|$ and $Y < 0$, then the arctangent approximation is given by

$$\tan^{-1}\left(\frac{Y}{X}\right) \approx -\frac{\pi}{2} - \frac{X/Y}{1 + 0.28125(X/Y)^2} \quad \text{(rad)}. \tag{9.17}$$

When $|X| \geq |Y|$, $X < 0$ and $Y \geq 0$,

$$\tan^{-1}\left(\frac{Y}{X}\right) \approx \pi + \frac{Y/X}{1 + 0.28125(Y/X)^2} \quad \text{(rad)}. \tag{9.18}$$

When $|X| \geq |Y|$, $X < 0$ and $Y < 0$,

$$\tan^{-1}\left(\frac{Y}{X}\right) \approx -\pi + \frac{Y/X}{1 + 0.28125(Y/X)^2} \quad \text{(rad)}. \tag{9.19}$$

Direct implementation of the rational approximation for arctangent function requires two dividers, one adder, one squarer and one constant multiplier. In [15], the constant 0.28125 is represented by a sum of two power-of-two terms, namely $0.28125 = 1/4 + 1/32$. Hence, Equation (9.15) can be reformulated as

$$\tan^{-1}\left(\frac{Y}{X}\right) \approx \frac{XY}{X^2 + (1/4)Y^2 + (1/32)Y^2} \quad \text{(rad)}. \tag{9.20}$$

A circuit diagram of the arctangent rational approximation is depicted in Figure 9.22 and there are two squarers, one multiplier, one divider and two adders.

Petitioner Samsung Ex-1026, 0200

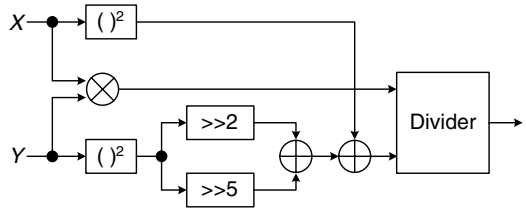


Figure 9.22. Implementation of the rational approximation for arctangent function

CORDIC

CORDIC, short for COordinate Rotation DIgital Computer, has been widely used in computing vector rotation in the rectangular coordinates. The basic concept underlying the CORDIC algorithm is to decompose a desired rotation angle into a set of elementary rotation angles. The elementary rotation angle is chosen so that it can be accomplished by simple shift-and-add operations. Consider the desired rotation angle θ, which is decomposed as

$$\theta = \sum_{i=0}^{N-1} \mu_i \theta_e(i), \tag{9.21}$$

where the i-th elementary rotation angle, $\theta_e(i)$, is given by

$$\theta_e(i) = \tan^{-1}(2^{-i}), \tag{9.22}$$

and $\mu_i \in \{+1, -1\}$ is used to determine the direction of rotation [16].

In order to calculate the phase of a complex signal, $Z = X + jY$, the CORDIC algorithm operates in the angle accumulation mode, also known as the backward rotation mode. Its goal is to rotate the vector $[X \;\; Y]^T$ back to the x-axis and, at the same time, accrue the angle between them. The iterations can be described as:

- Initialization: $X(0) = X,\ Y(0) = Y,\ \theta'(0) = 0$
- Direction decision:

$$\mu_i = -\text{sgn}(Y(i)) \tag{9.23}$$

- Micro rotation:

$$\begin{bmatrix} X(i+1) \\ Y(i+1) \end{bmatrix} = \begin{bmatrix} 1 & -\mu_i 2^{-i} \\ \mu_i 2^{-i} & 1 \end{bmatrix} \cdot \begin{bmatrix} X(i) \\ Y(i) \end{bmatrix} \tag{9.24}$$

- Angle accumulation:

$$\theta'(i+1) = \theta'(i) - \mu_i \tan^{-1}(2^{-i}) \tag{9.25}$$

- Go back to Equation (9.23).

Petitioner Samsung Ex-1026, 0201

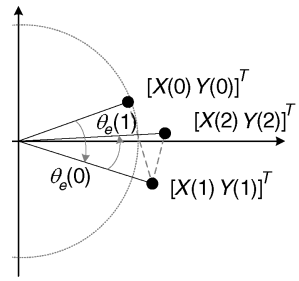


Figure 9.23. Trajectory of the vector during the CORDIC process using the angle accumulation mode

The function $\mathrm{sgn}(\cdot)$ is +1 if its argument is greater than or equal to zero; otherwise, it is set to -1. Figure 9.23 shows the trajectory of the vector after two iterations. It is clear that rotation changes the norm of the vector. The vector is multiplied by a factor $\sqrt{1+2^{-2i}}$ in each iteration. At the end of the process, the magnitude of the vector has been amplified by

$$\prod_{i=0}^{\infty} \sqrt{1+2^{-2i}} = 1.6468\cdots. \tag{9.26}$$

In this case, since only the angle is desired, magnitude scaling is not necessary.

The maximum of the accumulated angle is given by

$$\theta_{\mathrm{MAX}} = \sum_{i=0}^{\infty} \tan^{-1}(2^{-i}) = 1.7433\cdots(99^\circ). \tag{9.27}$$

From the above, it can be seen that θ_{MAX} is slightly greater than $\pi/2$. Thus, the angle of a vector in the first and fourth quadrants can be represented by a sum of the elementary angles. If the vector lies outside of the two quadrants, namely $X < 0$, then it must be mapped to the two quadrants before the CORDIC procedure. Afterwards the accumulated angle must be subtracted from $+\pi$ or $-\pi$, depending on the sign of Y, to obtain the correct arctangent result.

The angle approximation error of N CORDIC iterations is

$$\delta = \theta - \sum_{i=0}^{N-1} \mu_i \theta_e(i), \tag{9.28}$$

which is bounded by the smallest elementary rotation angle, $\theta_e(N-1)$ [17]. So, the minimum number of CORDIC iterations can be easily derived according to the required accuracy.

Hardware implementation of the CORDIC arctangent function contains the micro-rotation stage in Equation (9.24) and the angle accumulator in Equation (9.25). The circuit diagrams of these two modules are shown in Figure 9.24. The micro-rotation stage consists of two barrel shifters and adders/subtractors. The barrel shifter is capable of N-bit right shift and it

Petitioner Samsung Ex-1026, 0202

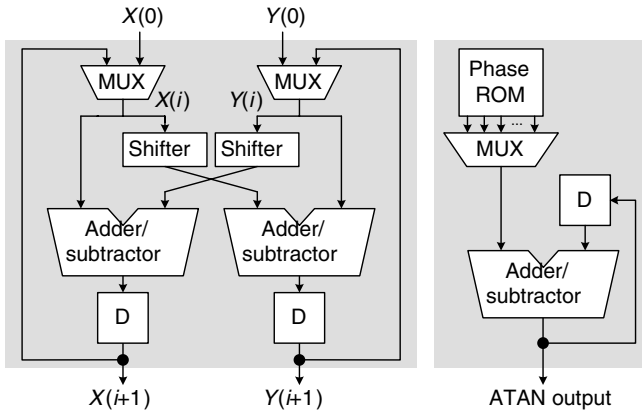


Figure 9.24. Basic CORDIC rotation stage and angle accumulator

is controlled by the iteration count. The add/subtract operation, depends on the sign of $Y(i)$. The angle accumulator also performs a simple add/subtract operation, depending on $Y(i)$. N such CORDIC rotation stages can be cascaded to construct one complete CORDIC angle accumulation mode circuit, as shown in Figure 9.25, where the initial stage maps the input vector into the first and the fourth quadrants.

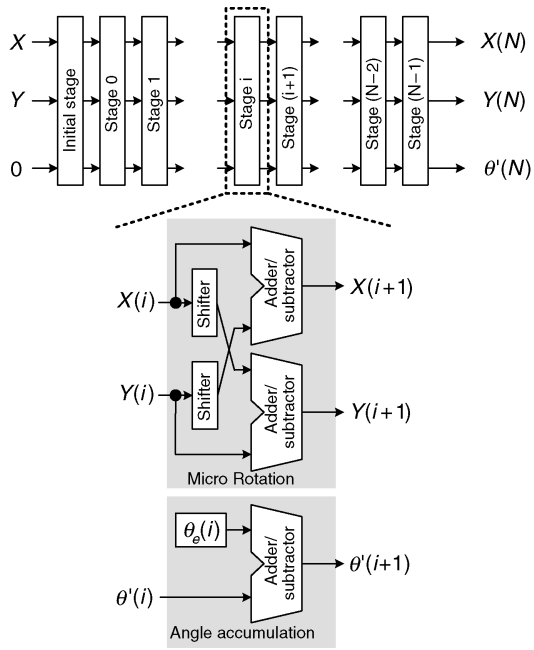


Figure 9.25. Parallel implementation of several CORDIC micro-rotation stages

Petitioner Samsung Ex-1026, 0203

9.4.2 Magnitude Function

CORDIC

The magnitude function of a complex signal, $Z = X + jY$, is defined as

$$|Z| = \sqrt{X^2 + Y^2}. \tag{9.29}$$

It specifies the length from that complex value to the origin in the complex plane.

The CORDIC algorithm can also calculate the magnitude of a complex signal. Scaled magnitude is a by-product when the CORDIC algorithm derives the phase of a complex number. Figure 9.23 clearly indicates that as a vector $[X \;\; Y]^T$ is being rotated back to the horizontal axis, the magnitude of the rotated vector gets scaled by a constant in each step. As a result, if $Y(N) = 0$ after N iterations, then

$$X(N) = \sqrt{X^2 + Y^2} \cdot \prod_{i=0}^{N-1} \sqrt{1 + 2^{-2i}}. \tag{9.30}$$

Consequently, the magnitude of $X + jY$ can be obtained by

$$\sqrt{X^2 + Y^2} = \frac{X(N)}{\prod_{i=0}^{N-1} \sqrt{1 + 2^{-2i}}} = S(N) \cdot X(N). \tag{9.31}$$

The scaling factor $S(N)$ can be computed in advance, since the number of iterations is known and fixed. As such, $S(N)$ can be converted into a canonical sign-digit representation:

$$S(N) = \sum_{k=1}^{K} s_k 2^{-k}, \tag{9.32}$$

where $s_k = \pm 1$, and the scaling can be realized by shift-and-add operations.

Since the scaled result after finite micro-rotations, $S(N) \cdot X(N)$, is approximated to the magnitude of $\sqrt{X^2 + Y^2}$, as shown in Figure 9.26, if a residual angle δ exists, the percentage approximation error is given by

$$\begin{aligned} \frac{|X + jY| - S(N)X(N)}{|X + jY|} &= \frac{S(N)\sqrt{X(N)^2 + Y(N)^2} - S(N)X(N)}{S(N)\sqrt{X(N)^2 + Y(N)^2}} \\ &= 1 - \frac{X(N)}{\sqrt{X(N)^2 + Y(N)^2}} \\ &= 1 - \cos(\delta). \end{aligned} \tag{9.33}$$

From Equation (9.28), it is clear that

$$\delta < \theta_e(N-1) = \tan^{-1}\left(\frac{1}{2^{N-1}}\right). \tag{9.34}$$

Petitioner Samsung Ex-1026, 0204

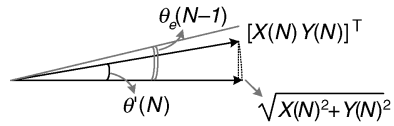


Figure 9.26. Magnitude error due to residual phase of finite CORDIC rotation iterations

As a result, the percentage approximation error is bounded by

$$\frac{|X+jY| - S(N)\,X(N)}{|X+jY|} < 1 - \cos\left(\tan^{-1}\left(\frac{1}{2^{N-1}}\right)\right) = 1 - \frac{2^{N-1}}{\sqrt{2^{2N-2}+1}}. \tag{9.35}$$

If $N = 8$, the approximation error will be around 3×10^{-5}.

A scaling-free CORDIC algorithm has been proposed in [18]. Unlike the conventional CORDIC algorithm, the scaling-free CORDIC algorithm allows $\mu_i = 1$ only, which means that the angle is decomposed as a pure summation of the elementary angles. These elementary angles $\theta_f(i)$ are selected to preserve the magnitude of the input vector $[X \;\; Y]^T$. Compared to Equation (9.24), its micro-rotation step becomes

$$\begin{bmatrix} X(i+1) \\ Y(i+1) \end{bmatrix} = \begin{bmatrix} 1 - 2^{-(2i+1)} & 2^{-i} \\ -2^{-i} & 1 - 2^{-(2i+1)} \end{bmatrix} \cdot \begin{bmatrix} X(i) \\ Y(i) \end{bmatrix} \tag{9.36}$$

for clockwise rotation. Its hardware implementation is depicted in Figure 9.27. Two extra shifters and two extra subtractors are needed. Note, however, that for all stages with $i \geq N/2$, the additional shifters and subtractors can be omitted because right shift of the input vector by $2i + 1$ bits can be neglected.

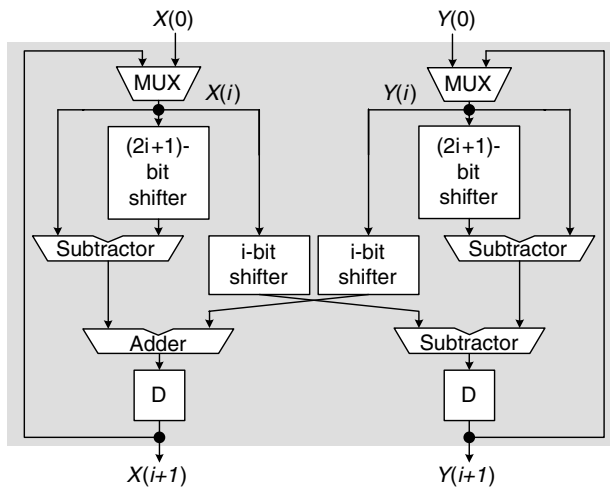


Figure 9.27. Scaling-free CORDIC micro-rotation stage

Petitioner Samsung Ex-1026, 0205

The scaling-free CORDIC algorithm eliminates the scale factor and thus provides an effective solution to magnitude-related vector rotation problems. Nevertheless, it suffers from a small angle range, which restricts its usage in general applications. Methods that expand the angle computation range have been developed. Further discussion can be found in [18,19].

Linear Approximation

A simple linear approximation to the magnitude function is proposed in [20]:

$$|Z| = \sqrt{X^2 + Y^2} \approx \max(|X|, |Y|) + \frac{\min(|X|, |Y|)}{2}, \tag{9.37}$$

where $\max(\cdot)$ and $\min(\cdot)$ represent the maximum and minimum of its two arguments, respectively. Although the above approximation is simple for hardware implementation, the approximation error can be as high as 12%, as shown in Figure 9.28(a), making it unsuitable for many applications that require more precision. Note that only the first octant is considered, since the magnitude of any general $[X \;\; Y]^T$ can be derived from some equivalent cases in the first octant.

In order to reduce the approximation error, the magnitude curve in the first octant is partitioned into more segments. A two-segment linear approximation can be given by

$$\sqrt{X^2 + Y^2} \approx \begin{cases} a_0 X + b_0 Y & 0 \le Y/X \le 1/2 \\ a_1 X + b_1 Y & 1/2 < Y/X \le 1 \end{cases}. \tag{9.38}$$

A four-segment linear approximation becomes

$$\sqrt{X^2 + Y^2} \approx \begin{cases} c_0 X + d_0 Y & 0 \le Y/X \le 1/4 \\ c_1 X + d_1 Y & 1/4 < Y/X \le 1/2 \\ c_2 X + d_2 Y & 1/2 < Y/X \le 3/4 \\ c_3 X + d_3 Y & 3/4 < Y/X \le 1 \end{cases}. \tag{9.39}$$

Note that the segments are chosen so as to facilitate hardware implementation.

The coefficients are chosen with a view to minimizing the maximum approximation error and representing them with the minimum number of power-of-two terms. For the two-segment approximation, $(a_0, b_0, a_1, b_1) = (1, 3/16, 3/4, 11/16)$, where the minimum power-of-two term is set to 2^{-4}. With these parameters, the two-segment linear approximation attains a maximum approximation error of 2.2%. Similarly, the coefficients of the four-segment linear approximation are set to

$$(c_0, d_0, c_1, d_1, c_2, d_2, c_3, d_3) = \left(1, \frac{13}{128}, \frac{15}{16}, \frac{45}{128}, \frac{27}{32}, \frac{69}{128}, \frac{3}{4}, \frac{85}{128}\right), \tag{9.40}$$

where the minimum power-of-two term is restricted to 2^{-7}. In this case, the maximum approximation error of the four-segment linear approximation is 0.52%. The approximation error distributions of the two-segment and four-segment linear approximation methods with the above coefficient settings are depicted in Figure 9.28(b).

The coefficients for the above linear approximation have been designed in the form of the sum of power-of-two terms so as to enable simple hardware realization using shift-and-add

Petitioner Samsung Ex-1026, 0206

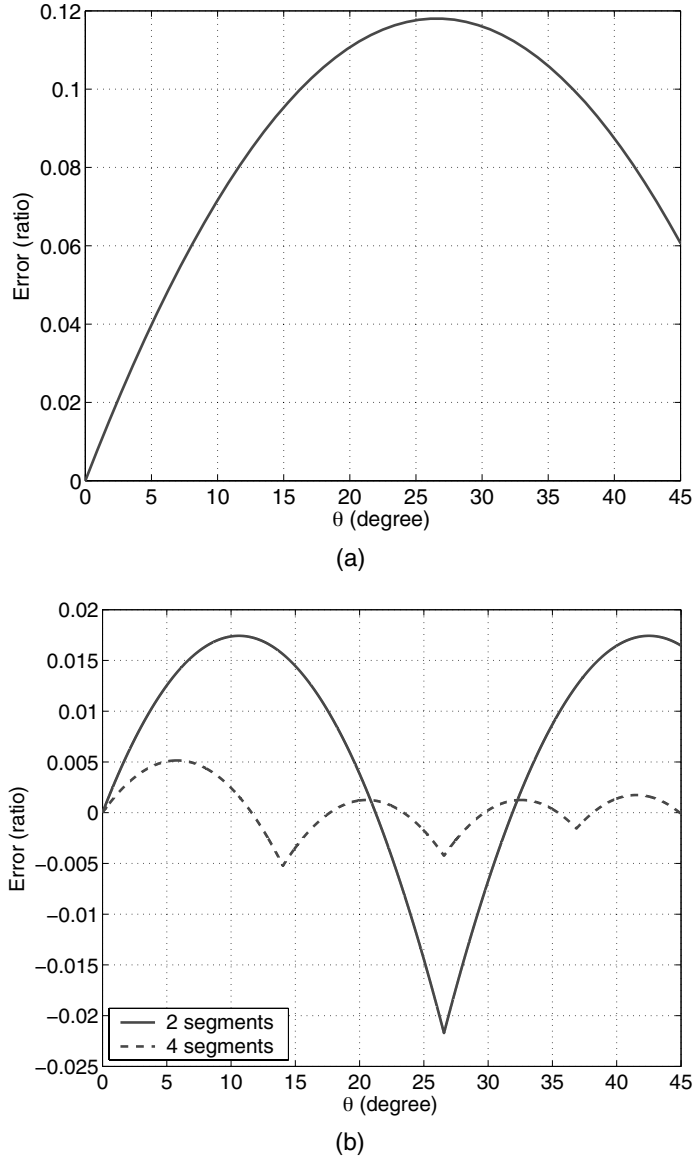


Figure 9.28. Error distributions of piecewise linear approximation to the magnitude function: (a) one segment and (b) two and four segments

circuit structure. The coefficients in the sum-of-power-of-two format are listed in Tables 9.4 and 9.5. To save even more hardware, terms of the real-part or imaginary-part coefficients must be shared as much as possible. Figure 9.29(a) and (b) depict the two circuits for two-segment and four-segment linear approximation of the magnitude for a vector in the first

Petitioner Samsung Ex-1026, 0207

Table 9.4. Coefficients of two-segment linear approximation to the magnitude function

Segment	Range	Coeff. of real part	Coeff. of imag. part
0	$0 \leq Y/X \leq 1/2$	1	$2^{-3} + 2^{-4}$
1	$1/2 < Y/X \leq 1$	$2^{-1} + 2^{-2}$	$2^{-1} + 2^{-3} + 2^{-4}$

octant, respectively. For a vector in other octants, another circuit that identifies and maps the vector to a corresponding vector in the first octant is designed. In addition, the segment identification can be easily implemented in two shifters and two comparators. The above two circuits are integrated and illustrated in Figure 9.29(c).

9.4.3 Comparison

Table 9.6 compares two approaches to the implementation of the arctangent function. The rational approximation, albeit its higher complexity, is favored in most cases due to its low latency and high accuracy. However, the CORDIC algorithm is still suitable when the throughput of the arctangent function is low and a high degree of accuracy is required. In this case, the CORDIC algorithm is advantageous in this area because one unit of CORDIC hardware can compute one CORDIC iteration in one clock cycle and very accurate arctangent results can be obtained in several clock cycles.

Table 9.7 lists the comparison between the multi-segment linear approximation and the CORDIC algorithm for the magnitude function computation. Both of them involve only simple shift-and-add operations. The four-segment linear approximation is slightly more complicated due to the need for more adders. But it works quite fast, considering that its accuracy is more than satisfactory in many applications. The CORDIC algorithms, again, is able to provide very accurate results given a sufficient number of iterations. In the case in which both the arctangent function and the magnitude function are needed, the CORDIC algorithm is obviously a better choice because of hardware sharing.

9.5 Circuits for Polar-to-Rectangular Conversion

In the previous section, circuits that transform a signal from the rectangular coordinates to the polar coordinates have been discussed. Now, circuits for inverse conversion from the polar coordinates to the rectangular coordinates will be described. Their applications include direct digital frequency synthesizers (DDFS), numeric control oscillators (NCO) and twiddle-factor generators in FFT. Conventionally, a ROM-based look-up table storing sine and cosine values and addressed by the input phase is a straightforward solution. By

Table 9.5. Coefficients of four-segment linear approximation to the magnitude function

Segment	Range	Coeff. of real part	Coeff. of imag. part
00	$0 \leq Y/X \leq 1/4$	1	$2^{-4} + 2^{-5} + 2^{-7}$
01	$1/4 < Y/X \leq 1/2$	$2^{-1} + 2^{-2} + 2^{-3} + 2^{-4}$	$2^{-2} + 2^{-4} + 2^{-5} + 2^{-7}$
10	$1/2 < Y/X \leq 3/4$	$2^{-1} + 2^{-2} + 2^{-4} + 2^{-5}$	$2^{-1} + 2^{-5} + 2^{-7}$
11	$3/4 < Y/X \leq 1$	$2^{-1} + 2^{-2}$	$2^{-1} + 2^{-3} + 2^{-5} + 2^{-7}$

Petitioner Samsung Ex-1026, 0208

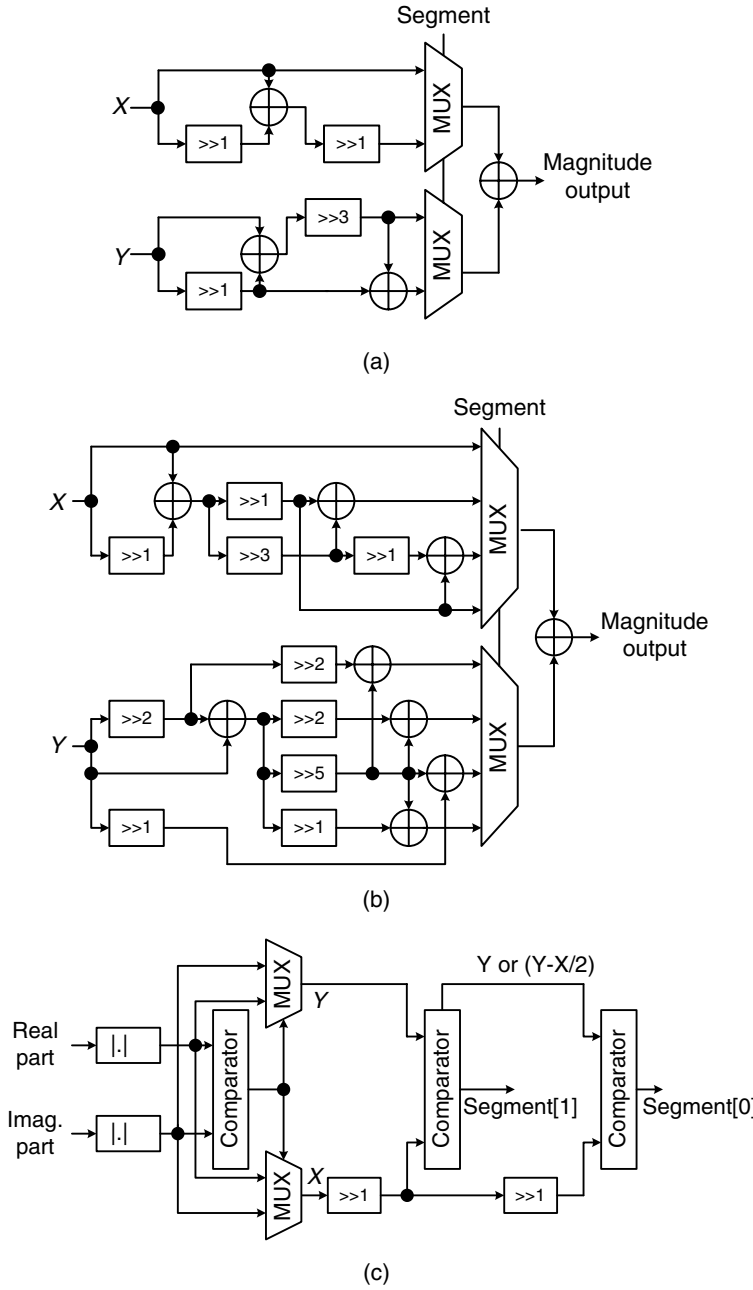


Figure 9.29. Circuits of (a) two-segment vector magnitude linear approximation, (b) four-segment vector magnitude linear approximation and (c) general vector mapping to the first octant and segment identification

Petitioner Samsung Ex-1026, 0209

Table 9.6. Comparison between the rational approximation and the CORDIC algorithm for arctangent function

	Rational approximation	CORDIC
Latency	Short	Long
ROM	No	N words
Error	$<0.3°$	$<0.44°(N = 8)$
Hardware	Adder, multiplier, square function, divider	Adder/subtractor, barrel shifter
Complexity	High	Low

Table 9.7. Comparison between the multi-segment linear approximation and the CORDIC algorithm for magnetude function

	Multi-segment linear approximation		CORDIC
	Two-segment	Four-segment	
Latency	Short	Moderate	Long
Error	2.2%	0.52%	0.003% ($N = 8$)
Hardware	Adder, shifter	Adder, shifter	Adder/subtractor, barrel shifter
Complexity	Low	Medium	Low

exploiting the eight-way symmetry of the sine and cosine waveforms, ROM tables storing the cosine/sine values of phases within [0, $\pi/4$] are sufficient (see Figure 9.30). The resolution of the phase is $2\pi/2^L$, while the precision of the amplitude is R bits. Three most significant bits (MSBs) of the phase, ϕ, are used to compute the final outputs from the two

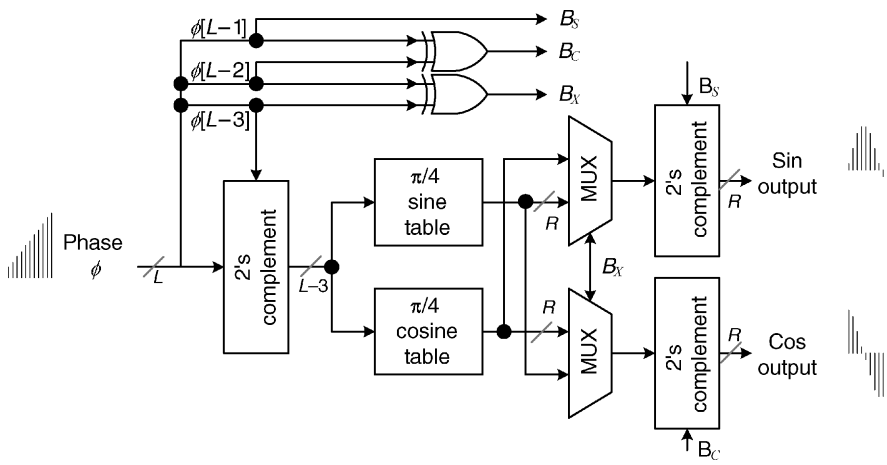


Figure 9.30. Quadrature phase-to-amplitude conversion exploiting eight-way symmetry of the sine/cosine waveforms

Petitioner Samsung Ex-1026, 0210

tables, while the remaining $L-3$ bits are used as the index to look up the tables. Denote $\phi[i]$ as the i-th binary bit of ϕ, and let three control signals, B_S, B_C and B_X, be defined as

$$\begin{aligned} B_S &= \phi[L-1], \\ B_C &= \phi[L-1] \oplus \phi[L-2], \\ B_X &= \phi[L-2] \oplus \phi[L-3], \end{aligned} \tag{9.41}$$

where $\oplus$ denotes the exclusive OR operation. These control signals represent sign change of the sine output, sign change of the cosine output and swapping between the sine and cosine outputs, respectively.

The look-up table approach has limited application in those cases that require high-frequency/phase resolution, since the ROM size grows exponentially with the word-length of the phase, L. In the following, recent works that implement high-resolution polar-to-rectangular conversion with reduced ROM sizes will be introduced.

9.5.1 Trigonometric Approximation

Angle Decomposition

The table look-up approximation to arctangent function usually involves large ROMs that store the sine/cosine values. One alternative is to break down one large table to several smaller ones. In [21], one such alternative arctangent approximation method is proposed. At first, assume without loss of generality that the phase is in the first octant, namely $0 \le 2\pi\phi \le \pi/4$, and ϕ has a word-length of $L-3$ bits (see Figure 9.30). Also let $\phi = \phi_A + \phi_B$. Then, according to trigonometric identities, one has

$$\begin{aligned} \sin(2\pi\phi) &= \sin(2\pi\phi_A)\cos(2\pi\phi_B) + \cos(2\pi\phi_A)\sin(2\pi\phi_B), \\ \cos(2\pi\phi) &= \cos(2\pi\phi_A)\cos(2\pi\phi_B) - \sin(2\pi\phi_A)\sin(2\pi\phi_B). \end{aligned} \tag{9.42}$$

Usually, the decomposition is done so that ϕ_A represents the more significant bits of the phase while ϕ_B represents the less significant bits. Hence, ϕ_B is relatively small and $\cos(2\pi\phi_B) \approx 1$. The trigonometric formulas become

$$\begin{aligned} \sin(2\pi\phi) &\approx \sin(2\pi\phi_A) + \cos(2\pi\phi_A)\sin(2\pi\phi_B), \\ \cos(2\pi\phi) &\approx \cos(2\pi\phi_A) - \sin(2\pi\phi_A)\sin(2\pi\phi_B). \end{aligned} \tag{9.43}$$

Suppose ϕ_B have a word-length of M bits; then, ϕ_A has a word-length of $L-M-3$ bits. Thus, instead of two tables, each with 2^{L-3} words, three smaller tables for $\sin(2\pi\phi_A)$, $\cos(2\pi\phi_A)$ and $\sin(2\pi\phi_B)$ will suffice. By setting $M \approx L-M-3$, the number of memory words can be reduced to about only the square root of that in Figure 9.30.

Quadruple Angle Formula

Another ROM-less phase to the sinusoid amplitude conversion technique uses the quadruple angle formula [22]. The double-angle formula for cosine function takes the form of

$$\cos(2\theta) = 2\cos^2(\theta) - 1 = 1 - 2\sin^2(\theta). \tag{9.44}$$

Petitioner Samsung Ex-1026, 0211

The quadruple-angle formula is obtained from the above formula as

$$\cos(4\theta) = 2\cos^2(2\theta) - 1 = 1 - 8\sin^2(\theta)(1 - \sin^2(\theta)). \tag{9.45}$$

Again, due to the symmetry of the sinusoidal waveform, the range of 4θ is set to within 0 to $\pi/2$, equivalently $0 \le \theta \le \pi/8$. Then, the approximation that $\sin(\theta) \approx \theta$ can be applied when θ is small and the above equation is reduced to

$$\cos(4\theta) \approx 1 - 8\theta^2(1 - \theta^2), \quad 0 \le \theta \le \pi/8. \tag{9.46}$$

Based on the trigonometric quadruple angle approximation, for $0 \le 2\pi\phi \le \pi/4$, a low-error approximation to $\cos(2\pi\phi)$ is given by

$$\cos(2\pi\phi) \approx (485/512)[1 - 8x^2(1 - x^2)] + (27/512)[1 - 8x^2(1 - x^2)]^2|_{x=\frac{3135}{8192}\phi}. \tag{9.47}$$

The approximation error of the above formula is smaller than 2^{-13} [22].

9.5.2 Polynomial Approximation

Dual Slope

In addition to using trigonometric identities, polynomial approximation is also widely used in trigonometric function generation. Polynomial approximation has been popular because:

- extensive literature exists on polynomial approximation;
- polynomial coefficients can be optimized given the criterion; and
- simple hardware architecture can be readily developed.

Consider sine/cosine function approximation of phases in the interval $[0, \pi/4]$. This phase range can be partitioned into 2^M segments and a polynomial of a given degree is used to fit the trigonometric function in each segment. The dual-slope technique is a special piecewise linear approximation with a reduced memory requirement [23]. It makes use of the continuity requirement to decrease the number of parameters (4×2^M) by 25%. In [24], a set of minimum mean-squared error (MMSE) parameters for sine/cosine approximation that use the dual-slope piecewise linear approach are derived. This implementation achieves an accuracy of about 2^{-12}.

Polynomial Hyperfolding

Instead of using many polynomials, one for each segment, one single polynomial can be adopted for trigonometric function approximation. This approach eliminates the ROM table that stores coefficients of polynomials for the segments. Second-order and third-order polynomials with optimized coefficients can be designed to approximate sine and cosine functions [25]. The coefficients are selected with a view to maximizing the absolute value of the spurious-free dynamic range (SFDR)–an index that measures the purity of sine/cosine waveforms, as well as minimizing the required hardware complexity.

Petitioner Samsung Ex-1026, 0212

Two second-order polynomials for first-octant sine and cosine approximation take the form of

$$\begin{aligned} \sin\left(\frac{\pi}{4}\phi\right) &\approx -0.004713 + 0.838015 \cdot \phi - 2^{-3} \cdot \phi^2, \\ \cos\left(\frac{\pi}{4}\phi\right) &\approx +0.995593 - 0.011408 \cdot \phi + (-2^{-2} - 2^{-5}) \cdot \phi^2, \end{aligned} \tag{9.48}$$

where $0 \le \phi \le 1$. Two optimal third-order polynomials for the same purpose are given by

$$\begin{aligned} \sin\left(\frac{\pi}{4}\phi\right) &\approx +0.00015005 + 0.77436217 \cdot \phi - 0.00530040 \cdot \phi^2 \\ &\quad + \frac{-2^{-2} + 2^{-5}}{3} \cdot \phi^3, \\ \cos\left(\frac{\pi}{4}\phi\right) &\approx +0.98423596 + 0.00452969 \cdot \phi - 0.32417224 \cdot \phi^2 \\ &\quad + \frac{2^{-3} - 2^{-5}}{3} \cdot \phi^3. \end{aligned} \tag{9.49}$$

A polynomial hyperfolding technique can be used to efficiently compute the third-order polynomial in the above equation by flattening all the sums of partial products [25]. In this technique, ϕ is replaced by its N-bit binary representation, namely $\phi = \sum_{i=0}^{N-1} x_i 2^i$. Then, ϕ in the polynomial can be expanded using the binary representation and all the terms with one, two, three binary variables can be collected. Since the coefficients are constants, canonical sign-digit representation can be employed to convert the coefficients into weighted sums of power of two. Hence, the polynomial computation is converted into evaluation of the sum of partial products, which can be efficiently processed by the well-known Wallace tree architecture.

As an example, if the second-order polynomial approximation uses a phase resolution of 11 bits and an amplitude precision of up to 2^{-9} and the third-order polynomial approximation uses 14-bit phase resolution and 12-bit amplitude precision, then the numbers of total partial products after flattening are around 30 and 130, respectively. When compared with the direct implementation approach, the polynomial hyperfolding technique saves more than 60% of partial products.

9.5.3 Comparison

Table 9.8 shows the comparison of different architectures for phase-to-amplitude conversion. Both the double-angle approximation and the piecewise linear approximation trade ROM size for more complicated arithmetic circuits. Although the table look-up technique entails more storage, it is still popular for storing twiddle factors in FFT processors [9,26]. This is because the twiddle factors are fixed and have equal phase spacing. For FFT processors of moderate sizes, special techniques can be employed and the ROM size reduced [26].

The quadruple-angle approximation and the high-order polynomial hyperfolding techniques completely eliminate the ROM table and are suitable for the cases in which the ROM cannot be implemented economically. However, the prices for such accurate results in

Petitioner Samsung Ex-1026, 0213

Table 9.8. Comparison of various phase-to-amplitude conversion techiques

	ROM* (Bits)	Complexity	Flexibility
Table Lookup	24576	Low	High
Double Angle Approx.	832	Medium	Medium
Piecewise Linear Approx. (64 segments)	1216	Medium	Medium
Quadruple Angle Approx.	–	High	Medium
2nd-Order Polynomial Hyperfolding Approx.	–	Medium	Low
3rd-Order Polynomial Hyperfolding Approx.	–	High	Low

*evaluated at 14-bit phase resolution and 12-bit amplitude precision.

quadruple angle approximation are three multipliers. Similarly, the third-order polynomial hyperfolding technique generates many partial products to be summed by Wallace trees. As to design flexibility, arbitrary look-up tables can be easily generated once the requirements in amplitude precision and phase resolution are determined. On the other hand, the word-lengths of the input/output signals of the arithmetic units in the double-angle approximation, quadruple-angle approximation and piecewise linear approximation all need be decided according to specified quantization error. In addition, the polynomial hyperfolding implementation requires custom design of the Wallace trees each time the word-length of the phase is changed, making it less desirable for hardware re-use.

Bibliography

[1] S. He and M. Torkelson, 'Designing pipeline FFT processor for OFDM (de)modulation', in *Proc. of IEEE URSI International Symposium on Signals, Systems, and Electronics*, Sep. 1998, pp. 257–262.

[2] B. G. Jo and M. H. Sunwoo, 'New continuous-flow mixed-radix (CFMR) FFT processor using novel in-place strategy', *IEEE Trans. Circuits Syst. I*, vol. 52, no. 5, May 2005, pp. 911–919.

[3] B. M. Baas, 'A low-power, high-performance, 1024-point FFT processor', *IEEE J. Solid-State Circuits*, Vol. 34, No. 3, Mar. 1999, pp. 380–387.

[4] J. O'Brien, J. Mather and B. Holland, 'A 200 MIPS single-chip 1K FFT processor', in *Proc. of IEEE International Solid-State Circuits Conference*, Vol. 36, 1989, pp. 166–167.

[5] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Eaglewood Cliffs, NJ: Prentice Hall, 1975.

[6] E. E. Swartzlander, V. K. Jain and H. Hikawa, 'A radix-8 wafer scale FFT processor', *J.VLSI Signal Processing*, May 1992, pp. 165–176.

[7] H. L. Groginsky and G. A. Works, 'A pipeline fast fourier transform', *IEEE Trans. Comput.*, vol. C-19, no. 11, Nov. 1970, pp. 1015–1019.

[8] A. M. Despain, 'Fourier transform computer using CORDIC iterations', *IEEE Trans. Comput.*, Oct. 1974, pp. 993–1001.

[9] Y. T. Lin, P. Y. Tsai and T. D. Chiueh, 'Low-power variable-length fast Fourier transform processor', *IEE Proc.-Comput. Digit. Tech.*, vol. 152, no. 4, Jul. 2005, pp. 499–506.

[10] E. K. Tsern and T. H. Meng, 'A low-power video-rate pyramid VQ decoder', *IEEE J. Solid-State Circuits*, vol. 31, no. 11, Nov. 1996, pp. 1789–1794.

Petitioner Samsung Ex-1026, 0214

[11] N. Shibata, M. Watanabe and Y. Tanabe, 'A current-sensed high-speed and low-power first-in–first-out memory using a wordline/bitline-swapped dual-port SRAM cell', *IEEE J. Solid-State Circuits*, vol. 37, no. 6, Jun. 2002, pp. 735–750.

[12] P. C. Hsieh, J. S. Jhuang and T. D. Chiueh, 'Design of a low-power delay line using gated-driver tree', in *Proc. of the VLSI/CAD Symposium*, Taiwan.

[13] R. Hossain, L.D. Wronski and A. Albicki, 'Low power design using double edge triggered flip-flops', *IEEE Trans. VLSI Syst.*, vol. 2, no. 2, Jun. 1994, pp. 261–265.

[14] M. Shams, J. Ebergen and M. Elmasry, 'Modeling and comparing CMOS implementations of the c-Element', *Tech. Rep. of Dep. Computer Science, Univ. of Waterloo, Waterloo, Ont., Canada*, CS-98-12, May 1998.

[15] R. Lyons, 'Another contender in the arctangent race', *IEEE Signal Processing Mag.*, Jan. 2004, pp. 109–110.

[16] Y. H. Hu, 'CORDIC-based VLSI architectures for digital signal processing', *IEEE Signal Processing Mag.*, July 1992, pp. 16–35.

[17] Y. H. Hu, 'The quantization effects of the CORDIC algorithm', *IEEE Trans. Signal Processing*, vol. 40, no. 4, Apr. 1992, pp. 834–844.

[18] A. S. Dhar and S. Banerjee, 'An array architecture for fast computation of discrete hartley transform', *IEEE Trans. Circuits Syst.*, vol. 38, no. 9, Sep. 1991, pp. 1095–1098.

[19] K. Maharatna, S. Banerjee, E. Grass, M. Krstic and A. Troya, 'Modified virtually scaling-free adaptive CORDIC rotator algorithm and architecture', *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 11, Nov. 2005, pp. 1463–1474.

[20] P. B. Denyer and D. Renshaw, *VLSI Signal Processing: A Bit-Serial Approach*. Boston, MA: Addison-Wesley, 1985.

[21] F. Curticapean and J. Niittylahti, 'A hardware efficient direct digital frequency synthesizer', in *Proc. of IEEE International Conference on Electronics, Circuits and Systems*, Sep. 2001, pp. 51–54.

[22] C.-C. Wang, Y.-L. Tseng, H.-C. She, C.-C. Li and R. Hu, 'A 13-bit resolution ROM-less direct digital frequency synthesizer based on a trigonometric quadruple angle formula', *IEEE Trans. VLSI Syst.*, vol. 12, no. 9, Sep. 2004, pp. 895–900.

[23] D. De Caro and A. G. M. Strollo, 'High-performance direct digital frequency synthesizers in 0.25 μm CMOS using dual slope approximation', *IEEE J. Solid-State Circuits*, vol. 40, no. 11, Nov. 2005, pp. 2220–2227.

[24] A. G. M. Strollo, D. De Caro, E. Napoli and N. Petra, 'Direct digital frequency synthesis with dual-slope approach', in *Proc. of Eur. Solid-State Circuit Conference*, Sep. 2003, pp. 397–400.

[25] D. De Caro, E. Napoli and A. G. M. Strollo, 'Direct digital frequency synthesizers with polynomial hyperfolding technique', *IEEE Trans. Circuits Syst. II*, vol. 51, no. 7, Jul. 2004, pp. 337–344.

[26] Y. W. Lin, H.-Y. Liu and C.-Y. Lee, 'A 1-GS/s FFT/IFFT processor for UWB applications', *IEEE J. Solid-State Circuits*, vol. 40, no. 8, Aug. 2005, pp. 1726–1735.

Petitioner Samsung Ex-1026, 0215

10

System Examples

OFDM baseband receivers include several key modules implementing signal-processing tasks of diverse nature. In addition, OFDM system parameter design is a complicated process of trading off different factors, such as channel delay spread, spectral efficiency, transmission rate and others. System examples offer concrete illustration of design considerations as well as implementation details and, hence, can help designers, experienced and novice, in their receiver design efforts.

Baseband receiver design of OFDM systems involves quite broad and complicated knowledge and experiences. Until now, this book has treated several important issues related to this work, yet no introduction to design technology is complete without actual examples. In this chapter, two OFDM receiver IC developments will be described. The first example demonstrates a receiver implemented for a wireless communication system that combines the multi-carrier OFDM modulation and the multiple-access capability of the code-division multiple access (CDMA). The second example is a wireless OFDM receiver IC for *cognitive radio* applications. In this wireless LAN-like OFDM receiver, dynamic spectrum access, MIMO techniques and channel assessment functions are integrated to meet the requirements of cognitive radio applications.

10.1 MC-CDMA Downlink Receiver IC

The downlink MC-CDMA receiver is designed for a next-generation mobile cellular communication system. Taking advantages of both multi-carrier modulation and CDMA techniques, the MC-CDMA system can handle frequency-selective fading channels and offer high data rate transmission in a multi-user environment. Moreover, the *orthogonal variable spreading factor* (OVSF) codes [1] spread the original data symbol over several subcarriers, making frequency diversity reception and multi-rate transmission possible. In light of the above considerations, MC-CDMA has been proposed as a possible solution for the beyond-3G (B3G) or the fourth-generation (4G) cellular communication systems [2].

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0216

Table 10.1. Proposed downlink MC-CDMA system parameters. Reproduced by permission of © 2005 IEEE

RF frequency	2 GHz
Signal bandwidth	5 MHz
Sampling frequency	5.76 MHz
FFT size	1024
Subcarrier spacing	5.625 KHz
Guard interval	11.1 μs
Symbol time	188.9 μs
Number of data subcarriers	768
Number of pilot subcarriers	33
Maximum data rate (uncoded)	21.7 Mbps

10.1.1 System Description

In Table 10.1, important parameters of the proposed downlink MC-CDMA system are summarized [3]. Note that the same RF frequency and signal bandwidth as those of the 3G standards, namely 2 GHz and 5 MHz, are used. This is because the major objective of this downlink MC-CDMA system is to enhance the downlink data rate of the current 3G W-CDMA cellular communication system in urban areas. Moreover, two guard bands, each with 5% bandwidth of the signal band are reserved. To facilitate future dual-mode receiver design, the ADC sampling rate is set to 5.76 MHz–1.5 times the chip rate in the W-CDMA system (3.84 MHz).

According to the channel models provided by 3GPP [4], the maximum excess delays are 2.14 μs and over 10 μs, respectively, in typical urban and bad urban areas. Moreover, the highest mobility supported is up to 120 km/hr. As a result, the guard interval needs to be longer than 10 μs. Also, a maximum Doppler frequency of 222 Hz limits the OFDM symbol duration to be shorter than 200 μs. Given the sampling rate and the symbol duration, the DFT size is set to 1024. Among the 1024 subcarriers, 768 subcarriers are used to transmit data and 33 pilot subcarriers are uniformly distributed. The system supports three signal constellations, namely QPSK, 16-QAM and 64-QAM, and the maximum uncoded data rate reaches 21.7 Mbps. For multiple-access purposes, the *orthogonal variable spreading factor* (OVSF) codes spread the user data onto a number of subcarriers. Also note that to facilitate smooth transition in the future, the system uses the same FFT size, guard interval ratio and guard band ratio as those in the feasibility study of the 3GPP UTRAN OFDM enhancement [5].

10.1.2 Transmitter and Receiver Design

MC-CDMA Transmitter

The transmitter of the MC-CDMA system is very similar to the transmitter of an OFDM communication system except for multiple user data streams and the OVSF code spreading. Figure 10.1(a) depicts the block diagram of the proposed MC-CDMA baseband transmitter. Data from each user first pass through a constellation mapper, get spread by the OVSF code

Petitioner Samsung Ex-1026, 0217

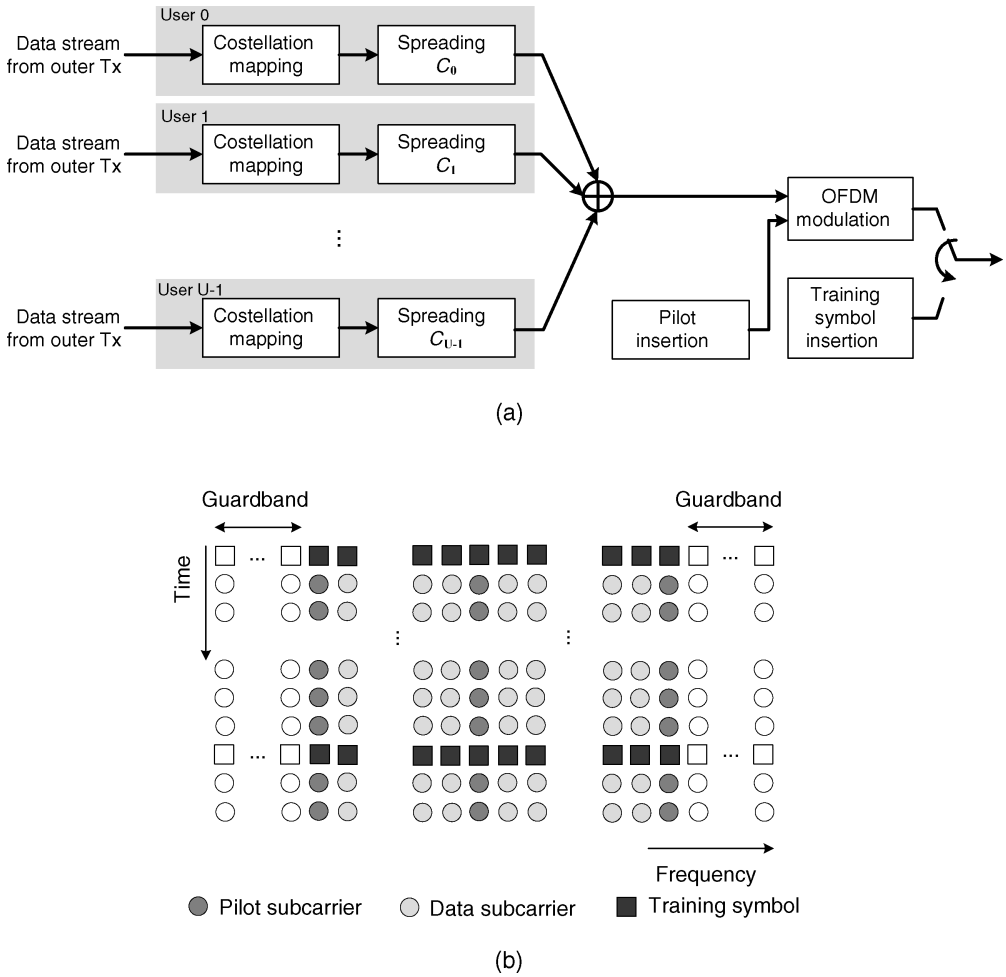


Figure 10.1. (a) Block diagram of the MC-CDMA baseband transmitter (reproduced by permission of © 2005 IEEE) and (b) allocation of pilot subcarriers and training symbols

of the user, and then are combined with signals from other users. A user may transmit several pieces of data in one MC-CDMA symbol, since the length of the OVSF code may be less than 768. To achieve maximum frequency diversity, the transmitter spreads one user's data to subcarriers that are as far apart as possible. To wit, neighboring subcarriers will carry pieces of data from different users [6].

Comb-type uniformly distributed pilot subcarriers are inserted into the transmitted symbols, as they can help the receiver to perform the tasks of synchronization and channel estimation. An OFDM modulator then transforms the frequency-domain signals to time-domain MC-CDMA signals. Note that to guarantee reliable communication, in addition to pilot subcarriers, the MC-CDMA system also transmits training symbols periodically in the time domain. The allocation of pilot subcarriers and training symbols is depicted in

Petitioner Samsung Ex-1026, 0218

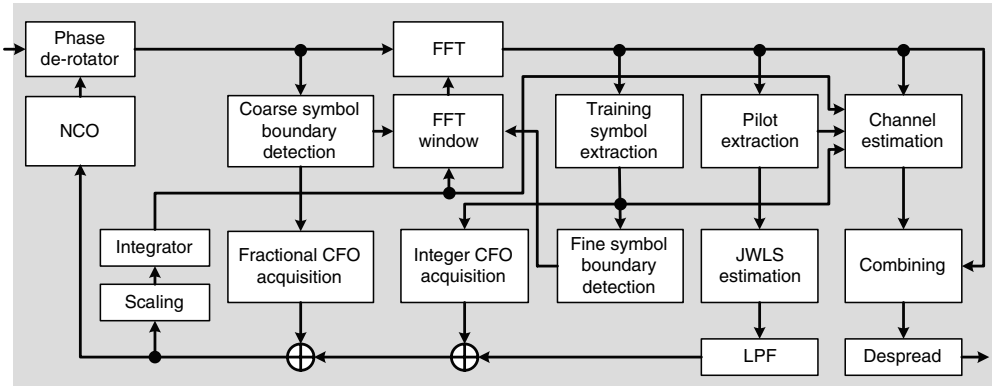


Figure 10.2. Block diagram of the MC-CDMA receiver. Reproduced by permission of © 2005 IEEE

Figure 10.1(b). The role of the training symbols is crucial to the receiver, especially during initial synchronization. A training symbol has all its odd-numbered subcarriers zero and so it contains two identical halves in the time domain [7]. With such periodic training symbols, the receiver can achieve more robust symbol boundary detection than other OFDM systems that use only a cyclic prefix. Moreover, the even-numbered subcarriers in the training symbol are differentially encoded with a *pseudo-noise* (PN) code to expedite integer carrier frequency offset acquisition.

MC-CDMA Receiver

The baseband receiver for the proposed downlink MC-CDMA communication system is very similar to a regular OFDM baseband receiver except for the combining (despreading) block that extracts a particular user's information. Figure 10.2 depicts the block diagram of such a baseband receiver. As is clearly seen, integer CFO/fractional CFO are estimated, combined and compensated in a carrier recovery loop with an NCO, a loop filter and a de-rotator. The SCO is derived from the estimated CFO and compensated in the frequency domain. The FFT window block controls the FFT input interval with information provided by the coarse symbol boundary detection and fine symbol boundary detection blocks and the overflow/underflow of SCO accumulation. After initial acquisition, a subsequent tracking mechanism which enhances the synchronization accuracy and thus communication performance is implemented [3]. For this purpose, the *weighted least squares* (WLS) algorithm block estimates the CFO/SCO and sends the result to the loop filter. Finally, the channel estimation and combining blocks restore faded subcarrier data and despread the equalized frequency-domain signals to reconstruct the detected information for the user.

After the automatic gain control circuit has settled, the baseband receiver starts to detect the coarse symbol timing by searching for training symbols. The receiver adopts a delay correlator with a delay of $N/2$, where N is the FFT size. The output of the delay correlation

Petitioner Samsung Ex-1026, 0219

is then passed to a moving average block, whose output peak indicates the end of a training symbol. Assume that the normalized CFO is given by ϵ, so $\Delta f = \epsilon f_S$. Then, the phase shift between the two identical halves of the training symbol is $\pi\epsilon$. As mentioned in Chapter 5, the normalized CFO is composed of the fractional part and the integer part. The maximum likelihood estimation of the fractional CFO, $\hat{\epsilon}_f$, can be derived simultaneously when the symbol timing is decided.

Integer CFO causes index shift in the frequency-domain subcarriers. Since the data on the even-numbered frequency-domain subcarriers of the training symbol are differentially encoded by a PN code, the frequency-domain PN-correlator approach mentioned in Section 5.3.2 is applied. After integer CFO acquisition, the receiver activates fine symbol timing detection and adjusts the symbol boundary to an ISI-free position. The frequency-domain phase rotation approach introduced in Section 5.3.1 is adopted due to its simplicity. Note that the integer CFO block and the fine symbol timing block share one complex multiplier.

The MC-CDMA signal, being based on the OFDM modulation, is also susceptible to synchronization errors. When the orthogonality among subcarriers is destroyed, both *inter-carrier interference* (ICI) and *multiple-access interference* (MAI) will be induced [8]. As such, it is imperative for the MC-CDMA receiver to incorporate accurate synchronization error estimation/compensation to guarantee acceptable performance. To deal with non-stationary channels and possible initial CFO/SCO estimation error, CFO as well as SCO must be continuously estimated and compensated. The joint WLS algorithm described in Chapter 5 is adopted for this purpose [9]. In practice, the carrier frequency and the sampling frequency are generated from the same oscillation source. Thus, CFO and SCO are related by a constant, making the estimation simpler. Furthermore, when the carrier frequency is much higher than the subcarrier spacing, the formula of joint WLS estimation in Section 5.3.3 can be simplified to

$$\hat{\delta} \approx \frac{\sum_m w_m \theta_m}{2\pi \frac{N+N_g}{N} f_c T \sum_m w_m}. \tag{10.1}$$

The estimated oscillator error is often corrupted by noise. Therefore, the receiver adopts a loop filter to remove the out-of-band noise. This filtered residual synchronization error is combined with initially acquired fractional/integer CFO. Then, a numerical-controlled oscillator and a phase de-rotator compensate the carrier frequency offset in the time domain. The SCO, being proportional to the CFO, is derived through a scaling block and then compensated in the frequency domain. Of course, overflows and underflows in the sampling phase offset need to be processed by the FFT window block.

The rest of the receiver obtains the final detection of the received signal by channel estimation, equalization and despreading of the equalized subcarrier data. Due to the variety of channel environments, the receiver has two channel estimation modes, and switches between them accordingly. *Least squares* (LS) channel estimation [10] from the training symbol is applied when the channel is stationary. On the other hand, in fast-fading channels, a shifted raised-cosine frequency-domain channel interpolation technique mentioned in Section 6.3.2 will be used [11]. A sixth-tap channel interpolator estimates the channel

Petitioner Samsung Ex-1026, 0220

responses in the data subcarriers as

$$\begin{aligned}\hat{H}_k &= \sum_{m=\lfloor k/D \rfloor-2}^{\lfloor k/D \rfloor+3} \tilde{H}_{mD} W_{k-mD} \\ &= e^{-j2\pi dk/N} \sum_{m=\lfloor k/D \rfloor-2}^{\lfloor k/D \rfloor+3} (\tilde{H}_{mD} e^{j2\pi dmD/N}) \\ &\quad \cdot \frac{\sin(\pi(k-mD)/D)}{\pi(k-mD)/D} \cdot \frac{\cos(\pi\beta(k-mD)/D)}{1-4\beta^2((k-mD)/N)^2},\end{aligned} \tag{10.2}$$

where $\tilde{H}_{mD}$ is the channel response of the m-th pilot subcarrier; D is the pilot subcarrier spacing; β is the roll-off factor; and d is the corresponding time shift.

In the MC-CDMA system, the signals that correspond to a piece of user information symbol are spread among a number of data subcarriers in an MC-CDMA symbol. The MC-CDMA baseband receiver will have to first equalize the data subcarriers and then combine them judiciously to arrive at the most signal energy from the related subcarriers for the final decision. As in the OFDM communication systems, the data on each subcarrier of the MC-CDMA signal suffer only flat fading though the channel is frequency-selective faded. Thus, a simple one-tap *frequency-domain equalizer* (FEQ) is sufficient. The proposed receiver uses the *threshold orthogonality restoring combining* (TORC) method [12], which strikes a balance between complexity and performance.

In TORC, the equalizer coefficient for subcarrier k can be expressed as

$$G_k = \begin{cases} 1/\hat{H}_k & \text{if } |\hat{H}_k| \geq h_{THR} \\ \hat{H}_k^*/|\hat{H}_k| & \text{if } |\hat{H}_k| < h_{THR} \end{cases}, \tag{10.3}$$

where h_{THR} is a threshold, below which a subcarrier is deemed in deep fade. To reconstruct the user symbol, the equalized signals are despread using that user's OVSF code.

10.1.3 Circuit Design

Word-Length Determination

Since most arithmetic circuits in the implemented receiver IC are based on fixed-point arithmetic, each signal in the receiver must be quantized to a digital signal with a certain word-length. To determine the word-lengths of the I/O signals in the baseband receiver modules, many functional simulations of the proposed receiver are conducted. The simulated SNR at the despreader output works as the performance metric. A word-length that induces little or no SNR degradation is chosen for each I/O signal. The implementation loss due to quantization noises in the fixed-point-arithmetic baseband receiver is smaller than 0.5 dB when the symbol error rate is at 10^{-5}. Signal word-lengths of several key signals in the MC-CDMA baseband receiver are listed in Table 10.2.

Petitioner Samsung Ex-1026, 0221

Table 10.2. Signal word-lengths of the major functional blocks in the MC-CDMA baseband receiver

Signal	Word-length
ADC output	12
Derotator output	14
FFT output	16
Channel estimator output	14
TORC output	12
Despreading output	14

FFT Module

Fast Fourier transform is a key component in the MC-CDMA baseband receiver, since it is both computation-intensive and communication-intensive. The FFT size (1024) is quite large in the receiver; therefore, the radix-2^3 architecture mentioned in Chapter 9 is adopted [13,14]. The FFT module needs only three complex multipliers and associated twiddle factor tables. A low-complexity complex multiplier with only three real multiplications and five real additions/subtractions is used [15]:

$$(A + jB)(C + jD) = [C \times (A - B) + B \times (C - D)] + j[D \times (A + B) + B \times (C - D)]. \tag{10.4}$$

In addition, the symmetry in the twiddle factors is exploited and the look-up tables store the sine and cosine values from 0 to $\pi/4$.

In the FFT architecture, several delay buffers of various lengths are required. Longer delay buffers are implemented using two single-port SRAM modules and some control logics, which have been introduced in Chapter 9. However, delay buffers shorter than 64 are implemented by flip-flops configured in a pointer-based configuration, which is also described in Chapter 9. A shared ring counter with some logic gates can address all the pointer-based delay buffers and save more power and complexity.

The decimation-in-frequency FFT outputs are in the bit-reverse address order. They need to be rearranged into the normal order for further frequency-domain signal processing, such as channel interpolation and despreading. Conventionally, two N-word SRAM modules working in the ping-pong fashion are used. One is being written in the bit-reverse order, while data in the other are read out in the normal order. After one cycle of N words, the two SRAM modules exchange roles. Following the concept of the in-place strategy in the memory-based FFT, a more complex, yet area-efficient, implementation using two $N/2$-word SRAM modules can be used.

Arctangent Module

In the proposed baseband receiver, several arctangent functions are needed. In the design, one arctangent module is shared among initial fractional CFO acquisition, fine symbol boundary detection and WLS CFO/SCO estimation. Note that in all three applications, the

Petitioner Samsung Ex-1026, 0222

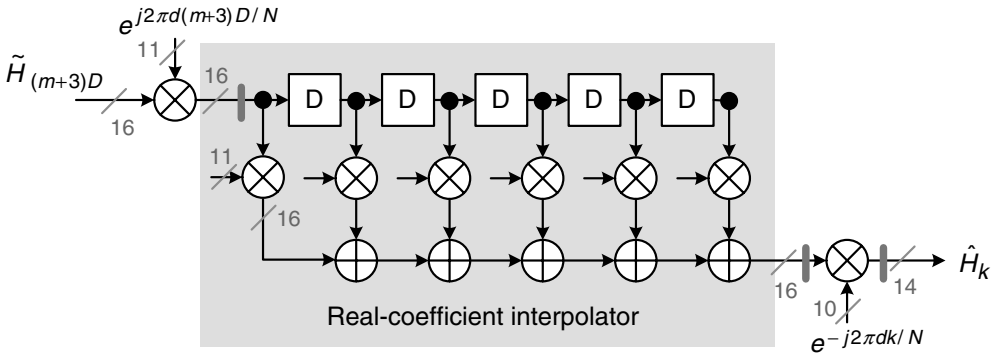


Figure 10.3. Architecture of the frequency-domain channel interpolator. Reproduced by permission of © 2005 IEEE

arctangent function has several clock cycles to complete. As a result, a folded version of the CORDIC hardware introduced in Chapter 9 is adopted.

Channel Estimation

There are two modes of channel estimation: one is the LS channel estimator in quasi-stationary channels; the other is the frequency-domain channel interpolator for fast fading channels. In the mobile cellular environments, the received signals are fast faded. Hence, accurate and up-to-date channel estimation is of paramount importance for reliable transmission. The receiver uses the shifted raised-cosine interpolator introduced in Section 6.3.2. The interpolator circuit, shown in Figure 10.3, takes in the channel responses of pilot subcarriers and applies a partial phase shift that corresponds to an offset in the time window. The result is fed to a six-tap real-coefficient interpolator, whose output is multiplied by another phase shift term that corresponds to the aforementioned time window offset.

Equalizer

The TORC equalizer coefficients in Equation (10.3) can be reformulated as

$$G_k = \begin{cases} \dfrac{1}{|\hat{H}_k|}\angle(\hat{H}_k^*) & \text{if } |\hat{H}_k| \geq h_{THR} \\ \angle(\hat{H}_k^*) & \text{if } |\hat{H}_k| < h_{THR} \end{cases}. \tag{10.5}$$

In the proposed TORC circuits, two pipelined CORDIC structures are used, as shown in Figure 10.4. One CORDIC rotates the estimated complex channel response of the k-th subcarrier back to the x-axis in the complex plane. Meanwhile, another CORDIC rotates the received frequency-domain data of that subcarrier by the same amount. Without explicitly deriving the channel phase and with only eight stages of micro-rotation in both CORDIC structures, the frequency-domain data are equalized in phase. In addition, one more initial stage is needed to rotate the input by multiples of $\pi/2$ so that the resultant output will be in

Petitioner Samsung Ex-1026, 0223

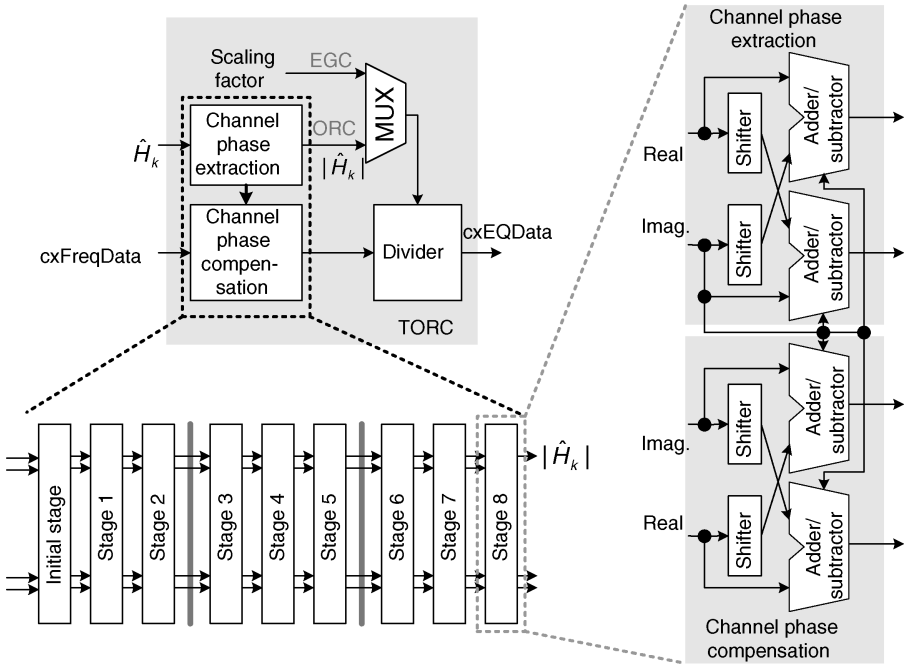


Figure 10.4. Block diagram of TORC equalization. Reproduced by permission of © 2005 IEEE

the first quadrant. The magnitude of the signal to be equalized needs also to be properly scaled. Since the CORDIC hardware computes the channel gain magnitude as a by-product, a divider scales the phase-equalized signal using this magnitude when $|\hat{H}_k| \geq h_{THR}$. Otherwise, the phase equalized signal is divided by the CORDIC scaling factor.

In the circuit design of this baseband receiver IC, several ideas have been adopted. The delay buffer design and a new bit-reverse ordering architecture save about 6 mW. The new channel estimator and the TORC equalizer also save about 10 mW of power. Finally, by turning off unused modules, more than 4 mW can be saved.

10.1.4 Experimental Results

The entire MC-CDMA baseband receiver is synthesized using a cell-based design flow and placed as well as routed by commercial software. The chip is then implemented in 0.18-μm one-poly and six-metal CMOS technology. The logic gate count and the transistor count of the chip are around 336 K and 2.1 millions, respectively. Only 12 SRAM modules (81 K bits) are used and they are placed near the periphery of the core to prevent malfunction caused by IR drops. The core size of this chip is 2.6×2.6 mm. Figure 10.5 shows a photograph of the fabricated baseband receiver chip. The fabricated chip was tested at the nominal operating frequency of 5.76 MHz. Thanks to power-saving techniques and the voltage scaling, this chip consumes only 9.9 mW from a supply voltage of 1.1 V. A summary of MC-CDMA baseband receiver IC is given in Table 10.3.

Petitioner Samsung Ex-1026, 0224

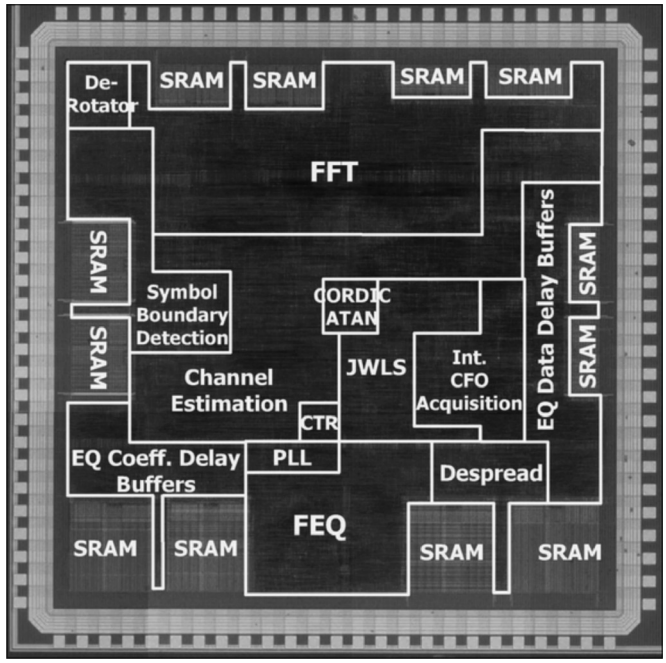


Figure 10.5. Die photograph of the proposed MC-CDMA downlink baseband receiver IC. Reproduced by permission of © 2005 IEEE

The chip's receiving performance in a multipath fading channel is also measured. The channel model used is the typical urban power-delay profile specified in the 3GPP technical document [4], in which a large number of paths ensure realistic frequency domain correlation properties. The chip's *symbol error rate* (SER) performance is almost identical to the functional simulation results. With 32 active users and 226 Kbps per user, the receiver can reach an uncoded SER on the order of 10^{-5}, given enough SNR. This corresponds to an

Table 10.3. Summary of the MC-CDMA baseband receiver chip. Reproduced by permission of © 2005 IEEE

Technology		0.18-μm 1P6M CMOS
Package		CQFP-100
Size (mm^2)	Core	2.6×2.6
	Chip	3.16×3.16
Gate count		335,823
Transistor count		2,093,944
SRAM size (bits)		81 K
Power consumption		9.9 mW @ 1.1 V, 5.76 MHz
Max. oscillator deviation		10 ppm

Petitioner Samsung Ex-1026, 0225

aggregate bit rate of more than 7.3 Mbps. In another measurement, the receiver IC with signals under different mobility is verified. With 32 active users and 120 km/hr, this receiver chip achieves satisfactory uncoded SER of around 10^{-2}. In conclusion, the proposed MC-CDMA downlink baseband receiver chip provides a low-power, high-bit-rate solution to mobile cellular communication applications.

10.2 MIMO–OFDM Cognitive Radio Receiver IC

The second system example shows the design of a MIMO–OFDM WLAN *cognitive radio* (CR) receiver IC. As wireless applications become more and more popular, the demand for bandwidth gets stronger. However, though spectrum is a scarce resource, it is not used efficiently most of the time in many areas. Cognitive radio techniques have been proposed to exploit such phenomena and mitigate the spectrum shortage problem [16]. Cognitive radio is a radio system that can sense the spectral utilization and utilize the unused spectrum without interfering with primary users. At the same time, a cognitive radio system also self-adapts the transmission parameters to increase spectral efficiency. In 2003, the Federal Communication Commission (FCC) in the United States of America initiated the commercialization of cognitive radio by issuing relevant notifications. The latest working group (WG) organized by the IEEE 802 LAN/MAN Standards Committee, IEEE 802.22 wireless regional area network (RAN) WG, mandates cognitive radio as its core technology. In addition, several CR systems have already been proposed for operation in the TV bands [17].

In this example, the proposed OFDM-based cognitive radio system not only implements channel assessment, but also provides different transmission modes to accommodate dynamic spectrum allocation. In addition, two MIMO techniques are incorporated to either obtain diversity gain or enhance transmission rate. They are the *space-time block code* (STBC) [18,19] and the *Vertical Bell Laboratory layer Space Time* (VBLAST) scheme [20]. The architecture of the proposed CR system is shown in Figure 10.6. The proposed CR system is fashioned after the IEEE 802.11a/g wireless LAN and the major distinctions include an eight-channel polyphase filter and a frequency-tunable bandpass filter (BPF) in the RF front-end, channel assessment and multi-band long preamble/data transmission in the baseband. These features are indispensable in the sensing/filtering of primary user signals and the transmission in dynamically allocated bands. This example focuses on the baseband receiver, major features of which are summarized below:

- supporting 270 dynamic spectrum allocation patterns;
- offering sensing, SNR and delay spread estimation capability;
- providing VBLAST and STBC MIMO decoding schemes;
- four clock frequencies–20/40/80/160 MHz;
- variable-length FFT–64/128/256/512; and
- data rate of up to 1.404 Gbps.

10.2.1 System Overview

The proposed CR system operates at either the lower or the upper 5-GHz UNII bands with 80-MHz and 160-MHz bandwidths, respectively. The adopted OFDM parameters are compatible with those in the IEEE 802.11a/g standard, except that the FFT size and the

Petitioner Samsung Ex-1026, 0226

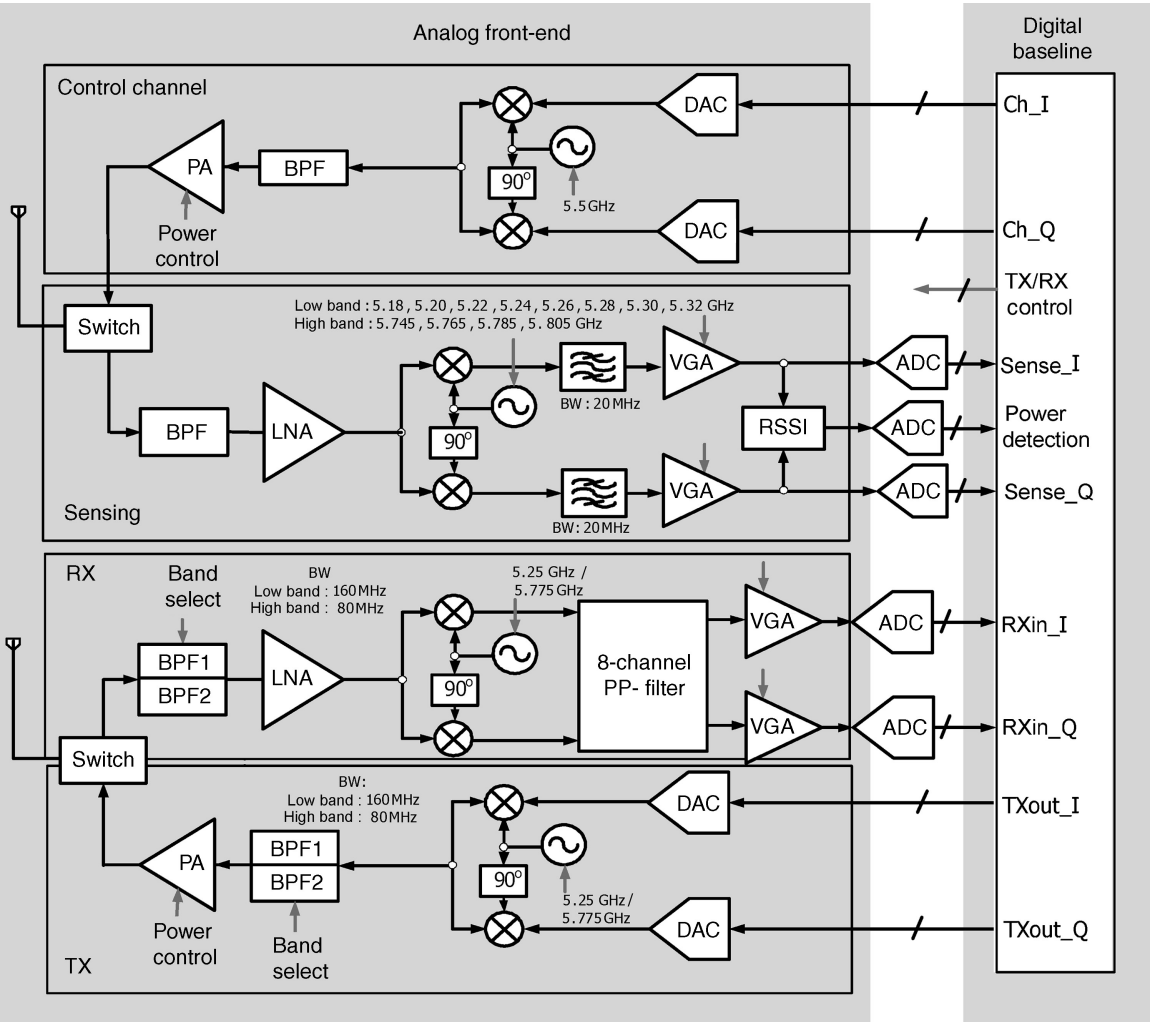


Figure 10.6. Block diagram of the analog front-end and digital baseband of the proposed cognitive radio system

Petitioner Samsung Ex-1026, 0227

Table 10.4. CR system parameters

Band	Lower UNII	Upper UNII
Frequency range (GHz)	5.15–5.35	5.725–5.825
Number of data subcarriers	48–468	48–244
Bandwidth (MHz)	20/40/80/160	20/40/80
FFT length	64/128/256/512	64/128/256
Modulation	QPSK/16-QAM/64-QAM OFDM	
Subcarrier spacing	312.5 KHz	
Symbol duration	4 μs	

number of subcarriers have more options due to variable bandwidth. Table 10.4 lists key parameters of the signal in the proposed CR system.

Packet Format

The packet format of the CR system is shown in Figure 10.7. As with the IEEE 802.11a/g packets, 10 short preambles start a CR packet. Then two or four long preamble symbols, depending on the mode (SISO or MIMO) used, are transmitted. These are followed by the SIGNAL field and the data symbols. In the MIMO mode, the Alamouti code [18] is adopted to encode the long preambles, making channel estimation more reliable.

Spectral Design

In the CR system, due to dynamic spectrum allocation, the data/pilot/guard subcarriers can have many possible configurations. A single 20-MHz band requires 12 subcarriers reserved for guard bands. Nonetheless, some or even all of the available 20-MHz bands can be contiguous. As such, relatively fewer guard bands are necessary, and the unused guard-band subcarriers can be freed for data transmission. In the proposed system, the number of available data subcarriers is given by

$$N_{ds} = 60N_b - 12N_{ck}, \tag{10.6}$$

where N_{ds}, N_b and N_{ck} are the number of data subcarriers, number of available 20-MHz bands and number of consecutive chunks of available spectrum, respectively. The number of pilot subcarriers are $4N_b$ and N_{ds} ranges from 48 to 468, corresponding to $N_b = 1$ and 8, respectively. Therefore, with 64-QAM modulation, 468 data subcarriers and two-antenna spatial multiplexing, the maximum achievable data rate of the CR system is 1404 Mbps.

10.2.2 Architecture and Circuit Design

The architecture of the CR baseband receiver is shown in Figure 10.8, consisting of three major parts: initial synchronization, CFO/SCO tracking and frequency-domain data recovery. The receiver first examines the availability of a particular targeted band, turning

Petitioner Samsung Ex-1026, 0228

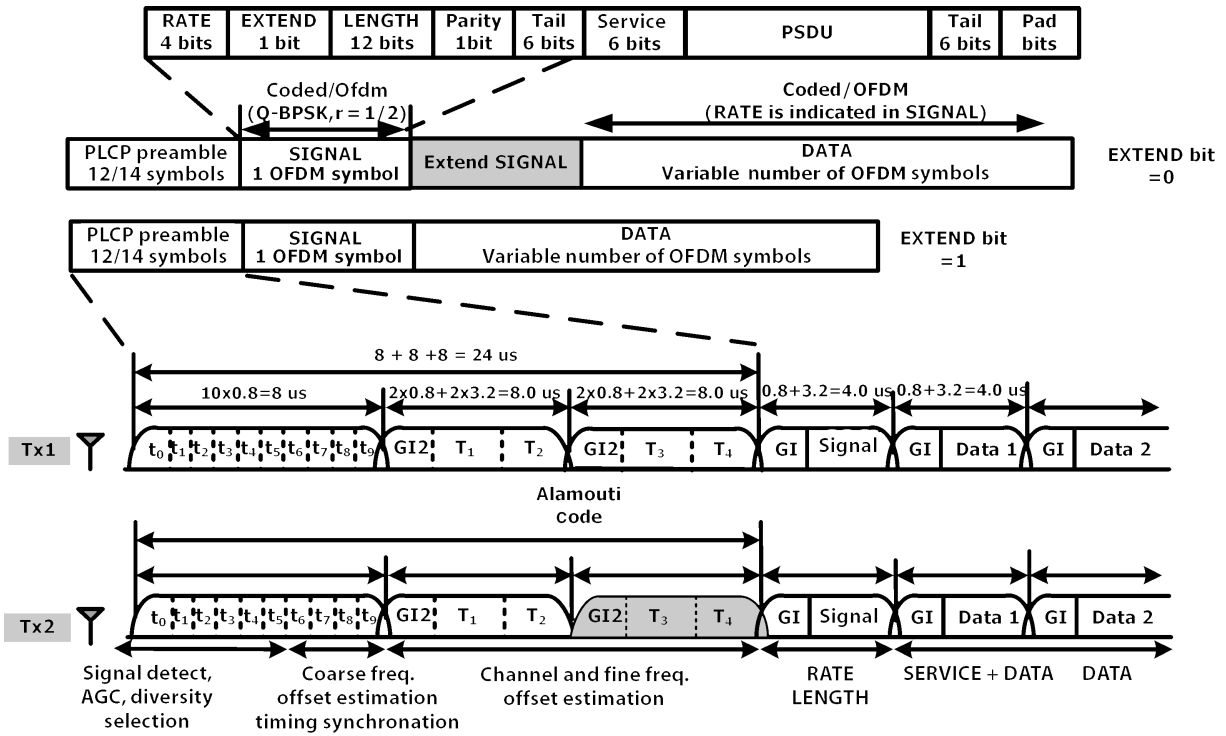


Figure 10.7. Packet format of the CR system

Petitioner Samsung Ex-1026, 0229

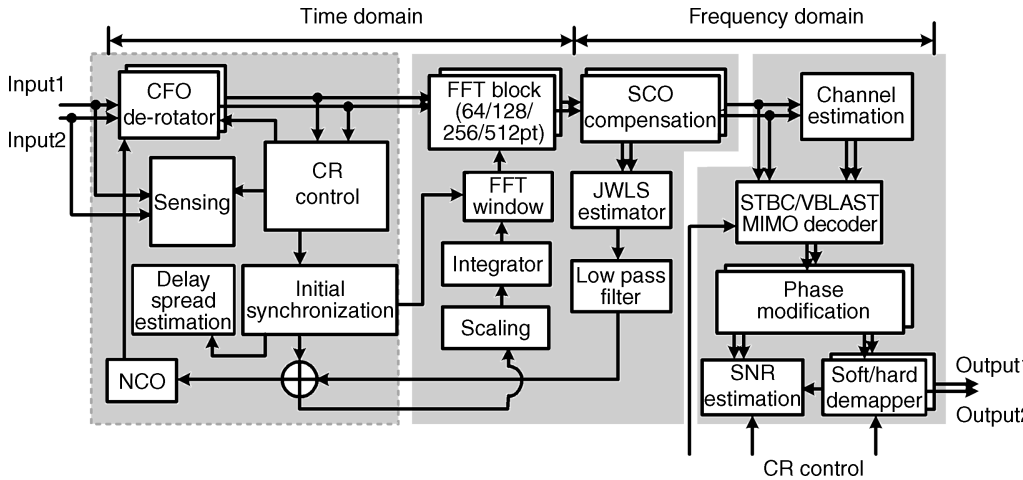


Figure 10.8. Architecture of the CR baseband receiver

on only sensing and CR control modules. After verifying that the targeted band is idle, the receiver starts coarse and fine timing synchronization, during which the coarse and fine CFO estimation result can also be obtained. The joint weighted least square (JWLS) algorithm introduced in Section 5.3.3 is adopted for the CFO/SCO tracking loops. The CFO-compensated time-domain signal is transformed to the frequency domain by the FFT block and the SCO effect compensated. Channel estimates can be extracted from the compensated signal. Subsequently, with the estimated channel matrix, the STBC/VBLAST processing block performs MIMO decoding. Finally, the phase modification block and soft/hard demapper produce the receiver output signal to the forward error correction (FEC) decoder.

Most of the circuits in the CR receiver are similar to those of the MC-CDMA baseband receiver IC described in the previous section. There are, however, some new circuits, namely the MIMO detector and the SNR estimator, which will be described next.

VBLAST/STBC MIMO Detector

This receiver supports two types of MIMO detection: VBLAST for spatial multiplexing data and the STBC decoder for STBC-coded data. Both algorithms require considerable hardware due to their arithmetic complexity. As such, common hardware that can perform both MIMO detection schemes is devised. In Figure 10.9(a), the detector is configured to work as a VBLAST detector for two-antenna spatially multiplexed OFDM subcarrier data. Signal $Z^{(q)}$ is the received signal from the q-th receive antenna and $H^{(q,p)}$ is the channel gain from the p-th transmit antenna to the q-th receive antenna in the 2×2 MIMO channel. Another circuit first finds the path with stronger energy, whose result is then used to control the

Petitioner Samsung Ex-1026, 0230

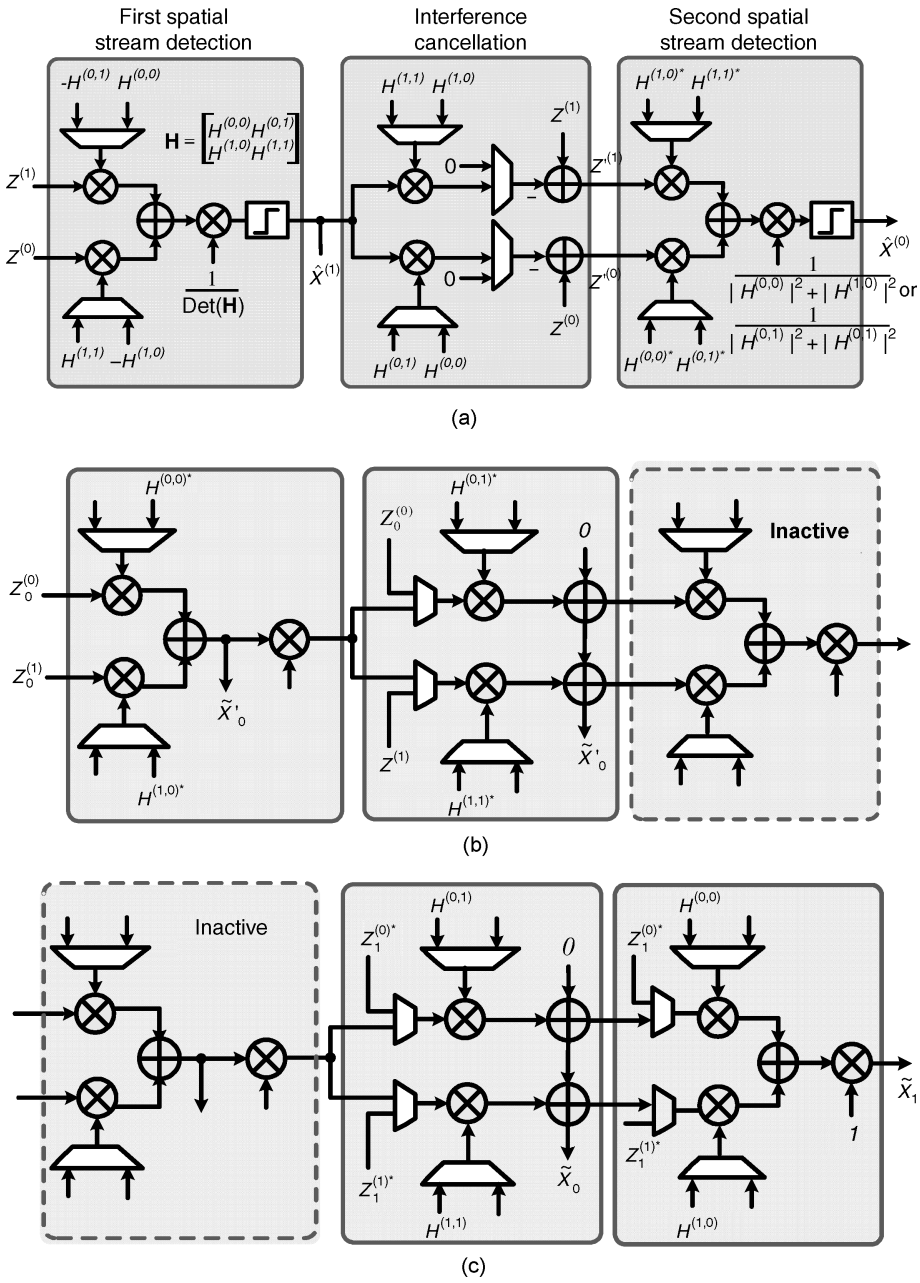


Figure 10.9. Configurations of the proposed MIMO detector working (a) as the VBLAST detector, (b) as the STBC decoder (first phase) and (c) as the STBC decoder (second phase)

Petitioner Samsung Ex-1026, 0231

operation of the MUXs in the decoder to select the correct channel gains. In the first stage, the QPSK data transmitted from the stronger spatial stream (assuming transmit antenna 1) are detected according to the following formula:

$$\hat{X}^{(1)} = \mathrm{sgn}\left(\frac{-H^{(1,0)}Z^{(0)} + H^{(0,0)}Z^{(1)}}{H^{(0,0)}H^{(1,1)} - H^{(0,1)}H^{(1,0)}}\right), \tag{10.7}$$

where $\hat{X}^{(1)}$ is the detected signal from the first transmit antenna. After detecting one of the transmitted signals, the *ordered successive interference cancellation* (OSIC) principle is then applied to eliminate the interference component corresponding to $X^{(1)}$, as illustrated in the second stage of Figure 10.9(a). The data transmitted from the antenna with smaller energy can then be detected in the third stage, using

$$\hat{X}^{(0)} = \mathrm{sgn}\left(\frac{H^{(0,0)*}Z'^{(0)} + H^{(1,0)*}Z'^{(1)}}{|H^{(0,0)}|^2 + |H^{(1,0)}|^2}\right), \tag{10.8}$$

where $Z'^{(0)}$ and $Z'^{(1)}$ are the received signals with the components from transmit antenna 1 cancelled.

Figure 10.9(b) and (c) illustrate the two phases of the detector operation when configured for STBC equalization. Without explicitly showing the pipelining overhead, the MIMO detector circuit requires two clock cycles to decode two received STBC coded vectors in two OFDM symbols: $[Z_0^{(0)} \;\; Z_0^{(1)}]$ and $[Z_1^{(0)} \;\; Z_1^{(1)}]$. Note that each vector contains two signals–one from receive antenna 0 and the other from receive antenna 1.

During the first clock cycle (phase), the first stage and second stage process the received vectors in two consecutive OFDM symbols, while the third stage is idle. The partial detections for X_0 and X_1 are given by

$$\begin{aligned} \tilde{X}_0{}' &= H^{(0,0)*}Z_0^{(0)} + H^{(1,0)*}Z_0^{(1)}, \\ \tilde{X}_1{}' &= H^{(0,1)*}Z_0^{(0)} + H^{(1,1)*}Z_0^{(1)}. \end{aligned} \tag{10.9}$$

In the second clock cycle (phase), the detected X_0 and X_1 are computed using the second and third stages, while the first stage is deactivated. The formulas in the second clock cycle are

$$\begin{aligned} \tilde{X}_0 &= \tilde{X}_0{}' + H^{(0,1)}Z_1^{(0)*} + H^{(1,1)}Z_1^{(1)*}, \\ \tilde{X}_1 &= \tilde{X}_1{}' - H^{(0,0)}Z_1^{(0)*} - H^{(1,0)}Z_1^{(1)*}. \end{aligned} \tag{10.10}$$

SNR Estimator and Delay Spread Estimator

To estimate the SNR without incurring too much computation complexity, all signal samples are mapped to one cluster. Hence, each detected subcarrier signal sample is mapped to the first quadrant and then shifted by a vector that makes the quantized decision become the

Petitioner Samsung Ex-1026, 0232

QPSK signal point in the first quadrant. Signal variance is then computed by subtracting a scaled power of the sum of signal samples from the sum of squared signal samples. Dividing the signal power by the variance then gives the SNR estimate. This approach is more efficient, in terms of hardware and storage, than the canonical approach of computing the signal average and then summing the squared differences between the signal samples and the average.

The matched filter structure presented in Section 5.3.1 is adopted for fine symbol timing detection in this CR receiver. The matched filter output waveform can also be used to estimate the channel delay spread, as this waveform indicates the channel impulse response. The root-mean-square (RMS) delay spread is computed according to the following formula:

$$\tau_{\mathrm{RMS}} = \left(\sum_r \frac{|\hat{h}_r|^2}{\sum_r |\hat{h}_r|^2} \tau_r^2 \right) - \left(\sum_r \frac{|\hat{h}_r|^2}{\sum_r |\hat{h}_r|^2} \tau_r \right)^2, \tag{10.11}$$

where $\hat{h}_r$ and τ_r are the estimated channel response and the delay of the r-th path, respectively.

Hardware Implementation Techniques

Major hardware-saving and power-reduction techniques applied in the receiver range from the algorithm level, the architecture level, to the circuit level. They include:

- quantized matched filter coefficients;
- no-look-up-table linear-approximated sin/cos functions;
- hardware-shared STBC/VBLAST detector;
- delay-buffer memory module sharing; and
- single-port SRAM for long delay buffers and register file for short delay buffers.

The above techniques not only make the receiver more compact and consume less power, but also make it possible to operate the receiver IC at the nominal 160-MHz clock rate.

10.2.3 Experimental Results

In light of testability, power ring width and the number of power pads are properly decided under power simulation results at the maximum clock frequency. Furthermore, scan chain insertion and SRAM-BIST are included. The receiver IC is realized through the cell-based approach using a commercial automatic place and route tool. Using 0.18-μm single-poly six-metal (1P6M) CMOS technology, the chip integrates 280K gates in die that measure $3.58 \times 3.59\,\mathrm{mm}^2$ ($2.69 \times 2.69\,\mathrm{mm}^2$ core size). The microphotograph of the proposed MIMO-OFDM-based CR receiver chip is illustrated in Figure 10.10. A summary of the chip is given in Table 10.5.

Measured results indicate that the receiver chip can tolerate the SCO/CFO effects quite well, with implementation loss of about 0.5–2 dB. The dynamic spectrum access capability

Petitioner Samsung Ex-1026, 0233

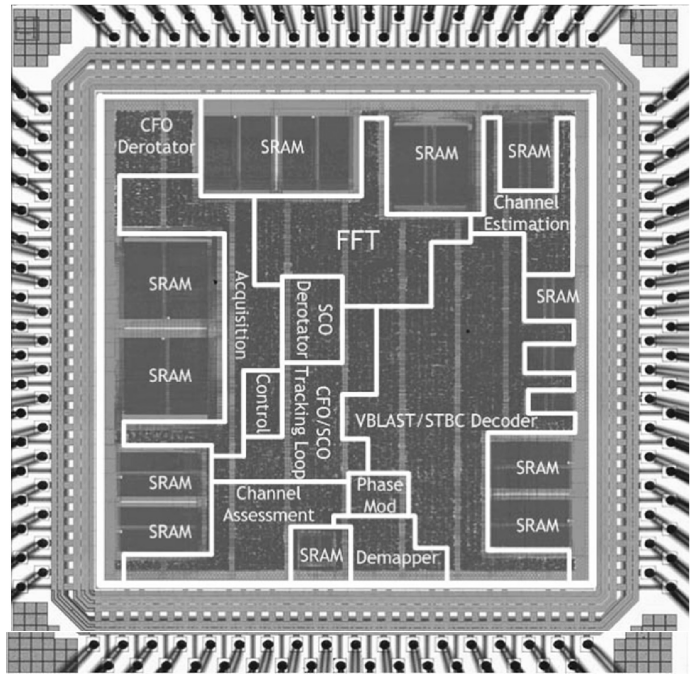


Figure 10.10. Die photo of the proposed MIMO–OFDM CR baseband receiver IC

is demonstrated by comparing the BER of the chip's performance when running under different spectrum allocation patterns. The receiver IC can successfully support up to 1.404-Gbps data transmission when occupying 160-MHz bandwidth using 64-QAM modulation. Finally, the channel SNR and delay spread estimation performances are measured. The channel delay spread estimations were tried 50 times and an estimation error of around 5% was achieved as channel SNR increased. The best SNR estimation performance is achieved naturally under the noise-only channel and a BPSK modulation, with an average error of about 2 dB. In the other channel scenarios and modulation schemes, the estimator, although

Table 10.5. Summary of the MIMO–OFDM CR baseband receiver chip

Clock frequency	20/40/80/160 MHz
Power supply	1.8 V
Power dissipation	720 mW @ 160 MHz
Process technology	0.18-μm 1P6M CMOS
Gate count	280 K
Die size	$3.58 \times 3.59\,\text{mm}^2$
Package	144-pin CQFP
Maximum data rate	1.404 Gbps

Petitioner Samsung Ex-1026, 0234

it exhibits a larger degree of deviation, still reports SNR estimates that scale rather linearly with the actual SNR. As such, this estimator can assess *relative* channel SNR quite reliably.

In conclusion, the proposed MIMO–OFDM-based CR system can support an overall of 270 configurations of spectrum allocation. With the integration of sensing, channel SNR estimation and delay spread estimation functions, the baseband CR receiver can complete basic CR operation requirements. The baseband receiver chip is fabricated in a 0.18-μm 1P6M CMOS process, with an area of 12.8 mm^2. The power consumption of the chip under a 160-MHz clock rate and a 1.8-V power supply is 720 mW, when achieving a maximum data rate of 1.404 Gbps. Finally, not only does the proposed MIMO–OFDM-based system serve as a promising candidate for the increasingly important CR communication, but also the CR baseband receiver IC provides a useful design prototype for building OFDM-based CR communication systems.

Bibliography

[1] 3GPP, 'Spreading and modulation (FDD)', 3G TR 25.213 v3.3.0, Oct. 1999.

[2] S. Hara and R. Prasad, 'Overview of multicarrier CDMA', *IEEE Commun. Mag.*, vol. 35, no. 12, Dec. 1997, pp. 126–133.

[3] P. Y. Tsai, H. Y. Kang and T. D. Chiueh, 'Design of a baseband transceiver for multi-carrier CDMA communication', *EURASIP J. Applied Signal Processing*, no. 11, July 2005, pp. 1645–1655.

[4] 3GPP, 'Deployment Aspects', TR 25.943, v2.1.0, Jun. 2001.

[5] 3GPP, 'Technical specification group radio access network, feasibility study for orthogonal frequency division multiplexing (OFDM) for UTRAN enhancement', TR 25.892, V6.0.0 (2004–06).

[6] K. Fazel and L. Papke, 'On the performance of convolutionally-coded CDMA/OFDM for mobile communication system', in *Proc. of IEEE Personal, Indoors, and Mobile Radios Communications*, Sep. 1993, pp. 468–472.

[7] T. M. Schmidl and D. C. Cox, 'Robust frequency and timing synchronization for OFDM', *IEEE Trans. Commun.*, vol. 45, Dec. 1997, pp. 1613–1621.

[8] Y. Kim, S. Choi, C. You and D. Hong, 'BER computation of an MC-CDMA System with carrier frequency offset', in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Mar. 1999, pp. 2555–2558.

[9] P. Y. Tsai, H. Y. Kang and T. D. Chiueh, 'Joint weighted least squares estimation of carrier frequency offset and timing offset for OFDM systems over multipath fading channel', *IEEE Trans. Veh. Technol.*, vol. 54, Jan. 2005, pp. 211–224.

[10] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson and P. O. Borjesson, 'On channel estimation in OFDM systems', in *Proc. of IEEE Vehicular Technology Conference*, Jul. 1995, pp. 815–819.

[11] P. Y. Tsai and T. D. Chiueh, 'Frequency-domain interpolation-based channel estimation in pilot-aided OFDM systems', in *Proc. of IEEE Vehicular Technology Conference*, vol. 1, May 2004, pp. 420–424.

[12] R. Le Gouable and M. Helard, 'Performance of single and multi-user detection techniques for a MC-CDMA system over channel model used for HIPERLAN2', in *Proc. of IEEE International Symposium on Spread Spectrum Techniques and Applications*, 2000, pp. 718–722.

[13] L. Jia, Y. Gao, J. Isoaho and H. Tenhunen, 'A new VLSI-oriented FFT algorithm and implementation', in *Proc. of IEEE ASIC Conference*, 1998, pp. 337–341.

[14] Y. T. Lin, P. Y. Tsai and T. D. Chiueh, 'Low-power variable-length fast Fourier transform processor', *IEE Proc. -Comput. Digit. Tech.*, vol. 152, no. 4, Jul. 2005, pp. 499–506.

[15] A. Wenzler and E. Luder, 'New structures for complex multipliers and their noise analysis', in *Proc. of IEEE International Symposium on Circuits and Systems*, vol. 2, May 1995, pp. 1432–1435.

[16] J. Mitola III and G. Q. Maguire Jr, 'Cognitive radio: making software radios more personal', *IEEE Pers. Commun.*, vol. 6, no. 4, Aug. 1999, pp. 13–18.

[17] D. Cabric, S. M. Mishra and R. W. Broderson, 'Implementation issues in spectrum sensing for cognitive radios', in *Proc. of Asilomar Conference on Signal, System and Computers*, 2004, pp. 772–776.

Petitioner Samsung Ex-1026, 0235

[18] S. M. Alamouti, 'A simple transmit diversity technique for wireless communications', *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, Oct. 1998, pp. 1451–1458.
[19] V. Tarokh, H. Jafarkhani and A. R. Calderbank, 'Space–time block coding for wireless communications: Performance results', *IEEE J. Select. Areas Commun.*, vol. 17, no. 3, Mar. 1999, pp. 451–460.
[20] G. J. Foschini, 'Layered space–time architecture for wireless communication in a fading environment when using multi-element antenna', *Bell Labs Technical Journal*, vol. 1, no. 2, 1996, pp. 41–59.

Petitioner Samsung Ex-1026, 0236

Index

8PSK 9
adaptive filter 137
adaptive modulation and coding (AMC) 4
adjacent channel interference (ACI) 15, 180, 181
amplitude shift keying (ASK) 9, 13–14
analog-to-digital converter (ADC) 176, 178, 179, 180–181
automatic gain control (AGC) 181

Bell Labs Layered Space-Time, diagonal (D-BLAST) 161
Bell Labs Layered Space-Time, horizontal (H-BLAST) 162
Bell Laboratory layer Space Time, vertical (V-BLAST) 161, 164
Barker codes 37
beamforming 144
belief propagation 54
best linear unbiased estimator (BLUE) 100–101
beyond 3G (B3G) 233
bit reverse 198, 201, 239
bit-true design 173, 189, 191–192
blind estimator 116
branch metrics 51–52
butterfly 197, 200–202, 203, 205–206

canonical sign digit 221, 230
capacity 144, 145–146
carrier frequency offset (CFO) 68–69, 76, 86, 88, 100–107
carrier phase error 86
c-element 213
channel estimation 115–116
channel impulse response (CIR) 96, 97, 98, 120, 126
chip 38
chip rate 38–39
clipping error 176, 178, 183
clipping noise 176, 178
code rate 42, 48, 49, 53
cognitive radio 243
coherence bandwidth 63, 118
coherence distance 66–67, 118
coherence time 68
coherent demodulation 85
common phase error (CPE) 70, 107
constraint length 49
continuous-phase frequency shift keying (CPFSK) 11, 14
convolutional code 48–50
coordinate rotation digital computer (CORDIC) 218–220, 221–223, 225
correlator bank 98, 100
crest factor (CF) 176
cyclic codes 44–45
cyclic prefixing (CP) 20, 24–25, 91–92

OFDM Baseband Receiver Design for Wireless Communications Tzi-Dar Chiueh and Pei-Yun Tsai
© 2007 John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82234-0

Petitioner Samsung Ex-1026, 0237

data-aided estimator 116
DC offset 70, 72–73, 76
decimation-in-frequency FFT 196, 198
decimation-in-time FFT 196–198
delay and correlate (DC) 92–93, 152
de-spreading 35, 38
digital modulation 9, 11
digital audio broadcasting (DAB) 2, 26
digital video broadcasting (DVB) 2, 3, 28
distribution, Gaussian 122, 154, 178, 179
distribution, multimodal 179
diversity 143, 146–147, 161–162
diversity, frequency 147, 233, 235
diversity, polarization 147
diversity, receive 143, 147
diversity, space 147
diversity, time 147
diversity, transmit 143, 147
Doppler spectrum 65–66, 78, 79, 135
dynamic range 23, 173–174, 176, 178, 180, 183, 184

effective number of bits (ENOB) 176, 178
eigen-decomposition 123, 145, 146
energy leakage 121, 130
equalizer 138–139, 240
erasures 31, 48, 53
error locations 48
error values 48

fast fading 67, 103, 104, 118, 138, 139, 150
fast Fourier transform (FFT) 195–208
FFT, radix-2 196–198
FFT, radix-4 199–200
FFT, radix-8 200–201
finite field 45–46
finite precision effect 174, 175, 181, 187, 188
FIR filter 98, 181, 187
flat fading 26, 63–64, 147
format propagation 192
frequency-domain equalizer (FDE) 138, 238
frequency-selective fading 4, 17–18, 63–64, 132

gated clock 212
Gaussian filter 15
generator matrix 42, 43, 53, 54
Gold code 37
guard band 18, 20, 26, 28, 31, 32, 125, 126, 234, 245
Guassian-filtered frequency shift keying (GFSK) 40

Hamming distance 41, 51
horizontal step 54–55

inter-carrier interference (ICI) 18, 20, 68, 70, 72, 88–89, 108, 110, 139, 140
inter-modulation distortion (IMD) 74
interpolative design 190
interpolator 76, 109, 110, 111, 130, 131, 134, 135, 240
inter-symbol interference (ISI) 20, 87, 89, 90, 91, 92, 94, 96, 98
IQ Imbalance 70–72, 76

jitter 70, 109

least-squares (LS) 124–125, 237, 240
line of sight (LOS) 60, 61, 67
linear code 42, 43, 44, 45
Linear feedback shift register (LFSR) 44
linear least squares (LLS) 105
link adaptation 4
local annotation 192
local oscillation (LO) 40, 68, 70, 71, 85, 108
look-up table (LUT) 185–186
low-density parity-check (LDPC) 53–54

matched filter 96–98, 100, 102, 103, 250
maximal-length sequence 35–37
maximum Doppler frequency 65, 67, 78, 234
maximum likelihood (ML) 51, 93, 94, 100, 124–125, 154–155, 161, 166–167
Meggit decoder 45
memory-based FFT architecture 201–202
multiple input multiple output (MIMO) 4, 143–144, 145
MIMO detector 148, 247, 249
MIMO-OFDM 147–149
minimum free distance 50, 53
minimum mean squared error (MMSE) 93, 94, 122–123, 128–129, 131, 133–134, 138, 163–164
minimum shift keying (MSK) 11, 14–15

Petitioner Samsung Ex-1026, 0238

multi-frequency network 2
multi-carrier CDMA (MC-CDMA) 233–238
multi-path delay commutator 203, 205, 206
multipath fading 60, 61, 74, 87, 120, 147, 242
multiple access 31, 32
multiple access, code-division (CDMA) 3, 4, 33, 35, 37
multiple access, frequency-division (FDMA) 7, 31, 32, 33, 34
multuipleac access, time-division (TDMA) 31, 32, 33, 34
multiple input single output (MISO) 143, 146
multiple-access interference (MAI) 237

non-data-aided estimator 116
numerical-controlled oscillator (NCO) 108, 195, 225, 236

ordered successive interference cancellation (OSIC) 164, 165, 167, 249
orthogonal frequency-division multiplexing (OFDM) 18, 19–20, 22, 23, 24, 25, 26, 28

parity-check matrix 42–44, 53–54
path loss 60–61
peak-to-average power ratio (PAPR) 23–24, 178, 180
phase noise 68, 70, 76, 88–89, 107
pipeline FFT architecture 202, 208
PN sequence 154
power delay profile (PDP) 62–63
primitive element 46
primitive polynomial 35, 36, 37, 45
processing gain 39
propagation error 183–185, 187, 188
pseudo random (PN) code 35, 37, 38, 39, 40, 97, 104, 236, 237
puncturing 53

quadrature amplitude modulation (QAM) 11, 13, 14, 19
quantization error 176, 186, 231
quantization noise 176, 177, 178, 184, 188
quaternary PSK (QPSK) 9, 15, 40

radio resource management (RRM) 4
raised-cosine filter 15, 17
ring counter 211, 212, 213, 214, 216
RMS delay spread 63, 250
RMS Doppler spread 65
RMS angle spread 68
round-off error 182

Sampling clock offset (SCO) 69, 76, 86, 89, 91, 104–108, 110, 112, 155
shadowing 60, 61
shift registers 36, 37, 45, 48, 49, 50, 208, 209, 211
signal constellation 9, 11, 74, 234
signal-to-interference-and-noise-ratio (SINR) 174
single frequency network (SFN) 2
single input multiple output (SIMO) 143, 146
single input single output (SISO) 143, 145, 146
single-path delay feedback 205–208
singular value decomposition (SVD) 168, 169
slow fading 67, 118
space-frequency block code (SFBC) 4, 143
space-time block code (STBC) 4, 143, 243, 247, 249, 250
space-time trellis code (STTC) 4, 143
space-time-frequency block code (STFBC) 143
spatial multiplexing 2, 144, 158, 161, 245, 247
spatial de-correlation 167–168
sphere decoding 166–167
spread, angle 67–68
spread, delay 62–63, 249–250
spread, Doppler 64–65, 80
spreading factor 38–39, 233, 234
spurious free dynamic range (SFDR) 229
symbol timing offset 86, 90, 99
synchronization 85–87, 90–91, 149, 151
syndromes 43, 44, 45, 47, 48
systematic code 42

threshold orthogonality restoring combining (TORC) 238, 240
trellis 50–52
truncation error 182, 183, 185, 187
twiddle factors 200, 201, 203, 206, 230

vertical step 54, 56
Viterbi algorithm 51–52

Petitioner Samsung Ex-1026, 0239

weighted least squares (WLS) 106, 107, 155, 236, 237, 239, 247
wide-sense stationary (WSS) 120
wireless local area network (wireless LAN) 5, 6–7, 22, 28
wireless metropolitan area network (wireless MAN) 7, 28
wireless network systems 5
wireless personal area network (wireless PAN) 5, 6, 40
wireless wide area network (wireless WAN) 8

zero padding (ZP) 20
zero-forcing (ZF) 163, 164