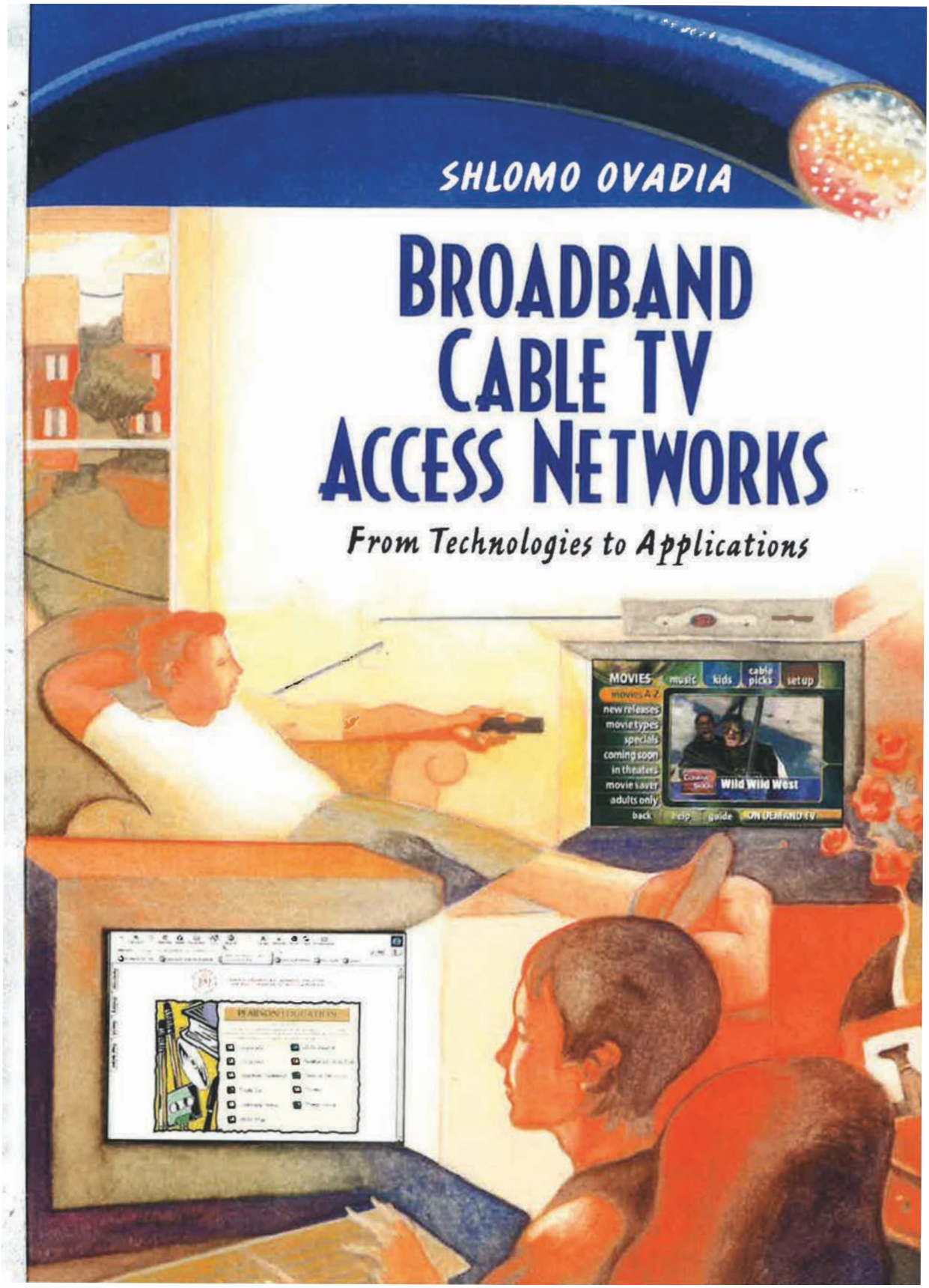




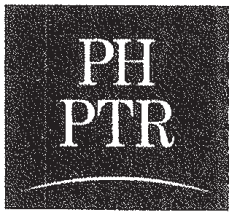

DISH 1034
Dish Network v. Entropic Commc'ns
IPR2024-00393

# BROADBAND CABLE TV ACCESS NETWORKS:

## FROM TECHNOLOGIES TO APPLICATIONS

**SHLOMO OVADIA**

*Intel Corporation*







Prentice Hall PTR
Upper Saddle River, NJ 07458

*Library of Congress Cataloging-in-Publication Data*

Ovadia, Shlomo
Broadband Cable TV Access Networks—From Technologies to Applications/ Shlomo Ovadia.
Includes bibliographical references and index.
ISBN 0-13-086421-8
1. Cable TV Access. I. Title

QA9.64 .M46 2000511.1--dc21 00-054756

*Production Editor:* Rose Kernan
*Acquisitions Editor:* Bernard Goodwin
*Editorial Assistant:* Michelle Vincenti
*Marketing Manager:* Dan DePasquale
*Manufacturing Manager:* Alexis R. Heydt
*Cover Designer:* Nina Scuderi
*Cover Design Direction:* Jerry Votta
*Composition:* Shlomo Ovadia
*Cover:* TV image is courtesy of DIVA




© 2001 Prentice Hall PTR
Prentice-Hall, Inc.
Upper Saddle River, NJ 07458

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the author and publisher.

The publisher offers discounts on this book when ordered in bulk quantities. For more information contact: Corporate Sales Department, Prentice Hall PTR, One Lake Street, Upper Saddle River, NJ 07458. Phone: 800-382-3419; Fax: 201-236-7141; E-mail: corpsales@prenhall.com
Intel is not affiliated with this book

Printed in the United States of America

ISBN 0-13-086421-8

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Pearson Education Asia Pte. Ltd.
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

*Library of Congress Cataloging-in-Publication Data*

Ovadia, Shlomo
Broadband Cable TV Access Networks—From Technologies to Applications/ Shlomo Ovadia.
Includes bibliographical references and index.
ISBN 0-13-086421-8
1. Cable TV Access. I. Title

QA9.64 .M46 2000511.1--dc21 00-054756

*Production Editor:* Rose Kernan
*Acquisitions Editor:* Bernard Goodwin
*Editorial Assistant:* Michelle Vincenti
*Marketing Manager:* Dan DePasquale
*Manufacturing Manager:* Alexis R. Heydt
*Cover Designer:* Nina Scuderi
*Cover Design Direction:* Jerry Votta
*Composition:* Shlomo Ovadia
*Cover:* TV image is courtesy of DIVA




© 2001 Prentice Hall PTR
Prentice-Hall, Inc.
Upper Saddle River, NJ 07458

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the author and publisher.

The publisher offers discounts on this book when ordered in bulk quantities. For more information contact: Corporate Sales Department, Prentice Hall PTR, One Lake Street, Upper Saddle River, NJ 07458. Phone: 800-382-3419; Fax: 201-236-7141; E-mail: corpsales@prenhall.com
Intel is not affiliated with this book

Printed in the United States of America

ISBN 0-13-086421-8

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Pearson Education Asia Pte. Ltd.
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

*To my parents, Mazal and Moshe,*
*My wife, Nina, and our children,*
*Yafit, Shlomit, Binyamin, and Leora*

תם ונשלם שבח לאל בורא עולם

# CONTENTS

**Preface** xiii

**1 Broadband Hybrid Fiber/Coax Access Networks Overview** 1

*1.1 Introduction* 2
*1.2 Traditional Cable TV Networks* 5
*1.3 Two Way HFC Access Networks* 6
1.3.1 Subscriber's Home Terminal 6
1.3.2 Cable Modems 9
1.3.3 IP Telephony 10
*1.4 Competing Access Technologies* 12
1.4.1 Asymmetric Digital Subscriber Line (ADSL) 12
1.4.2 Fiber-In-The-Loop (FITL) 15
1.4.3 Direct Broadcast Satellite (DBS) 17
1.4.4 Multichannel Multipoint Distribution Service (MMDS) 18
References 19

**2 Basic Cable TV Background: Modulation, Signal Formats and Coaxial Cable Systems** 21

*2.1 Analog Modulated Video Signal Formats* 21
2.1.1 NTSC and AM-VSB Video Signals 21
2.1.2 NTSC Signal Test Parameters 25
2.1.3 PAL and SECAM Video Signals 27
*2.2 Digital Video and Audio Signals* 29
2.2.1 MPEG-1 Standard 29
2.2.2 MPEG-2 Standard 33
2.2.3 MPEG and AC-3 Audio 35
2.2.4 MPEG-4 Standard 36
2.2.5 Other Digital Video Standards 37
*2.3 Cable TV Frequency Plans* 37
*2.4 Coaxial Cable TV Components and Systems* 42
2.4.1 Coaxial Cable 42
2.4.2 RF Amplifiers 44

2.4.3 Taps 47
*2.5 Multichannel Coaxial Cable TV Systems* *47*
2.5.1 CNR of a Single and Cascaded Amplifiers 47
2.5.2 Nonlinear Distortions: CSO, CTB, and XMOD 49
2.5.3 Multipath Reflections (Echoes) and Group Delay 50
2.5.4 AM Hum Modulation 55
*2.6 Cable TV Return-Path Transmission Characteristics* *56*
2.6.1 Return-Path Noise Sources 56
2.6.2 Return-Path Noise Filtering 60
References 61

**3 Directly Modulated Cable TV Lightwave Laser Transmitters 63**

*3.1 Semiconductor Laser Diodes* *63*
3.1.1 Basic Laser Physics Concepts 63
3.1.2 Semiconductor Laser Structures 68
3.1.3 Light-Current Characteristic 70
*3.2 DFB and Multiple-Quantum-Well (MQW) Laser Diodes* *72*
3.2.1 DFB Laser Diodes 72
3.2.2 Multiple-Quantum-Well (MQW) Lasers 73
*3.3 Laser Dynamic Characteristics* *74*
3.3.1 Small-Signal Response 75
3.3.2 Large-Signal Response Circuit Model 79
*3.4 Noise in Laser Diodes* *82*
3.4.1 Relative Intensity Noise (RIN) 82
3.4.2 Laser Phase Noise 84
*3.5 DFB Laser Transmitter* *85*
3.5.1 DFB Laser Transmitter System Design 85
3.5.2 Optical Isolators 87
3.5.3 Thermoelectric Cooler (TEC) Design and Operation 88
3.5.4 Linearization Methods 90
3.5.5 DFB Laser Transmitter Performance Requirements 91
*3.6 Return-Path Laser Transmitters* *92*
3.6.1 Mode-Partition Noise and Mode-Hopping Noise 93
3.6.2 Performance Requirements 94
References 96

**4 Externally Modulated Cable TV Lightwave Laser Transmitters 99**

*4.1 $LiNbO_3$ Optical Modulators 100*
4.1.1 Basic Operation of LiNbO3 Intensity Modulators 100
4.1.2 Distortion Characteristics of MZI and BBI Modulators 104
*4.2 Linearization Methods of Optical Modulators 106*
4.2.1 Feedforward Linearization Method 107
4.2.2 Predistortion Linearization Method 108
4.2.3 Linearizer Circuits 110
*4.3 Optical Linearization Methods 112*
4.3.1 Optical Dual Parallel Linearization Method 112
4.3.2 Optical Dual Cascade Linearization Method 114
*4.4 Externally Modulated Laser Transmitter Design 116*
4.4.1 Externally Modulated YAG Laser Transmitter 116
4.4.2 Externally Modulated DFB Laser Transmitter 119
References 120

**5 Lightwave Receivers for Cable TV Networks 123**

*5.1 p-i-n Photodiode 123*
*5.2 Noise Sources in Lightwave Receivers 129*
5.2.1 Shot Noise 129
5.2.2 Thermal Noise 130
5.2.3 Laser RIN Noise 132
*5.3 Carrier-To-Noise Ratio at the Receiver 133*
*5.4 Nonlinear Behavior of p-i-n Photodetectors 135*
*5.5 Basic Cable TV Receiver Design Configurations 138*
5.5.1 Low- and High-Impedance Front-End Receiver Design 140
5.5.2 Transimpedance Front-End Receiver Design 141
5.5.3 High Performance Receiver Design for Cable TV 143
References 145

**6 Optical Fiber Amplifiers for Cable TV Networks 147**

*6.1 Optical Fiber Amplifier Components 147*
6.1.1 Wavelength-Division Multiplexers (WDMs) 147
6.1.2 Erbium-Doped Fibers (EDFs) 154
6.1.3 Pump Lasers 161
*6.2 Basic EDFA System Configurations 163*
*6.3 Amplifier Noise and CNR Calculation 165*
6.3.1 Optical Fiber Amplifier Noise 166

6.3.2 CNR and Noise-Figure Calculation 167
6.3.3 Noise-Figure Measurement 169
*6.4 EDFA Requirements for Cable TV Networks* *170*
6.4.1 EDFA's Noise-Figure Requirement 170
6.4.2 EDFA's CNR Requirement 175
6.4.3 Gain-Flattened EDFAs 176
References 178

**7 RF Digital QAM Modems** **183**

*7.1 RF QAM Modem Building Blocks* *183*
*7.2 MPEG Transport Framing* *185*
*7.3 Reed-Solomon Codes* *188*
*7.4 Interleaver/Deinterleaver* *191*
*7.5 Trellis-Coded Modulation (TCM)* *193*
7.5.1 Punctured Convolutional Coding 193
7.5.2 Viterbi Decoding 195
7.5.3 TCM for 64/256-QAM Modulation 198
7.5.4 Differential Precoder 200
*7.6 Randomizer/Derandomizer* *201*
*7.7 M-ary QAM Modulator Design and Operation* *202*
7.7.1 Baseband Shaping Filter 202
7.7.2 Building Blocks of M-ary Modulator 204
*7.8 M-ary QAM Receiver Design and Operation* *206*
*7.9 Adaptive Equalizer* *208*
*7.10 Carrier and Timing Recovery* *214*
7.10.1 Carrier Recovery in QAM Receivers 214
7.10.2 Timing Recovery in QAM Receivers 217
*7.11 MER and EVM* *219*
7.11.1 MER and EVM Definition 219
7.11.2 MER and EVM Test Procedure 220
*7.12 BER of M-ary QAM Signals in AWGN Channel* *223*
References 228

**8 Subscriber Home Terminals** **231**

*8.1 Digital Set-Top Box Building Blocks* *231*
*8.2 Cable TV RF Tuner* *233*
*8.3 Out-of-Band (OOB) Receiver* *235*
8.3.1 OOB Randomizer 237
8.3.2 OOB Reed-Solomon Coding 238

8.3.3 OOB Interleaver 238
8.3.4 OOB QPSK Mapping 239
*8.4 RF QAM Transceiver* *239*
8.4.1 RF Upstream FEC 242
8.4.1.1 Upstream R-S Coding 242
8.4.1.2 Upstream Randomizer 242
*8.5 MPEG Video/Audio Demultiplexer and Decoder* *243*
8.5.1 VBI Retriever and Decoder 245
*8.6 Conditional Access and Control* *246*
8.6.1 Digital Encryption/Decryption Basics 246
8.6.2 Access Control Basics 248
8.6.3 Renewable Security 249
*8.7 Graphics Processor* *250*
8.7.1 Basic 3D Graphics Concepts and Techniques 251
8.7.2 On-Screen Video and 2D/3D Graphics Rendering Requirements 255
*8.8 Set-Top Box CPU and Memory* *257*
8.8.1 Set-Top Box CPU 257
8.8.2 Set-Top Box Memory 258
*8.9 Advanced Set-Top Box with Built-in DOCSIS Cable Modem* *260*
*8.10 M-QAM Transmission Impairments in HFC Networks* *261*
8.10.1 Set-Top Box Front-End Losses 262
8.10.2 QAM Transmitter Losses 266
8.10.3 Additive White Gaussian Noise (AWGN) 266
8.10.4 Multipath Reflections 267
8.10.5 Amplitude and Group Delay Variations 270
8.10.6 Burst and Impulse Noise 271
8.10.7 AM Hum Modulation 273
8.10.8 64/256-QAM System Budget Link 273
References 276

**9 Transmission Impairments in Multichannel AM/QAM Lightwave Systems** **279**

*9.1 Clipping-Induced Nonlinear Distortions* *280*
9.1.1 Asymptotic Statistical Properties of Clipping Noise 281
9.1.2 BER of M-ary QAM Channels Due to Clipping Noise 283
9.1.3 Comparison to Experimental Results 287
9.1.4 BER of QAM Channel Due to "Dynamic Clipping" Noise 290
9.1.5 Clipping Noise Reduction Methods 292
*9.2 Bursty Nonlinear Distortions* *294*
*9.3 Multiple Optical Reflections* *299*
9.3.1 Interferometric Noise due to DRB 300

9.3.2 Laser RIN Due to Multiple Discrete Optical Reflections 302
9.3.3 Multiple Reflections Effect on AM-VSB Channels 303
9.3.4 Multiple Reflections Effect on QAM Channels 307
*9.4 Dispersion-Induced Nonlinear Distortions* *309*
*9.5 Optical Fiber Nonlinear Effects* *312*
9.5.1 Stimulated Brillouin Scattering Effect 312
9.5.2 Self and External Phase Modulation Effects 316
9.5.3 Stimulated Raman Scattering Effect 319
9.5.4 Cross-Phase Modulation Effect 322
*9.6 Polarization-Dependent Distortion Effects* *323*
References 326

**10 EDFA-Based WDM Multichannel AM/QAM Video Lightwave Access Networks 331**

*10.1 Architecture and Performance of Multichannel AM-VSB/QAM Video Lightwave Trunking Networks* *331*
10.1.1 Multichannel AM/QAM Video Lightwave Trunking Systems 333
10.1.2 Multichannel AM-VSB Video Lightwave Trunking Results 334
10.1.3 Multichannel AM-VSB/QAM Video Lightwave Trunking Results 336
10.1.4 Differential Detection in AM-VSB Video Trunking Systems 340
10.1.5 SPM and EPM Effects in AM Video Trunking Systems 342
10.1.6 Polarization Effects in AM-VSB Video Trunking Systems 344
10.1.7 Gain Tilt Distortion in AM Video Trunking Systems 346
*10.2 The Problem with the Current HFC Networks* *348*
*10.3 DWDM Downstream Access Network Architecture* *349*
*10.4 DWDM Upstream Access Network Architecture* *352*
10.4.1 Frequency-Stacking Scheme 352
10.4.2 Digitized Return-Path Transport 354
References 357

**11 Data-Over-Cable Interface Specifications (DOCSIS) Protocol 359**

*11.1 DOCSIS Communication Protocol* *359*
*11.2 Downstream PHY Layer* *361*
11.2.1 Downstream PMD Sublayer 362
*11.3 Upstream PHY Layer* *363*
11.3.1 Upstream Channel Parameters and Requirements 365
11.3.2 Burst Profiles 365
11.3.3 Burst Timing 368
11.3.4 Upstream Spurious Power Outputs 370

11.3.5 Burst Frame Structure 371
*11.4 Downstream Transmission Convergence Sublayer* *372*
*11.5 Media Access Control (MAC) Layer* *373*
11.5.1 MAC Frame Format 374
11.5.2 MAC Management Messages 375
*11.6 Random Access and Contention Resolution Methods* *380*
11.6.1 Random Access Methods 380
11.6.2 p-Persistence with Binary Exponential Backoff Algorithm 382
*11.7 MAC Layer Protocol Operation* *383*
*11.8 Quality of Service (QoS) and Fragmentation* *384*
11.8.1 Basic Concepts and Operation 385
11.8.2 Upstream Service Flow Scheduling 386
11.8.3 Upstream Fragmentation 388
*11.9 CM and CMTS Interaction* *390*
References 392

**12 Digital Set-Top Box Software Architecture and Applications 395**

*12.1 Digital Set-Top Box Software Architecture* *395*
12.1.1 Real-Time Operating System (RTOS) 397
12.1.2 Set-Top Box Middleware 398
*12.2 Native Applications* *402*
12.2.1 Electronic Program Guides (EPGs) 402
12.2.2 Parental Control 404
*12.3 TV-Based Interactive Applications* *405*
12.3.1 Pay-Per-View (PPV) and Impulse PPV (IPPV) 406
12.3.2 Enhanced Broadcast, Interactive, and Targeted Advertisements 407
12.3.3 Video-on-Demand (VOD) and Near VOD (NVOD) 407
*12.4 Internet-Based Applications* *409*
12.4.1 Set-Top Box Web Browsers 409
12.4.2 E-mail 411
12.4.3 E-Commerce 412
12.4.4 Home Banking, Education, and Gaming 413
*12.5 Integrated Set-Top Box Applications* *414*
References 416

**Appendix A Comparison of DAVIC and DOCSIS Specifications 419**

**Appendix B International Cable TV Frequency Plans 423**

*B.1 CCIR System B/G Frequency Plan* 423
*B.2 CCIR System I Frequency Plan* 425
*B.3 CCIR System D Frequency Plan* 427

**Appendix C Satellite Transponder Parameters for Cable TV Networks** **429**

**Acronyms** **431**

**Index** **437**

**Biography** **447**

# PREFACE

Broadband cable TV access networks have been going through a dramatic transformation worldwide since the late 1980s. The technology vision, which was articulated by several visionary industry leaders such as Bill Gates of Microsoft, of "information at your finger tips" has started to become a reality. With the invention and emergence of the Internet and Intranet as the information superhighway, cable subscribers and small businesses did not want to be left behind and were eager for high-speed access.

Traditionally, cable TV networks were broadband coaxial networks that offered one-way broadcast of analog video channels. The fundamental understanding of the physics of various opto-electronics devices and components led to the invention and development of key fiber-optics transmission technologies such as high-power directly and externally modulated DFB laser transmitters operating at 1310 nm and 1550 nm, optical fiber amplifiers, and optical receivers. These fiber-optics technologies transformed cable TV network architecture to 750-MHz and greater bandwidth hybrid fiber/coax (HFC) networks. Furthermore, the development of low-cost highly integrated communication chips, modules, and systems such as quadrature amplitude modulation (QAM) modulators, transceivers, and MPEG encoders and decoders, enabled cable operators to introduce many digital video programs to the home using a digital set-top box. Instead of low-speed Internet access using a dial-up modem, high-speed Internet access through cable TV networks is possible using a cable modem either as a stand-alone unit or built into a digital set-top box.

The continuing insatiable desire at sometimes accelerating pace for more "bandwidth" and more services on demand has forced the cable TV operators to rethink their two-way HFC network architecture, paving the road toward two-way dense wavelength division multiplexed (DWDM) cable TV networks.

From 1992 to 1996 as a research scientist at Bell Communications Research (Bellcore), I had the privilege to work with many world-class scientists and participate in the emerging fiber-optics and QAM receiver technology studies. With a built-in HFC test-bed, my colleagues and I were able to provide valuable technical analysis and auditing services to the Regional Bell Operating Companies such as Pacific Bell and Ameritech, as well as cable TV equipment manufacturers. My exposure to the hardware and software development of digital set-top boxes and cable modems came during my work as a principal scientist at Digital

Network Systems, General Instrument from 1996 to 2000. In particular, I was exposed to the various discussions and debates on the different cable TV network and set-top box requirements while participating at the various cable TV standards meetings such as DOCSIS, IEEE802.14, and OpenCable.

The purpose of this book is to provide the reader with the basic understanding of today's two-way HFC cable TV network technologies and their evolution toward DWDM network architectures. This book, which can be used as the basis for a graduate-level material, is intended for engineers, scientists, cable TV professionals, and students interested in learning more about the existing and emerging cable TV technologies and applications.

The book is organized into five main sections:

- The first section, which consists of Chapters 1 and 2, provides an overview of the two-way HFC network with the competing access technologies, including digital subscriber line (DSL), fiber-in-the-loop (FITL), direct broadcast satellite (DBS), and multichannel multipoint distribution service (MMDS). Chapter 2 provides basic cable TV systems background, including different analog and digital video modulation formats, cable TV frequency plans, coaxial cable components and systems, and return-path transmission characteristics.
- The second section of the book, which consists of Chapters 3, 4, 5, and 6, introduces the key fiber-optics transmission technologies, including directly and externally modulated laser transmitters such as distributed feedback (DFB) and YAG laser transmitters operating at both 1310 nm and 1550 nm, optical receivers, and Erbium-doped optical fiber amplifiers (EDFAs).
- The third section of the book, which consists of Chapters 7 and 8, discusses the hardware architecture with the key components, features, and functionality of end-user home terminals, including cable modems and digital set-top boxes. Chapter 8 also includes a budget link analysis for digital set-top box in the presence of various HFC network impairments.
- The fourth section, which consists of Chapters 9 and 10, provides an in-depth analysis of single and multiple-wavelength fiber-optics transmission impairments over HFC and DWDM networks carrying both AM-VSB and QAM channels. The emerging two-way DWDM network architecture is also discussed in Chapter 10.
- The last section, which consists of Chapters 11 and 12, discusses the transmission protocol, the software architecture, and the various applications for the end-user home terminals. This includes the DOCSIS protocol for cable modems, native applications such as electronic program guides (EPGs), TV-based and Internet-based applications such as video-on-demand (VOD), e-mail, and e-commerce for digital set-top boxes.

I would like also to acknowledge many people who have contributed to my professional career. These include Professors Hyatt Gibbs and Murray Sargent III from the Optical Sciences Center, University of Arizona, who supervised my Ph.D. dissertation on optical bista-

ble devices and non-linear optics, and Professor Chi Lee, University of Maryland. I also indebted to Dr. Chinlon Lin, William T. Anderson, and Burt Unger from Bellcore, who not only introduced me to the various technical issues of broadband fiber-optics transmission systems and networks, but also supported my work in the HFC test-bed lab at Bellcore. At General Instrument/Motorola corporation, I had the pleasure to work with various individuals including Mark Kolber, Victor Hou, Reem Safadi, Qin Zhang, Robert Mack, Steve Kimble, and Matt Waight, who helped me further understand the various technical aspects of the operation of cable TV networks as well as the OpenCable standard. In particular, I would like to acknowledge the efforts of Mark Kolber, Victor Hou, Petr Peterka, and Qin Zhang, along with Keith Williams, Philip Antoney, and Ron Hranac who graciously agreed to review portions of this book. Also, I would like to express my sincere thanks to the management in the Cable Network Operation business unit of Network Communication Group at Intel for their encouragement and support of my efforts to publish this book.

Finally, I am grateful to my parents, Moshe and Mazal, for their love and encouragement over the years. I am also exceedingly thankful to my wife Nina and our children, Yafit, Shlomit, Binyamin, and Leora, for their love and patience while I was working many nights and weekends over the last two years to write this book.

Shlomo Ovadia
*San Jose, California*

## BIOGRAPHY

Shlomo Ovadia has earned his B.Sc. in Physics from Tel-Aviv University in 1978, and his M.Sc. and Ph.D. in Optical Sciences from the Optical Sciences Center, University of Arizona in 1982 and 1984, respectively. He spent two years as a postdoctoral fellow at the Electrical Engineering Department, University of Maryland, investigating different III-V optoelectronics materials and devices. In 1987, Shlomo joined IBM at East Fishkill as an optical scientist developing various IBM optical communications and storage products. He joined Bellcore in 1992, where he developed an HFC test-bed, and studied the transmission performance of multichannel AM/QAM video transmission systems. As a project manager and a senior scientist, Shlomo provided technical analysis and consulting services to the Regional Bell Operating Companies as well as to various cable TV equipment vendors. In 1996, Shlomo joined General Instrument as a principal scientist in Digital Network Systems division, where he was developing the next-generation digital set-top boxes for both domestic and international markets. In April 2000, Shlomo joined Intel's Cable Network Operation business unit in San Jose, California, as a principal system architect developing communication products such as cable modems. He is a Senior Member of IEEE/LEOS/Comsoc with more than 60 technical publications and conference presentations. Shlomo served on the technical committees of many IEEE/LEOS conferences, and he is a regular reviewer for various IEEE publications such as Photonics Technology Letters and Journal of Lightwave Technology. Shlomo has eight pending patents, and his personal biography is included in the Millennium edition of Who's Who in Science and Engineering (2000/2001).

# CHAPTER 1

# BROADBAND HYBRID FIBER/COAX ACCESS NETWORKS OVERVIEW

Broadband access networks have been dramatically evolving for more than a decade to offer consumers and businesses a variety of services such as analog and digital video, high-speed data, and telephony services at their "finger tips." Various competing enabling access technologies have also been developed since the late 1980s to address the continuously growing consumer appetite for these services. The goal of this book is threefold. The first goal is to explain to the reader the architectural evolution of cable TV access networks to fulfill the consumer requirements for a wide range of entertainment and interactive information services. The cable TV access architecture has been evolving from a one-way broadband coaxial network to a two-way hybrid fiber/coax network to a dense wavelength-division-multiplexed (DWDM) network. The second goal is to introduce the reader to the technology behind all the key building blocks of cable TV access networks, and discuss their characteristics and performance requirements. This includes fiber-optics transmission technologies such as directly and externally modulated laser transmitters, optical receivers, and optical fiber amplifiers as well as digital communication technologies such as quadrature amplitude modulation modems and digital set-top boxes. Using these technologies, the third goal of this book is to explain to the reader the high-speed data transmission protocol for cable modems as well as the software architecture and the wide range of interactive applications for digital set-top boxes. These applications include native applications such as electronic program guides, TV-based interactive applications such as video-on-demand, and Internet-based applications such as e-mail and e-commerce.

Chapter 1 provides a brief overview of cable TV access networks based on hybrid fiber/coax architecture, and introduces the reader to the key end-user home terminals such as digital set-top boxes and cable modems. To complete this discussion, it is necessary to briefly explain the emerging competing broadband access technologies with two-way cable TV networks such as digital subscriber line, fiber-in-the-loop, direct broadcast satellite, and fixed wireless services.

## 1.1 Introduction

During the past two decades, telecommunications networks have been evolving in separate ways. In general, telecommunications networks can be divided into three different groups as follows: (A) community antenna television (CATV) network[1], (B) local and wide-area computer network, and (C) public-switched telephone network (PSTN). These telecommunication networks were truly independent networks since they delivered unique services that otherwise where not available in the other type of network. Thus, traditional cable TV networks did not offer subscribers telephony or high-speed data services, and PSTN networks did not offer broadcast analog or digital video services. In the middle of the 1990s, two strong independent forces have played a significant role in changing the telecommunication landscape. First, easy and low-cost access to the Internet opened the information superhighway to many subscribers and businesses to electronic commerce, on-line shopping, advertisement, and information services that are easy to use, quick, and tax-free. The second major force was the passing of the U.S Telecommunications Act in 1996. The main idea was to deregulate the telecommunications industry by allowing telephone companies (local and long-distance), wireless operators, cable TV operators, and broadcasters to enter

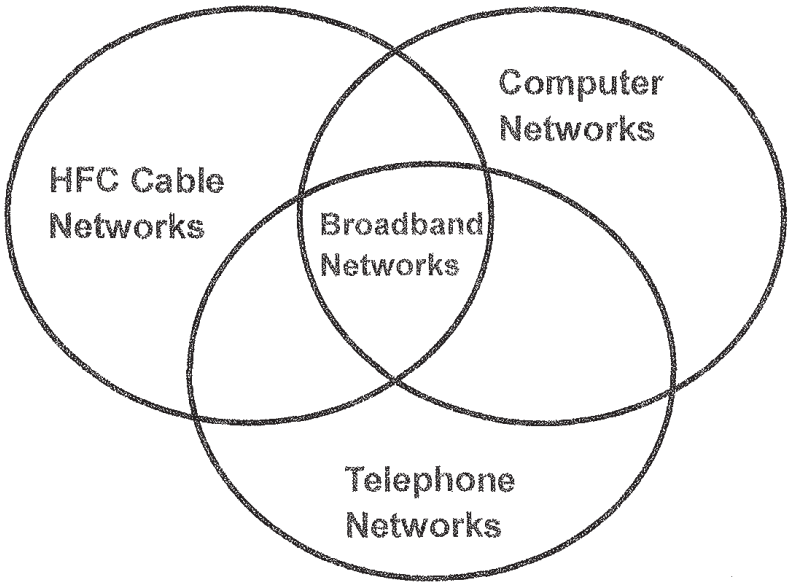



Figure 1.1 The convergence of HFC cable TV networks, computer networks and public-switched telephone networks into the emerging broadband networks carrying video, data, and voice services.

[1] The term "CATV" is generally accepted today to mean simply cable television or cable TV.

each other's markets. This act has opened the door to a merger mania among the different companies. Furthermore, utilities companies such as gas companies and even railway companies wanted to lay fiber optics cables along their rights-of-way and build their own telecommunications networks. Figure 1.1 shows the convergence of the three telecommunications networks into a broadband network to deliver a wide-range of entertainment and in- formation services [1]. However, as we will see later in this chapter (Section 1.4), there are many economical, business, and regulatory factors in addition to the selected technologies, which determine the feasibility of building such a telecommunications network.

Cable TV networks have been evolving from one-way broadcast of analog video channels to two-way interactive hybrid fiber/coax (HFC) networks delivering analog/digital video channels and high-speed data. The traditional tree-and-branch broadband coaxial network, which was supported by RF technology, served well the cable TV operators for the type of services it provided, namely, broadcast or point-to-multipoint services. The cascades of RF amplifiers became very long (30–40 amplifiers), which in turn reduced the quality and availability of the transmitted amplitude modulation vestigal sideband (AM-VSB) video channels and increased customer dissatisfaction. The use of terrestrial microwave links to reduce long cascades of RF amplifiers improved the transmission performance of the broadcast analog video channels. Significant improvements in fiber-optics transmission technology since the late 1980s fueled the explosive growth of the cable TV industry [2–5]. The introduction of 550 MHz directly modulated (DM) DFB laser transmitters and optical receivers operating in the 1310 nm wavelength band changed the traditional architecture of coaxial networks. HFC networks enabled robust transmission of broadcast analog video channels via standard single-mode fiber (SMF) to the optical fiber-node, resulting in a significant reduction in the number of cascaded RF amplifiers. In addition, cable TV operators were able to consolidate the deployment of headend equipment using fiber-optic rings to connect between the master headend and secondary headends or hubs in major metropolitan markets. Thus, cable TV operators were able to cut cost and further improve the quality and availability of the original broadcast services.

The development of various key systems such as quadrature amplitude modulation (QAM) modulators, low-cost QAM receivers, and moving picture-expert group (MPEG)[2] digital video encoders and decoders enabled cable operators to offer tens of new digital video services along with the traditional AM-VSB video channels using a digital set-top box (STB) at the subscriber's home [6]. The rapid deployment of 750 MHz HFC networks also allowed cable TV operators and some telecom services to provide competitive access to consumers and various businesses in major markets.

In the middle of the 1990s, the existing HFC network architecture had started to evolve in a new direction. This evolution can be attributed to the following forces in the marketplace:

(A) An explosive demand for high-speed data (e.g., Internet) access in residential areas

[2] Detailed overview of the various MPEG standards can be found in Chapter 2, Section 2.2.

(B) The delivery of highly targeted interactive digital services
(C) Increased competition from various telecommunication and direct-broadcast-satellite (DBS) service operators
(D) Improved cable TV network scalability with reduced life-cycle cost of various fiber and coaxial components and systems
(E) Improvements in fiber-optics technologies, particularly laser transmitters and receivers, and cable TV network management

These requirements and market forces have been forcing cable TV operators to rethink their existing HFC network architecture, and are paving the road toward DWDM cable TV access networks [7]. This book is organized in such a way as to enable the reader to understand step-by-step the technology, operation, and applications of the different key systems in HFC and DWDM networks. The architecture and services of two-way HFC networks are reviewed and contrasted with competing access technologies in the remainder of this chapter. Chapter 2 explains the different signal formats for analog and digital video and audio, and the basic characteristics of coaxial cable TV networks. The operation and design principles of directly modulated DFB laser transmitters (Chapter 3) and externally modulated laser transmitters (Chapter 4), which are used to transmit the sub-carrier multiplexed (SCM) analog and digital video signals from the cable TV headend up to the fiber node through a SMF, are discussed in detail. Chapter 5 discusses the theory and operation of optical receivers, which are located at the fiber node, and are used to convert the optical signals to electrical signals. Chapter 6 explains the operation principles and use of optical fiber amplifiers in cable TV networks. The design and use of RF digital modulators and demodulators such as M-ary QAM and 8/16-VSB, which are among the key systems in the HFC network, are explained in Chapter 7. Chapter 8 discusses the design and operation of the primary subscriber's home terminals, namely, a STB and a cable modem (CM), which are the enabling technologies to receive and transmit information over HFC networks. To understand the requirements for robust transmission on HFC and DWDM networks, the various types of transmission impairments are analyzed both theoretically and experimentally in Chapter 9. In Chapter 10, the reader is now ready for the discussion of the downstream and upstream architecture of the next generation of broadband cable TV access networks, namely, DWDM networks. To enable the operation of CMs, standard communication protocols are needed. First, CableLabs' multimedia cable network system (MCNS) data-over-cable service interface specification (DOCSIS) protocol for CMs is discussed with emphasis on the physical (PHY) layer and medium-access-control (MAC) layer in Chapter 11. Chapter 12 discusses the software architecture of a digital STB and the different types of applications, including native applications, TV-based interactive applications, and Internet-based applications.

## 1.2 Traditional Cable TV Networks

As the name implies, HFC networks are a hybrid combination of fiber optics and coaxial cables used for signal transmission and distribution. Fig. 1.2 shows a typical major metropolitan HFC network architecture. Analog and digital video signals from various sources such as satellite transponders, terrestrial broadcast, and video servers are multiplexed and transmitted via an SMF ring from the cable TV master headend to typically four or five primary hubs or headends. Each primary headend or hub has as many as 100,000 homes passed. Up to four or five secondary hubs and local headends, where each primary headend has up to 25,000 homes passed, are used for additional distribution of the multiplexed analog and digital video signals. To reduce the necessity to duplicate the same broadcast video channels at the various primary or secondary headends, the analog or digital channels from the master headend can be shared over the backbone network. The backbone network is typically built with a ring architecture using either synchronous optical network (SONET) with the associated SONET add-drop multiplexers (ADMs), or some proprietary technology. The SONET specification, which was proposed by Bellcore [8] and standardized by the American National Standards Institute (ANSI), defines a hierarchy of standard digital data

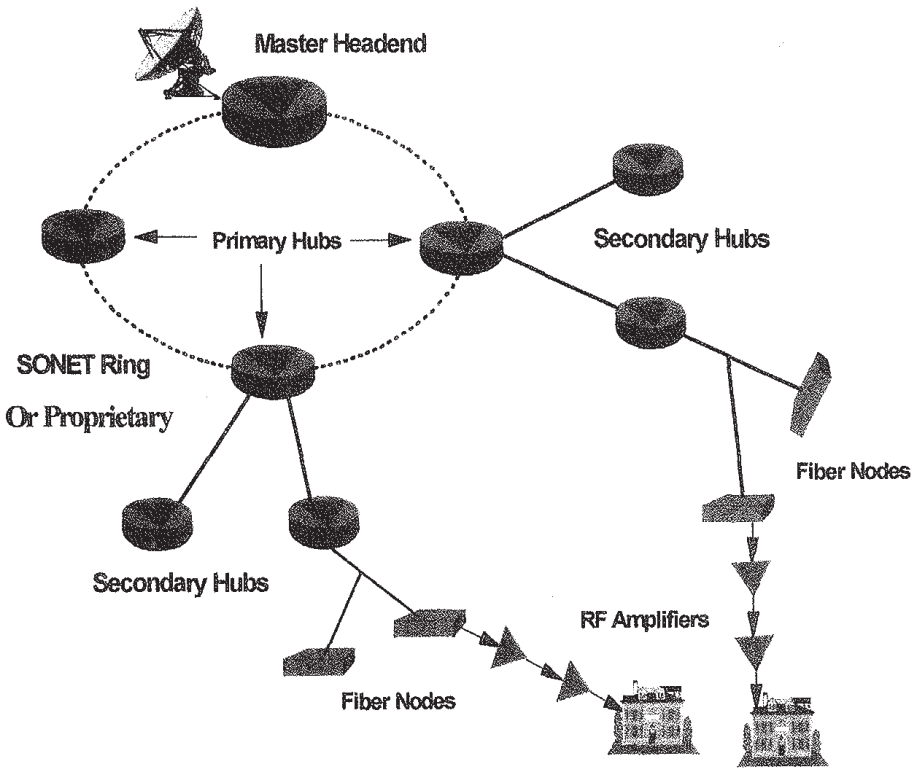



Figure 1.2 Typical HFC network architecture for a large metropolitan area.

rates from 51.84-Mb/s or OC-1 (optical carrier level 1) to integer multiples of this rate. In a typical SONET network, the analog video signals are digitized, modulated, time-division-multiplexed (TDM), and transmitted as digital baseband signals at various rates ranging from OC-12 (622-Mb/s) to OC-48 (2488-Mb/s). Statistical TDM, which can be used as an alternative to synchronous TDM, increases overall bandwidth usage by dynamically allocating time slots on demand to the different variable bit-rate services transported over SONET [9]. To further reduce deployment cost, most cable TV operators have selected to use SONET-compatible equipment, relying on proprietary network interfaces. The analog and digital video signals are further distributed to the various fiber nodes using a tree-and-branch architecture as shown in Figure 1.2. At the fiber node the optical signals are converted to electrical signals and transmitted to the subscriber through coaxial cables using different types of RF amplifiers and taps. The fiber node size, which is defined by the number of homes passed (HP), can vary from a "small node size" of about 100 HP to a "very large node size" of about 2000 HP. As we will see in both Chapter 2 and Chapter 10, the size of the fiber node has significant implications on the performance of the HFC network, particularly in the return-path or upstream spectrum. In general, a large fiber node size reduces upstream channel availability and peak user data throughput because of the time varying ingress noise as well as other impairments from all the homes passed.

## 1.3 Two Way HFC Access Networks

The evolution of one-way HFC cable networks into two-way broadband HFC access networks has been enabled by the introduction of three new systems as follows: (A) advanced subscriber's home terminal or STB, (B) CM, and (C) IP telephony systems operating over HFC access networks. Each of these technologies enabled the introduction of new types of services that were not possible before. In the following section, the basic functionality and the different type services are reviewed. The hardware system architectures, including the operation principles of all the primary building blocks, as well as the performance requirements of both a digital STB and a CM, are discussed in detail in Chapter 8.

### 1.3.1 Subscriber's Home Terminal

Subscriber home terminals, which include analog and digital STBs, are probably the most important component in the HFC access network. Digital STBs represent the connection between the present analog TVs and the advanced digital TVs of the future. With the introduction of analog STBs since the 1970s and 1980s, subscribers were able to receive broadcast analog TV channels via the HFC network instead of the terrestrially broadcast channels. Cable operators offered subscribers impulse-pay-per-view (IPPV) services of movies and special events using telephone return for interactive operation. The IPPV subscribers were able to place an order for a particular program or service directly through a cable home terminal at almost any time. An extended credit from the cable operator, which was dispensed

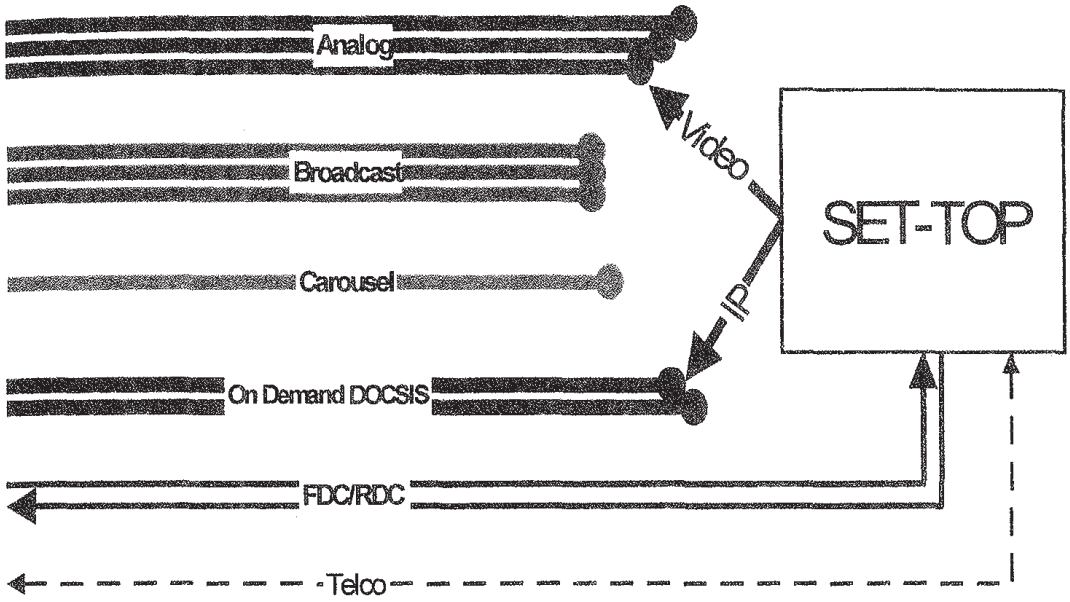



Figure 1.3 Illustration of the different types of video and data services that can be accessed through an advanced digital STB, which has a built-in cable modem and dual cable TV tuners.

by the home terminal, allowed the subscriber impulse purchases. With IPPV, the billing system calculates the credit limit of a given subscriber based upon the cable operator guidelines and payment history of the subscriber. The credit limit is transmitted from the cable headend to the subscriber's home terminal, where IPPV purchases by the subscriber are enabled up to the available credit limit. With each purchase, the available credit is reduced until the cable operator renews the credit balance.

The introduction of digital STBs in 1996 by both General Instrument and Scientific Atlanta opened the door for cable operators to offer subscribers new array of broadband services [10]. The field-installed digital STB provided the following basic functions:

- Channel tuning of digital (MPEG-2) and analog video services in the downstream bandwidth
- Demodulation of the received digital channel
- Modulation of return-path digital signals
- Encryption and decryption of the selected digital services
- Subscriber management signaling from the cable TV headend
- Subscriber user interface

The recently deployed advanced digital STBs from Motorola/General Instrument, which have dual cable TV tuners and a built-in CM, enable subscribers to "watch" TV and "surf"

the web simultaneously over the HFC network. The various types of video and data services that can be accessed with the advanced digital STB are illustrated in Figure 1.3. Services such as clear and encrypted analog/digital video services, digital pay-per-view (PPV), digital IPPV, near-video-on-demand (NVOD), and video-on-demand (VOD) can be received using one of the STB tuners. The VOD service allows subscribers to order a movie from a centralized library using a downloadable electronic program guide (EPG) and navigation system. This system, which is delivered through an out-of-band (OOB) channel[3] and provides menu or icon-driven access to the resources of the network and the STB, is protected for secure access of resource areas using a personal identification number (PIN). The VOD service provides the subscriber with VCR-like functionality just as if it was a rented movie played at his home VCR. In contrast, the NVOD service allows the subscriber to order a movie from a limited selection of regularly scheduled playing movies at fixed time intervals, say every 30 minutes, without the VCR functionality. The downstream OOB channel, which carries all the specific information for the various in-band programs, is operating at 2.048-Mb/s using quadrature-phase-shift-keying (QPSK) modulation and is typically located in the 70-MHz to 130-MHz frequency range [11].

**Table 1.1 Average and peak rates to the digital STB for various types of traffic. CBR is constant bit rate traffic.**

| Traffic type | Average rate | Peak rate |
|---|---|---|
| Entitlement control messages | 200-kb/s | CBR |
| MPEG-2 compressed video and audio | 3-Mb/s per active subscriber | 9-Mb/s per active subscriber |
| Applications (e.g., Web browsing) | 100-kb/s per active subscriber | 5-Mb/s per active subscriber |
| Broadcast data carousel | 0.1- 5.0-Mb/s | CBR |
| Analog broadcast | N/A | N/A |

The second tuner in the advanced STB provides two-way high-speed Internet access using DOCSIS protocol. Additional new services such as IP telephony and IP videophone, and interactive games will be supported in the near future [12]. Thus, the advanced STB acts as a residential gateway (RGW), delivering Internet protocol (IP) packets to external devices such as IP voice or videophone. Table 1.1 shows the average and peak bit-rate to the digital STB for the various types of traffic.

The advanced STB also provides pass-through, processing, and decoding of high-definition TV (HDTV) signals. All the previously mentioned applications may utilize the 2D/3D graphics presentation capabilities of the STB when such applications are delivered to

[3] The downstream OOB channel is sometimes called forward data channel (FDC), while the upstream OOB channel is sometimes called reverse data channel (RDC).

a TV receiver/monitor either as baseband video or as RF modulated video. The advanced STB has a number of input/output interfaces including universal serial bus (USB), IEEE 1394 interface ("FireWire"), and Ethernet connection. The IEEE 1394 is a high-performance serial bus standard, which was developed for real-time transfer of multimedia applications at 100-Mb/s, 200-Mb/s, and 400-Mb/s rates over cable [13].[4] The IEEE 1394 standard is a scalable, flexible, easy to use, low-cost digital interface that provides the necessary integration between different consumer electronics devices such as digital camcorder, digital STB, digital TV, videophone, and personal computers (PCs).

### 1.3.2 Cable Modems

Cable modems (CMs) are the most exciting enabling technology that provides two-way high-speed access to the Internet over HFC networks. Since the development of the first World Wide Web software by Tim Berners-Lee and the 1993 release of the Mosaic web-browser software, the number of web sites has grown exponentially around the world, doubling about every three months. The realization of the enormous potential of these on-line global information resources available to every user at the click of his computer mouse has created the so-called global information superhighway. To access the Internet, telephone modems operating at a typical rate of 28.8 to 56-kb/s have been used. The existing telephone networks, which were not designed to handle the massive amount of data traffic, became congested, and thus accessing the multimedia-rich web pages took a relatively long time. In contrast, CM technology operating over two-way HFC networks can provide downstream data rates more than 30-Mb/s, which is more than 1000 times faster than a typical telephone modem, and upstream data rates more than 10-Mb/s. These data rates are also more than 200 times faster than the basic data rates (128-kb/s) available from the LECs using the so-called integrated services digital network (ISDN). The CM has an enormous economic advantage over ISDN services since cable TV operators and Internet service providers (ISP) charge only a small fraction of the access fees charged by the LECs. Most CMs typically in residential areas are configured for asymmetric data traffic. This is because most users are using the web to download multimedia-rich applications while using the upstream only for lower bit-rate applications such as e-mail, file transfer, and simple keystroke commands requesting downloads. CMs operating at symmetric data rates are typically needed in a corporate campus Intranet, where each CM is shared among various networked computers.

Given the enormous importance of CMs in two-way HFC networks, different aspects of their system design, operation, and protocol are discussed throughout this book. Chapter 7 reviews and explains the RF system architecture, design, and operation of the CM in HFC networks. The standard CM transmission protocols, which are known as DOCSIS 1.0 and 1.1, are discussed in detail in Chapter 11[14]. Finally, a technical proposal for the next-generation advanced physical layer specification for DOCSIS titled "Advanced TDMA

[4] The IEEE 1394 was conceived by Apple Computer and then developed within the IEEE 1394 Working Group.

Proposal for HFC Upstream Transmission" has been recently submitted to CableLabs for approval [15]. The goals of the proposal are (A) to increase upstream capacity with additional robustness against cable plant impairments over the existing DOCSIS modems, and (B) to achieve full backward compatibility with existing DOCSIS modems.

### 1.3.3 IP Telephony

The term IP telephony refers to the use of IP to carry voice over telecommunications networks [16]. It is also frequently called voice-over-IP (VoIP). It should be noted that IP networks carry data independent of the physical layer. These networks can also be categorized into *circuit-switched* networks and *packet-switched* networks. Circuit-switched networks, which are also called connection-oriented networks, operate by forming a dedicated connection (e.g., circuit) between two physical points in the network. In packet-switched networks, which are also called connectionless networks, the data is segmented into small blocks called packets and transmitted by forming virtual connections between any points in the network. Due to the Internet's explosive growth, IP has essentially become the standard packet-switched layer protocol for both local-area networks (LANs) and wide area networks (WANs).

As we learned earlier in Section 1.1, the integration of telephony services into residential broadband HFC networks carrying both video and data has enormous advantages for the information superhighway [17]. The IP telephony can be implemented using either an IP phone or a POTS phone connected to a modified CM or a digital STB. An IP phone is a relatively new device, which instead of connecting to a proprietary private branch exchange[5] (PBX) port, it connects to a standard Ethernet port in a CM, or a digital STB, or a home PC. An IP phone can operate as a standard IP device with its own IP address, and has the built-in ability to provide voice compression. In order to connect a POTS phone to a CM or a digital STB, the analog voice signals needs to be converted to short packets of digital audio. Currently, new interface modules are being developed that plug into a CM or STB and provide this functionality. These IP packets are sent through the HFC network using the DOCSIS protocol, which has built-in a guaranteed quality of service (QoS) for voice telephony [14]. It should be pointed out that DOCSIS 1.1 capable CMs have not been deployed as of this writing since CableLabs has not certified them. Some manufacturers, to enable QoS VoIP, have deployed CMs using a subset of the DOCSIS 1.1 protocol, which is called DOCSIS 1.0+.

For a mass-market deployment of IP telephony service to residential customers, seamless integration between PSTN and IP networks has to occur. Figure 1.4 shows, for example, the proposed integration between PSTN and IP networks [18]. A RGW, which is sometimes called embedded client, terminates an analog POTS connection to a phone or PBX. The RGW is responsible for supporting real-time transport protocol (RTP), which is used for the

[5] PBX is a telephone company's central office switch that routes many incoming and outgoing calls to their proper destinations.

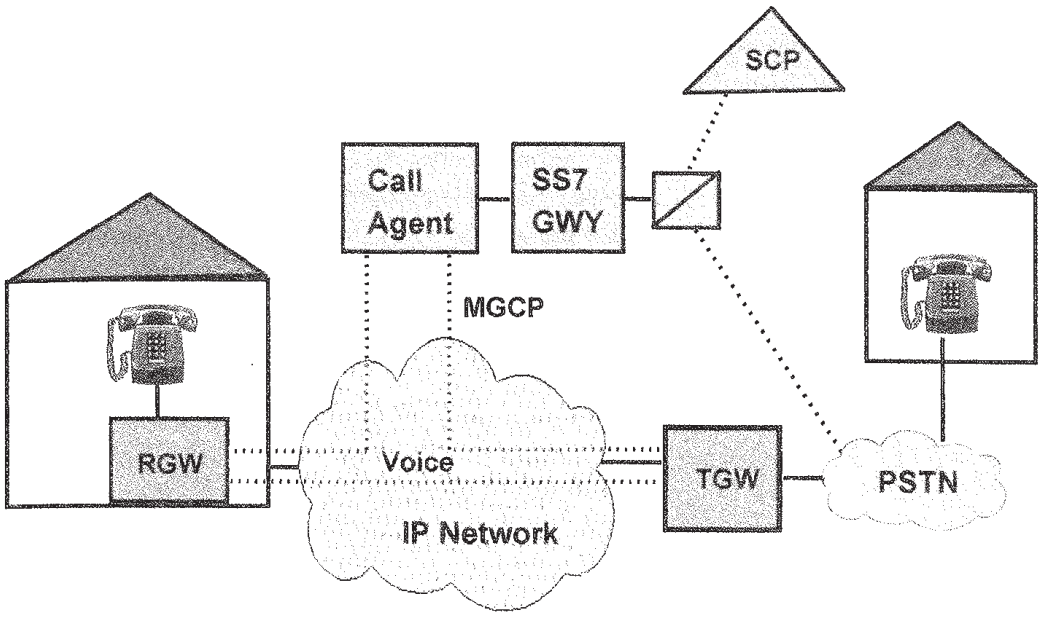



Figure 1.4 Integration of MGCP with PSTN for IP telephony services. After Ref. [18] (© 1999 IEEE).

voice signals end-to-end communications over IP networks. The RGW also supports an interface to IP networks through the HFC access network. A trunking gateway (TGW) is responsible for bridging between the PSTN and the IP network. The TGW converts the TDM voice signals, which were originated from the PSTN, to RTP packets. Similarly, calls that originated from the IP side of the network and terminated at the PSTN are converted to the proper TDM voice format. The TGW also supports the media gateway control protocol (MGCP), which is the pending standard interface between the TGW and call agents [19]. Using MGCP messages, the call agent instructs both the RGWs and TGWs on how to proceed with the call processing. The call agent also handles the so-called signaling-system number 7 (SS7) signaling between TGWs that connect the PSTN with the IP network [20]. SS7 is a reliable open-architecture protocol that divides a telephone call into two separate components: (A) the signaling component, which contains address information for call setup and teardown, and (B) the voice and data component for communication between the originator and terminator of a call. These two components are transmitted on separate facilities. The facility for carrying signaling information is a signaling link or data link, and the facility for carrying voice and data traffic is a voice trunk. Telephone switches use the SS7 ISDN user part (ISUP) protocol for communicating with each other when setting up calls that span more than one switch. The call agent sends ISUP messages over SS7 for setting up calls between the IP network and the PSTN. A PacketCable network-based call signaling (NCS) protocol is currently being drafted by CableLabs [21]. The NCS protocol, which is part of

PacketCable, defines the control of VoIP in RGWs by the call agents. The MGCP assumes that the different call agents can synchronize with each other in order to send the proper commands to each RGW under their control. For further details on DOCSIS protocol, see Chapter 11.

The IP telephony requirements over enterprise networks are quite different from those for residential broadband networks. In 1996, the ITU standard organization introduced ITU-T Recommendation H.323, which defines the components, procedures, and protocols necessary to provide multimedia communications over LANs and IP networks that do not provide a guaranteed QoS [22]. H.323 terminals, which may carry real-time multimedia services over LANs or IP networks, may be integrated into a PC or as a standalone device such as a videophone.

The PSTN carries voice as a 64-kb/s stream using pulse-code-modulation (PCM) for both A-law and μ-law or as a compressed voice signal using ITU-T speech coding standards such as G.727 or G.729. On the other hand, IP networks set up the voice path for each call using call agents, which control the resources of the RGWs and TGWs, using RTP since there are tight constraints of the QoS that must be delivered. The H.323 gateway provides the necessary voice format conversion between the PSTN and the IP networks. In addition, the H.323 gateway terminates the call on the IP network side. However, existing H.323 gateways do not support SS7 signaling. The ITU-T standard defines the procedures and protocol by which network elements in the PSTN exchange information over a digital signaling network to effect call setup, routing, and control.

## 1.4 Competing Access Technologies

There are various competing access technologies that can bring 'broadband' services to the subscriber's home. This section provides a brief overview of the various competing access technologies with their advantages and disadvantages.

### 1.4.1 Asymmetric Digital Subscriber Line (ADSL)

Asymmetric digital subscriber line (ADSL) technology uses existing twisted-pair copper telephone wires to provide the required bandwidth for broadband services such as Internet access, video conferencing, interactive multimedia, and VOD. The ADSL technology is designed to solve the severe bottleneck that currently exists in telephone networks between the central office (CO) and the end customer [23]. ADSL can deliver data rates ranging from 64-kb/s up to 8.192-Mb/s for downstream channels and from 16-kb/s up to 768-kb/s for upstream channels, while simultaneously providing plain old telephone service (POTS) [24].

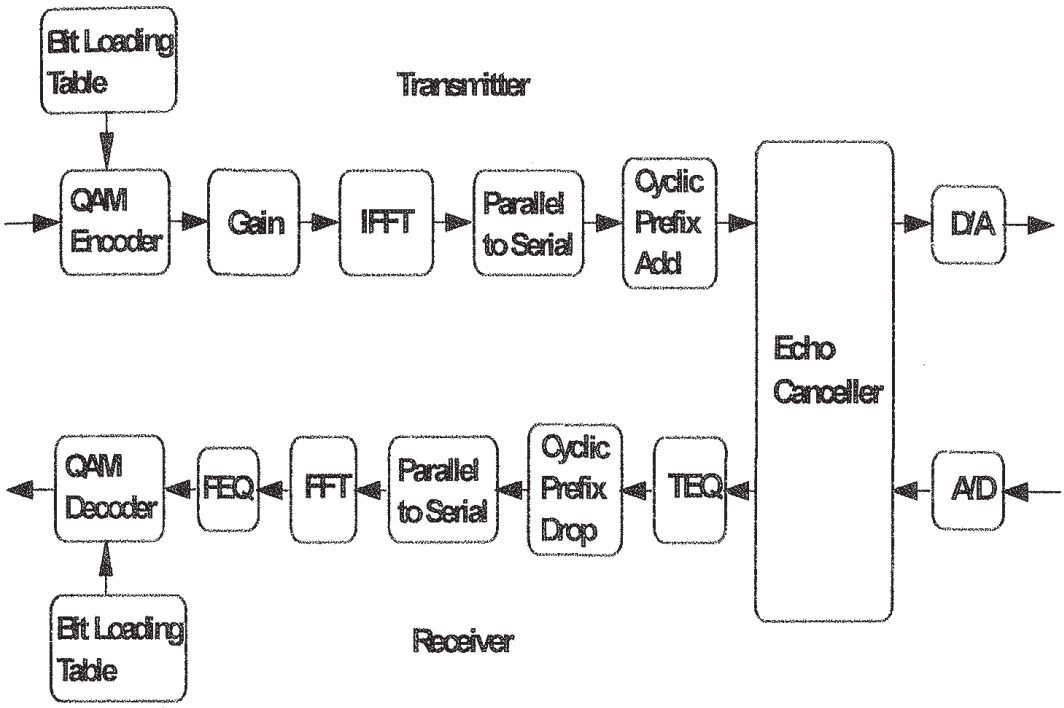



Figure 1.5 Basic block diagram of ADSL modem.

Discrete multi-tone (DMT) modulation technology has been standardized by the ANSI T1 committee for ADSL transmission, due to its unique ability to overcome the severe interference that exists in the residential twisted copper-pair environment. The DMT technique divides the input data stream into N sub-channels, which are typically referred to as "tones," each with equal bandwidth but at a different center frequency. 256 frequency tones are available for downstream channels and 32 tones for upstream channel are used, where each tone has a bandwidth of 4.3125-kHz and the same frequency separation between two successive tones. Each of the tones is QAM modulated on a separated RF carrier. The RF carrier frequencies are multiples of one basic frequency. The available spectrum band ranges from about 20-kHz to 1.104-MHz, while below 20 kHz is reserved for voice service (POTS). Noise and channel conditions are constantly measured for each tone separately, to achieve optimal transmission at any time. Figure 1.4 shows a simplified block diagram of an ADSL modem. The QAM encoding is done according to a bit-loading table, which defines the number of bits carried by each tone. Thus, using higher QAM constellation modes in those tones with high SNR provides higher bandwidth efficiency (b/s/Hz). The bit-loading table is calculated during startup according to the actual measured SNR to allow optimal use of channel capacity. Loading is limited by the ADSL standard to 2–15 bits per

tone. When a given bit rate is required, the bits are allocated to the carriers in such a way that the sum of all the bits on all carriers matches the desired rate, and the probability of error on each carrier is about the same. When the maximum available bit rate is needed, each carrier is allocated with the maximum number of bits that can be transmitted without errors, based on the measured SNR of that carrier. This mode is typically referred to as the rate-adaptive mode. This mode allows higher SNR and data rates for subscribers who are living close to the CO, and lower SNR and data rates for subscribers who are far away from the CO due to additional line attenuation and noise. From a practical point of view, transmission distance for an ADSL service is typically limited to about 17,500 feet from the CO.

Another key element in the ADSL modem design is the use of fast Fourier transform (FFT) and inverse FFT (IFFT). The IFFT allows one to create the sum of N tones, where each tone is independently QAM modulated. At the DMT transmitter, the time-domain samples at the output of the IFFT (2N) are converted to a serial bit stream and fed to cyclic prefix (CP). The CP is designed to separate the transmitted symbols in time to decrease inter-symbol interference (ISI). At the DMT receiver, a time equalizer (TEQ) and a frequency equalizer (FEQ) are used to minimize ISI, cochannel interference, and reduce the signal amplitude and phase distortion.

To support two-way traffic, ADSL modems divide the available bandwidth using a frequency division-multiplexing (FDM) scheme, where different bands are assigned for the downstream and upstream data. Another method is to use echo cancellation (EC). EC is accomplished by generating an exact replica of the transmitted signal that leaks into the receiver. Upon subtraction of the near-end echo-replica the received far-end signal can be processed as if its only impairment has been the channel-induced noise sources. EC in ADSL modem must be designed with different sampling rates for upstream and downstream due to the asymmetric data transmission rates. The ADSL modem design also implements forward-error-correction (FEC) with an interleaver to combat up to 500-μs burst noise.

ADSL modems usually include a POTS splitter, which enables simultaneous access to voice telephony and high-speed data access. The POTS splitter provides a low-pass filter between the copper line node and the telephone node, and provides a high-pass filter between the copper line node and the ADSL modem node. Thus, between the ADSL node and the telephone node the splitter attenuates all signals. An active or passive POTS splitter can be implemented. The advantage of the active POTS splitter is that it enables simultaneous telephone and data access. However, the telephone service fails if there is a power outage or modem failure with an active POTS splitter. In contrast, a passive POTS splitter maintains lifeline telephone access even if the modem fails (because of a power outage, for example), and is typically used for a regular analog voice channel.

One of the key functions of the POTS splitter is to block impulsive noise coming from the home telephone or narrowband switch at the CO into the ADSL modem. Telephone companies are finding out that residential wiring is a huge source of impulsive noise on the ADSL upstream path. The impulsive noise originates from the ring trip signals, electrical switches, electrical appliances and tools, and powerline-switching transients coupled to the home tele-

phone lines. The filter also blocks the ADSL modem signals from going into the telephone set, reducing the quality of the line. The impedance of telephone wires varies significantly between lines. This depends heavily on the length and gauge of the wire, and also on the telephone itself, when the line is short. The addition of two filters, one at the CO and one at the home may be more complicated, particularly if one is trying to fit just one design to all cases. Most POTS splitter designs are passive due to their reliability. They do not require an external power, offer better protection from lightning, and support lifeline services. However, active POTS splitter designs may be required in countries outside the United States.

To address the need of reduced ADSL modem complexity, a new ADSL standard called G.Lite or "splitterless" DSL has been developed by the ADSL working group. In October 1998, the G.Lite standard was renamed as the G.922.2 standard. The G.Lite technology is capable of providing data rates up to 1.5-Mb/s for downstream channels and up to 386-kb/s for upstream channels. One of the main drivers for the G.Lite standard was to eliminate the installation of the POTS splitter by a telephone company technician, and thus reduce deployment cost.

Very-high-bit-rate DSL (VDSL) and high-bit-rate DSL (HDSL) modems are related technologies that are expected to be used for video distribution and interactive multimedia applications for the residential consumer market [24]. There are many unresolved technical and business issues that impede the mass deployment of these modems.

### 1.4.2 Fiber-In-The-Loop (FITL)

Fiber-in-the-loop (FITL) access technology, which uses optical fibers in a star architecture (point-to-multipoints), consists of a family of network architectures such as fiber-to-the-node (FTTN), fiber-to-the-curb (FTTC) and fiber-to-the-home (FTTH) [25]. The FITL systems were intended to be compatible with typical local exchange carrier (LEC) service, transmission, and operation. The initial FITL architecture is the FTTC as shown in Figure 1.5. A FITL network consists of a host digital terminal (HDT) with subtending broadband network units (BNUs) in a star architecture that is managed by the HDT. The HDT provides the necessary transmission and operations interface of the FITL system to the rest of the LEC network. For example, the HDT may separate locally switched traffic and nonlocally switched traffic for routing management. Broadband services such as Internet, interactive multimedia, and telephony are transmitted to the HDT, which may be located at the CO or a remote site, as baseband digital signals. This is in contrast with HFC networks, where the broadband services are RF modulated. At the HDT, the baseband digital signals are switched and sent to various broadband network units via optical fibers in a star architecture[6]. The BNU is located in the customer neighborhood and serves multiple customers. The BNU, which terminates the optical fiber from the HDT, performs optical-to-electrical

[6] Broadband network unit is also sometimes called Optical Network Unit (ONU).

conversion and other essential functions. The electrical signals are then transmitted to the subscriber's home via coaxial cable or over twisted copper pair wires. A network interface unit at the side of the subscriber's home separates the video signals, data signals, and telephony signals as shown in Figure 1.6. The digital video signals are demultiplexed and decoded using a set-top box.

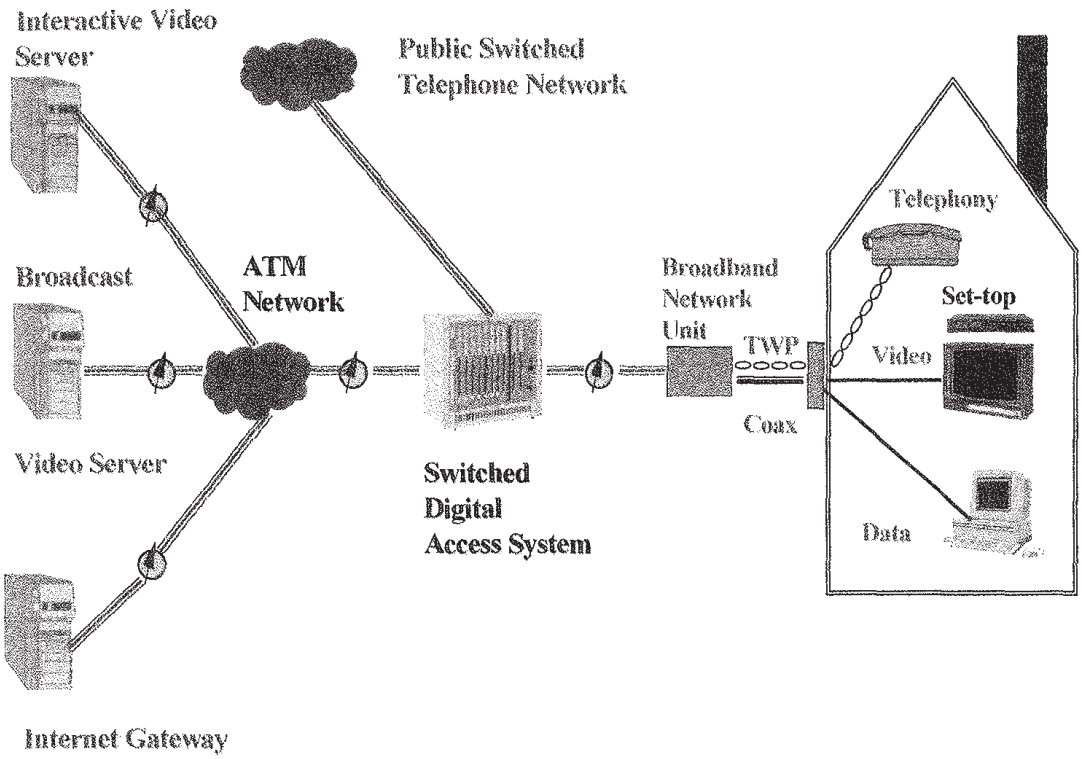



Figure 1.6 Basic fiber-to-the-curb network architecture.

Another FITL architecture is FTTH. As the name implies, the optical fiber replaces the coaxial cable or the twisted copper pair wires from the BNU to the subscriber's home. The difference between the FFTC and FTTH architecture lies in the location of the BNU.

In FTTH architecture, the BNU is located at the subscriber's home. Thus, the need to use twisted copper pair technology outside the subscriber's home is eliminated in the FTTH architecture. The portion of the network between the HDT and the BNUs becomes a passive optical fiber network, which is essential for its future upgrade. The large available bandwidth (> 2-GHz) of the optical fiber to deliver broadband services to the subscriber's home is one of the main advantages of this all-digital fiber architecture. With the use of WDM

technology, FTTH systems are capable of delivering multi-gigabit, e.g. OC-48 or SONET-compatible broadband services to each subscriber. In addition, since each BNU is located at the subscriber's home, no external power or additional maintenance is needed. The main disadvantage of the FTTH architecture is the relatively high cost of each BNU, and the initial installation costs of the optical fibers. Despite this disadvantage, the FTTH architecture has a very promising potential to deliver multi-gigabit broadband services to each subscriber's home.

### 1.4.3 Direct Broadcast Satellite (DBS)

While ADSL and FITL access technologies use wireline networks to provide broadband services, direct-broadcast-satellite (DBS) technology is based on various geosynchronous satellites that provide multichannel digital video programming services to subscribers equipped with DBS receivers [26]. DirecTV and EchoStar (also known as the Dish network) dominate the DBS industry and offer several hundred digital video channels. DirecTV service, which was launched in the summer of 1994, consists of an 18" satellite antenna dish and a digital integrated receiver and decoder (IRD), which tune and decode the selected digital channel. The DirecTV service is distributed by three high-power HS 601 satellites (DBS-1, DBS-2, and DBS-3). Each satellite features 16 120-watt Ku-band transponders with DBS-2 and DBS-3 each configured to provide 8 transponders at 240 watts. DBS-1 delivers up to 60 channels and over 20 channels of programming from USSB®. Recently, USSB was acquired by DirecTV and merged its service offering. With DBS-1, DBS-2, and DBS-3, DirecTV service has several hundred video and audio channels of programming. All three satellites are colocated in geosynchronous orbit 22,300 miles above the earth at 101 degrees west longitude.

EchoStar Communications Corporation, which was founded in 1980, was granted a DBS orbital slot at 119 degrees west longitude, and three years later, the DISH (Digital Sky Highway) network brand name was created. On December 28, 1995, EchoStar successfully launched its first DBS, EchoStar I, and DISH network came to life on March 4, 1996. The second DBS satellite, EchoStar II was successfully launched on September 10, 1996 and is also orbiting at 119°. On October 5, 1997 EchoStar launched the satellite EchoStar III into an orbital slot at 61.5 degrees west longitude. EchoStar IV was launched from Kazakhstan on May 8, 1998. These four satellites give DISH network delivery capacity for over 250 channels of digital video, audio, and data services. The satellite-delivered programming service, which is extracted from various video sources, is first digitized, encrypted, and uplinked to the orbiting DBS satellites. Then, these DBS satellites retransmit the video signals back down to every subscriber's home with the properly oriented receiver dish. Each of the transponders on the DBS-1 satellite can transmit MPEG-2 video streams at a rate more than 23 Mb/s, while the DBS-2 and DBS-3 satellites can transmit about 30 Mb/s. For further details on the MPEG standard, please see Chapter 2, Section 2.3.

### 1.4.4 Multichannel Multipoint Distribution Service (MMDS)

Multichannel multipoint distribution service (MMDS) access technology is another wireless technology that is based on terrestrial broadcast of analog and digital video channels. In the early 1960s, the FCC began setting aside certain RF frequencies, which were called instructional television fixed service (ITFS), for distance learning. In the 1970s, various equipment vendors petitioned the FCC to allow commercial use of part of the ITFS spectrum. The industry began by offering one and two video channel service to consumers, which was known as multipoint distribution service (MDS). Since many of the ITFS frequencies were not used in many geographical locations, the MMDS operators petition the FCC to reallocate the ITFS frequencies for MMDS use. When additional frequencies in the ITFS spectrum were allocated for video services, MDS was renamed as MMDS.[7] In September 1998, the FCC ruled for two-way MMDS. The MMDS spectrum has been allocated worldwide in various RF frequency bands from 2-GHz to 2.7-GHz.

**Table 1.2 MMDS frequency plan for downstream and upstream channels.**

| Downstream Center Frequency (MHz) | Downstream Channel # | Upstream Center Frequency (MHz) | Upstream Channel # |
|---|---|---|---|
| 2599 | E1 | 2675 | H3-A |
| 2605 | F1 | 2677 | H3-B |
| 2611 | E2 | 2679 | H3-C |
| 2617 | F2 | 2681 | G4-A |
| 2623 | E3 | 2683 | G4-B |
| 2629 | F3 | 2685 | G4-C |
| 2635 | E4 | 2687 | H4-A |
| 2641 | F4 | 2689 | H4-B |
| 2647 | G1 | 2151 | MDS1-A |
| 2653 | H1 | 2153 | MDS1-B |
| 2659 | G2 | 2155 | MDS2-A |
| 2665 | H2 | 2157 | MDS2-A |
| 2671 | G3 | 2159 | MDS2-B |

Table 1.2 shows the MMDS frequency plan in the United States for both downstream and upstream channels. In addition, the FCC has allocated additional RF spectrum in 900-MHz, 2.4-GHz, 5.8-GHz, and 24-GHz bands for unlicensed use to encourage competition. The unlicensed bands are referred sometimes as industrial scientific and medical (ISM) bands.

[7] MMDS is sometimes called "wireless cable."

The wireless network operators would use the unlicensed spectrum to offer broadband services such as Internet, telephony, and interactive multimedia to businesses at cost-effective rates compared with HFC cable networks.

The basic MMDS architecture consists of MMDS radio transmitters mounted on radio towers with antennas, a subscriber premises antenna, a downconverter, and an STB. Each geographical serving area is divided into slightly overlapping cells, each with a typical radius of 40-km. For robust transmission, line-of-sight between the transmitter and receiver antenna is typically required. Since line-of-sight is not always available, the primary channel impairment mechanism in MMDS systems is multipath signal fading.

A related access technology to the MMDS is called hybrid fiber wireless (HFW) or hybrid fiber radio (HFR). This architecture is similar to HFC where a central headend transmits broadband services to various RF cell sites via SMF cables, except the connection from the RF cell site to the subscriber is accomplished via two-way MMDS. There are many advantages to such architecture. First, it is a more reliable two-way transmission between the subscriber and the headend compared with a traditional MMDS architecture. Second, it reduces RF site installation and maintenance costs (less equipment at each site). Third, this architecture is well suited for large-scale deployment in metropolitan areas where fiber networks may already be installed.

## References


1. Special issue on "Residential Broadband Services and Networks," *IEEE Communications Magazine*, **35**, no.6 (1997).
2. R. Olshansky, V. A. Lanzisera, and P. M. Hill, "Subcarrier Multiplexed Lightwave Systems for Broadband Distribution," *IEEE Journal of Lightwave Technology* **7**, 1329–1324 (1989).
3. W. I. Way, "Subcarrier Multiplexed Lightwave System Design Considerations for Subscriber Loop Applications," *IEEE Journal of Lightwave Technology* **7**, 1806–1818 (1989).
4. T. E. Darcie, "Subcarrier Multiplexing for Lightwave Networks and Video Distribution Systems," *IEEE Journal on Selected Areas in Communications* **8**, 1240–1248 (1990).
5. J. A. Chiddix, J. A. Vaughan, and R. W. Wolfe, "The Use of Fiber Optics in Cable Communications Networks," *IEEE Journal of Lightwave Technology* **11**, 154–166 (1993).
6. W. S. Ciciora, "Inside the Set-top Box," *IEEE Spectrum Magazine*, 70–75 (1995).
7. S. Ovadia and C. Lin, "Performance Characteristics and Applications of Hybrid Multichannel AM-VSB/M-QAM Video Lightwave Transmission Systems," *IEEE Journal of Lightwave Technology* **16**, 1171–1186 (1998).
8. TR-NWT-00253, *Synchronous Optical Network (SONET) Transport System: Common Generic Criteria*, Issue 2 (Bellcore, December 1991).

9. U. Black and S. Waters, *SONET and T1: Architectures for Digital Transport Networks*, PTR Prentice Hall, New York (1997).
10. A. N. Nair, "Interactive Television Broadcasting and Reception," *Digital Consumer Electronics Handbook*, R. K. Jurgen, Editor-in-Chief, McGraw-Hill, New York (1997).
11. SCTE DVS/178, "Cable System Out-Of-Band Specifications," (1998).
12. T. C. Kwok, "Residential Broadband Internet Services and Applications Requirements," *IEEE Communications Magazine*, 76–83 (1997).
13. IEEE 1394-1995 High Performance Serial Bus Interface Standard (1998).
14. CableLabs, MCNS Data-Over-Cable Service Interface Specifications: Radio Frequency Interface Specification SP-RFIv1.1-I01-990311 (1999).
15. Broadcom and Texas Instruments, "Advanced TDMA Proposal for HFC Upstream Transmission," Rev. 1.0, December (1999).
16. D. Minoli and E. Minoli, *Delivering Voice over IP Networks*, John Wiley & Sons, New York (1998).
17. Jim Forster, "An integrated Architecture for Video, Voice, and IP services over SONET and HFC Cable Systems," *SCTE Proceedings Manual of Conference on Emerging Technologies*, 155–199 (1998).
18. C. Huitema, J. Cameron, P. Mouchtaris, and D. Smyk, "An Architecture for Residential Internet Telephony Service," *IEEE Network* **13**, 50–56 (1999).
19. M. Arango et al., "Media Gateway Control Protocol (MGCP)," Internet Engineering Task Force (IETF) Internet draft, October (1998).
20. F. Cuervo et al., "SS7-Internet Interworking Architectural Framework," Internet Engineering Task Force (IETF) Internet draft, July 9 (1998).
21. CableLabs, PacketCable Network-Based Call Signaling Protocol Specification, PKT-SP-EC-MGCP-I02-9901201.
22. ITU-T Recommendation H.323, *Visual Telephone Systems and Equipment for Local Area Networks which Provide a Non-Guaranteed Quality of Service* (1996).
23. W. Y. Chen and D. L. Waring, "Applications of ADSL to Support Video Dial Tone in the Copper Loop," *IEEE Communications Magazine*, 102 (1994).
24. D. Rauschmayer, *ADSL/VDSL Principles: A Practical and Precise Study of Asymmetric Digital Subscriber Lines and Very High Speed Digital Subscriber Lines*, MacMillan technical publishing (1998).
25. TR-NWT-000909, *Generic Requirements and Objectives for Fiber in the Loop*, Issue 2 (Bellcore, December 1993).
26. See for example http://www.dbsdish.com/ Web site, which has the latest news and information on the DBS industry.
27. S. B. Moghe and D. Urban, "Broadband Wireless Services Based on Combined Use of MMDS, LMDS and Unlicensed Band Spectrum," *Technical Papers of the Fifth Annual Technical Symposium Wireless Communication Association International* (1998).
28. J. C. Cornelius, "Design and Implementation of a Two Way Data System in an Existing Analog market," *Technical Papers of the Fifth Annual Technical Symposium Wireless Communication Association International* (1998).

# CHAPTER 2

# BASIC CABLE TV BACKGROUND: MODULATION SIGNAL FORMATS AND COAXIAL CABLE SYSTEMS

As we learned in chapter 1, HFC networks have evolved from traditionally broadcasting only multichannel analog video signals to a two-way mixture of analog and digital signals carrying both digital video and high-speed data. This chapter provides the basic technical background of cable TV systems carrying analog and digital video signals. Section 2.1 will explain the most commonly used analog video modulation formats, while Section 2.2 will discuss the different digital video and audio signal standards. The primary cable TV frequency plans that are used in the U.S. are reviewed in Section 2.3. In Section 2.4, the basic characteristics of multichannel coaxial systems are reviewed in terms of the properties of their components such as coaxial cables, taps, and RF amplifiers. The different degradation mechanisms over multichannel coaxial cable plants such as second- and third-order nonlinear distortions, multipath micro-reflections or echoes, group delay, and AM hum modulation are reviewed in Section 2.5. The last section will discuss the sources of upstream noise, including ingress noise, and their impact on the return-path transmission over cable TV networks.

## 2.1 Analog Modulated Video Signal Formats

### 2.1.1 NTSC and AM-VSB Video Signals

The National Television Systems Committee (NTSC) initiated the basic monochrome TV standard in the U.S. in 1941 by broadcasting 525 interlaced lines at 60 fields per second, which was designated as system M by the Commite' Consulatif International Radiocommunications (CCIR). In 1953, the NTSC of the Electronic Industries Association (EIA) established color TV standards, which are now in use for terrestrial broadcasting and cable TV transmission systems in North America, Japan, and many other countries. This color TV system was designed to be compatible with the monochrome (black and white) TV systems that existed previously [1]. The specifications of a composite NTSC video signal, as described by NTSC, include a 525-line interlaced scan at horizontal frequency of 15,734.26

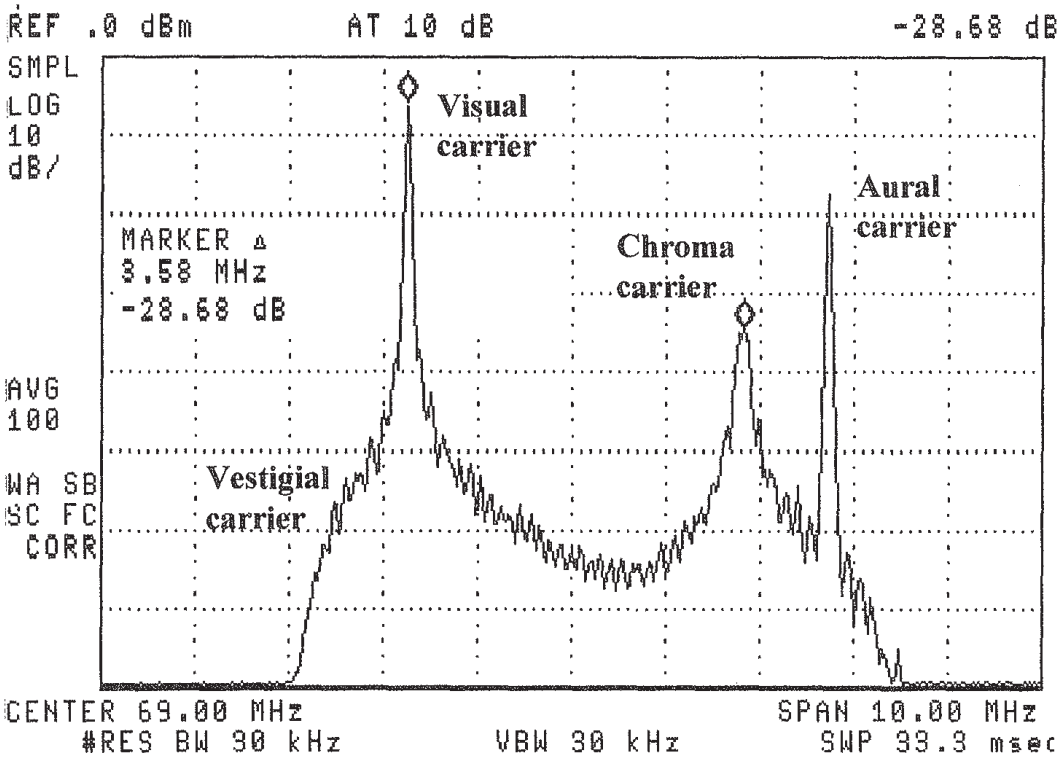



Figure 2.1 Measured frequency spectrum of NTSC video signal at channel 4, showing the luma, chroma, and aural carriers.

Hz; 15,750 Hz for monochrome TV, and a vertical scan frequency of 59.94 Hz (60 Hz for monochrome TV).

A complete video image as seen on a TV monitor is called a frame, which consists of two interlaced vertical fields with 262.5 lines each. The TV image is scanned at a vertical frequency of 59.94 Hz such that the lines of field 2 are interlaced with the lines of field 1 to create the desired 525 lines frame at a repetition rate of 29.97 Hz. Historically, the 60-Hz vertical scan frequency was selected for monochrome TV sets to match the 60-Hz power line rate such that any power-related distortions would appear stationary. For color TV, both the horizontal and vertical scan frequencies have been slightly reduced from the monochrome display case to allow for the interference beat between the chrominance carrier and the aural carrier to be synchronized to the video signal. Figure 2.1 shows the modulated RF spectrum of an NTSC video signal based on AM-VSB format at an RF visual carrier frequency of 67.25 MHz (channel 4). The brightness portion of the video signal, which contains all the information of the picture details, is commonly called the luminance or visual carrier. The color portion of the video signal, which contains information about the picture hue (or tint) and color saturation, is commonly called chrominance or *chroma* carrier. The

chroma carrier is located 3.579545 MHz above the luma carrier. The hue information, which is contained in the phase angle of the 3.58-MHz chroma carrier, allows one to distinguish between the different colors such as red (R), green (G), blue (B), etc. Notice that white, gray, and black are not hues. Color saturation indicates the amount of white light dilution of the hue, and often is expressed in percentages. Thus, 100%-saturated color is a pure hue, which is undiluted by white light. The saturation information is carried by the amplitude magnitude of the chroma carrier. Since the human eye response varies from one hue to another, different amplitudes for different colors are required for 100% color saturation. For more details on the 1931 Commission Internationale de l'Eclairage (CIE) colorimetric standard, the reader is encouraged to look at these references [1–3].

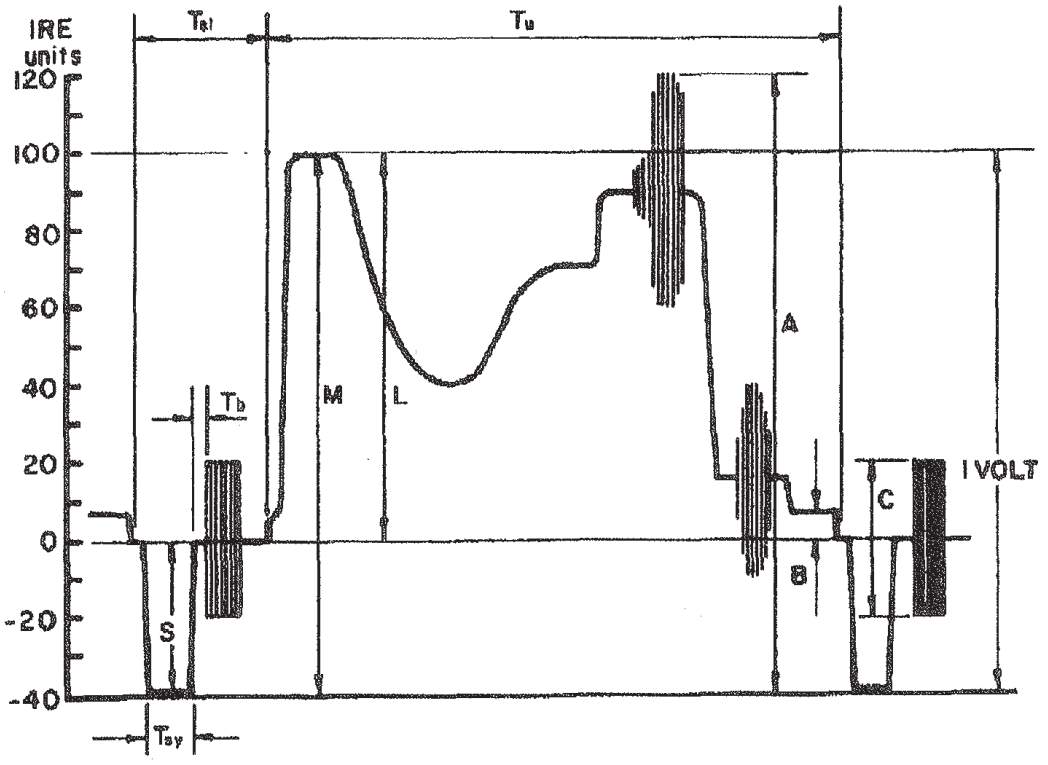



Figure 2.2 Time domain waveform of standard NTSC composite color video signal, showing its components as follows: (A) the peak-to-peak amplitude of the composite color video signal, (B) the difference between the black level and blanking level, (C) the peak-to-peak amplitude of the color burst, (L) nominal value of luminance signal, (M) peak-to-peak amplitude of monochrome video signal, (S) amplitude of synchronizing signal; $T_b$, duration of breezeway, $T_{sl}$, duration of line blanking period, $T_{sy}$, duration of synchronizing pulse, $T_u$, duration of active line period. After Ref. [2].

It is well known that the human eye is more sensitive to the color green than blue. Thus, the white color or visual carrier can be constructed from the three primary colors, namely $R$, $G$, and $B$, each with different weighting normalized to 1.0 as follows:

$$Y = 0.3R + 0.59G + 0.11B = 1.0 \tag{2.1}$$

Using this relationship, the color information of any point can be coded into two chosen color difference signals: $(R - Y = V)$ and $(B - Y = U)$. The NTSC system employs quadrature amplitude modulation (QAM) of the 3.58-MHz chroma carrier to convey the color difference information of the composite chroma signal. This method of color broadcast provides a compatible signal that can be reproduced on black-and-white as well as color TV sets.

Figure 2.1 also shows the aural (sound) carrier, which is located 4.5 MHz above the visual carrier with frequency deviation of ±25 kHz for monaural audio. The analog sound carrier is frequency modulated (FM), and in HFC networks must be maintained between 10 dB to 17 dB below the visual carrier amplitude according to Federal Communications Commission (FCC) requirements [4]. Notice the 0.75-MHz lower sideband from the luma carrier, which is typically called the vestigial sideband. This is the portion of the lower sideband of the composite video signal that remained as a tradeoff between minimizing the NTSC signal bandwidth and preserving picture details as much as possible.

To gain further understanding about NTSC signals, let us examine the time-domain waveform of standard NTSC composite video as shown in Figure 2.2 [2]. The standard NTSC composite video signal is specified to be 1-volt peak-to-peak (p-p) from the tip of the horizontal synchronization pulse to 100% white, and is divided into 140 equal parts called IRE units. The horizontal blanking pedestal, which is the 0 IRE reference level, starts as the sweeping electron beam of the picture tube reaches slightly over the extreme right-hand side edge of the screen. This prevents illumination of the TV screen during retrace, namely, until the electron beam deflection circuits are reset to the left edge of the TV screen and ready to start another line scan. In a typical TV set, the blanking level is "blacker than black" to assure no illumination during retrace. This difference, which is 7.5 IRE units, is shown as (B) in Figure 2.2. The tip of the horizontal synchronization pulse, which is abbreviated as the sync pulse, is at –40 IRE, and about 0.3 volt (p-p). Thus, the various gray levels of the luma carriers are divided between +7.5 IRE units and +100 IRE units (100% white).

During the vertical blanking interval, the first 21 lines are not displayed as the TV set prepares to receive a new field. This is commonly called the vertical blanking interval (VBI). The FCC requires that broadcasting TV stations include special vertical interval test signals (VITS) in each TV field [5]. The VITS are typically inserted between lines 10 and 21 to allow in-service testing of the TV broadcasting equipment. The FCC composite and the NTC-7[1] test signals have the necessary video signal components to perform the FCC re-

[1] NTC report number 7 (1976) was prepared by the Network Transmission Committee of the Video Transmission Engineering Advisory Committee, a joint committee of television network broadcasters and the Bell system.

quired tests for NTSC signals. Figure 2.3 shows, for example, the FCC composite test signal, which consists of a line bar, a 2T pulse, a chroma pulse, and a modulated five-riser staircase signal. The primary difference between these two test signals is the test sequence in which the video signal components are presented.

The relationship between the amplitude of the composite video signal and the percentage of amplitude modulation of the RF carrier is also unique. TV transmission uses a negative modulation such that the sync pulse produces the maximum RF peak-to-peak amplitude modulated envelope of the composite signal, while the white portion of the signal (+100 IRE units) produces the RF minimum amplitude of the modulated envelope. Negative modulation means that brightest portion of the composite video signal is represented by the minimum RF power. The depth of AM modulation of the composite video signal is specified to vary between 82.5 and 90% with a nominal value of 87.5%. The depth of modulation is less than 100% in order to allow the sound carrier to be recovered for even during the brightest portion of the composite video signal. This relationship between the amplitude and depth of modulation is advantageous, in order to minimize the visual effect of noise interference, which typically shows up as "snow" in the picture, at the weakest portion of the signal.

Figure 2.2 also shows four main time intervals in the composite video signal waveform. The horizontal blanking pedestal consists of the front porch, which lasts 1.47 μs, the -40 IRE sync pulse ($T_{sy}$) with duration of 4.89 μs, and a 4.4 μs back porch at the blanking level. During the back porch period, color is being generated in 8 to 10 cycles of 3.58 color burst at a specific reference phase for all the TV display lines.

### 2.1.2 NTSC Signal Test Parameters

There are various parameters, which are often quoted in the literature or in product specifications, to measure the quality of an NTSC video signal after its transmission. In this section a short overview of the most important test parameters of NTSC video signal is provided. The main NTSC video signal test parameters are

- Signal-to-(weighted)-noise-ratio (SNR)
- Differential gain (DG)
- Differential phase (DP)
- Chrominance-to-luminance delay inequality (CLDI)

The NTC-7 SNR is the ratio of the peak luma carrier level, 714 mV or 100 IRE, to the weighted rms noise level contained in 4-MHz bandwidth. The peak luma carrier level must be measured across a terminating impedance of 75 ohms. The weighted noise is used to compensate for the visual responses of the human eye, namely, different viewers perceive greater picture degradation at some frequencies more than other frequencies. A special luma weighted filter (NTC-7) per CCIR Recommendation 567 is primarily used in the U.S. The carrier-to-noise-ratio (CNR) is approximately equal to the SNR (within 0.5 dB) for

AM-VSB video signals. The CNR measurement on the spectrum analyzer is done with 30-kHz resolution bandwidth (RB), 100-Hz or 300-Hz video bandwidth (VB), and in automatic sweep mode.

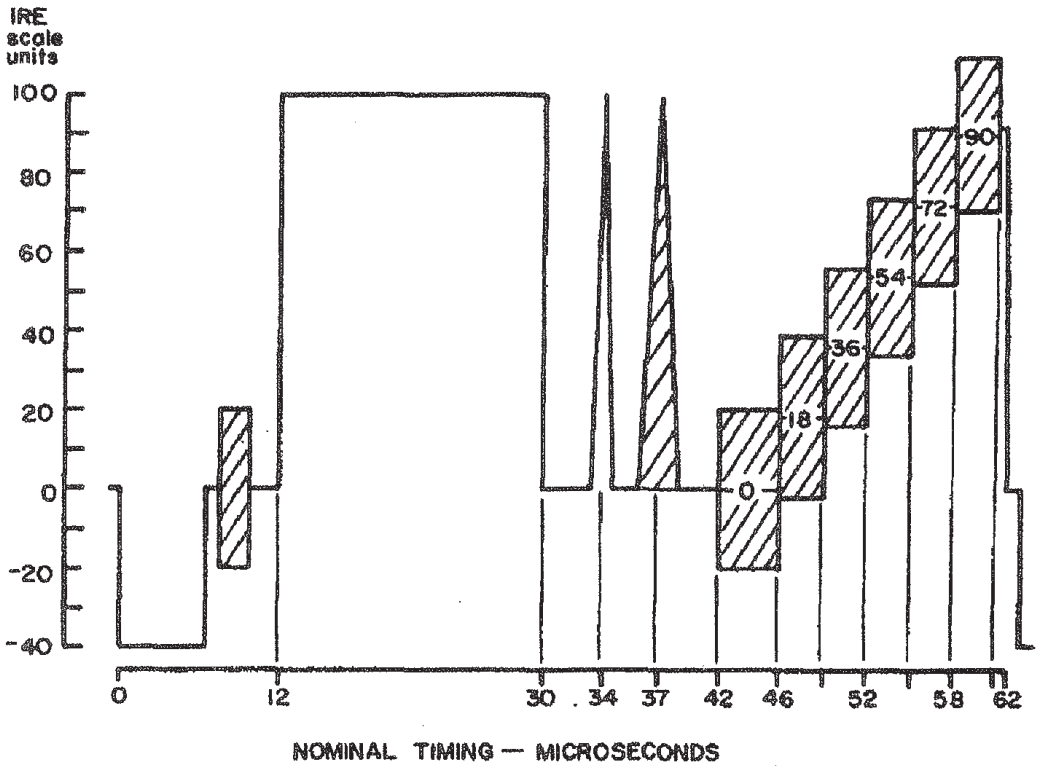



Figure 2.3 FCC Composite test signal within one horizontal scan period. After Ref. [2].

There are several noise correction factors that are best explained through an example. Suppose one measures an uncorrected CNR of 63 dB. Then, the corrected CNR is calculated as follows: CNR = 63 dB – (21.25 + 2.5 – 0.52 –2.2) = 41.97 dB, where the 21.25 dB is the conversion factor from a RB of 30 kHz to 4 MHz, and the 2.5 dB is log detect Rayleigh noise correction factor for the spectrum analyzer. Since most spectrum analyzers have a Gaussian shaped video filter, the measured 3-dB noise bandwidth is larger than an ideal rectangular filter. Thus, the corrected noise needs to be reduced by about 0.52 dB. In addition, the 2.2 dB is the spectrum analyzer noise floor correction factor, assuming the noise floor drops only 4 dB when disconnecting the coaxial cable from the spectrum analyzer. According to the FCC Section 76.605, the measured video signal at the input to the subscriber's terminal is required to have a CNR equal to or greater than 43 dB.

The differential gain (DG) is the change in amplitude of the chroma signals (measured at 3.58 MHz) as the luminance level changes. It is measured as the difference in amplitude

between the largest and smallest component of the chroma signal, divided by the largest component and expressed in % or dB of the largest component. There are two types of modulated staircase signals, namely, the previously mentioned FCC composite test and NTC-7. The NTC-7 tests are intended for baseband video measurements. The FCC composite test signal should be used when using a video test generator. However, the generator is not needed if the channel to be tested already contains the FCC VITS. In the FCC Section 76.605, the DG is required not to exceed 20% for the received NTSC signal at the subscriber's terminal.

The differential phase (DP) is the change in the phase of the chroma signal as the amplitude of the luma signal changes. The same test signals to measured DG should be used to measure DP. According to FCC Section 76.605, the DP for the color carrier is measured as the largest phase difference in degrees between each chroma component and a reference component (the component at the blanking level of 0 IRE) should not exceed ±10 degrees.

The chrominance-to-luminance delay inequality (CLDI) is the change in time delay of the chroma signal relative to the luminance signal of an NTSC video signal as measured at the output of a modulator or processing unit at the cable headend. According to the FCC Section 76.605, the CLDI shall be within 170 ns. This test is typically done inserting the VITS signal before it is received at the cable TV headend.

### 2.1.3 PAL and SECAM Video Signals

Two other color TV standards, which are mostly used in various countries outside of North America, are the phase alternation line (PAL) standard and the sequentiel couleur avec memoire (SECAM) standard. There are different variations of the PAL standard, which are mostly used in European countries and are designated by different letters such as PAL "I" (United Kingdom), PAL "B", PAL "G", PAL "H" (continental Europe), and PAL "M" (Brazil) [1].

Like the NTSC system, the PAL system employs QAM modulation of the chroma carrier to convey the color-difference information as a single composite chroma signal. The $(R-Y)$ color-difference carrier is phase inverted (180°) on alternate lines to improve the picture performance for different impairment conditions. For example, suppose that for a red hue, the $(R-Y)$ color component in the NTSC signal has a phase of +90°. In a PAL signal, one line is identical to the NTSC signal, but on the next line, the phase of this $(R-Y)$ component is switched to +270°, returning to +90° on the following line, etc. This phase-alternation line of the $(R-Y)$ component gives the PAL system its name. The color burst signal, which is used by the NTSC system during the horizontal synchronization time interval, is eliminated in the PAL system because of the alternating phase of line by line. The "simple" PAL system relies upon the human eye to average the line-by-line color switching process. Thus, the picture on a PAL TV can be degraded by 50-Hz line beats caused by system nonlinearities introducing visible luminance changes at the line rate. To solve this problem, an accurate and stable delay element was used to store the chroma signal for one complete line period in what is sometimes called PAL-deluxe (PAL-D).

Similar to the NTSC system, the PAL system has an equal bandwidth (1.3 MHz at 3 dB) for the two color-difference components, namely *U* and *V*. There are small variations of the PAL system among different countries, particularly where 7-MHz and 8-MHz channel bandwidths are used. Table 2.1 summarizes the key technical specifications for both NTSC and PAL video signals.

**Table 2.1 Basic technical specifications for NTSC and PAL video signals**

| Parameter/Format | NTSC | PAL |
|---|---|---|
| Number of lines per picture | 525 | 625 |
| Field (vertical Scan Frequency | 59.94 Hz | 50 Hz |
| Horizontal Scan Frequency | 15,734.26 Hz | 15,625 Hz |
| Channel Bandwidth (MHz) | 6 | 6 System "N"<br>7 System "B"<br>8 System "I" |
| Nominal Video Bandwidth (MHz) | 4.2 | 4.2 System "N"<br>5 Systems "B", "G", "H"<br>6 Systems "D", "K", "L" |
| Color Subcarrier relative to the vision carrier (MHz) | 3.579545 | 4.433618 Systems "I", "B", "G", "H"<br>3.575611 System "M" |
| Sound Sub-carrier relative to vision carrier (MHz) | 4.5 | 4.5 System "N"<br>5.5 Systems "B", "G", "H"<br>5.9996 System "I" |
| Sound Modulation | FM | FM |
| Vestigial Sideband (MHz) | 0.75 | 0.75 Systems "B", "G", "N"<br>1.25 Systems "H", "I", "L" |
| Gamma | 2.2 | 2.8 |

The PAL system "I", which has an 8-MHz channel bandwidth, has added a new digital sound carrier, which is called *near-instantaneous companding audio multiplex* (NICAM). The NICAM carrier is located 6.552 MHz above the visual carrier (exactly nine times the bit-rate), and is differentially QPSK modulated at a bit rate of 728 kb/s and it is 20 dB below the visual carrier peak-sync power. Since the new sound carrier is located close to the analog FM sound carrier as well as the adjacent channel luminance carrier, the digital sound signal is scrambled before it is being modulated.

The SECAM system, which was adopted by France and the USSR in 1967, has several common features with the NTSC system such as the same luminance carrier given by equation 2.1 and the same color-difference $U$ and $V$ components. But this approach is considerably different from both the NTSC and PAL systems in the way the color information is being modulated onto the composite chroma carrier. First, the color-difference components are transmitted alternately in time sequence from one successive line to the next with the same visual carrier for every line. Since there is odd numbers of lines in a frame, any given line carries the $(R-Y)$ component on one field and the $(B-Y)$ component on the next field. Second, the color information in the $(R-Y)$ and $(B-Y)$ components is carried using FM. Similar to the PAL-D receiver, an accurate and stable delay element is used in synchronization with the switching process to have a simultaneous existence of the $(B-Y)$ and the $(R-Y)$ signals in a linear matrix to form the $(G-Y)$ color-difference component. In addition, the bandwidth of the chroma components U and V is reduced to 1-MHz. It should also be noted that the SECAM system employs AM modulation of the sound carrier as opposed to the FM modulation that is used in both NTSC and PAL systems.

## 2.2 Digital Video and Audio Signals

Digital coding of NTSC composite signals typically requires a raw bit rate of about 115 Mb/s, assuming analog-to-digital conversion of 8-bit and 4 times sampling of the chroma carrier. Obviously, this is a very high transmission rate to stuff in a 6-MHz channel bandwidth. To overcome this limitation, a standard for digital video (sequences of images in time) and audio compression has been developed by the MPEG (moving pictures experts group) committee under ISO (the international standards organization) [6-9]. The MPEG technology includes different patents from many companies and individuals around the world. However, the MPEG committee only deals with the technical standards and does not address the intellectual property issues. Related standards for still image compression are called joint photographic experts group (JPEG) and joint bilevel image experts Group (JBIG), which is used for binary image compression such as faxes. Other non-MPEG standards for video/audio are discussed in Section 2.2.3.

### 2.2.1 MPEG-1 Standard

A picture element or pixel is formed by $Y$, $U$, $V$ sample values. Thus, if all the three components ($Y$, $U$, $V$) use the same sampling grid, each pixel has three samples (see Section 2.2.2). The first result of the MPEG committee was called MPEG-1. MPEG-1 compression starts with a relatively low-resolution video sequence of about 352 by 240 pixels by 29.97 frames/s (US), but with the original compact disk (CD) quality audio. The video compression method relies on the human eye's inability to resolve high frequency color changes in a picture. Thus, the color images are converted to the $YUV$-space, and the two-chroma components ($U$ and $V$) are decimated further to 176 by 120 pixels. The 352x240x29.97 rate is

derived from the CCIR-601 digital television standard, which is used by professional digital video equipment. 720 by 243 by 59.94 fields per second are used, where the fields are interlaced when displayed. Note that fields are actually acquired and displayed at a 1/59.94 second apart. The chroma components are 360 by 243 by 59.94 fields a second interlaced. This 2:1 horizontal chroma decimation is called 4:2:2 sampling. The source input format for MPEG-1, called SIF, is CCIR-601 decimated by 2:1 in the horizontal direction, 2:1 in the time direction, and an additional 2:1 in the chroma vertical direction. For PAL and SECAM video standards where the display frequency is 50-Hz, the number of lines in a field is increased from 243 or 240 to 288, and the display rate is reduced to 50 fields/s or 25 frames/s. Similarly, the number of lines in the decimated chroma components is increased from 120 to 144. Since 288.50 = 240.60, the two formats have the same source data rate. Section 2.2.2 has additional discussion on the chroma subsampling for MPEG-2 video.

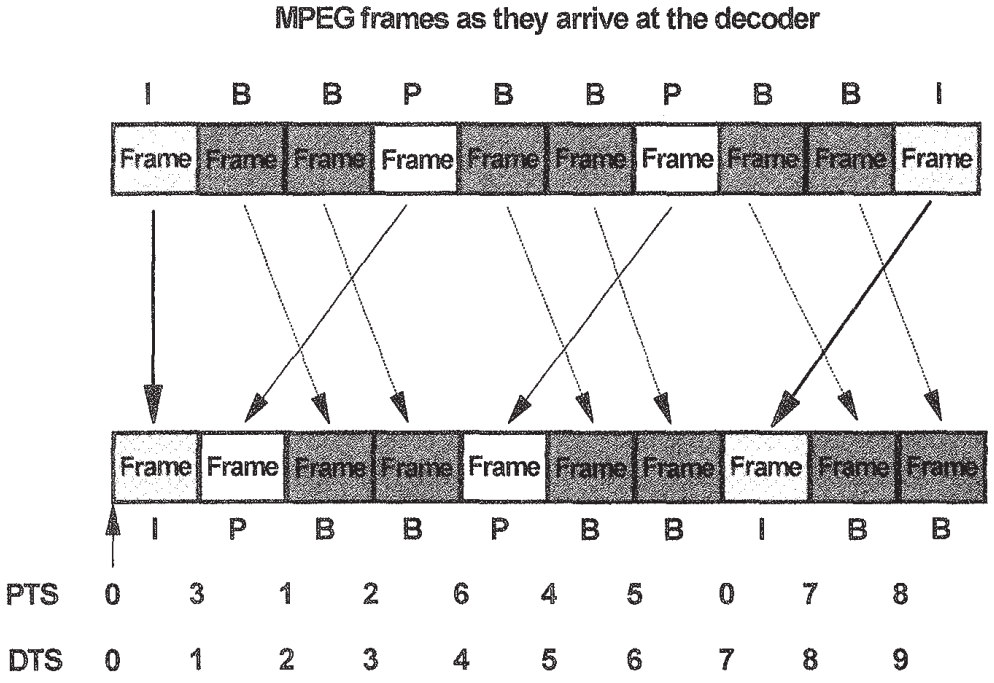



Figure 2.4 PES frames as they arrive at the decoder and as they are displayed, showing the corresponding PTS and DTS.

Macroblock is the basic building block of an MPEG picture. It consists of 16x16-luminance sample together with one 8x8 block of samples for the two-color components. The basic method is to predict image motion from frame to frame, and then to use the discrete cosine transform (DCT) to organize the redundancy in the spatial directions. The

DCTs are done on 8x8 pixel blocks, and the motion predictions are done on the luminance signals (*Y*) in 16x16 pixel blocks.

The second step of the encoding process is called *entropy coding*. The task of the entropy encoder is to transmit a particular stream of symbols in as few bits as possible. The technique for making optimal integer code assignments is called Huffman coding, which is used to code the result of the DCT coefficients, the motion vectors, and the quantization parameters with fixed tables. The DCT coefficients have a special Huffman table that is two-dimensional in that one code specifies a run-length of zeros and the nonzero value that ended the run. Another statistical coding model that is used to code the motion vectors and the frequency-independent DCT components is called differential pulse code modulation (DPCM). Here, the difference between each picture-element (pel) and a prediction is calculated from neighboring pel values already transmitted. Pel differences are considerably less correlated than the original pel values, and can be coded independently with reasonable efficiency.

There are three types of coded frames. Intra (I) frames are simply frames coded as still images, not using any past history, while predicted (P) frames are predicted from the most recently reconstructed I or P frames. Each macro-block in a P frame either comes with a vector and DCT coefficients for the closest match in the last I or P frame, or can be intra-coded (as with the I frames) if there was no good match. Bidirectional (B) frames are predicted from the closest two I or P frames, one in the past and one in the future. Thus, one is trying to search for the matching blocks in those frames, and see which one works best. In other words, given the 16x16 block in the current frame that one is trying to code, what is the closest match to that block in a previous or future frame? Beside the picture data, the video encoder attaches header information for each group of B, P, and I frames, known as group of pictures (GOP) to form the elementary video stream. A packetized elementary stream (PES) is formed by attaching additional header information such as program clock reference (PCR), optional encryption, packet priority levels, etc. to the elementary stream. The encoder or the MPEG multiplexer provides each PES with the so-called presentation time stamp (PTS) and decode time stamp (DTS). The DTS tells the decoder when to decode each frame, while the PTS tells the decoder when to display each frame. Before the PES is ready to be transmitted, a transport stream (TS) is formed. The packet structure for the MPEG-TS will be discussed in the next section on MPEG-2.

Figure 2.4 shows the ordering of B, P, and I frames for decoding with the corresponding PTS and DTS. The sequence of decoded frames usually goes like IBBPBBPBBPBBIBBPBBPB..., where there are 12 frames from one I frame to the next (U.S. and Japan only). This is based on a random access requirement that you need a starting point at least once every 0.4 seconds or so. The ratio of P frames to B frames may vary from one encoder to another. For the decoder to work, one would have to send that first P frame before the first two B frames, so the compressed data stream ends up looking like the example in Figure 2.4. One has to decode the I frame, then decode the P frame, keep both of those in memory, and then decode the two B frames. The I frame is probably displayed while the P frame is decoded, and the B-frames are displayed while they are decoded, and

then the P frame is displayed as the next P frame is decoded, and so on. Since bidirectional macroblock predictions are an average of two macroblocks blocks, noise is reduced at low bit rates.

Table 2.2 shows the typical P-, B-, and I-frame size (in kbits) for both MPEG-1 and MPEG-2 video signals. At nominal MPEG-1 video (352x240x30, 1.15 Mbit/sec) rates, it is said that B frames improve SNR by as much as 2 dB. However, at higher bit rates, B frames become less useful since they inherently do not contribute to the progressive refinement of an image sequence (i.e., not used as prediction by subsequent coded frames).

**Table 2.2 Typical frame sizes (in kilobits) for MPEG-1 and MPEG-2 compressions.**

| Compression Method | I frame | P frame | B frame | Average |
|---|---|---|---|---|
| MPEG-1 SIF @ 1.15 Mb/s | 150 kb | 50 kb | 20 kb | 38 kb |
| MPEG-2 @ 4.0 Mb/s | 400 kb | 200 kb | 80 kb | 200 kb |

The MPEG-1 coding was originally designed only for progressive frames display (e.g., computer monitors). In order to display the MPEG-1 syntax on an analog TV set, the video frames need to be interlaced. There are two methods that can be applied to interlaced video that maintain syntactic compatibility with MPEG-1. In the field concatenation method, the encoder model can carefully construct predictions, and prediction errors that realize good compression while maintaining field integrity (distinction between adjacent fields of opposite parity). Some preprocessing techniques also can be applied to the interlaced source video that would lessen the sharp vertical frequencies. This technique is not efficient, of course. On the other hand, if the original source was progressive (e.g., film), then it is relatively more trivial to convert the interlaced source to a progressive format before encoding. Reconstructed frames are reinterlaced in the decoder display process.

An essential element in the utilization of MPEG intercompression is called motion compensation. Motion compensation is used to minimize the effect of image movement from a reference picture to the predicted picture. Motion vectors are described in terms of horizontal and vertical image displacements. The key technical issues in using motion vectors are the precision of the motion vectors, the size of the image region assigned to a single motion vector, and the selection criteria for the best motion vector values. There are various techniques that are used to estimate the motion vectors such as mean absolute distortion (MAD) and mean square error (MSE).

One of the main misconceptions about MPEG-1 is that it has fixed or limited frame size (i.e., 352x240x29.97 frame/s or 352x288x25 frame/s). In fact, MPEG-1 can use any frame size, including CCIR-601 resolutions (704x480), with frame sizes as high as 4095x4095x60 frame/s. MPEG-2 is more limited since frame sizes must be multiples of 16.

Another misunderstood issue arises from the standard profile and is known as constrained parameters bit stream (CPB). The CPB is a series of restrictions that the MPEG-1 stream must meet, including bit rate and frame sizes. Most hardware decoders accept only streams that follow the CPB profile. One can encode with any bit rate or frame size and still have standard MPEG-1. However, this encoded MPEG-1 stream cannot be decoded and displayed with some of the existing decoder chips.

CPB is a limited set of sampling and bit-rate parameters designed to normalize computational complexity, buffer size, and memory bandwidth while still addressing the widest possible range of applications. CPB limits video images to 396 macroblocks (101,376 pixels) per frame if the frame rate is less than or equal to 25 frames/s, and 330 macroblocks (84,480 pixels) per frame if the frame rate is less or equal to 30 frames/s. Therefore, MPEG-1 video is typically coded at SIF dimensions (352x240x30 frames/s or 352x288x25 frames/s). The total maximum sampling rate is 3.8 Msamples/s including chroma. The coded video rate is limited to 1.862 Mbit/sec. In industrial practice, the bit-rate is the most often waived parameter of CPB, with rates as high as 6 Mbit/s in use.

Historically, CPB was an optimum point that barely allowed cost effective VLSI implementations using 0.8 μm technology. It also implied a nominal guarantee of interoperability for decoders and encoders. MPEG decoders, which were not capable of meeting SIF rates, were not considered to be true MPEG decoders. Currently, there are several ways of getting around CPB for SIF class applications and decoder. Thus, one should remember that CPB limits frames to 396 macroblocks (as in 352x288 SIF frames). Still within the constraints are sampling rates of 416x240x24 Hz, but this only aids NTSC (240 lines/field) displays. Deviating from 352 samples/line could throw off many decoder implementations that have limited horizontal sampling-rate conversion modes. From a practical perspective, many decoders are simply doubling the sampling rate from 352 to 704 samples/line via binary taps, which are simple shift-and-add operations. Future MPEG decoders will have arbitrary sample rate converters on-chip.

### 2.2.2 MPEG-2 Standard

The MPEG-2 standard for compressing both video and audio signals is similar to MPEG-1, and it is targeted toward more diverse applications such as all-digital transmission of broadcast TV, digital media storage, high-definition TV (HDTV), etc. The most significant enhancement over MPEG-1 is the addition of syntax for efficient coding of interlaced video signals, such as those that originated from electronic cameras. Several other more subtle enhancements (e.g., 10-bit DCT DC precision, nonlinear quantization, VLC tables, and improved-mismatch control) are included, which have a noticeable improvement on coding efficiency, even for progressive video. Other key features of MPEG-2 are scalable extensions, which permit the division of a continuous video signal into two or more coded bit streams, representing the video at different resolutions, picture quality (i.e., SNR), or picture rates. The MPEG-2 standard defines 4:2:0, 4:2:2, and 4:4:4 chroma sampling formats relative to the luminance. Figure 2.5 illustrates the difference between 4:2:0 and 4:2:0 sampling

formats. In 4:2:0 sampling, the chroma samples from two adjacent lines in a field are interpolated to produce a single chroma sample, which is spatially located half way between one of the original samples and the location of the same line but opposite field. The 4:2:0 sampling has many drawbacks, including significantly inferior vertical chroma resolution compared with a standard composite NTSC signal. In 4:2:2 sampling, the chroma is subsampled 2:1 horizontally but not vertically, with the chroma alignment with the luminance as shown in Figure 2.5. The 4:4:4 sampling has the same sampling to both chroma components, and the luminance with the same decomposition into interlaced fields, resulting in high-quality video images. MPEG-2 defines five different profiles as follows: simple profile (SP), main profile (MP), SNR scalable profile, spatially scalable profile, and high profile. The most

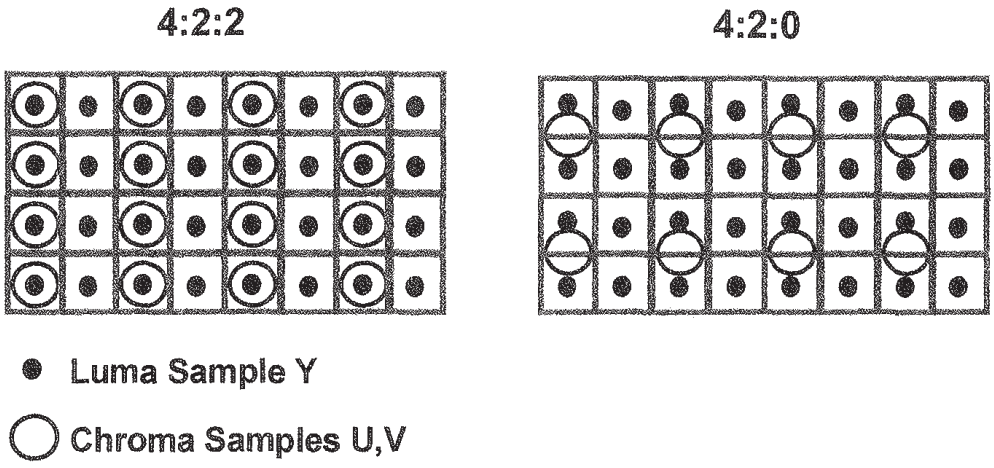



Figure 2.5 The 4:2:2 and 4:2:0 chroma components (U, V) sub-sampling relative to the luminance sample (Y).

**Table 2.3 MPEG-2 level bounds for the main profile.**

| MPEG Parameter | MP at HL | MP at H-14 | MP at ML | MP at LL |
|---|---|---|---|---|
| Samples/line | 1920 | 1440 | 720 | 352 |
| Lines/frame | 1152 | 1152 | 576 | 352 |
| Frame/sec | 60 | 60 | 30 | 30 |
| Luma Rate (samples/s) | 62,668,800 | 47,001,600 | 10,368,000 | 3,041,280 |
| Bit Rate (Mb/s) | 80 | 60 | 15 | 4 |
| VBV Buffer size (bits) | 9,781,248 | 7,340,032 | 1,835,008 | 475,136 |

useful profile is the MP, which consists of four levels: main level (ML), high-1440 level (H-14), main level (ML), and low level (LL). Table 2.3 provides the level bounds for the main profile.

The video buffering verifier (VBV), which is an idealized model of the MPEG-2 decoder, is used to constrain the instantaneous bit rate of the MPEG-2 encoder such that the average bit rate target is met without overflowing the decoder data buffer. All the bounds in Table 2.3 are upper bounds except for the VBV buffer size, which is a lower bound.

The use of B frames increases the computational complexity, bandwidth, delay, and picture buffer size of the encoded MPEG video since some of the macroblock modes require averaging between two macroblocks. At the worst case, memory bandwidth is increased an extra 16 MB/s (601 rate) for this extra prediction. An extra picture buffer is needed to store the future prediction reference. In addition, extra delay is introduced in the encoding process, since the frame used for backwards prediction needs to be transmitted to the decoder before the intermediate B-pictures can be decoded and displayed. Since the extra picture buffer pushes the decoder DRAM memory requirements past the magic 1-Mbyte threshold, several companies such as AT&T and General Instrument argued against the use of the B-frames. In 1991, General Instrument introduced DigiCipher I video coding, which is similar to MPEG-2 coding but uses smaller macroblock predictions with no B frames and with Dolby AC-1 audio. In 1994, General Instrument introduced the DigiCipher II specification, which supports both the full MPEG-2 video main profile syntax and DigiCipher II with the use of the Dolby AC-3 audio algorithm [7].

The transmission of MPEG 2 compressed video streams through HFC access networks using QAM modulators to the digital STB at the subscriber's home will be discussed in detail in Chapter 8.

### 2.2.3 MPEG and AC-3 Audio

The MPEG compression of audio signals, which is based on MUSICAM technology, is done using high-performance perceptual schemes [10]. These schemes specify a family of three audio coding schemes called layer-1 (MP1), layer-2 (MP2), and layer-3 (MP3) with increasing complexity and performance. Since the uncompressed digital stereo music in CD quality is about 1.5 Mb/s, one can achieve 4:1 compression or 384-kb/s for stereo signals using MP1, while still maintaining the original sound quality. Compression of 8:1 to 6:1 (192 to 256 kb/s) the stereo signals can be achieved with MP2, while MP3 provides 12:1 to 10:1 compression (112 kb/s to 128 kb/s) with essentially no loss in discernible quality. The audio coding scheme can be describes as "perceptual subband transform coding." The audio coding scheme makes use of the masking properties of the human ear to reduce the amount of data. The various "spectral" components of the audio signal are analyzed by calculating a modified DCT (MDCT) and applying a psychoacoustic model to estimate the audible noise threshold. The noise caused by the quantization process is distributed to various frequency bands in such a way that it is masked by the total signal, that is, it remains inaudible.

Another perceptual coding technology, which has been adopted by the advanced television systems committee (ATSC) as the standard audio for HDTV in the United States, is called AC-3 and was developed by Dolby Laboratories [11]. AC-3 technology can support up to eight channel configurations, ranging from mono to six discrete audio channels (left, center, right, left surround, right surround, and subwoofer). As mentioned before, DigiCipher II uses the Dolby Labs AC-3 compression algorithm for the audio source. The AC-3 encoder can support data rates ranging from 32-kb/s up to 640-kb/s at sample rates of 32 kHz, 44.1 kHz, and 48 kHz. For further details, see reference [12].

### 2.2.4 MPEG-4 Standard

Unlike MPEG-1/2, where the scope and technology were well defined when the project started, MPEG-4 was born to answer the emerging needs of various new applications ranging from interactive audiovisual services to remote monitoring and control [13]. Thus, the MPEG-4 goal is to provide a flexible and extensible standard for the convergence of interactive multimedia applications, which currently are not addressed by the existing standards.

To enable the content-based interactive functionality, the MPEG-4 video standard introduces the concept of video object planes (VOPs). It is assumed that each frame of an input video sequence is segmented into a number of arbitrarily shaped image regions (e.g., VOPs), where each of the regions may possibly cover particular image or video content of interest (i.e., describing physical objects or content within scenes). In contrast to the video source format used for MPEG-1 and MPEG-2, the video input to be coded by the MPEG-4 verification model is no longer considered a rectangular region. The input to be coded can be a VOP image region of arbitrary shape, and the shape and location of the region can vary from frame to frame. Successive VOPs belonging to the same physical object in a scene are referred to as video objects (VOs), that is, a sequence of VOPs of possibly arbitrary shape and position. The shape, motion, and texture information of the VOPs belonging to the same VO is encoded and transmitted or coded into a separate video object layer (VOL). In addition, the bit stream must include the relevant information to identify each of the VOLs, and how the various VOLs are composed at the receiver in order to reconstruct the entire original sequence. This allows the separate decoding of each VOP and the required flexible manipulation of the video sequence. Notice that the video source input assumed for the VOL structure either already exists in terms of separate entities (i.e., is generated with chroma-key technology) or is generated by means of on-line or off-line segmentation algorithms.

It is expected that MPEG-4 video coding will eventually support all the functionalities already provided by MPEG-1 and MPEG-2, including the provision to efficiently compress standard rectangular sized image sequences at varying levels of input formats, frame rates and bit rates. In addition, content-based functionality will be assisted.

### 2.2.5 Other Digital Video Standards

There are several non-MPEG video standards such as ITU-T Recommendation H.261 [14] and H.263 [15], which actually were developed before the MPEG standard. The first digital video standard was developed for visual applications by the CCITT group XV for integrated services digital network (ISDN) services, operating at px64 kb/s for p = 1,..,30. The H.261 standard has many elements in common with MPEG-1. The image dimensions were restricted to two sizes, common intermediate format (CIF) with 360x288 (Y) and 180x144 (*U* and *V*), and quarter CIF (QCIF). The H.261 syntax consists of four layers as follows: (A) picture layer, which has a picture header followed by 3 or 12 group of blocks (GOBs) with each GOB consisting of 33 macroblocks and a header, (B) GOBs layer, (C) macroblock layer, and (D) block layer, which is similar to the block layer of MPEG-1. The macroblock header provides information about the position of the macroblock relative to the position of the just coded macroblock. To simplify the compression standard, macroblocks between frames can be skipped if the motion-compensated prediction is sufficiently good.

Another similar non-MPEG video teleconferencing standard is H.263, which is based on H.261 standard for low bit-rate applications. The H.263 standard has four optional modes that enhance its functionality, including unrestricted motion vector mode and arithmetic coding mode instead of variable length codes, advanced prediction mode, and coding P and B frames as one unit.

## 2.3 Cable TV Frequency Plans

The frequency plan of cable TV channels in the U.S. is specified by the FCC. FCC rules in Part 76 specify frequencies to in accordance with the channel allocation plan set forth by with the Electronics Industry Association (EIA) [16]. The nominal channel spacing is 6-MHz, except for the 4-MHz frequency gap between channels 4 and 5. Table 2.4 shows the standard (STD), incrementally related carrier (IRC), and harmonically related carrier (HRC) cable TV frequency plans in the U.S.

In the STD cable TV frequency plan, which is similar to the frequency plan of terrestrially broadcasted TV channels, all the visual carriers except channels 5 and 6 are located 1.25 MHz above the lower edge of the channel boundary in 6-MHz multiples (1.25 + 6*N*) MHz. The visual carriers for channels 5 and 6 are located 0.75 MHz below the 6-MHz multiples. Notice that the visual carrier in the following channel groups 14–15, 25–41, and 43–53 have a 12.5-kHz frequency offset relative to the rest of the channels.

In the IRC frequency plan, all the visual carriers except for channels 42, 60, and 61 are located 1.2625 MHz above the lower edge of the channel boundary in 6-MHz multiples (1.2625 + 6*N*) MHz. Thus, the visual carrier in the IRC frequency plan has 12.5-kHz frequency offset compared with the STD frequency plan. This frequency offset was selected to

minimize interference in the 25-kHz radio channels, which are used for communications by airport control towers and aircraft navigation equipment based on FCC rulings.[2]

In the HRC frequency plan, all the visual carriers except for channels 60 and 61 are located essentially at the lower channel boundary in 6.0003-MHz multiples. As with the IRC plan, the 300-Hz incremental band increase was selected to minimize the interference in the 25-kHz radio channels used for aviation. The cable TV channel numbers are often designated according to electronic industry association (EIA) [16]. As we will see in Section 2.4,

**Table 2.4 Standard, HRC, and IRC cable TV frequency plans in the US. After Ref. [16].**

| Channel Designation | Visual Carrier Frequency (MHz) | | |
|---|---|---|---|
| EIA | STD | HRC | IRC |
| T7* | 7.0000 | | |
| T8* | 13.0000 | | |
| T9* | 19.0000 | | |
| T10* | 25.0000 | | |
| T11* | 31.0000 | | |
| T12* | 37.0000 | | |
| T13* | 43.0000 | | |
| 2 | 55.2500 | 54.0027 | 55.2625 |
| 3 | 61.2500 | 60.0030 | 61.2625 |
| 4 | 67. 2500 | 66.0033 | 67.2625 |
| 5 | 77. 2500 | N/A | |
| 6 | 83. 2500 | N/A | |
| 5 | N/A | 78.0039 | 79.2625 |
| 6 | N/A | 84.0042 | 85.2625 |
| 7 | 175. 2500 | 174.0087 | 175.2625 |
| 8 | 181. 2500 | 180.0090 | 181.2625 |
| 9 | 187. 2500 | 186.0093 | 187.2625 |
| 10 | 193. 2500 | 192.0096 | 193.2625 |
| 11 | 199. 2500 | 198.0099 | 199.2625 |
| 12 | 205. 2500 | 204.0102 | 205.2625 |
| 13 | 211. 2500 | 210.0105 | 211.2625 |
| 14 | 121. 2625 | 120.0060 | 121.2625 |
| 15 | 127. 2625 | 126.0063 | 127.2625 |
| 16 | 133. 2625 | 132.0066 | 133.2625 |
| 17 | 139. 2500 | 138.0069 | 139.2625 |
| 18 | 145. 2500 | 144.0072 | 145.2625 |

[2] See for example Code of Federal Regulations part 87.421 about the aviation channels and FCC rulings.

* Based on a conventional plan, which was originated in about 1960.

| | | | |
|---|---|---|---|
| 19 | 151. 2500 | 150.0075 | 151.2625 |
| 20 | 157. 2500 | 156.0078 | 157.2625 |
| 21 | 163. 2500 | 162.0081 | 163.2625 |
| 22 | 169. 2500 | 168.0084 | 169.2625 |
| 23 | 217. 2500 | 216.0108 | 217.2625 |
| 24 | 223. 2500 | 222.0111 | 223.2625 |
| 25 | 229. 2500 | 228.0114 | 229.2625 |
| 26 | 235.2625 | 234.0117 | 235.2625 |
| 27 | 241.2625 | 240.0120 | 241.2625 |
| 28 | 247.2625 | 246.0123 | 247.2625 |
| 29 | 253.2625 | 252.0126 | 253.2625 |
| 30 | 259. 2625 | 258.0129 | 259.2625 |
| 31 | 265. 2625 | 264.0132 | 265.2625 |
| 32 | 271. 2625 | 270.0135 | 271.2625 |
| 33 | 277. 2625 | 276.0138 | 277.2625 |
| 34 | 283. 2625 | 282.0141 | 283.2625 |
| 35 | 289. 2625 | 288.0144 | 289.2625 |
| 36 | 295. 2625 | 294.0147 | 295.2625 |
| 37 | 301. 2625 | 300.0150 | 301.2625 |
| 38 | 307. 2625 | 306.0153 | 307.2625 |
| 39 | 313. 2625 | 312.0156 | 313.2625 |
| 40 | 319. 2625 | 318.0159 | 319.2625 |
| 41 | 325. 2625 | 324.0162 | 325.2625 |
| 42 | 331. 2750 | 330.0165 | 331.2750 |
| 43 | 337. 2625 | 336.0168 | 337.2625 |
| 44 | 343. 2625 | 342.0171 | 343.2625 |
| 45 | 349. 2625 | 348.0174 | 349.2625 |
| 46 | 355. 2625 | 354.0177 | 355.2625 |
| 47 | 361. 2625 | 360.0180 | 361.2625 |
| 48 | 367. 2625 | 366.0183 | 367.2625 |
| 49 | 373. 2625 | 372.0186 | 373.2625 |
| 50 | 379. 2625 | 378.0189 | 379.2625 |
| 51 | 385. 2625 | 384.0192 | 385.2625 |
| 52 | 391. 2625 | 390.0195 | 391.2625 |
| 53 | 397. 2625 | 396.0198 | 397.2625 |
| 54 | 403.2500 | 402.0201 | 403.2625 |
| 55 | 409.2500 | 408.0204 | 409.2625 |
| 56 | 415.2500 | 414.0207 | 415.2625 |
| 57 | 421.2500 | 420.0210 | 421.2625 |
| 58 | 427.2500 | 426.0213 | 427.2625 |

| | | | |
|---|---|---|---|
| 59 | 433.2500 | 432.0216 | 433.2625 |
| 60 | 439.2500 | 438.0219 | 439.2625 |
| 61 | 445.2500 | 444.0222 | 445.2625 |
| 62 | 451.2500 | 450.0225 | 451.2625 |
| 63 | 457.2500 | 456.0228 | 457.2625 |
| 64 | 463.2500 | 462.0231 | 463.2625 |
| 65 | 469.2500 | 468.0234 | 469.2625 |
| 66 | 475.2500 | 474.0237 | 475.2625 |
| 67 | 481.2500 | 480.0240 | 481.2625 |
| 68 | 487.2500 | 486.0243 | 487.2625 |
| 69 | 493.2500 | 492.0246 | 493.2625 |
| 70 | 499.2500 | 498.0249 | 499.2625 |
| 71 | 505.2500 | 504.0252 | 505.2625 |
| 72 | 511.2500 | 510.0255 | 511.2625 |
| 73 | 517.2500 | 516.0258 | 517.2625 |
| 74 | 523.2500 | 522.0261 | 523.2625 |
| 75 | 529.2500 | 528.0264 | 529.2625 |
| 76 | 535.2500 | 534.0267 | 535.2625 |
| 77 | 541.2500 | 540.0270 | 541.2625 |
| 78 | 547.2500 | 546.0273 | 547.2625 |
| 79 | 553.2500 | 552.0276 | 553.2625 |
| 80 | 559.2500 | 558.0279 | 559.2625 |
| 81 | 565.2500 | 564.0282 | 565.2625 |
| 82 | 571.2500 | 570.0285 | 571.2625 |
| 83 | 577.2500 | 576.0288 | 577.2625 |
| 84 | 583.2500 | 582.0291 | 583.2625 |
| 85 | 589.2500 | 588.0294 | 589.2625 |
| 86 | 595.2500 | 594.0297 | 595.2625 |
| 87 | 601.2500 | 600.0300 | 601.2625 |
| 88 | 607.2500 | 606.0303 | 607.2625 |
| 89 | 613.2500 | 612.0306 | 613.2625 |
| 90 | 619.2500 | 618.0309 | 619.2625 |
| 91 | 625.2500 | 624.0312 | 625.2625 |
| 92 | 631.2500 | 630.0315 | 631.2625 |
| 93 | 637.2500 | 636.0318 | 637.2625 |
| 94 | 643.2500 | 642.0321 | 643.2625 |
| 95 | 91.2500 | 90.0045 | 91.2625 |
| 96 | 97.2500 | 96.0048 | 97.2625 |
| 97 | 103.2500 | 102.0051 | 103.2625 |
| 98 | 109.2750 | 108.0250 | 109.2750 |

| | | | |
|---|---|---|---|
| 99 | 115.2750 | 114.0250 | 115.2750 |
| 100 | 649.2500 | 648.0324 | 649.2625 |
| 101 | 655.2500 | 654.0327 | 655.2625 |
| 102 | 661.2500 | 660.0330 | 661.2625 |
| 103 | 667.2500 | 666.0333 | 667.2625 |
| 104 | 673.2500 | 672.0336 | 673.2625 |
| 105 | 679.2500 | 678.0339 | 679.2625 |
| 106 | 685.2500 | 684.0342 | 685.2625 |
| 107 | 691.2500 | 690.0345 | 691.2625 |
| 108 | 697.2500 | 696.0348 | 697.2625 |
| 109 | 703.2500 | 702.0351 | 703.2625 |
| 110 | 709.2500 | 708.0354 | 709.2625 |
| 111 | 715.2500 | 714.0357 | 715.2625 |
| 112 | 721.2500 | 720.0360 | 721.2625 |
| 113 | 727.2500 | 726.0363 | 727.2625 |
| 114 | 733.2500 | 732.0366 | 733.2625 |
| 115 | 739.2500 | 738.0369 | 739.2625 |
| 116 | 745.2500 | 744.0372 | 745.2625 |
| 117 | 751.2500 | 750.0375 | 751.2625 |
| 118 | 757.2500 | 756.0378 | 757.2625 |
| 119 | 763.2500 | 762.0381 | 763.2625 |
| 120 | 769.2500 | 768.0384 | 769.2625 |
| 121 | 775.2500 | 774.0387 | 775.2625 |
| 122 | 781.2500 | 780.0390 | 781.2625 |
| 123 | 787.2500 | 786.0393 | 787.2625 |
| 124 | 793.2500 | 792.0396 | 793.2625 |
| 125 | 799.2500 | 798.0399 | 799.2625 |
| 126 | 805.2500 | 804.0402 | 805.2625 |
| 127 | 811.2500 | 810.0405 | 811.2625 |
| 128 | 817.2500 | 816.0408 | 817.2625 |
| 129 | 823.2500 | 822.0411 | 823.2625 |
| 130 | 829.2500 | 828.0414 | 829.2625 |
| 131 | 835.2500 | 834.0417 | 835.2625 |
| 132 | 841.2500 | 840.0420 | 841.2625 |
| 133 | 847.2500 | 846.0423 | 847.2625 |
| 134 | 853.2500 | 852.0426 | 853.2625 |
| 135 | 859.2500 | 858.0429 | 859.2625 |
| 136 | 865.2500 | 864.0432 | 865.2625 |
| 137 | 871.2500 | 870.0435 | 871.2625 |
| 138 | 877.2500 | 876.0438 | 877.2625 |

| | | | |
|---|---|---|---|
| 139 | 883.2500 | 882.0441 | 883.2625 |
| 140 | 889.2500 | 888.0444 | 889.2625 |
| 141 | 895.2500 | 894.0447 | 895.2625 |
| 142 | 901.2500 | 900.0450 | 901.2625 |
| 143 | 907.2500 | 906.0453 | 907.2625 |
| 144 | 913.2500 | 912.0456 | 913.2625 |
| 145 | 919.2500 | 918.0459 | 919.2625 |
| 146 | 925.2500 | 924.0462 | 925.2625 |
| 147 | 931.2500 | 930.0465 | 931.2625 |
| 148 | 937.2500 | 936.0468 | 937.2625 |
| 149 | 943.2500 | 942.0471 | 943.2625 |
| 150 | 949.2500 | 948.0474 | 949.2625 |
| 151 | 955.2500 | 954.0477 | 955.2625 |
| 152 | 961.2500 | 960.0480 | 961.2625 |
| 153 | 967.2500 | 966.0483 | 967.2625 |
| 154 | 973.2500 | 972.0486 | 973.2625 |
| 155 | 979.2500 | 978.0489 | 979.2625 |
| 156 | 985.2500 | 984.0492 | 985.2625 |
| 157 | 991.2500 | 990.0495 | 991.2625 |
| 158 | 997.2500 | 996.0498 | 997.2625 |

the nonlinear distortions in both the HRC and IRC frequency plans are reduced compared with the STD plan.

## 2.4 Coaxial Cable TV Components and Systems

As we learned in Chapter 1, the basic architecture of HFC networks to deliver multichannel analog video/audio signals from the fiber node to the subscriber's home consists of coaxial cables, cascaded amplifiers, and taps. The transmission characteristics of these components such as transmission loss, frequency response, amplifier gain, and nonlinear distortions need to be considered when designing such a network. Furthermore, when RF digital channels are added, linear distortions such as group delay and amplitude variation within each 6-MHz channel also must be considered. The following section provides a brief overview of these properties.

### 2.4.1 Coaxial Cable

Network distribution coaxial cables typically consist of a copper-clad aluminum wire, which is the inner conductor, an insulating layer such as foam polyethylene, a solid aluminum shield, which serves as the outer conductor, and a PVC jacket. Subscriber drop cables are manufactured with a copper-clad steel center conductor and a combination of aluminum

braid and aluminum-polypropylene-aluminum (APA) tap shield. For specialty applications such as plenum installations, the coaxial cable jacket is often made with polytretrafluoro-ethylene (PTFE) materials. The transmission characteristic of coaxial cables depends mainly on the transmission frequency and on the diameter of the cable. The primary loss mechanisms in coaxial cables are the frequency and temperature dependence of the inner conductor and the dielectric outer conductor losses. In particular, the transmission loss of the inner conductor in coaxial cables at RF frequencies is influenced by the so-called *skin effect* [17]. When DC current flows through a conductor, the current is uniformly distributed throughout the cross-section of the conductor. As the RF frequency is increased, the current tends to crowd around the conductor surface, which reduces the effective current cross-section, resulting in a given conductor having higher impedance at higher RF frequencies.

There are three different types of coaxial cables that are used in the distribution system: trunk, feeder, and drop cables. The transmission loss of trunk cables, which have a typical diameter from 1/2" to 1", increases from 0.89 dB at 50 MHz to 3.97 dB at 750 MHz per 100 m for 1" cable. The feeder cable has typically a smaller diameter than the trunk cable, and is used to connect between the line-extender amplifiers to the tap. The drop cable typically has

**Table 2.5 Maximum loss for drop cable (dB/100 ft at 68F) with four different cable diameters as a function of frequency. To obtain loss in dB/100 m, multiply by 3.281. After Ref. [16].**

| Frequency (MHz) | 59 Series Foam | 6 Series Foam | 7 Series Foam | 11 Series Foam |
|---|---|---|---|---|
| 5 | 0.86 | 0.58 | 0.47 | 0.38 |
| 30 | 1.51 | 1.18 | 0.92 | 0.71 |
| 40 | 1.74 | 1.37 | 1.06 | 0.82 |
| 50 | 1.95 | 1.53 | 1.19 | 0.92 |
| 110 | 2.82 | 2.24 | 1.73 | 1.36 |
| 174 | 3.47 | 2.75 | 2.14 | 1.72 |
| 220 | 3.88 | 3.11 | 2.41 | 1.96 |
| 300 | 4.45 | 3.55 | 2.82 | 2.25 |
| 350 | 4.80 | 3.85 | 3.05 | 2.42 |
| 400 | 5.10 | 4.15 | 3.27 | 2.60 |
| 450 | 5.40 | 4.40 | 3.46 | 2.75 |
| 550 | 5.95 | 4.90 | 3.85 | 3.04 |
| 600 | 6.20 | 5.10 | 4.05 | 3.18 |
| 750 | 6.97 | 5.65 | 4.57 | 3.65 |
| 865 | 7.52 | 6.10 | 4.93 | 3.98 |
| 1000 | 8.12 | 6.55 | 5.32 | 4.35 |

a smaller diameter than the feeder cable, and is used between the tap and the subscriber home terminal. Table 2.5 shows the maximum drop cable loss (dB/100 ft.) at 68°F as a function of RF frequency for the following nominal cable diameters: 0.240" (59 series foam), 0.272" (6 series foam), 0.318" (7 series foam), and 0.395" (11 series foam). Notice the nearly 10 times increase in the cable loss from 5 MHz to 1 GHz for 59 series foam cable. The coaxial cable losses (in dB) are reasonably proportional to the square root of the frequency [17]. Cable plants are usually specified to operate over a wide temperature range from –40°C to +70°C. In general, the cable loss slowly increases linearly with temperature, since the cable attenuation increases with temperature.
A useful parameter to remember is the so-called cable loss ratio (CLR) given by

$$CLR = \sqrt{\frac{f_1}{f_2}} \tag{2.2}$$

where $f_1$ and $f_2$ are two different RF frequencies. For example, approximate the cable loss at 55 MHz when the cable loss at 450 MHz is 20 dB. The cable loss (CL) at 55 MHz is given by $CL = 20 \cdot (55/450)^{1/2} = 6.99$ dB.

### 2.4.2 RF Amplifiers

There are also three different types of cable TV RF amplifiers, depending on their location in the coaxial portion of the HFC network. Figure 2.6 shows the respective locations of trunk, bridger, and line-extender amplifiers in the traditional tree-and-branch architecture of the coaxial portion of the HFC network. Trunk amplifiers, which are typically spaced 20–22 dB from one another, are moderate-gain amplifiers with a typical output of 30 to 36 dBmV that are used to provide high CNR with low nonlinear distortions (< –80 dBc), particularly at the high-frequency cable TV channels (> 300 MHz). Feeder amplifiers are used not only for downstream delivery of analog and digital video channels, but also to split the transmitted signals to up to four feeder cables as shown in Figure 2.7. The output power of a bridger amplifier is typically in the range of 40–50 dBmV, which is about 12 dB higher than that of trunk amplifiers. However, higher nonlinear distortions are present at the output of bridger and line extender amplifiers. To reduce the effect of nonlinear distortions on the transmitted video signals and to maintain the flatness of the entire band, a maximum of two to four line- extender amplifiers are used, depending on the number of taps between the line extender amplifiers. Line-extender amplifiers, which are typically spaced between 120 m to 350 m, are used in the vicinity of the subscriber's home. For two-way cable TV systems, the downstream or forward video channels in the United States are placed between 52 MHz and 860 MHz, while the return-path or upstream band is between 5 MHz and

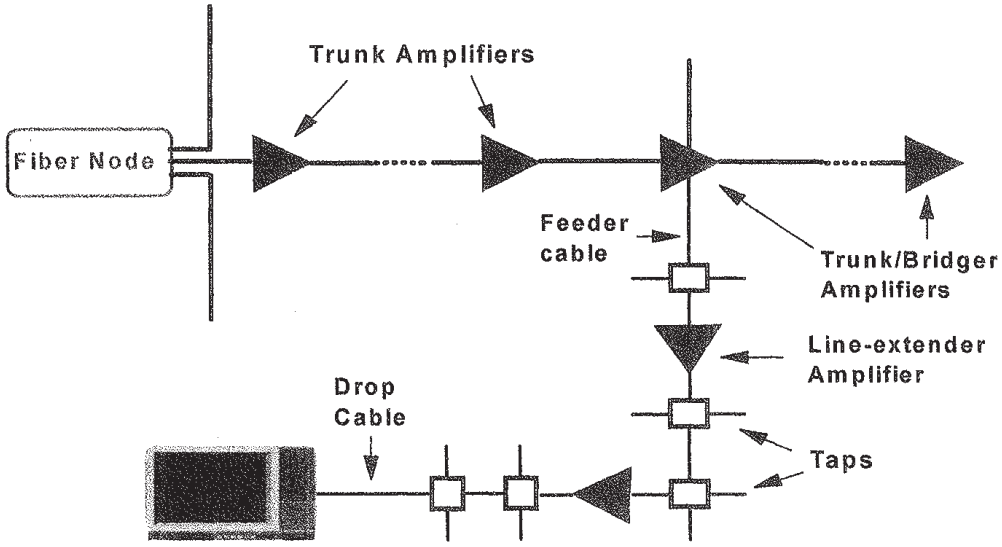



Figure 2.6 Tree-and-branch architecture of the coaxial portion of the HFC networks showing the different types of cables and amplifiers that are used.

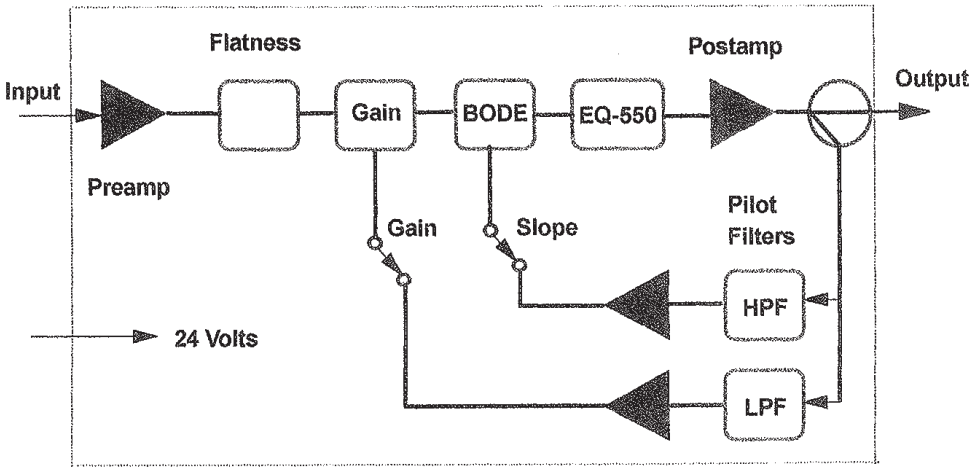



Figure 2.7 Simplified block diagram of trunk amplifier, showing the automatic gain and slope controls and equalizer up to 550 MHz.

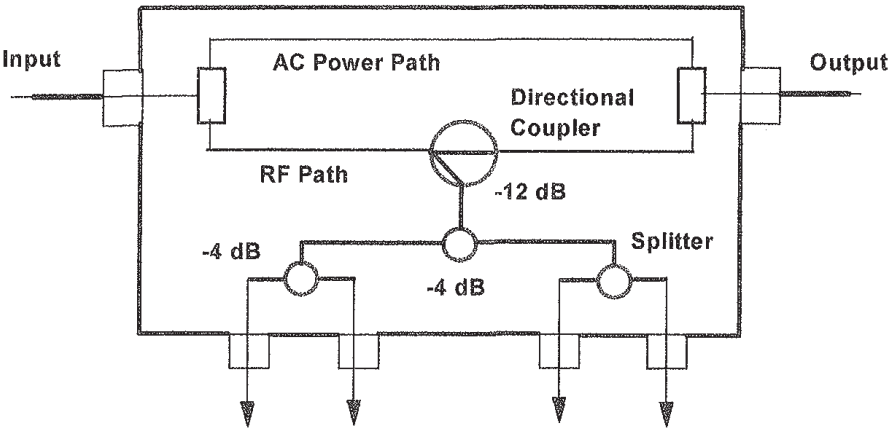



Figure 2.8 Simplified block diagram of four-way 20-dB tap.

42 MHz.[3] This partition is done using a special filter known as *diplex* filter, which has a typical isolation between these bands better than 60 dB. The diplexer filter is three-port device, with high, low, and common ports. The common-to-low port, which is marked as "L," is essentially a low-pass filter that allows the return-path signals to be transmitted. The common-to-high port, which is marked as "H," is essentially a high-pass filter that allows the forward channels to be transmitted. In a typical two-way trunk or bridger amplifier, the forward signals pass through the H ports, while the upstream signals are pass through the L ports.

A simplified block diagram of a trunk amplifier is shown in Figure 2.7. Since coaxial cables have a strong frequency-dependent loss (see Table 2.5 for example), the amplitude of the transmitted video channels must be equalized in order to maintain flatness across the transmitted RF spectrum. Forward equalizers are designed to compensate for fixed lengths of coaxial cables. By introducing additional attenuation at the lower frequencies, the equalizer allows the trunk amplifier to maintain a known slope across its transmission band. In addition, some trunk amplifiers are equipped with Bode equalizers to compensate for changes in the cable loss caused by temperature variations.

Trunk amplifiers typically use automatic gain control (AGC) and/or automatic slope control (ASC) circuitry. Typical gain and slope control ranges are 6 to 10 dB up to 750 MHz. The AGC and ASC modules in the trunk amplifier detects a sampled signal of standard pilot channels at the amplifier output, which is used to create the appropriate voltage to control the gain and/or slope of the amplifier. The standard pilot frequencies vary among the differ-

[3] Prior to 1994, the upstream band ended at 30 MHz. Other countries have different split frequencies such as Australia 65/85; Japan and New Zealand 55/70; India and Eastern Europe 30/48; Western Europe, Ireland, and United Kingdom 65/85.

ent manufacturers. All cable TV amplifiers use some variation of push-pull circuitry to minimize second-order distortions. Both Feedforward and power-doubling technology are used for improved distortion performance.

### 2.4.3 Taps

Figure 2.5 shows a simplified block diagram of a four-way 20-dB tap. Taps are used to split the transmitted signals to drop cables as shown in Figure 2.8. A typical tap consists of a RF directional coupler and power splitters. The directional coupler diverts a specific amount of the input signal power, while the power splitter splits that signal to typically two, four, or eight subscriber ports. The power loss between the input and the output ports is called is called *insertion loss*, while the power loss between different output ports is called *isolation loss*. Tap insertion loss is nominally independent on frequency or temperature. High isolation loss is essential for two-way cable systems in order to prevent the upstream signals from one customer to leak into the forward signals of another customer. Typical isolation in a tap is about 20 dB for both forward and return-path bands. Taps are marked by their tap value, which is the power ratio of the signal at the tap to the input signal. Common values are from 4 dB to 35 dB, in 3-dB steps. Taps are typically housed in a die-cast aluminum alloy housing for aerial or ground mounting with a weather seal gasket to prevent entry of moisture and dust.

## 2.5 Multichannel Coaxial Cable TV Systems

### 2.5.1 CNR of a Single and Cascaded Amplifiers

One of the most important parameters that characterize the transmission performance of a cable TV system is the carrier-to-noise ratio (CNR). The CNR at the output of a single RF amplifier is given by

$$CNR(dB) = P_{out} / k_B TB + 59.16 - F - G \tag{2.3}$$

where $P_{out}$ is the output power from the amplifier, $k_B$ is Boltzmann's constant ($1.38 \times 10^{-23}$ joule/K), $T$ is the effective Kelvin temperature of the amplifier, $B$ is the signal noise bandwidth (4-MHz), and $F$ and $G$ are the amplifier noise figure and gain in dB. The value −59.16 dBmV is the 75-ohm impedance room temperature thermal noise in 4-MHz bandwidth. Typical noise figures for trunk amplifiers range from 7 to 10 dB at an input signal level of +10 dBmV and gain of 20 dB.

Now, consider a cable TV system with a chain of $N$ unlike cascaded amplifiers but with the noise bandwidth, where the $N_{th}$ amplifier has a noise figure $F_N$ and amplification $G_N$ as

shown in Figure 2.9. Then, the overall system noise figure F (expressed as ordinary power ratio) is given by

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \ldots + \frac{F_N - 1}{G_1 G_2 \cdots G_{N-1}}$$

In the simple case of $N$ identical RF amplifiers, Equation (2.4) is reduced to $N{\cdot}F$ where $F$ is the noise figure of each amplifier. The overall noise figure in dB is given by

$$NF_N(dB) = 10 \cdot \log(N \cdot F) \tag{2.5}$$

Then, the overall system CNR is given by

$$CNR_N = CNR - 10 \cdot \log(N)$$
(2.6)

For example, if a cable TV system has four cascaded amplifiers with CNR of 56 dB for a single amplifier, then the overall CNR after the fourth amplifier is 50 dB.

The overall system $CNR_N$ for a system of dislike amplifiers as described by Equation (2.4) is simply given by

$$CNR_N(dB) = -10 \cdot \log\left[10^{-CNR_1/10} + 10^{-CNR_2/10} + \ldots + 10^{-CNR_N/10}\right] \tag{2.7}$$

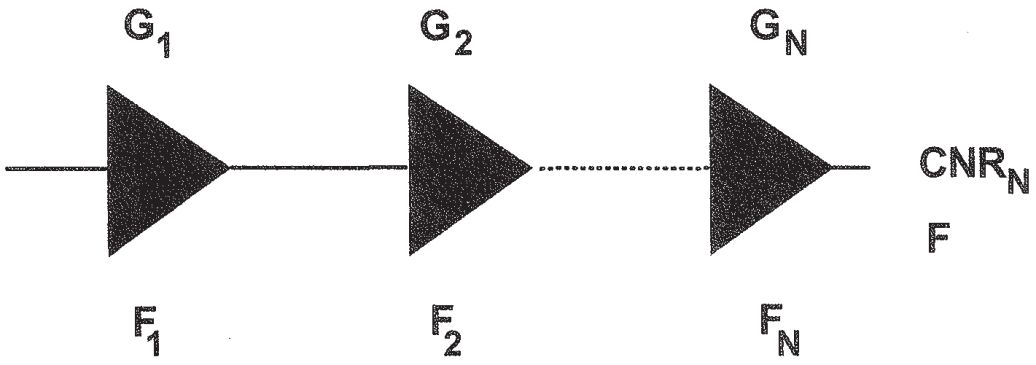



Figure 2.9 N cascaded amplifier configuration for calculating the overall system CNR and noise figure (F).

### 2.5.2 Nonlinear Distortions: CSO, CTB, and XMOD

When multiple band-limited signals such as AM-VSB modulated video signals are transmitted through a nonlinear component or system such as an RF amplifier or a laser transmitter, various nonlinear distortions (NLD) will be generated. The most important and strongest second-order intermodulation ($IM_2$) products are given by $A \pm B$, where $A$ and $B$ are two arbitrary RF frequencies, which are always 6-dB higher than the $2A$ or $2B$ distortion products. In a cable TV system, the summation of all possible different $IM_2$ products that are present in a particular channel is commonly called composite second-order (CSO) distortion. The CSO distortion beats in the standard cable TV frequency plan are 1.25 MHz and 0.75 MHz above and below the luminance carrier of a particular channel. The CSO distortion beats at ± 0.75 MHz is a result of channels 5 and 6, which are offset 2 MHz from the 6 MHz multiples, mixing with the other channels in the standard frequency plan.

In the IRC frequency plan, the CSO distortion beats are only located at ± 1.25 MHz above or below the luminance carrier of a particular channel. Due to the frequency offset of the CSO distortion from the luminance carrier, it appears as slowly moving diagonal stripes on a TV picture. In the HRC frequency plan, the CSO and CTB distortion beats are frequency-coincident with the visual carrier for coherent channel cable TV systems. As we will see later, this is one of the main advantages of the HRC frequency plan compared with the other frequency plans.

For a typical cable TV system with $N$ dislike amplifiers, the overall CSO is given by

$$CSO_N = -15 \cdot \log_{10}\left[10^{-CSO_1/15} + 10^{-CSO_2/15} + \ldots + 10^{-CSO_N/15}\right] \tag{2.8}$$

The dominant third-order intermodulation ($IM_3$) products are $A+B-C$, $A-B+C$, and $A-B-C$ where $A < B < C$, and $A$, $B$, and $C$ are three arbitrary RF frequencies. The weakest third-order intermodulation product is the 3A product, which is 15.56 dB weaker than the ABC distortion products. The $2A \pm B$ and $A \pm 2B$ products are 9.54 dB weaker than the $ABC$ products. In a cable TV system, the summation of all possible different third-order intermodulation products that fall onto a particular channel is commonly called composite triple beat (CTB) distortion. Thus, the CTB distortion is deduced by measuring the two-tone $IM_3$ (i.e., $2A \pm B$), counting the number of $IM_3$ products that fall in a particular channel, and adding 6 dB.

For equally spaced luminance carriers, the estimated number of CTB beats in any channel is given by

$$N_{CTB} = \frac{(N-1)^2}{4} + \frac{(N-M)\cdot(M-1)}{2} - \frac{N}{2} \tag{2.9}$$

where $N$ is the total number of channels and $M$ is the number of channels being measured. For $N>>1$ and in the middle of the band, Equation (2.9) reduces to $3N^2/8$, while at the edge of the band, Equation (2.9) reduces to $N^2/4$. Notice that the number of CTB beats at the band edge is 2/3 of the number of CTB beats in the middle of the band, independent of the number of channels. The overall CTB distortion (in dB) is 10·log ($N_{CTB}$). In Equation (2.9), it is also assumed that the luminance carriers are not phase locked. If the carriers are phase locked, then the CTB beats add coherently, and the overall CTB is 20·log ($N_{CTB}$).

For a typical cable TV system with $N$ different amplifiers, the overall CTB is given by

$$CTB_N(dB) = -20 \cdot \log\left[10^{-CTB_1/20} + 10^{-CTB_2/20} + \ldots + 10^{-CTB_N/20}\right] \tag{2.10}$$

Thus, a 1 dB change in the amplifier output will change the CTB distortion ratio by 2-dB (the beat product itself changes by 3 dB). For every double in the number of amplifiers with identical CTB distortion, the overall CTB ratio degrades by 6 dB.

Cross modulation (XMOD) distortion is another type of cable TV distortion. It occurs when a group of video carriers are modulating other video carriers in a multichannel video system. The origin of XMOD distortions is the similar to the CSO/CTB distortions. These distortions are usually generated when an RF coaxial amplifier or other active device is overloaded or driven beyond its compression point such that its gain becomes nonlinear. In older cable TV networks, XMOD distortions often produced picture interference that resembled a windshield wiper effect. However, in current cable TV networks with a large number of AM-VSB channels and high operating levels, the effect of XMOD distortions are typically masked by CSO/CTB distortions, and thus do not impose a serious problem to the cable TV operators.

The relative magnitude of the CSO, CTB, and XMOD distortions for a given channel is also an important consideration. Unfortunately, there is no simple answer. However, as we will learn in Chapters 3 and 4, the relative magnitude of the CSO, CTB, and XMOD distortions in a given fiber link depend primarily on the type of the laser transmitter used and the linearization techniques used to suppress these distortions.

### 2.5.3 Multipath Reflections (Echoes) and Group Delay

Multiple multipath reflections in the coaxial portion of the HFC network can severely degrade the propagating analog or digital signals before they reach the subscriber home. Multipath reflections occur when two or more propagation paths exist between the transmitter and receiving sites. The various reflections relative to the directly transmitted signal as measured at the receiver are called *echoes*, which are characterized by their amplitude attenuation and time delay. The echoes result from reflections off man-made or natural structures, repeaters, or the use of multiple transmitters.

The effect of multipath echoes on analog NTSC signals is quite different from that on digital signals. Let us first focus on the effect on the transmitted analog NTSC signals. For NTSC signal, a ghostlike image is horizontally displaced from the main image by an amount proportional to the time delay of the reflected signal. The effect of multipath reflections can be seen directly on an analog TV screen as ghosting, horizontally displaced from the main image. This can be used to provide an estimate of both the time delay and amplitude of the multipath echoes. The duration of each horizontal TV line is about 63.5 μs, with about 11 μs being used for the horizontal sync and blanking interval, providing about 52.5 μs for the actual video image. Because of TV set over-scan and simplicity, let us round off the video signal duration to 50 μs. Assuming a single dominant echo with a time delay less than 50 μs, its time delay can be estimated by multiplying the percentage of TV screen displaced by 50 μs. For example, if a ghost is displaced 25% from the main signal, the multipath delay is approximately 12.5 μs.

The amplitude of the multipath can also be determined by comparing the amplitude of the ghost to the amplitude of the original signal, using a video waveform monitor with the appropriate triggering to view only the VITS signals. The magnitude of the multipath echo can be determined as 20·log (% of amplitude/100). For example, if multipath reflection generates a ghost that is 25% the amplitude of the original waveform, then its amplitude is –12 dBc, or 12 dB below the desired signal. This method can be used effectively to measure the amplitude of the echoes for one or more of the transmitted analog channels.

Multiple echo degradation is not seen in the digitally demodulated picture until some "threshold" level of the digital signal is reached, resulting in a loss of the receiver synchronization. Uncorrected multipath echoes introduce intersymbol interference (ISI) that results in a closure of the eye pattern, making the signal more susceptible to decoding errors. Using an adaptive equalizer in the digital receiver can minimize the digital degradation effect caused by multiple echoes. However, strong multiple echoes can cause the digital receiver to lose synchronization. Outside the time range of the receiver's adaptive equalizer, the effect of multiple echoes is perceived as additional noise and causes degradation to the received SNR. The topic of how an adaptive equalizer in a digital receiver works to mitigate the effect of multiple echoes will be discussed in detail in Chapter 7.

If the time delay of the multipath echoes is beyond the range of the adaptive equalizer in the digital receiver, then the adaptive equalizer is not able to combat the generated ISI. In this case, an ordinary spectrum analyzer can be used to characterize multipath on digital signals. Figure 2.10 shows, for example, the simulated spectrum of a 64-QAM channel in a 6-MHz band. In the presence of Gaussian noise, the spectrum of the 64-QAM signal is essentially flat across most of the symbol rate bandwidth. If, however, multipath reflections are present, constructive and destructive interference of the reflected paths with the direct path will cause ripples in the otherwise flat spectrum. The time delay for a single multipath echo can be estimated by taking the inverse of the measured frequency spacing of the ripples on a spectrum analyzer. The ripple spacing is measured from peak-to-peak or from null-to-null.

It turns out that for a single echo, its magnitude (in dB) with respect to the transmitted signal can be estimated according to the following approximation:

$$EM(dB) = 20 \cdot \log\left[\frac{10^{\Delta_{pp}/20} - 1}{10^{\Delta_{pp}/20} + 1}\right] \tag{2.11}$$

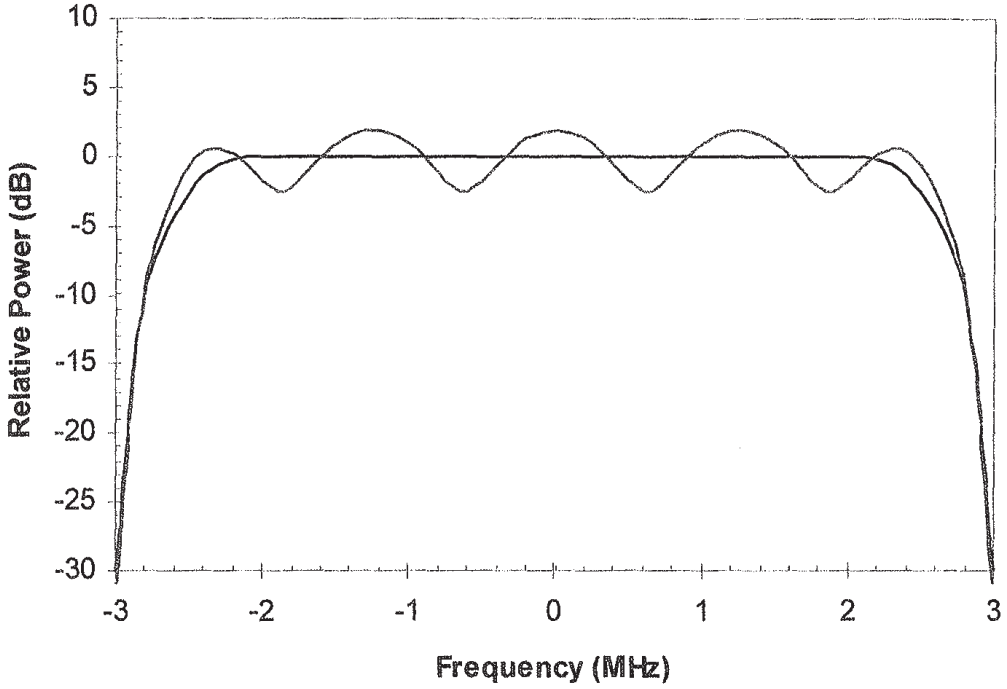



Figure 2.10 Simulated effect of a single echo with magnitude of −12 dBc and time delay of 0.8 μs relative to the main signal on the RF spectrum of a 64-QAM signal. After Ref. [9].

**Table 2.6 The corresponding echo profile for Figure 2.10.**

| Echo # | Time Delay | Magnitude | Phase |
|---|---|---|---|
| 1 | 0.2-μs | −11 dBc | 180° |
| 2 | 0.4-μs | −14 dBc | 180° |
| 3 | 0.8-μs | −17 dBc | 180° |
| 4 | 1.2-μs | −23 dBc | 180° |
| 5 | 2.5-μs | −32 dBc | 180° |

where $\Delta_{pp}$ is the measured signal peak-to-valley amplitude variation (in dB) on a spectrum analyzer. Figure 2.10 shows also, the simulated effect on the spectrum of a 64-QAM signal caused by a single multipath echo with a time delay of 0.8 μs and with a magnitude of 12 dB below the transmitted signal. Note that the frequency spacing between adjacent ripples is 1.25 MHz and the peak to valley amplitude variation of 4.5 dB.

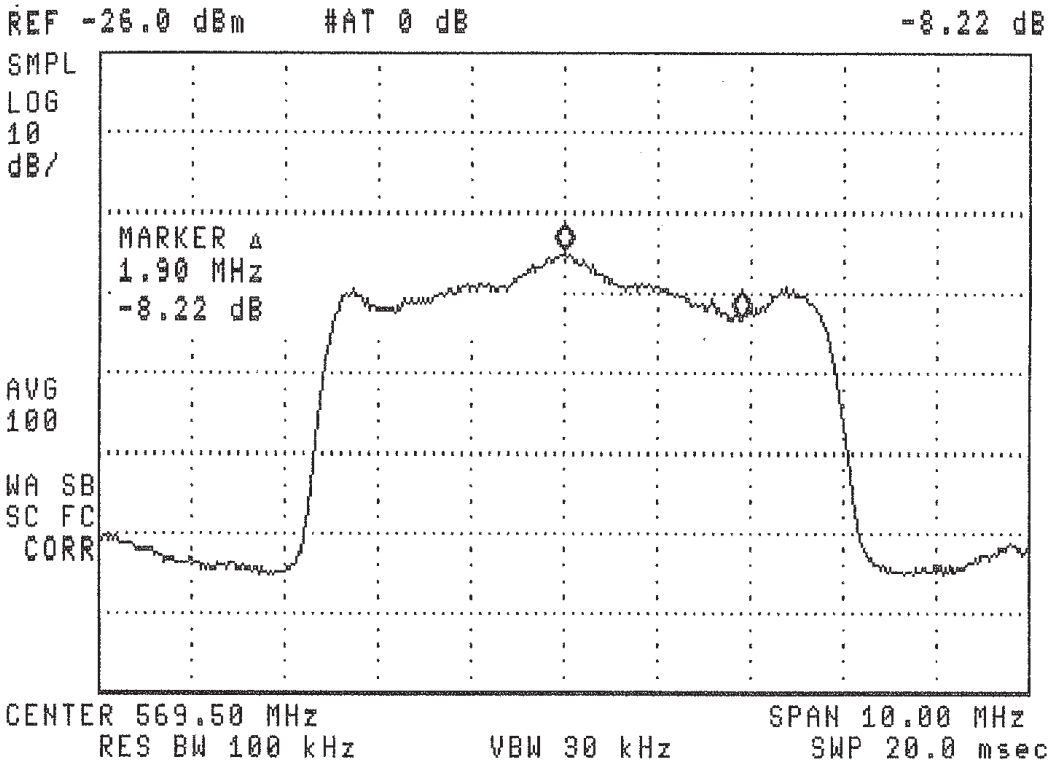



Figure 2.11 Measured 256-QAM spectrum at center frequency of 569.5 MHz in the presence of five-multipath echoes, according to Table 2.6.

When multiple multipath echoes degrade a QAM signal, the echo analysis using a spectrum analyzer becomes more difficult. This is because the ripple pattern is a complex superposition of the different echoes' time delay and magnitude, requiring a Fourier analysis. Figure 2.11 shows, for example, the measured spectrum of a 256-QAM signal at 569.5 MHz in the presence of five multipath echoes, where their time delay, amplitude, and phase are shown in Table 2.6. The QAM receiver in this measurement had an adaptive equalizer with 8 symbol-spaced feed-forward (FFE) taps and 24 symbol-spaced decision-feedback (DFE) taps. One of the FFE taps is used as the center tap, leaving 7 FFE taps for leading echoes and 24 DFE taps for lagging echoes. The total theoretical range of the adaptive equalizer is from −1.3 μs to +4.48 μs (symbol duration ≅ 186.5 ns). Notice the large 8.22-dB amplitude ripple (i.e., peak-to-valley variation) in the spectrum of the 256-QAM signal due to the pres-

ence of multiple echoes. In fact, the measured 256-QAM BER had increased from $10^{-8}$ to $1.3 \cdot 10^{-3}$ at a corrected CNR of 31.2 dB in the presence of these multipath echoes.

For multipath echoes with a short time delay (less than 0.2 μs), less than one ripple may be created in the 6-MHz spectrum. This means that multipath echoes with short-time delay, which are caused by relatively close reflections, have the effect of attenuating or amplifying entire channels, that is, it creates ripples across a wide frequency band spanning many adjacent channels. Consequently, it may actually be beneficial to those channels that are increased in level, but harmful to those channels that are attenuated.

Another important related parameter is the so-called *group delay*. The group delay is related to the frequency-dependent phase of various RF components in a cable TV plant such as diplex filters, equalizers, and impedance matching transformers. Mathematically, the complex multipath transfer function for a transmitted analog or digital channel can be described as

$$H(\omega) = |H(\omega)| \cdot \exp[i\phi(\omega)] \tag{2.12}$$

where $A(\omega) = -20 \cdot \log[\,|H(\omega)|\,]$ is the signal attenuation function (in dB), and $\phi(\omega)$ is the phase function of the propagating signal. The delay distortion or group delay is basically the instantaneous phase slope with respect to frequency, and it is given by

$$GD(\omega) = -\frac{d\phi(\omega)}{d\omega} = -\frac{1}{2\pi} \cdot \frac{d\phi}{df} \tag{2.13}$$

where $f$ is the frequency in Hz. A critical parameter for robust transmission is the slope of the group delay across the transmission channel band. If the group delay is constant within the desired channel, there is no group-delay distortion. The term group-delay variation (GDV) is referred to simply as the maximum change (peak-to-valley) in the group delay across a given channel. For a cable TV system with $N$ cascaded amplifiers, the total GDV across a given channel would be the sum of GDV from each amplifier in the link. In the small echo regime (for a single dominant echo), the maximum in-band GDV (in μs) can be approximated as

$$GDV(\mu s) = 2\tau \cdot 10^{-r/20} \tag{2.14}$$

where $r$ is the echo magnitude (in dB) and $\tau$ is the echo delay (in μs) relative to the transmitted signal. Coaxial components such as amplifiers and taps typically generate GDV due to nonflat amplitude variations at different frequencies. For example, the GDV should be less than 0.2 μs per 1 MHz of channel bandwidth, which corresponds to an amplitude variation of 0.5 dB for a single echo with 0.1-μs time delay. The DOCSIS 1.1 specifications, for example, assume that the GDV of a typical cable TV network is 75 ns for a 6-MHz down-

stream channel, and 200 ns/MHz for an upstream channel. Even without the presence of multipath echoes, the use of various filters in the coaxial amplifiers typically generates GDV because of their nonflat frequency response.

There are various discrete multipath reflection models such as IEEE802.14 [8] and DOCSIS 1.0 [9] for both forward and return-path channels in a cable TV plant, which have been recently discussed. While the DOCSIS 1.0 model assumes a single dominant echo, the IEEE802 model breaks the echo power into multiple echoes within the time-delay range. Table 2.7 summarizes the maximum echo power and time-delay bounds for both downstream and upstream channels according to the DOCSIS model, assuming a single dominant echo.

**Table 2.7 Multipath echo model according to DOCSIS 1.0 standard.**

| Echo Time Delay | Echo Magnitude (downstream) | Echo Magnitude (upstream) |
|---|---|---|
| 0 to 0.5 μs | −10 dBc | −10 dBc |
| ≤ 1.0 μs | −15 dBc | −20 dBc |
| ≤ 1.5 μs | −20 dBc | −30 dBc |
| > 1.5 μs | −30 dBc | −30 dBc |

Note that placing the echo power at the maximum of the delay range produces the largest GDV, but not necessarily the worst effect of the transmitted signal. Breaking the echo power into multiple echoes within the time-delay range lowers the GDV, but actually has a worse effect on the transmitted signal due to higher peak-to-rms ratio for combining multiple echoes compared with a single echo.

### 2.5.4 AM Hum Modulation

As we learned in Section 2.1, the vertical scan frequency of a TV image was originally selected to match the AC power frequency (60 Hz). Modulation distortion at 60 Hz or harmonics of the fundamental powerline frequency, which is called "hum," is the amplitude distortion of the transmitted signals caused by the modulation of the signal by the power source. The AM hum modulation is defined as the percentage of the peak-to-peak interference compared with the rms value of the sync peak level of the visual RF signal [2]. According to FCC regulations, Part 76, Section 76.605, the magnitude of AM hum modulation at the subscriber's terminal is required not to exceed 3% peak-to-peak of the visual signal level [4]. The major sources of power line hum (60 Hz and 120 Hz) are amplifiers with defective power supplies and overloaded system power supplies. However, other cable system components, even passive devices, can introduce hum modulation distortion under certain conditions. For additional discussion of AM hum modulation, please see Section 8.10.7.

## 2.6 Cable TV Return-Path Transmission Characteristics

The standard frequency plan for upstream transmission is shown in Table 2.4. The "T-channel" plan originated in about 1960 to identify the video channels that could be converted in a block to the FCC standard channel assignment T7 to T13. The current return-path or upstream transmission band is from 5 MHz to 42 MHz in the United States (5 to 65 MHz in Europe). Recently, it has been proposed by the same cable operators that the return-path band be located above the downstream channels in the 900-MHz to 1-GHz band. Although this upstream band is almost three times as large as the 5 to 42-MHz band, coaxial cable losses are almost ten times larger than at the 5–42 MHz band and exhibit strong frequency dependence. (See Table 2.5.) On the other hand, the upstream noise in these RF frequencies is significantly less than in the 5–42-MHz band. To date, cable TV operators, because of the cost to upgrade their network to 1 GHz, have mostly ignored this proposal.

### 2.6.1 Return-Path Noise Sources

The noise sources that impair the transmitted upstream signals in cable TV systems include ingress noise, common-path distortion, laser transmitter, and optical receiver noise. Ingress noise is the most important and dominant noise source in the return-path portion of the HFC network, and can be divided into three general types:

- Narrowband shortwave signals, primarily from radio and radar stations, that are transmitted terrestrially and coupled to the return-path band at the subscriber's home or in the cable TV distribution plant.
- Burst noise, which is generated by various manmade and naturally occurring sources, has time duration longer than the $(\text{symbol rate})^{-1}$.
- Impulse noise, which is similar to burst noise, but has time duration shorter than $(\text{symbol rate})^{-1}$ such that the receiver impulse response is effectively being measured.

The propagation of narrowband interference signals depends on the atmospheric conditions as well as the 10.7-year solar cycle. Increased solar sunspot activity has been correlated with increases in the maximum useable frequency because of ionization in the upper layers of the atmosphere. There are various methods for reducing ingress in the return-path band at the subscriber's location, which will be discussed later. Burst and impulse noise are generated from various manmade sources such as electric motors and power-switching devices. Although these sources produce burst/impulse noise events in the 60-Hz to 2-MHz portion of the spectrum, their harmonics show up in the 5–42-MHz upstream frequency band. Naturally occurring burst/impulse noise events include lightning, atmospherics, and electrostatic discharge, which typically extend from 2 kHz up to 100 MHz. Figure 2.12 shows a typical return-path spectrum as measured at a cable headend with the average signal level in dBmV. Notice the presence of various ingress peaks, particularly below 10 MHz.

As we will see later, for robust data transmission using quadrature-phase-shift-keying (QPSK) modulation, a 16–20 dB margin above the noise level is needed, as indicated by the horizontal solid line in Figure 2.12.

The time-varying ingress noise level also depends on the number of subscribers connected to the cable TV headend. Figure 2.13 shows the average ingress noise level as a function of the number of subscribers for T7, T8, T9, and T10 upstream channel frequencies (see Table 2.4), which is based on CableLabs field measurements [7]. The worst-case ingress levels can exceed +10-dBmV within a 100-kHz bandwidth. Two distinct trends emerge from Figure 2.13. First, the ingress noise levels are generally higher in the low-frequency region of the upstream band. Second, ingress noise levels increase for cable TV distribution plants with larger number of subscribers per fiber node. The solid curves in Figure 2.13 can be approximated by $A \cdot \log(N) - B$, where $N$ is the number of subscribers, $A = 9$, and $B = 28$, 33, 37, and 40 for T7 (top curve), T8, T9, and T10 channels, respectively.

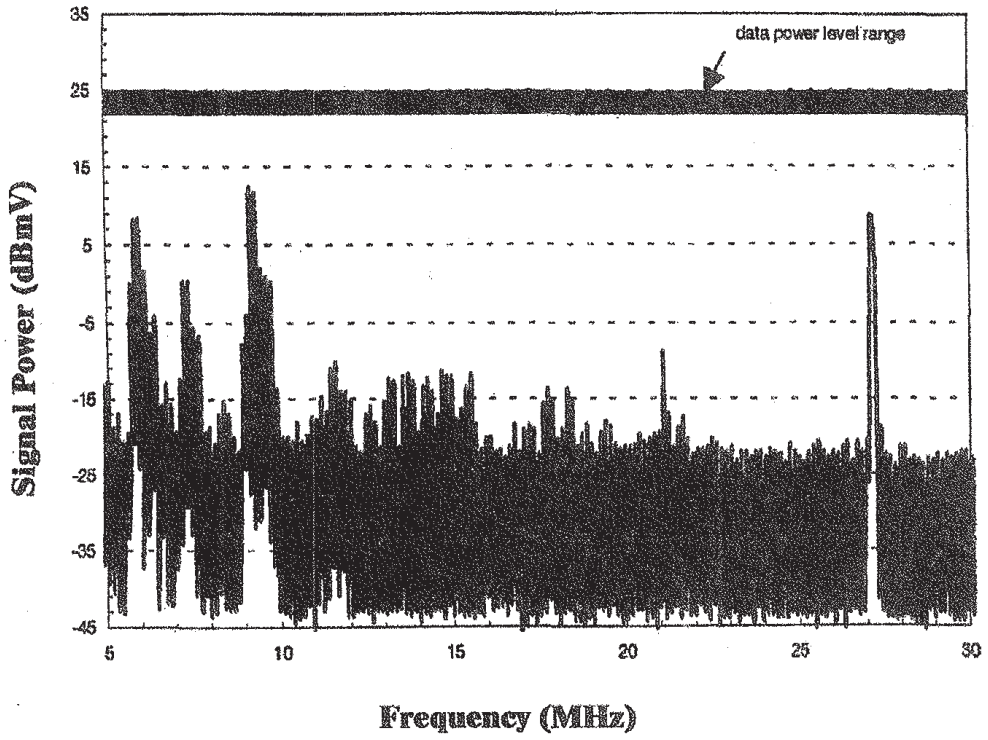



Figure 2.12 Typical return-path spectrum (5-30-MHz) as measured at the cable TV headend, with averaged signal levels (in dBmV). Notice the pronounced ingress peaks, particularly below 10 MHz. After Ref. [22] (© 1995 IEEE).

Another important noise source is common-path distortion, originating from various nonlinearities in the cable TV plant such as oxidized connectors and bad amplifiers. The common-path distortion appears as discrete noise peaks in the return-path spectrum spaced by 6 MHz (8 MHz in PAL systems). This distortion can be contained in properly maintained cable TV plants. Other types of noise sources are the upstream laser transmitter noise and optical receiver noise, which will be discussed in detail later.

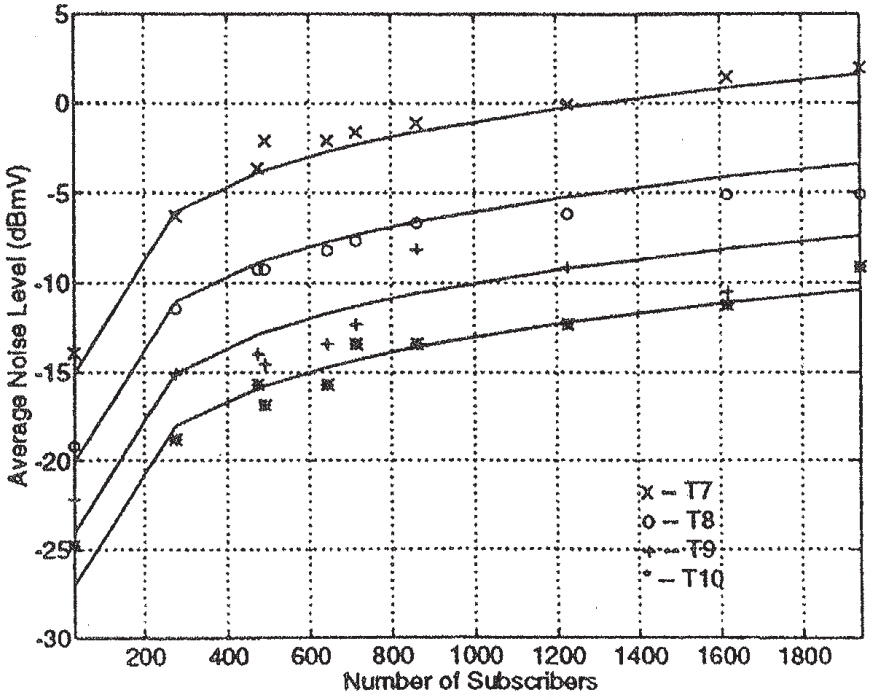



Figure 2.13 Average ingress level (in dBmV) as a function of the number of subscribers for T7, T8, T9, and T10 upstream channel frequencies. After Ref. [24].

The observed increase in the ingress noise levels as the number of homes passed is increased is due to the so-called *noise funneling* effect. This effect is based on the assumption that the unwanted noise signals are located at the subscriber's location with time-dependent amplitude. Since the unwanted signals are uncorrelated, for example, Gaussian noise, the signals sum noncoherently. If the signals from each home passed were equal, then the noise-funneling factor would be $10 \cdot \log_{10}(N)$, where $N$ is the number of homes passed. It turns out that according to the empirically derived expression based on Figure 2.13, the noise-funneling factor is slightly smaller than one obtains from a noncoherent noise power sum.

The time-varying narrowband interferers in the return-path cable systems limits channel availability for high-speed data transmission application. To quantify this parameter, automated spectral measurements of upstream ingress noise were taken every one to five minutes over an extended period of time (48 to 72 hours). The channel availability is the percentage of the time in which a channel of a specified bandwidth is available for transmission for a given modulation format such as QPSK or 16-QAM. For example, to transmit 256-kb/s, using QPSK modulation with FEC over a 192-kHz channel bandwidth, the required carrier-to-noise-plus-interference C/(N+I) must be at least 15.8 dB at BER of $10^{-7}$. Thus, if the C/N or CNR (in the presence of Gaussian noise only) is maintained at 21 dB, a single narrowband interferer as high as 10 dB below the equivalent unmodulated carrier (corresponding to C/I = 10 dB) can be tolerated.

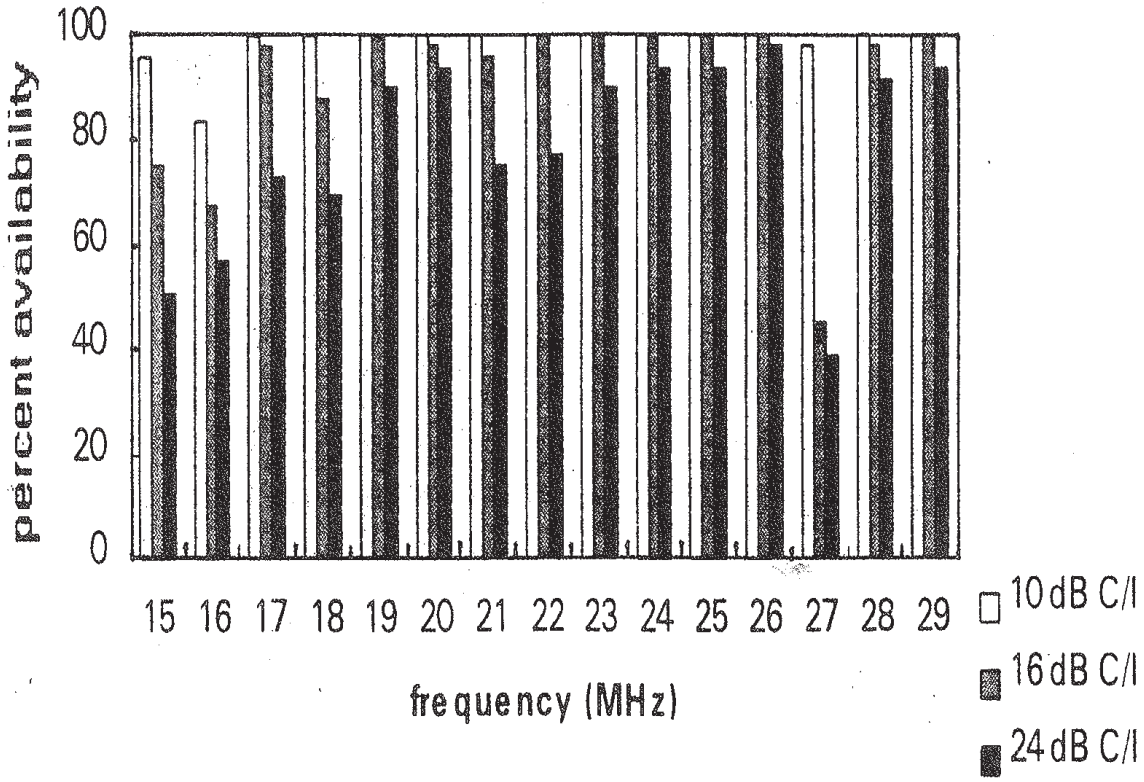



Figure 2.14 Percentage of 1-MHz channel availability in the 15 to 30-MHz frequency range for a 1500 home node. After Ref. [22] (© 1995 IEEE).

In general, the upstream channel availability depends primarily on the number of homes connected to the fiber node, the channel bandwidth, and selected upstream frequency. Based

on Figure 2.10, it is clear that by decreasing the number of homes passed in the fiber node the noise funneling effect is reduced. Also, selecting an upstream frequency above, say 15 MHz, is likely to improve the upstream channel availability based on Figure 2.12. Choosing a relatively wide upstream channel band, say 1 MHz or more, is also likely to reduce the channel availability. Figure 2.14 demonstrates, for example, the availability of the 1 MHz channel in the 15- to 30-MHz frequency range of the upstream spectrum for a single fiber node with 1500 homes passed [22]. Notice the reduction in channel availability as the C/I is increased from 10 dB to 24 dB.

### 2.6.2 Return-Path Noise Filtering

The appearance of ingress noise and other narrowband interference on the return-path cable network has introduced a new challenge for cable TV operators. Many companies have been working on developing methods to combat the effect of the ingress noise from reaching the cable TV headend. These methods include active or passive filtering techniques anywhere along the return-path channel from the subscriber to the fiber node.

One method, for example, consists of using a blocking filter, nominally between 15 to 40 MHz, between an in-home splitter and a coaxial termination unit (CTU) at the side of the home. The signals from the subscriber home terminals such as CM and STB can be transmitted from inside the home, but a low-pass filter prevents any signals originating from the home in the 15 to 42-MHz frequency band to enter the return-path cable network. Upstream-transmitted signals in the 15 to 42-MHz band can be added in the CTU after the blocking filter. Although this method reduces the amount of ingress noise coming from each home, it still allows the relatively high ingress spectral region of the return-path (5 to 15 MHz) to enter the cable network, and the method may be costly to implement. Another proposed method is to use a low-pass filter at the side of the house such that the filter is off except when the subscriber is transmitting data upstream.

Another method is using a bandpass filter with a switch at the side of the subscriber home. The upstream ingress noise from the subscriber home is filtered when the subscriber is not transmitting data. The disadvantage of this method is the large time delays from the switches when there are many subscribers connected to a fiber node ($\approx$1200) who are trying to transmit data upstream.

In addition to the use of filters and similar approaches, upstream ingress and other impairments can largely be controlled by the use of quality materials, proper subscriber drop installation practices, good network maintenance programs, and aggressive signal leakage monitoring and repair [26]. This latter item has the benefit of reduced ingress where a leak exists and outside signals can also enter the network. To make all of these efforts successful requires effective training and quality control programs.

## References


1. G. H. Hutson, P. J. Shepherd, and W. S. J. Brice, *Color Television System Principles, Engineering Practice and Applied Technology*, McGraw-Hill, Europe (1990).
2. *NCTA Recommended Practices for Measurements on Cable Television Systems*, 2nd Ed., National Cable Television Association, Washington, D.C. (1993).
3. John D. Lenk, *Lenk's Video Handbook: Operation and Troubleshooting*, McGraw-Hill (1991).
4. *Code of Federal Regulations, Title 47, Telecommunications, Part 76, Section 76.605, Cable Television Service.* Federal Communications Commission Rules and Regulations, Washington, D.C. (1991).
5. *National Association of Broadcasters Engineering Handbook*, 7th Ed., Washington, D.C. (1985).
6. L. Chariglione, "MPEG: A Technological Basis for Multimedia Applications," *IEEE Multimedia* **2**, 85–89 (1995).
7. *MPEG Video Compression Standard*, Edited by J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall, Chapman and Hill, N.Y. (1997).
8. B. G. Haskell, A. Puri, and A. N. Netravili, *Digital Video: An Introduction to MPEG-2*, Chapman & Hall, New York (1996).
9. See http://www.mpeg.org/ Web site for many technical information and news on the different MPEG standards.
10. L.Chiariglione, "MPEG and Multimedia Communications," *IEEE Transaction on CSVT* 7, 1, (Feb. 1997).
11. U.S. Advanced Television Systems Committee, *Digital Audio Compression (AC-3),* draft ATSC standard (1994).
12. S. Vernon, "Design and Implementation of AC-3 Coders," *IEEE Transaction on Consumer Electronics* **41**, no.3 (1995).
13. ITU-T Recommendation H.261: Video Codec for Audiovisual Services at px64 kbits (1993).
14. ITU-T Recommendation H.263: Video Coding for Low Bit-rate Communication (1995).
15. See http://drogo.cselt.stet.it/mpeg/, which is the official MPEG committee Web site, for information on MPEG-4 and MPEG-7 standards.
16. Electronic Industries Association, "Cable Television Channel Identification Plan," EIA IS-132, May (1994).
17. A. S. Taylor, "Characterization of Cable TV Networks as the Transmission Media for Data," *IEEE Journal on Selected Areas in Communications* **SAC-3**, 255–265 (1985).

18. S. Ovadia and Q. Zhang, "Robust Adaptive Equalization in DVB 64/256-QAM Receiver for CATV Applications," *ICSPAT Proceedings*, **1**, 332–336 (1998).
19. M. Kolber and M. Ryba, "Measuring Multipath in Wireless Cable Environment," *RF Design Magazine* **2**, 52–74 (1999).
20. T. J. Kolze, "Proposed HFC Channel Model," General Instrument IEEE802.14 contribution, IEEE802.14-96/196; Enschede, Netherlands (1996).
21. CableLabs, "Data-Over-Cable Service Interface Specifications (DOCSIS): Radio Frequency Interface Specification," version 1.0 (1997).
22. C. A. Eldering, N. Himayat, and F. C. Gardner, "CATV Return Path Characterization for Reliable Communications," *IEEE Communications Magazine* **8**, 62–69 (1995).
23. CableLabs, "Two Way Cable Television System Characterization," (1995).
24. M. D. Carangi, W. Y. Chen, K. Kerpez, and C. F. Valenti, "Coaxial Cable Distribution Plant Performance Simulation," *SPIE* vol. 2609, 215–226 (1995).
25. F. Edgington, "Preparing for in-service Video Measurements," *Communications Engineering and Design* **6**, 94–100 (1994).
26. R. Hranac, "Mystified by Return Path Activation? Get Your Upstream Fiber Links Aligned," *Communications Technology Magazine*, May (2000).

# CHAPTER 3

# DIRECTLY MODULATED CABLE TV LIGHTWAVE LASER TRANSMITTERS

Semiconductor laser diodes are the primary light sources in both downstream and upstream transmission over hybrid HFC and the emerging DWDM cable TV networks. However, as the cable TV networks are evolving from the traditional HFC architecture toward DWDM architecture, the laser transmitters must have different performance specifications, depending on the type of traffic being transmitted. Thus, to understand the laser transmitter design and its performance requirements, we shall start with a review of the basic laser physics concepts, including the Fabry-Perot (FP) laser, the distributed-feedback (DFB) laser, and multiple-quantum-well laser diodes. In Section 3.3, we will review the laser modulation characteristics, including the laser rate equations, small-signal frequency response, and equivalent circuit model. The two fundamental laser noise mechanisms, namely, the laser intensity and phase noise are reviewed in Section 3.4. The DFB laser transmitter design, which includes the packaged laser diode module with an optical isolator and a thermoelectric cooler, pre-distortion circuit, as well as its performance requirements for downstream transmission are reviewed in Section 3.5. Finally, Section 3.6 provides an overview on the performance requirements of return-path laser transmitters and their limitation due to their mode-partition noise and mode-hopping noise.

## 3.1 Semiconductor Laser Diodes

To understand the operation of semiconductor lasers, particularly DFB laser diodes, let us review first the basic laser physics concepts. Then, the different laser diode structures, including gain-guided and index-guided laser diodes as well as their light-versus-current characteristics are discussed.

### 3.1.1 Basic Laser Physics Concepts

A semiconductor laser diode is essentially a forward-biased p-n junction, which is formed by bringing into physical contact n-type and p-type semiconductor material. An n-type or p-type semiconductor is made by doping it with some impurities whose atoms have an excess valence electron (i.e., donors) or one less electron (i.e., acceptors) compared with the semiconductor atoms. In an n-type semiconductor, the excess electrons occupy the

conductor atoms. In an n-type semiconductor, the excess electrons occupy the conduction band states, which are typically empty for an undoped or an intrinsic semiconductor. As the doping level is increased, the quasi-Fermi level[1], which lies in the middle of the bandgap for an intrinsic semiconductor, moves toward the conduction band. Similarly, the quasi-Fermi level moves toward the valence band for the p-type semiconductor. Since the Fermi level must be continuous in thermal equilibrium, electrons and holes must diffuse across the p-n junction, and the ionized impurities create a built-in potential [1]. The space-charged region across the p-n junction, where the number of the mobile carriers is essentially 0, is called the *depletion region.* When the p-n junction is forward biased by an external source, the built-in potential is reduced, resulting in an electric current flow. The current density of a forward-bias p-n junction with voltage $V$ is given by [1]

$$J = J_s\left[\exp(qV / k_B T) - 1\right] \tag{3.1}$$

where $J_s$ is the saturation current density, which depends on the diffusion coefficients of both the electrons and the holes. The electrons and the holes, which are simultaneously present in the depletion region, can radiatively recombine to generate photons through either spontaneous or stimulated emission processes. In the spontaneous emission process, which is an incoherent process, the photons are randomly emitted in all directions with no phase relation among them. In contrast, the stimulated emission process is a coherent process, where the emitted photons not only have the same energy as the excitation photon, but also propagate in the same direction. In addition to the radiative recombination processes, electrons and holes can recombine nonradiatively. Nonradiative recombination processes include recombination at trap and defect sites, surface, and Augur recombination [2]. In Augur recombination, which is important in long-wavelength semiconductor lasers, the electron-hole recombination energy is given to another electron or hole as kinetic energy rather than being emitted as photons. The internal quantum efficiency of the p-n junction, which is given by [2]

$$\eta_{\text{int}} = \frac{R_{rr}}{R_{rr} + R_{nr}} = \frac{\tau_{nr}}{\tau_{rr} + \tau_{nr}} \tag{3.2}$$

is the radiative recombination rate ($R_{rr}$) relative to the total recombination rate. The radiative and nonradiative recombination times are defined as $N/R_i$, where $N$ is the carrier density, and i = rr or nr. In direct bandgap semiconductors such as GaAs, where the conduction band minimum and the valence band maximum occur at the same electron wave vector ($\vec{k}$), the probability of radiative recombination is large ($\eta_{\text{int}} \approx 1$). In contrast, $\eta_{\text{int}} << 1$ for indirect bandgap semiconductors such as Si or Ge.

[1] The Fermi level is the energy at which the probability of occupation by an electron is equal to ½.

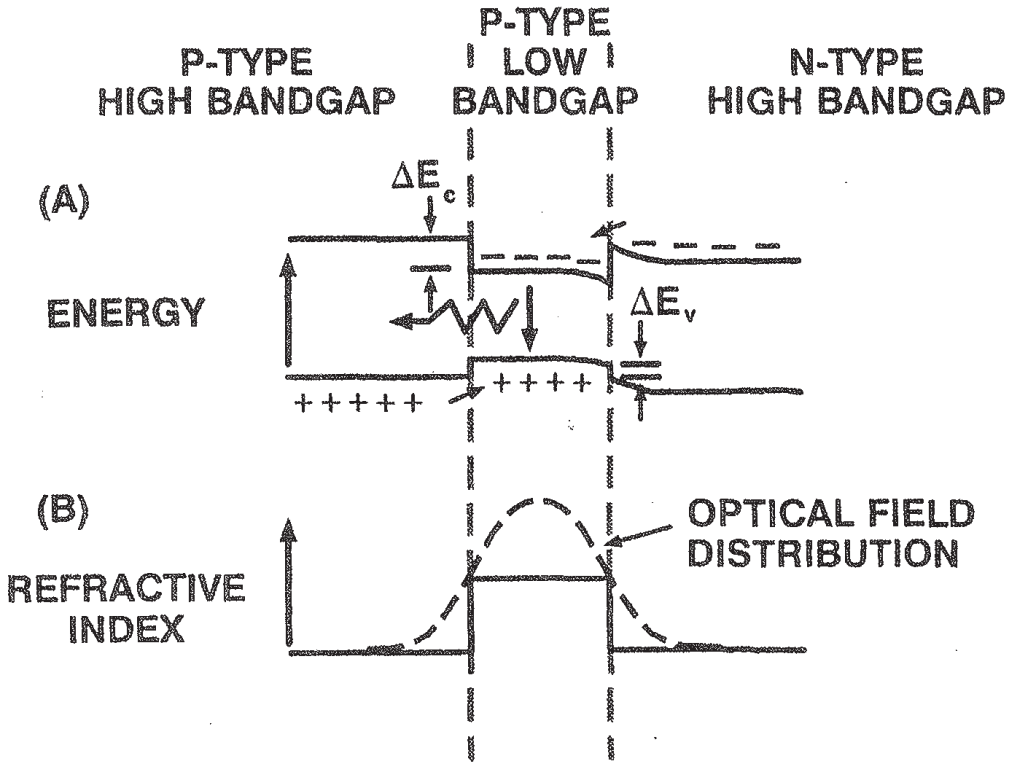



Figure 3.1 Schematic of the simultaneous confinement of the charged carriers and the generated optical mode in a double-heterostructure design. After Ref. [3] (© 1991 IEEE).

So far, we have discussed p-n junctions that are homojunctions, namely, the same semiconductor material is used on both sides of the junction. The problem with homojunctions is that the mobile carriers are not confined to the immediate vicinity of the junction due to diffusion, resulting in lower carrier densities. Inserting a thin semiconductor layer between the p-type and n-type layers with the same lattice constant but with a smaller bandgap and higher index of refraction solves this problem. Such a semiconductor structure is called a heterostructure. The thin semiconductor layer with a typical thickness of about 0.3 μm or less is called the active layer since light is generated inside it. The active layer provides two important functions. First, it confines the electrons and the holes due to the bandgap energy difference between the active layer and the surrounding layers, which are called cladding layers. Second, it also confines the optical modes like a dielectric waveguide since its index of refraction is higher than the cladding layers. The number of optical modes can be con trolled by the active layer thickness. Figure 3.1 illustrates the simultaneous confinement of the charged carriers and the generated optical mode to the active layer in a double-heterostructure (DH) design [3].

In order to obtain stimulated emission from the heterostructure, one has to overcome the absorption in the semiconductor material. To achieve this goal, it is required to have population inversion, namely, sufficiently high density of electrons in the conduction band and holes in the valence band. To obtain stimulated emission, the emitted photon energy has to satisfy the following condition

$$E_g < h\nu < E_{Fc} - E_{Fv} \equiv \Delta E_F \tag{3.3}$$

Equation (3.3) states that the emitted photon energy has to be larger than the bandgap energy, but smaller than the difference in the quasi-Fermi level energies in the conduction and valence bands. Optical gain can be achieved in the active layer when the injected carrier density in the active layer is sufficiently high to overcome the absorption. When the optical gain is sufficiently high, the peak optical gain dependence on the carrier density N can be approximated as [2]

$$g_p = g_0[N - N_0] \tag{3.4}$$

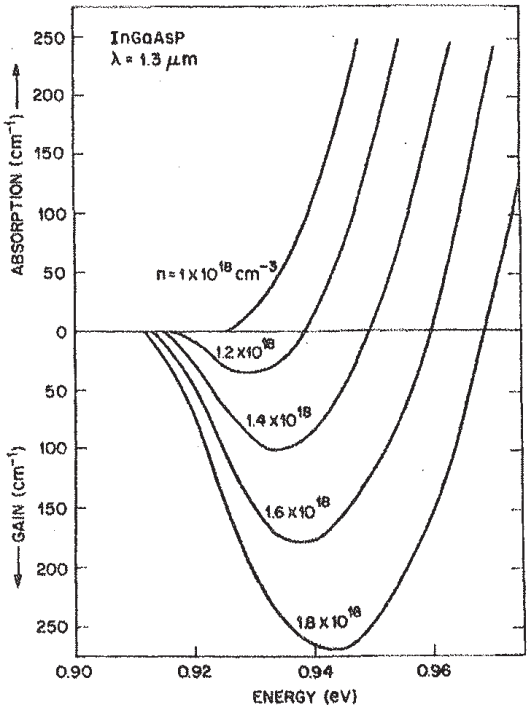



Figure 3.2 Calculated optical gain coefficient versus the photon energy at various carrier densities for 1.3-μm InGaAsP active layer, after Ref. [4] (© 1980 American Institute of Physics).

where $N_0$ is the carrier density at transparency, which is typically in $1–1.5\cdot10^{18}$ $cm^{-3}$ for InGaAsP lasers, and $g_0$ is the differential gain coefficient. In general, the optical gain is a function of both the carrier density and the photon energy g(N,ν). Figure 3.2 shows the calculated optical gain at various carrier densities as a function of the emitted photon energy for 1.3-μm InGaAsP active layer [4,5]. The differential gain coefficient can be increased by at least a factor of two in multiple quantum well (MQW) laser diodes (see Section 3.2.2).

Thus, we learned that in order to obtain optical gain, the injected carrier density has to be sufficiently high. However, a semiconductor laser diode is not an optical amplifier, but an optical oscillator. Thus, the optical feedback mechanism is required to achieve optical oscillations. The optical feedback is provided by cleaving both facets of the semiconductor laser, and applying high-reflectivity coatings to one or both facets. An optical cavity, which is formed by placing a gain medium between two optical mirrors, is often called a *Fabry-Perot* (FP) cavity. In the case of semiconductor lasers, the cleaved facets act as mirrors with reflectivity of $(n–1/n+1)^2$ for $n = 3.5$, or about 30% reflectivity.

To achieve lasing, the two following conditions must be satisfied:

- The optical gain in the cavity has to be equal to the loss.
- Optical feedback is needed, which is provided by the cleaved facets in FP lasers.

The laser threshold condition is the minimum optical gain that is necessary to overcome the FP cavity losses and operate the laser. To obtain the laser threshold condition, let us assume an electric field with amplitude $E_0$, angular frequency ω, and propagation constant $\beta = n\omega/c$, inside a cavity with length $L$, and mirror reflectivities of $R_1$ and $R_2$. During one cavity round trip, the electric field amplitude changes by $(R_1\cdot R_2)^{1/2}\exp(-\alpha_i L)$, where $\alpha_i$ is the internal loss of the cavity due to free-carrier absorption and scattering. At the same time, the amplitude of the electric field also increases as $\exp[g/2(2L)]$, where $g$ is the optical gain of the cavity, and its phase changes by $2\beta\cdot L$. Thus, after one round trip and in steady state, the electric field should remain unchanged, namely

$$E_0 = E_0 \exp(g_{th}L)\sqrt{R_1R_2}\exp(-\alpha_i L)\cdot\exp(2j\beta L) \tag{3.5}$$

From the real part of Equation (3.5), we obtain the first lasing condition at threshold:

$$g_{th} = \alpha_i - \frac{1}{2L}\ln[R_1R_2] = \alpha_{cavity} \tag{3.6}$$

Equation (3.6) states that the cavity losses must equal the laser gain at threshold. For GaAs lasers with cavity length $L$ = 300-μm, $\alpha_i$ = 10 $cm^{-1}$, $R$ = 0.36, the threshold gain is 44 $cm^{-1}$. The second lasing condition is obtained from the imaginary part of Equation (3.5):

$$2\beta_m L = 2m\pi \tag{3.7}$$

where $m$ is an integer. Thus, the laser wavelength must satisfy the following condition:

$$\lambda_m = 2nL/m \tag{3.8}$$

Equation (3.8) defines the longitudinal modes or the FP modes of the semiconductor laser. The longitudinal mode spacing is given by

$$\Delta\lambda = \frac{\lambda^2}{2[n-\lambda(dn/d\lambda)]L} = \frac{\lambda^2}{2n_g L} \cong \frac{\lambda^2}{2nL} \tag{3.9}$$

where $n_g$ is the group index, which can be approximated to the index of refraction of the active layer if its wavelength dependence can be ignored. Using the above example for GaAs lasers with $\lambda$ = 850-nm and $n$ = 3.5, the mode spacing is 0.385-nm. Since the gain bandwidth of a FP semiconductor laser is very broad (≅ 30-nm), it can support many longitudinal modes, where each mode can carry a significant amount of the optical power relative to the main mode. Such a laser is often called a *multimode* semiconductor laser. When the light from a multimode laser is coupled into a SMF, each of the longitudinal modes propagates at slightly different speed due to the fiber dispersion, resulting in a limited propagation distance. Consequently, single-wavelength semiconductor lasers are needed, which are discussed in Section 3.2.

### 3.1.2 Semiconductor Laser Structures

Semiconductor laser diodes can be divided into two main categories: gain-guided and index-guided lasers. Let us first discuss the structures of gain-guided lasers. Figure 3.3 shows, for example, gain-guided semiconductor laser in a stripe geometry [4]. In a simple DH semiconductor laser, there is no mechanism for lateral confinement of the injected carriers in the active layer. Consequently, the threshold current, which is proportional to threshold current density ($J_{th}$) times the active layer area ($LxW$), can be as high as 300-mA. One of the approaches to solve this problem was to deposit an oxide stripe ($SiO_2$) on top of the p-type layer where carriers can be injected [7]. Another approach, for example, was to deposit an n-type layer on top of the p-layer and diffuse Zn over the central region, which converts the n-type to a p-type layer [8]. Thus, the narrow strip (≤ 5 μm) along the cavity laterally confines the injected carriers and the optical modes. These types of semiconductor laser designs have several problems. The stripe geometry leads to a spatially varying carrier density in the lateral direction, which is determined by the carrier diffusion. In addition, the optical gain, which peaks at the center of the stripe, may be significantly reduced beyond the stripe to large carrier absorption. Since the confinement of the optical modes is governed by the

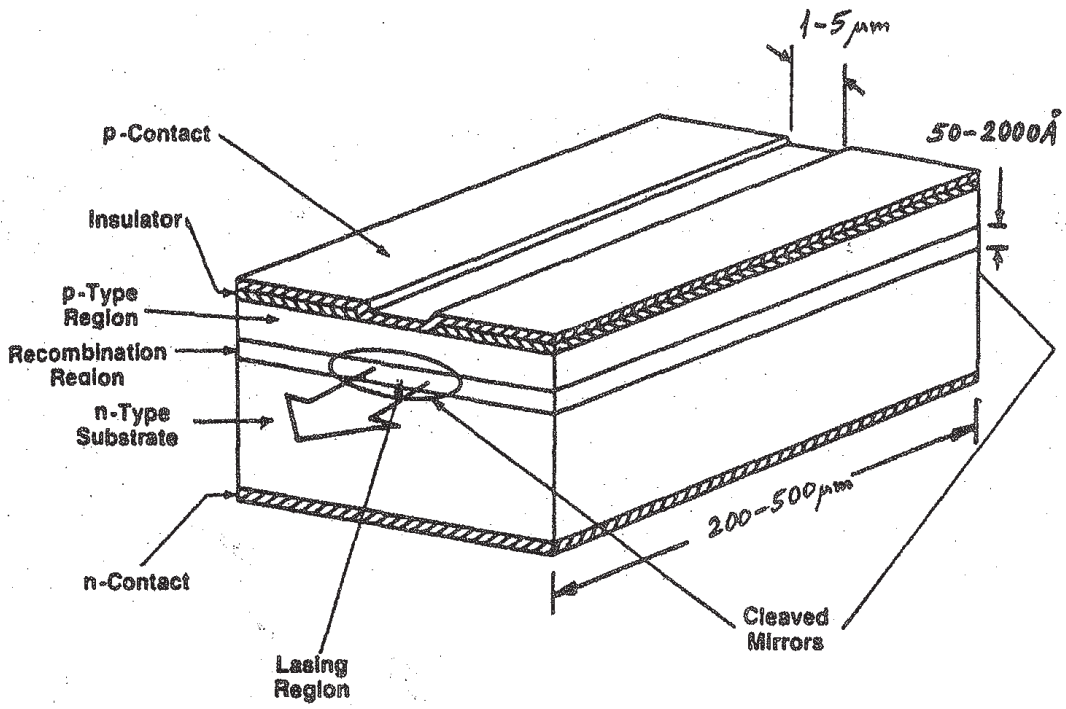



Figure 3.3 A Gain-guided semiconductor laser in a stripe geometry showing its main features, including the lasing region, n-type and p-type regions and contacts, and the cleaved mirrors. After Ref. [6].

optical gain, such lasers are called gain-guided semiconductor lasers. One of the main disadvantages of gain-guided lasers is their light-versus-current nonlinearity characteristic as the current is increased. This nonlinearity, which has been attributed to a transition to higher order modes, limits the maximum amount of optical power that can be coupled into a SMF [2].

In index-guided lasers, lateral confinement of the injected carriers and the optical modes is achieved by forming a waveguide in the lateral direction with refractive index distribution. Index-guided lasers can be subdivided into two classes: weakly index-guided and strongly index-guided semiconductor lasers. An example of a weakly index-guided design is a *ridge-waveguide* laser, where the ridge is formed by etching parts of the p-type layer and depositing an $SiO_2$ layer to block the current flow and create the index guiding. The lateral index guiding is because of the higher index of refraction of the ridge-waveguide (p-type layer) compared with the surrounding $SiO_2$ layer. Although such designs produce a refractive index change of only about 0.01, they are attractive due to their relatively simple design. An example of strongly index-guided design is a *buried-heterostructure* (BH) laser, where the active layer is buried from all sides by layers with a lower index of refraction [2,9]. BH semiconductor lasers typically exhibit a very linear light-versus-current characteristic with

relatively low threshold (≈ 10-mA), modulation bandwidth in excess of 2 GHz, and can operate in the fundamental transverse mode even at high power.

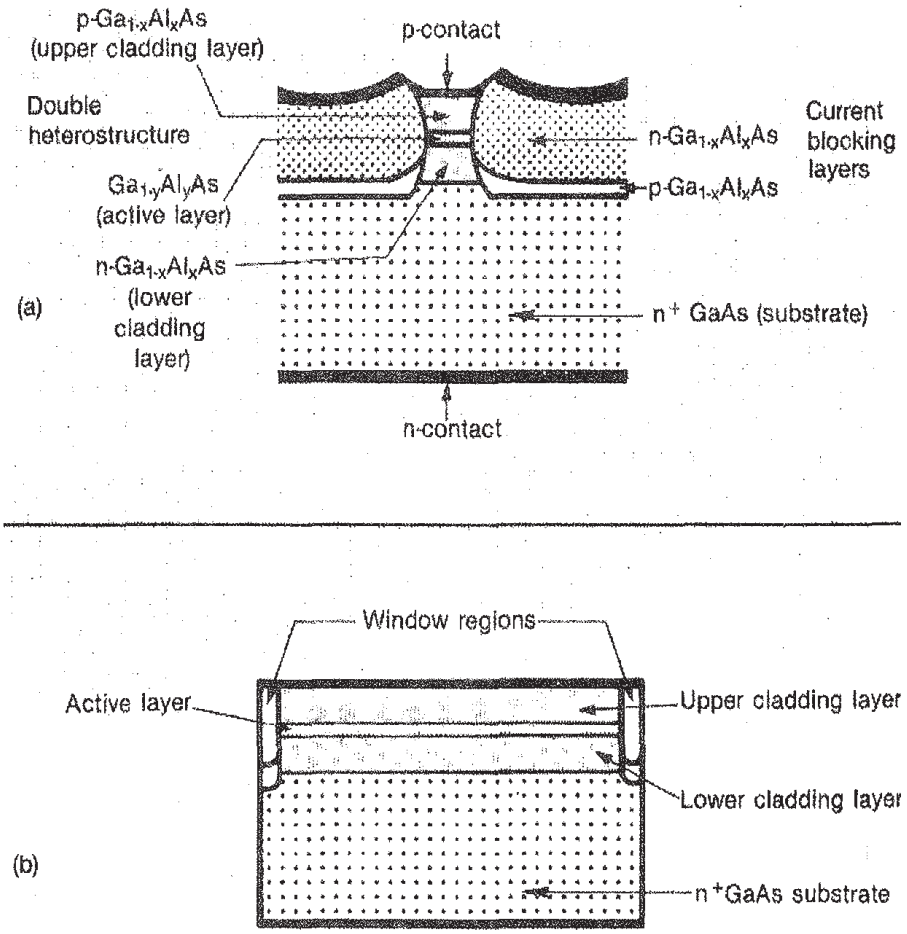



Figure 3.4 (a) Cross section, and (b) side view of window buried-heterostructure semiconductor laser. After Ref. [9].

### 3.1.3 Light-Current Characteristic

Generally speaking, the most fundamental characteristic of a laser diode is its output optical power versus the injected current or the *L–I* characteristic. Figure 3.5 shows the measured *L–I* curves (CW) at operating temperatures ranging from 10°C up to 130°C for a 1.3 μm double-channel planar BH (DCPBH) InGaAsP laser [2]. Two important observations can be made. First, the laser threshold current at room temperature degrades significantly from

about 20-mA to more than 100-mA as the operating temperature is increased. In fact, the threshold current was found experimentally to increase exponentially with temperature as:

$$I_{th}(T) = I_0 \exp[T / T_0] \tag{3.10}$$

where $I_0$ is a constant, and $T_0$ is a characteristic temperature often used to describe the temperature sensitivity of the threshold current. For example, $T_0$ is typically in the range of 50 to 70°K for InGaAsP lasers compared with 100–150°K for GaAs lasers. The exponential increase of the threshold current with temperature is primarily attributed to the reduction of the optical gain with temperature (gain $\propto 1/T$). Because of the high temperature sensitivity of InGaAsP lasers, their temperature is often controlled in a laser transmitter system using a thermoelectric cooler (see Section 3.5.3).

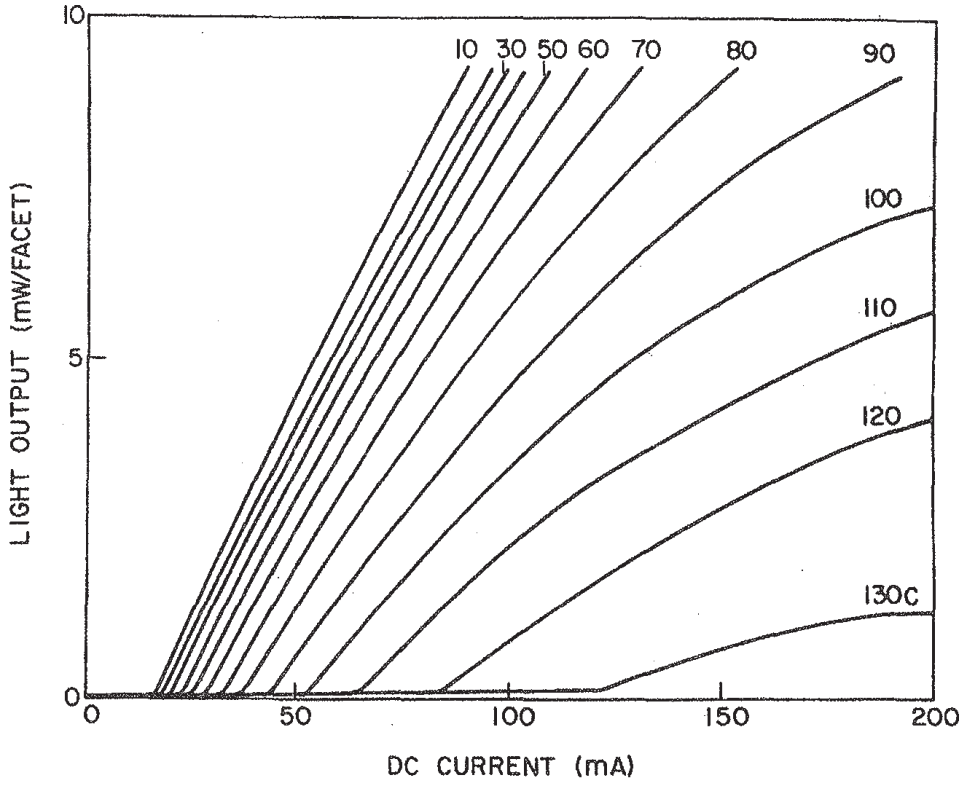



Figure 3.5 Measured L–I curve at various operating temperatures for 1.3-μm InGaAsP DCPBH semiconductor laser. After Ref. [2] (with kind permission from Kluwer Academic Publishers).

The second important observation from Figure 3.5 is the slope ($dP/dI$) of the $L$–$I$ curve, which is being considerably reduced as the operating temperature is increased. Above threshold ($I > I_{th}$), the slope efficiency is given by [2]

$$\frac{dP}{dI} = \frac{h\nu}{q}\eta_D = \frac{h\nu}{q}\left[\frac{\alpha_i \alpha_m}{\alpha_i + \alpha_m}\right] \tag{3.11}$$

where $\eta_D$ is the so-called differential quantum efficiency, which measures the efficiency of the laser to convert current to output light, and αm is the mirror loss. Using Equation (3.6), the differential quantum efficiency can be written as

$$\eta_D = \eta_{\text{int}} \frac{\ln[1/R_1R_2]}{2\alpha_i L + \ln[1/R_1R_2]} \tag{3.12}$$

As the operating temperature is increased, the internal quantum efficiency as defined in Equation (3.2) is reduced due to the exponential temperature dependence of the nonradiative recombination processes such as Auger recombination and leakage currents in BH lasers

## 3.2 DFB and Multiple-Quantum-Well (MQW) Laser Diodes

### 3.2.1 DFB Laser Diodes

As we have learned in Section 3.1.1, the cleaved facets of the semiconductor cavity provide the optical feedback in FP lasers. Another type of optical feedback is a *distributed feedback* (DFB) throughout the semiconductor cavity. The distributed feedback is achieved by periodic perturbation of the refractive index from the built-in diffraction grating. The DFB mechanism is a mode-selective mechanism that allows oscillation to occur only for the mode that satisfies the Bragg condition:

$$\lambda_B = \frac{2n \cdot \Lambda}{m} \tag{3.13}$$

where $n$ is the refractive index of the mode, Λ is the grating spatial period, and $m$ is an integer representing the Bragg diffraction order. Typically $m = 1$ is used because of the strongest coupling between the forward and backward oscillating modes, but sometimes $m = 2$ is used. For a 1.55-μm DFB laser, the gating period is only 235 nm for $m = 1$ and $n = 3.3$.

The grating in DFB lasers is typically etched onto either side of the active layer, creating periodic variations in the waveguide layer next to the active layer, which translate to refractive index modulation of the propagating modes along the cavity. It should be pointed out that the cleaved facets of the DFB laser act as mirrors whose reflectivity is maximized for the Bragg wavelength that satisfies Equation (3.13).

In order to obtain a single longitudinal operation, the cavity losses have to be significantly higher for the side modes than the main mode. The longitudinal mode with the smallest cavity losses can reach threshold and becomes the dominant mode. However, depending on the coupling coefficient between the forward and backward modes over the grating length ($\kappa L$), two dominant modes may coexist in the DFB laser cavity. DFB lasers with two competing longitudinal modes with a typical power difference of about 10 dB are undesirable from an optical transmission point of view. To discriminate between the two modes, the grating is shifted by $\lambda_B/4$ in the middle of the laser cavity to produce a $\pi/2$ phase shift to one of the modes. The $\pi/2$ phase shift detunes the Bragg wavelength from the gain peak, thus allowing only one mode to exist. Furthermore, the mode detuning is adjusted on the shorter wavelength side of the gain peak in order to obtain a higher differential gain. These types of DFB lasers are called $\lambda/4$-shifted DFB lasers [2].

### 3.2.2 Multiple-Quantum-Well (MQW) Lasers

In conventional DH semiconductor lasers, the active layer thickness (0.1–0.3 μm) is sufficiently thin to confine both the carriers and the optical mode, but its electronic and optical properties remain the same as the bulk semiconductor material. If the active layer thickness is further reduced to about 10 nm or less, its electronic and optical properties are drastically changed since the three-dimensional free-electron motion is reduced to two-dimensional motion. Consequently, the energy levels of the confined carriers are quantized in the z-direction, and given by

$$E_n = \frac{h}{4\pi m^*}\left[\frac{n\pi}{L_z}\right]^2 \tag{3.14}$$

where $h$ is Planck's constant, $m^*$ is the effective mass of the carrier, $L_z$ is the quantum-well thickness, and $n$ is an integer. The carriers are free to move in the x and y directions. The quantization of the energy levels changes the density of states from a parabolic dependence to a step-like behavior as shown in Figure 3.6 [3]. Notice that the density of states is constant at each energy sub-band, rather than gradually increasing from zero. This means that there is a group of electrons with nearly the same energy that is available to recombine with a group of holes of nearly the same energy. Therefore, a much higher optical gain can be achieved in these MQW structures than in conventional DH designs. Consequently, the required carrier density to reach threshold ($N_0$) is lower, resulting in very low threshold currents.

Strained-layer (SL) QW lasers are an important class of QW lasers with an active layer whose lattice constant is slightly different from the cladding layers and the substrate. These lasers, which have been extensively investigated over the last decade, exhibit many desirable characteristics, including very low threshold currents, narrower CW and modulated linewidth than conventional QW lasers, high relaxation oscillation frequency, and low

nonlinear distortions [3]. In fact, threshold currents as low as 1-mA or less have been demonstrated for 1.3-μm SL-MQW InGaAsP semiconductor lasers operating at room temperature [10, 11]. The improved laser characteristics have been explained by the changes in the band structure caused by the lattice mismatch-induced strain. The induced strain splits the heavy-hole and the light-hole valence bands at the Γ point of the Brillouin zone, which is the minimum bandgap energy in the direct bandgap semiconductors [3]. One of the widely used material systems for SL-MQW lasers is $In_xGa_{1-x}As$ QW layers with InGaAsP barriers whose composition is lattice match to InP. For $x < 0.53$, the QW active layer is under a tensile stress, while for $x > 0.53$, the QW active layer is under a compressive stress.

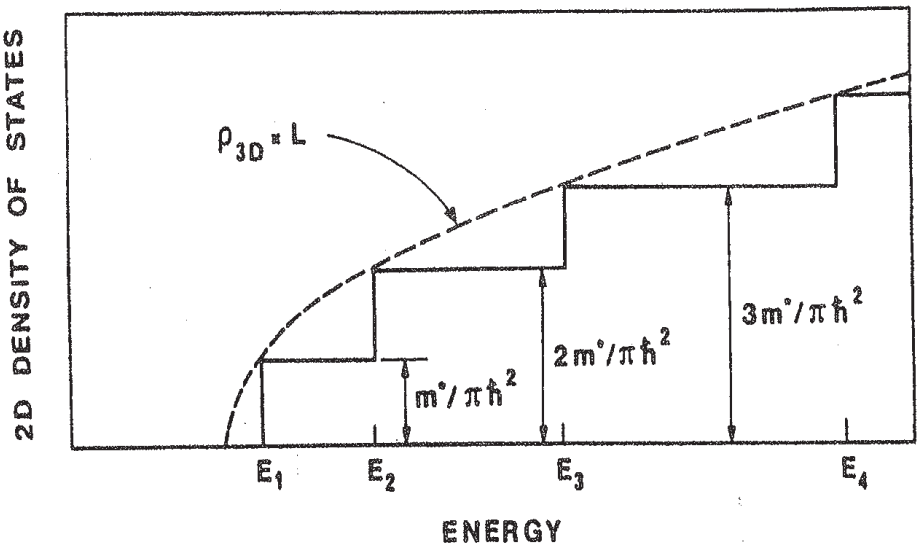



Figure 3.6 Density of states of a two-dimensional confined carrier in a quantum well compared to that of a bulk semiconductor. After Ref. [3] (© 1991 IEEE).

## 3.3 Laser Dynamic Characteristics

Until now, only the static characteristics of semiconductor lasers have been discussed. In the following subsections, we will examine the dynamic characteristics of semiconductor lasers and define key parameters that play an important role in the laser transmitter design for cable TV transmission. First, the temporal behavior of a semiconductor laser is analyzed using the modal rate equations. The aim of this section is to summarize some of the important results and concepts rather than to present detailed derivations of the various equations.

### 3.3.1 Small-Signal Response

The temporal behavior of semiconductor lasers is governed by the laser rate equations, which describe the interaction between the injected carriers and the photons inside the active region. Let us assume a single-longitudinal laser diode such as DFB laser with uniform distribution of electrons with density $N$ over the volume $V$ (= $d$x$L$x$W$) of the active layer, and photon density $S$ for the lasing mode. The single-mode rate equations can be written as [2]:

$$\frac{dN}{dt} = \frac{I}{qV} - \frac{N}{\tau_n} - \Gamma v_g g_0 (N - N_0)(1 - \varepsilon S)S \tag{3.15}$$

$$\frac{dS}{dt} = \Gamma v_g g_0 (N - N_0)(1 - \varepsilon S)S - \frac{S}{\tau_p} + \Gamma \gamma_{sp} \frac{N}{\tau_n} \tag{3.16}$$

where $\tau_n$ and $\tau_p$ are the electron and photon lifetimes, $\Gamma$ is the carrier confinement factor in the active layer, $v_g$ is the group velocity, $\gamma_{sp}$ is the fraction of the spontaneous emission coupled into the cavity mode. In Equation (3.15), the electron density rate of change is proportional to (A) the injected electron rate, (B) the electron-density rate reduction due to spontaneous emission, and (C) the rate reduction due to stimulated emission of the lasing mode. Notice that the mode gain has been modified from Equation (3.4) to include the effect of gain compression. Without gain compression, the net rate of peak gain due to stimulated emission is given by

$$G = v_g g_0 (N - N_0) \tag{3.17}$$

However, due to an important phenomenon called spectral-hole burning (SHB), the optical gain is compressed, requiring modifications to Equation (3.17). SHB is a small-localized reduction of the optical gain curve at the lasing mode wavelength [12]. It occurs when the rate of stimulated emission is very high and the large number of available electrons in the conduction band cannot immediately fill the localized reduction in the number of conduction band states. This is because the intraband scattering processes such as electron-electron and electron-phonon scattering have a relaxation time on the order of 1-ps [13]. The gain rate reduction caused by SHB is proportional to the stimulated emission rate, and can be written as

$$\Delta G = v_g g_0 (N - N_0) S \cdot \varepsilon \tag{3.18}$$

where $\varepsilon$ is the gain compression coefficient, which is also called the nonlinear gain parameter. Thus, the total net gain rate is given by

$$G(n,S) = v_g g_0 (N - N_0) - \Delta G = v_g g_0 (N - N_0)(1 - \varepsilon S) \tag{3.19}$$

The rate of change for the photon density in Equation (3.16) is proportional to (A) the fraction Γ of photons confined in the active layer due to stimulated emission, (B) the reduction of the photon density due to internal absorption or transmission through the facets, and (C) the rate increase of the photon density caused by the coupling of the spontaneous emission into the lasing mode. Table 3.1 summarizes the typical values of DFB laser diodes that are used for cable TV transmission.

**Table 3.1 Typical values for 1.3-μm InGaAsP DFB laser diodes.**

| Parameter | Typical Values |
|---|---|
| Active Layer Volume | 2-μmx0.2-μmx300-μm |
| Electron Lifetime ($\tau_n$) | 1–3-ns |
| Photon Lifetime ($\tau_p$) | 1–3-ps |
| Optical Mode Confinement Factor (Γ) | 0.3–0.5 |
| Electron Density at Transparency ($N_0$) | $1–1.5x10^{18}$ $cm^{-3}$ |
| Gain Compression Coefficient ($\varepsilon$) | $1–5x10^{-17}$ $cm^{-3}$ |
| Differential Gain Coefficient ($g_0$) | $2–3x10^{-16}$ $cm^2$ |

Due to the nonlinear behavior of the rate equations, it is necessary to solve them numerically. However, analytical solution can be obtained for the small-signal modulation response. The small-signal response solution is important for cable TV applications since it reveals the design parameters that need to be controlled in order to obtain high modulation bandwidth of DFB laser diodes. To obtain an analytical solution, let us assume that the general solution to the rate equations is in the form

$$\begin{aligned} N &= n_0 + ne^{j\omega t} \\ S &= S_0 + se^{j\omega t} \\ I &= I_0 + ie^{j\omega t} \end{aligned} \tag{3.20}$$

where ω is the angular frequency of the modulating signal, and $n_0$, $S_0$, and $I_0$ are the steady-state values for the electron density, photon density, and current modulation, respectively. Substituting Equation (3.20) into the rate equations, equating terms on both sides of the rate equations with exp(jω$t$), and neglecting second-order small-signal products, the intensity modulation response is given by [14]

$$H(\omega)=\frac{s(\omega)/S_0}{i(\omega)/I_0}=\frac{B\omega_R^2}{B\omega_R^2-\omega^2+j\omega\left[\dfrac{\Gamma\gamma_{sp}I_{th}}{qVS_0}+\dfrac{1}{\tau_n}+S_0\left(v_g g_0+\dfrac{\varepsilon}{\tau_p}\right)\right]+\dfrac{\Gamma\gamma_{sp}I_{th}}{qVS_0\cdot\tau_n}+\dfrac{\gamma_{sp}+\varepsilon S_0}{\tau_n\tau_p}} \tag{3.21}$$

where $I_{th}$ is the threshold current, $B = 1-\varepsilon S_0$, and $\omega_R$ is the so-called laser resonance frequency or relaxation oscillation frequency, which is given by

$$\omega_R=\left[\frac{v_g g_0 S_0}{\tau_p}\right]^{1/2}=\left[\frac{\Gamma v_g g_0}{qV}(I-I_{th})\right]^{1/2} \tag{3.22}$$

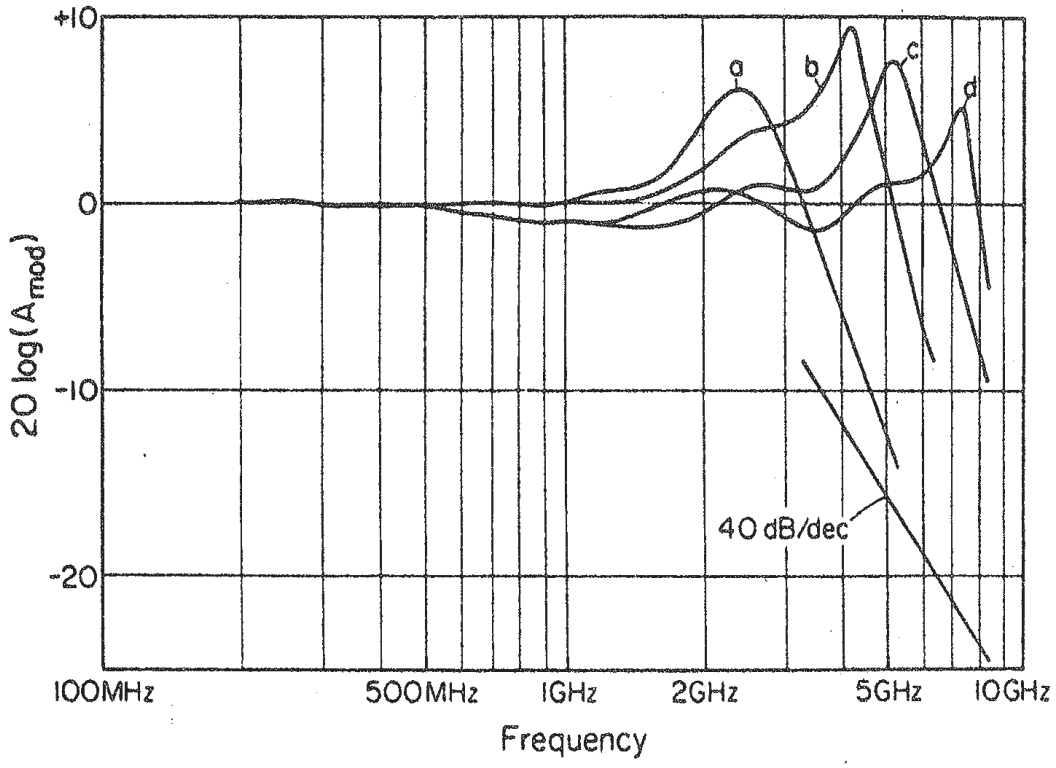



Figure 3.7 Measured modulation characteristics of 120-μm cavity of BH laser on semi-insulating substrate at optical output powers of (a) 1-mW, (b) 2-mW, (c) 2.7-mW, and (d) 5-mW. After Ref. [15] (© 1985 IEEE).

According to Equation (3.21), the small-signal modulation response of a semiconductor laser behaves as a standard second-order low-pass transfer function with a resonance frequency $\omega_R$ and with a damping factor, whose magnitude is given by the coefficient of the $j\omega$ term. For laser operation above threshold, $\gamma_{sp} = 0$, the $1/\tau_n$ and $S_0 g_0$ terms in the damping coefficient are small and can be neglected, and Equation (3.21) simplifies to

$$H(\omega) = \frac{\omega_R^2}{\omega_R^2 - \omega^2 + j\omega\left[S_0 \varepsilon / \tau_p\right]} \tag{3.23}$$

The electrical power spectrum of the small-signal frequency response is given by

$$|H(\omega)|^2 = \frac{B^2 \omega_R^4}{\left[B\omega_R^2 - \omega^2\right]^2 + \omega^2\left[S_0 \varepsilon / \tau_p\right]^2} \tag{3.24}$$

In most lasers the damping coefficient is much smaller than the resonance frequency. Consequently, the 3 dB modulation frequency can be obtained from Equation (3.24) as [15]

$$\omega_{3dB} = \omega_R\sqrt{1+\sqrt{2}} \cong 1.55\omega_R \tag{3.25}$$

where the approximation $B^{1/2}\omega_R \approx \omega_R$ was used in Equation (3.24). If, on the other hand, the damping coefficient is large, the height of the resonance peak will be reduced, and its frequency will be different than $\omega_R$. Notice that the 3 dB modulation frequency is proportional to $(I-I_{th})^{1/2}$, which has been verified by many experimental results [16]. In addition, from Equation (3.24), the power spectrum at the resonance frequency is inversely proportional to $S_0$, which means that the resonance peak is reduced due to gain saturation as the output power is increased. Figure 3.7 shows, for example, the measured modulation characteristics of 120-μm cavity of BH semiconductor laser on semi-insulating substrate at optical output powers of (a) 1-mW, (b) 2-mW, (c) 2.7-mW, and (d) 5-mW [15]. Notice that the low frequency region (up to 1 GHz) of the frequency response of the BH laser is essentially flat. This region can be used for SCM lightwave transmission. In addition, above the resonance frequency, the magnitude of the frequency response approaches asymptotically a slope of –40-dB/decade.

It is well known that the laser behavior is highly nonlinear near its resonance frequency. For SCM lightwave transmission, the RF carriers must be sufficiently far away from the resonance frequency in order to minimize any dynamically induced nonlinear distortions. Therefore, the semiconductor laser designer must find ways to expand the laser modulation bandwidth, increasing the frequency separation between the operating frequency range and the resonance frequency. According to Equation (3.22), the laser modulation bandwidth can be increased by (A) increasing the output power by operating at higher bias current, (B)

increasing the differential gain coefficient $g_0$, and (C) reducing the active layer volume. In fact, recent results have shown that SL-MQW DFB lasers have 30% and 90% higher resonance frequency compared with MQW DFB lasers and conventional DFB lasers, respectively.

### 3.3.2 Large-Signal Response Circuit Model

Generally speaking, it is convenient to model the high-speed characteristics of a packaged semiconductor laser as a two-port system as shown in Figure 3.8 [14]. It consists of three subsections: (A) the package or mount parasitics, (B) the parasitics associated with the semiconductor chip, and (C) the intrinsic laser namely, the active region and the cavity. The package parasitics include, for example, bond-wire inductance and some capacitance between the input terminals, which can be entirely eliminated by monolithically integrating the laser with the drive circuit elements, but can also behave as nonlinear elements for some types of drive circuitry. The laser chip parasitics are the stray capacitance and resistance of the various semiconductor layers surrounding the active region, which typically behave as linear current in cases such as input step current waveform. The laser chip parasitics, which greatly vary among different laser designs, are diverting the high-frequency components of the drive current away from the intrinsic laser. Figure 3.9 illustrates, for example, the equivalent circuit model of the package and chip parasitics for a 1.3-μm InGaAsP etched mesa BH laser [14]. The chip parasitics take the form of a resistance in series combined with a shunt capacitance. $L_p$ and $R_p$ represent the bond-wire inductance and resistance, while $C_s$ and $R_s$ are the parasitic capacitance and resistance of the intrinsic laser. The current source $I_L$ models the DC leakage current around the active region, and $R_{IN}$ is the source resistance.

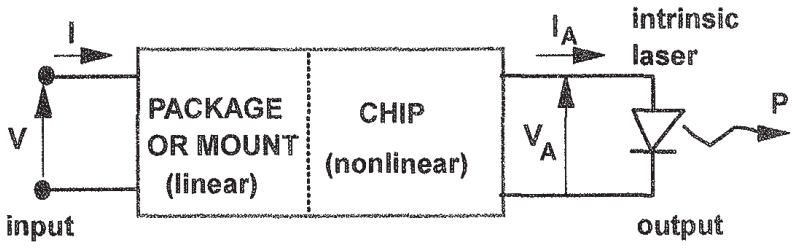



Figure 3.8 Two-port model of semiconductor laser with the package and chip parasitics. After Ref. [14] (© 1985 IEEE).

The small-signal analysis, which was introduced in the previous section, does not apply to a directly modulated DFB laser transmitter for cable TV applications. For a large-signal

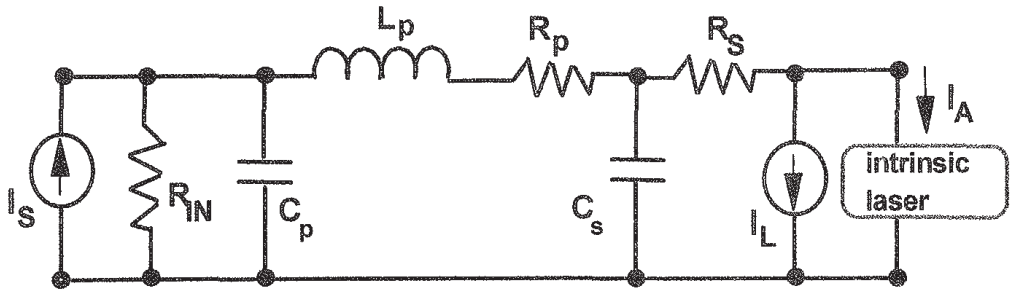



Figure 3.9 Equivalent circuit model of the laser's package and chip parasitics. After Ref. [14] (© 1985 IEEE).

response, the rate Equations (3.15) and (3.16) must be solved numerically. Another convenient method is to transform the rate equations into an equivalent circuit model, which can be solved using standard circuit-analysis techniques. Various authors have developed a number of large-signal models [14, 17]. The following example demonstrates the circuit-modeling methodology for a large-signal response. Figure 3.10 shows the equivalent circuit model of an intrinsic semiconductor laser [14]. To transform the rate Equations (3.15) and (3.16) to the circuit model, the equivalent spontaneous recombination current is defined as:

$$I_{spon} = \frac{qV \cdot N}{\tau_n} = I_s \exp(qV_A / \theta k_B T) \tag{3.26}$$

where $\theta \cong 2$, and $I_s$ the saturation current. The equivalent stimulated emission current is defined as

$$I_{stim} = qV \cdot v_g g_0 (N - N_0)(1 - \varepsilon S) \tag{3.27}$$

The photon loss and storage are modeled as a resistor $R_{ph}$ and capacitance $C_{ph}$ defined as

$$R_{ph} = \frac{\tau_p}{qV} \tag{3.28}$$

$$C_{ph} = qV \tag{3.29}$$

Using the above definitions, the rate Equations (3.15) and (3.16) become

$$\tau_n \frac{dI_{spon}}{dt} = I - I_{spon} - I_{stim} \tag{3.30}$$

$$C_{ph}\frac{dS}{dt} = I_{stim} - \frac{S}{R_{ph}} + \gamma_{sp} I_{spon} \tag{3.31}$$

Notice that the diode shown in Figure 3.10 models the spontaneous recombination current. The stimulated emission current as well the other current sources in Figure 3.10 model *dN/dt* in Equation (3.15). The stimulated emission current as well as the spontaneous emission coupled into the lasing mode is modeled by two different current sources.

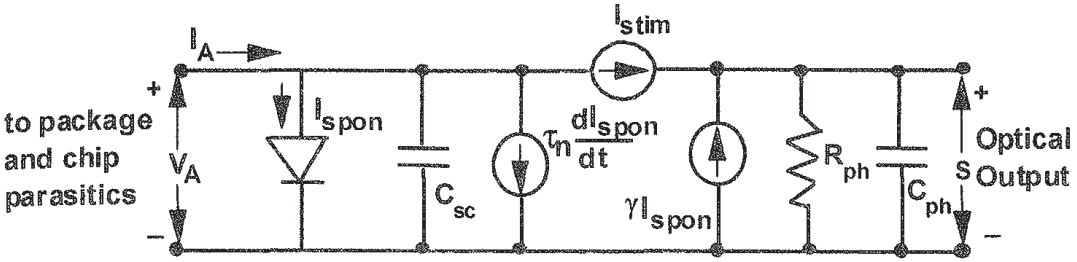



Figure 3.10 Large-signal circuit model of intrinsic laser. After Ref. [14] (© IEEE 1985).

Direct amplitude modulation of a semiconductor laser is always accompanied by a phase modulation. The variation of the electron density ($N$) in the active region during the current modulation also produces a change in the refractive index in the gain medium. The parameter $\alpha$, which is called the linewidth enhancement factor, is defined as [18, 19]

$$\alpha = -\frac{4\pi}{\lambda}\left[\frac{\left(\frac{dn}{dN}\right)}{\left(\frac{dg}{dN}\right)}\right] \tag{3.32}$$

which is the ratio of the changes in the real part to the imaginary part of the refractive index. Thus, as the optical power varies with the current modulation, it produces changes in both the real and imaginary parts of the refractive index, dynamically shifting the optical frequency of the longitudinal mode. The so-called frequency chirp is approximately related to the time-dependent optical power as [20]

$$\Delta\nu = -\frac{\alpha}{4\pi}\left[\frac{d}{dt}\ln P(t) + \kappa P(t)\right] \tag{3.33}$$

where $\kappa = 2\Gamma\varepsilon/V\eta_d h\nu$. According to Equation (3.33), the frequency chirp consists of a "transient chirp" (first term) that arises from the frequency relaxation oscillation during the onset of the laser pulse, and an "adiabatic chirp" (second term) that arises from the damping of the relaxation oscillation. The result is a wavelength shift between the high and low power points in the optical waveform. At high modulation frequencies, lasers with strong damping are likely to have good chirp characteristics due to the reduction in the rate of change of *ln[P(t)]*. However, excessive damping may lead to a reduced modulation bandwidth. Consequently, the laser designer must compromise between the amount of damping of the relaxation oscillation and the "transient chip." As we will see in Chapter 9, the frequency chirp of directly modulated DFB lasers has important implications for cable TV transmission. The interaction between the frequency chirp of DFB lasers and the fiber chromatic dispersion produces an unacceptable amount of nonlinear distortions, limiting the transmission distance that can be used.

## 3.4 Noise in Laser Diodes

The subject of noise in semiconductor lasers has an important implication for cable TV transmission. The following sections will discuss the two fundamental noise mechanisms in the laser output, namely, its intensity and phase noise. The laser intensity and phase noise can be analyzed by adding the so-called *Langevin* noise sources to the single-mode rate equations [21, 22]. The purpose of the following subsections is to summarize some of the important results and their implications for laser transmitters in cable TV networks.

### 3.4.1 Relative Intensity Noise (RIN)

The output power of semiconductor laser fluctuates around its steady-state value due to quantum fluctuations in the electron density as well as spontaneous emission events that are converted to intensity noise. The statistical properties of the laser noise can be studied by adding to each of the rate equations the corresponding Langevin noise source. Assuming a spatially uniform laser, which is driven by Langevin forces $F_N$ and $F_S$ for the carrier density and photon density, respectively, the stochastic rate equations can be written as follows [21, 22]:

$$\begin{aligned} \dot{N} &= I/qV - G(N)S - \gamma_e N + F_N(t) \\ \dot{S} &= G(N)S - \gamma_p S + R_{sp} + F_S(t) \end{aligned} \tag{3.34}$$

where $\gamma_e = 1/\tau_N$, $\gamma_p = 1/\tau_p$, and $R_{sp} = \gamma_{sp} n_{sp} N/\tau_N$ is the spontaneous emission rate. The solution to the stochastic rate equations is considerably simplified if the Langevin forces are assumed to be Gaussian random processes with zero mean, and under the Markovian approximation satisfies the following relation [21]:

$$\left\langle F_i(t)F_j(t')\right\rangle = 2D_{ij}\delta(t-t') \tag{3.35}$$

where the angle brackets denote ensemble average, and $D_{ij}$ are the corresponding diffusion coefficients. It can be shown using a rigorous analysis that the nonzero diffusion coefficents are $D_{SS} = R_{sp}S$, and $D_{NN} = R_{sp}S + \gamma_N N$ [2], where $N$ and $S$ are the steady-state values. With these assumptions, the stochastic rate equations can be solved in the frequency domain using Fourier analysis. The laser relative intensity noise (RIN) is defined as [2]

$$RIN = \frac{\left\langle |\delta S(\omega)|^2 \right\rangle}{S^2} \tag{3.36}$$

where $\delta S$ is the output power fluctuation from the average power value $S$. The laser RIN can be written as [2]

$$RIN(\omega) = \frac{2R_{sp}\left[\Gamma_N^2 + \omega^2 + G_N^2 S^2\left(1+\dfrac{\gamma_N N}{R_{sp}S}\right)\right]}{S\left[\left(\omega_R^2 - \omega^2\right)^2 + 2\Gamma_R^2\left(\omega_R^2 + \omega^2\right) + \Gamma_R^4\right]} \tag{3.37}$$

where $\Gamma_R \equiv (\Gamma_N + \Gamma_S)/2$ is the relaxation oscillation decay rate, $\Gamma_N = \gamma_N + N(\partial\gamma_N/\partial N) + G_N S$ is the small-signal decay rate, $\Gamma_S = R_{sp}/S - G_S S$ is the photon decay rate, $G_N = \Gamma v_g[\partial g/\partial N]$, $G_S = \Gamma v_g[\partial g/\partial S]$. Figure 3.11 illustrates, for example, the RIN spectra of a self-pulsating ridge-waveguide QW laser with a 15-μm absorber at bias currents of 48% and 90% above threshold [23]. At low frequencies ($\omega << \omega_R$), the laser RIN is almost frequency independent, but it is significantly enhanced in the vicinity of $\omega = \omega_R$. At a given frequency, the RIN decreases with the bias current as $(I-I_{th})^{-3}$. As the bias current is increased, the RIN decreases more slowly as $1/(I-I_{th})$. As we will discuss in Chapter 5, the laser RIN imposes an upper limit on the maximum achievable CNR at the fiber node receiver. Consequently, the RIN of DFB laser transmitters, which are used for analog video transmission, are typically equal to –155 dB/Hz or better. It should be pointed out that the laser RIN can significantly degraded by multiple optical reflections [24]. This issue is discussed in detail in Section 9.3.

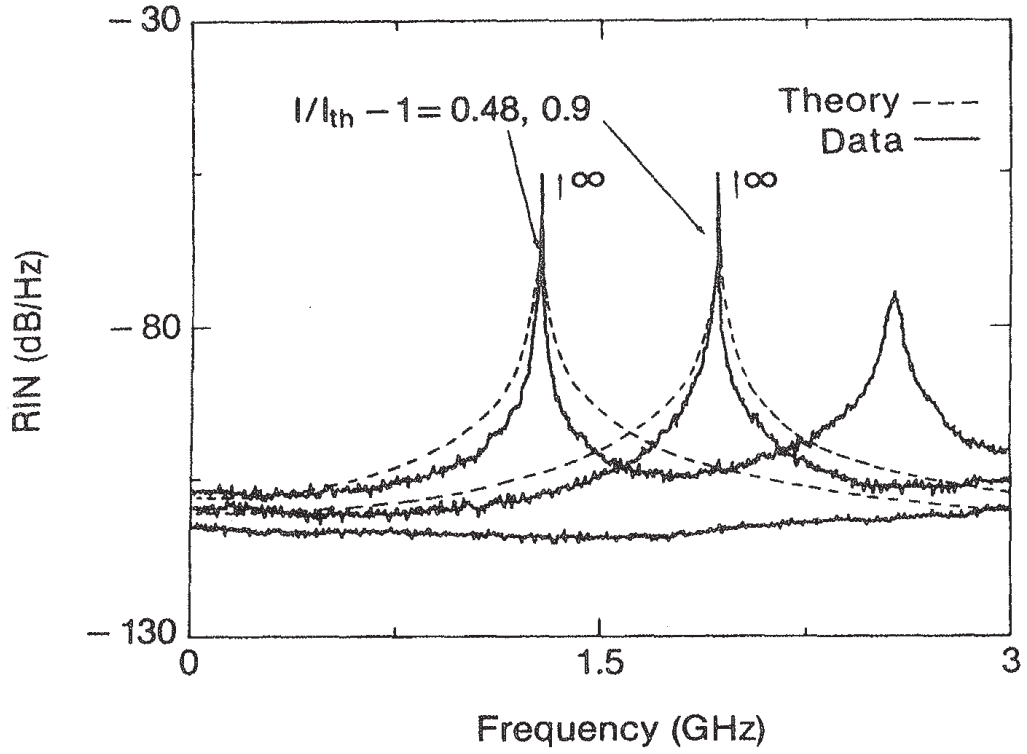



Figure 3.11 Measured and calculated RIN spectra of a self-pulsating ridge-waveguide QW laser with a 15-μm absorber at bias currents of 48% and 90% above threshold. The bottom trace is the instrument noise floor spectra. After Ref. [23] (© 1992 IEEE).

### 3.4.2 Laser Phase Noise

It is well known that spontaneous emission events in the laser cavity change both the phase and amplitude of the optical field. Coupling of the spontaneous emission into the lasing modes as well as fluctuation in the electron density induce changes in both the real and imaginary parts of the refractive index, and produce phase noise. Following the theoretical work by C. Henry, the spectral linewidth of a semiconductor laser due to spontaneous emission can be written as [18, 19, 25]

$$\Delta\nu = \frac{v_g \cdot h\nu \cdot n_{sp}(\alpha_i + \alpha_m)\alpha_m(1+\alpha^2)}{8\pi S} \tag{3.38}$$

where $\alpha_i$ and $\alpha_m$ are internal and mirror losses, respectively; $\alpha$ is the linewidth enhancement factor given by Equation (3.32), and $n_{sp}$ is the spontaneous emission factor. According to Equation (3.38), the laser linewidth is inversely proportional to the output power and it is enhanced by the factor $(1 + \alpha^2)$. The factor $\alpha$ is about 2 for QW lasers and about 4–7 for InGaAsP lasers. The smaller $\alpha$ factor is attributed to the quantum size effect in MQW lasers [26]. The unmodulated or CW linewidth of semiconductor DFB lasers is typically in the range of 0.5 MHz to 10 MHz [19]. When a DFB laser is directly modulated, the modulated linewidth, which is typically greater than 1 GHz, is considerably larger than the CW linewidth. The larger DFB laser linewidth is undesirable, particularly for relatively long-distance transmission (> 30-km).

## 3.5 DFB Laser Transmitter

### 3.5.1 DFB Laser Transmitter System Design

In order to design a DFB laser transmitter for cable TV applications, the laser diode must be properly packaged, coupled into an SMF, and mounted with additional components and circuitry. Figure 3.12 shows the three primary subsystems of a typical DFB laser transmitter.

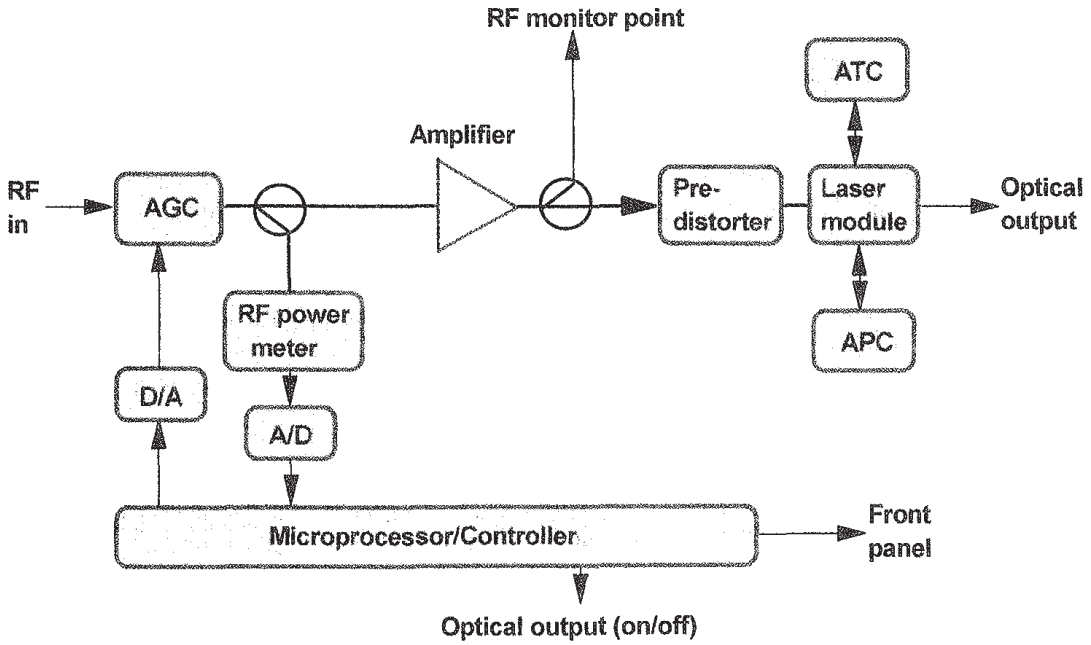



Figure 3.12 Basic block diagram of a directly modulated DFB laser transmitter.

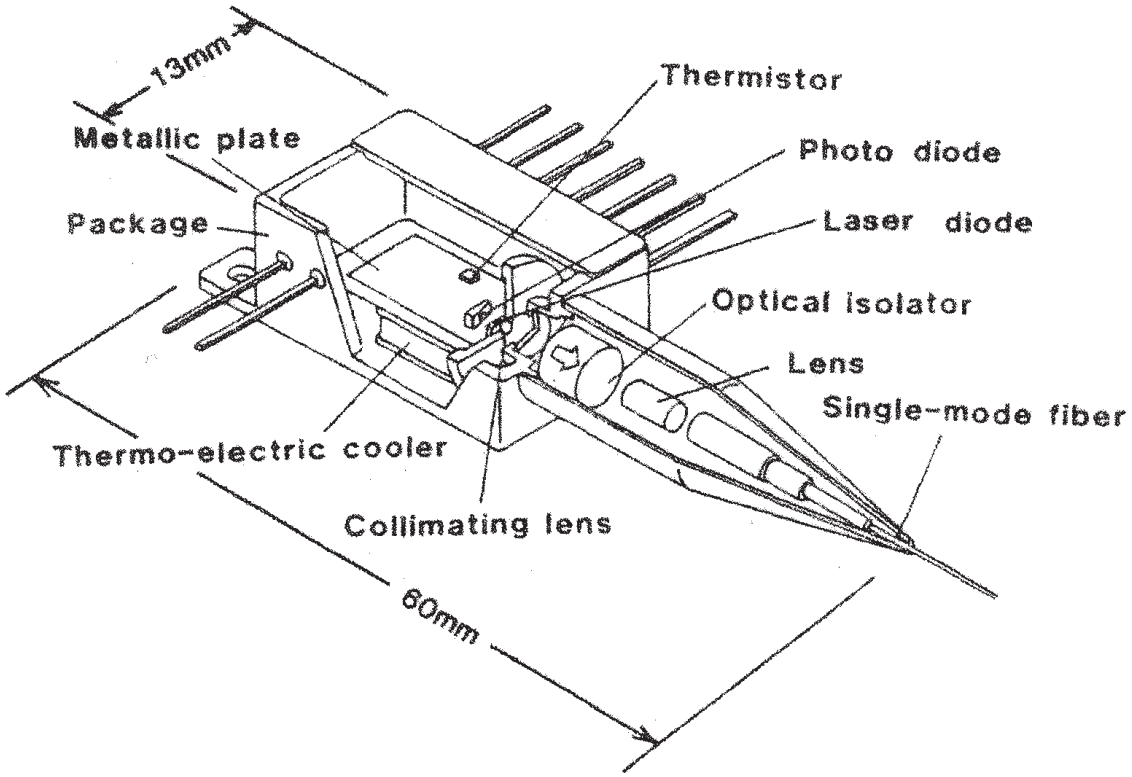



Figure 3.13 Three dimensional view of a DFB laser diode module configuration with a single-mode fiber pigtail. After Ref. [27] (© 1990 IEEE).

The front-end RF section functions to amplify the RF input signal to the proper amplitude before modulating the laser diode. It consists of one or more RF preamplifiers with a variable attenuator and RF power detector. The laser diode section contains the packaged laser diode module with all the necessary circuitry to control its operating point. This includes the automatic gain control (AGC), linearization circuitry, automatic power control (APC), and automatic temperature control (ATC). The role of the AGC circuitry, which is controlled by the microprocessor, is to maintain the RF input level per channel to a fixed level for an optimum operation. It turns out that this feature is desired by cable operators since with no AGC, the laser transmitter performance can be severely degraded, depending on the RF input level. The linearization or predistortion circuitry suppresses the generated nonlinear distortions by the laser module in order to meet the transmission requirements at the fiber node (see Section 3.5.1). The ATC circuitry controls the thermoelectric cooler (TEC), which is necessary for stable temperature operation of the DFB laser (see Section 3.5.3).

Finally, the microprocessor/controller subsystem controls all the critical and auxiliary functions within the laser transmitter, including the status of alarm indicators, front-panel display. To maintain the laser transmitter peak performance, the microprocessor can access the optimum operating point, which is typically stored in a nonvolatile memory, and adjusts

the gain of the RF signal driving the laser. In addition, the microprocessor can report the laser transmitter status to an external status monitoring system. The most important and expensive part of the laser transmitter is the packaged DFB laser diode module. Figure 3.13 shows, for example, a three-dimensional schematic view of a DFB laser diode mounted in a hermetically sealed butterfly package with an SMF pigtail [27]. The DFB laser diode, which is mounted on a metallic plate with a back-facet photodiode, is in direct contact with the TEC. Collimating optics is typically used in front of the laser to provide efficient coupling to the SMF.

### 3.5.2 Optical Isolators

The packaged laser module often includes another important component, namely, optical isolators. Optical isolators are passive, nonreciprocal optical devices based on the Faraday effect. In 1842, Michael Faraday discovered that the plane of polarized light is rotated while transmitting through glass that is contained in a magnetic field. The direction of rotation is independent of the direction of the light propagation, and only dependent on the direction of the magnetic field. Generally speaking, optical isolators can be divided into two categories,

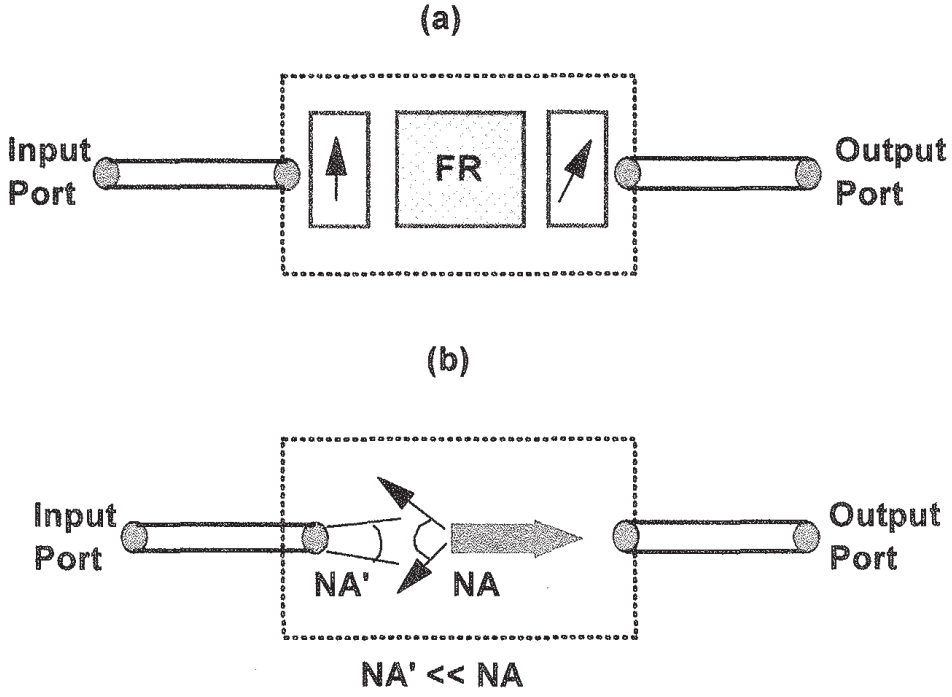



Figure 3.14 Schematic configurations of (a) polarization-sensitive optical isolators, and (b) polarization-insensitive optical isolators. FR = Faraday rotator, NA = numerical aperture. After Ref. [28] (© 1994 IEEE).

polarization-sensitive and polarization-insensitive [28]. Figure 3.14 illustrates the operating principles of the two types of optical isolators, which will be referred to as type "A" and type "B" [29]. In a type A optical isolator, shown in Figure 3.14(a), a Faraday rotator with a cylindrical magnet, which applies the magnetic field in the light direction, is sandwiched between a polarizer and an analyzer. When a light beam is passed through the Faraday rotator, its polarization plane is rotated by 45°. Since the polarization plane of the analyzer is already rotated by 45° relative to the polarizer, the propagating beam passes through the isolator with almost no attenuation loss. If a reflected light with an arbitrary polarization is incident on the analyzer, only the reflected light, which matches the polarization of the analyzer, will pass through it. Then, the Faraday rotator will rotate the polarization plane by 45° in the opposite direction, resulting in a 90° rotation relative to the polarizer. Thus, the polarizer blocks the backward-reflected light. The primary disadvantages of a type A isolator are its polarization sensitivity, which may result in additional insertion losses if the input light is randomly polarized, and its packaging reliability issues.

A type B optical isolator, which is polarization insensitive, consists of a Faraday rotator sandwiched between two wedge-shaped birefringent prisms with a collimating lens on each side. When an input light with an arbitrary polarization in incident on the first birefringent wedge-shaped prism, it is separated into ordinary and extraordinary beams caused by the different refractive indices of the prism. Since the second birefringent prism is already aligned to match the 45° rotation by the Faraday rotator, the ordinary and extraordinary beams become parallel, and the output light is coupled to the fiber with minimal attenuation loss. The backward-reflected light is also separated into ordinary and extraordinary beams, which are rotated 45° by the Faraday rotator in the opposite direction. After the second birefringent prism, the existing backward-reflected ordinary and extraordinary beams are diverging with a large angle between them, and thus are not coupled into the input fiber.

For operation in the 1.3 μm and 1.55 μm, garnet crystals such as *yttrium-iron-garnet* (YIG) and bismuth-substituted rare-earth iron garnet (BIG) are most commonly used for high-performance isolators [28]. Single-stage commercial isolators typically provide about 40-dB optical isolation and as much as 60-dB for dual-stage configurations with an insertion loss less than 0.3-dB (single stage).

### 3.5.3 Thermoelectric Cooler (TEC) Design and Operation

A TEC module is a small solid-state device that operates as a heat pump, utilizing the so-called Peltier effect, which was discovered in 1834, to move heat [30]. Generally speaking, when electrical current passes through the junction of two different types of conductors, a temperature gradient is produced. The Peltier effect requires the development of semiconductors that are good electrical conductors but poor conductors of heat. For example, heavily doped n-type and p-type bismuth telluride is primarily used as the semiconductor material for the TEC. Figure 3.15 illustrates a schematic configuration of a typical TEC assembly. The TEC consists of a number of p-type and n-type pairs (couples) connected

electrically in series and sandwiched between two ceramic plates. The purpose of the ceramic plates on each side of the TEC elements is to prevent shorting of the laser diode by the TEC circuit. When connected to a DC power supply, the current causes the heat to move from one side of the TEC to the other, creating a "hot" side and a "cold" side on the TEC. In a typical application, the cold side of the TEC is exposed to the laser diode to be cooled and the hot side is exposed to the heat sink, which dissipates the heat to the environment. The dissipated heat at the hot side ($Q_H$) is simply the sum of the removed heat from the cold side ($Q_C$) and the input power to the TEC, which can be written as

$$Q_H = Q_C + I_{TEC}^2 R \tag{3.39}$$

where R is the total resistance of the serially connected TEC elements. To achieve temperature stability over the operation of the laser diode, a closed-loop temperature control is needed. In this approach, the controlled cold-side temperature is compared to some reference temperature, which is typically the ambient or the hot-side temperature.

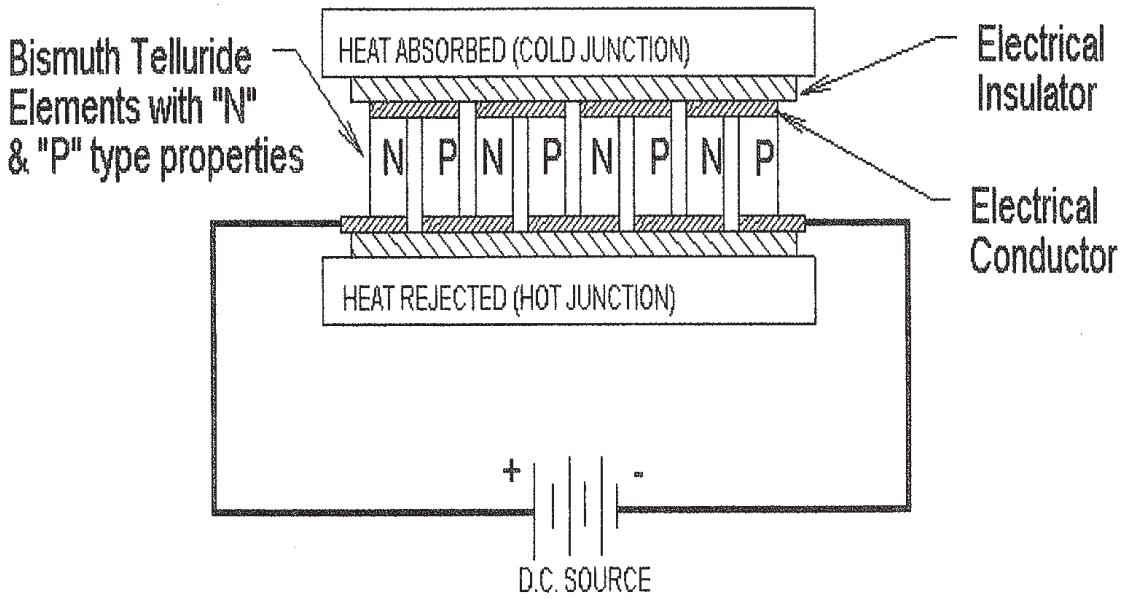



Figure 3.15 A schematic of typical thermoelectric module assembly (courtesy of Melcor Corporation). Note that the bismuth telluride elements are electrically in series and thermally in parallel.

### 3.5.4 Linearization Methods

As mentioned before, the nonlinear distortions generated by the semiconductor laser are often too large to meet the performance requirements for cable TV networks. Employing a predistortion or linearization circuitry before the laser module as shown in Figure 3.12 typically solves this problem. There are several linearization or predistortion methods to suppress the effect of CSO and CTB distortions in DM-DFB laser transmitters [31,32]. Figure 3.16 shows, for example, a block diagram of an electronic predistortion circuit to minimize the effect of second- and third-order nonlinear distortions [31]. The RF input signal is split into two paths with the primary path applied directly to the DFB laser with some time delay to compensate for the delays in the secondary path. A predistortion circuit in the secondary path generates second- and third-order distortions with equal amplitude and opposite phase to the nonlinear distortions generated by the DFB laser at the desired operating point. The phases of signals in the secondary path are synchronized before they are recombined with the delayed RF signal at the primary path to produce the modulating signal with the nonlinear distortions. A push-pull amplifier with unbalanced bias current can be used, for example, to generate the required even and odd harmonic distortions. There are other types of linearization methods such as a feedforward linearization and an optical linearization, which are sometimes used for an EM-DFB laser transmitter (see Sections 4.3 and 4.4).

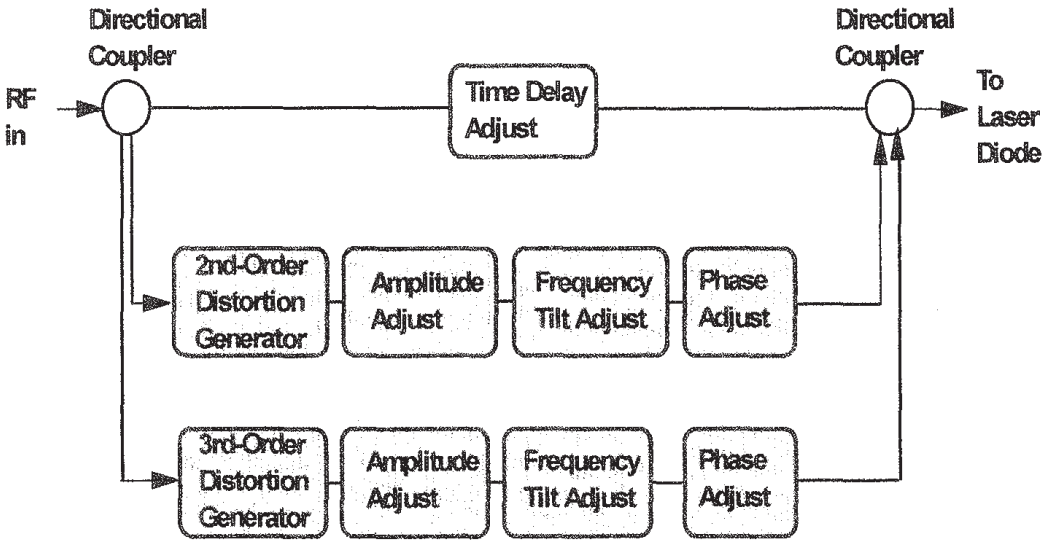



Figure 3.16 Block diagram of a predistortion circuit that minimizes the effect of second- and third-order distortions in DFB laser transmitters. After Ref. [31].

### 3.5.5 DFB Laser Transmitter Performance Requirements

In order for cable operators to transmit downstream multiple AM/QAM video channels over their networks, the DFB laser transmitter must be able to comply with cable TV transport requirements as well as environmental conditions [33]. In particular, the DFB laser transmitter is required to provide the following:

(A) Single longitudinal mode with high side-mode suppression ratio (> 30-dB)
(B) High-output optical power with high differential quantum efficiency
(C) Low RIN (< −155 dBc/Hz)
(D) Low nonlinear distortions over the cable TV spectrum (50–860 MHz)
(E) Stable operating characteristics over the required temperature and humidity conditions
(F) Small frequency chirp for 1550-nm operation ( < 100 MHz/mA)

Generally speaking, the laser transmitter is designed to transport multiple AM/QAM video signals with the maximum possible CNR for a given optical received power at the fiber node. To achieve the maximum CNR for a given optical link budget, the output optical power as well as the optical modulation index per channel ($m$) must be as large as possible to overcome the laser RIN, fiber losses, and receiver noise. On the other hand, the maximum optical modulation index is limited to avoid unacceptably large nonlinear distortions and a clipping effect [34,35]. This is because when transporting multiple AM-VSB channels, the composite RF input signal is clipped by temporarily driving the laser current below threshold. The static clipping behavior, which will be discussed in detail in Chapter 9, increases the laser's nonlinear distortions and imposes an upper limit on the channel capacity. Consider a SCM system with $N$ channels, each with modulation index m and a total photocurrent at the receiver $I(t)$. For large $N$, the photocurrent $I(t)$ can be modeled as a Gaussian random process with mean $I_p$ and standard deviation $\sigma_p = mI_p(N/2)^{1/2}$. The normalized RMS modulation index can be defined as

$$\mu = \frac{\sigma_p}{I_p} = m \cdot \sqrt{\frac{N}{2}} \tag{3.40}$$

It can be shown that the carrier-to-nonlinear distortions caused by clipping for small μ is approximately given as [36]

$$C/NLD = \frac{1}{\Gamma}\sqrt{\frac{\pi}{2}}\mu^{-3}\left(1+6\mu^2\right)\exp(1/2\mu^2) \tag{3.41}$$

where Γ (≈1/2) is the fraction of the distortion power within the cable TV band. The resultant CNR is the sum of the CNR and the C/NLD according to Equation (2.10). Thus, for a given CNR at the fiber node receiver, the AM-VSB channel capacity ($N$) can be determined

using Equation (3.41). This upper limit of the number of transmitted channels due to nonlinear clipping distortions is sometimes called the Saleh limit [36].

Table 3.2 summarizes the required performance characteristics of a directly modulated DFB laser transmitter for multiple AM/QAM channel downstream transmission over cable TV networks [37]. For example, SL-MQW InGaAsP DFB laser diodes operating at 1310 nm have been shown to meet these performance requirements.

**Table 3.2 Typical characteristics of DFB laser transmitter for multichannel AM/QAM video transmission over cable TV network.**

| Parameter | Specification |
|---|---|
| Optical Wavelength | 1310 nm or 1550 nm ± 10 nm |
| Optical Power | 3 mW to 14 mW |
| Operating Bandwidth | 50 to 860 MHz |
| RF Input Level Range (video signals) | +10 to +25 dBmV per channel |
| Optical Modulation Index | 3 to 4% per channel |
| CNR (for 80 channel loading) | 52 dB (with 0-dBm at the receiver) |
| CSO Distortion | –62 dBc |
| CTB Distortion | –65 dBc |
| Frequency Response Variations | ± 1 dB (50 to 860-MHz) |
| Return Loss | > 16 dB (at RF input) |
| Operating Temperature range | 0° to +50°C |
| Storage Temperature Range | –40° to +70°C |
| Relative Humidity | Maximum 85% noncondensing |

## 3.6 Return-Path Laser Transmitters

So far, we have discussed directly modulated DFB laser transmitters for multiple AM/QAM downstream transmission over cable TV networks. To enable a mass deployment of high-speed multimedia return-path transmitters, low-cost, preferably uncooled, DFB laser transmitters or even FP laser transmitters operating at either 1310-nm or 1550-nm wavelengths are needed [38]. For example, high-performance uncooled 1.3-μm $Al_xGa_yIn_{1-x-y}As$/InP SL-MQW lasers suited for fiber-in-the-loop and return-path applications have been developed recently. The use of the AlGaInAs/InP material system instead of the conventional $Ga_xIn_{1-x}As_yP_{1-y}$/InP material system provides a better carrier confinement under high temperature operation. Preliminary test data of these lasers show a mean-time-to-failure of 9.4 years at operating temperature of 85°C with an output power of more than 5-mW. One of the main problems in these quasi-single-mode or multimode lasers is the observation of mode-partition noise and mode-hopping noise. In the next section, both noise mechanisms are

reviewed, and their impact on return-path laser transmitter design for digital transmission is discussed.

### 3.6.1 Mode-Partition Noise and Mode-Hopping Noise

A mode-partition noise (MPN) arises when the optical power is distributed among many different laser modes such that the instantaneous power in each individual mode can fluctuate, but the total power in all the modes remains relatively unchanged. Figure 3.17 shows the calculated and measured intensity fluctuations of a dominant mode of a multimode AlGaAs laser [39, 40]. The intensity noise spectral density of the dominant mode is essentially unchanged at low frequencies (< 50-MHz), and decreases at 6-dB per octave above the low frequency region. Generally speaking, at the low-frequency region, the laser intensity noise can be very larger and tends to mask the MPN. However, at higher frequencies, say above 1-GHz, the MPN is the dominant noise, masking the laser intensity noise. The reason semiconductor lasers are subjected to MPN is that the different optical modes are competing with each other for the same optical gain. The MPN introduces several problems

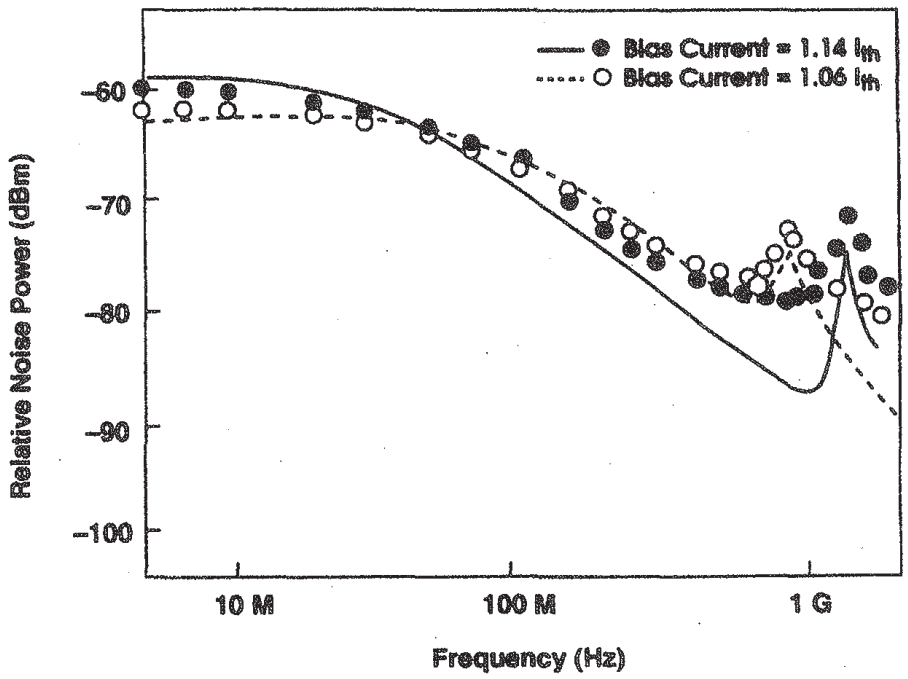



Figure 3.17 Calculated (lines) and Measured (symbols) intensity fluctuations of a dominant mode of a multimode AlGaAs laser. After Ref. [39] (© 1994 IEEE).

that limit the performance of multiple-longitudinal-mode FP lasers. First, the laser's mode distribution was found to vary cyclically with the bias current and temperature, which can dramatically alter the laser transmitter performance [41]. Second, when a multiple longitudinal-mode FP laser is used with a dispersive fiber, the instantaneous power fluctuations of the individual modes become separated in time because each mode experiences a different propagation delay [42]. Consequently, the power fluctuations of the various modes at any given time do not cancel each other. Thus, using multimode FP lasers limits the transmission of video signals in an SCM lightwave system. The MPN problem is usually eliminated with the use of single-longitudinal mode DFB laser diodes.

Another type of noise, which has been observed in multiple modes or quasi-single-mode FP lasers, is called mode-hopping noise (MHN). MHN is attributed to random hopping of among the various oscillating modes of the laser when the laser bias current or temperature is varied. The mode-hopping mechanism can cause a significant intensity noise enhancement (< −130 dB/Hz), particularly at low frequencies (< 30-MHz). Controlling the bias current and/or temperature of the semiconductor lasers can eliminate the mode-hopping mechanism. Another way to suppress the MHN can be achieved by introducing a saturable absorber in index-guided semiconductor lasers, for example, by using Te-doping of the n-type cladding layer of AlGaAs/GaAs lasers [43].

### 3.6.2 Performance Requirements

Directly modulated semiconductor uncooled FP lasers can be used to transport digital channels such as QPSK and/or 16-QAM in the return-path portion of the cable TV networks [37]. The return-path laser transmitters are not required to provide the same CNR or linearity as the downstream laser transmitters since they are not intended to transport AM-VSB video channels. The optical signals can be transmitted at either 1310-nm or 1550-nm. For cost-effective laser transmitters, the TEC or the optical isolator in the laser diode module can be eliminated [44, 45]. Figure 3.18 shows, for example, the measured uncoded BER versus the optical modulation index per channel ($m$) for a differentially encoded QPSK channel transmitted over 24-km of a standard SMF using an uncooled FP laser with a RIN of −127-dBc/Hz. Notice that at room temperature, an error-free QPSK transmission can be achieved over a wide range of optical modulation indices, from about 1% to 11%. The operating range of optical modulation indices is reduced by 5-dB at a higher temperature (≈80°C) indicating an upper limit on an upstream channel capacity. Table 3.3 summaries the performance specifications for return-path laser transmitters. For example, assuming a 256-kb/s QPSK channel, the required upstream C/(N+I) must be equal or greater than 20-dB, which includes a 4-dB margin caused by the simultaneous presence of multiple impairments.

The upstream impairments include a laser RIN, thermal noise, and optical reflections, as well as cumulative ingress noise from all the homes connected to a given fiber node. Since the return-path transmitters are typically installed inside the downstream optical receiver at the fiber node, they are required to operate over a much larger temperature range than the downstream DFB laser transmitters. The performance requirements of the return-path laser

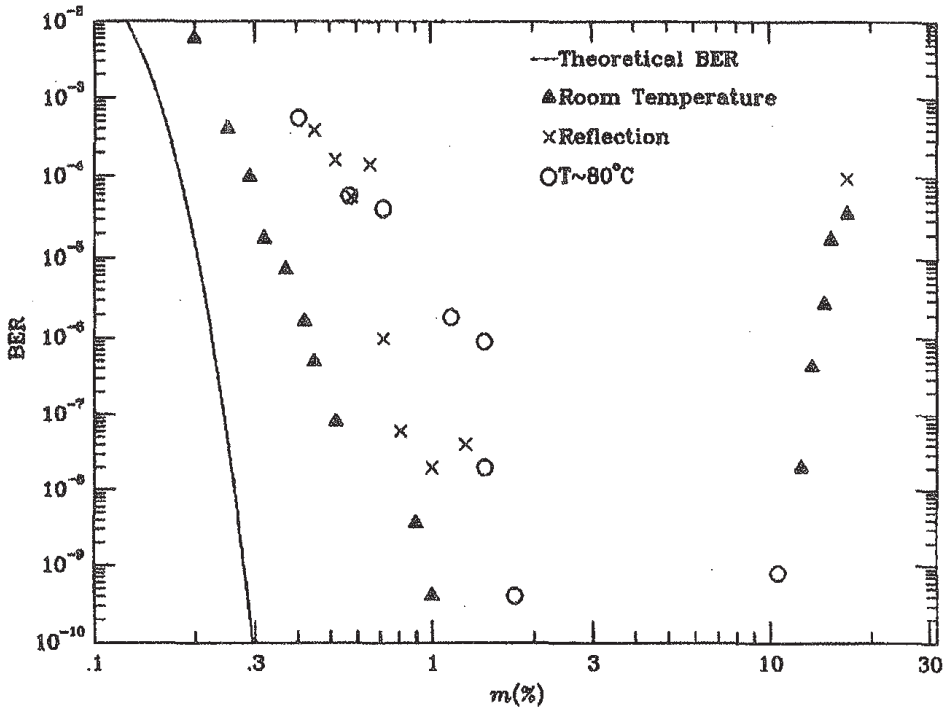



Figure 3.18 BER versus the optical modulation index per channel (m) for differentially encoded QPSK for an uncooled FP laser diode. After Ref. [44] (© IEEE 1996).

**Table 3.3 Typical characteristics for return-path laser transmitters.**

| Parameter | Specification |
|---|---|
| Optical Wavelength | 1310 nm or 1550 nm ± 10-nm |
| Optical Power | 3 to 5 mW |
| Operating Bandwidth | 5 to 42-MHz |
| RF Input Level Range (video signals) | +10 to +25 dBmV per channel |
| Dynamic Range | 15-dB |
| CNR (QPSK channels) | ≥ 16 dB at BER ≤ $10^{-7}$ |
| C/(N+I) QPSK channels | ≥ 20 dB[2] at BER ≤ $10^{-7}$ |
| Laser Transmitter Spurious Output Level | ≤ –25 dBc |
| Return Loss | > 16 dB (at RF input) |
| Operating Temperature range | –20° to +65°C |
| Storage Temperature Range | –40° to +70°C |

[2] Includes the simultaneous presence of all the impairments in the upstream portion of the cable TV network.

| | |
|---|---|
| Relative Humidity | Maximum 85% noncondensing |

transmitters change when the return-path portion of the cable TV access network is digitized, particularly in a DWDM network. For more discussion, see Section 10.4.2.

## References


1. S. M. Sze, *Semiconductor Devices Physics and Technology*, John Wiley & Sons (New York, 1985).
2. G. P. Agrawal and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*, Van Nostrand Reinhold Company (New York, 1986).
3. T. P. Lee, "Recent Advances in Long Wavelength Semiconductor Lasers for Optical Fiber Communications," *Proceedings of IEEE* (1991).
4. N. K. Dutta, "Calculated Absorption, Emission, and Gain in $In_{0.72}Ga_{0.28}As_{0.6}P_{0.4}$," *Journal of Applied Physics* 51, 6095–6100 (1980).
5. N. K. Dutta and R. J. Nelson, *Journal of Applied Physics* 53, 74 (1982).
6. E. Kapon, Semiconductor Lasers: Physics and Technology, lecture notes (1988).
7. K. Oe and K. Sugiyama, *Japanese Journal of Applied Physics* 47, 3533 (1976).
8. K. Oe, S. Ando, and K. Sugiyama, "GaInAsP/InP Planar Stripe Lasers Prepared by Using Sputtered SiO2 Film as a Zn Diffusion Mask," *Journal of Applied Physics* 51, 43–49 (1980).
9. J. Ungar, N. Bar-Chaim, and I. Ury, "High-Power GaAlAs Window Lasers," *Laser and Applications* **9**, 111–114 (1985).
10. K. Uomi, T. Suchiya, M. Komori, A. Oka, K. Shinoda, and A. Oishi, "Extremely Low Threshold (0.56 mA) Operation in 1.3-μm InGaAsP/InP Compressive-Strained-MQW Lasers," *IEE Electronic Letters* **30**, 2037–2038 (1994).
11. H. Nobuhara, T. Inoue, T. Watanabe, K. Takana, T. Odagawa, T. Abe, and K. Wakao, "1.3-μm Wavelength, Low-Threshold Strained Quantum Well Laser on p-Type Substrate," *IEE Electronic Letters* **30**, 1292–1293 (1994).
12. G. P. Agrawal, "Effects of Gain and Index Nonlinearities on Single-Mode Dynamics in Semiconductor Lasers," *IEEE Journal of Quantum Electronics* **26**, 1901 (1990).
13. M. Asada and Y. Suematsu, "Density Matrix Theory of Semiconductor Lasers with Relaxation Broadening Modal-Gain and Gain-Suppression in Semiconductor Lasers," *IEEE Journal of Quantum Electronics* **QE-21**, 434–442 (1985).
14. R. S. Tucker, "High-Speed Modulation of Semiconductor Lasers," *IEEE Journal of Lightwave Technology* **LT-3**, 1180–1192 (1985).

15. K. Y. Lau and A. Yariv, "Ultra-High Speed Semiconductor Lasers," *IEEE Journal of Quantum Electronics* **QE-21**, 121–138 (1985).
16. P. L. Derry, T. R. Chen, Y. H. Zhuang, J. Paslaski, M. Mittelstein, K. Vahala, and A. Yariv, "Spectral and Dynamic Characteristics of Buried-Heterostructure Single Quantum Well (Al,Ga) As Lasers," *Applied Physics Letters* **53**, 271–273 (1988).
17. I. Habermayer, "Nonlinear Circuit Model for Semiconductor Lasers," *IEEE Journal of Optical and Quantum Electronics* **13**, 461–468 (1984).
18. C. H. Henry, "Theory of the Linewidth of Semiconductor Lasers," *IEEE Journal of Quantum Electronics* **QE-18**, 259–264 (1982).
19. M. Osinski and J. Buus, "Linewidth Broadening Factor in Semiconductor Lasers - An Overview," *IEEE Journal of Quantum Electronics* **QE-23**, 9–29 (1987).
20. T. L. Koch and R. A. Linke, "Effect of Nonlinear Gain reduction on Semiconductor Laser Wavelength Chirping," *Applied Physics Letters* **48**, 613–615 (1986).
21. K. Vahala and A. Yariv, "Semiclassical Theory of Noise in Semiconductor Lasers-Part I," *IEEE Journal of Quantum Electronics* **QE-19**, 1096–1101 (1983).
22. K. Vahala and A. Yariv, "Semiclassical Theory of Noise in Semiconductor Lasers-Part II," *IEEE Journal of Quantum Electronics* **QE-19**, 1102–1109 (1983).
23. S. Ovadia and K. Y. Lau, "Low-frequency Relative Intensity Noise in Self-Pulsating Ridge-Waveguide Quantum Well Lasers," *IEEE Photonics Technology Letters* **4**, 336–338 (1992).
24. W. I. Way et al., "Multiple-Reflection-Induced Intensity Noise Studies in a Lightwave System for Multichannel AM-VSB Television Signal Distribution," *IEEE Photonics Technology Letters* **2**, 360–362 (1990).
25. C. H. Henry, "Theory of the Phase Noise and Power Spectrum of a Single Mode Injection Laser," *IEEE Journal of Quantum Electronics* **QE-19**, 1391–1397 (1983).
26. K. Uomi, S. Sasaki, T. Tsuchiya, M. Okai, M. Aoki, and N. Chinone, "Spectral Linewidth Reduction by Low Spatial Hole Burning in 1.5 μm Multi-Quantum Well λ/4 Shifted DFB Lasers," *IEE Electronics Letters* **26**, 52–53 (1990).
27. A. Takemoto, H. Watanabe, Y. Nakajima, Y. Sakakibara, S. Kakimoto, J. Yamashita, T. Hatta, and Y. Miyake, "Distributed Feedback Laser Diode and Module for CATV Systems," *IEEE Journal of Selected Areas in Communications* **8**, 1359–1364 (1990).
28. K. Kikushima, K.-I. Suto, H. Yoshinaga, and E. Yoneda, "Polarization Dependent Distortion in AM-SCM Video Transmission Systems," *IEEE Journal of Lightwave Technology* **12**, 650–657 (1994).
29. SR-NWT-002855, *Optical Isolators: Reliability Issues and Proposed Tests*, Issue 1 (Bellcore, December 1993).
30. P. W. Shumate, "Semiconductor Laser Transmitters," *Optoelectronic Technology and Lightwave Communication Systems*, Ed. Chinlon Lin, Van Nostrand-Reinhold (New Jersey, 1990).
31. H. A. Blauvelt and H. L. Loboda, "Predistorter for Linearization of Electronic and Optical Signals," U.S. Patent Number 4,992,754, February (1991).

**32.** M. Nazarathy, C. H. Gall, and C. Y. Kuo, "Predistorter for High frequency Optical Communications Devices," U.S. Patent Number 5,424,680, June (1995).

**33.** H. Yonetani, I. Ushijima, T. Takada, and K. Shima, "Transmission Characteristics of DFB Laser Modules for Analog Applications," *IEEE Journal of Lightwave Technology* **11**, 147–153 (1993).

**34.** C. Y. Kou, "Fundamental Second-order Nonlinear Distortions in Analog AM CATV transport Systems Based on Single Frequency Semiconductor Lasers," *IEEE Journal of Lightwave Technology* **10**, 235–243 (1992).

**35.** K. Maeda, H. Nakata, and K. Fujito, "Analysis of BER of 16QAM Signal in AM/16QAM Optical Transmission System," *IEE Electronics Letters* **29**, 640–641 (1993).

**36.** A. A. M. Saleh, "Fundamental Limit on Number of Channels in Subcarrier Multiplexed Lightwave CATV Systems," *IEE Electronics Letters* **25**, 776–777 (1989).

**37.** GR-2853-CORE, *Generic Requirements for AM/Digital Video Laser Transmitters and Receivers*, Issue 2 (Bellcore, December 1995).

**38.** C. E. Zah, R. Bhat, B. N. Pathak, F. Favire, W. Lin, M. C. Wang, N. C. Andreadakis, D. M. Hwang, M. A. Koza, T. P. Lee, Z. Wang, D. Darby, D. Flanders, and J. J. Hsieh, "High Performance Uncooled 1.3-μm $Al_xGa_yIn_{1-x-y}As/InP$ Strained-Layer Quantum-Well Lasers for Subscriber Loop Applications," *IEEE Journal of Quantum Electronics* **30**, 511–523 (1994).

**39.** G. J. Meslener, "Mode-Partition Noise in Microwave Subcarrier transmission Systems," *IEEE Journal of Lightwave Technology* **12**, 118-126 (1994).

**40.** T. Ito, S. Machida, K. Nawata, and T. Ikegami, "Intensity Fluctuations in Each Longitudinal Mode of a Multimode AlGaAs Laser," *IEEE Journal of Quantum Electronics* **13**, 574–579 (1977).

**41.** G. J. Meslener, "Temperature Dependence of Distribution, Intensity Noise, and Mode-Partition Noise in Subcarrier Multiplexed Transmission System," *IEEE Photonics Technology Letters* **4**, 939–941 (1992).

**42.** R. H. Wentworth, G. E. Bodeep, and T. E. Darcie, "Laser Mode Partition Noise in Lightwave Systems Using Dispersive Optical Fiber," *IEEE Journal of Lightwave Technology* **10**, 84–89 (1992).

**43.** N. Chinone, T. Kuroda, T. Ohtoshi, T. Takahashi, and T. Kajimura, "Mode Hopping Noise in Index-Guided Semiconductor and its Reduction by Saturable Absorbers," *IEEE Journal of Quantum Electronics* **21**, 1264–1270 (1985).

**44.** S. L. Woodward and G. E. Bodeep, "Uncooled Fabry-Perot Lasers for QPSK Transmission," *IEEE Photonics Technology Letters* **7**, 558–560 (1995).

**45.** S. L. Woodward, V. Swaminathan, G. E. Bodeep, and A. K. Singh, "Transmission of QPSK Signals Using Unisolated DFB Lasers," *IEEE Photonics Technology Letters* **8**, 127–129 (1996).

# CHAPTER 4

# EXTERNALLY MODULATED CABLE TV LIGHTWAVE LASER TRANSMITTERS

Direct modulation of a laser diode has several limitations, depending on the operating wavelength. One of the key limitations is the induced nonlinear distortion due to the interaction between the frequency chirp of directly modulated DFB laser transmitters operating at 1550-nm and the dispersion of standard SMF (see Chapter 9). Transmission at 1550-nm wavelength band is desirable in order to take advantage of the low fiber losses and the use of Erbium-doped fiber amplifiers (EDFAs). The higher required optical power from DFB laser transmitters operating at 1310-nm compared with the 1550-nm due to higher SMF losses imposes another limitation. In addition, the performance of directly modulated DFB laser transmitters can also be severely degraded in the presence of multiple optical reflections (see Chapter 9). These problems, which significantly limit the transmission distance over standard SMF, can be overcome by using an externally modulated laser transmitter with an optical modulator.

We shall start Chapter 4 with a review of the basic operation and distortion characteristics of $LiNbO_3$ optical modulators, which are the most widely used in the cable TV industry. In Section 4.2, the different electronic linearization methods and circuits are reviewed. There are also optical linearization methods, including optical dual parallel linearization and optical dual cascade linearization, which are discussed in Section 4.3. Finally, we will discuss the design and performance characteristics of the two different types of externally modulated laser transmitters in Section 4.4.

There are two different types of commercially available externally modulated laser transmitters operating at 1.3-μm and 1.55-μm. The first type is externally modulated diode-pumped Nd:YAG laser operating at 1319-nm, which has many desirable characteristics such as extremely narrow linewidth (< 200-kHz) and very low laser RIN (< −165-dBc/Hz) [3]. The second type is an externally modulated DFB laser transmitter operating at 1550-nm. Since the mid-1990s, high-power (> 20 mW) with low RIN (<−155-dBc/Hz) SL-MQW DFB laser diodes operating 1550-nm have been developed to meet the cable TV performance requirements. The early generation of DFB laser diodes had limited output power levels, which were reduced due to the insertion loss of the $LiNbO_3$ modulator. Consequently, EDFAs were incorporated in the laser transmitter design. Chapter 6 is devoted to the design, operation, and characteristics of EDFAs, since they are one of the key building blocks of cable TV networks.

## 4.1 $LiNbO_3$ Optical Modulators

Optical modulators such as Lithium Niobate ($LiNbO_3$) modulators are one of the key components in the externally modulated laser transmitter. Consequently, their basic operation, design, and distortion characteristics are discussed in this section.

### 4.1.1 Basic Operation of $LiNbO_3$ Intensity Modulators

Lithium Niobate ($LiNbO_3$) is an important dielectric material for many applications in optical communications, optical sensors, and signal processing due to its unique physical properties [1]. One of the most important properties of $LiNbO_3$ is its large electro-optic coefficient. The linear electro-optic effect, which is also called the Pockels effect, is a change in the material refractive index when an electric field is applied [2]. Figure 4.1 shows schematic cross-sections of a ridge-type $LiNbO_3$ waveguide modulator with the relative orientation of the electrodes to the waveguide. When a voltage $V$ is applied to the electrodes, which are placed alongside the waveguide with a gap $D$, an internal electric field is created whose magnitude is approximately given by: $|E| = V/D$. If the electric field is applied in the z-direction ($E_z$), one can obtain the largest electro-optic induced refractive index change, which is given by

$$\Delta n \cong -n^3 r_{33} \cdot \Gamma\left[\frac{V}{2D}\right] \tag{4.1}$$

where $n$ is either the ordinary or the extraordinary refractive index of the waveguide, $r_{33}$ is the largest electro-optic coefficient, and $\Gamma$ is the overlap integral between the RF and optical mode fields. To generate $E_z$, the electrode orientation relative to the waveguide depends on the $LiNbO_3$ crystal orientation. The crystal orientation is often specified as a "cut," which is perpendicular to the flat surface on which the waveguide is fabricated. For example, in X-cut $LiNbO_3$, the electric field is propagating in the y direction along the waveguide as shown in Figure 4.1(a). The resulting phase shift from the refractive index change over the effective waveguide length is given by

$$\Delta\phi = \frac{2\pi \cdot \Delta n \cdot L}{\lambda} = -\pi n^3 r_{33} \cdot \Gamma\left[\frac{V \cdot L}{\lambda D}\right] \tag{4.2}$$

In order to obtain the largest phase shift, one needs to optimize the waveguide geometry, namely, minimizing the lateral waveguide parameter $D/\Gamma$ and maximizing the waveguide length. However, if the electrode gap, which is typically in the range of 10–20-μm, is too small, the electrode capacitance may be too large, resulting in a reduced modulation bandwidth and an increased amount of DC drift [3]. It should be pointed out that both the driving

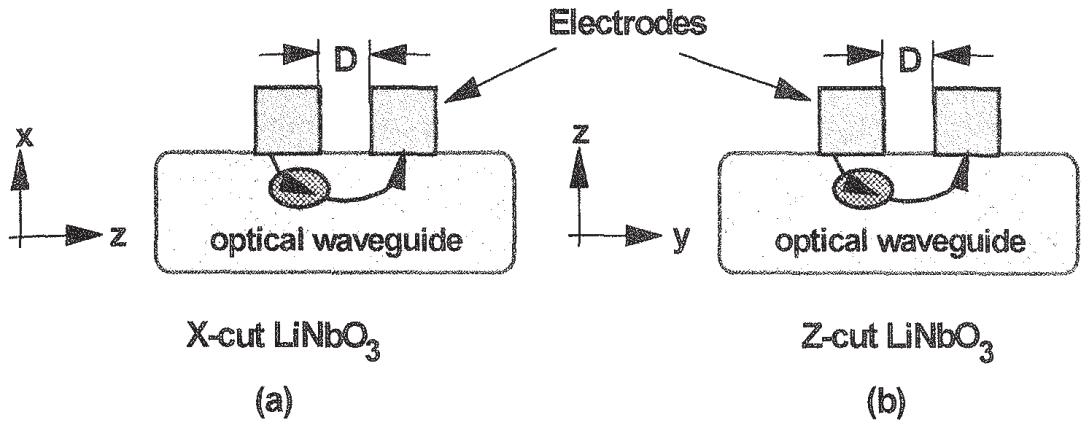



Figure 4.1 Schematic cross sections of ridge-type $LiNbO_3$ waveguide with the corresponding electrodes with their orientation for (a) X-cut, and (b) Z-cut $LiNbO_3$ waveguides.

voltage and the modulation bandwidth are proportional to $(1/L)$. Since cable TV networks typically operate below 1-GHz, reducing the driving voltage is very desirable. The required voltage to switch the optical output from on to off, which is also called *half-wave voltage* ($V_\pi$), is typically in the range of 3–5 V [3, 4].

Generally speaking, integrated optical modulators are constructed by patterning optical waveguides onto the $LiN_bO_3$ substrate using standard photolithography techniques, and depositing the metal electrodes on top of the waveguide. The two primary methods to fabricate the optical waveguides on the $LiNbO_3$ substrate are annealed proton exchange (APE) and titanium indiffusion [4, 5]. Although both methods produce waveguides with somewhat similar characteristics, the recently developed APE has been more widely adopted for the following reasons. First, the APE waveguide supports only a single polarization light because the APE process increases the extraordinary refractive index ($n_e$), while the ordinary refractive index ($n_o$) decreases slightly. Typical polarization extinction ratios are greater than 50 dB. Thus, one can avoid additional waveguide loss due to the different polarization modes, but it also means that one must properly align the input polarization to the waveguide for efficient coupling. Second, the APE method can produce a much large refractive index change, which has advantages for more compact waveguide designs.

Optical modulators can be divided into two main configurations, which are called a lumped circuit-element modulator and a traveling-wave modulator. In the lumped circuit-element modulator configuration, the RF input signal terminates at a pair of electrodes in parallel with a matched load. The problem with this modulator configuration is that its modulation bandwidth is primarily limited by the product of the load resistance and the electrode capacitance. However, reducing the electrodes' length or increasing their gap also increases the driving voltage, which is undesirable. For operation below 1-GHz, a possible

solution can be obtained by using a matching circuitry between the RF signal source and the electrodes. Another approach is to use the traveling-wave modulator, where the electrodes are constructed along the optical waveguide and acting as transmission lines whose impedance is designed to match the RF source. The electrodes are terminated with a matched load. The propagating RF input signal along the transmission line interacts with the co-propagating optical wave inside the waveguide, resulting in efficient phase modulation of the optical wave along the waveguide. Consequently, the modulation bandwidth is not limited by the RC time constant, but by the inherent velocity mismatch between the propagating optical wave and RF signal.

For transmission over cable TV networks, externally modulated laser transmitters use primarily optical amplitude or intensity modulators. The most widely used intensity modulators, which are illustrated in Figure 4.2, are Mach-Zehnder interferometer (MZI) modulators and balanced-bridge interferometer (BBI) modulators. In the MZI modulator, a single input waveguide is split into two waveguides by a 3 dB Y junction, and then recombined by

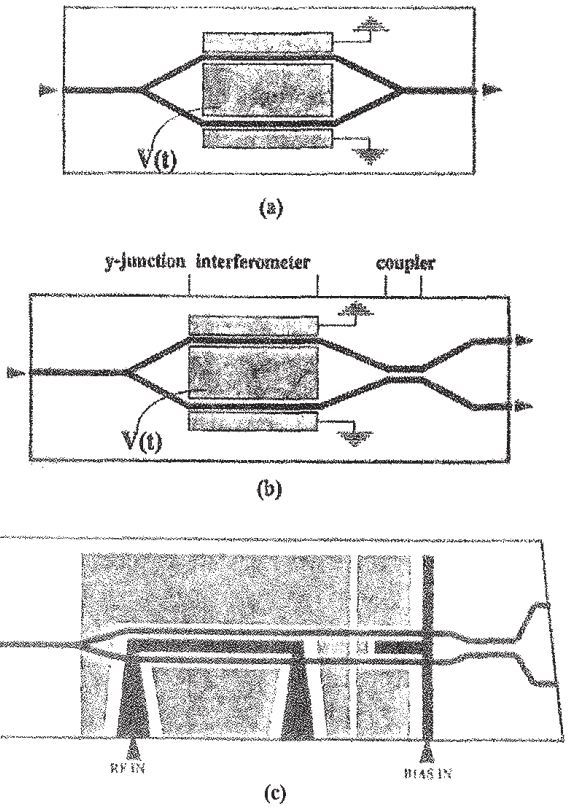



Figure 4.2 Schematic of different optical intensity modulators: (a) Mach-Zehnder interferometer, (b) balanced-bridge interferometer modulator, and (c) state-of-the-art balanced-bridge interferometer modulator, including traveling-wave electrodes, a separate DC bias electrode, and slanted facets. After Ref. [3] (© 1993 IEEE).

a second Y junction into a single output waveguide. In the traveling-wave MZI modulator, the RF modulation signal is traveling along the transmission lines causing a phase shift between the two arms. The phase-modulated optical wave at each arm of the waveguide can recombine either constructively or destructively, resulting in amplitude modulation of the output optical wave. The modulated optical power output from a MZI modulator can be written as:

$$P(V) = \langle P \rangle \left[ 1 + \sin\left( \frac{\pi V}{V_\pi} + \phi_b \right) \right] \tag{4.3}$$

where $\langle P \rangle = P_{in}/2L_m$ is the average optical output power, $P_{in}$ is the input CW optical power, $L_m$ is the excess loss of the modulator, $V_\pi$ is the half-wave voltage, and $\phi_b$ is the bias phase.

One of the main problems with external modulated transmitter was the relatively high insertion loss of $LiNbO_3$ optical modulators. The MZI modulator insertion loss factors include a coupling loss at the input facet of the waveguide (≈ 0.5-dB), an inherent 3 dB waveguide loss at the Y-junction combiner, and an optical propagation loss, which is typically negligible. There is an additional 3-dB loss since the modulator has to be biased at a half-power point (see explanation later), and an electrode loading loss due to impedance mismatch with the RF signal source. Thus, the typical insertion loss of the MZI modulator is about 7 dB.

To understand the inherent 3-dB loss at the Y-junction combiner, let us assume that the optical field distribution in the two arms is represented as a superposition of symmetric mode and antisymmetric mode. The symmetric mode is gradually coupled into the single output waveguide, while the antisymmetric mode, which cannot be supported by the output waveguide, radiates into the substrate. Thus, 50% (3 dB) of optical power is being wasted in the substrate. This limitation of the MZI modulator is overcome by incorporating two optical output ports by replacing the Y-junction combiner with a 2x2 directional coupler as shown in Figure 4.2(b). In this type of intensity modulator, which is called the balanced-bridge interferometer (BBI) modulator, the 2x2 directional coupler mixes the two optical waves and routes them to the two optical outputs. Consequently, both optical outputs can be utilized, and the 3-dB inherent loss of the MZI modulator is being eliminated. The modulated optical power of the BBI modulator at each output port ($P_\pm$) is given by

$$P_\pm = \langle P \rangle \left[ 1 \pm \sin(\theta + \phi_b) \right] \tag{4.4}$$

where $\theta = \pi V/V_\pi$ is the normalized modulating voltage. The two transfer functions of the BBI modulator are plotted in Figure 4.3. Notice that the sum of the two optical outputs from the MZI modulator is constant and equal to $2\langle P \rangle$. Thus, compared with the MZI modulator, the BBI modulator has twice the optical power utilization efficiency. Figure 4.2 (c) shows a schematic of state-of-the art BBI modulator, which includes a traveling-wave electrode design, a separate DC bias electrode to separate the RF and DC paths, and slanted input and

output facets to minimize the effect of interferometric noise. The advantage of the traveling-wave electrode design is a flatter frequency response over the operating bandwidth that is critical for the modulator linearization (see Section 4.3).

The ideal transfer characteristics of the BBI modulator may be modified due to wavelength and polarization effects. If the laser source is not a single-mode laser, the resulting transfer function will be a superposition of Equation (4.4) with a different $V_\pi$ and $\phi_b$ for each wavelength. This is undesirable for a cable TV transmission system. In addition, if there is a polarization mismatch between the launched optical beam and the BBI modulator, significant optical coupling losses may occur. Consequently, a polarization-maintaining fiber is typically used to couple the light into the single-mode optical modulator waveguide.

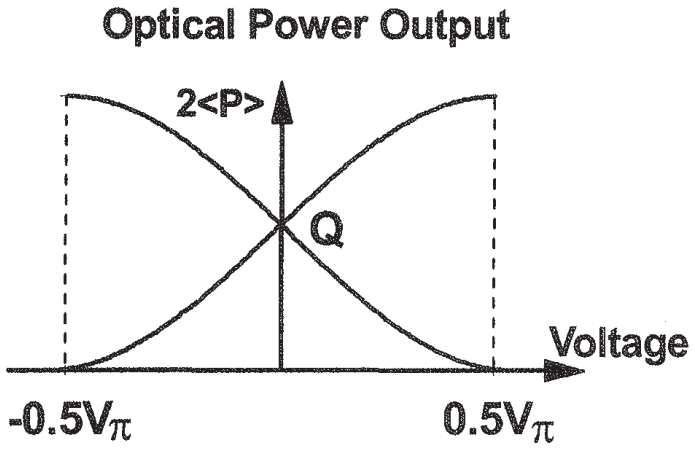



Figure 4.3 The optical power versus the modulating voltage for the two output ports of the balanced-bridge interferometer modulator. After Ref. [3] (© 1993 IEEE).

### 4.1.2 Distortion Characteristics of MZI and BBI Modulators

The linearity characteristics of MZI and BBI modulators are key for the operation of these modulators in the transmission of analog video signals. Following Reference [3], let us define the AC component of the modulator output as

$$p = \frac{P - \langle P \rangle}{\langle P \rangle} = \sin(\theta + \phi_b) \tag{4.5}$$

Let us assume a multichannel cable TV modulation signal consisting of $N$ unmodulated carriers plus a DC bias term $V_B$, which can be written as:

$$V(t)=\sum_{i=1}^{N} V_0 \cdot \cos\left[\omega_i t+\psi_i\right]+V_B \tag{4.6}$$

and it is applied to Equation (4.5). Expanding sin(θ) and cos(θ) in a power series and retaining terms up to third order, Equation (4.5) becomes

$$p=\cos(\phi_T)\left[\beta\sum_{i=1}^{N}\cos(\omega_i t+\psi_i)-\frac{\beta^3}{6}\left(\sum_{i=1}^{N}\cos(\omega_i t+\psi_i)\right)^3\right]+\sin(\phi_T)\left[1-\frac{\beta^2}{2}\left(\sum_{i=1}^{N}\cos(\omega_i t+\psi_i\right)^2\right] \tag{4.7}$$

where $\beta=\pi V/V_\pi$, and $\phi_T=\phi_b+\pi V_B/V_\pi$ is the total retardation bias, which is the sum of the intrinsic and applied DC phase biases. In Equation (4.7), the linear term in β is the desirable term, the term in $\beta^2$ is the CSO component, while the term in $\beta^3$ is the CTB component. Notice that if $\phi_T=0$, the CSO term is canceled, the CTB term is maximized, and the ratio of the linear term and the CTB term is independent on the bias point $\phi_T$. Consequently, the ideal bias point of the BBI modulator is the inflection point $\phi_T=0$, which is also known as the quadrature point (Q), as shown in Figure 4.3. Furthermore, when the BBI modulator is biased at the Q point, all the even orders of the intermodulation products are canceled. In practice, temperature variations and stress may cause the intrinsic bias to slowly drift with time constants of minutes and even hours [6]. This problem is typically solved using a tracking DC voltage $V_B$, which is applied to the bias electrode, whose magnitude is given by

$$V_B(t)=-\frac{\phi_b(t)\cdot V_\pi}{\pi} \tag{4.8}$$

This condition is satisfied using a parametric feedback control system for CSO suppression, and it is discussed in Section 4.4. If the BBI modulator is biased at the Q point, its normalized output from Equation (4.7) is reduced to:

$$p=p^{(1)}+p^{(3)}=\beta\sum_{i=1}^{N}\cos(\omega_i t+\psi_i)-\frac{\beta^3}{6}\left[\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\cos\left(\omega_{i_1}t+\psi_{i_1}\right)\cdot\cos\left(\omega_{i_2}t+\psi_{i_2}\right)\cdot\cos\left(\omega_{i_3}t+\psi_{i_3}\right)\right] \tag{4.9}$$

where the first term $p^{(1)}$ is the linear term, and the second term $p^{(3)}$ is the CTB contribution, which can be further expanded to yield

$$p^{(3)}=-\frac{\beta^3}{24}\sum_{\pm}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\cos\left[\left(\omega_{i_1}\pm\omega_{i_2}\pm\omega_{i_3}\right)t+\psi_{i_1}\pm\psi_{i_2}\pm\psi_{i_3}\right] \tag{4.10}$$

where the ± implies summation over all four possible pairs of signs. To evaluate the different combinations of summation in Equation (4.10), one has to remember that the cable TV

channels are spaced 6 MHz apart from channel 2, and many of the CTB products could potentially fall on the same beat frequency $\omega_b$ whenever the following condition is satisfied:

$$\omega_b = \omega_{i_1} + \omega_{i_2} + \omega_{i_3} \tag{4.11}$$

where the following convention is used for negative indices: $\omega_{-i} = -\omega_i$. In addition, the beat contribution depends on whether the various beats are added coherently (i.e., equal phases) or incoherently. For a given triple $\{i_1, i_2, i_3\}$ with different indices, the beats associated with all of the other five possible permutations of these three indices are mutually coherent. Thus, the total beat contribution associated with a given combination of three different indices with an arbitrary order is six times larger than any individual beat associated with any particular order. Similarly, for a given triple $\{i_1, i_1, i_2\}$ with only two distinct indices, the beats associated with the other two possible permutations of these indices are mutually coherent. Thus, the total beat contribution associated with a triplet combination with only two distinct indices is three times larger than any other individual beat with a particular index permutation. Using these results, the CTB component of the modulator output is given by

$$p^{(3)} = \sum_{\pm}\left[6\sum_{i_1=1}^{N-2}\sum_{i_2=i_1+1}^{N-1}\sum_{i_3=i_2+1}^{N} B(i_1,\pm i_2,\pm i_3) + 3\sum_{i_1=1}^{N-1}\sum_{i_2=i_1+1}^{N}\left[B(i_1,\pm i_2,\pm i_2) + B(i_2,\pm i_1,\pm i_1)\right] + \sum_{i=1}^{N} B(i,\pm i,\pm i)\right] \tag{4.12}$$

where the triple-beat mixing product B is defined as

$$B(i_1,i_2,i_3) = -\frac{\beta^3}{24}\cos\left[\left(\omega_{i_1} + \omega_{i_2} + \omega_{i_3}\right)t + \psi_{i_1} + \psi_{i_2} + \psi_{i_3}\right] \tag{4.13}$$

The CTB component of the normalized modulator output can easily be evaluated using computer simulations.

## 4.2 Linearization Methods of Optical Modulators

Many linearization methods of optical modulators have been proposed and implemented in externally modulated laser transmitters [7–11]. Generally speaking, the linearization methods can be divided into feedforward linearization and predistortion linearization categories. The operation principle of each of the linearization methods is explained in the following subsection with a practical implementation example. The intention here is not to cover all the possible implementations, but to explain the main advantages and disadvantages of the different methods. Optical linearization methods of external modulators, including optical

feedforward linearization, dual parallel linearization, and dual cascade linearization methods are discussed separately in Section 4.3.

### 4.2.1 Feedforward Linearization Method

Figure 4.4 shows a practical implementation block diagram of the feedforward linearization method for the BBI modulator [3]. The RF input signal is split before the modulator and it is delayed and amplified relative to the applied RF signal. A small fraction of the output power from each arm of the optical modulator is also split and detected using a simple photodetector. The RF error signal, which is generated by combining the delayed RF signal with the photodetector output signal using a broadband 180° RF hybrid comparator, is applied to an auxiliary DFB laser. The DFB laser must be tuned to a wavelength close to the source laser, but it is not allowed to have the same wavelength in order to prevent coherent interference. The optical output of each of the auxiliary DFB lasers is then combined with

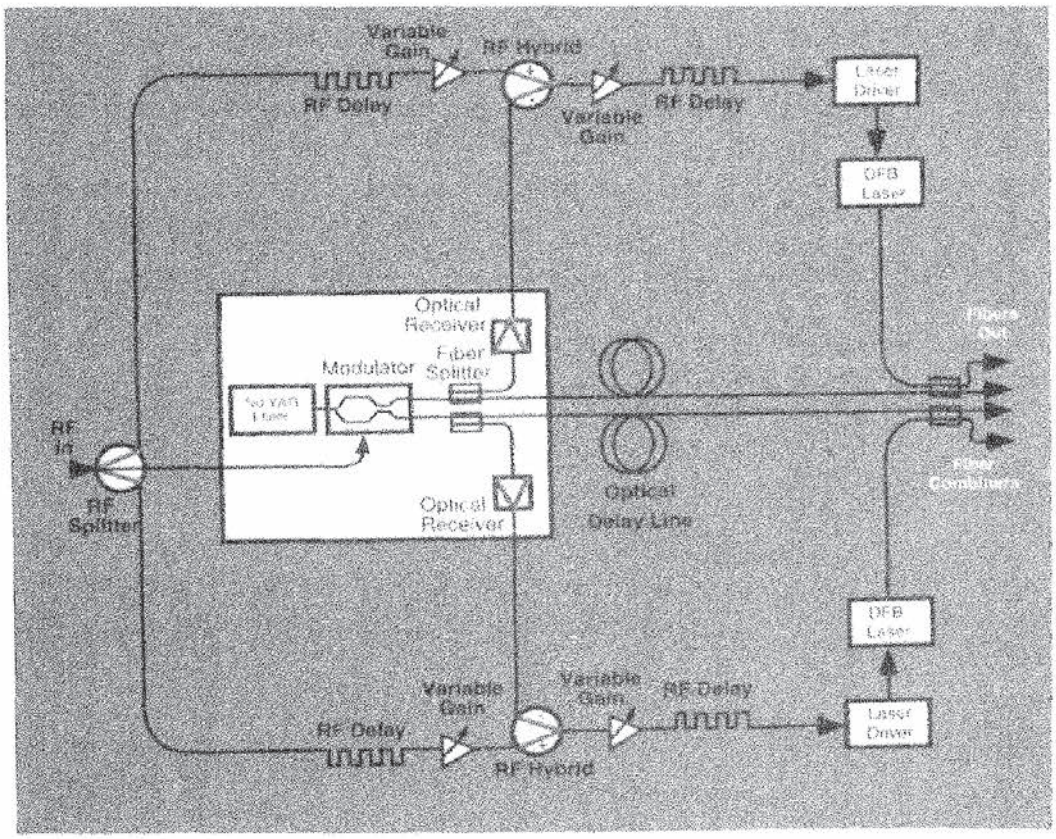



Figure 4.4 Block diagram of the feedforward linearization method for the balanced bridge interferometer modulator. After Ref. [3] (© 1993 IEEE).

the corresponding arm of the BBI modulator output. It should be pointed out that the auxiliary DFB lasers have to be sufficiently powerful to provide a signal swing four times larger than the correction signal swing in the output fiber, assuming a 20/80% optical coupler. For example, the required DFB laser peak power is of 2.72-mW, assuming 60 NTSC channels at an optical modulation index of 3.6% per channel.

To compensate for the modulator nonlinearity, the feedforward linearization method requires frequency flattening and gain matching at the auxiliary photodetector, DFB laser, and its drive circuitry. When combining the optical signals from two different lasers operating at two different wavelengths, a differential time delay is generated due to the fiber dispersion, which also has to be compensated. The differential delay is proportional to the fiber length.

### 4.2.2 Predistortion Linearization Method

The predistortion linearization method, which is simpler than the feedforward linearization, takes advantage of the stable and consistent transfer function characteristics of the MZI/BBI modulators. The basic concept behind the predistortion linearization method is illustrated in

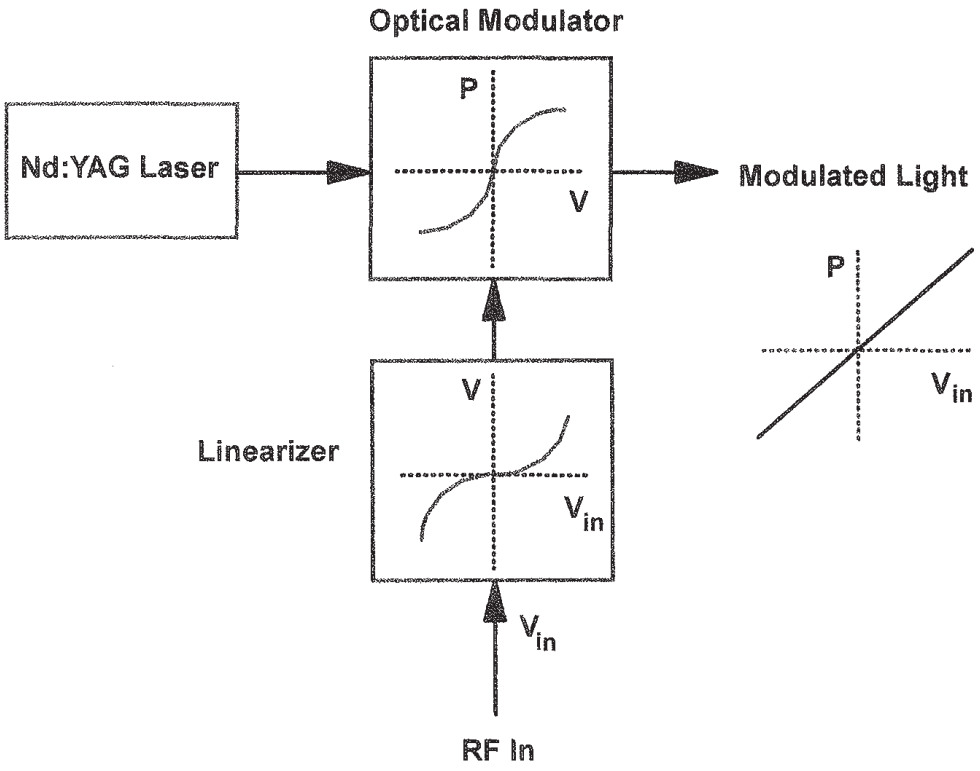



Figure 4.5 Principle of electronic predistortion linearization method. After Ref. [3] (© 1993 IEEE).

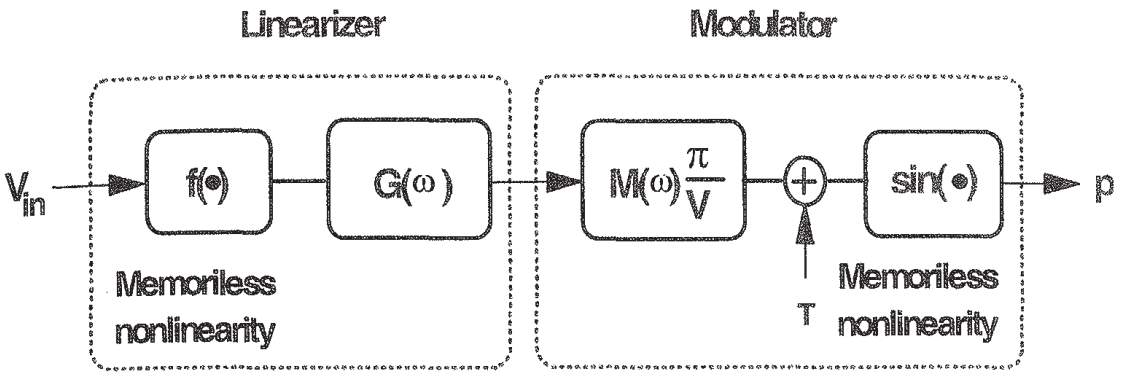



Figure 4.6 General model of the predistortion linearization method used in externally modulated laser transmitters. After Ref. [3] (© 1993 IEEE).

Figure 4.5. The linearized light-versus-voltage of the optical modulator is obtained by inserting a linearizer between the transmitter RF input and the modulator with an arcsine-shaped transfer characteristic. The general model of the predistortion linearization method employed in externally modulated laser transmitters in shown in Figure 4.6 [3]. Assuming that the modulator is biased at the Q point ($\phi_T = 0$), and both the modulator and linearizer can be described by memoriless nonlinearity, then the linearizer transfer characteristic can be written as a power series with odd terms:

$$V = f(V_{in}) = \alpha_1 V_{in} + \alpha_3 V_{in}^3 + \alpha_5 V_{in}^5 + \ldots \tag{4.14}$$

where $\alpha_i$ (i = 1,2,3,…) represents the small signal gain coefficients. Furthermore, assuming M(ω) = 1, the normalized optical power output is given by

$$p = \sin(K \cdot V) = \sin\left[K \cdot \alpha_1 \cdot V_{in} + K \cdot \alpha_3 \cdot V_{in}^3 + K \cdot \alpha_5 \cdot V_{in}^5 + \ldots\right] \tag{4.15}$$

where $K = \pi G/V\pi$ is the effective transfer gain of the linearizer, assuming $G(\omega)$ is a constant $G$. Using the power series expansion of sin(•), and neglecting fifth-order or higher terms for simplicity, we obtain

$$p = K \cdot \alpha_1 \cdot V_{in} + \left[K \cdot \alpha_3 - \frac{\alpha_1 \cdot K^3}{6}\right] V_{in}^3 + O\left(V_{in}^5\right) \tag{4.16}$$

To cancel the CTB contribution of the modulator, $G$ has to satisfy the following condition:

$$G = \frac{V_\pi}{\pi}\sqrt{\frac{6\alpha_3}{\alpha_1}} \tag{4.17}$$

Since fifth- and higher odd order distortions are still present, a practical linearizer design has to be further optimized to minimize their contribution at the fiber node receiver. In the general model, the effective transfer function $K(\omega)$ becomes

$$K(\omega) = G(\omega)\frac{\pi}{V_\pi}M(\omega) \tag{4.18}$$

If the power series expansion neglects fifth- or higher order terms, the linearization problem is equivalent to analyzing a system consisting of a linear filter $K(\omega)$ sandwiched between two memoriless nonlinearities. It follows from such an analysis that a flat frequency response of the modulator is required for good frequency linearization. The frequency response flatness can be affected by various factors such as parasitics of the linearizer circuits, RF amplifier ripples, electrode reactance and losses, and a large velocity mismatch between the RF and optical fields in the modulator waveguide.

### 4.2.3 Linearizer Circuits

So far, we have learned about the predistortion linearization theory and operation principles.

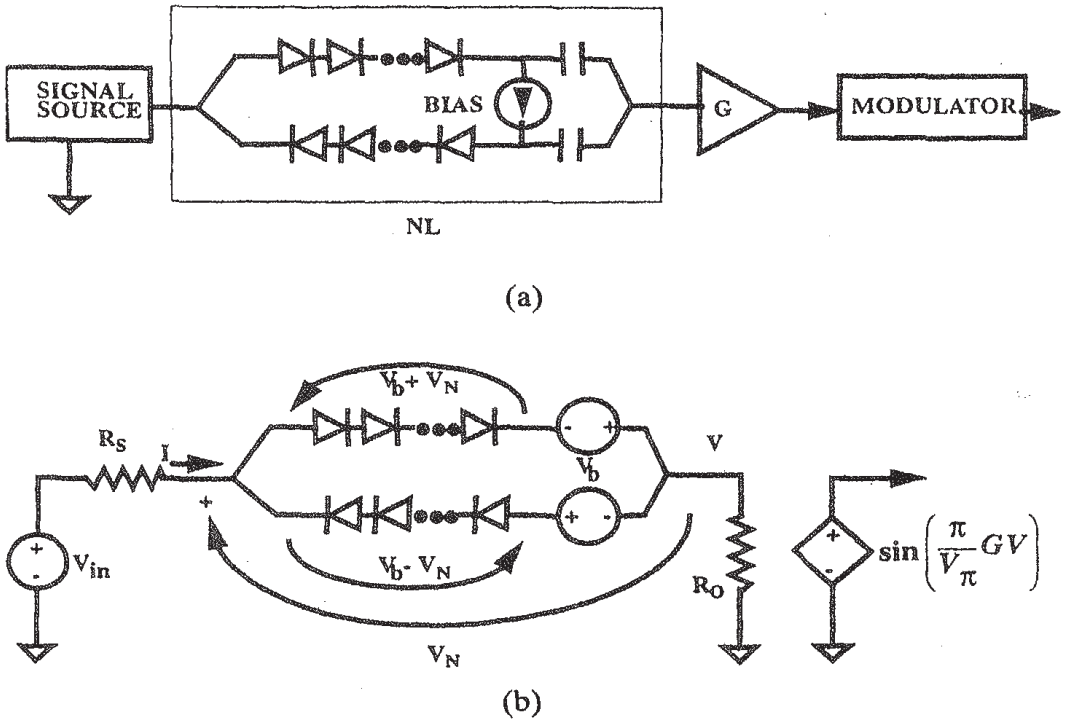



Figure 4.7 Predistortion circuit (a) and the equivalent circuit model (b). After Ref. [5] (© 1990 IEEE).

Let us examine several examples of linearizer circuits for external optical modulators. Figure 4.7 shows an example of a linearizer circuit and the equivalent circuit model [9]. The circuit takes advantage of a parallel combination but in opposite directions of two or more balanced diodes to suppress the magnitude of third-order distortion. The DC current source biases the diodes to a given bias, and the capacitors are charged to first approximation to a fixed large voltage and are modeled as a voltage source as shown in Figure 4.7(b). Using the well-known diode characteristic, $I = I_s\exp(V/V_T)$, where $I_s$ is the leakage current, $V_T = k_B T/e$, and the output current through the resistor $R_0$ is given by:

$$I = I_s\left[e^{(V_B+V_N)/NV_T} - e^{(V_B-V_N)/NV_T}\right] = 2I_s \cdot e^{V_B/NV_T}\sinh\left[\frac{V_N}{NV_T}\right] = 2I_B\sinh\left[\frac{V_N}{NV_T}\right] \tag{4.19}$$

where $N$ is the number of diodes in each branch, $V_B$ and $V_N$ are defined in Figure 4.7(b), and $I_B$ is the current flowing through the diode with no applied signal. The voltage-current characteristic can easily be obtained by inverting Equation (4.19):

$$V_N = NV_T \cdot a\sinh\left[\frac{I}{2I_B}\right] \tag{4.20}$$

Using Equation (4.20), we can expand $V_N$ in power series as

$$V_N = r_1 \cdot I + r_3 \cdot I^3 + \ldots = \left[\frac{NV_T}{2I_B}\right]I - \left[\frac{NV_T}{48I_B^3}\right]I^3 + \ldots \tag{4.21}$$

From the equivalent circuit model, the input voltage is related to the diode current by

$$V_{in} = R_s \cdot I + (r_1 \cdot I + r_3 \cdot I^3) + R_0 \cdot I = (R_0 + R_s + r_1)I + r_3 \cdot I^3 \tag{4.22}$$

Inverting Equation (4.22) yields the following power series

$$I = g_1 \cdot V_{in} + g_3 \cdot V_{in}^3 = \frac{V_{in}}{R_s + R_0 + r_1} - \left[\frac{r_3}{(R_s + R_0 + r_1)^4}\right]V_{in}^3 \tag{4.23}$$

The linearizer circuit output voltage is given by

$$V_{out} = R_0 \cdot I = \alpha_1 \cdot V_{in} + \alpha_3 \cdot V_{in}^3 = R_0 \cdot g_1 \cdot V_{in} + R_0 \cdot g_3 \cdot V_{in}^3 \tag{4.24}$$

where the coefficients $g_1$ and $g_3$ are given by Equation (4.23). Equation (4.24) provides the dependence of the output voltage on the bias current and the number of diodes, allowing one to complete the linearizer circuit design by applying Equation (4.17) to determine the linearization gain.

## 4.3 Optical Linearization Methods

Optical linearization can also be used instead of electronic linearization circuit. Optical linearization methods can be classified into three categories as follows: (A) an optical dual parallel linearization, (B) an optical dual cascade linearization, and (C) an optical feedforward linearization. The operation principles as well as the practicality of each method will be briefly reviewed in the following sections.

### 4.3.1 Optical Dual Parallel Linearization Method

The optical dual parallel linearization method has been proposed by Korotky et al. [11] and implemented by Brooks et al. [12]. A schematic configuration of an MZI modulator using the optical dual parallel linearization method is shown in Figure 4.8. In this configuration, two MZI modulators, which are called "primary" (top) and "secondary" (bottom), are biased at two 180° out-of-phase Q points such that their optical output power is given by Equation (4.4). In practice, the static phase of each MZI modulator may be different. Consequently, a

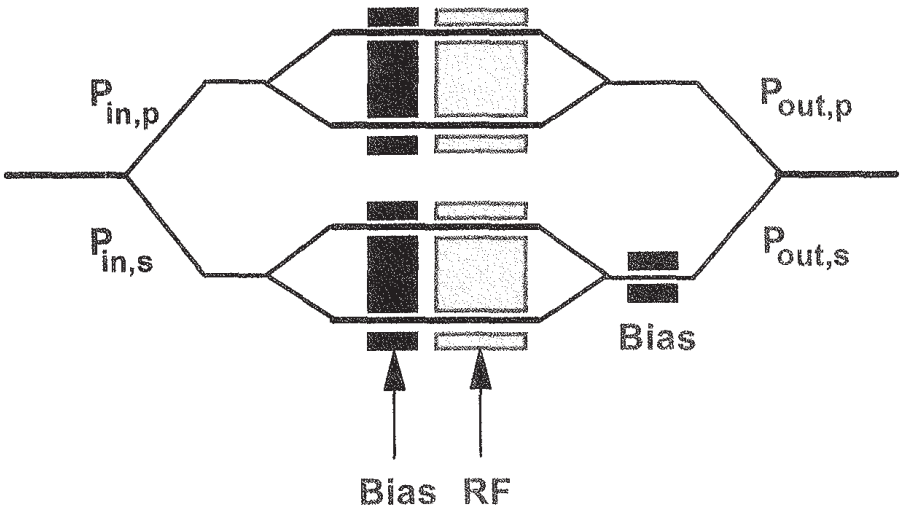



Figure 4.8 External modulator configuration employing the dual parallel linearization method. After Ref. [10] (© 1993 IEEE).

phase modulator at the output of the secondary MZI modulator is used to maintain the required $\pi$ phase shift between this modulator and the primary MZI modulator. In order to understand the operation principal of the dual parallel linearization method, let us assume that the RF drive signal is asymmetrically split between the two modulators with ratio $1:\alpha^2$ ($\alpha > 1$) such that the secondary MZI modulator operates at higher modulation depth and generates greater distortion than the primary MZI modulator. Assuming both modulators are biased at their Q point with $\pi$ relative phase shift, the output optical power of the primary and secondary MZI modulators is given by

$$P_{primary}(t) = \langle P_p \rangle \left[1 - \sin\left(\frac{\pi \cdot V(t)}{V_\pi}\right)\right] \approx \langle P_p \rangle \left[1 - \left(\frac{\pi \cdot V(t)}{V_\pi}\right) + \frac{1}{6}\left(\frac{\pi \cdot V(t)}{V_\pi}\right)^3\right] \tag{4.25}$$

$$P_{secondary}(t) = \langle P_s \rangle \left[1 + \sin\left(\frac{\alpha \cdot \pi \cdot V(t)}{V_\pi}\right)\right] \approx \langle P_s \rangle \left[1 + \left(\frac{\alpha \cdot \pi \cdot V(t)}{V_\pi}\right) - \frac{1}{6}\left(\frac{\alpha \cdot \pi \cdot V(t)}{V_\pi}\right)^3\right] \tag{4.26}$$

where $\langle P_p \rangle$ and $\langle P_s \rangle$ are the average optical output power of the primary and secondary MZI modulators. The combined output power from both modulators is given by

$$P_{total} = \langle P_p \rangle + \langle P_s \rangle - \left[\frac{\pi \cdot V(t)}{V_\pi}\right]\left(\langle P_p \rangle - \alpha \langle P_s \rangle\right) + \frac{1}{6}\left[\frac{\pi \cdot V(t)}{V_\pi}\right]^3 \left(\langle P_p \rangle - \alpha^3 \langle P_s \rangle\right) \tag{4.27}$$

The condition to cancel the third-order distortion of the MZI modulator is easily obtained from Equation (4.27) as

$$\langle P_p \rangle = \alpha^3 \langle P_s \rangle \tag{4.28}$$

Substituting Equation (4.28) into Equation (4.27), the magnitude of the fundamental component of the signal $P^{(1)}$ is given by

$$P^{(1)} = \frac{\pi |V(t)|}{V_\pi} \langle P_p \rangle \left[1 - \frac{1}{\alpha^2}\right] \tag{4.29}$$

and the magnitude of the DC component $P^{(0)}$ is given by

$$P^{(0)} = \langle P_p \rangle \left[1 + \frac{1}{\alpha^3}\right] \tag{4.30}$$

Using Equations (4.29) and (4.30), we find that the optical modulation index after the linearization is reduced by

$$\frac{1-\alpha^{-2}}{1+\alpha^{-3}} = \frac{\alpha(\alpha-1)}{1-\alpha+\alpha^2} \tag{4.31}$$

compared with the unlinearized MZI modulator. For example, for $\alpha = 3$, the optical modulation index is reduced to about 86% of that of a single MZI modulator.

This linearization method is not as practical as it may seems since the accuracy of the relative phase shift between the two MZI has to be controlled extremely tightly by the phase modulator. For example, to achieve a CSO distortion of –65-dBc with $\alpha = 3.3$, the relative phase shift must be $|\Delta\phi| < 0.06°$, and to achieve a –70 dBc CTB distortion, the relative phase shift must be $|\Delta\phi| < 0.54°$. Another possible solution is to replace the phase modulator with two independent optical sources that must be accurately maintained at a constant optical power ratio. However, the cost of two optical sources as well as the use of a secondary MZI modulator to linearize the primary MZI modulator makes this solution prohibitively expensive and unpractical.

### 4.3.2 Optical Dual Cascade Linearization Method

Another optically linearized modulator configuration is shown in Figure 4.9 [17–20]. The dual stage modulator consists of Y-junction splitter, two push-pull phase modulators with their separate RF electrodes and DC bias voltages, and two directional couplers, resulting in two output waveguide ports. For a 2x2 directional coupler modulator, the output optical power from the upper-waveguide port, for example, is modified from Equation (4.4) as:

$$P_1 = \langle P \rangle \left[1 - \sin(\theta + \phi_b)\sin(2\Gamma)\right] \tag{4.32}$$

where $\theta = \pi V/V_\pi$ is the normalized modulating voltage, $\Gamma = \kappa L$, $\kappa$ is the coupling coefficient between the two waveguides, and $L$ is the interaction length. In the dual cascade optically linearized modulator, the bias voltages are required to maintain the modulator operation at the Q point. To analyze the performance of this optical linearization method, let us assume that the RF input signal is split into a normalized voltage signal $\theta$ for $RF_1$ and $\rho\theta$ for $RF_2$. In addition, let us assume that the couplings between the two branches of the first and second directional couplers are $\Gamma_1$ and $\Gamma_2$, respectively. It can be shown that by multiplying the transfer matrix for each section of this modulator its output power transfer function is given by

$$P(V) = \langle P \rangle \begin{Bmatrix} 1 - \sin(2\Gamma_2)\cos(\theta + \phi_1)\sin(\rho\theta + \phi_2) \\ -\left[\sin(2\Gamma_1)\cos(2\Gamma_2) + 2\Gamma_1 \cdot \sin(2\Gamma_2)\cos(\rho\theta + \phi_2)\right]\sin(\theta + \phi_1) \end{Bmatrix} \tag{4.33}$$

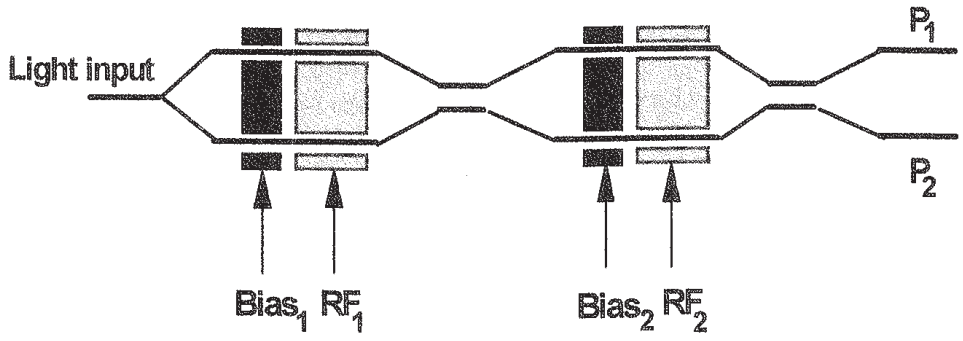



Figure 4.9 Basic configuration of a dual cascade optically linearized modulator. After Ref. [19] (© 1995 IEEE).

where $\phi_1$ and $\phi_2$ are the bias offset angles. The transfer function $P(V)$ can be linearized by first biasing the modulator at the Q point, namely, setting $\phi_1 = \phi_2 = 0$, to ensure that even higher order distortion contributions are eliminated. The third-order distortions can be canceled by requiring that $d^3P/dV^3 = 0$. This can be achieved by either fixing both $\Gamma_1$ and $\Gamma_2$ with an adjustable $\rho$ or vice versa. From a practical perspective, it is easier to fix $\Gamma_1 = \Gamma_2 = \Gamma$ when the waveguide is fabricated, allowing $\rho$ to vary. Consequently, the modulator transfer function is reduced to

$$P(V) = \langle P \rangle \left\{ 1 - \frac{1}{2}\sin(4\Gamma)\sin(\theta)\left[1 + \cos(\rho\theta)\right] - \sin(2\Gamma)\cos(\theta)\sin(\rho\theta) \right\} \tag{4.34}$$

The optimum coupling values for this modulator are in the range of 58°–65°. For $\rho = 0$, there is no linearization since the transfer function is pure sinusoidal. The optimum linearization of Equation (4.33) is obtained for $\rho = 0.5$. However, in practice, a slightly larger value of $\rho$ is optimal due to the contribution of the odd higher order distortions. The optimal value of $\rho$ is determined by the number of cable TV channels to be modulated and optical modulation index per channel.

This optical linearization method has two main disadvantages. First, due to asymmetric losses between the upper and lower branches of each of the directional couplers, it is difficult to obtain the same linearization for both optical outputs simultaneously compared with a single optimized output. Consequently, only one linearized optical output is typically being implemented. Second, this method has double the insertion losses compared with a single linearized stage modulator, which may limit the maximum optical output power.

## 4.4 Externally Modulated Laser Transmitter Design

There are two primary types of externally modulated laser transmitters that are used in the cable TV industry and will be reviewed here. The first type, which became commercially available in the early 1990s, is an externally modulated (EM) *yttrium aluminum garnet* (YAG) laser transmitter operating at 1319-nm. The second type, which became commercially available later, is an EM-DFB laser transmitter operating at 1550-nm.

### 4.4.1 Externally Modulated YAG Laser Transmitter

The first type of an EM laser transmitter, which was developed in the early 1990s, is the high-power Nd:YAG laser transmitter, which operates at 1319-nm [3]. Figure 4.10 shows a simplified energy level diagram of a YAG crystal doped with Nd atoms. When pumped with 808-nm light, the Nd atoms are excited to pump bands, which decay nonradiatively to a metastable state. The stimulated emission at 1319-nm is emitted from the metastable state as shown in Figure 4.10. The Nd:YAG laser cavity, which is shown in Figure 4.11, consists of a high-power (0.5–1 W) AlGaAs diode array with a broad active area mounted on a TEC, a Nd:YAG crystal, and micro-optic collimating and coupling lenses to shape and focus the 1319-nm light. Using selective coatings on the Nd:YAG crystal and output mirror, the 1319-nm light is confined to a resonant cavity overlapping the pump volume. Only a small fraction (1–2%) of the intracavity light is coupled through the output mirror. The output po-

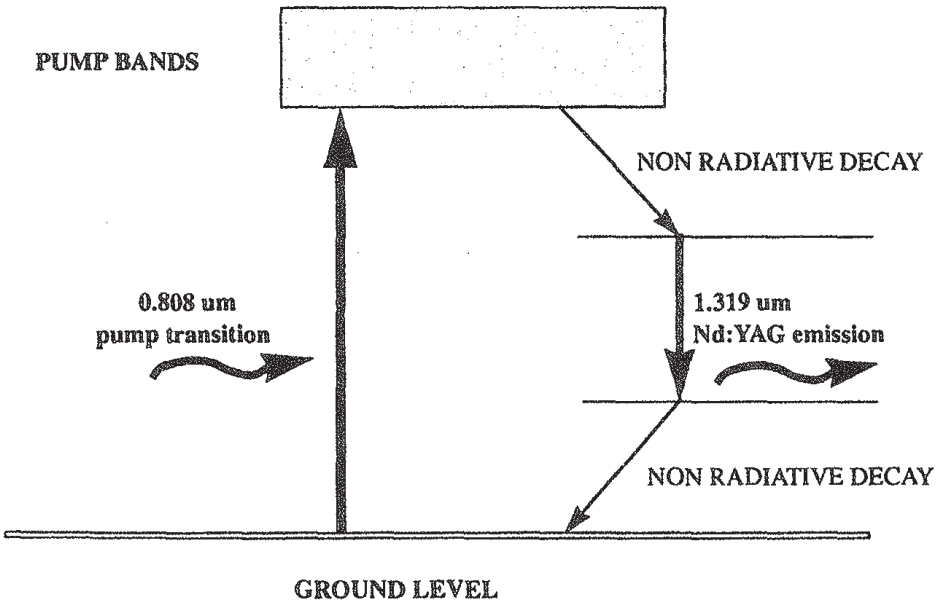



Figure 4.10 Energy-level diagram of Nd:YAG. After Ref. [3] (© 1993 IEEE).

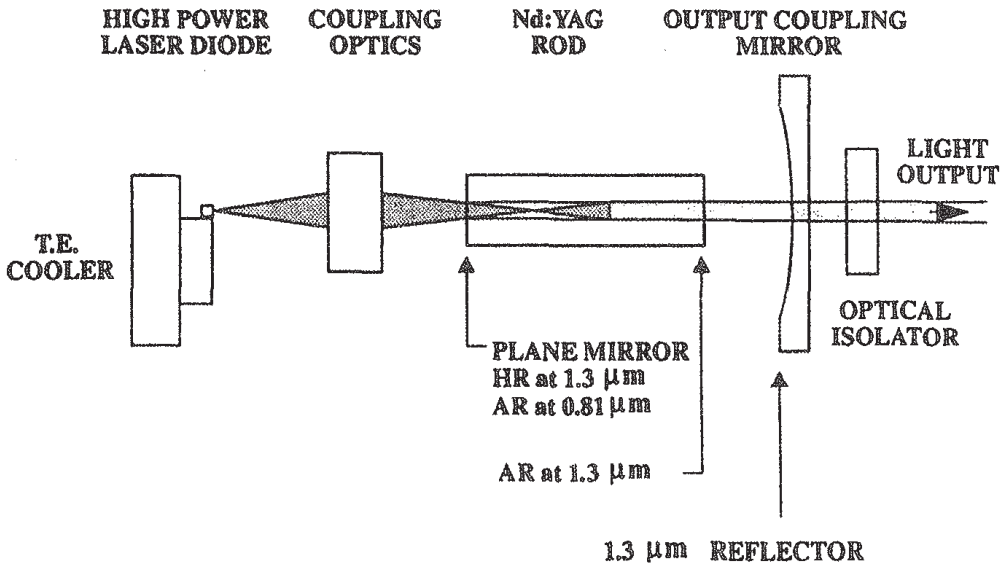



Figure 4.11 Schematic of laser diode array-pumped Nd:YAG laser cavity. After Ref. [3] (© 1993 IEEE).

wer from Nd:YAG laser typically ranges from 50 to more than 200 mW. The laser output beam is a almost perfectly shaped Gaussian beam, which allows efficient coupling to single-mode waveguides. In addition, the output beam is linearly polarized with an extinction ratio typically better than 23 dB. Due to the very low cutoff frequency (≈ 200-kHz) of the Nd:YAG crystal, the high-frequency noise is filtered, resulting in a very low RIN (< –165 dBc/Hz). The external optical modulator typically has a total loss of about 8 dB, including a 3-dB loss due to the DC bias at the quadrature point, a 3-dB loss due to the directional coupler, waveguide-fiber coupling and waveguide propagation losses. A block diagram of the EM Nd:YAG laser transmitter with predistortion linearization and parametric CSO and CTB control is illustrated in Figure 4.12. To achieve the long-term CSO and CTB performance requirements of the EM linearized laser transmitter, parametric feedback loops are implemented. In order to maintain the modulator bias at the Q point ($\phi_T = 0$), an error signal, which is proportional to $\phi_T$, is detected and processed to generate a corrected bias voltage to the modulator. A separate closed loop is used to control the CTB performance by control ling the relationship between the gain and the nonlinearity of the linearizer. For an optical budget link of 13 dB, assuming 0 dBm at the optical receiver with 77 AM-VSB video channels, the CSO and CTB distortions are typically less than –70-dBc and –65-dBc using the EM Nd:YAG laser transmitter.

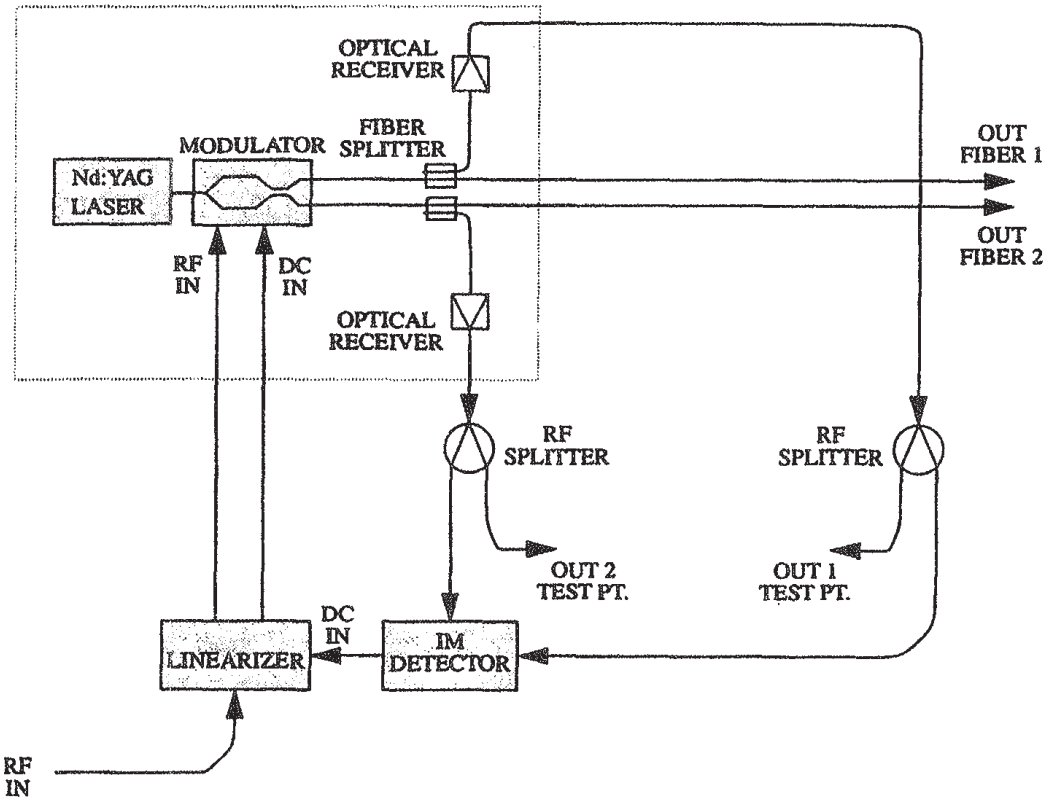



Figure 4.12 Block diagram of externally modulated Nd:YAG laser transmitter with predistortion linearization and parametric CSO and CTB control. After Ref. [3] (© 1993 IEEE).

Cascading predistorted externally modulated laser transmitters has an additional advantage in terms of CTB performance compared with directly modulated DFB laser transmitters [21]. When cascading two directly modulated DFB laser transmitter based links, the CTB distortions behave according to Equation (2.30), which is also called the "20 log" rule. This rule means that if the CTB distortion of each link is –65 dBc, the total CTB distortion of the cascaded link is degraded by 6-dB to –59-dBc. It turns out that the CTB distortions of cascaded predistorted EM YAG laser transmitters follow instead the "10 log" rule. This is because the distortion beats are randomized and uncorrelated with one another between the two externally modulated links. Therefore, the resultant CTB distortion of the cascaded link is degraded by only 3-dB to –62-dBc. In general, the CTB distortions are added according to the "X log" rule, where X can vary from 0 to 20, depending on the type of the type of linearization scheme and the cable TV channel frequency. The CTB performance advantage of the EM YAG laser transmitters was a very popular solution for cable TV supertrunk applications with links exceeding 30-km before low-cost high-power EDFAs became readily available.

### 4.4.2 Externally Modulated DFB Laser Transmitter

An externally modulated DFB laser transmitter operating at 1550-nm is the most widely used transmitter for broadcasting and narrowcasting applications over cable TV networks. Figure 4.13 shows a simplified block diagram of an externally modulated DFB laser transmitter operating at 1550-nm with a built-in boost optical amplifier. The optimal RF input signal range to the external modulator is controlled by the AGC circuitry, while the modulator bias point is controlled by the microprocessor. As was discussed previously in Chapter 1, to meet the CNR performance requirements for many cable TV network configurations, it is necessary to significantly increase the output optical power of these transmitters by incorporating a booster fiber amplifier. However, the maximum optical output power of these laser transmitters that can be launched into the fiber may be limited by a stimulated Brillouin scattering (SBS) effect [22]. The SBS optical power limit, which is discussed in detail in Chapter 9 can be increased by applying a single frequency tone whose frequency is more than twice the maximum transmitted signal frequency (≈ 2-GHz), or by dithering the optical frequency of the DFB laser transmitter [23].

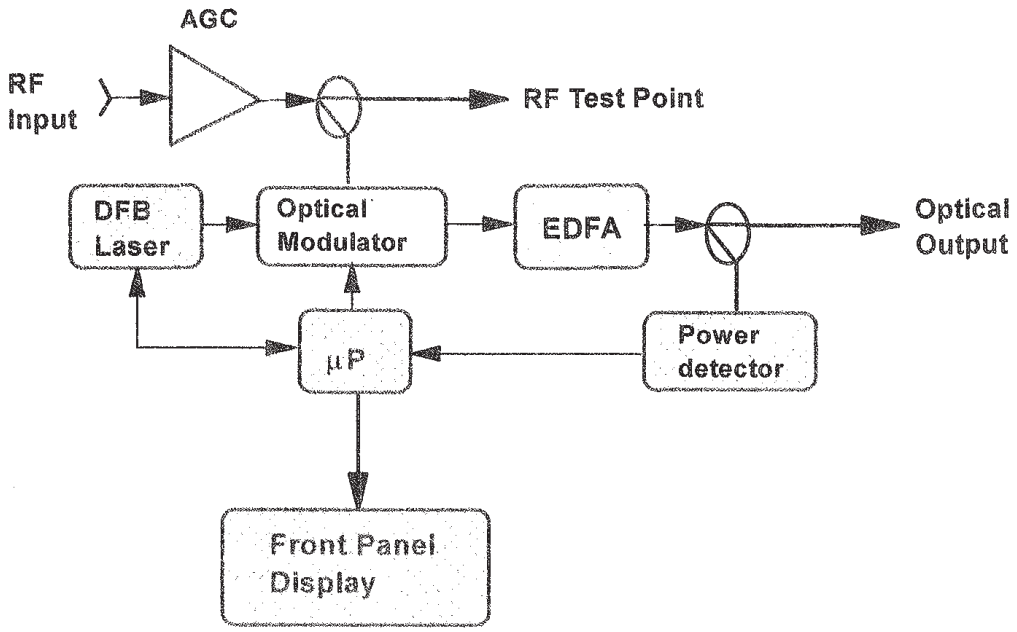



Figure 4.13 A simplified block diagram of a 1550-nm externally modulated DFB laser transmitter with a built-in EDFA.

Figure 4.14 shows, for example, the measured worst-case CSO and CTB distortions as a function of the optical modulation index per channel for externally modulated optically linearized 1550-nm DFB laser transmitter [24]. Seventy-nine simulated AM video channels

were transmitted 50-km over a standard SMF. The optimal modulation index per channel of this transmitter is 3.1% to meet the CTB distortion requirement of –65 dBc. Notice that the CTB distortion rapidly degrades as the optical modulation index increases beyond its optimal value. Consequently, the maximum possible CNR becomes limited at the fiber node receiver.

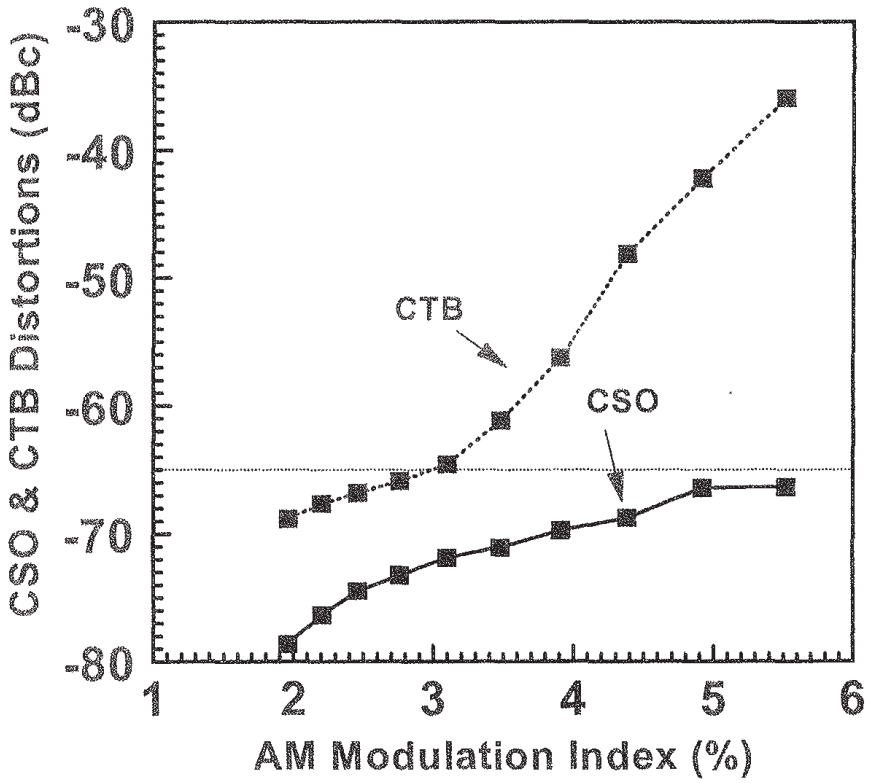



Figure 4.14 Measured worst-case CSO and CTB distortions versus the optical modulation index per channel using an 1550-nm externally modulated optically linearized DFB laser transmitter. After Ref. [24] (© 1998 IEEE).

## References

1. R. S. Weis and T. K. Gaylord, "Lithium Niobate: Summary of Physical Properties and Crystal Structure," *Applied Physics A* **37**, 191–203 (1985).

2. M. Nazarathy, J. Berger, A. J. Ley, I. M. Levi, and Y. Kagan, "Progress in Externally Modulated AM CATV Transmission Systems," *IEEE Journal of Lightwave Technology* **11**, 82–105 (1983).
3. F. W. Willems, W. Muys, and J. S. Leong, "Simultaneous Suppression of Stimulated Brillouin and Interferometric Noise in Externally Modulated Lightwave AM-SCM Systems," *IEEE Photonics Technology Letters* **12**, 1476–1478 (1994).
4. H. Skeie, "An Optically Linearized Modulator for CATV Applications," *Proceedings of SPIE*, **2291**, 227–238 (1994).
5. R. B. Childs and V. A. O'Byrne, "Multichannel AM Video Transmission Using a High-Power Nd:YAG Laser and Linearized External Modulator," *IEEE Journal on Selected Areas in Communications* **8**, 1369–1376 (1990).
6. R. S. Cheng, W. L. Chen, and W. S. Wang, "Mach-Zehnder Modulators with Lithium Niobate Ridge Waveguides Fabricated by Proton-Exchange Wet-Etch and Nickel Indiffusion," *IEEE Photonics Technology Letters* **7**, 1282–1284 (1995).
7. H. Nagata, K. Kiuchi, and T. Saito, "Studies of Thermal Drift as a Source of Output Instabilities in Ti:$LiNbO_3$ Optical Modulators," *Journal of Applied Physics* **75**, 4762–4764 (1994).
8. P. G. Suchoski Jr. and G. R. Boivin, "Reliability and Accelerated Aging of LiNbO3 Integrated Optic Fiber Gyro Circuits," OE Fibers, Components, Networks, and Sensors conference, Boston, September 8-11 (1992).
9. H. Skeie, An Optically Linearized Modulator for CATV Applications," *Proceedings of SPI,* **2291**, 227–238 (1994).
10. K. D. LaViolette, "CTB Performance of Cascaded Externally Modulated and Directly Modulated CATV Transmitters," *IEEE Photonics Technology Letters* **8**, 281–283 (1996).
11. W. Wang, R. Tavlykaev, and R. V. Ramaswamy, "Bandpass Traveling-Wave Mach-Zehnder Modulator in $LiNbO_3$ with Domain Reversal," *IEEE Photonics Technology Letters* **9**, 610–612 (1997).
12. M. Seino, T. Nakazawa, Y. Kubota, M. Doi, T. Yamane, and H. Hakogi, "A low DC-drift Ti:$LiNbO_3$ Modulator Assured over 15 Years," *Technical Digest of Optical Fiber Communications Conference*, Post-deadline paper PD-3 (1992).
13. D. A. Atlas, "On the Overmodulation Limit in Externally Modulated Lightwave AM-VSB CATV Systems," *IEEE Photonics Technology Letters* **8**, 697–699 (1996).
14. G. C. Wilson, "Optimized Predistortion of Overmodulated Mach-Zehnder Modulators with Multicarrier Input," *IEEE Photonics Technology Letters* **9**, 1535–1537 (1997).
15. J. L. Brooks, G. S. Maurer, and R. A. Becker, "Implementation and Evaluation of a Dual Parallel Linearization System for AM-SCM Video Transmission," *IEEE Journal of Lightwave Technology* **11**, 34–41 (1993).

16. S. K. Korotky and R. M. deRidder, "Dual Parallel Modulation Schemes for Low-Distortion Analog Optical Transmission," *IEEE Journal on Selected Areas in Communications* **8**, 1377–1381 (1990).
17. H. Skeie and R. V. Johnson, "Linearization of Electro-Optic Modulators by a Cascade Coupling of Phase Modulating Electrodes," Integrated Optical Circuits, *SPIE proceedings*, **1583**, 153–164 (1991).
18. W. K. Burns, "Linearized Optical Modulator with Fifth Order Correction," *IEEE Journal of Lightwave Technology* **13**, 1724–1727 (1995).
19. D. J. M. Sabido IX, M. Tabara, T. K. Fong, C.-Li Lu, and L. Kazovsky, "Improving the Dynamic Range of a Coherent AM Analog Optical Link Using a Cascaded Linearized Modulator," *IEEE Photonics Technology Letters* **7**, 813–815 (1995).
20. J. D. Farina, B. R. Higgins, and J. P. Farina, "New Linearization Technique for Analog Fiber-Optic Links," Technical Digest, Optical Fiber Communications conference, paper ThR6, San Jose, CA (1996).
21. M. Nazarathy and Y. Simler, "Integrated Networks for CATV Transmission and Distribution," *Cable Telecommunication Engineering* **16**, Numbers 3 and 4 (1994).
22. X. P. Mao, G. E. Bodeep, R. W. Tkach, A. R. Chraplyvy, T. E. Darcie, and R. M. Derosier, "Brillouin Scattering in Externally Modulated Lightwave AM-VSB CATV Transmission Systems," *IEEE Photonics Technology Letters* **4**, 287–289 (1992).
23. F. W. Willems, W. Muys, and J. S. Leong, "Simultaneous Suppression of Stimulated Brillouin Scattering and Interferometric Noise in Externally Modulated Lightwave AM-SCM Systems," *IEEE Photonics Technology Letters* **12**, 1476–1478 (1994).
24. S. Ovadia and C. Lin, "Performance Characteristics and Applications of Hybrid Multichannel AM/M-QAM Video Lightwave Transmission Systems," *IEEE Journal of Lightwave Technology* **16**, 1187–1207 (1998).

# CHAPTER 5

# LIGHTWAVE RECEIVERS FOR CABLE TV NETWORKS

The primary role of optical receivers in cable TV networks, which are typically located at both the fiber nodes and at the cable TV headends, is to convert the transmitted optical signals to RF signals. To understand the design and performance requirements of lightwave receivers, let us first discuss the physics of a p-i-n photodetector and define its basic parameters. In Section 5.2, we will review the three primary noise sources in optical receivers, namely, shot noise, thermal noise, and laser relative-intensity noise (RIN). Using this knowledge, the calculation of the SNR or CNR at the optical receiver for an SCM fiber-optics system is explained in Section 5.3. Section 5.4 will explain the nonlinear behavior of the p-i-n photodetector in order to be effectively used in cable TV networks. Finally, the basic cable TV receiver designs, including low- and high-impedance front-end designs, transimpedance front-end design, and various circuit configurations for high-performance operation are discussed in Section 5.5.

## 5.1 p-i-n Photodiode

Photodetectors are semiconductor devices that convert optical signals to electrical signals. The p-i-n photodiode, which is the most commonly used photodetector, consists of p-type and n-type semiconductors separated by lightly doped intrinsic semiconductor material. In n-type semiconductors, the majority carriers are electrons, while in p-type semiconductors, the majority carriers are holes. In an intrinsic semiconductor, which contains only small amounts of impurities compared with the thermally generated electron and holes, the density of electrons (per unit volume) in the conduction band is equal to the density of the holes in the valence band ($n = p = n_i$).

Before we examine the characteristics of p-i-n photodiodes, let us first explain the basic concepts of p-n photodiodes. A p-n junction is formed by physically joining uniformly doped p-type and n-type semiconductors. The importance of the p-n junction will be clear later. When an electric field is applied to an n-type semiconductor in thermal equilibrium (assuming uniform doping), the electrons will be accelerated between collisions in the opposite direction of the field. The composite displacement of an electron is due to its random thermal motion and the drift velocity is due to the electric field. The drift velocity can be

written as $-\mu_n E$, where the factor $\mu_n$ is called the electron mobility. A similar expression can be written for the holes in the valence band. In general, the equations for the drift current densities of the holes and electrons can be written as [1]

$$J_n^{drift} = -q \cdot n(x,t) \cdot v_n(E)$$
$$J_p^{drift} = q \cdot p(x,t) \cdot v_p(E) \tag{5.1}$$

where $q$ is the electronic charge, $n(x,t)$ and $p(x,t)$ are the densities of the electrons and holes, respectively, and $v_n(E)$ and $v_p(E)$ are the electric-field dependent hole and electron drift velocities, respectively, which will be discussed in Section 5.4.

Another important process in semiconductors is called the diffusion process, which describes the spatial variation of the carrier concentration in the semiconductor material. When an electric field is applied to a semiconductor at uniform temperature, assuming only that the electron and hole densities vary along the x-axis, then the total current density of the electrons and holes is simply the sum of the drift and diffusion components, and can be written as [2]:

$$J_n = J_n^{drift} + J_n^{diff} = qn \cdot \mu_n \cdot E + qD_n \cdot \frac{dn}{dx}$$
$$J_p = J_p^{drift} + J_p^{diff} = q\mu_p \cdot pE - qD_p \cdot \frac{dp}{dx} \tag{5.2}$$

where, $D_n$ and $D_p$ are the electron and hole diffusion constants, respectively, which are related to the electron and hole mobilities through the *Einstein relation* as follows [2]:

$$D_i = \left[\frac{kT}{q}\right]\mu_i \tag{5.3}$$

where i = $n$ or $p$. Notice that the diffusion current is proportional to the spatial derivative of the electron density. Now, consider a direct band-gap semiconductor in thermal equilibrium, where the relation $np = n_i^2$ is valid. When the p-n junction is forward-biased and/or a light is incident on the p-type semiconductor, the thermal equilibrium is disturbed (i.e., $np > n_i^2$) by the injection of additional carriers, which are called excess carriers. The process of recombination of the injected minority carriers with the majority carriers restores the p-n junction to thermal equilibrium. The recombination process can be radiative with the emission of photons or nonradiative. The rate of direct recombination is proportional to the number of available electrons in the conduction band and holes in the valence band. To maintain thermal equilibrium, the carrier generation rate $G(x,t)$ must be equal to the carrier recombination rate $R(x,t)$. Combining the effects of carrier drift, diffusion, recombination,

and generation, the operation of the p-n junction in one dimension is governed by the continuity equations as follows [1]:

$$\frac{\partial n(x,t)}{\partial t} = G(x,t) - R(x,t) + v_n \frac{\partial n}{\partial x} + n\frac{\partial v_n}{\partial x} + \frac{1}{q}\frac{\partial J_n^{diff}}{\partial x}$$
$$\frac{\partial p(x,t)}{\partial t} = G(x,t) - R(x,t) - v_p \frac{\partial p}{\partial x} - p\frac{\partial v_p}{\partial x} - \frac{1}{q}\frac{\partial J_p^{diff}}{\partial x} \tag{5.4}$$

where $v_n$ and $v_p$ are the electron and hole velocities, respectively. The continuity equations are linear only if the carrier velocities (i.e., $v_n$ and $v_p$) are independent of the carrier densities.

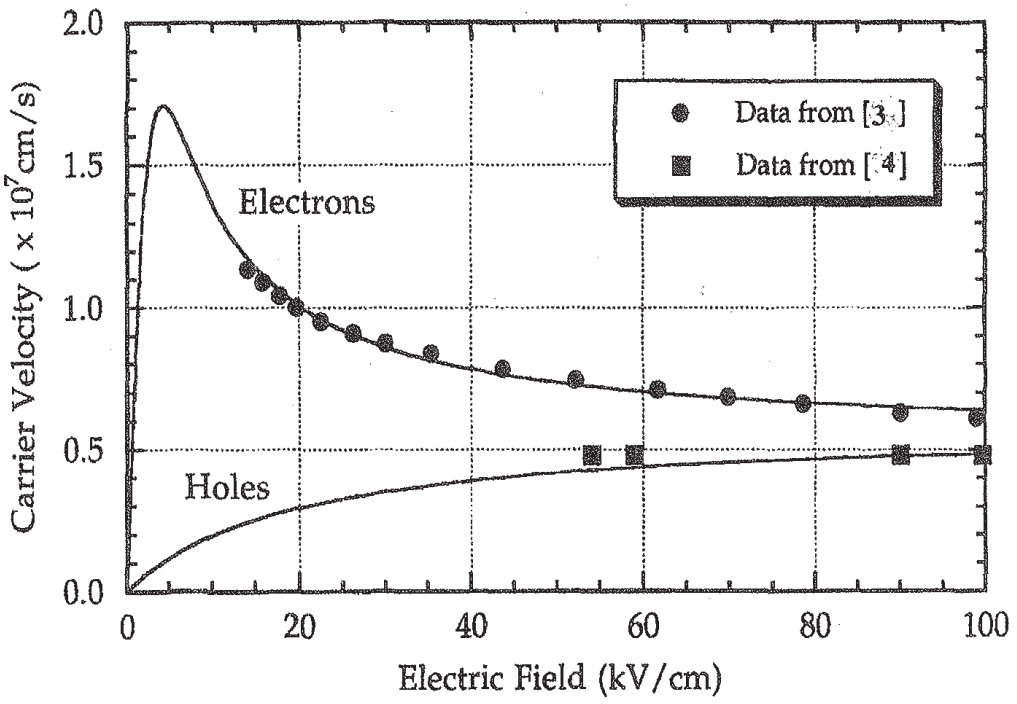



Figure 5.1 Measured and fitted electron and hole velocities versus the electric field in InGaAs. After Ref. [1] (© IEEE 1996).

Figure 5.1 shows the measured and fitted electron and hole velocities versus the electric fields in n-type InGaAs at room temperature. The electron velocity measurements have been fitted using the following empirical expression [3]:

$$v_n(E) = \frac{E\left[\mu_n + v_n^{hf} \cdot \beta |E|\right]}{1 + \beta \cdot E^2} \tag{5.5}$$

where $v_n^{hf}$ is the high-field electron velocity, and $\beta$ is a fitting parameter. Similarly, hole velocity measurements have also been fitted using the following expression [4]:

$$v_p(E) = \frac{\mu_p \cdot v_p^{hf} \cdot E}{\left[(v_p^{hf})^\gamma + (\mu_p \cdot E)^\gamma\right]^{1/\gamma}} \tag{5.6}$$

where $v_p^{hf}$ is the high-field hole velocity, and $\gamma$ is a temperature-dependent factor. The measured low-field electron and hole mobilities are 8000 $cm^2/Vs$ and 300 $cm^2/Vs$, respectively [3, 4]. Notice that for $\beta = 0$, the carrier drift velocities become the low-field carrier mobility times the applied electric field.

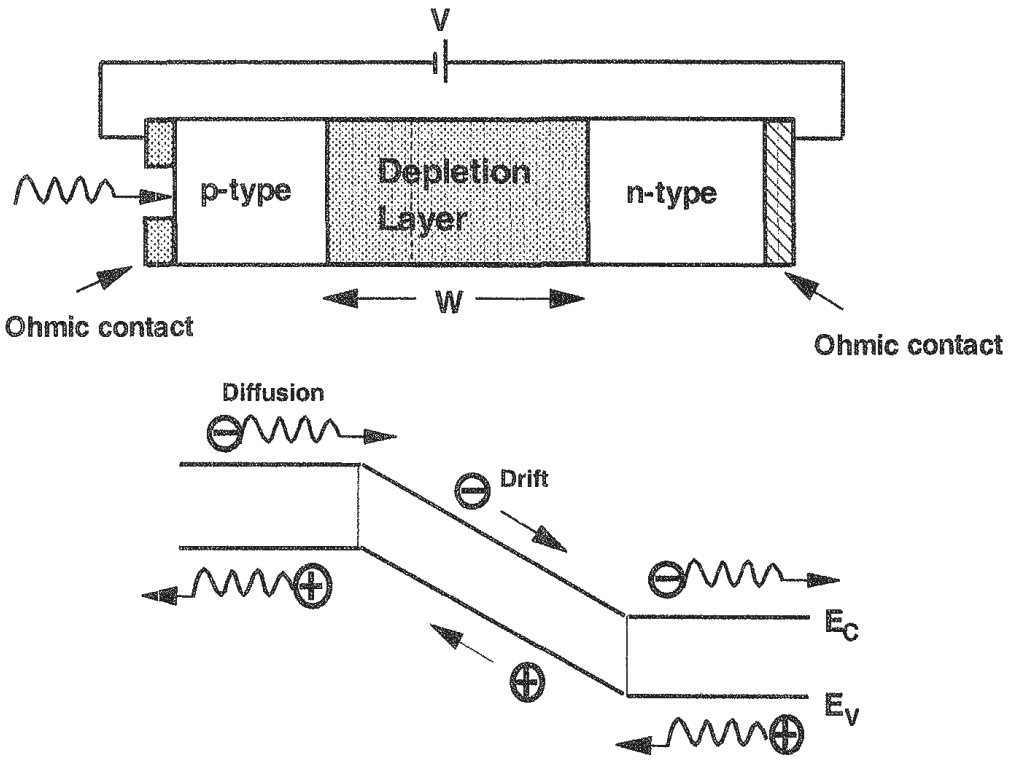



Figure 5.2 (a) Operation of p-i-n photodiode under reverse bias, and (b) energy-band diagram under reverse bias.

The large carrier concentration gradient across the p-n junction, holes from the p-side diffuse into the n-side, and electrons from the n-side diffuse into the p-side. Consequently, a negative space charge is formed on the p-side and a positive space charge is formed at the n-side, creating an electric field pointing from the n-side to the p-side. The electrostatic potential $\psi$ distribution across the p-n junction is given by the Poisson's equation [1, 2]

$$\frac{d^2\psi}{dx^2} = -\frac{dE}{dx} = -\frac{q}{\varepsilon_s}\left[N_D - N_A + p - n\right] \tag{5.7}$$

where $N_D$ and $N_A$ are the densities of ionized donor and acceptor dopants present in the crystal, respectively, and $\varepsilon_s$ is the dielectric constant. The total electrostatic potential difference between the p-side and n-type at thermal equilibrium is given by

$$V_{bi} = \psi_n - \psi_p = \frac{k_B T}{q} \ln\left[\frac{N_A N_D}{n_i^2}\right] \tag{5.8}$$

where $V_{bi}$ is called the built-in potential. The space-charged region across the p-n junction, where the number of the mobile carriers is essentially zero, is called the depletion region, and its width $W$ is given by [2]:

$$W = \left[\frac{2\varepsilon_s}{q}\left(V_{bi} - V\right)\left(N_A^{-1} + N_D^{-1}\right)\right]^{1/2} \tag{5.9}$$

where $V$ is the applied voltage across the p-n junction. The built-in potential value, which depends on the specific semiconductor material, is, for example, 1.1 volts for GaAs.

According to Equation (5.9) the depletion-region width is reduced when the p-n junction is forward-biased, namely, the positive voltage is applied to the p-side relative to the n-side. To increase the depletion-region width, a layer of lightly doped intrinsic semiconductor is inserted between the p-n junction. Figure 5.2(a) shows the operation of such a structure, which is called p-i-n photodiode under reverse bias. The corresponding energy-band diagram, which is illustrated in Figure 5.2(b), shows the electron and hole movement through the diffusion and drift processes. Increasing the thickness of the intrinsic layer allows one to optimize the width of the depletion layer based on the photodiode requirements. It should be pointed out that increasing the intrinsic layer thickness also increases the photodetector response time, since it takes longer for the carriers to drift across the depletion layer. When light is incident on the p-side of the p-i-n junction, electron-hole pairs are created through absorption in the depletion region, where the large electric field separates the electrons and holes to the n-side and p-side, respectively. The generated photocurrent is proportional to the incident light power, where the proportionality factor $R$ is called the *responsivity* (A/W) of the photodetector. The quantum efficiency ($\eta$) of the photodetector is defined as the ratio

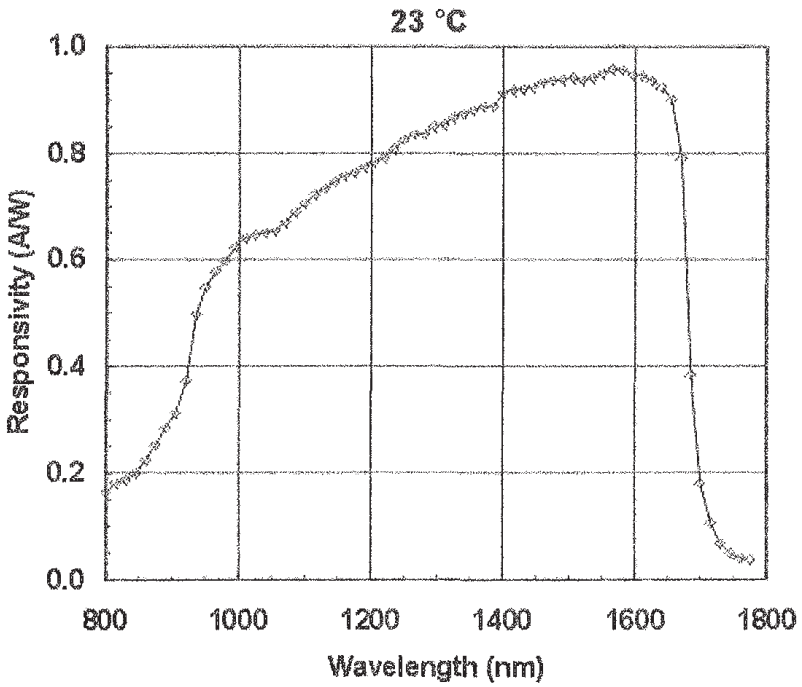



Figure 5.3 Typical spectral response of InGaAs p-i-n photodetector at 23°C. Courtesy of JDS Uniphase, Inc., Epitaxx division.

of the generated electron rate to the incident photon rate. The responsivity is related to the quantum efficiency through the following relation [5]:

$$R(\lambda) = \frac{q\eta(\lambda)}{h\nu} = \frac{q\lambda\eta(\lambda)}{hc} \tag{5.10}$$

where the wavelength dependence of the quantum efficiency is through the absorption coefficient $\alpha$. The measured spectral responsivity at room temperature of a typical InGaAs p-i-n photodetector is shown in Figure 5.3. Notice that the spectral responsivity of these photodetectors peaks around 1550-nm at 0.9-A/W, making them particularly useful for cable TV networks. Assuming the incident light has a normal incidence with reflectivity $r$, and each absorbed photon generates an electron-hole pair, the quantum efficiency is given by

$$\eta(\lambda) = (1-r)[1-\exp(-\alpha(\lambda)\cdot W] \tag{5.11}$$

According to Equation (5.11), almost 100% quantum efficiency can be achieved by increasing the depletion-region width. Notice that, as expected, $\eta \rightarrow 1$ as $\alpha \rightarrow 0$, assuming $r = 0$. To be useful in lightwave communication systems, photodetectors have to operate below the so-called cutoff wavelength ($\lambda < \lambda_c$), where $\alpha$ approaches zero. Both InGaAs and Ge semiconductors have high quantum efficiency (> 85%) in the 1.3-μm to 1.55-μm region. However, compared with InGaAs photodetectors, Ge photodetectors require high reverse bias voltages around 30V, and their quantum efficiency drops rapidly above 1.55-μm. Consequently, InGaAs photodetector are typically preferred.

It should be pointed out that the performance of p-i-n photodetectors could be improved by using a double-heterostructure design, where the intrinsic layer is sandwiched between n-type and p-type layers of a different semiconductor material such that the light is absorbed only in the intrinsic layer. For example, n-type and p-type layers of InP are transparent to the light in the 1.3–1.55-μm wavelength region, but are strongly absorbed by InGaAs intrinsic layer of the photodetector. By applying antireflection coating to the front facet of the photodetector, its quantum efficiency can be increased to almost 100%. Practical p-i-n photodetectors need ohmic contacts not only to apply the necessary voltage bias, but also to connect to other electronic devices. Ohmic contact is a metal-semiconductor contact that has a negligible contact resistance relative to the bulk resistance of the semiconductor.

## 5.2 Noise Sources in Lightwave Receivers

There are three fundamental noise mechanisms in p-i-n photodetectors when the incident optical signal is converted into electrical current. The noise mechanisms are (A) shot noise, (B) thermal noise, and (C) laser RIN noise. In the following subsections, we will review each of these noise mechanisms, which will better allow us to discuss the signal-to-noise ratio (SNR) or carrier-to-noise-ratio (CNR) of the received signal.

### 5.2.1 Shot Noise

Shot noise in photodetectors is a quantum noise, which is due to the random generation of electron-hole pairs when the photodetector is illuminated by photons. The shot-noise phenomenon was first investigated by Campbell in 1909 [6], Schottky in 1918 [7], and by many other researchers since then [8–9]. To derive the noise variance of the photocurrent generated in response to an optical signal with constant amplitude, let us make the following assumptions:

- The probability of generating a single electron-hole pair in a very small time interval $\Delta t$ is proportional to $\Delta t$.
- The probability of generating more than a single electron-hole pair in $\Delta t$ is negligible.
- The electron-hole pair generation events are statistically independent.

Based on these assumptions, the probability of generating exactly $n$ electron-hole pairs per unit time is described by Poisson statistics, and is given by

$$p(n) = \frac{N_0^n \cdot e^{-N_0}}{n!} \tag{5.12}$$

where $N_0$ is the average number of received photons in the time interval $\Delta t$, which is equal to $P_{in}\Delta t/h\nu$, where $P_{in}$ is the incident optical power and $h\nu$ is the photon energy. Let us also assume that every photon generates an electron-hole pair at the receiver (100% quantum efficiency), then, the average photocurrent is simply $qN_0$. The noise variance of the photocurrent at the receiver per unit frequency bandwidth is given by [5]

$$\sigma_{shot}^2 = \left\langle \delta i_R^2 \right\rangle - \left\langle \delta i_R \right\rangle^2 = q^2 \left[ \left\langle n^2 \right\rangle - \left\langle n \right\rangle^2 \right] = q^2 \cdot N_0 = q \cdot I_R \tag{5.13}$$

Mathematically speaking, $\delta i_R$, which is the photocurrent fluctuation, was assumed to be a stationary random process with Poisson statistics. Equation (5.13) is also the spectral density of the shot noise, which is frequency independent. If only positive frequencies are considered, the single-sided spectral density becomes $2qI_R$. Under reverse-biased operation, the dark current ($I_d$), which is the residual photocurrent with no light due, also adds to the photodetector shot noise. Thus, the total photocurrent shot noise variance per unit frequency bandwidth is given by:

$$\sigma_{shot}^2 = 2q\left(I_R + I_d\right) \tag{5.14}$$

### 5.2.2 Thermal Noise

The electrons move randomly in any conductor due to a finite temperature, which manifests itself as random fluctuations in the current even when no electrical voltage is applied. The random photocurrent fluctuations cause random voltage noise over a load-resistor terminal. The thermal noise is also called Johnson noise or Nyquist noise after the two scientists who first analyzed its behavior experimentally and theoretically [10, 11]. To calculate the effect of thermal noise, one has to recall Planck's relationship for the average energy per mode per unit frequency interval of a blackbody radiation, which is given by the following equation:

$$S_p(\nu) = \frac{2h|\nu|}{\exp\left[\dfrac{h|\nu|}{k_B T}\right] - 1} \tag{5.15}$$

where the factor of 2 accounts for the double-sided spectral density, $k_B$ (= $1.38\text{x}10^{-23}$ J/°K) is the Boltzmann constant, and $T$ is the absolute temperature [12]. Assuming low frequencies

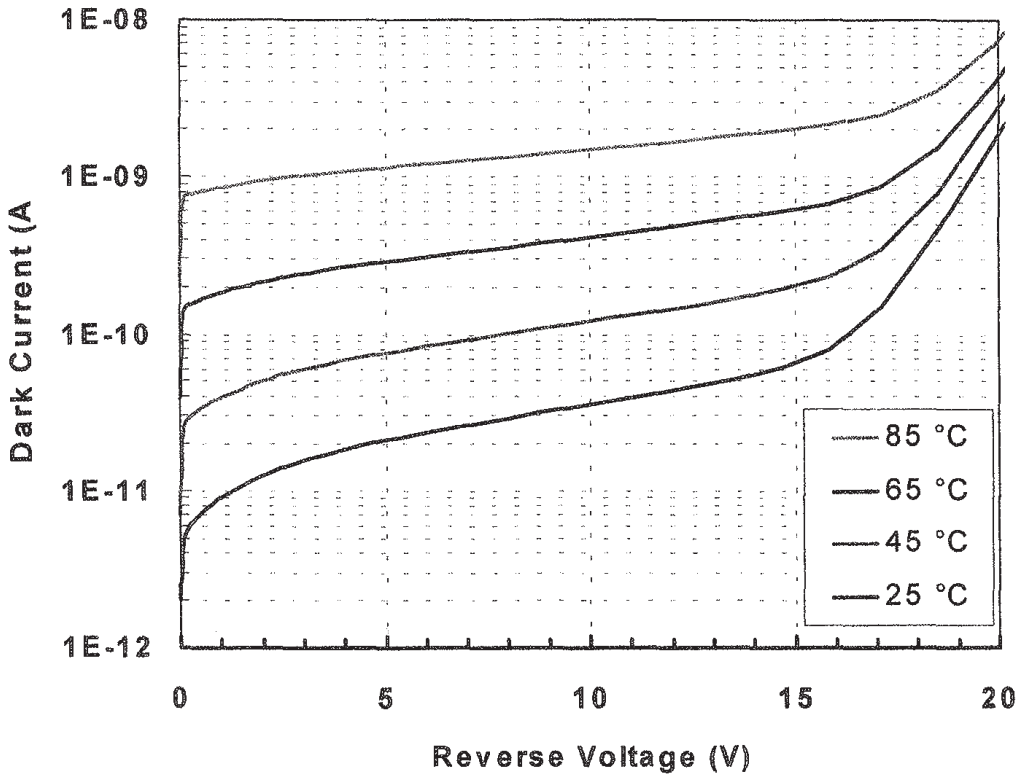



Figure 5.4 Measured dark current versus the reverse voltage at 25° C, 45° C, 65° C, and 85° C for InGaAs p-i-n photodetector. Courtesy of JDS Uniphase, Inc., Epitaxx division.

($h\nu << k_B T$) operation, and converting the voltage fluctuations to current fluctuations, the double-sided spectral density expression simplifies to

$$S_p(\nu) = \frac{2k_B T}{R_L} \tag{5.16}$$

where $R_L$ is the load resistor. The open-circuit single-sided spectral density of the photocurrent is given by

$$S_p(\nu) = \frac{4k_B T}{R_L} = \left\langle \delta i_{th}^2 \right\rangle \tag{5.17}$$

If an RF amplifier with a noise figure $F$ is connected directly to the p-i-n photodetector, then the photocurrent variance per unit frequency interval due to thermal noise is given by

$$\sigma_{th}^2 = \langle \delta i_{th}^2 \rangle = \frac{4k_B T \cdot F}{R_L} \tag{5.18}$$

At room temperature with a 50-ohm load and a preamplifier with a noise figure of 3, $\sigma_{th}$ = 31.5–pA/√Hz. The effect of thermal noise is demonstrated in Figure 5.4, which shows the measured dark current versus the reverse-bias voltage of a typical InGaAs photodetector at various temperatures. In particular, at low reverse-bias voltages, the dark current increases by almost three orders of magnitude for a 60° C temperature increase.

The thermal noise is sometimes expressed in terms of another useful parameter called noise-equivalent power (NEP), which is defined as the minimum optical power per unit of bandwidth that is required to produce SNR = 1. Therefore, the NEP can be written as

$$NEP = \frac{h\nu\left[\langle \delta i_{th}^2 \rangle\right]^{1/2}}{q\eta} = \frac{h\nu}{q\eta}\left[\frac{4k_B T \cdot F}{R_L}\right]^{1/2} \tag{5.19}$$

The NEP is useful to estimate the required optical power for a given SNR if the noise bandwidth B is known. Using the NEP definition, the noise-equivalent photocurrent $N_R$ can also be defined as NEPR, which has typical values in the range 1–10-pA/√Hz.

### 5.2.3 Laser RIN Noise

The laser relative intensity noise (RIN) is due to the laser spontaneous emission and fluctuations in the electron density. The laser RIN is defined as

$$RIN \equiv \frac{\langle \delta i_{RIN}^2 \rangle}{I_R^2} \tag{5.20}$$

where $\langle \delta i_{RIN}^2 \rangle$ is the photocurrent spectral density due to the laser RIN. Thus, the photocurrent variance due to the laser RIN is given by

$$\sigma_{RIN}^2 = \langle \delta i_{RIN}^2 \rangle = I_R^2 \cdot RIN \tag{5.21}$$

It should be pointed out that the laser RIN has frequency dependence similar to the small-signal modulation response of a DFB laser as was discussed in Chapter 3. In a lightwave communication system, the laser RIN is replaced by the system RIN, which includes the contribution of the various system elements such as the fiber RIN, laser RIN, and EDFA RIN.

## 5.3 Carrier-to-Noise Ratio at the Receiver

The performance of photodetectors in a lightwave communication system is typically expressed using the signal-to-noise ratio (SNR) or carrier-to-noise ratio (CNR). Let us assume an SCM communication system with $m$ modulation index per RF channel with a DC photocurrent of $I_R$ and an effective noise bandwidth $B$ at the receiver. Then, using the CNR (SNR) definition, the CNR can be written as

$$CNR = \frac{\langle i_R^2 \rangle}{\left[\sigma_{shot}^2 + \sigma_{th}^2 + \sigma_{RIN}^2\right] \cdot B} \tag{5.22}$$

where $\langle i_R^2 \rangle = (mI_R)^2/2$ is the mean-square signal photocurrent. Substituting Equations (5.13), (5.18), and (5.21) for the shot noise, thermal noise, and RIN noise, respectively, in Equation (5.22) for the CNR at the photodetector, one obtains [13]

$$CNR = \frac{(mI_R)^2}{2B\left[I_R^2 RIN + 2q(I_R + I_d) + \dfrac{4k_B TF}{R_L}\right]} \tag{5.23}$$

To gain further insight into Equation (5.23), let us analyze the CNR in the following three cases. Figure 5.5 illustrates the CNR behavior versus the received photocurrent according to Equation (5.23) as well as the thermal noise, shot noise, and laser RIN contributions to the CNR. We assumed that the photodetector was operating at room temperature with a preamplifier with a noise figure of 3, a 10-kΩ load resistor, laser RIN = −155 dBc/Hz, a 4% modulation index for the transmitted RF channel, a 4-MHz noise bandwidth, and the photodetector dark current was neglected. In most practical cases in which the incident optical power is very small (< −10-dBm), the thermal noise dominates over both the shot noise and the laser RIN in a p-i-n photodetector as shown in Figure 5.5. Therefore, the CNR becomes:

$$CNR = \left[\frac{R_L (mR)^2}{8k_B TBF}\right] P_{in}^2 \tag{5.24}$$

Equation (5.24) shows that the CNR increases as the square of the input optical power in the thermal noise limit. Furthermore, increasing the load resistor and reducing the noise figure of the amplifier can improve the CNR.

Another interesting limit is the shot noise limit, which is where the shot noise dominates over both the thermal noise and the laser RIN. In this limit, the CNR becomes

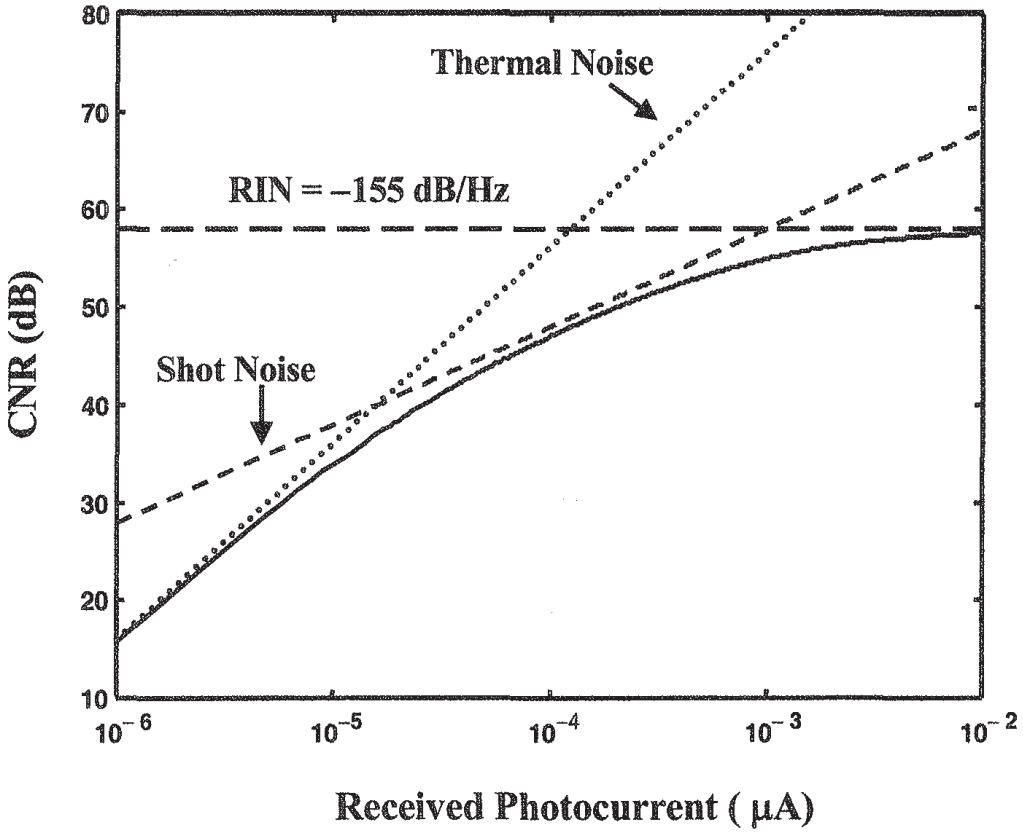



Figure 5.5 Calculated CNR and its components due to shot noise, thermal noise, and laser RIN noise versus the received photocurrent.

$$CNR = \left[\frac{\eta m^2}{4h\nu B}\right] P_{in} \tag{5.25}$$

Notice that the CNR increases linearly with the optical input power. Figure 5.5 shows that the shot noise competes with the thermal noise when a high-impedance load resistor is used (see Section 5.5.1).

The third limit is the laser RIN limit, which is where the laser RIN dominates over both the thermal and the shot noise. In this limit, the CNR does not depend on the photocurrent in the receiver and can be improved by reducing the laser RIN. This limit plays an important role at high optical-input power levels (> 0-dBm), where the CNR at the receiver is upper limited by the laser RIN. Consequently, to maximize the CNR at the fiber-node

receiver, directly modulated DFB laser transmitters with RIN of −155 dBc/Hz or less are typically required.

## 5.4 Nonlinear Behavior of p-i-n Photodetectors

As we have learned in Chapter 2, the transmission of AM-VSB video channels over cable TV networks imposes strict requirements on the magnitudes of the CSO and CTB distortions. This is because visual picture degradation of the transmitted analog channels can be observed if the magnitude of the nonlinear distortions at the subscriber's home is not sufficiently small. To achieve this requirement, the nonlinear distortions not only from the laser transmitter but also from the optical receiver must be sufficiently small to meet the CSO and CTB distortion requirements at the fiber nodes.

The need to develop high-fidelity analog and communications systems has produced many reports on the nonlinearities of p-i-n photodetectors [1, 14–16]. In particular, Williams, et al. analyzed the nonlinear behavior of a single-heterostructure InGaAs p-i-n photodetector [1]. Two primary mechanisms were identified in this device with intrinsic-region length of 0.95-μm, and they are (A) space-charge-induced nonlinearities, and (B) nonlinearities associated with absorption in undepleted semiconductor regions. When the photodetector is illuminated with a high-power optical beam, the carrier velocities and diffusion constant become a function of the carrier densities. The space-charge fields change the carrier velocities according to Equations (5.6) and (5.7) as well as the diffusion constants. Thus, for electric fields below 50-kV/cm, the dominant mechanism for the photodetector nonlinearity is the space-charge effect. As the electric field increases above 10-kV/cm (see Figure 5.1), the dependence of the carrier velocities decreases. Consequently, the contribution of the space-charge field to the photodetector second-order harmonic power is monotonically reduced as shown in Figure 5.6. Contrary to previous work, it was found that the photodetector nonlinearities are insensitive to the load resistor, except for very short devices (0.1–0.2-μm) or large incident beam spot size.

The dominant mechanism for the photodetector nonlinearity at high electric fields is carrier absorption in the undepleted highly doped InGaAs p-contact region, which is adjacent to the edge of the depletion. In contrast to the space-charge field-induced nonlinearity, the p-region carrier absorption mechanism is almost not affected by changing the electric field or the incident beam spot size. This is because the applied bias mainly affects the electric field in the photodetector intrinsic region. Also, the number of available free holes for transport in the p-region is significantly larger than the number of carriers generated by the incident light beam. Figure 5.6 summarizes the dominant regions for the measured second-order harmonic power as a function of the reverse-bias voltage. To minimize the effects of nonlinearities, p-i-n photodiodes should be designed with short intrinsic regions, not long ones. A shorter intrinsic region of the photodiode lends itself to a higher, more uniform, electric field. In addition, the photodiode should be illuminated with uniformly filled light beam spot size and packaged with proper thermal heat sinking.

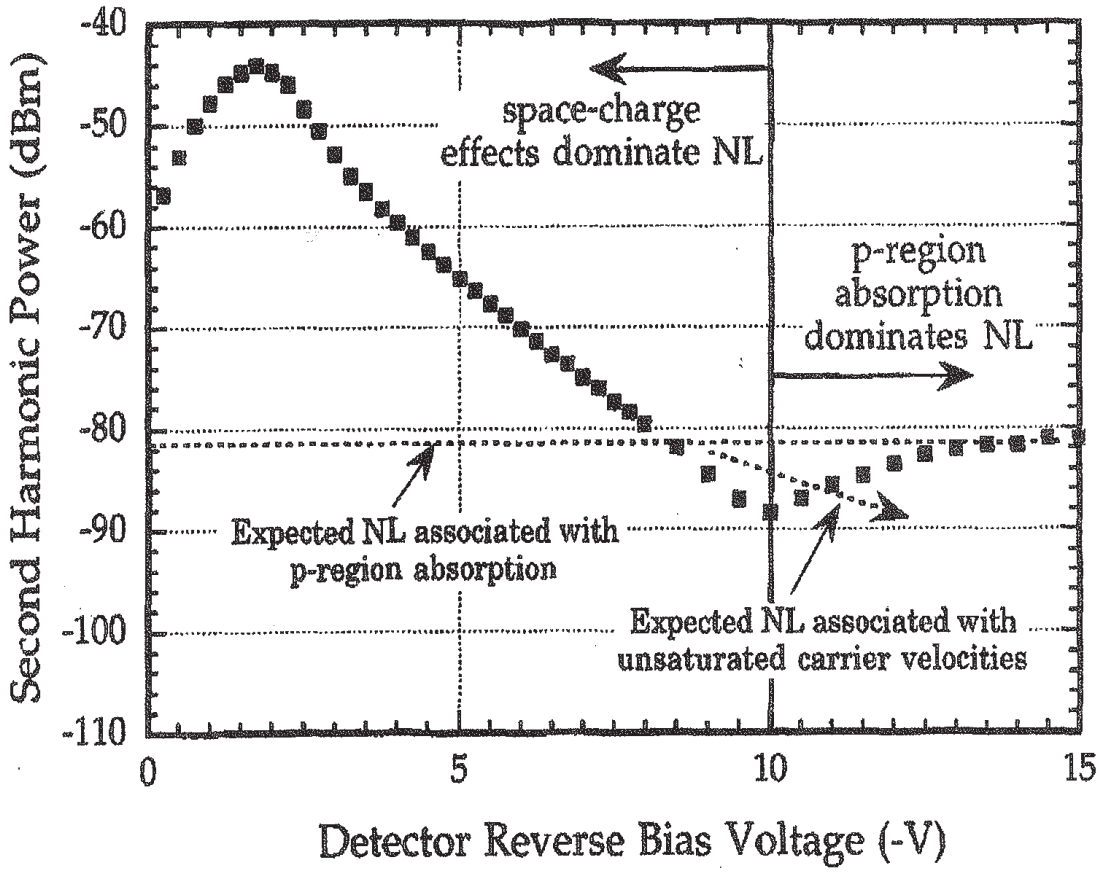



Figure 5.6 Measured second harmonic power versus the photodetector reverse-bias voltage, showing the regimes for the two dominating nonlinear mechanisms. After Ref. [1] (© 1996 IEEE).

The optical cable TV receiver at the fiber node is required not only to maintain its low second- and third-order nonlinear distortions over the reverse-bias voltage range, but also over the optical input power range as well as over the operating temperature range. For example, Figure 5.7(a) shows the measured second-order intermodulation distortion ($IM_2$) versus the input optical power for an InGaAs p-i-n photodetector operating at reverse bias voltage of –12V with 50-ohm load. The $IM_2$ was measured using two frequency tones at 400-MHz and 450.25-MHz, each with an optical modulation index of 0.7. Notice that the second-order intermodulation distortion of this photodetector is –75-dBc or less, even at a high optical input power level. This means that the photodetector nonlinear distortion contributions can essentially be neglected if the transmitted distortions are –65-dBc or higher. Figure 5.7(b) shows the measured second-order intermodulation distortions versus the reverse-bias voltage for the same InGaAs p-i-n photodetector for an incident optical power of 0-dBm using the

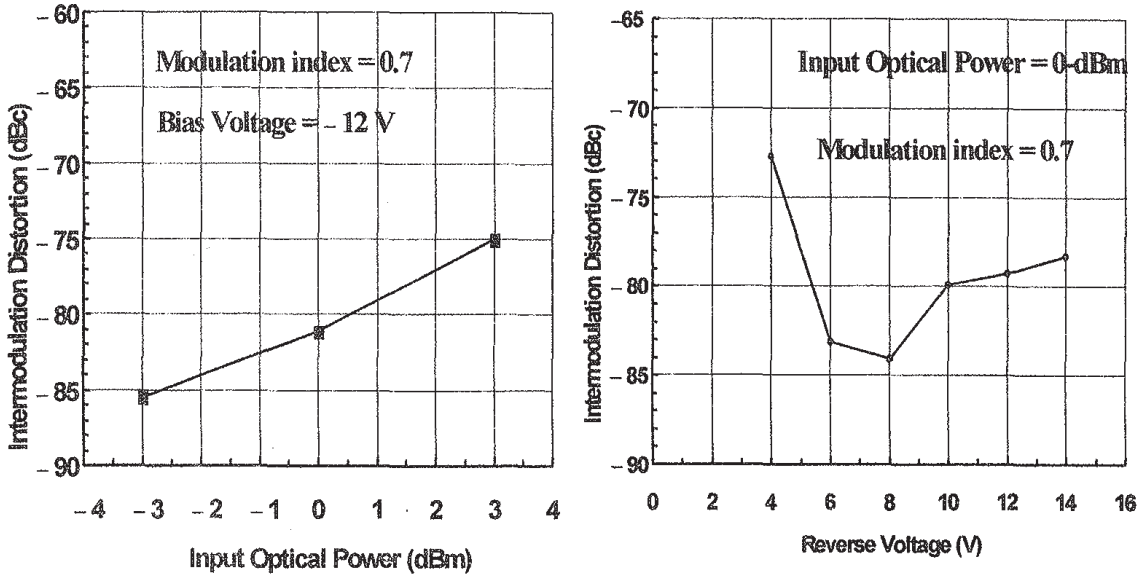



Figure 5.7 Measured second-order intermodulation distortions versus the optical input power and versus the reverse bias voltage for InGaAs p-i-n photodetector operating at modulation index of 0.7. Courtesy of JDS Uniphase, Inc., Epitaxx division.

same frequency tones as in Figure 7.7(a). This result indicates that if the photodetector is reverse-biased below about –8 volts, the nonlinear distortions start to increase rapidly.

**Table 5.1 Typical and maximum specifications of an InGaAs p-I-n photodetector used in 860-MHz cable TV networks.**

| Parameter | Specification |
|---|---|
| Reverse-Bias Voltage | 12 V |
| Maximum Forward and Reverse Current | 10 mA |
| Maximum Power Dissipation | 100 mW |
| Minimum Responsivity | 0.9 A/W at 1550-nm<br>0.85 A/W at 1310-nm |
| Maximum Second-Order Intermodulation ($IM_2$) | –75 dBc |
| Maximum Dark Current | 1 nA |
| Maximum Capacitance (with grounded case) | 0.35-pF |
| Back Reflection | –45 dB |
| 3-dB Bandwidth (into 50-ohm load) | 3-GHz |
| Operating and Storage Temperature | –40°C to +85°C |

Table 5.1 summarizes the typical and maximum performance parameters of an InGaAs p-i-n photodetector, which is normally used in the optical fiber nodes for 860-MHz cable TV networks. The photodetector is assembled into a hermetically sealed package with an antireflective-coated lens for efficient coupling to a fiber. Since the packaged photodetector is required to operate over wide temperature range, its operating bandwidth may be severely degraded due to optical misalignments and other packaging parasitics.

To minimize the impact of the nonlinear distortions of the photodiode in the optical receiver, the interface between the p-i-n photodetector and the preamplifier as well as the design of the preamplifier must be optimized. The basic front-end receiver design configurations as well as the overall receiver design considerations in cable TV networks are discussed in the next section.

## 5.5 Basic Cable TV Receiver Design Configurations

There are two primary categories of optical receivers, depending on their applications. Digital receivers are designed to receive digital baseband optical signals, and analog receivers are designed to receive RF or microwave signals, such as SCM analog video channels. Consequently, the performance requirements of digital receivers are different from those of analog receivers. Digital receivers require flat response at very low frequencies as well as high frequencies, depending on the transmitted data rates. For example, SONET receivers operating at OC-48 rate (2.488-Gb/s) must have flat response down to about 300-Hz. Thus, DC-coupled receiver design is needed for these applications. In contrast, analog receivers require flat response only in the frequency range of interest, namely, 50 to 860-MHz for downstream receivers and 5 to 42-MHz for upstream receivers. In addition, the overall analog receiver design, as was discussed in Section 5.3, must have good linearity to meet the strict CSO and CTB distortion requirements generated by the transmitted AM-VSB video signals.

Figure 5.8 illustrates a simplified block diagram of an optical receiver used in cable TV networks. The basic optical receiver configuration can generally be divided into two primary sections: (A) the front-end receiver section, and (B) the signal control and amplification section. The front-end section of the receiver consists of the p-i-n photodiode, a low-noise preamplifier, and some kind of interface circuitry between them. The role of the interface circuitry is to match the impedance between the photodiode and the preamplifier by minimizing the effect of stray capacitance as well as the input capacitance of the preamplifier. The output amplifier provides the necessary low impedance to match the coaxial-cable load. There are various types of interface circuits, depending on whether the front-end design is low impedance, high impedance, or transimpedance. For example, an impedance-matching transformer with some additional circuitry is used in a transimpedance front-end design in a fiber-node receiver. This is because the low-frequency response (i.e., less than 50-MHz) can be essentially ignored, while the high-frequency response is limited to about 860-MHz.

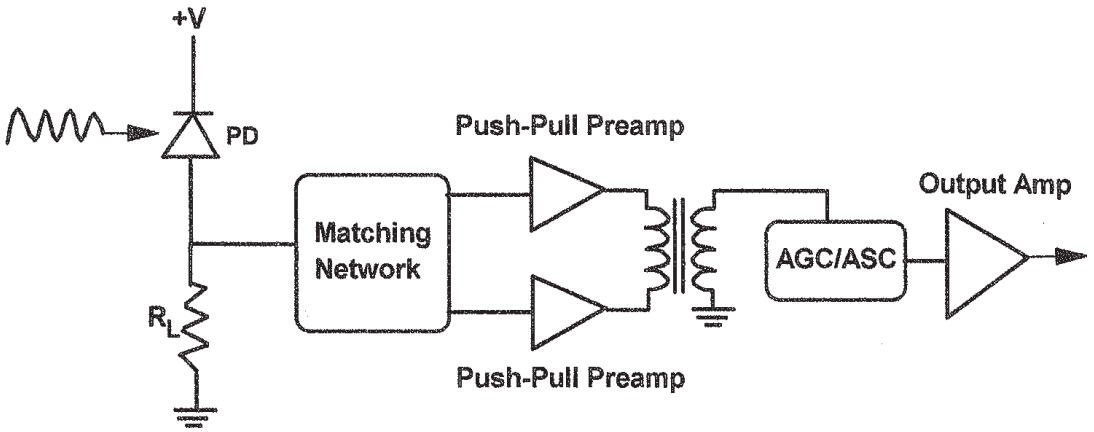



Figure 5.8 Simplified block diagram of a cable TV receiver using push-pull preamplifiers. AGC: automatic gain control. ASC: automatic slope control.

The role of the preamplifier is to convert the photocurrent to voltage with low noise and low nonlinear distortions. The preamplifiers can be, for example, bipolar-junction transistor (BJT) amplifiers or GaAs-junction field-effect transistor (JFET) amplifiers. Without the proper front-end receiver design, the photocurrent amplification may result in enhancement of the nonlinear distortions from the many transmitted analog video signals. The most common solution to this problem involves the use of dual matching preamplifiers in a push-pull configuration as shown in Figure 5.8. The push pull configuration works as follows. Let us assume that the input voltages from the matching network are 180° out of phase, namely, $V_{2,\text{in}} = -V_{1,\text{in}}$. The output voltage from each amplifier can be written as

$$V_1^{out} = b_1 \cdot V_{1,in} + b_2 \cdot V_{1,in}^2 + O(V_{1,in}^3)$$
$$V_2^{out} = -b_1 \cdot V_{1,in} + b_2 \cdot V_{1,in}^2 - O(V_{1,in}^3) \tag{5.26}$$

where third-order or higher order distortions are neglected. At the second transformer after the preamplifiers, the output voltages are recombined 180° out of phase, to yield $2b_1V_{1,\text{in}}$. Thus, the result of one amplifier pushing while the second pulling is to cancel the effect of second-order nonlinear distortions. The push pull configuration typically reduces the second-order nonlinear distortions by about 20 dB. The obvious disadvantage of this configuration is that it cannot provide any compensation for the third-order nonlinear distortions.

The push pull configuration offers a gain advantage in bipolar amplifiers, in which emitters are difficult to ground. For heterojunction bipolar transistor (HBT) amplifiers fabricated on an insulating substrate, adjacent push-pull emitters can share a common ballast resistor, thereby raising the gain. The penalty of push pull HBT design is the increased layout and

circuit complexity. Another advantage of push pull amplifiers is better isolation between adjacent HBTs, allowing the packaging of a higher gain amplifier. In emitter-coupled logic digital circuitry, a differential output is typically used in order to minimize the kick in the supply line. This supply line desensitization is important in complex ICs that operate above a few hundred MHz.

The receiver configuration following the front-end section consists of additional amplification stages with automatic gain and slope controls. As we learned in Section 2.4.2, the amplitude of the transmitted video signals must be equalized in order to maintain flat frequency response across the transmitted cable TV bandwidth.

In the following two subsections, we will discuss the front-end design considerations of analog receivers that are commonly used in optical communications systems. An improperly designed front-end circuit will often suffer from excessive noise associated with ambient light focused onto the detector.

### 5.5.1 Low- and High-Impedance Front-End Receiver Design

The front-end design of the photodetector is an essential part of the overall receiver design needed to meet the necessary performance requirements over cable TV networks. The simplest way to convert the photocurrent to voltage is to terminate the reverse-biased photodetector with a load resistor $R_L$, which is followed by a preamplifier as shown in Figure 5.9(a). The equivalent circuit model is shown in Figure 5.9(b), where the p-i-n photodetector is modeled as a current source in parallel with some shunt capacitance $C_p$. High-performance InGaAs p-i-n photodetectors are available today with shunt capacitance as low as 0.3-pF. The total capacitance, $C_T$, is equal to $C_p$ plus the stray capacitance that is associated with the input impedance of the preamplifier. According to Equation (5.23), a large $R_L$ is needed to increase the output CNR at low optical-input power levels. Such a front-end receiver design is called a high-impedance front end. However, the available modulation response bandwidth is proportional to $1/2\pi C_T R_L$. Thus, there is a trade-off consideration between achieving a large bandwidth and high CNR. One common way to extend the high-frequency response of the high-impedance design is to use a voltage equalizer after the preamplifier as shown in Figure 5.9(b). As the operating frequency is increased, the effective voltage drop across $R_1$ is reduced due to reduced reactance of $C_1$, while the voltage drop across $R_2$ is increased, resulting in increased frequency response. The output voltage divided by the input current or the transfer function of the equivalent circuit model with equalization is given by [17]

$$H(f) = \frac{V_R(f)}{I_R} = \frac{G \cdot R_L \cdot R_2}{R_1 + R_2} \tag{5.27}$$

and the extended frequency response is given by [17]

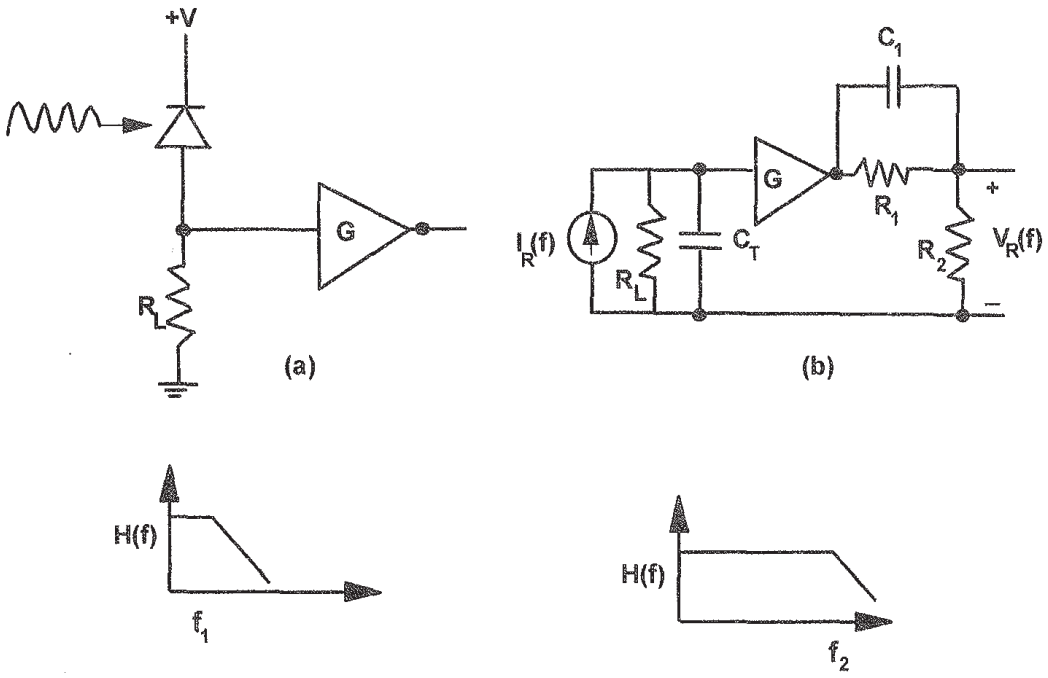



Figure 5.9 (a) A schematic of high-impedance front-end design, and (b) equivalent circuit model with equalization. After Ref. [17].

$$f_2 = \frac{R_1 + R_2}{2\pi R_1 R_2 C_1} = f_1\left[1 + \frac{R_1}{R_2}\right] \tag{5.28}$$

Even with the frequency equalization, the high-impedance front-end design has some problems. Because of the large load resistor, the high-impedance front-end design does not have a wide dynamic range. In addition, the high-impedance front-end design suffers from nonlinearity, particularly at high optical-input power levels. Reducing the load resistor to 50 or 75-ohms, which is sometimes called a low-impedance front-end design, improves the receiver linearity but also increases the noise level, resulting in a lower CNR. High-impedance front-end design is advantageous for applications in which a sensitive (i.e., low-noise) receiver with narrow dynamic range and frequency response is needed. The problems associated with high-impedance front-end receiver design are solved by transimpedance front-end receiver design, which is discussed in the next section.

### 5.5.2 Transimpedance Front-End Receiver Design

Figure 5.10 shows a schematic of the transimpedance front-end receiver design with the equivalent circuit model. In this design, the feedback resistor $R$, which is connecting the

output voltage to the input of an inverting amplifier, replaces the load resistor $R_L$ in the high-impedance front-end design. The amplifier acts as a buffer and produces an output voltage proportional to the photodiode current. The feedback resistor should be large to minimize the noise and maximize the output voltage $V_R$. The equivalent circuit model (Figure 5.10(b)) shows a feedback capacitance $C$, which can be much lower than the input capacitance, in order to improve the high-frequency response of the photodetector. The transfer function of the equivalent circuit model can be written as [17]

$$H(f) = \frac{V_R(f)}{I_R(f)} = \frac{-R}{1 + j2\pi fR\left[C + C_T G^{-1}\right]} \tag{5.29}$$

where the 3-dB frequency response is given by

$$f_c = \frac{G}{2\pi R_f\left[C_T + GC\right]} \tag{5.30}$$

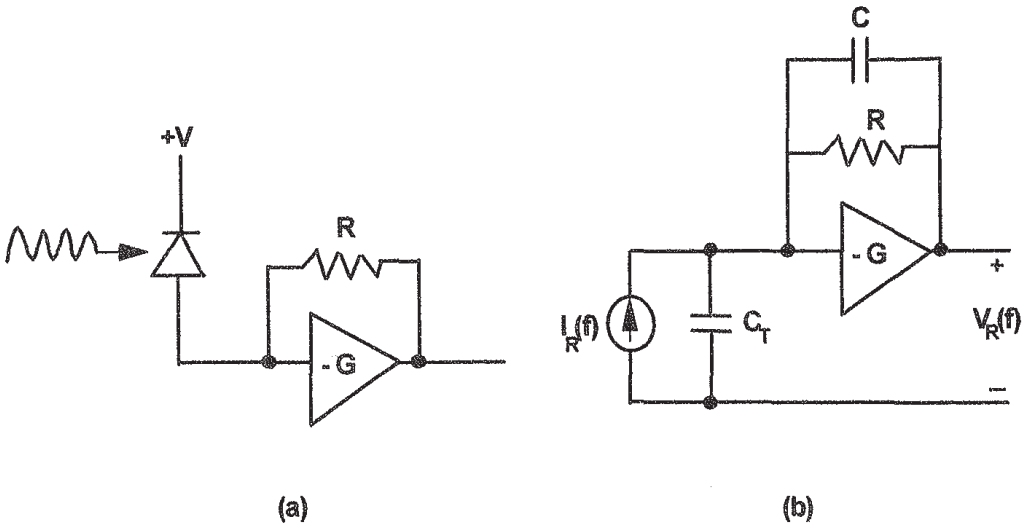



Figure 5.10 (a) A transimpedance front-end receiver design schematic, and (b) the equivalent circuit model. After Ref. [17].

It is apparent from Equation (5.30) that the reduced effective capacitance of the transimpedance front-end design allows the circuit to work at much higher frequencies. The other advantages of the transimpedance front-end designs are (A) an improved dynamic range over the high-impedance front-end design, (B) no need for external equalization circuits, and (C) an improved sensitivity over the low-impedance front-end design. The primary disadvantage of the transimpedance front-end receiver is a higher noise level than the high-impedance front-end receiver. In addition, similar to the high-impedance front-end design,

the circuit still uses a fixed resistor to convert the current to a voltage and is thus prone to saturation and interference from ambient light.

### 5.5.3 High-Performance Receiver Design for Cable TV

So far, we have shown that the transimpedance front-end design is advantageous for optical receivers operating over cable TV networks. However, the different circuit configurations of the front-end design have not been addressed. In general, two types of circuit configurations are used for transimpedance front-end preamplifiers in optical receivers: an inverter type [18], and a cascode type [19]. Figure 5.11(a) shows a typical circuit diagram of a transimpedance front-end photodetector that has two stages. The inverter-input stage consists of FET $Q_2$ with an active load $Q_1$, where the signal is inverted at the input to FET $Q_3$. The transimpedance preamplifier configuration consists of the input stage together with FETs $Q_3$ and $Q_4$, and the feedback resistor $R_f$. The second stage at the output of the transimpedance preamplifier, which consists of FETs $Q_5$ and $Q_6$, acts as a buffer stage to provide 50- or 75-ohms output impedance to the following receiver section. There are two voltage sources, namely, $V_{dd}$ (5 V) and $V_{ss}$ (–2.5 V). Although this configuration is attractive due to its simplicity, it suffers from a large Miller capacitance. It can be shown that the equivalent-input noise current variance for FET preamplifiers is given by [20]

$$\langle i_a^2 \rangle = \frac{4k_B T \cdot I_2 \cdot B}{R_f} + 2qI_{gate} I_2 \cdot B + \frac{4k_B T \cdot \Gamma}{g_m} (2\pi C_T)^2 \cdot I_3 \cdot B^3 \tag{5.31}$$

where $g_m$ and $\Gamma$ are the transconductance and noise factor of the FET at the input stage, respectively, $B$ is the receiver noise bandwidth, and $I_2$ and $I_3$ are normalized noise-bandwidth Personick integrals, which depend on the input and output pulse shapes [21]. The first term in Equation (5.31) accounts for the thermal noise of the load resistor, while the last term accounts for the thermal noise associated with the FET channel. The second term in Equation (5.31) arises from the FET gate leakage current. Another noise term, which is associated with FET 1/f noise, was neglected. The equivalent-input noise current using either BJT or JFET shows a cubic dependence on the photodetector bandwidth. However, the noise figures of silicon BJT are much higher, typically in the 7–9-dB range, than that of low-noise GaAs JFETs, which are in the 2–4-dB range. This is the reason why JFET preamplifiers are used, particularly in high-speed optical receivers.

The input capacitance $C_T$ is the sum of the photodiode capacitance and the FET capacitance CFET. The problem is that the FETs at the input stage of the preamplifier are operating in a common-emitter or common-source configuration, and thus suffer from the so-called Miller effect [20]. The input capacitance of the preamplifier consists of the gate-source capacitance and the Miller capacitance, which is the gate-drain capacitance multiplied by the FET gain. Thus, due to the Miller capacitance of the preamplifier, its input capacitance is increased, which reduces the preamplifier high-frequency response.

Minimizing the Miller capacitance by using the cascode preamplifier configuration as shown in Figure 5.11(b) solves this problem.

The cascode circuit configuration consists of a common-source or common-emitter FET $Q_3$, and a common-gate or common-base FET $Q_2$. Notice that the common-gate FET $Q_2$ controls the gain of the common-emitter FET $Q_3$. By matching the transconductance of $Q_2$ and $Q_3$, the current gain of the common-emitter FET can be unity, which minimizes the input capacitance to the preamplifier, producing a large feedback resistance. Further improvements in the transimpedance front-end receiver design can be achieved using monolithically integrated receivers, which is beyond the discussion here.

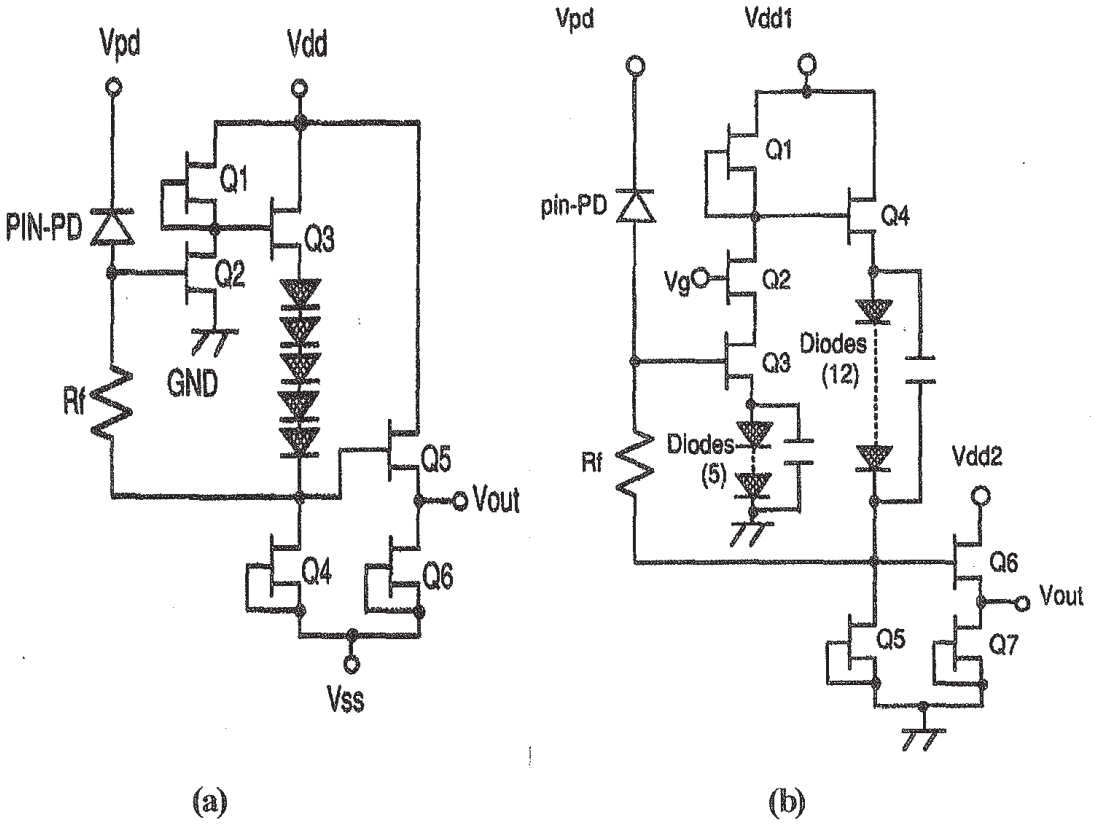



Figure 5.11 Transimpedance circuit configurations for photoreceivers (a) inverter-type preamplifier with multiple power supply, and (b) cascode preamplifier. After Ref. [18] (© 1996 IEEE).

## References


1. K. J. Williams, R. D. Esman, and M. Dagenais, "Nonlinearities in p-i-n Microwave Photodetectors," *IEEE Journal of Lightwave Technology* **14**, 84–96 (1996).
2. S. M. Sze, *Semiconductor Devices Physics and Technology*, John Wiley & Sons, New York (1985).
3. T. H. Windhorn et al., "The Electron Velocity-Field Characteristics for n-InGaAs at 300K," *IEEE Electron Device Letters*, **EDL-3**, 18 (1982).
4. P. Hill et al., "Measurement of Hole Velocity in n-type InGaAs," *Applied Physics Letters* **50**, 1260 (1987).
5. G. P. Agrawal, *Fiber-Optic Communication Systems*, John Wiley & Sons, New York (1992).
6. N. Campbell, "Discontinuities in Light Emission," *Proc. Cambr. Phil. Soc.*, **15**, 310–328, 1909.
7. W. Schottky, "Über Spontane Stromschwankungen in Verschiedenen Elektrizitätsleitern," *Annalen der Physik*, **57**, 541–567 (1918).
8. E. N. Gilbert and H. O. Pollak, "Amplitude Distribution of Shot Noise," *Bell System Technical Journal*, **39**, 333–350 (1960).
9. J. E. Mazo and J. Salz, "On Optical Data Communication via Direct Detection of Light Pulses," *Bell System Technical Journal*, **55**, no. 3, 347–369, (1976).
10. J. B. Johnson, Physical Review **32**, 97 (1928).
11. H. Nyquist, Physical Review **32**, 110 (1928).
12. F. Grum and R. J. Becherer, *Optical Radiation Measurements Volume 1: Radiometry*, Academic Press (New York, 1979).
13. T. E. Darcie, "Subcarrier Multiplexing for Lightwave Networks and Video Distribution Systems," *IEEE Journal on Selected Areas in Communications* **8**, 1240–1248 (1990).
14. T. Ozeki and E. H. Hara, "Measurement of Nonlinear Distortion in Photodiodes," *IEE Electronics Letters* **12**, 80 (1976).
15. M. Dentan and B. deCremoux, "Numerical Simulation of the Nonlinear response of a p-I-n Photodiode Under High Illumination," *IEEE Journal of Lightwave Technology* **8**, 1137 (1990).
16. R. R. Hayes and D. L. Persechini, "Nonlinearity of p-i-n Photodetectors," *IEEE Photonics Technology Letters* **5**, 70–72 (1993).
17. T. V. Muoi, "Optical Receivers," *Optoelectronic Technology and Lightwave Communications Systems*, Ed. Chinlon Lin, Chapter 16, Van Nostrand-Reinhold, Princeton (1989).
18. J. Yoshida, Y. Akahori, M. Ikeda, N. Uchida, and A. Kozen, "Sensitivity Limits of Long-Wavelength Monolithically Integrated p-i-n JFET Photoreceivers," *IEEE Journal of Lightwave Technology* **14**, 770–779 (1996).

19. N. Uchida, Y. Akahori, M. Ikeda, A. Kohzen, J. Yoshida, T. Kokubun, and K. Suto, "622 Mb/s High-Sensitivity Monolithic InGaAs-InP pin-FET Receiver OEIC Employing a Cascode Preamplifier," *IEEE Photonics Technology Letters* **3**, 540–542 (1991).
20. R. G. Smith and S. D. Personick, *Semiconductor Devices for Optical Communication*, Springer-Verlag, New York (1982).
21. B. L. Kasper and J. C. Campbell, "Multigigabit-per-second avalanche photodiode lightwave receivers," *IEEE Journal of Lightwave Technology* **5**, 1351–1364 (1987).

# CHAPTER 6

# OPTICAL FIBER AMPLIFIERS FOR CABLE TV NETWORKS

As we learned in Chapter 1, optical fiber amplifiers, particularly Erbium-doped fiber amplifiers (EDFAs), have been a key enabling technology in the revolution of cable TV access networks. The discussion in this chapter on EDFAs is not intended to be exhaustive, but to explain their basic characteristics and performance requirements for cable TV networks. We shall start this chapter with the review of the technology and operating principles of the key optical-fiber amplifier components, including WDMs, Erbium-doped fibers (EDFs), and pump lasers. In Section 6.2, the basic EDFA system configurations are introduced, including single-stage and dual-stage EDFA system configurations. The noise characteristics of optical amplifiers in different operating regions as well as the SNR and the amplifier's noise figure are derived for SCM lightwave transmission systems in Section 6.3. Finally, we will discuss in Section 6.4 the EDFA selection process for various applications such as long-distance super-trunking and broadband networks based on passive optical networks. In addition, the required EDFA's CNR and gain flatness for a single- and for multiple-wavelength cable TV networks are discussed.

Other types of optical amplifiers operating at 1300-nm, including praseodymium-doped fluoride fiber amplifiers, Raman fiber amplifiers, and semiconductor optical amplifiers, will not be discussed here. From a practical point of view, there are major technical and economical challenges that currently prevent these optical amplifiers from being used in cable TV networks.

## 6.1 Optical-Fiber Amplifier Components

The key EDFA system components are WDM multiplexers (WDM muxes) and demultiplexers (WDM demuxes), EDFs, and pump lasers. The next subsection discusses the characteristics and technologies for muxes and demuxes.

### 6.1.1 Wavelength-Division Multiplexers

Optical transmission through conventional SMF is typically characterized by the fiber's wavelength-dependent attenuation (dB/km) and chromatic dispersion (ps/nm/km). There are

two low-attenuation "windows" around 1310-nm and 1550-nm with average optical losses of 0.35-dB/km and 0.21-dB/km, respectively, for telecommunication applications [1, 2]. The total corresponding available bandwidths around 1310-nm and 1550-nm are 60-nm and 110-nm. The available electrical bandwidth can easily be calculated as follows:

$$\Delta f = -\left[\frac{c}{\lambda^2}\right]\cdot\Delta\lambda \tag{6.1}$$

For $\Delta\lambda$ = 1-nm, the electrical bandwidth is 172-GHz and 122-GHz around 1310-nm and 1550-nm, respectively. Thus, a standard SMF can provide an enormously large electrical bandwidth, particularly around 1550 nm (172-GHz x110 = 18,920-GHz ≅ 19-THz). Since it is impossible for a single-wavelength laser transmitter or a photodetector to utilize this enormous bandwidth, multiple single-wavelength laser transmitters with the corresponding WDM muxes, demuxes, and/or optical filters are used. Generally speaking, there are two categories of WDM filters, namely, passive and active. However, for cable TV access networks, passive WDM filters are preferred due to their lower cost and ease of use. Let us briefly review the different types of passive and active WDM filters.

There are four main technologies for passive WDM filters. They are as follows: (A) fused-fiber couplers, (B) graded-index (GRIN) rods and multilayer filters, (C) planar waveguides, and (D) fiber Bragg gratings. Fused-fiber couplers, which are commercially very common WDM devices, are made when two fibers are fused together; the coupling between the fiber cores determines the device response. The coupled powers in each fiber, namely, $P_1$ and $P_2$ can be described using the basic coupled-mode equations [3] as $P_0\sin^2(\kappa L)$ or $P_0\cos^2(\kappa L)$, where $P_0$ is the launch optical power in fiber 1, $\kappa$ is the coupling coefficient between the two fibers and $L$ is the coupling length. For example, by selecting a specific $\kappa L$ such that $P_2(\lambda_1) = P_0$ and $P_2(\lambda_2) = 0$, only one wavelength is passed through output port 2. The advantage of these devices is that by varying the coupling length, different coupling ratios such as 10/90% or 50/50% can easily be obtained. In addition, these devices can easily be scaled to 1 x $N$ or $M$ x $N$ couplers for DWDM applications. From a practical stand point, the packaging of these couplers can affect their reliability. In particular, the coupling region of these devices must be protected from environmental changes such as humidity and temperature.

The second type of passive WDM filter technology is a multilayer dielectric stack, which is typically placed between a pair of GRIN glass rods [4, 5]. The GRIN rods are used to collimate and efficiently couple the light beam from one fiber to another. The multilayer filters can be reflective or transmission types, depending on the thickness and refractive index at the designed wavelength. Thus, the reflected or transmitted light from the multilayer dielectric stack can be added constructively at one wavelength, and destructively at another wavelength. Such WDM filters, which are sometimes called interference filters, can be cascaded to produce multiple wavelength channels with desirable characteristics such as low insertion loss, polarization insensitivity, and a high isolation (> 30-dB) between channels. The first major application of these filters was 1550/1480-nm and 1550/980-nm WDM

muxes and demuxes for EDFAs [5]. Using cascaded WDM interference filters, eight-channel 200-GHz frequency spacing WDM demux with an insertion loss less than 5.5-dB and channel bandwidth greater than 0.5-nm has been achieved [5]. The advantage of these filters is that they provide flat amplitude across each channel with a low polarization-dependent loss (PDL). With the proper design and packaging, this type of WDM filter can be as reliable as fused-fiber filters.

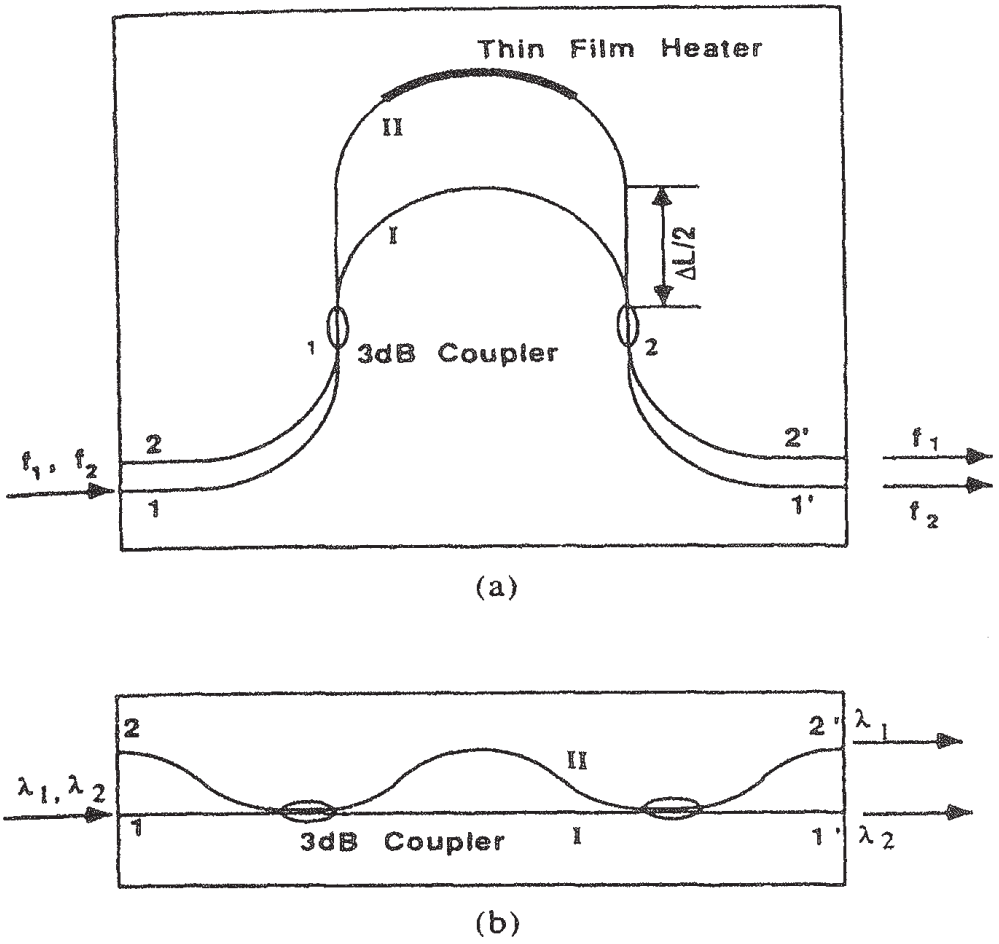



Figure 6.1 Basic configurations of MZ multiplexer/demultiplexer with (A) narrow and (B) wide wavelength spacing. After Ref. [6] (© 1990 IEEE).

The third type of passive filter technology is planar waveguides, which are fabricated on silica or silicon substrates using semiconductor-manufacturing processes. Figure 6.1 shows the basic configurations of a 2x2 Mach-Zehnder (MZ) multiplexer/demultiplexer [6]. It consists of two inputs and two output ports, two 3-dB couplers, and two waveguide arms, where there is a fixed length difference of $\Delta L$. The MZ configurations shown in 6.1(a) and 6.2(b) are designed for narrow and wide wavelength spacing (i.e., large or small $\Delta L$), respec-

tively. A thin-film heater is placed on one of the waveguide arms, which acts as a phase shifter, because the light-path difference of the heated waveguide arm changes due to the refractive index change. This method is used for a precise frequency tuning. The output power from ports 1 and 2 is $P_0\cos^2(\pi n\Delta L/\lambda)$ and $P_0\sin^2(\pi n\Delta L/\lambda)$, respectively, where $P_0$ is the injected power into port 1. In order to obtain $\lambda_1$ out of port 2 and $\lambda_2$ out of port 1, the injected wavelengths have to satisfy the following condition:

$$\frac{\pi n\cdot\Delta L}{\lambda_1}=\frac{\pi}{2}+\frac{\pi n\cdot\Delta L}{\lambda_2} \tag{6.2}$$

The required wavelength spacing is given by

$$\Delta\lambda=\lambda_2-\lambda_1=\frac{\lambda_1\cdot\lambda_2}{2n\cdot\Delta L} \tag{6.3}$$

The corresponding frequency spacing is simply $c/2n\Delta L$. For example, a path-length difference of 20-mm can provide an optical frequency spacing of 5-GHz, or a wavelength spacing of 0.04-nm at 1550-nm. On the other hand, if $\lambda_1$ = 1310-nm, and $\lambda_2$ = 1550-nm, then the path-length difference is only $\Delta L$ = 2.87-μm.

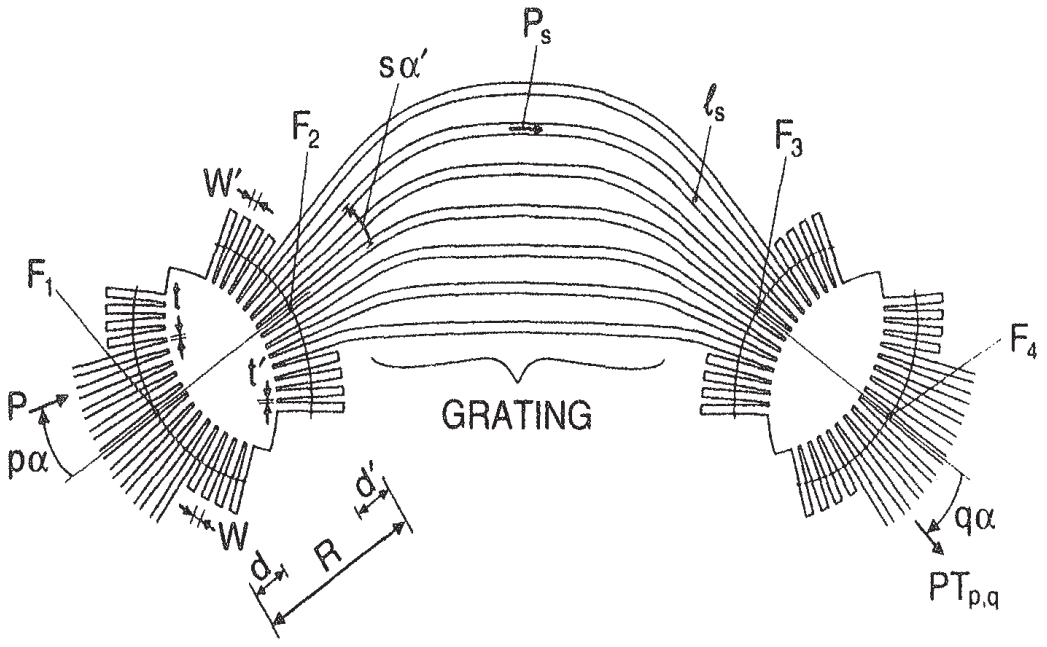



Figure 6.2 NxN multiplexer using two identical couplers and a grating. After Ref. [8] (© 1990 IEEE).

A generalized MZ multiplexer is shown, for example, in Figure 6.2 [7, 8]. An integrated *N*x*N* multiplexer fabricated using an $SiO_2$/Si waveguide, which consists of two identical star couplers combined with *M* waveguides acting as a grating between the couplers. Each star coupler consists of two confocal arrays of radial waveguides with foci $F_1/F_2$ and $F_3/F_4$. The waveguides are made from strips of a constant refractive index $n_2$ separated by strips of refractive index $n_1$. The multiplexer operates as follows: The input to the *p*th port is radiated in the free-space region, and most of its power is accepted by the grating. Some of this power is then transferred to the *q*th output port. The grating provides a constant path-length

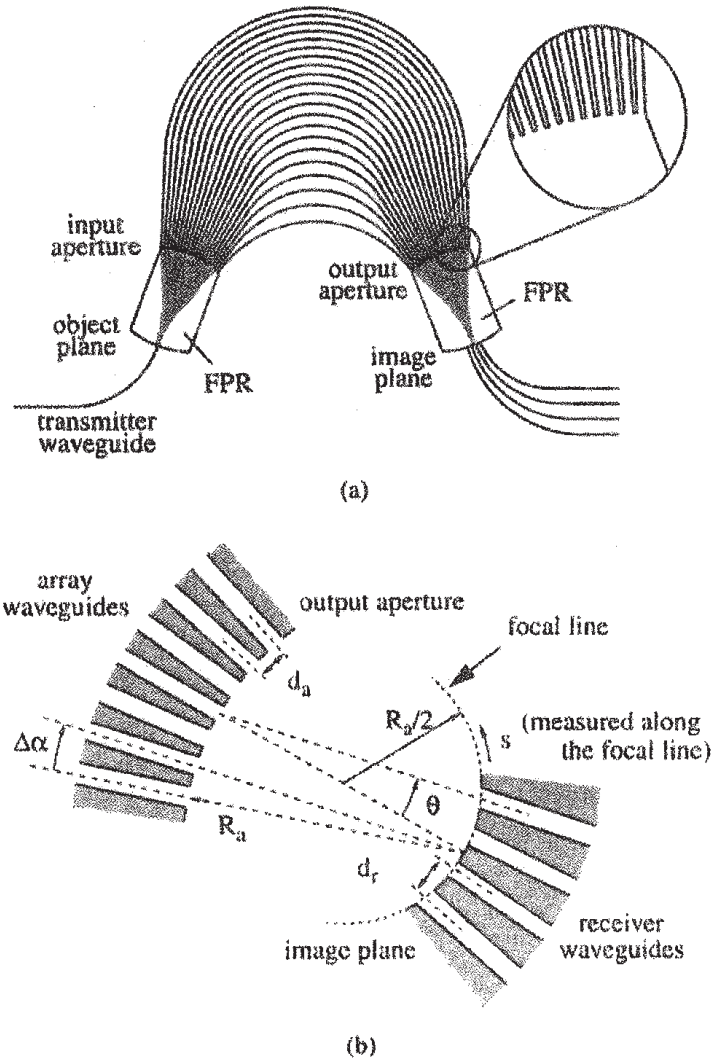



Figure 6.3 (a) Schematic layout and (b) receiver side geometry of the AWG demultiplexer. After Ref. [9] (© 1996 IEEE).

difference between adjacent paths such that the phase difference $\phi = \phi_s - \phi_{s-1}$ becomes independent of port $s$. The transmitted power coefficient varies periodically with $1/\lambda$, producing maximum transmission peaks at wavelengths for which $\phi$ is an integer multiple of $2\pi$. Such a multiplexer/demultiplexer, which can be scaled up to about 50 input and output wavelength channels, can also be used as a wavelength router. Other types of integrated-optics WDM devices are grating-based or phased array-based devices, which are also called array waveguide gratings (AWGs) [9]. These are imaging devices since they image the field of an input waveguide onto an array of output waveguides in a dispersive way. The schematic layout of the AWG multiplexer is shown in Figure 6.3(a) and its receiver geometry in Figure 6.3(b). Upon entering the free-propagation region (FPR), the laterally confined beam becomes divergent. At the input aperture, the propagating beam is coupled into the waveguide array to the output aperture. The length of the individual waveguides is selected such that the optical path difference between adjacent waveguides is equal to an integer multiple of the central wavelength ($\lambda_c$) of the AWG demultiplexer. This condition can be written as [10]

$$\frac{2\pi}{\lambda_C}\left[n_g \cdot \Delta L + n_{FPR} \cdot d_a\left(\theta_i + \theta_a\right)\right] = 2\pi m \tag{6.4}$$

where $n_g$ and $n_{FPR}$ are the effective refractive indices of the waveguide and the FPR, respectively, $d_a$ is the adjacent waveguide separation, and $\theta_i$ and $\theta_a$ are the incident and diffracted angles at the input and output aperture of the AWG demux, respectively. For the central wavelength, the fields in the individual waveguides will arrive at the output aperture with equal phase, and the field distribution at the input aperture will be reproduced at the output aperture. Angular dispersion is obtained by linearly increasing the length of the array waveguides. The outgoing beam for each wavelength is tilted, and the focal point is shifted along the image plane as shown in Figure 6.3(b). Consequently, spatial separation is obtained for different wavelengths at the receiver waveguides. It is found that the frequency response of the AWG is a periodic function with a so-called free-spectral range (FSR), where the frequency shift is equal to $2\pi$. The FSR can be obtained from Equation (6.4) as [9]

$$\Delta f_{FSR} = f_1 - f_2 = \frac{c}{n_g \cdot \Delta L} \tag{6.5}$$

where $f_1$ and $f_2$ correspond to adjacent order (i.e., $m$ and $m+1$) frequencies. The AWG demux can be used in other applications such as WDM routers and receivers. For example, multiple-channel WDM receivers can be achieved by integrating the AWG demux with a photodiode array [11].

Another type of passive filter technology is based on Bragg gratings, which are permanently written into the fiber by using either UV light or an excimer laser with a phase mask [12, 13]. The fiber Bragg grating (FBG) filters are formed by a periodic modulation of the refractive index of the fiber core as shown in Figure 6.4(a). When the transmitted light

beam is incident on the FBG, the partially reflected beams recombine constructively only if they satisfy the following condition:

$$\frac{\lambda}{n_{av}} = 2d \sin\theta \tag{6.6}$$

where $\lambda/n_{av}$ is the wavelength of the propagating beam in the fiber, $d$ is the spatial period of the grating, and $\theta$ is the angle of incident of the light beam as shown in Figure 6.4(b). The maximum reflection condition occurs when $\theta = \pi/2$, namely, all the reflected beam recombined constructively, and we have

$$\lambda_B \equiv \lambda / n_{av} = 2d \tag{6.7}$$

Equation (6.7) is called the Bragg condition, which states that the Bragg wavelength ($\lambda_B$) is simply equal to twice the spatial period of the grating. Thus, for a maximum reflection of 1550-nm light, the spatial period of the grating should be 0.53-μm, assuming $n_{av} = 1.47$ for a standard SMF. The FBG concept can be applied to fiber couplers to build optical bandpass filters or routers. For example, an FBG can be fabricated on each output arm of a 1xN fiber coupler in order to transmit or reflect only specific wavelengths. An FBG can also be used as a dispersion-compensation device in both digital and SCM lightwave transmission systems [14, 15].

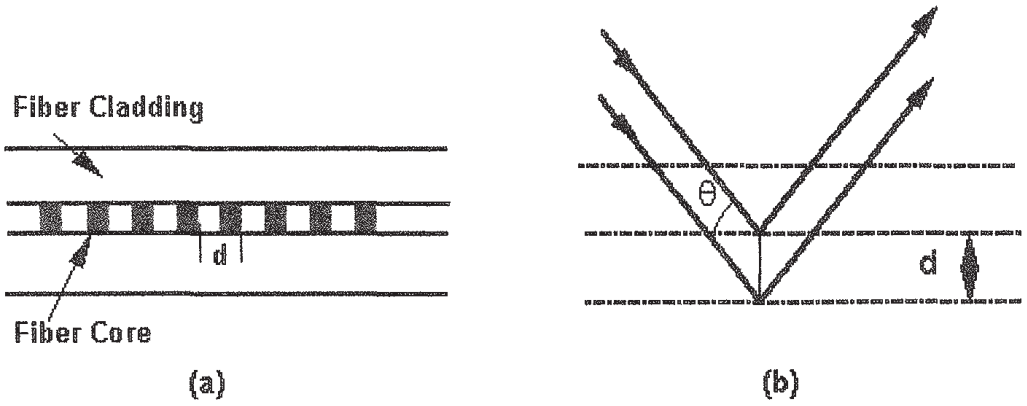



Figure 6.4 (a) Fiber Bragg grating formation by refractive-index modulation, and (b) incident and diffracted light beams at an arbitrary angle θ from the Bragg grating.

In general, active filters can be divided into three categories: (A) Fabry-Perot (FP) filters, (B) liquid-crystal-based filters, and (C) acousto-optic tunable filters (AOTFs). Both the FP and liquid-crystal-based filters, which are similar to the passive ones, are actively tuned by changing the mirror spacing using an RF voltage, and can provide wavelength spacing

between 0.5-nm and 10-nm [16, 17]. Another type of active filter is an all-fiber AOTF, whose spectral profile is electronically controllable [18]. The operation of the all-fiber AOTF can be explained as follows. When a flexural acoustic wave, which is generated by a coaxial acoustic transducer, propagates along a standard SMF, it creates antisymmetric microbends that travel along the fiber and introduces periodic refractive-index perturbations. This creates coupling between the input symmetric fundamental mode and the antisymmetric cladding mode when the acoustic wavelength is the same as the beat length between the two modes, which is the phase-matching condition. Thus, the optical filtering is achieved since the coupled light in the cladding mode is significantly attenuated. The all-fiber AOTF acts as an optical notch filter, where the center optical wavelength and the rejection amplitudes are tunable by adjusting the frequency and voltage of the applied RF signals to the acoustic transducer, respectively. As we will see in Section 6.4.2, the all-fiber AOTF can be used for gain flattening of EDFAs.

### 6.1.2 Erbium-Doped Fibers (EDFs)

Using rare-earth ions such as Erbium ($Er^{3+}$) to dope glass fiber produces the necessary gain medium in the EDFA system. Figure 6.5 shows the energy-level diagram of Erbium ions in silica fibers, showing both absorption and radiative transitions. The transition wavelengths are in nanometers and indicated only for transitions experimentally observed in silicate and flurozirconate Er-doped fiber. There are many different pump laser wavelengths such as 1480-nm, 980-nm, 800-nm, and 670-nm, which can be used to achieve population inversion. The notation of the possible states of a multi-electron atom such as Er is done according to ${}^{2S+1}L_J$, where $L = 0,1,2,3,\ldots$, corresponding to the letters S, P, D, F, G,..., respectively, is the overall angular momentum, $2S + 1$ is the number of spin configurations, and $J$ is the total angular momentum with $2J + 1$ possible states. For example, the notation for the $Er^{3+}$ ground state is ${}^4I_{15/2}$, which corresponds to $L = 6$, has a multiplicity $2J + 1 = 8$, and a spin multiplicity $2S + 1 = 4$. Figure 6.5 shows a simplified three-level energy model for the Erbium ions. First, the Erbium ions are excited from the ground state (level 1) to some higher energy state (level 3) by absorbing the pump photons. This process is sometimes called excited state absorption (ESA) [19]. The pump rate is $R_{13}$, and the stimulated emission to the ground state is $R_{31}$. The Erbium ions stay at level 3 for an extremely short time (≈1-ps), and quickly relax to a metastable state (level 2), which is characterized by a long lifetime ($\tau \cong 10$-ms for silica and fluoride-based glasses), predominantly through a nonradiative process. The metastable state (level 2) decays to the ground state through either a stimulated emission at a rate of $W_{21}$ or through a spontaneous emission with $A_{21} = 1/\tau$. The propagating signal is amplified through stimulated emission around 1540-nm, depending on the type of glass medium and the concentration of other doping elements. The absorption process by the ground state from higher energy states is sometimes called ground state absorption (GSA). In addition, there is a stimulated absorption between the ground state and level 2 at a rate of $W_{12}$. Generally speaking, optical amplifiers are more useful if they have a larger gain bandwidth.

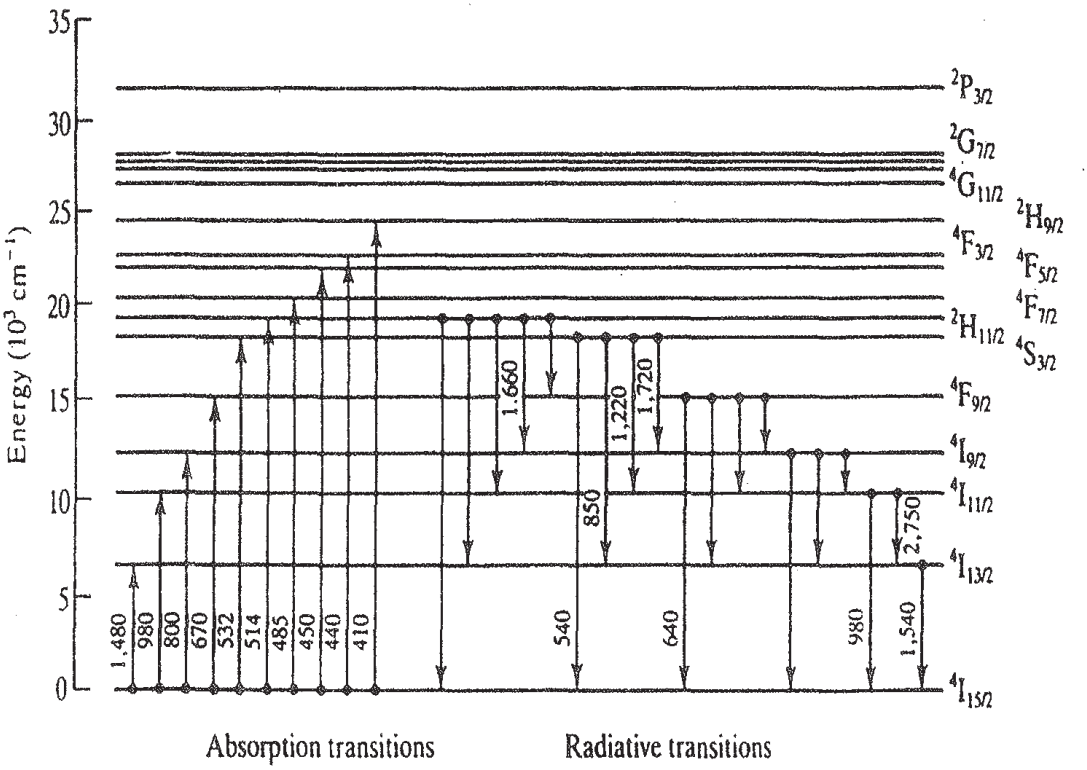



Figure 6.5 Energy level diagram of Er-doped glass, showing absorption and radiative transitions, where the transition wavelengths are in nanometers and indicated only for transitions experimentally observed in silicate and flurozirconate Er-doped fiber. After Ref. [19] (reprinted by permission of John Wiley & Sons, Inc.).

This requirement allows a larger number of optical channels to be multiplexed and transmitted over the communications network. It also relaxes the wavelength tolerance requirement for a single-channel system. It was found that by co-doping the silica-based glass with aluminum (Al) and/or germanium (Ge) together with Erbium ions, one could significantly increase the amplifier 3-dB gain bandwidth to about 40-nm [20]. Recently, a Tellurite-based EDFA with a gain exceeding 20-dB over 80-nm bandwidth from 1530-nm to 1610-nm has been demonstrated [21]. The relatively wide gain bandwidth of these Al/Ge silica-based glasses is attributed to the Stark splitting of the laser transitions induced by the Stark effect. The host's crystalline electric field removes the degeneracy of the energy levels. If, for example, levels 1 and 2 are split into manifolds of $m_1$ and $m_2$ nondegenerate sublevels, respectively, then the corresponding laser transition between these two levels is now made of the superposition of $m_1 m_2$ possible transitions. Thus, the laser transition is broadened. Since the Stark sublevels have small energy gaps and are therefore strongly coupled by the effect

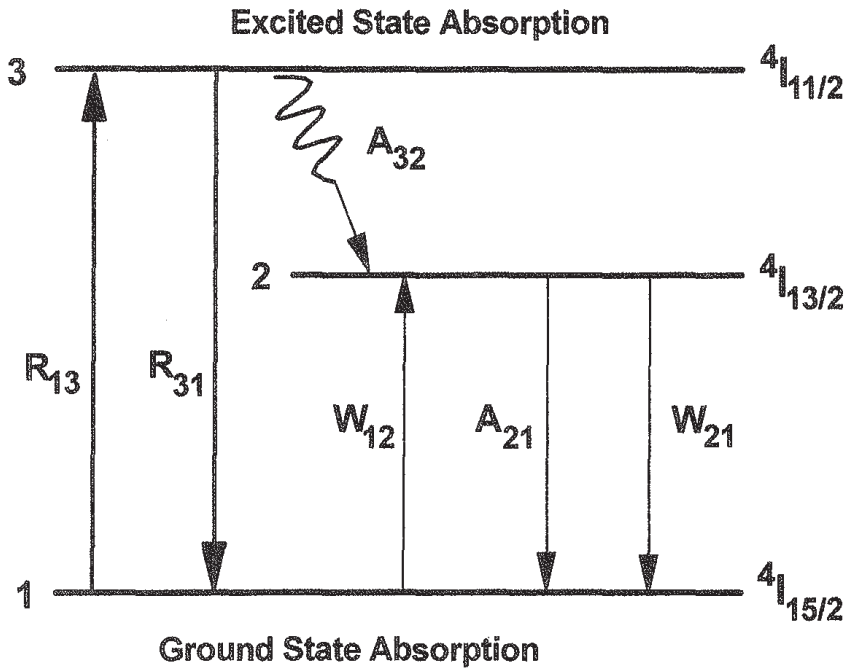



Figure 6.6 Simplified three-level energy model for Erbium. After Ref. [19] (reprinted by permission of John Wiley & Sons, Inc.).

of thermalization, the overall laser line exhibits homogeneous saturation characteristics. However, due to the presence of various defects and dislocations in the host crystals and the amorphous structure of glass hosts, the electric field that each dopant atom experiences varies from site to site, changing the energies of the Stark sublevels [19]. Thus, the spectral lines resulting from the superposition of all of the dopant atom contributions in such a host are inhomogeneously broadened. The inhomogeneous broadening of the emission spectra has been modeled by the Voigt profile, which is a convolution between the Lorentzian line shape and the Gaussian distribution, and has provided a good agreement with the experimental data [22].

The simplified three-level model of the Erbium ions can be analyzed using the rate equations. Let us denote the Erbium ions density by $N = N_1 + N_2 + N_3$, where $N_1$, $N_2$, and $N_3$ are the fractional ion density in energy levels 1, 2, and 3, respectively. The corresponding rate equations for these fractional population densities are given by [19]

$$\begin{aligned}\frac{dN_1}{dt} &= -\left[R_{13}+W_{12}\right]N_1+\left[W_{21}+A_{21}\right]N_2+R_{31}N_3\\ \frac{dN_2}{dt} &= -\left[W_{21}+A_{21}\right]N_2+W_{12}N_1+A_{32}N_3\\ \frac{dN_3}{dt} &= -\left[R_{31}+A_{32}\right]N_3+R_{13}N_1\end{aligned} \tag{6.8}$$

Let us now consider the steady-state regime, where the population densities are time invariant, namely, $dN_i/dt = 0$ (i = 1,2,3). To simplify the solution to Equations (6.8), let us make the following two assumptions:

1. $A_{32} >> R_{13}, R_{31}$, which means that the nonradiative decay rate $A_{32}$ dominates over the pumping rates to level 3. Thus, $N_3 \approx 0$, level 3 ions rapidly relax to level 2.
2. $A_{32} >> A_{21}$, which means that the lifetime of the Erbium ions in level 2 is significantly longer than level 3.

With these two assumptions, the steady-state solution to Equation (6.8) is given by

$$N_1 = N\frac{1+W_{21}\tau}{1+R\tau+\left[W_{12}+W_{21}\right]\tau} \tag{6.9}$$

$$N_2 = N\frac{\left[R+W_{12}\right]\tau}{1+R\tau+\left[W_{12}+W_{21}\right]\tau} \tag{6.10}$$

where $R = R_{13}$. The result of the steady-state population densities is important for the calculation of the amplifier's gain coefficient, and will be used later on.

To understand the amplifier gain, one has to consider the pump power, signal power, and amplified spontaneous emission (ASE) power distribution along the EDF length. The length of the EDF typically varies from a few meters to about 30 meters. The spatial evolution of the signal, pump, and ASE power along the fiber, where transverse power variations are neglected, can be described as [23]

$$\frac{dP_s(z,t)}{dz} = P_s\Gamma_s\left[\sigma_{se}N_2-\sigma_{sa}N_1\right]-\alpha_s P_s \tag{6.11}$$

$$\frac{dP_p^{\pm}(z,t)}{dz} = \pm P_p^{\pm}\Gamma_p\left[\sigma_{pe}N_2-\sigma_{pa}N_1-\frac{\alpha_p}{\Gamma_p}\right] \tag{6.12}$$

$$\frac{dP_{ase}^{\pm}(z,t)}{dz} = \pm P_{ase}\Gamma_s\left[\sigma_{se}N_2-\sigma_{sa}N_2-\frac{\alpha_s}{\Gamma_s}\right]\pm 2\sigma_{se}N_2\Gamma_s h\nu_s\Delta\nu_s \tag{6.13}$$

where $\sigma_a$ and $\sigma_e$ are the absorption and emission cross sections for the pump or the signal, and $\Gamma_s$ and $\Gamma_p$ are the signal-to-core overlap and pump-to-core overlap factors, respectively. The "+" superscript refers to both the pump and ASE copropagating with the signal, and the "–" superscript refers to when they are counterpropagating with the signal. The second term in Equation (6.13) represents the ASE power produced in the amplifier per unit length with its bandwidth $\Delta\nu$ for both polarization states (factor of 2). The loss terms $\alpha_s$ and $\alpha_p$ represent the internal loss for the signal and pump wavelengths, respectively. Neglecting these terms, Equation (6.13) can be simplified to

$$\frac{dP_{ase}^{\pm}(z,t)}{dz} = \pm\left[P_{ase}^{\pm} + 2n_{sp}\cdot h\nu_s\cdot\Delta\nu_s\right]g(z) \tag{6.14}$$

where the spontaneous emission factor is defined as

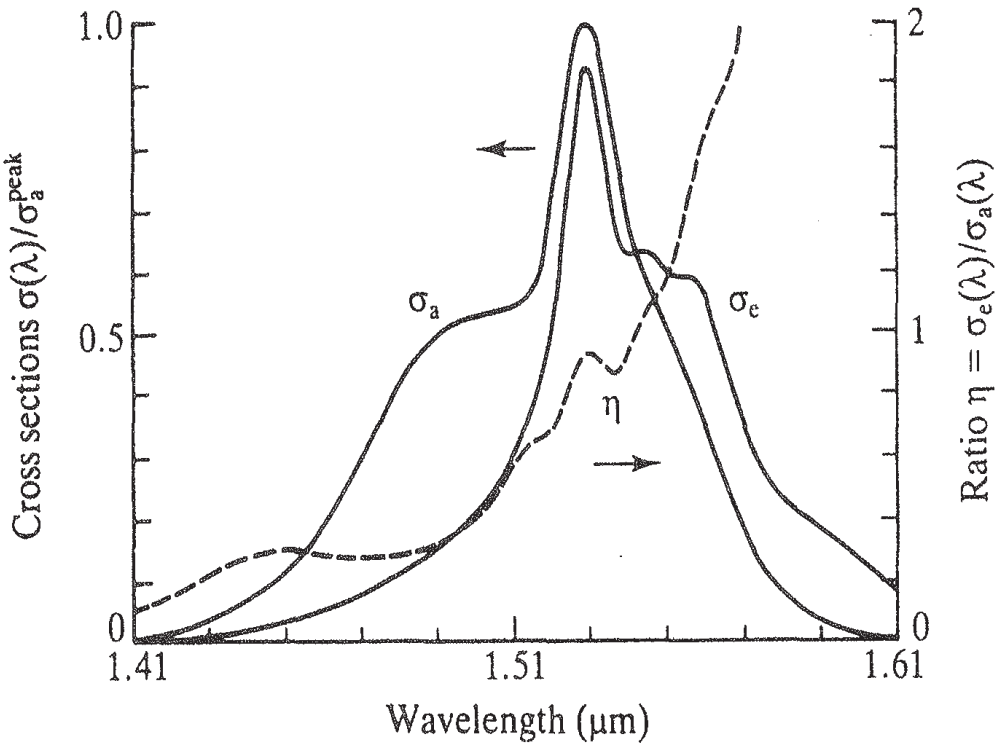



Figure 6.7 Wavelength dependence of absorption and emission cross sections ($\sigma_e$ and $\sigma_a$) as well as their ratio ($\eta$s) for a typical alumino-germanosilicate Er-doped fiber. After Ref. [19] (reprinted by permission of John Wiley & Sons, Inc.).

$$n_{sp} = \frac{\eta_s N_2}{\eta_s N_2 - N_1} \tag{6.15}$$

where $\eta_s = \sigma_{se}/\sigma_{sa}$, and $g(z) = \Gamma_s[\sigma_{se}N_2 - \sigma_{sa}N_1]$ is the gain coefficient independent of $z$, assuming the amplifier is pumped uniformly by a sufficiently strong pump. The steady-state solution to Equation (6.14) for an amplifier with a length $L$ is given by

$$P_{ase}(z) = e^{gL}P_{ase}(0) + g\int_0^L e^{g(L-z)}\left[2n_{sp}h\nu\Delta\nu\right]dz = 2n_{sp}[G-1]h\nu\Delta\nu \tag{6.16}$$

where $G = \exp(gL)$ is the amplifier gain, and we have assumed the initial condition $P_{ase}(0) = 0$, which is typically the case for a single-stage amplifier. Thus, Equation (6.16) provides the ASE power along a uniformly pumped single-stage amplifier with length $L$. The steady-state solution for $P_p$ and $P_s$ can be obtained by substituting Equations (6.9) and (6.10) for $N_1$ and $N_2$, respectively, into Equations (6.11) and (6.12). By allowing the pumping power to

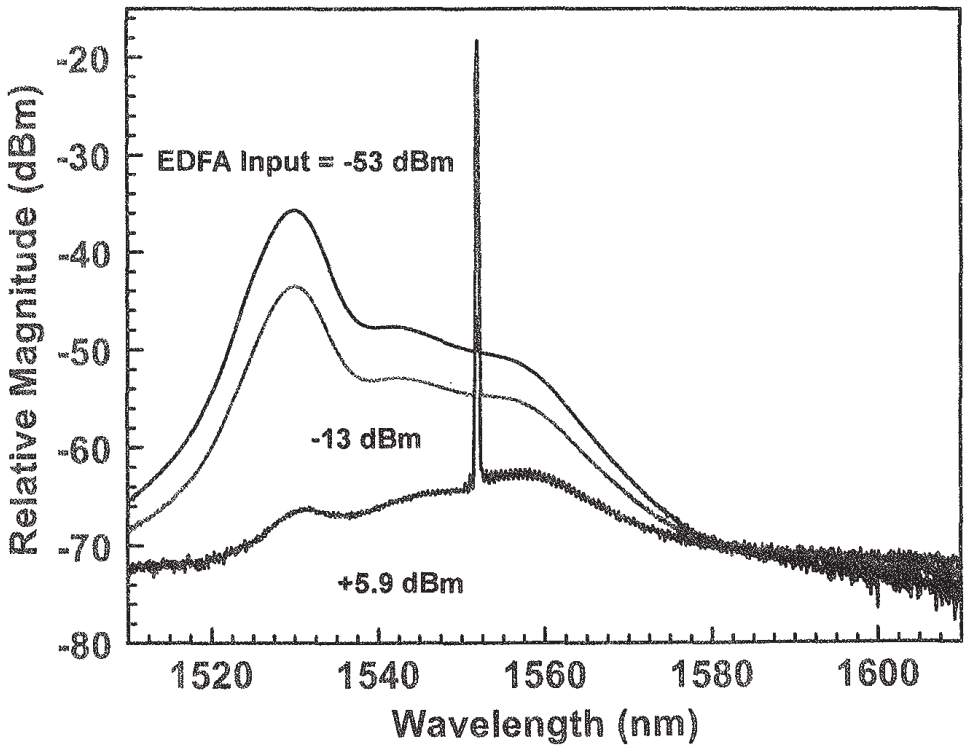



Figure 6.8 Measured optical spectra of a single-stage EDFA at various optical input power levels. Notice the dominant ASE peak at 1530-nm at very low optical input power level.

approach infinity ($R_{13} \rightarrow \infty$), Equation (6.11) for the signal-power spatial evolution simplifies to

$$\frac{dP_s}{dz} \cong \Gamma_s N \sigma_{se} P_s \tag{6.17}$$

Equation (6.17) implies that when complete population inversion is achieved, the signal gain is proportional to $\Gamma_s N \sigma_{se}$. Assuming the input signal power to the EDFA is $P_s(0)$, the maximum unsaturated gain for an EDFA with length $L$, which is described by the three-level model, is given by

$$G = \frac{P_s(L)}{P_s(0)} = \exp\left[\Gamma_s \cdot N \cdot \sigma_e \cdot L\right] \tag{6.18}$$

Thus, according to Equation (6.18), there is an optimal EDF length for a given pump power to achieve maximum signal gain. Furthermore, as the EDF length becomes longer than the optimal length, the pump power is not sufficiently strong to deplete the ground-state population, resulting in a partial absorption of the amplified signal. The spatial evolution of $P_p$ and $P_{ase}$ can be obtained numerically by substituting Equations (6.9) and (6.10) into Equations (6.12) and (6.13), respectively.

So far, we have not discussed the spectral shape of the EDFA gain, which is given by

$$g(\lambda) = \Gamma_s\left[\sigma_e(\lambda)N_2 - \sigma_a(\lambda)N_1\right] \tag{6.19}$$

The wavelength dependence of the absorption and emission cross sections as well as their ratio for a typical alumino-germano-silicate EDF is shown in Figure 6.7. Let us make two important points. First, the emission-to-absorption cross-section ratio ($\eta_s$) is typically larger at the longer wavelength region (i.e., > 1550-nm) than at the shorter wavelength region. This means that as the pump power is increased, optical amplification will be achieved first at the longer wavelength region. Second, the maximum overlap between the absorption and emission cross sections occurs around 1530-nm, which minimizes the net gain given by Equation (6.19). This means that in the small-signal regime one expects to find the ASE peak around 1530-nm in the EDFA gain spectra as shown in Figure 6.8. As the input signal to the EDFA increases, the gain spectrum near the saturating input signal ($\lambda_s$ = 1552-nm) compresses unevenly across the EDFA band. In fact, the ASE peak compresses about ten times more than the gain peak at the longer wavelength (> 1552-nm) region. The result is shifting of the gain-spectrum peak to around 1557-nm, while the ASE peak around 1530-nm becomes a secondary peak. The tilted amplifier gain spectrum has a significant impact on the transmission of AM video channels in an SCM lightwave system (see the discussion in Section 10.5). The uneven gain compression can be explained by the following simple

argument. For change in the manifold populations induced by the saturating signal $\Delta N_1 = -\Delta N_2$, the local gain coefficient decreases as

$$\Delta g(\lambda) = \Gamma_s\left[\sigma_e(\lambda)\Delta N_2 - \sigma_a(\lambda)\Delta N_1\right] = \Gamma_s\left[\sigma_e(\lambda) + \sigma_a(\lambda)\right]\Delta N_2 \tag{6.20}$$

Thus, the maximum gain change occurs where the sum of the emission and absorption cross sections is maximum, which typically occurs at a wavelength of 1531-nm for aluminosilicate EDFAs. It should be pointed out that the amplifier gain is primarily limited by the amplifier self-saturation by the ASE and laser oscillations. As the amplifier gain is increased, the stimulated emission by the ASE is also enhanced, until it competes with the pumping rate, located near the EDF ends. Thus, the amplifier gain becomes saturated. However, increasing the pump power also increases the amplifier gain. Oscillations in the amplifier occur when the optical gain is larger than the return losses of any reflecting element in the path of the ASE signal. Consequently, the reflected back-and-forth ASE in the amplifier operates as a laser, and the laser oscillation saturates to a steady state.

### 6.1.3 Pump Lasers

Semiconductor laser diodes are considered to be the most practical sources for lightwave communications systems due to their compact size, high gain coefficient (dB/mW), and excellent reliability. From a historical perspective, 1480-nm pump lasers were first employed in EDFA systems in the early 1990s due to their proven reliability [24]. With the development of the necessary packaging technologies with improved reliability, 980-nm pump lasers were used in EDFA systems by 1994 [25]. The performance of a particular EDFA design can be characterized by introducing the so-called EDFA gain coefficient (dB/mW), which is defined as the peak ratio of the small-signal gain to the injected pump power. Figure 6.6 compares the highest recorded gain coefficients (dB/mW) as a function of the input pump power using different pump bands near 1480-nm, 980-nm, 827-nm, 664-nm, and 532-nm.

The best gain efficiency of 11-dB/mW is obtained using 980-nm pump lasers, while the worst gain efficiency of 1.3-dB/mW is obtained using 830-nm pump lasers. Optimization of the EDFA gain coefficient requires (A) confining the $Er^{3+}$ ions doping to the center of the fiber core, and (B) reducing the pump mode size. Er-doping confinement concentrates the ions at the fiber center where the pump has its highest intensity. Reducing the pump mode size increases the pump intensity for a given pump level. Notice that the gain efficiency using 1480-nm pump lasers is only 60% of that obtained using 980-nm pump lasers. This is mainly due to incomplete population inversion (about 70%) with 1480-nm pump lasers compared with a complete inversion with 98-nm pump lasers. It should be pointed out here that because 980-nm pumped EDFA systems achieve complete population inversion, their noise figure is inherently lower than 1480-nm pumped EDFAs, approaching the quantum limit of 3-dB. The dB/mW gain coefficient increases as the length of the EDF is increased. However, it does not increase indefinitely, but levels off due to the amplifier self-saturation

effect. As the maximum gain of the amplifier is increased with longer EDF, more ASE power is generated at both amplifier ends, resulting in gain saturation. In addition to high output powers, pump lasers are required to produce stimulated emissions in a single transverse mode for efficient coupling into standard SMFs, but with multiple longitudinal modes (FP lasers). The three primary pump laser categories are (A) InGaAsP and multiple-quantum-well (MQW) InGaAs lasers operating at 1480-nm, (B) strained-layer MQW InGaAs lasers operating at 980-nm, and (C) GaAlAs lasers operating at 820-nm. 1480-nm pump lasers based on InGaAs/InGaAsP MQW producing a CW output power of 250-mW have been reported [24]. 980-nm pump lasers with improved reliability, which were based on strained-layer InGaAs MQW lasers, produced a maximum output power over 320-mW [25]. Because of the limited output power from conventional FP lasers, a new type of laser device based on a monolithically integrated master oscillator power amplifier (MOPA) has been developed [26, 27]. The MOPA device produced a single-lobed diffraction limited beam with CW output power up to about 2.2-W at 850-nm and 980-nm. However, the reliability of such devices is still under ongoing research.

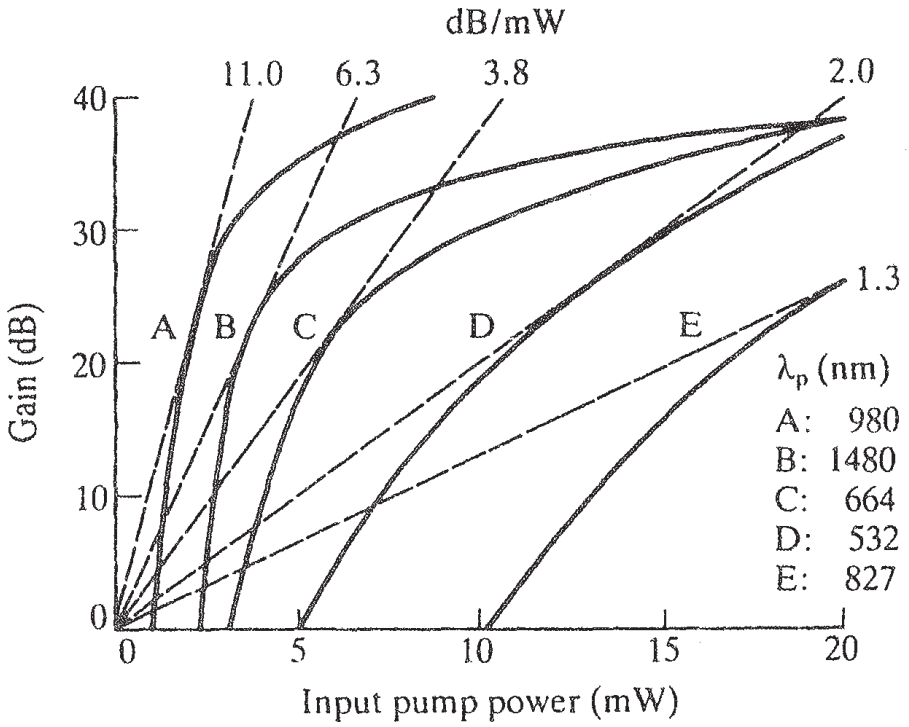



Figure 6.9 Record experimental EDFA gain versus input pump power characteristics using different pump bands near 532-nm, 827-nm, 664-nm, 980-nm, and 1480-nm. After Ref. [19] (reprinted by permission of John Wiley & Sons, Inc.).

## 6.2 Basic EDFA System Configurations

Three basic EDFA system configurations are illustrated in Figure 6.10 [19]. All of the EDFA system configurations consist of pump laser diodes; one or two WDM filters, which are also called wavelength-selective couplers (WSCs); a section of an EDF; and in-line optical isolators (OI). The first EDFA configuration [Figure 6.10(a)] has unidirectional forward pumping, since the pump wavelength is copropagating forward with the signal wavelength relative to the EDF. EDFs usually have smaller cores and higher numerical apertures (NAs) compared with standard SMFs. Thus, splicing between two dissimilar fibers must be done using nonstandard splicing techniques. In order to couple the light from the pump laser diodes into the SMF, different types of microlenses are used, depending on their characteristics. These microlenses can be, for example, spherical, aspherical, graded-index, or even fabricated directly on the SMF by melting the end of a tapered fiber [28, 29]. There are four basic requirements for good light coupling between pump lasers and SMF, and they are as follows: (A) large NA, (B) matching focal length for both the fiber transverse modes and the laser, (C) low spherical aberrations, and (D) low surface reflections. The last requirement is

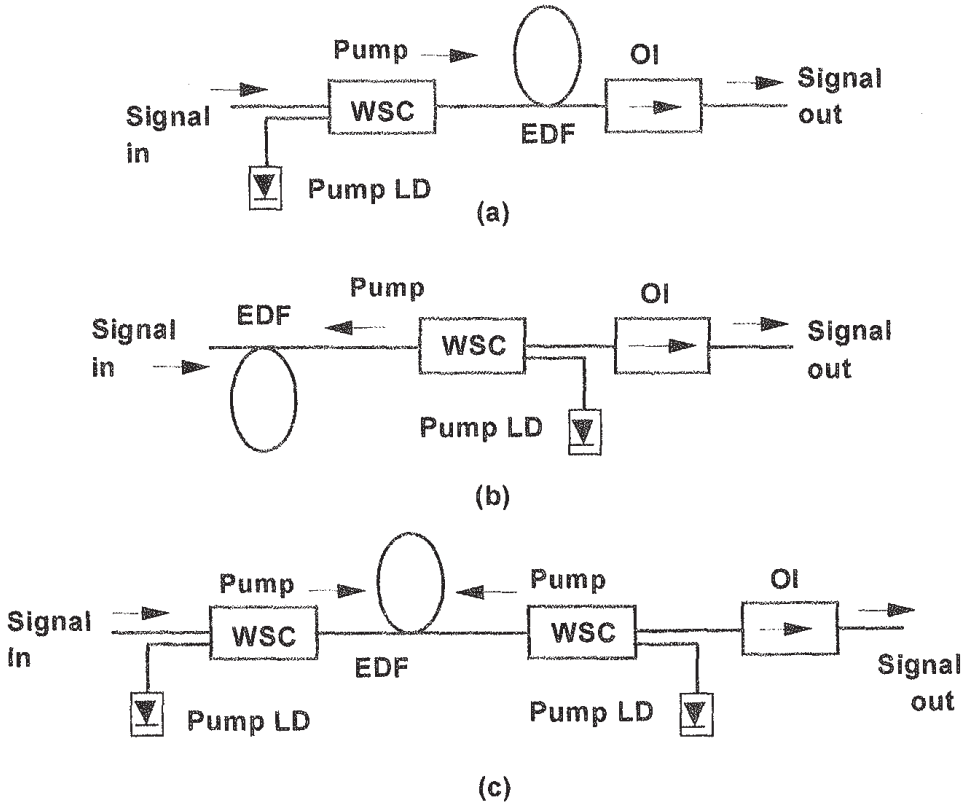



Figure 6.10 Basic EDFA configurations. After Ref. [19] (reprinted by permission of John Wiley & Sons, Inc.).

typically satisfied using antireflection (AR) coating. High LD-SMF coupling efficiency may provide additional power margin for the EDFA system. The proper design of the WDM filters also may be critical to the operation of the EDFA system. There are many requirements for an ideal WDM filter, depending on the particular technology. The ideal WDM filter should have (I) low coupling losses at both the pump and signal wavelengths, (II) polarization insensitivity, (III) low backward reflections, (IV) high-power damage threshold, (V) large optical signal bandwidth, (VI) temperature stability, (VII) compact size, and (VIII) cost-effectiveness.

Another important passive component is the OI, which prevents the backward-scattered light from coupling into the EDFA. In the second EDFA system configuration [Figure 6.10(b)], the propagating signal gain is obtained using unidirectional backward pumping. Thus, the injected signal is counterpropagating to the pump signal. As we will learn later, the lowest EDFA noise figure is always achieved when the pump signal is propagating in the same direction as the signal [19]. In the third EDFA system configuration [Figure 6.10(c)],

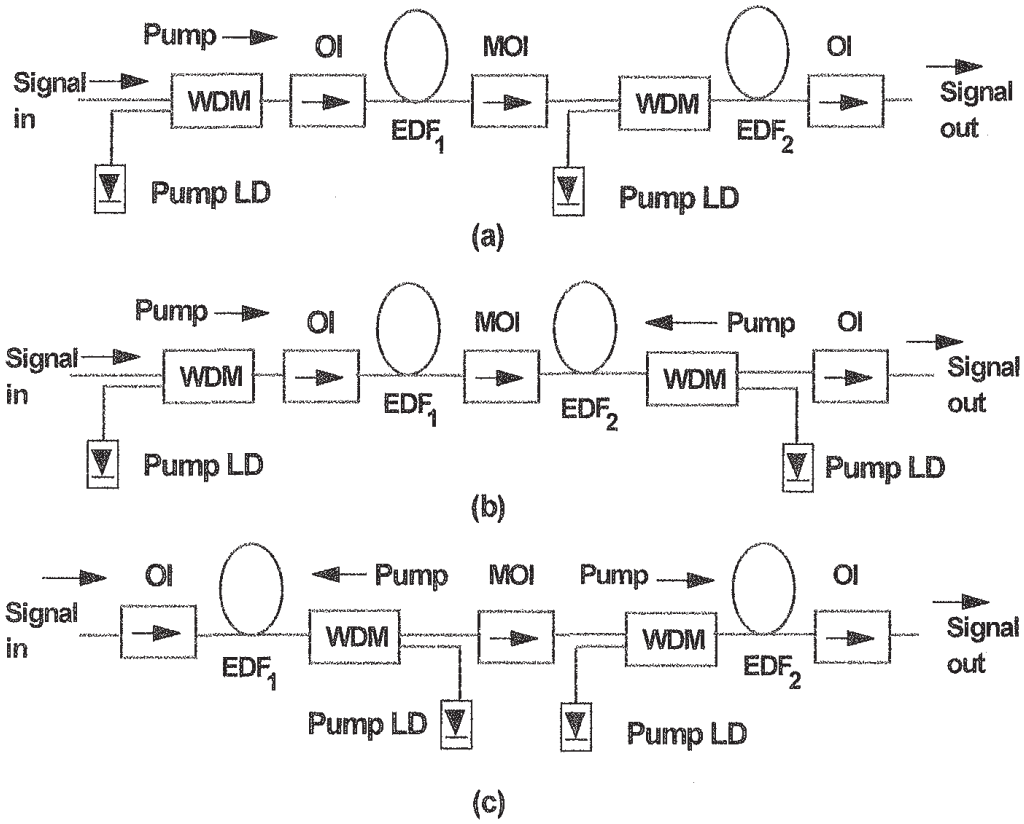



Figure 6.11 Dual-stage EDFA system configurations: (a) unidirectional pumping for both stages, (b) unidirectional pumping for first stage and backward pumping for the second stage, (c) the reverse-configuration (b).

bidirectional forward and backward pumping is employed from both ends of the EDF. Since the output power of today's 980-nm pump lasers is somewhat limited, polarization beam splitters (PBSs) may be used for a polarization-combined bidirectional-pumping scheme. From a practical aspect, it should be pointed out that a generic EDFA system also contains various electronic control subsystems including signal power control, pump power control, and temperature control. For example, the signal power control is achieved by tapping a small portion of the amplifier output power to a photodetector, where its voltage can be used to regulate the signal power using a microprocessor. The pump power can also be controlled by tapping a small portion of the pump power to another photodetector, where its output voltage regulates the power supply voltage.

So far, we have discussed only single-stage EDFA configurations. There are also dual-stage EDFA system configurations, in which two EDF sections are placed in tandem. These EDFA systems are designed to provide very high gains (G > 40 dB) and high output powers (> 20-dBm), particularly for DWDM cable TV networks. Figure 6.11 shows three possible dual-stage EDFA system configurations. In the first EDFA system configuration [Figure 6.11(a)] unidirectional forward pumping is employed in for both stages. In the second EDFA system configuration [Figure 6.11(b)], unidirectional forward pumping is employed for the first stage and a unidirectional backward pumping is employed in the second stage. In the third EDFA system configuration [Figure 6.11(c)], unidirectional backward pumping is employed for the first stage and unidirectional forward pumping is employed for the second stage. A midway OI and/or a narrow optical bandpass filter are typically employed with this EDFA design. The role of the midway OI is to suppress the buildup of an ASE peak at 1530-nm. The backward-propagating ASE from the second stage is prevented by the midway OI from entering into the first stage. The result is an improved EDFA gain and noise performance. The optimal location of the midway OI must be determined using numerical simulations. The combination of both the midway OI and optical filter is the most efficient way to suppress the amplifier self-saturation caused by ASE. Additional possible EDFA configurations, which have not been shown here, include bidirectional forward and backward pumping for each stage, and unidirectional backward pumping for each stage.

## 6.3 Amplifier Noise and CNR Calculation

Understanding noise generation in optical amplifiers and its impact on amplifier performance is a critical issue in lightwave communication networks. First, we will review how the new noise terms for optical amplifiers are generated. Then, the optical and electrical methods to measure amplifier noise figure will be discussed. Third, the calculation of the CNR for SCM lightwave systems will be explained.

### 6.3.1 Optical Fiber Amplifier Noise

In lightwave communication systems, optical amplifiers are used to amplify the propagating signals as in-line amplifiers, as power booster amplifiers (to the laser transmitter), or as pre-amplifiers (to the receiver). In order to determine the SNR or the CNR of the signal in these communication systems, one must analyze the noise contributions of the optical amplifier. Let us assume an optical amplifier with gain $G$ so that the output power is related to the input power by $P_{out} = GP_{in}$. Furthermore, let us assume that we have an optical receiver following the optical amplifier with an average DC photocurrent given by

$$I_R = \left[\frac{q \cdot \eta}{h\nu}\right] G \cdot P_{in} = R \cdot G \cdot P_{in} \tag{6.21}$$

where $q$ is the electronic charge, $\eta_D$ is the quantum efficiency of the photodetector, and we have assumed 100% coupling efficiency between the amplifier and the photodetector. Using Equation (6.16) for a single-stage amplifier, the photocurrent at the receiver due to the ASE is given by

$$I_{ASE} = R \cdot P_{ASE} = 2R \cdot n_{sp} \cdot h\nu[G-1]\Delta\nu \tag{6.22}$$

The current variance in the photodetector due to the ASE shot noise is given by

$$\sigma_{ASE}^2 = \left\langle i_{ASE}^2 \right\rangle = 2qI_{ASE} = 4q^2 \cdot \eta \cdot n_{sp}[G-1]\Delta\nu \tag{6.23}$$

Of course, there is also a shot-noise component due to the photocurrent $I_R$. Let us define a related quantity to the ASE power called the single-sided spectral density of the ASE, which is nearly constant and can be written as

$$S_{sp}(\nu) = \frac{P_{ASE}}{2\Delta\nu} = n_{sp}[G-1]h\nu \tag{6.24}$$

The effect of the ASE is to add noise fluctuations to the amplified power, which are converted to current fluctuations during the photodetection process. The primary contributions to the optical receiver noise come from the beating of the spontaneous emission with the signal, which is called a signal-spontaneous beat noise, and with itself, which is called a spontaneous-spontaneous beat noise. Using Equation (6.24), the variance of the photocurrent due to the signal-spontaneous beat noise is given by

$$\sigma_{S-ASE}^2 = \left\langle i_{S-ASE}^2 \right\rangle = 4I_R\left[R \cdot S_{ASE}\right] = 4\frac{(q\eta)^2}{h\nu} n_{sp} \cdot G[G-1]P_{in} \tag{6.25}$$

where a factor of 2 in Equation (6.25) is used because one needs to consider a double-sided ASE spectral density, and another factor of 2 is used to include both light polarizations. Notice that the signal-spontaneous (S-ASE) beat noise is independent of the amplifier optical bandwidth ($\Delta\nu$). Consequently, one cannot remove the ASE without eliminating the desired signal at the same time. In contrast, the spontaneous-spontaneous beat noise is proportional to the amplifier bandwidth. This means that a narrow bandpass optical filter can be used to remove this ASE beat noise. The variance of the spontaneous-spontaneous (ASE-ASE) beat noise is given by

$$\sigma^2_{ASE-ASE} = \left\langle i^2_{ASE-ASE} \right\rangle = 4\left[R \cdot S_{ASE}\right]^2 \Delta\nu = 4\left[q \cdot \eta \cdot n_{sp}(G-1)\right]^2 \cdot \Delta\nu \tag{6.26}$$

where the factor of 4 in Equation (6.26) is used for the same considerations as in Equation (6.25). When the amplifier gain G is high (i.e., small input power levels), the current variance of both the S-ASE and ASE-ASE beat noise is proportional to $G^2$, while the ASE shot noise is only proportional to $G$. Consequently, the ASE shot noise contribution can be neglected compared with the ASE-ASE and S-ASE beat noise contributions in the photodetector. We will use this approximation in the calculation of the SNR in the next section.

### 6.3.2 CNR and Noise-Figure Calculation

Now we are ready to analyze the CNR of fiber-optics communication system, particularly for cable TV applications. Let us assume an SCM communication system with an in-line single-stage optical amplifier, with $m$ modulation index per RF channel, and DC photocurrent of $I_R$. Then, using the CNR definition in chapter 5, the CNR can be written as

$$CNR = \frac{\left\langle i_R^2 \right\rangle}{\left[\sigma^2_{shot} + \sigma^2_{th} + \sigma^2_{RIN} + \sigma^2_{S-ASE} + \sigma^2_{ASE-ASE}\right]B} \tag{6.27}$$

where $\langle i^2_R \rangle = (mI_R)^2/2$ is the mean-square signal photocurrent. The photodetector shot-noise variance is given by the following equation

$$\sigma^2_{shot} = \left\langle i^2_{shot} \right\rangle = 2q(I_R + I_{ASE}) = 2q^2\eta\left[\frac{G \cdot P_{in}}{h\nu} + 2n_{sp}(G-1)\Delta\nu\right] \tag{6.28}$$

From a practical aspect, it is convenient to express the CNR given by Equation (6.27) in terms of the amplifier noise figure. The amplifier noise figure is defined as

$$F = \frac{SNR_{in}}{SNR_{out}} = \frac{CNR_{in}}{CNR_{out}} \tag{6.29}$$

where "in" and "out" refer to the SNR or the CNR measured at the output of the photodetector with and without the optical amplifier, respectively. Without the optical amplifier, the photodetector is essentially shot-noise limited. Consequently, the $CNR_{in}$ can be written as

$$CNR_{in} = \frac{(mI_R)^2}{4qI_R} = \frac{m^2 \cdot \eta \cdot P_{in}}{4h\nu} \tag{6.30}$$

In the presence of the optical amplifier, the photodetector is primarily limited by the shot noise and the signal-spontaneous beat noise. The ASE-ASE beat noise and the ASE shot noise contributions can be neglected compared with the signal-ASE beat noise contribution. The contribution to the inverse CNR caused by the signal-ASE beat noise can be written as

$$CNR_{S-ASE}^{-1} = \frac{8h\nu \cdot n_{sp}}{m^2 \cdot P_{in}}\left[1 - \frac{1}{G}\right] \tag{6.31}$$

The resultant CNR at the photodetector with the optical amplifier can be calculated according to

$$CNR_{out}^{-1} = CNR_{shot}^{-1} + CNR_{S-ASE}^{-1} = \frac{4h\nu}{m^2 \cdot P_{in} \cdot G}\left[1 + 2n_{sp}(G-1)\right] \tag{6.32}$$

Substituting Equations (6.30) and (6.32) into Equation (6.29), we get

$$F = 2n_{sp}\left[1 - \frac{1}{G}\right] + \frac{1}{G} \tag{6.33}$$

For a high-gain operation ($G >> 1$), the amplifier noise figure can be approximated by $2n_{sp}$. If we further assume an ideal optical amplifier with a complete population inversion, then $n_{sp} = 1$. This means that in the quantum noise limit, the amplifier has a noise figure of 3-dB [$= 10 \cdot \log_{10}(2)$]. Furthermore, neglecting the right-hand side term $1/G$ from Equation (6.33), the optical amplifier noise figure can be written as

$$F = 2n_{sp}\left[1 - \frac{1}{G}\right] \tag{6.34}$$

Substituting the noise figure expression given by Equation (6.34) in Equation (6.27) for the CNR, we obtain [30]

$$CNR = \frac{m^2}{2B\left[RIN + \frac{2q}{I_R} + \left(\frac{N_R}{I_R}\right)^2 + 2h\nu\left(\frac{F}{P_{in}}\right) + \Delta\nu\left(\frac{F \cdot h\nu}{P_{in}}\right)^2\right]} \tag{6.35}$$

where B is the photodetector electrical bandwidth, RIN is the overall communication system RIN, and $N_R$ is the photodetector noise equivalent current. Equation (6.35) is very useful expression since it describes the CNR dependence in terms of two primary measurable parameters, namely, the amplifier noise figure ($F$) and its optical input power. Furthermore, Equation (6.35) suggests that the transmitted CNR of the RF channel after the in-line EDFA is primarily governed by the signal-ASE and ASE-ASE beat noise at low optical input levels (< −10-dBm). In this input power regime, the amplifier noise figure approaches the quantum noise limit of 3-dB. If the noise figure is nearly unchanged when the EDFA is operating in saturation (optical input power ≥ 0-dBm), the AM CNR becomes limited by the receiver's thermal and shot noise at a given detected optical power. However, if the amplifier noise figure increases monotonically with the optical input power, then it sets the upper limit on the AM CNR as seen from Equation (6.35). As we will see later, robust SCM transmission of AM-VSB video channels often requires the EDFAs to operate in saturation.

### 6.3.3 Noise-Figure Measurement

The noise figure of an optical amplifier can be measured using either the optical or electrical methods. The optical measurement method uses an optical spectrum analyzer and/or an optical power meter. It is easily done, and provides accurate results in the small-signal regime (< −10 dBm). On the other hand, the electrical method, which uses an electrical spectrum analyzer, is a more accurate method when the amplifier operates in saturation. However, it is more complicated to measure.

The optical method is based on Equation (6.16) for the ASE power and Equation (6.34) for the amplifier noise figure in the small-signal regime. Substituting Equation (6.34) in Equation (6.16), one obtains

$$P_{ASE}(L) = F \cdot G \cdot h\nu \cdot \Delta\nu \tag{6.36}$$

Thus, the noise figure can be calculated according to Equation (6.36) by measuring amplifier small-signal gain $G$ ($G >> 1$) and ASE power. The ASE power can be measured by turning the input optical signal off and placing an optical bandpass filter with bandwidth $\Delta\nu$, centered at the signal wavelength. As the input signal power is increased, the amplifier saturates and ASE power becomes compressed. The compressed ASE power cannot be measured using this method, since when the input signal is turned off the compressed ASE becomes uncompressed. By adding the polarization nulling technique, one can still use the optical method to measure the compressed ASE power. The polarization-nulling technique uses a polarization controller and/or a polarizer between the optical amplifier and the optical

spectrum analyzer to null the input signal [31]. However, there are two problems with the optical method when the amplifier is operating in saturation. First, residual spectral components of the ASE signal may still exist, creating a measurement uncertainty. Second, the measured noise figure may be overestimated due to a phenomenon called polarization hole burning (PHB) [32]. The magnitude of the compressed ASE noise measured in the "orthogonal" polarization may be significantly different than in the "parallel" polarization, particularly when the optical amplifier is operating in a deep saturation. Since only the compressed ASE with "parallel" polarization generates the signal ASE beat noise, the measured "orthogonal" ASE may not be very reliable under these conditions. Furthermore, interferometric noise effect caused by multiple optical reflections can add uncertainty to the optical measurement.

The electrical method is based on using Equations (6.34) and (6.35), requiring the use of an electrical spectrum analyzer, particularly when the amplifier is saturated. For an SCM lightwave communication system, the CNR of the transmitted RF channel is measured on an electrical spectrum analyzer both with and without the optical amplifier as the optical input power is increased. In the large-signal regime, the measured CNR of the RF channel is limited by the shot and thermal noise at the photodetector. The RIN in Equation (6.35) is the overall system RIN noise, including RIN from multiple reflections. Using Equation (6.34), the noise figure ($F$) can be easily be calculated at various optical input power levels to the amplifier. It should be pointed out that for DWDM cable TV networks, the EDFAs are typically operating in saturation, where it is advantageous to use the electrical method. The noise characteristics of EDFAs operating in the saturation regime are discussed in the next section.

## 6.4 EDFA Requirements for Cable TV Networks

### 6.4.1 EDFA's Noise-Figure Requirement

As we learned earlier, the CNR of the transmitted RF channel through a high-power in-line EDFA is limited by the EDFA noise in the small-signal regime. Moreover, the transmitted CNR may be further reduced in a cascaded EDFA link, which is often required to transport the AM video channels, due to EDFA noise buildup. Thus, selecting the proper EDFA system design configuration is essential to meet the overall transmission requirements. The selection process is best illustrated by the following comparison. Two in-line 980-nm pumped-EDFAs with different characteristics but with the same small-signal noise figure have been selected. The experimental setup for the multichannel AM-VSB video lightwave transmission system consisted of a 1552-nm EM-DFB laser transmitter with a built-in power amplifier, 75-km of a standard SMF with an in-line 980-nm pumped EDFA, and an optical receiver [30]. The EM-DFB laser transmitter has built-in mechanisms for a predistortion linearization and the suppression of SBS. The SBS effect converts the transmitted optical

signal in the fiber to a backward-scattered one, and thus limits the maximum optical power that can be launched into the standard SMF [33]. Two types of 980-nm pumped EDFA were used as in-line amplifiers, which are labeled EDFA type A and type B. EDFA type A is a single-stage single-pump amplifier with a 980-nm unidirectional forward pumping. EDFA type B is a dual-stage amplifier with a 980-nm unidirectional backward pumping for the first stage, and unidirectional forward pumping for the second stage. The saturated output power levels of both EDFA type A and B are 14.2-dBm and 16.2-dBm, respectively. Both EDFA types were designed to achieve simultaneously high output power with low NF and overall flat gain. Seventy-one AM-VSB channels (55.25- to 505.25-MHz) simulated by CW carriers from a multitone generator, were used to modulate the 1550-nm EM-DFB laser with an output power of +16-dBm. The AM-VSB video channels were transmitted through a standard SMF and in-line EDFAs, and then detected at 0-dBm optical power level at the optical receiver. Frequency-domain simulations were used to analyze the two different types of in-line EDFAs. The algorithm for the EDF was based on the quasi-analytical steady-state model, which was developed by Saleh and Jopson, for a two-level EDFA for which ESA can be neglected [34, 35]. This model gives accurate results for low-gain amplifiers (gain $\leq$ 20-dB) pumped at 980-nm or 1480-nm. However, at higher gains, the three-level model EDF, which was discussed in Section 6.1.2, was used. The various EDFA parameters used in the simulations are summarized in Table 1. The Erbium-doping profile of the EDF was identical for the two EDFA types. In addition, the isolator forward pump loss and backward loss were assumed to be infinite for both EDFA types. The optical spectra in the 100-nm band of each in-line EDFA in the AM video lightwave link were measured at various optical input power levels up to +5.9-dBm. For example, the measured and calculated optical spectra of EDFA type A at an optical input power of −13-dBm is shown in Figure 6.12 [30]. Notice that the ASE peak is located around 1530-nm. Excellent agreement is obtained between the measured and calculated optical spectra for both types of EDFAs based on the frequency- domain simulations at various optical input powers. The calculated gain for both EDFAs, which is reduced by a similar amount as the optical input power, is increased. EDFA type B has about 2-dB higher gain than EDFA type A at the same optical input power level. Since the saturated gain of both EDFAs remains high ($G >> 1$), the noise figure can still be approximated by the

**Table 6.1 Simulation parameters for 980 -nm pumped EDFA types "A" and "B".**

| EDFA Parameter | EDFA "A" | EDFA "B" |
|---|---|---|
| Pump Power (mW) | 100 | 90 |
| Pump Wavelength (nm) | 980 | 980 |
| Fiber Length (m) | 30 | 1st Stage = 25,<br>2nd Stage = 30 |
| WDM Pump Loss (dB) | 0.6 | 0.6 |
| WDM Signal Loss (dB) | 1 | 0.5 |
| Isolator Forward Signal Loss (dB) | 0.8 | 0.65 |

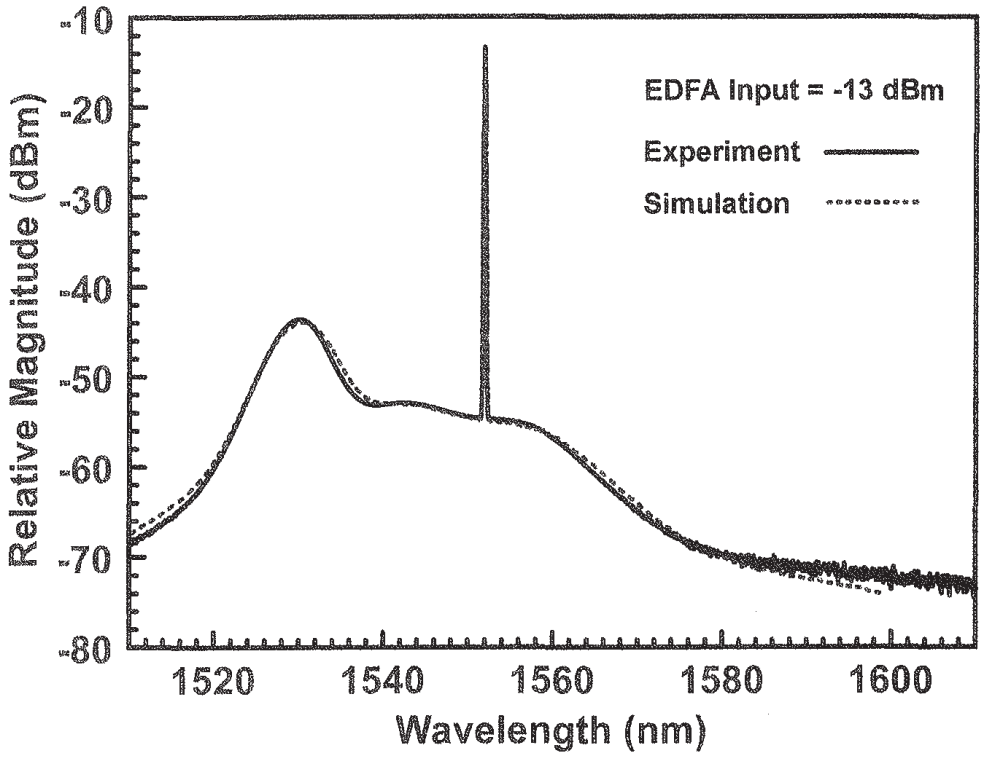



Figure 6.12 Measured and calculated optical spectrum of EDFA type A with an optical input power of –13-dBm. After Ref. [30] (© 1997 IEEE).

$2n_{sp}$ parameter. The noise figure was measured at the optical receiver with and without the in-line EDFA according to Equation (6.29) at various optical input power levels. Figure 6.13 shows the measured and calculated noise figure for both EDFA types as a function of the optical input power levels. The noise figure was calculated from the simulated gain profiles and $n_{sp}$ for each EDFA using Equation (6.34). The noise figure can also be calculated from the measured CNR of the transmitted analog channels using Equation (6.35) with the listed

**Table 6.2 System parameters for CNR calculation.**

| Parameter | Specification |
|---|---|
| Modulation Index per Channel (m) | 3.2% |
| Electrical Noise Bandwidth of AM Channel | 4-MHz |
| Optical Bandwidth of the EDFA | 40 nm |
| Receiver Photocurrent | 0.85-mA |
| Receiver Noise Equivalent Current | 7-pA/√Hz |
| RIN | –164 dBc/Hz |

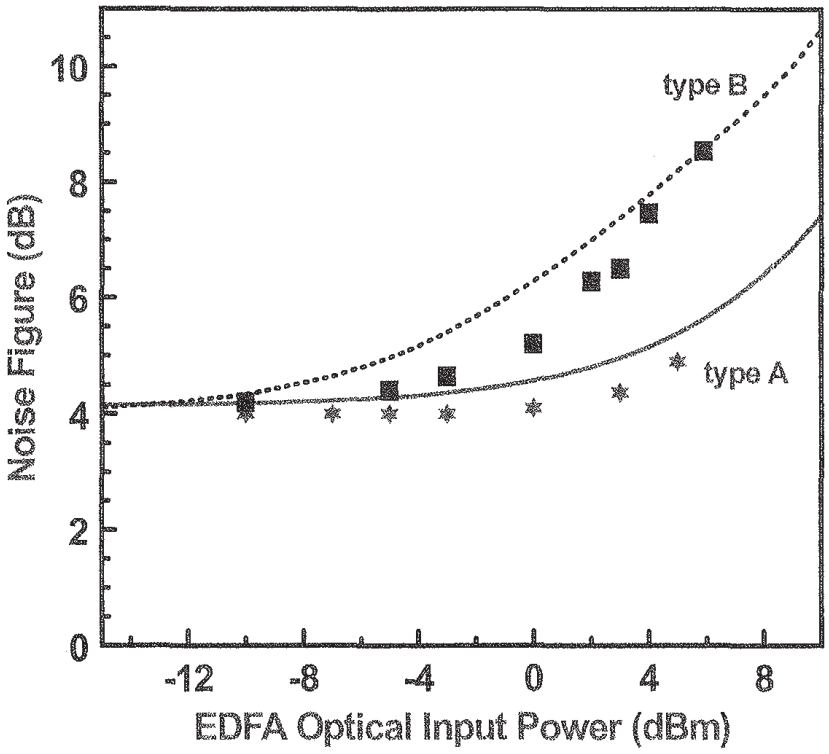



Figure 6.13 Measured (symbols) and calculated (lines) noise figure versus the EDFA optical input power level for EDFA type A and B. After Ref. [30] (© 1997 IEEE).

parameters in Table 6.2. Notice that although the small-signal noise figure of these EDFAs is nearly the same (= 4 dB), the noise figure of EDFA type B dramatically increases to 8.5-dB at high optical input power (= +6 dBm). In contrast, the noise figure of EDFA type A remains essentially unchanged even at these high optical input levels.

This noise figure behavior of these EDFAs can be explained as follows. The lowest EDFA noise figure is achieved when the pump propagates in the same direction as the signal, as is the case for EDFA type A. There is a much stronger inversion depletion of the $Er^{3+}$ ions at high optical input levels in the backward-pumped first stage of EDFA type B compared with EDFA type A. In addition, the midway isolator or filter as shown in Figure 6.11 (c) can be eliminated, since the ASE peak is probably suppressed when the EDFA operates in saturation. Consequently, a single-stage EDFA with unidirectional forward pumping is advantageous for analog video transmission over cable TV networks.

The noise figure of gain-saturated EDFAs has been the investigation subject of many studies [30, 36–38]. The reason is that the theoretical noise figure ($F$) based on Equation (6.34)

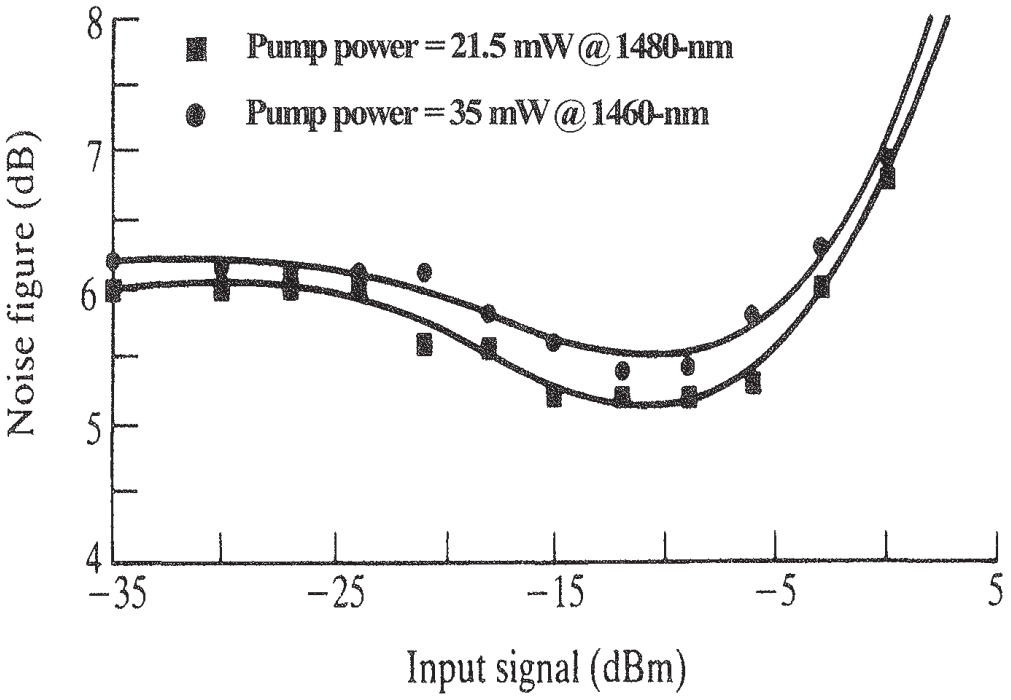



Figure 6.14 Measured and calculated saturated noise figure versus the optical input power for pump powers of 21.5-mW (circles) and 35-mW (square) in a both-end 1480-nm pumped EDFA. After Ref. [37] (© 1991 Optical Society of America).

is predicted to be approximately equal to $2n_{sp}$ if the EDFA saturated gain remains high. This may seem contradictory to various experimental results (i.e., Figure 6.13), which show that the noise figure is actually increasing. To resolve this apparent dilemma, one should remember that the noise figure derivation in the previous sections assumed a small-signal regime in which one has a linear amplification. However, in the saturated-gain regime, the EDFA operates as a nonlinear amplifier, where no rigorous definition has been developed. It has been proposed to define the saturated noise figure (SNF) parameter as $2n_{sp}^{min}(1 + X)$, where $X$ is an excess noise parameter due to saturation [36]. This definition is based on a simplified model with the following three assumptions. First, there is no ASE self-saturation, which means that signal-induced saturation is the main mechanism. Second, the pumping power is much larger than the amplifier saturating power, which is aimed at power amplifiers. Third, the pump power is uniformly distributed along the fiber, which is a reasonable approximation only when the fraction of the pump power absorbed in the EDFA is small. However, in most power amplifiers, this fraction can be large, say 60%, leading to potentially inaccurate noise figure predictions [19].

The noise figure of gain-saturated EDFA can sometimes exhibit a complex behavior in highly performing doped fibers. Figure 6.14 shows, for example, the noise figure behavior as a function of the optical input power to a both-end 1480-nm pumped EDFA [36]. In this example, the small-signal gain is 39 dB for a signal wavelength of 1550-nm. Notice that the saturated noise figure decreases by 1-dB compared with the small-signal value, exhibiting a minimum at an optical input power level near −10-dBm. At higher input power levels, the saturated noise figure increases to 7-dB at an optical input power of 0-dBm. The dip in the noise figure, which occurs at high-gain amplifiers, can be explained by the redistribution of the ASE power along the EDF due to the saturating signal. For sufficiently high optical input power, the backward traveling ASE, which degrades the noise figure, is reduced near the fiber ends. Thus, a higher medium inversion is achieved, and a lower noise figure is obtained.

### 6.4.2 EDFA's CNR Requirement

Based on these results, what is the required noise figure of a high-power in-line amplifier for an AM video lightwave links?

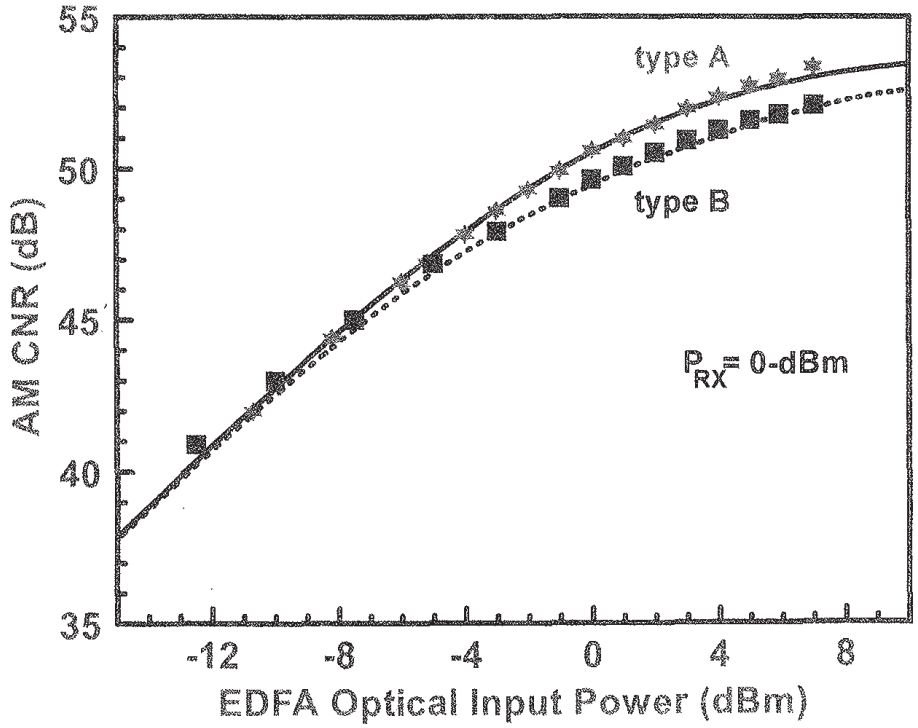



Figure 6.15 Measured (symbols) and calculated (lines) AM CNR versus the EDFA optical input power levels (in dBm) for EDFA types A and B. After Ref. [30] (© 1997 IEEE).

To answer this question, let us examine the following measurement. The AM CNR was first measured for SMF lengths up to 50-km without using the in-line EDFAs. The worst- case measured AM CNR at 325.25-MHz (channel 41) was 53.75-dB with low CSO and CTB distortions (≤ –65 dBc). Then, the measurements were repeated with type A or type B in-line EDFAs in the lightwave link with the same detected optical power of 0-dBm at the receiver. Figure 6.15 shows the worst-case measured and calculated AM CNR at channel 4 as a function of the EDFA's optical input power level (dBm) for the two EDFAs. Both the calculated noise figures and the AM CNR results are in good agreement with the measured results. Notice that the AM CNR difference between the two EDFAs is increased to 1.7-dB at an optical input power level of +6 dBm. Negligible degradation in both the CTB and CSO distortions in the AM video lightwave links was observed using these in-line EDFAs. The experimental results in Figures 6.13 and 6.14 suggest that an in-line amplifier with a noise figure of 5-dB or less is needed to achieve CNR greater or equal to 50-dB for analog video lightwave trunking links [39, 40].

To test this requirement, an AM-VSB video lightwave trunking link with a 39 dB link budget was constructed using two similar in-line EDFAs, where the optical input power for the first in-line EDFA was kept at +3-dBm and with 0-dBm at the receiver. With +3-dBm optical input power to the second in-line EDFA, the measured AM CNR was 50.4-dB using the single-stage EDFA, representing a CNR loss of 3.4-dB. In contrast, the measured AM CNR was only 48-dB using the dual-stage EDFA, or a CNR loss of 5.8-dB under the same operating conditions. Therefore, single-stage EDFAs are advantageous in AM video lightwave links, particularly when cascaded EDFAs are required to achieve the CNR requirement of the designed link. The performance characteristics of cascaded EDFA-based transmission links for both analog and QAM channels are further discussed in Chapter 10.

### 6.4.3 Gain-Flattened EDFAs

So far, we have discussed the EDFA requirements for single-wavelength cable TV networks. As we learned in Chapter 1, HFC cable TV access networks are evolving toward DWDM networks for targeted delivery of analog and digital video/data services. To support the delivery of digital QAM channels over DWDM networks, EDFAs with very broad bandwidths (tens of nanometers) and with flat gain (within 1-dB) over the amplifier bandwidth are required. Gain differences among digital channels transmitted at different wavelengths, particularly over a cascaded EDFA link, may result in a large difference in the received optical power, causing an unacceptable BER. In addition, gain equalization is also desirable for analog video transmission to suppress the CSO distortions induced by gain tilt (see Section 10.5) [41]. Let us review the recent research and development in this area. The main approaches to gain equalization of EDFAs can be divided into passive (or static) methods and active methods. The passive methods include the use of long-period gratings, which have a grating periodicity in the order of hundreds of microns and act as low-loss wavelength selective filters [42]. The problem with the passive methods is that they produce a flat gain only for a predetermined gain level, and produce an undesirable gain tilt when the

gain level is changed. Furthermore, there is no way to adapt the filter transmission profile when the EDFA operating conditions are changed. The active methods for EDFA gain equalization, which are based on active optical filtering, can produce a complex filter transmission response necessary to even out the EDFA gain for different input signals and pump powers. Early work has focused on using integrated-optic $LiNbO_3$ waveguide devices, where the acousto-optic coupling occurs between the TE and the TM polarization modes through a surface acoustic wave.

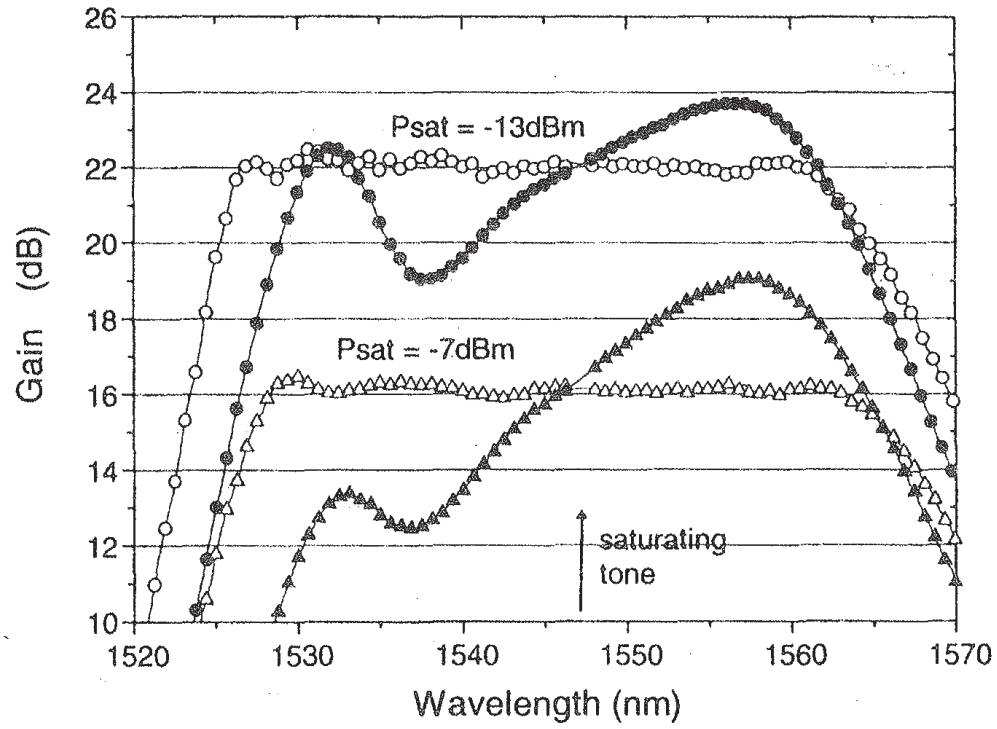



Figure 6.16 EDFA gain profiles with (filled circles) and without (open circles) equalization at –7-dBm and –13-dBm saturation levels. After Ref. [44] (© 1998 IEEE).

To obtain polarization-independent operation, complex polarization diversity configurations were required, resulting in various problems relating to the use of two independent acousto-optic polarization converters [43]. Recent works have demonstrated dynamic EDFA gain equalization using an all-fiber active gain-flattening filter, which consist of two AOTFs in tandem [18, 44]. Each AOTF was driven by three RF signals at different amplitudes and frequencies, which generates sinc-function notch profiles at variable wavelengths and rejection ratio, to produce acousto-optic mode conversion from the fundamental spatial

mode to the cladding modes. Figure 6.16 shows the measured gain profile with (open circles) and without (filled circles) gain equalization for two saturating signal powers of –7-dBm and –13-dBm [44]. Gain variations over 5-dB were reduced to less than 0.7-dB with gain bandwidth over 35-nm.

## References


1. G. P. Agrawal, *Fiber-Optic Communication Systems*, John Wiley & Sons (New York, 1992).
2. F. P. Kapron, "Transmission Properties of Optical Fibers," *Optoelectronic Technology and Lightwave Communication Systems*, Ed. Chinlon Lin, Van Nostrand-Reinhold (New Jersey, 1989).
3. D. Marcuse, *Theory of Dielectric Optical Waveguides*, Academic Press, New York, 2nd Ed. (1991).
4. J. Straus and B. Kawasaki, "Passive Optical Components," *Optical Fiber Transmission*, Ed. E. E. Basch, Howard W. Sams & Co., Indianapolis (1987).
5. B. Nyman, "Passive Components," *OFC Tutorial* (1998).
6. N. Takato, T. Kominato, A. Sugita, K. Jinguji, H. Toba, and M. Kawachi, "Silica-Based Integrated Optic Mach-Zehnder Multi/Demultiplexer Family with Channel Spacing of 0.01-250-nm," *IEEE Journal of Lightwave Technology* **8**, 1120–1127 (1990).
7. C. Dragone, "An N x N Optical Multiplexer Using a Planar Arrangement of Two Star Couplers," *IEEE Photonics Technology Letters* **3**, 812–815 (1991).
8. C. Dragone, C. A. Edwards, and R. C. Kistler, "Integrated Optics N x N Multiplexer on Silicon," *IEEE Photonics Technology Letters* **3**, 896–899 (1991).
9. M. K. Smit and C. Van Dam, "PHASAR-Based WDM Devices: Principles, Design and Applications," *IEEE Journal of Selected Topics in Quantum Electronics* **2**, 236–250 (1996).
10. W. I. Way, *Broadband Hybrid Fiber/Coax Access System Technologies*, Academic Press (San Diego, 1999).
11. M. R. Amersfoort, C. R. de Boer, B. H. Verbeek, P. Demeester, A. Looyen, J. J. G. M. vander Tol, and A. Kuntze, "Low-loss Phase-Array Based 4-Channel Wavelength demultiplexer Integrated with Photodetectors," *IEEE Photonics Technology Letters* **6**, 62–64 (1994).
12. J.-L. Archambault and S. G. Grubb, "Fiber Gratings in Lasers and Amplifiers," *IEEE Journal of Lightwave Technology* **15**, 1378–1390 (1997).
13. K. O. Hill, B. Malo, F. Bilodeau, D. C. Johnson, and J. Albert, "Bragg Gratings Fabricated in Monomode Photosensitive Optical Fiber by UV Exposure Through a Phase Mask," *Applied Physics Letters* **62**, 1035–1037 (1993).

14. W. H. Loh, R. I. Laming, N. Robinson, A. Cavaciuti, F. Vaninett, C. J. Anderson, M. N. Zervas, and M. J. Cole, "Dispersion Compensation over Distances in Excess of 500 km for 10-Gb/s Systems Using Chirped Fiber Gratings," *IEEE Photonics Technology Letters* **8**, 944–946 (1996).
15. D. Pastor, J. Capmany, and J. Marti, "Reduction of Dispersion Induced Composite Triple beat and Second-Order Intermodulation in Subcarrier Multiplexed Systems Using Fiber Grating Equalizers," *IEEE Photonics Technology Letters* **9**, 1280–1282 (1997).
16. C. M. Miller and J. W. Miller, "Wavelength-Locked, Two-Stage Fiber Fabry-Perot Filter for Dense Wavelength Division Demultiplexing in Erbium-Doped Fiber Amplifier Spectrum," *IEE Electronics Letters* **28**, 216–217 (1992).
17. K. Hirabayashi, H. Tuda, and T. Kurokawa, "New Structure of Tunable Wavelength-Selective Filters with a Liquid Crystal for FDM Systems," *IEEE Photonics Technology Letters* **3**, 741–743 (1991).
18. H. S. Kim, S. H. Yun, I. K. Kwang, and B. Y. Kim, "All-Fiber Acousto-Optic Tunable Notch Filter with Electronically Controllable Spectral Profile," *Optics Letters* **22**, 1476–1478 (1997).
19. E. Desurvire, *Erbium-Doped Fiber Amplifiers Principles and Applications*, John Wiley & Sons (New York, 1994).
20. W. J. Miniscalco, "Erbium-Doped Glasses for Fiber Amplifiers at 1500-nm," *IEEE Journal of Lightwave Technology* **9**, 234–250 (1991).
21. A. Mori, Y. Ohishi, M. Yamada, H. Ono, Y. Nishida, K. Oikawa, and S. Sudo, "1.5-μm Broadband Amplification by Tellurite-Based EDFAs," Technical Digest, *OFC Conference*, paper PD-1, Dallas, Texas (1997).
22. J. L. Zyskind, E. Desurvire, J. W. Sulhoff, and D. J. DiGiovanni, "Determination of Homogeneous Linewidth by Spectral Gain Hole-Burning in an Erbium-Doped Fiber Amplifier with $GeO_2$:$SiO_2$ Core," *IEEE Photonics Technology Letters* **2**, 869–871 (1990).
23. C. R. Giles and E. Desurvire, "Propagation of Signal and Noise in Concatenated Erbium-Doped Fiber Optical Amplifiers," *IEEE Journal of Lightwave Technology* **9**, 147–154 (1991).
24. H. Asano, S. Takano, M. Kawaradani, M. Kitamura, and I. Mito, "1.48-μm High-Power InGaAs/InGaAsP MQW LD's for Er-Doped Fiber Amplifiers," *IEEE Photonics Technology Letters* **3**, 415–417 (1991).
25. S. Ishikawa, "Recent Progress in Reliable 980-nm Pump Laser Diodes," *Technical Digest, OFC Conference*, paper ThC3 (1995).
26. X. F. Shum, R. Parke, G. Harnagel, R. Lang, and D. Welch, "1.2 W Single-Mode Fiber-Coupled MOPA at 980 nm," Technical Digest, *OFC Conference*, paper PD13 (1995).
27. S. O'Brien, R. Lang, R. Parke, J. Major, D. F. Welch, and D. Mehuys, "2.2-W Continuous-Wave Diffraction-Limited Monolithically Integrated Master Oscillator Power Amplifier," *IEEE Photonics Technology Letters* **9**, 440–442 (1997).

28. H. M. Presby, "Near 100% Efficient Fiber Microlenses," OSA Proceeding on Conference on Optical Fiber Communications, paper PD24 (1992).
29. H. Kuwahara, M. Sazaki, and N. Tokoyo, "Efficient Coupling from Semiconductor Lasers into Single-Mode-Fibers with Tapered Hemispherical Ends," *Applied Optics* **19**, 2578 (1980).
30. S. Ovadia, "CNR Limitations of Er-Doped Fiber Amplifiers in AM-VSB Video Lightwave Trunking Systems," *IEEE Photonics Technology Letters* **9**, 1152–1154 (1997).
31. J. Aspell, J. Federici, B. M. Nyman, D. L. Wilson, and D. S. Shenk, "Accurate Noise Figure Measurements of Erbium-Doped Fiber Amplifiers in Saturation Conditions," Technical Digest, OFC conference, paper ThA4 (1992).
32. D. W. Hall, R. A. Hass, W. F. Krupke, and M. J. Weber, "Spectral and Polarization Hole Burning in Neodymium Glass Lasers," *IEEE Journal of Quantum Electronics* **19**, 1704 (1990).
33. X. Mao, R. W. Tkach, A. R. Chraplyvy, R. M. Jopson, and R. M. Derosier, "Stimulated Brillouin Threshold Dependence on Fiber Type and Uniformity," *IEEE Photonics Technology Letters* **4**, 66–69 (1992).
34. A. A. M. Saleh, R. M. Jopson, J. D. Evankow, and J. Aspell, "Modeling of Gain in Erbium-Doped Fiber Amplifiers", *IEEE Photonics Technology Letters* **2**, 714–717 (1990).
35. R. M. Jopson and A. A. M. Saleh, "Modeling of Gain and Noise in Erbium-Doped Fiber Amplifiers", *SPIE proceedings of Fiber Laser Sources and Amplifiers III*, vol. **1851**, 114–119 (1991).
36. W. I. Way, A. C. VonLehman, M. J. Andrejco, M. A. Saifi, and C. Lin, "Noise Figure of Gain-Saturated Erbium-Doped Fiber Amplifier Pumped at 980-nm," Proceedings of Topical Meeting on Optical Amplifiers and Applications, 1990, paper TuB3, 134.
37. J. Marcerou, H. Fevrier, J. Hervo, and J. Auge, "Noise Characteristics of the EDFA in Gain Saturation Regimes," OSA Proceeding of Topical Meeting on Optical Amplifiers and Their Applications, paper THE1-1, 162–165 (1991).
38. R. G. Smart, J. L. Zyskind, J. W. Suhoff, and D. J. DiGiovanni, "An Investigation of the Noise Figure and Conversion Efficiency of 0.98 μm Pumped Erbium-Doped Fiber Amplifiers Under Saturated Conditions," *IEEE Photonics Technology Letters* **4**, 1261–1264 (1992).
39. W. I. Way, M. M. Choy, A. Yi-Yan, M. Andrejco, M. Saifi and C. Lin, "Multichannel AM-VSB Television Signal Transmission using an Erbium-Doped Optical Fiber Power Amplifier," *IEEE Photonics Technology Letters* **1**, 343–345 (1989).
40. S. Ovadia and C. Lin, "Performance Characteristics and Applications of Hybrid Multichannel AM-VSB/M-QAM Video Lightwave Transmission Systems," *IEEE Journal of Lightwave Technology* **16**, 1171–1186 (1998).
41. C. Y. Kuo and E. E. Bergmann, "Second-Order Distortion and Electronic Compensation in Analog Links Containing Fiber Amplifiers," *IEEE Journal of Lightwave Technology* **10**, 1751–1759 (1992).

42. A. M. Vensarkar, P. J. Lemaire, J. B. Judkins, V. Bhatia, T. Erdogan, and J. E. Sipe, "Long-Period Fiber Gratings as Band-Rejection Filters," *IEEE Journal of Lightwave Technology* **14**, 58–65 (1996).
43. F. Tian, C. Harizi, H. Herrmann, V. Reimann, R. Ricken, U. Rust, W. Sohler, F. Wehrmann, and S. Westenhofer, "Polarization-Independent Integrated Optical, Acoustically Tunable, Double-Stage Wavelength Filter," *IEEE Journal of Lightwave Technology* **12**, 1192–1197 (1994).
44. H. S. Kim, S. H. Yun, H. K. Kim, N. Park, and B. Y. Kim, "Actively Gain-Flattened Erbium-Doped Fiber Amplifier over 35 nm by Using All-fiber Acousto-optic Tunable Filters," *IEEE Photonics Technology Letters* **10**, 790–792 (1998).

# CHAPTER 7

# RF DIGITAL QAM MODEMS

The purpose of this chapter is to explain the design and operation principles of RF cable modulators and demodulators (modems) for transporting digital video and data signals over the HFC network using the quadrature amplitude modulation (QAM) format based on a well-known standard. The principles of design and operation of QAM modems are essential for the understanding of cable modems and digital set-top boxes, which are discussed in Chapter 8. We shall start with a brief review of all the building blocks of QAM modems. In Section 7.2, the MPEG transport framing is discussed, which enables the transmission of MPEG packets. The next four sections will explain each of the four layers in the forward-error-correction (FEC) coding scheme. The FEC layers include Reed-Solomon coding, Trellis coded modulation, interleaving, and randomization. The building blocks, operation, and design of M-ary QAM modulators are discussed in Sections 7.7 and 7.8. Three key building blocks of the QAM receiver, namely, adaptive equalizer, carrier, and timing recovery are explained in Sections 7.9 and 7.10. Two important parameters to test the fidelity of the QAM transmitter are introduced in Section 7.11. Finally, the BER of M-ary QAM signals in a Gaussian channel are derived in Section 7.12.

## 7.1 RF QAM Modem Building Blocks

Radio frequency (RF) digital transmitters and receivers constitute one of the most important systems in HFC networks to transport digital video and/or data signals from the cable TV headend to the subscriber and back to the headend. In particular, the QAM modulation format was selected as the standard format for cable TV access networks by well-known standard organizations DOCSIS [1], DAVIC [2], and ITU-T J.83 Annex B [3]. Figure 7.1 shows a simplified block diagram of a QAM transmitter and receiver operating over an HFC network. The QAM transmitter portion of the modem consists of a Reed-Solomon (R-S) encoder, a convolutional interleaver, a randomizer or scrambler, a Trellis encoder (TCM), a symbol mapper, an M-ary QAM modulator with differential encoder, and an RF up-converter. The QAM receiver portion of the modem consists of an RF tuner, a QAM demodulator, a symbol demapper, a TCM decoder, a convolutional deinterleaver, an R-S decoder, and a derandomizer (descrambler). All these QAM modem building blocks are explained in the following sections based on the DOCSIS 1.1 protocol [1], which is the primary standard used in the U.S. Figure 7.2 illustrates the FEC section of the QAM modem.

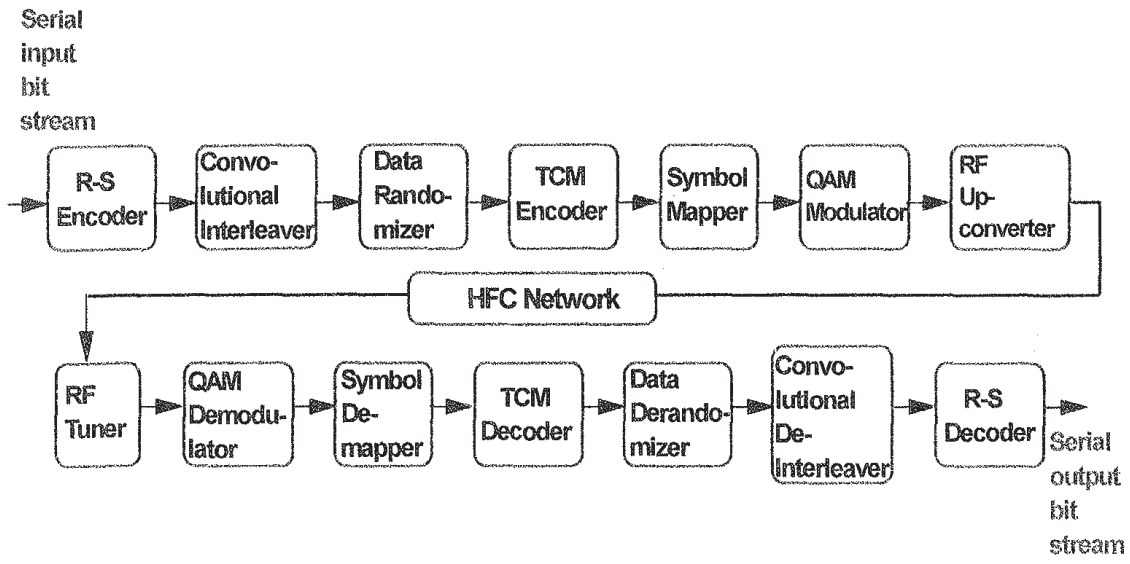



Figure 7.1 A simplified block diagram of RF QAM modem operating over HFC network.

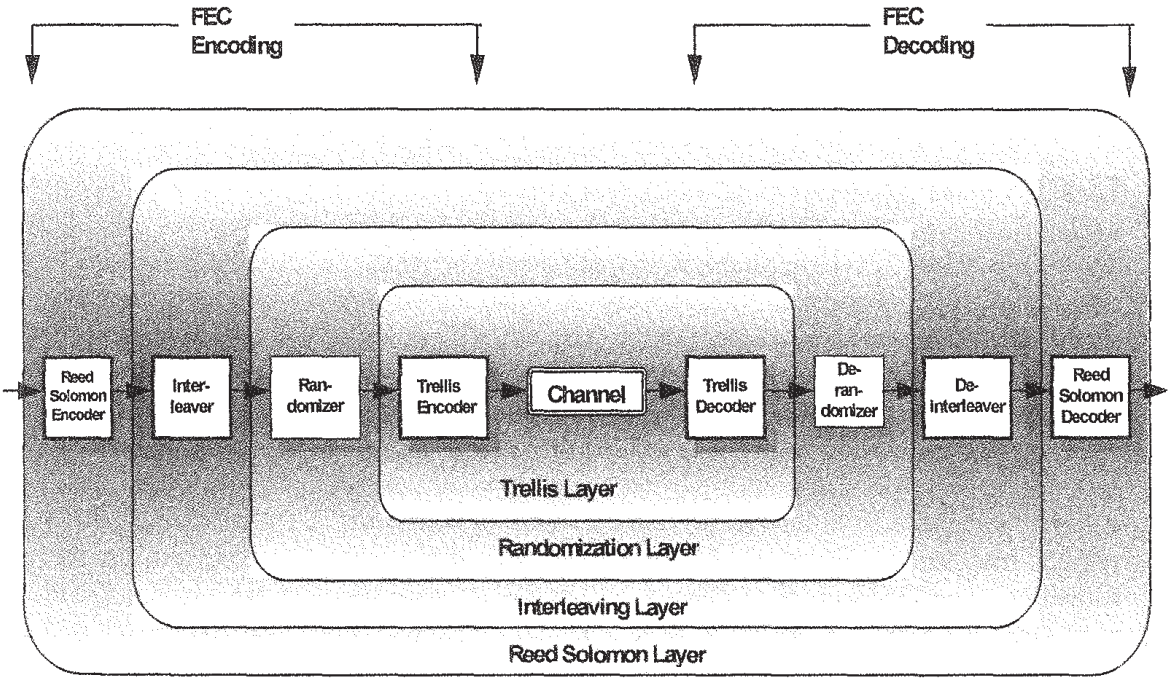



Figure 7.2 Four-layer block diagram of the forward-error-correction code. After Ref. [4].

The FEC section uses various types of error-correcting algorithms and techniques used to transport data reliably over the cable channel. It consists of four processing layers as follow:

- **Reed-Solomon (R-S) Coding** – This layer provides block encoding and decoding to correct up to three symbols within an R-S block.
- **Interleaving** – This layer evenly disperses the symbols in time, preventing a burst of symbol errors from being sent to the R-S decoder.
- **Randomization** – This layer randomizes the data on the channel to allow effective QAM demodulator synchronization.
- **Trellis Coding** – This layer provides convolutional encoding and the possibility of using soft-decision Trellis decoding of random channel errors.

Before the data are FEC encoded in the QAM transmitter and after the data are FEC decoded in the receiver can be applied, MPEG stream transport framing is needed. The following section explains this important processing layer.

## 7.2 MPEG Transport Framing

The MPEG transport framing, which is the outermost layer of processing, enables robust

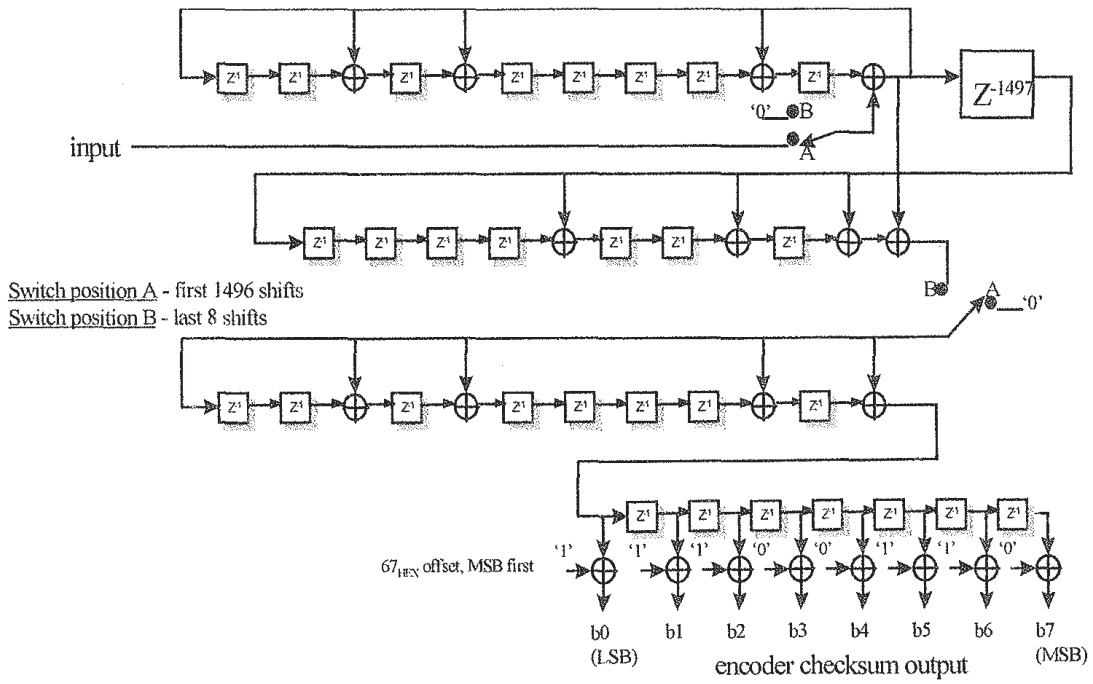



Figure 7.3 Checksum generator for the MPEG-2 sync byte encoder using LFSR. After Ref. [4].

transmission of the MPEG packets to the receiver [4]. This is accomplished by delivering the MPEG packet synchronization byte at the receiver output, and setting up the MPEG packet error flag for uncorrected errors that occurred during the transmission. The MPEG transport-framing block receives an MPEG-2 transport data stream consisting of a continuous stream of fixed length 188 byte packets (n =1504 bits), which is serially transmitted with the most-significant bit (MSB) first. The first byte of the MPEG packet is a sync byte, which has a constant value of $47_{HEX}$, and it is used for both synchronization and additional error correction above that provided by the R-S code in the FEC. The rest of the MPEG payload is data (k =1496 bits).

The problem of additional error detection and synchronization is solved using a parity checksum. Parity checksum is a coset of finite-impulse-response (FIR) parity check linear block code, which is substituted for this sync byte. Figure 7.3 shows the parity checksum generator structure for the MPEG-2 sync byte encoder using linear feedback shift register (LFSR), where all additions are assumed to be modulo 2. The parity checksum is computed by sending the 1496-bit MPEG-2 payload through the LFSR (Figure 7.3), which is described by the following polynomials:

$$f(x) = \left[1 + b(x) \cdot X^{1497}\right] / g(x)$$
$$g(x) = 1 + X + X^5 + X^6 + X^8 \qquad (7.1)$$
$$b(x) = 1 + X + X^3 + X^7$$

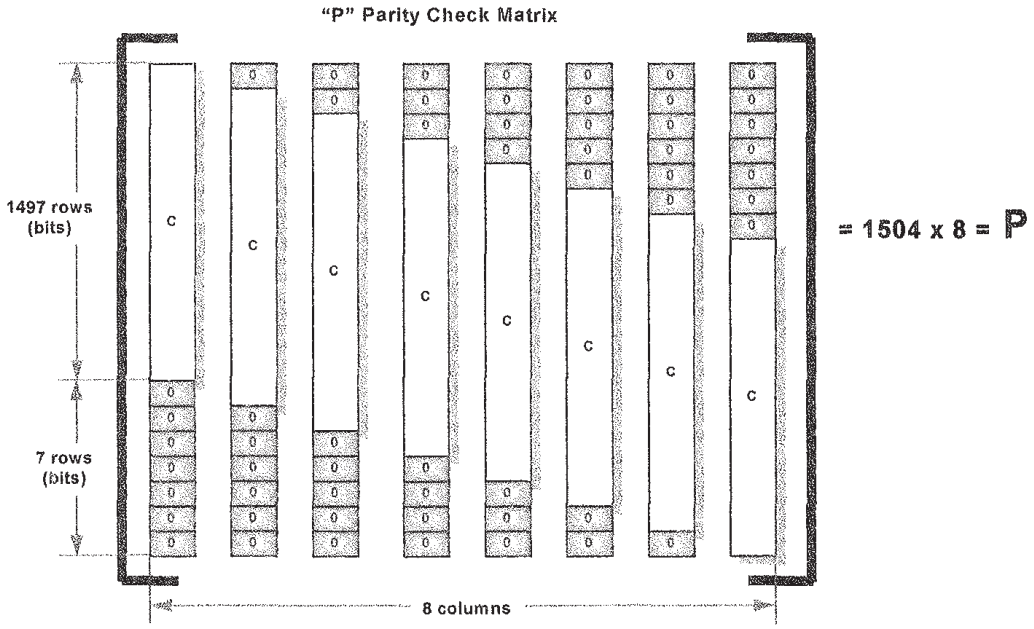



Figure 7.4 Structure of the parity check matrix P. After Ref. [4].

The parity checksum generator works as follows: before the MPEG-2 packet payload is shifted into the LFSR, the LFSR is initialized to zero. The input MPEG-2 packets are first divided by $g(x)$, and the output is split. Part of the output is delayed by 1497 bits and multiplied by $b(x)$, and then added to the non-delayed part to form $f(x)$. After the 1496 data bits are received, the encoder input is set to zero, and the switch position is changed from A to B. The eight parity bits are sequentially output by eight additional shifts of the LFSR content. 0x67 offset is added to the parity checksum byte at the MPEG-2 sync byte encoder to improve the auto-correlation properties, and to produce the 0x47 value during the parity checksum decoding at the receiver when a valid codeword is presented. A valid checksum is identified at the MPEG-2 decoder using a parity-check matrix, P, which is shown in Figure 7.4. The parity-check matrix consists of eight columns of 1497-bit vector, which we will call vector C. Vector C is calculated according to the $f(x)$ polynomial given by Equation 7.1. It is downshifted by one bit position as one proceeds from the leftmost column of the matrix to the right. The unoccupied bit positions in each data column of the parity-check matrix are filled with zeros. The syndrome generator at the decoder can be implemented using either FIR or infinite-impulse-response (IIR) filters. For example, Figure 7.5 shows an IIR filter implementation of the syndrome generator for the MPEG-2 sync byte at the decoder [5]. Note that the checksum at the MPEG-2 sync decoder is calculated always on the previous 187 bytes. Packet synchronization at the MPEG decoder is achieved as follows: the received MPEG-2 vector, which is called vector R, includes 187 bytes of MPEG-2 data followed by the checksum byte, yielding a total of 1504 bits. Multiplying (modulo 2) vector R by the parity-check matrix P yields an eight-bit-length vector S. A valid checksum is indicated when $S = [0100, 0111] = 47_{HEX}$.

The MPEG-2 decoder computes a sliding checksum on the serial data stream, using the detection of a valid codeword to detect the start of a packet. Once packet synchronization

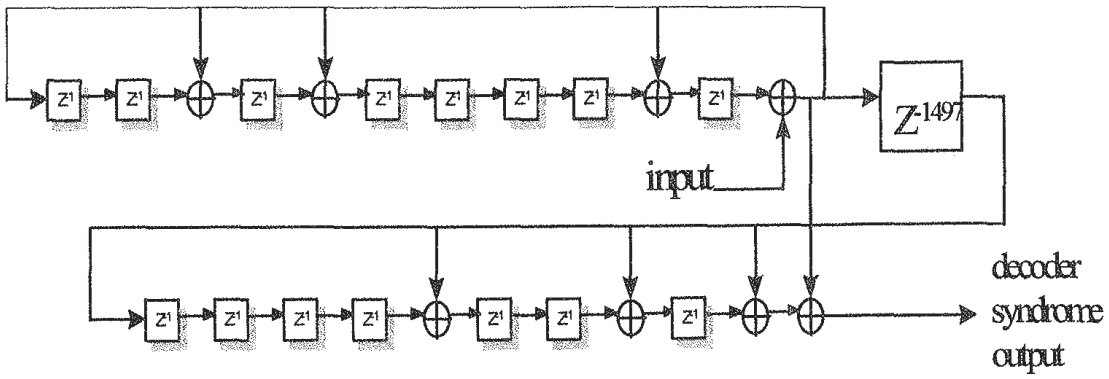



Figure 7.5 Syndrome generator for the MPEG-2 sync decoder. After Ref. [4].

has been achieved, the absence of a valid codeword at the beginning of the MPEG packet indicates a packet error. The error flag of the previous packet can then be set. To provide a standard MPEG-2 data stream, the normal sync word is reinserted in place of the parity checksum as the data is output from the decoder.

It should be pointed out that if other than MPEG-2 transport protocols such as ATM are used, this outer processing layer is removed or bypassed. The FEC layer is essentially transparent to all other transport protocols. The ATM header-error-control byte (HEC) typically provides the proper packet framing and error detection. Isochronous ATM streams are therefore carried transparently without overhead for MPEG or quasi-MPEG packet encapsulation.

## 7.3 Reed-Solomon Codes

The *nonbinary* Reed-Solomon (R-S) codes are a special subclass of the Bose-Chadhuri-Hocquenghem (BCH) codes, which are a large class of powerful random-error correcting cyclic codes [5]. A very useful feature of the R-S codes is that they achieve the largest possible code minimum distance for any linear code with the same encoder input and output lengths. For nonbinary codes, the distance between two code words is defined as the number of nonbinary symbols in which the sequences differ. For R-S codes, the minimum distance is given by $d_{min} = n-k+1$, where $n$ is the total number of code symbols and $k$ is the number of encoded data symbols in the encoded data block. Furthermore, the R-S code is capable of correcting up to $T$ symbol errors, which is given by $T = (d_{min}-1)/2 = (n-k)/2$, corresponding to no more than $2T$ parity-check symbols. R-S codes are particularly effective for burst error correction; they are effective for channels with memory. Relatively long R-S codes can be efficiently implemented in a digital receiver with typical hard-decision decoding algorithms. For a $T$ error-correcting R-S code with an alphabet of $2^m$ symbols, $n = 2^m-1$, and $k = 2^m-1-2T$, where m = 2,3,..., the R-S decoded-symbol-error probability can be approximated by

$$P_E \cong \frac{1}{n}\sum_{j=T+1}^{n} j\binom{n}{j}p^j(1-p)^{n-j} \tag{7.2}$$

where p is the symbol-error probability with no R-S coding.

For MPEG-2 transmission, the transport stream is R-S encoded using a (128,122) code over the so-called *Galois* field (GF) (128). This means that each R-S block consists of 122 information symbols and 6 parity symbols and thus has the capability of correcting up to T = 3 errors per R-S block. The same R-S code is used for both 64-QAM and 256-QAM formats, but with different Trellis coding. In addition, the FEC frame format is different for each modulation format. For the DAVIC standard, R-S T=8 (204,188) coding over GF(256) is used for both 64-QAM and 256-QAM modulation formats.

Figure 7.6 shows the packet frame format for 64-QAM modulation. For 64-QAM, an FEC frame consists of six symbols sync trailer, which is appended to the end of 60 R-S blocks, with each block containing 128 symbols. Each R-S and sync symbol consists of 7 bits.

Thus, each FEC frame consists of a total of 53,760 data bits and 42 frame sync trailer bits. For 256-QAM, an FEC frame consists of a 40-bit sync trailer, which is appended to the end of an 88 R-S block FEC frame. Each symbol in the R-S blocks is a 7-bit symbol. The total number of data bits in a 256-QAM FEC frame is 78,848 bits.

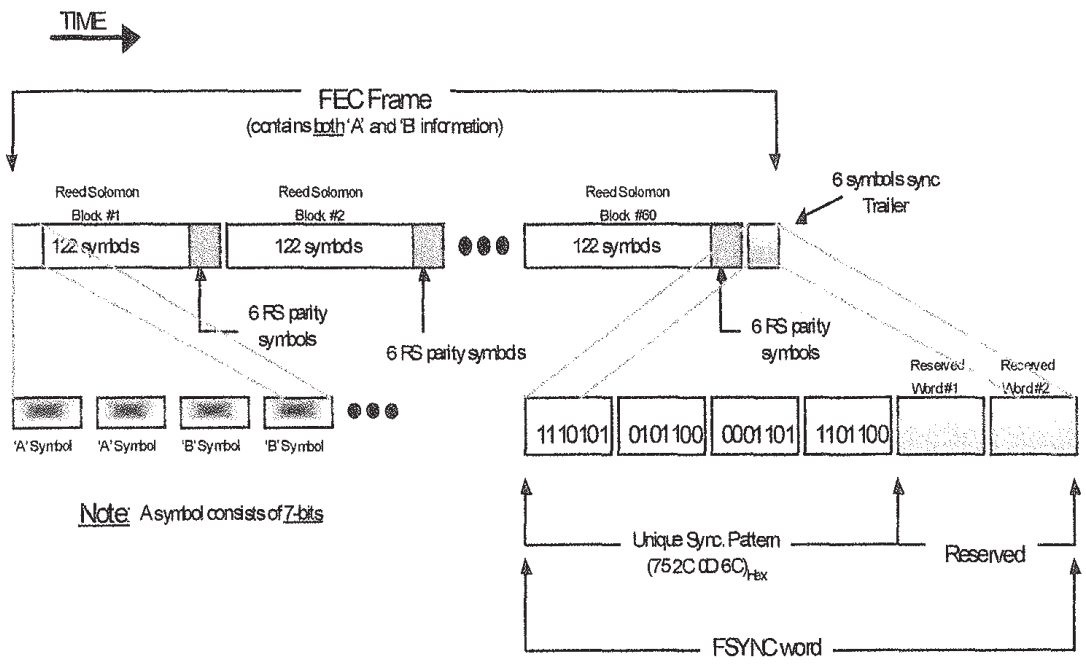



Figure 7.6 FEC frame packet format for a 64-QAM modulation. After Ref. [4].

For the ITU-T J.83 Annex B standard, the R-S encoder is implemented as follows: a systematic encoder is utilized to implement R-S T = 3 (128,122) code over the GF (128). The primitive polynomial used to form the field over GF(128) is:

$$p(X) = X^7 + X^3 + 1 \tag{7.3}$$

The generator polynomial used by the encoder is:

$$g(X) = (X+\alpha)(X+\alpha^2)(X+\alpha^3)(X+\alpha^4)(X+\alpha^5) \tag{7.4}$$
$$= X^5 + \alpha^{52}X^4 + \alpha^{116}X^3 + \alpha^{119}X^2 + \alpha^{61}X + \alpha^{15}$$

where $\alpha$ is the root of the primitive polynomial, namely $p(\alpha) = 0$. Notice that all the coefficients of g(X) are from GF(128). The message polynomial input to the encoder consists of 122, 7-bit symbols, and is described

$$m(X) = m_{121}X^{121} + m_{120}X^{120} + \ldots + m_1X + m_0 \tag{7.5}$$

This message polynomial is first multiplied by $X^5$; then, it is divided by the generator polynomial $g(X)$ to form a remainder, described by the following equation:

$$r(X) = \frac{X^5 m(X)}{g(X)} = r_4X^4 + r_3X^3 + r_2X^2 + r_1X + r_0 \tag{7.6}$$

This remainder constitutes five parity symbols, which are then added to the message polynomial to form a 127-symbol code word that is an even multiple of the generator polynomial. Then, the generated code word can be described by the following polynomial:

$$c(X) = m_{121}X^{126} + m_{120}X^{125} + m_{119}X^{124} + \ldots + r_4X^4 + r_3X^3 + r_2X^2 + r_1X + r_0 \tag{7.7}$$

A valid code word will have roots at the first through fifth powers of $\alpha$.

The last symbol of the transmitted R-S block is formed by extending the parity symbol (c*) as follows:

$$c^* = c(\alpha^6) \tag{7.8}$$

In other words, the code word is evaluated at the sixth power of $\alpha$. Thus, the extended code word would appear as follows:

$$c' = X \cdot c(X) + c^* = m_{121}X^{127} + m_{120}X^{126} + \ldots + r_1X^2 + r_0X + c^* \tag{7.9}$$

The structure of the R-S block, which illustrates the order of transmitted output symbols from the R-S encoder, is shown as follows:

$m_{121}$ $m_{120}$ $m_{119}$ ... $r_1$ $r_0$ c* (order sent is left to right)

For 64-QAM, the first four symbols of the FSYNC word contain the 28-bit "unique" synchronization pattern (1110101 0101100 0001101 1101100) or $(75\ 2C\ 0D\ 6C)_{HEX}$. The remaining two symbols are reserved for future applications. The FSYNC word will be inserted by the encoder and will be detected at the decoder. The decoder circuits search for this pattern and determine the location of the frame end when it is found.

For 256-QAM, the 40-bit frame sync trailer is divided as follows: 32 bits are the "unique" synchronization pattern (0111 0001 1110 1000 0100 1101 1101 0100) or (71 E8 4D D4 )$_{HEX}$, 3 bits are a control word that determines the size of the interleaver employed, and 5 bits are a reserved word that is set to zero.

Note that there is no synchronization relationship between the transmitted R-S block and transport data packets. Thus, MPEG-2 synchronization must be obtained independently from R-S frame synchronization. This keeps the FEC and transport layers decoupled and independent. The downstream 64-QAM and 256-QAM parameters for ITU-T J.83 Annex B and A are summarized in Tables 7.3 and 7.4 (Section 7.12).

## 7.4 Interleaver/Deinterleaver

The concept of a convolutional interleaver has been proposed by Ramsey [6] and Forney [7]. Interleaving the input symbols before transmission and deinterleaving them after reception allows the QAM receiver to spread the burst errors in time, allowing the receiver to handle them as if they were random errors. Thus, the QAM receiver transforms a channel with memory to a memoryless channel, enabling random-error-correcting codes such as an R-S code to be used effectively. Convolutional interleaving is employed in both 64-QAM and 256-QAM modulation formats. A convolutional interleaver is typically characterized by the number of its shift registers, which is also called the "depth" $I$ (symbols), and by the symbol delay increment $J$ per register. Due to memory cost and end-to-end delay for the transmitted symbols, it is advantageous to limit the interleaver $(I, J)$ values in certain applications.

The concept of convolutional interleaving is illustrated in Figure 7.7. The interleaving commutator position increments at the R-S symbol frequency, with a single symbol output from each position. With a convolutional interleaver, the R-S code symbols are sequentially shifted into the bank of 128 registers (the width of each register is 7 bits, which matches the R-S symbol size). Each successive register has J symbols more storage than the preceding register. The first interleaver path has a zero time delay, while the second path has $J$ symbol

**Table 7.1 Convolutional interleaver characteristics for 64-QAM and 256-QAM.**

| I | J | Burst Protection 64/256-QAM | Latency 64/256-QAM |
|---|---|---|---|
| 8 | 16 | 5.9 μs/4.1 μs | 0.22 ms/0.15 ms |
| 16 | 8 | 12 μs/8.2 μs | 0.48 ms/0.33 ms |
| 32 | 4 | 24 μs/16 μs | 0.98 ms/0.68 ms |
| 64 | 2 | 47 μs/33 μs | 2.0 ms/1.4 ms |
| 128 | 1 | 95 μs/66 μs | 4.0 ms/2.8 ms |
| 128 | 2 | 189 μs/132 μs | 8.0 ms/5.6 ms |
| 128 | 4 | 377 μs | 16.0 ms/11.2 ms |

period of delay, and the third path $2 \cdot J$ symbol periods of delay, and so on, up to the 128th path. Thus, the last interleaver register has $127 \cdot J$ symbol periods of delay. At the receiver, the time-delay process through each of the deinterleaver paths is reversed such that the total net delay for each R-S symbol is the same through the interleaver and deinterleaver.

Burst noise in the channel can cause a series of erroneous symbols or multiple errors in each symbol. These errors are spread over many R-S blocks by the deinterleaver such that the resultant symbol errors per block are within the range of the R-S decoder correction capability. For 64-QAM, in the absence of all other impairments, ignoring any error propagation from the Trellis decoder, the burst tolerance as measured at the input of the deinterleaver is 95-μsec. To accomplish this, $J$ is set equal to 1 in the interleaver.

For 256-QAM, the interleaver continues to employ a depth of $I = 128$, but here $J$ may be selected by a 3-bit field in the frame sync trailer to be an integer value between 1 and 8. Table 7.1 shows the burst noise protection and latency convolutional interleaver with different $I$ and $J$ values. The burst noise protection, which is sometimes called erasure correction (EC) of a convolutional interleaver, is given by [2]:

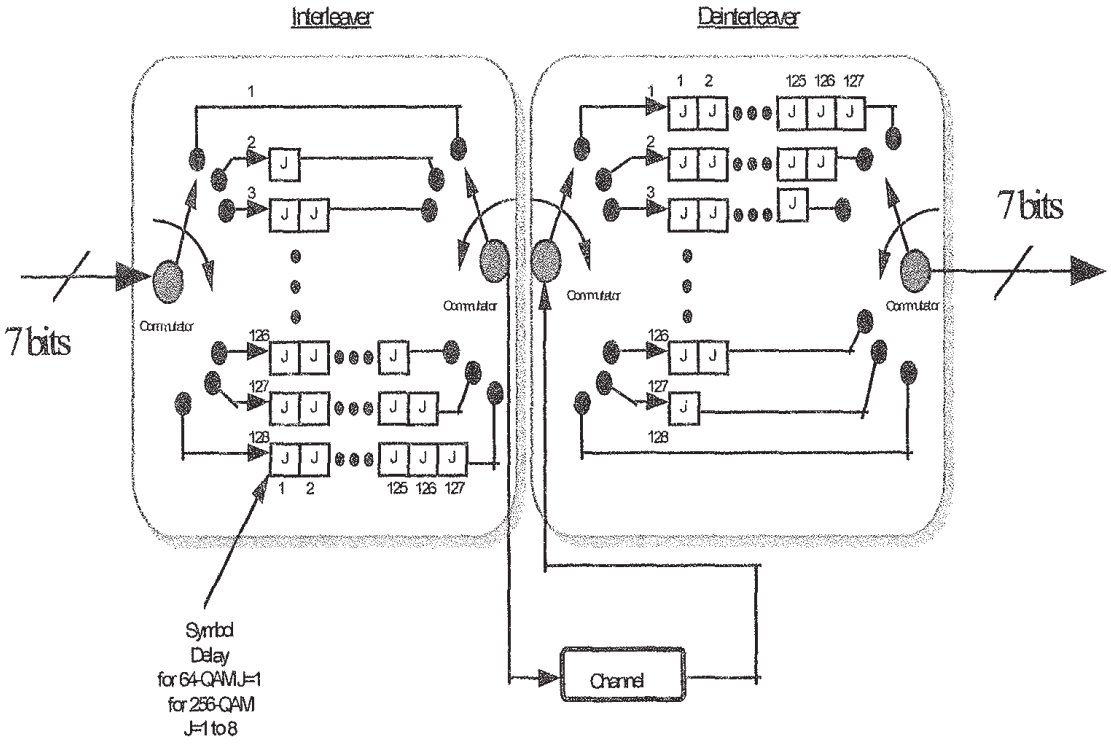



Figure 7.7 Interleaver and deinterleaver functional block diagram. After Ref. [4].

$$EC(\mu s) = \frac{(I \cdot J + 1)}{R_{RS}} \left\lceil \frac{T \cdot I}{N} \right\rceil \tag{7.10}$$

where $N$ is the R-S block size, and $R_{RS}$ is the R-S symbol rate. The end-to-end delay or latency is given by:

$$Latency(\mu s) = \frac{I \cdot J \cdot (I-1)}{R_{RS}} \tag{7.11}$$

From Equations (7.10) and (7.11), larger interleaver size (I·J) requires larger memory to perform the convolutional interleaving operation, causing a larger the end-to-end delay. The ITU-T J.83 Annex B standard for downstream QAM channels provides a burst tolerance up to 95-μs and 66-μs for 64-QAM and 256-QAM, respectively, for an $I$ = 128, $J$ = 1 interleaver setting. The performance of the interleaver in the QAM receiver to combat burst noise will be discussed in Chapter 8.

## 7.5 Trellis-Coded Modulation (TCM)

The discussion on trellis-coded modulation (TCM) is divided into three parts. First, we introduce the concept of convolutional coding with punctured pattern. Second, the Viterbi decoding process in the QAM receiver is reviewed. Then, the use of TCM in both 64-QAM and 256-QAM modulation formats for downstream transmission is discussed.

### 7.5.1 Punctured Convolutional Coding

Trellis coding is employed as the inner coding scheme, while R-S coding is employed as the outer coding scheme as shown in the QAM modem block diagram in Figure 7.1. The ITU-T J.83 Annex B standard for downstream QAM transmission, which is used in the U.S, employs a concatenated coding scheme. The combination of modulation with Trellis coding is termed *trellis-coded modulation* (TCM), which was first published and analyzed by Ungerboeck [8, 9]. It allows the introduction of redundancy to improve the SNR by increasing the symbol constellation without increasing the symbol rate or bandwidth.

The TCM coding provides significantly larger "free distance" or minimum Euclidean distance ($d_{free}$) than the minimum distance between uncoded modulation signals at the same information rate, bandwidth, and power. The convolutional encoder consists of a shift register with $k$·$K$ stages, where $K$ is called the *constraint length* of the code. The constraint length represents the number of $k$-bit shifts in which the encoder output is influenced by a single bit. At any given time, $k$ bits enter the shift register, and the contents of the last $k$ stages of the shift register are dropped. $n$ linear combinations of the shift register contents are computed to generate an encoded output waveform. The code rate of the convolutional encoder is $k/n < 1$, where $n-k$ is the number of redundant bits. Figure 7.8 shows, for exam-

ple, a 1/2-rate convolutional encoder with constraint length of 7, followed by a puncturing unit. The convolutional encoder connections between the shift register and each of the n-modulo-2 adders may be characterized by generating polynomials, as we learned in Section 7.2 on R-S codes. Each polynomial has a degree of $K-1$ or less, and its coefficients are either 1 or 0, depending on whether a connection exists or does not exist between the shift register and the particular modulo-2 adder. The generating polynomials are sometimes written in octal notation. For example, the generating polynomials for the convolutional encoder shown in Figure 10.5 are given by

$$g_1(X) = 1 + X^2 + X^3 + X^5 + X^6$$
$$g_2(X) = 1 + X + X^2 + X^3 + X^6 \qquad (7.12)$$

or 133 and 171 octal.

Punctured coding is an attractive method for implementing rate-selectable convolutional encoders using existing hardware. To create a punctured code, one first encodes the data using a $1/n$ convolutional encoder and then deletes some of the channel symbols at the encoder output, which is called "puncturing." The puncturing pattern is represented by "0" or "1" symbols, where a "1" symbol indicates that a channel symbol is to be transmitted and "0" indicates a channel symbol to be deleted. Figure 7.8 shows, for example, how to create

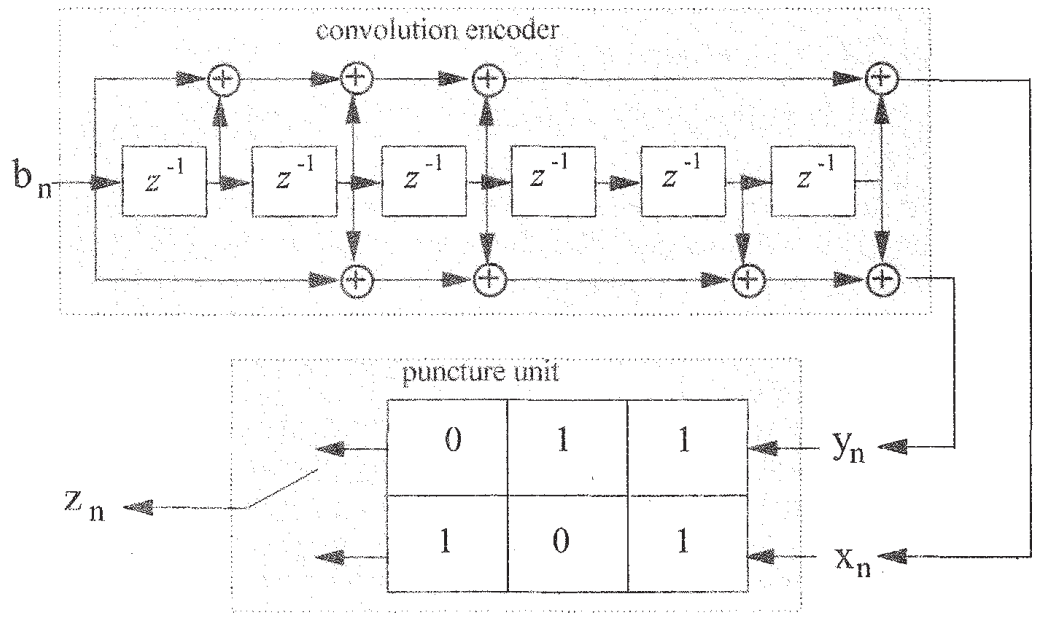



Figure 7.8 3/4-rate punctured convolutional encoder with constraint length K = 7 using a 1/2-rate convolutional encoder.

3/4-rate code from the 1/2-rate code using the puncturing pattern. Each of the columns of the puncturing pattern corresponds to a bit input to the encoder, while each bit in the pattern corresponds to an output-channel bit. The puncturing pattern can also be written in a matrix format as [P1, P2] = [011, 101]. Thus, the puncturing unit outputs 4 bits for every 3 bits input, resulting in an overall 3/4-rate convolutional encoder. Using the basic 1/2 code, this method allows the implementation of any rate $(n-1)/n$ that is slightly less efficient than the optimum code for that rate.

## 7.5.2 Viterbi Decoding

At the QAM receiver, the received bit stream needs to be decoded to recover the original data with an error correction code. The state of a $1/n$ convolutional encoder is defined as the content of the rightmost $K-1$ shift-registers. The output of the encoder is determined by the input bits and the state of the encoder. There are three common methods to characterize convolutional encoders. The first method is called a state diagram, which is a graphical representation of all the possible states and the possible transitions from one state to another. The problem with the state diagram is that it does not show the time evolution of the encoder transitions. The second method is called a tree diagram, which shows the time evolution of each successive input bit, where each tree branch describes an output branch word. The

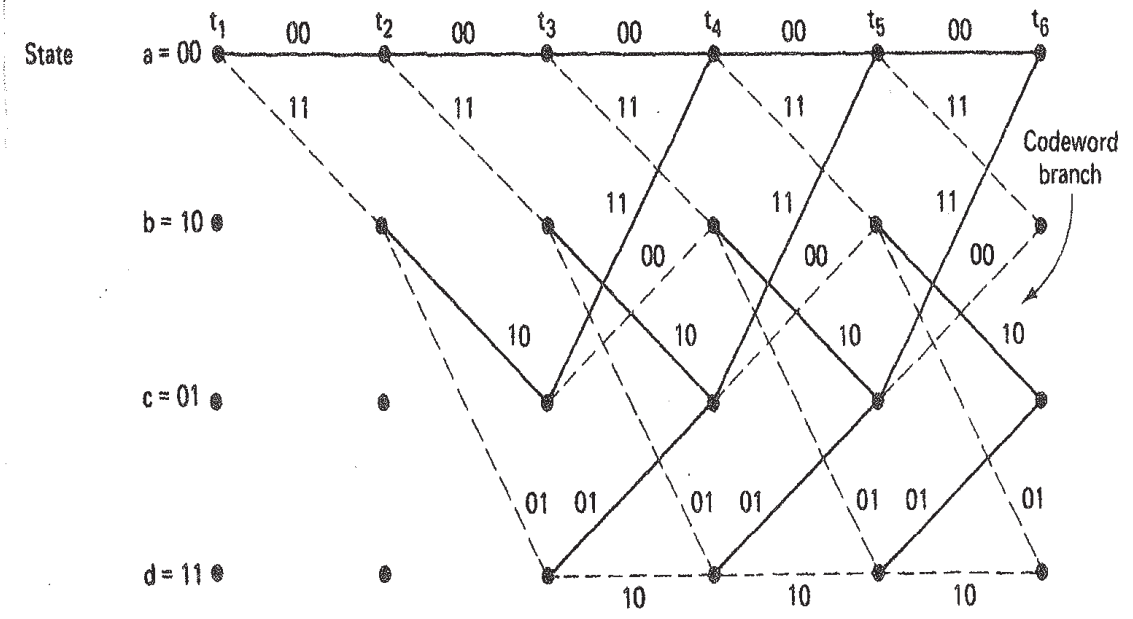



Figure 7.9 Trellis diagram for 1/2-rate convolutional code with K = 3, where the solid and dashed lines indicate the output generated by input bits "0" and "1", respectively. After Ref. [10].

drawback of the tree-diagram method is that the number of tree branches increases as $2^M$, where $M$ is the number of bits in the input sequence. The third method, which is also the most useful method, is called a trellis diagram, where the nodes of the trellis characterize the encoder states, and solid and dashed lines represent outputs generated by bits "0" and "1", respectively. At any given time, the trellis diagram requires $2^{M-1}$ nodes to represent the $2^{M-1}$ possible encoder states. Figure 7.9 illustrates the concept of a time-expended trellis diagram of 1/2-rate convolutional code with $K = 3$ ($d_{free} = 5$). The generating polynomials for this encoder are $g_1(X) = 1 + X + X^2$ and $g_2(X) = 1 + X^2$. Notice that the trellis diagram at each stage after the initial transition consists of four nodes, corresponding to four different states labeled *a*, *b*, *c*, and *d*. At each node in the trellis there are two incoming paths and two outgoing paths after the second stage. The optimum decoding method is maximum-likelihood decoding, where the decoder attempts to find the closest "valid" sequence to the received bit stream. The most important algorithm for maximum-likelihood decoding is the Viterbi algorithm [11,12]. The Viterbi decoder recognizes at any given time that certain trellis paths cannot possibly belong to the maximum-likelihood path, and thus it discards them. When binary convolutional code (BCC) with $k = 1$ and constraint length $K$ is decoded using the Viterbi algorithm, there are $2^{K-1}$ surviving paths at each stage, and $2^{K-1}$ add compare operations. For BCC in which $k$ bits each time are shifted into the encoder, there are $2^{k(K-1)}$ surviving paths at each stage, and $2^k$ paths merge at each node. Since there is only one surviving path at each node, namely, the minimum-distance path, the required number of computations at each stage increases exponentially with both $k$ and $K$. Thus, in practice, the Viterbi algorithm decoder is typically implemented only for small $K$ (< 10) and $k$.

The Viterbi decoder can be implemented using the so-called *hard-decision* or *soft-decision* decoding. In the hard-decision mode, the output of the QAM demodulator is quantized to either "1" or "0", and fed to the Viterbi decoder. If the demodulator output is quantized to more than two levels based on a Gaussian distribution, the decoding is called a soft-decision decoding. For example, eight levels (3-bits) of soft-decision Viterbi decoding reduces the required $E_b/N_o$ by about 2-dB in a Gaussian channel compared to a hard-decision (1-bit) Viterbi decoding at the same BER [13]. The cost of this improvement is the larger required memory size and also the decoding speed at the receiver.

An essential part of the TCM scheme, which was developed by Ungerboeck, is the set-partitioning concept [9]. Set partitioning divides a signal set successively into smaller subsets while increasingly maximizing the smallest free distance within each subset. The set partitioning can be carried out to the limit where each subset contains only a single point. However, in many practical cases, such set partitioning is not necessary.

The performance of TCM coding can be expressed in terms of coding gain relative to the uncoded modulation set. Thus, the coding gain ($G$) for a given TCM code is given by

$$G = \frac{\left[d_{\min}^2 / P_{av}\right]_{coded}}{\left[d_{\min}^2 / P_{av}\right]_{uncoded}} \tag{7.13}$$

where $d_{min}$ is the minimum Euclidean distance, and $P_{av}$ is the average power of the modulation. Table 7.2 shows, for example, the coding gains for different trellis-coded M-ary. QAM modulation formats [9]. $N_{fed}$ represents the number of signal sequences with the minimum distance that diverge at any state and after one or more transitions remerge at that state. The asymptotic coding gain represents the maximum possible coding gain for a given M-ary QAM modulation over the uncoded modulation. Notice that for a trellis-coded 64-QAM modulation, most of the coding gain (4.56-dB) can be achieved using only 16 states as indicated in Table 7.2. Viterbi decoding is also more tolerant than sequential decoding to metric table and receiver AGC errors. The inherent parallelism of the Viterbi decoder makes it easy to implement in hardware, an important consideration for high-speed decoding.

An important design parameter for a Viterbi decoder is the path memory length. An ideal Viterbi decoder would keep every possible path in memory and delay a final decision about the first bits of the packet until the very end. However, various simulations have shown that several constraint lengths back, the paths usually merge into a single maximum-likelihood candidate. Thus, a hardware implementation of a Viterbi decoder (unpunctured), which typically retains 4–5 constraint lengths of decoded data, is sufficient to achieve nearly the same uncorrected error rate as an ideal decoder. For a $K = 7$ code, a 32-bit path memory is a good match to a 32-bit word size for a digital signal processor (DSP) implementation. This is fast and efficient since 32-bit operations on a 32-bit DSP are no more costly than smaller operations.

**Table 7.2 Coding gains for different trellis-coded M-ary QAM modulations. After Ref. [9] (© 1987 IEEE).**

| Number of states | Code rate | G (16-QAM/ 8-QAM) (dB) | G (32-QAM/ 16-QAM) (dB) | G (64-QAM/ 32-QAM) (dB) | Asymptotic coding gain (dB) | $N_{fed}$ |
|---|---|---|---|---|---|---|
| 4 | 1/2 | 3.01 | 3.01 | 2.80 | 3.01 | 4 |
| 8 | 2/3 | 3.98 | 3.98 | 3.77 | 3.98 | 16 |
| 16 | 2/3 | 4.77 | 4.77 | 4.56 | 4.77 | 56 |
| 32 | 2/3 | 4.77 | 4.77 | 4.56 | 4.77 | 16 |
| 64 | 2/3 | 5.44 | 5.44 | 4.23 | 5.44 | 56 |
| 128 | 2/3 | 6.02 | 6.02 | 5.81 | 6.02 | 344 |
| 256 | 2/3 | 6.02 | 6.02 | 5.81 | 6.02 | 44 |

### 7.5.3 TCM for 64/256-QAM Modulation

For a 64-QAM modulation, the input to the trellis-coded modulator is a 28-bit sequence of four 7-bit R-S symbols, which are labeled in pairs of "A" symbols and "B" symbols. A block diagram of a trellis-coded modulator for 64-QAM modulation is shown in Figure 7.10. The trellis-coded modulator includes a punctured 1/2-rate BCC that is used to introduce the redundancy into the least-significant bits (LSBs) of the trellis group. The convolutional encoder is a 16-state nonsystematic 1/2-rate encoder with the generating polynomials $g_1$ = 010 101 and $g_2$ =011 111, or (25,37) in octal notation. Equivalently, one convolutional encoder can be described in terms of the generator matrix [1⊕$D^2$⊕$D^4$, 1⊕D⊕$D^2$⊕$D^3$⊕$D^4$]. The puncturing matrix for 4/5-rate is [P1, P2] = [0001; 1111], which produces a single serial bit stream. All 28 bits are assigned to a trellis group, where each trellis group forms 5 QAM symbols, as shown in Figure 7.10. Each trellis group is further divided into two groups: two upper or most-significant bits (MSBs) uncoded bit streams, and one lower or LSB-coded bit stream. Of the 28 bits in the trellis group, the four LSBs of the trellis group are fed to the BCC for coding, producing five coded bits. The remaining 10 bits are sent to the mapper

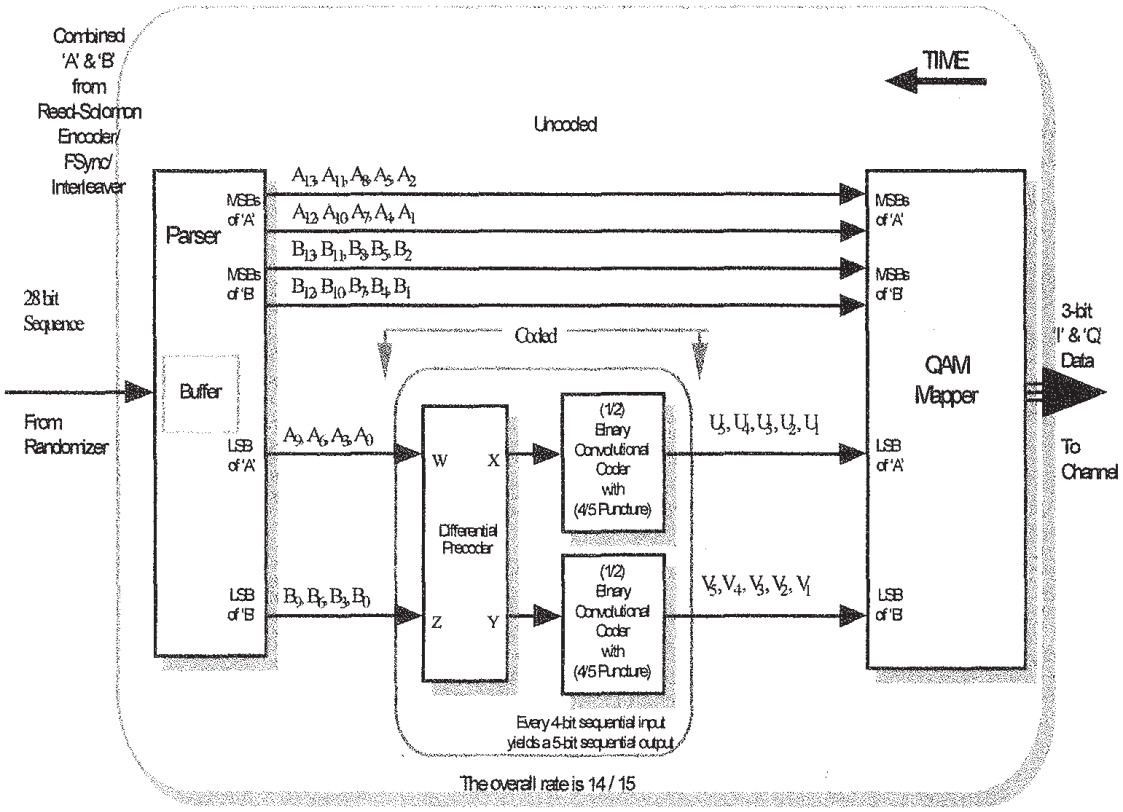



Figure 7.10 Block diagram of a 64-QAM trellis-coded modulation. After Ref. [4].

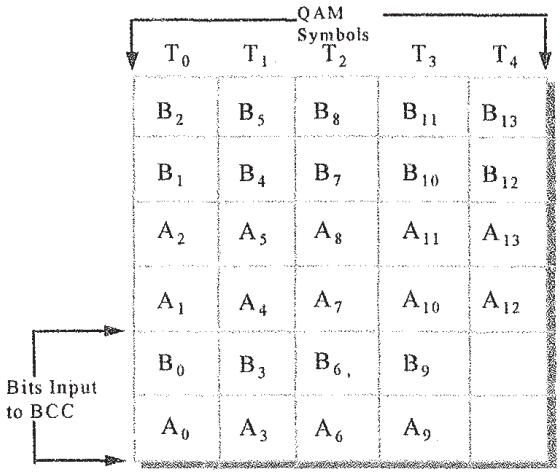



Figure 7.11 Trellis group for a 64-QAM modulation. After Ref. [4].

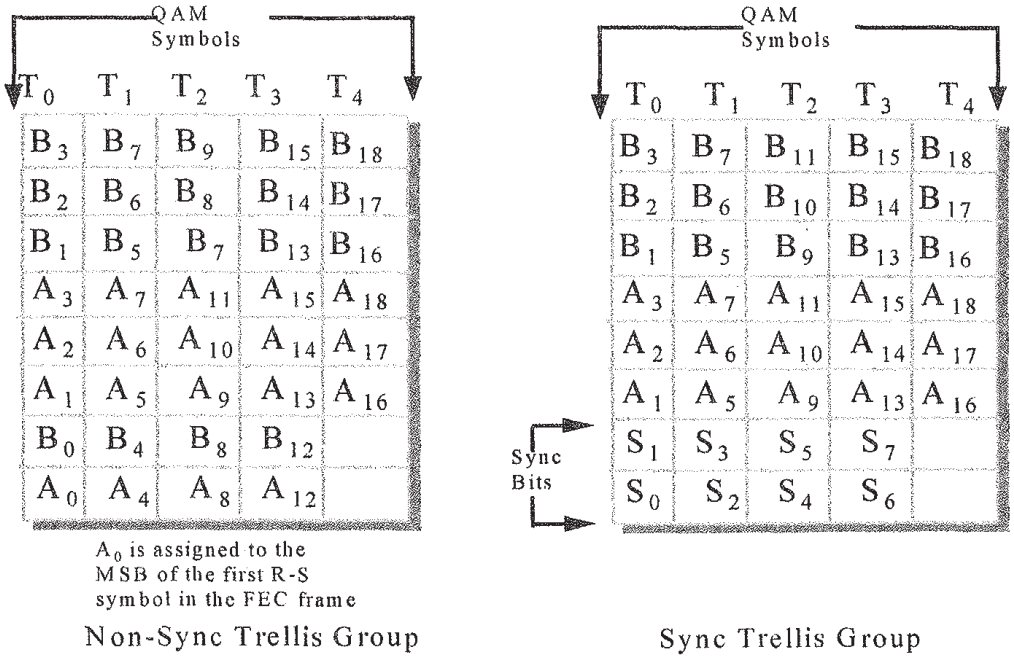



Figure 7.12 Synchronization and nonsynchronization trellis groups for a 256-QAM modulation. After Ref. [4].

uncoded. This will produce an overall output of 15 bits. Thus, the overall TCM rate for 64-QAM is 14/15. The trellis group is formed from R-S symbols as follows: for the "A" symbols, the R-S symbols are read, from MSB to LSB, $A_{10}$, $A_8$, $A_7$, $A_5$, $A_4$, $A_2$, $A_1$ and $A_9$, $A_6$, $A_3$, $A_0$, $A_{13}$, $A_{12}$, $A_{11}$. The four MSBs of the second symbol are fed into the BCC, one bit at a time, LSB first. The remaining bits of the second symbol and all the bits of the first symbol are fed into the mapper, uncoded, LSB first one bit at a time. The four bits sent to the BCC will produce five coded bits labeled $U_1$, $U_2$, $U_3$, $U_4$, $U_5$. The same process is done for the "B" bits. The process can be seen in Figure 7.10. With 64-QAM, four R-S symbols conveniently fit into one trellis group, and in this case the sync word is embedded in every position of the trellis group.

For a 256-QAM modulation, the TCM is using the same 1/2-rate BCC and the same 4/5 punctured matrix as a 64-QAM modulation. However, in this case all the sync information is embedded only in the trellis group LSBs. There are two different types of trellis groups in 256-QAM, namely the nonsynchronization and the synchronization groups, which will be labeled as nonsync and sync groups, respectively. Each trellis group will produce five QAM symbols; the nonsync group will contain 38 data bits while the sync group will contain 30 data bits and 8 sync bits. Figure 7.12 shows both the nonsync Trellis group and the sync trellis group. Since there are 88 R-S blocks per FEC frame, there will be a total of 2,076 trellis groups per frame. Of these trellis groups, 2,071 are nonsync trellis groups and five are sync trellis groups. The five sync trellis groups come at the end of the frame. The frame sync trailer is aligned to the trellis groups. The trellis group is further divided into two groups: one upper or MSB uncoded bit stream and one lower or LSB coded bit stream. The first bit of the nonsync trellis group is assigned to the MSB of the first R-S symbol in the FEC frame. The output from each BCC is the five parity bits labeled $U_1$ through $U_5$ and $V_1$ through $V_5$, respectively. To form trellis groups from R-S code words the R-S code words are serialized beginning with the MSB of the first codeword. The MSB of the first symbol is placed in position $A_0$ of the nonsync trellis group shown in Figure 7.12. Bits are then placed in trellis group locations, from R-S symbols, in the following order: $A_0$ $B_0$ $A_1$ ... $B_3$ $A_4$ $B_4$ ... $B_{16}$ $B_{17}$ $B_{18}$. For the sync trellis groups, the bits from the serialized R-S symbols begin at location $A_1$ instead of $A_0$; from each trellis group, four bits are coded by the BCC producing five coded bits. Each of the last five sync trellis groups in an FEC frame contain 8 sync bits, $S_0S_1...S_7$ in the frame sync trailer as shown in Figure 7.12.

Of the 38 input bits to the TCM modulator, two groups of 4 bits of the differentially precoded bit stream in a trellis group are separately encoded by the 1/2-rate BCC. Each BCC produces 5 coded bits, and the remaining bits are sent to the QAM mapper uncoded. This produces a total output of 40 bits, resulting in a 19/20-rate TCM for 256-QAM modulation.

### 7.5.4 Differential Precoder

The differential precoder shown in Figure 7.13 performs the 90° rotationally invariant trellis coding. Rotationally invariant coding is employed in both the 64- and 256-QAM modulators. The key for robust modem design is to have very fast recovery from a carrier phase

slip. In a nonrotationally invariant design, a carrier phase slip will require a major re-synchronization of the FEC, leading to a burst of errors at the FEC output.

The differential precoder allows the information to be carried by the change in phase, rather than by the absolute phase. For 64-QAM, the 3rd and the 6th bits of the 6-bit symbols are differentially encoded, and for 256-QAM, the 4th and 8th bits are differentially encoded. The differential precoder is described by the following coupled set of equations:

$$X_j = W_j + X_{j-1} + Z_j\left[X_{j-1} + Y_{j-1}\right] \qquad (7.14)$$
$$Y_j = Z_j + W_j + Y_{j-1} + Z_j\left[X_{j-1} + Y_{j-1}\right]$$

If you mask out the 3rd and the 6th bits in 64-QAM as in Figure 7.10 as well as the 4th and 8th bits in 256-QAM, the remaining bits have 90° rotational invariance, which is inherent in the labeling of the symbol constellation.

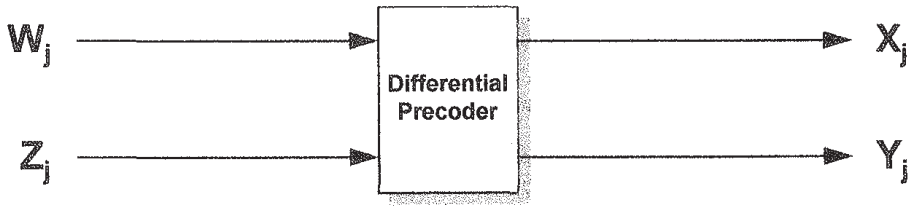



Figure 7.13 Differential precoder for TCM as shown in Figure 7.10.

## 7.6 Randomizer/Derandomizer

The randomizer is the third layer of processing in the FEC block diagram. The randomizer provides an even distribution of the symbols in the constellation, resulting in a flat QAM spectrum, which enables the QAM demodulator to maintain proper lock. The randomizer adds a pseudo-random noise (PN) sequence of 7-bit symbols to the transmitted signal to assure a random transmitted sequence.

For both 64- and 256-QAM, the randomizer is initialized during the FEC frame trailer and is enabled at the first symbol after the trailer. Thus, the trailer itself is not randomized. Initialization is defined as preloading to the all "1" state, for the randomizer structure shown in

Figure 7.14. The generating polynomial of the randomizer, which uses a linear feedback shift register specified by a GF ($2^7$), is defined as follows:

$$g(X) = X^3 + X + \alpha^3 \tag{7.15}$$

where $\alpha^7 + \alpha^3 + 1 = 0$. The corresponding randomizer structure is shown in Figure 7.14.

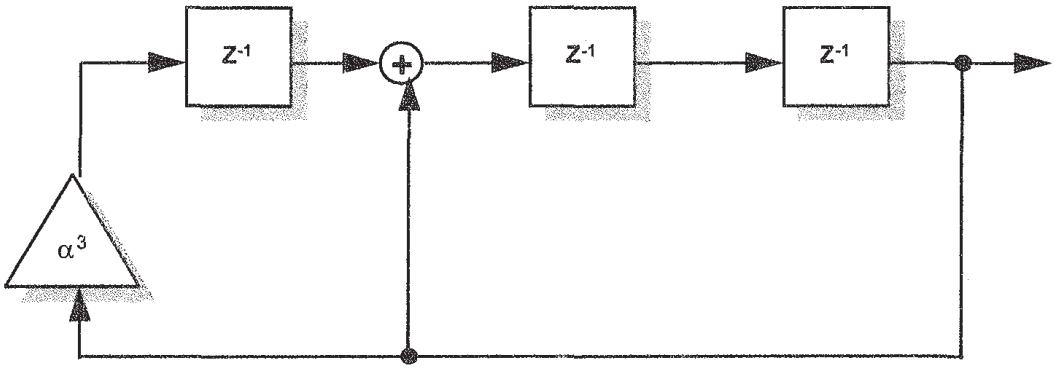



Figure 7.14 A QAM randomizer structure (7-bit byte scrambler).

## 7.7 M-ary QAM Modulator Design and Operation

### 7.7.1 Baseband Shaping Filter

One of the main tasks of a baseband digital communication system is to cancel the effect of the so-called *intersymbol interference* (ISI) [10]. Due to distortions in the transmitter, in the receiver, or in the channel, the received signals overlap with one another in time. In other words, the end of one received pulse interferes in the detection process with the beginning of the following pulse, generating ISI. Nyquist [14], who has investigated this problem, showed that the theoretical minimum system bandwidth needed to detect $R_s$ symbols/s without generating ISI is $R_s/2$ Hz. This condition is satisfied for an ideal filter with a rectangular transfer function H(f) with a Nyquist bandwidth $W = R_s/2 = 1/2T_s$. The impulse response of this filter, which is given by the inverse Fourier transform, is $h(t) = \text{sinc}(t/T_s)$. However, such a brick-wall filter is very difficult to implement since it requires an infinite time delay and would be very sensitive to small timing errors. The most commonly used filter in digital

systems, which smoothly approximates the Nyquist filter, is called a Nyquist *raised-cosine* (RC) filter, whose transfer function can be written as

$$H(f)=\begin{cases}1 \quad for \quad |f|<R_s(1-\alpha)/2 \\ \cos^2\left[\dfrac{\pi}{2}\dfrac{|f|-R_s(1-\alpha)/2}{\alpha R_s}\right] for \quad R_s(1-\alpha)/2\le|f|\le R_s(1+\alpha) \\ 0\, for \quad |f|>R_s(1+\alpha)\end{cases} \tag{7.16}$$

where $W = R_S(1 + \alpha)/2$ is the –6-dB bandwidth, and $\alpha$ is the so-called bandwidth roll-off factor, which represents the fractional excess bandwidth over the Nyquist bandwidth. The corresponding impulse response $h(t)$ of this filter is given by:

$$h(t)=R_s\cdot \sin c\left[R_s t\right]\cdot\frac{\cos\left[\pi\alpha R_s t\right]}{1-4(\alpha R_s t)^2} \tag{7.17}$$

where $\text{sinc}(x) = \sin(\pi x)/\pi x$. This filter is very popular since its impulse response h(t) goes through zero when it is sampled every $T_S$, and thus it has no ISI. It should be pointed out that $h(t)$ is a *noncausal* filter since its impulse response starts from t = $-\infty$. However, $h(t–t_0)$, which is a delayed version of $h(t)$ is causal and can be generated by a real filter. Notice from Equation (7.17) that $h(t)$ with a smaller roll-off factor ($\alpha$) has large oscillating amplitudes between positive and negative values.

A closely related filter to the Nyquist RC filter is the square-root-raised-cosine (SRRC) shaping filter. It is well known that when a communication system is operating in the presence of additive white Gaussian noise (AWGN), then, the BER is minimized by optimum partitioning of the transmitter and receiver filters. These transmit and receiver filters are called a *matched* filter pair since their transfer function product must yield the RC filter as described by Equation (7.16). Thus, the frequency response of each filter is simply the square root of $H(f)$ or the SRRC filter, whose impulse response is given by [10]

$$h_{SRRC}(t)=\sqrt{R_s}\cdot\frac{\sin\left[\pi R_s(1-\alpha)t\right]+\left(4\alpha R_s t\right)\cdot\cos\left[\pi R_s(1+\alpha)t\right]}{\pi R_s t\left[1-(4\alpha R_s t)^2\right]} \tag{7.18}$$

Figure 7.15 shows, for example, the frequency and amplitude response of an SRRC filter with 160 taps and with a 12% roll-off factor. Notice the side-lobs in the SRRC frequency response, which are caused by the truncated length of an ideal filter. There are different in-band and out-of-band amplitude variation requirements on a digitally implemented SRRC filter, particularly in the transmitter. For example, the MCNS DOCSIS 1.1 protocol requires

that the amplitude variations generated by SRRC filters in an upstream QAM transmitter be within ±0.3-dB for any frequency within $f_C \pm R_S/4$, and less than −30-dBc for any frequency within $f_C \pm 5R_S/8$, where $f_C$ is the channel center frequency [1].

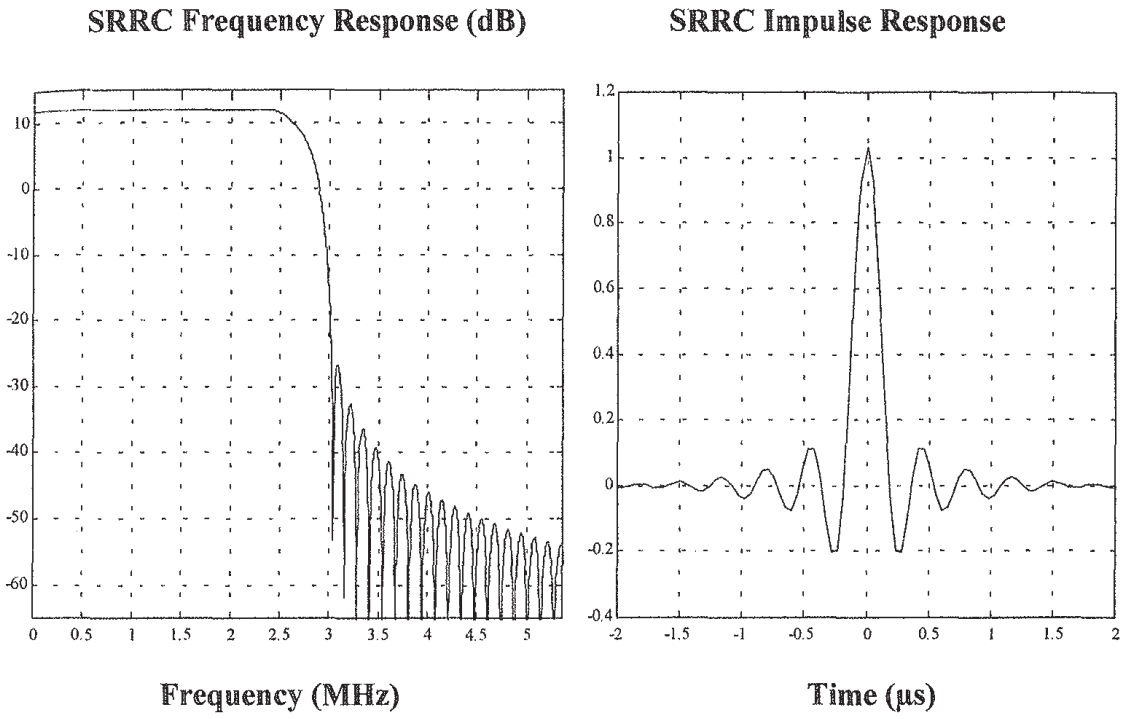



Figure 7.15 Frequency and impulse response of 160-taps SRRC filter with 12% roll-off factor.

## 7.7.2 Building Blocks of M-ary QAM Modulator

The task of an M-ary QAM modulator is to transform a baseband bit sequence to a QAM constellation with M possible amplitude and phase levels within a given bandwidth at IF frequency. There are various ways to digitally implement a conventional M-ary QAM modulator. Figure 7.16 shows, for example, a block diagram of a digitally implemented conventional QAM modulator. The serial input bit stream, which is typically a binary non-return-to-zero (NRZ) sequence, has $\log_2(M)$ bits per symbol. The input symbols are first demultiplexed into two branches, one in-phase (I branch) and one in-quadrature-phase (Q branch) by a serial-to-parallel converter, where each branch has $k = \log_2(M)/2$ of the bits per symbol. These bits need to be mapped to $2^K$ possible signal amplitudes, which can be done

in several ways. The preferred mapping is the one in which adjacent signal amplitudes differ by one binary bit. Such mapping is based on the fact that random errors due to a Gaussian noise in the demodulation of the desired signal will most likely affect the adjacent amplitudes to the transmitted signal amplitude. The result is only a single bit error. This mapping scheme is known as Gray coding [10]. Since M-ary QAM constellations have $\pi/2$ symmetry, the inherent $\pm\ \pi/2$ phase ambiguity in the QAM modulator needs to be resolved. This means that the QAM receiver can lock the carrier for every $\pi/2$ phase error. Differential encoding, as was explained in Section 7.4.4, encodes the next symbol based on the current input symbol and the previous encoded symbol.

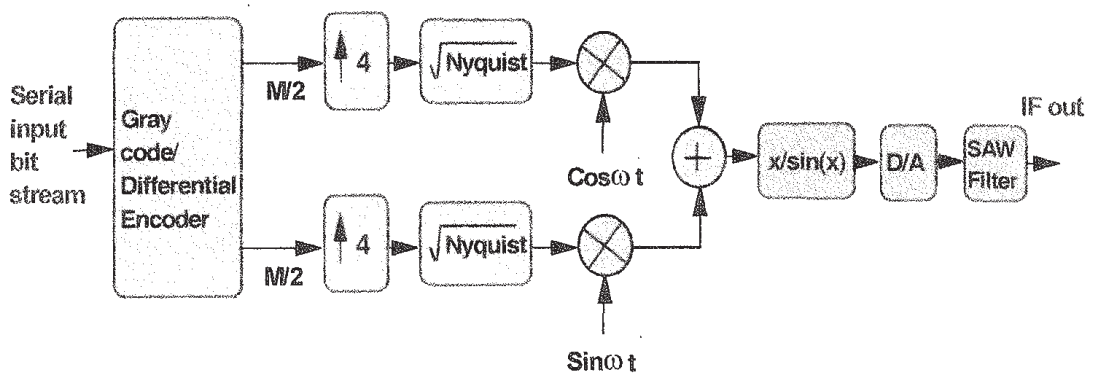



Figure 7.16 Simplified block diagram of a digitally implemented QAM modulator. After Ref. [15].

In many digital applications, the QAM modulator is required to operate at a given symbol rate, which is independent of the input symbol rate. For example, Figure 7.16 shows a 1x4 interpolation filter, which has four sample points per symbol and provides the needed sampling rate up-conversion. In the QAM receiver (Section 7.7), a 1x4 decimation filter is used. The advantage of using interpolation or decimation filters in the QAM modulator or demodulator is the elimination of the need to use high-speed digital multipliers and adders. Then, each of the differentially encoded baseband bit sequences is filtered using an SRRC shaping filter to combat ISI, which will be generated at the QAM receiver.

To form the QAM constellation, the I and Q branches of the M-ary QAM modulator are multiplied by $\sin(2\pi f_c t)$ and $\cos(2\pi f_c t)$, where $f_c$ represents the QAM carrier frequency. Mathematically, the M-ary QAM signal can be written as

$$S(t) = I(t)\cdot\cos(2\pi f_c t) + Q(t)\cdot\sin(2\pi f_c t) \tag{7.19}$$

where $I(t)$ and $Q(t)$ are the $2^K$ amplitude and phase levels of the M-ary QAM constellation. Notice that there is a $\pi/2$ phase shift between the I and Q branches. In practical applications, the quadrature oscillators can be implemented using the so-called d*irect digital frequency synthesizer* (DDFS) [16]. The DDFS can generate frequencies in steps of $\Delta f = F_S/2^N$, where $F_S$ is the modulator sampling frequency and $N$ is the number of bits used in the DDFS phase accumulator. The QAM carrier frequency $f_c$ can be generated by the DDFS by $f_c = W \cdot \Delta f$, where $W$ is a constant word that is added to the DDFS accumulator at $F_S$ rate [17].

After combining the I and Q branches, a x/sin(x) correction filter is applied before the digital-to-analog converter (DAC). In an ideal DAC, the output voltage is instantaneous for each sample. Thus, the output voltage between samples has to be interpolated to make the analog output continuous in both time and amplitude. If the Nyquist sampling criterion is used and the interpolation filter eliminates all the frequencies above the folding frequency, then the output is an exact replica of the original input signal, except for the quantization errors. For practical reasons, natural sampling instead of impulse sampling is used in the D/A converter. This means that the DAC output is typically held constant for the full duration between samples. The frequency transfer function of the interpolation filter causes a sin(x)/x error, typically less than 4-dB up to $f/2F_S$. The x/sin(x) filter is designed to compensate for the frequency roll-off [18].

When the M-ary QAM modulator is designed, the magnitude of all the discrete and spurious noise outside the QAM channel is particularly important. This is because when QAM channels are stacked right next to one another, a large amount of spurious and discrete noise of one QAM channel appears as an in-band noise, which will degrade the BER of the transmitted QAM channel. To overcome this possible limitation, a surface-acoustic-wave (SAW) filter is used at the IF frequency of the QAM modulator. The SAW filters have high insertion loss, typically about 20-dB, and they may be expensive. Despite that, SAW filters are used since they are FIR filters with a very flat in-band group delay. In North America, the IF frequency of QAM modulator is typically 43.75-MHz or 44-MHz, while in Europe the IF frequency is 36.125-MHz. The IF output of the QAM modulator is typically upconverted to the desired RF frequency in the 54 MHz to 860 MHz range.

## 7.8 M-ary QAM Receiver Design and Operation

The task of the QAM receiver is the inverse of the QAM transmitter, namely, to convert the received RF signal to a digital baseband bit sequence. According to Figure 7.1, the first building block of the QAM receiver is the RF tuner. Since the RF tuner is also an essential part of the digital set-top box, it will be discussed separately in Chapter 8 on subscriber home terminals. Figure 7.17 shows a simplified block diagram of a digitally implemented M-ary QAM receiver [19, 20]. The RF tuner output is an analog QAM signal typically at an IF of 44-MHz. The analog front end of the QAM receiver is shown, for example, in Figure 7.18 [19]. There are two widely used methods to digitize and down-convert the analog IF

signal before it is presented to the QAM demodulator. In the first method, the IF signal is down-converted to a secondary IF frequency equal to the QAM symbol rate (≅ 5 MHz) before it is digitized. The down-conversion process typically requires the use of a SAW filter, a programmable gain amplifier (PGA), a voltage-controlled oscillator (VCO) or a numerically controlled oscillator (NCO), and a mixer. Then, the mixer output is fed to an ADC with a typical 10-bit resolution. The PGA provides the necessary adjustable gain (up to 20-dB) before it is input to the sample-and-hold circuit in the ADC converter. The ADC samples the input signal at a higher sampling rate than the symbol rate, typically four times, in order to allow all digital quadrature down-conversion and digital filtering. Notice that the error signals generated from the tracking and acquisition loops can be used for example to drive two VCOs to adjust the input carrier frequency and timing accuracy as shown in Figure 7.18. Another implementation of the front-end section of the QAM receiver is to replace the two VCOs with fixed-frequency oscillators and adjust the input carrier frequency in the quadrature oscillators. In the second method, which is called direct IF sampling, the 44-MHz IF frequency is passed through a SAW filter and PGA before it is directly digitized and sampled typically by a 10-bit ADC to produce a digital IF signal. The ADC is locked to a very stable reference crystal. The digitally sampled IF signal is first converted to a baseband signal, normally using a multirate filter, and then it is separated into its I and Q

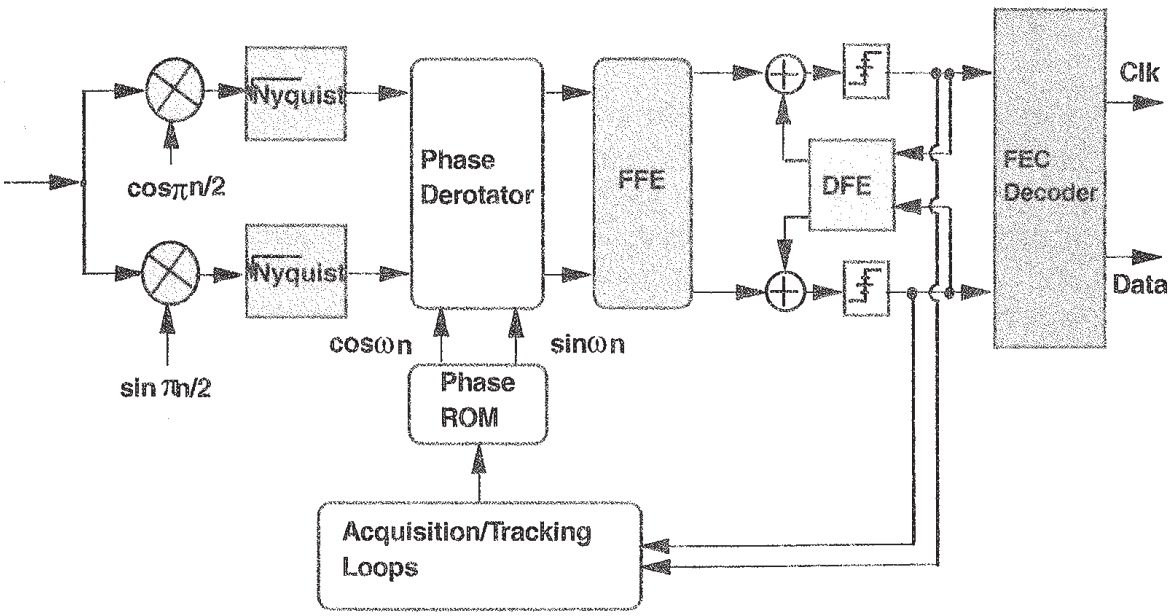



Figure 7.17 Simplified block diagram of a digitally implemented M-ary QAM receiver. After Ref. [19].

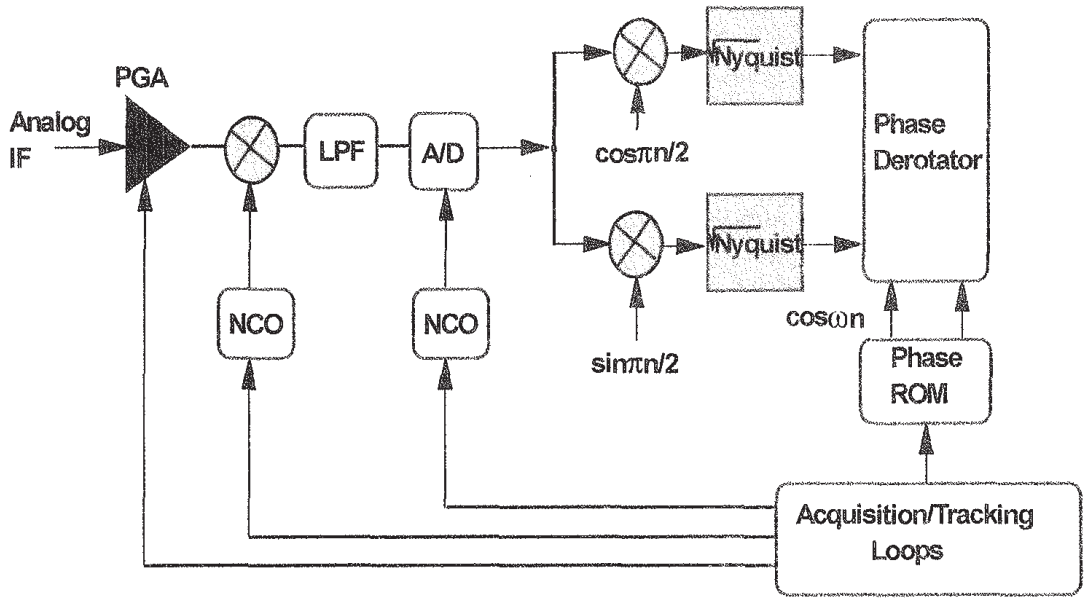



Figure 7.18 Block diagram of the analog front end of the QAM receiver. PGA = programmable gain amplifier. After Ref. [19].

components by multiplication with the quadrature oscillators, namely, the corresponding $\sin(2\pi f_c n)$ and $\cos(2\pi f_c n)$. The I and Q components of the baseband signals are filtered by dual SRRC filters. Since the received sampled signal may have a large amount of phase errors, a decision-directed phase-locked loop (PLL), which is sometimes called a *phase derotator* loop, is used to estimate the carrier phase magnitude and direction for proper compensation. The block in Figure 7.17, which is labeled acquisition and tracking loops, represents the QAM carrier frequency and timing recovery loops. Since the adaptive equalization, carrier recovery, and timing recovery are essential parts of the QAM receiver design and operation, they are discussed in more detail in the following sections. The operation of the FEC decoder block is more complicated than that of the FEC encoder block, and it will not be discussed here.

## 7.9 Adaptive Equalizer

One of the primary components of a QAM receiver is its adaptive equalizer (AE). The AE is designed to adapt its coefficients to minimize the ISI resulting from linear amplitude and phase distortions in the channel, which cannot be removed by the SRRC filters. In the following section, we will review the state-of-the-art adaptive equalization methods and prac-

tices that are employed in QAM receivers used in cable modems and digital set-top terminals.

In order to compensate for a channel's linear distortions, one will need to estimate the actual channel impulse response and adjust accordingly an equalizer filter in the receiver to equalize the channel. This problem was first proposed and addressed by R. W. Lucky [21–23]. The adaptation process adapts the coefficients of the equalizer filter to minimize the effect of ISI at the receiver output, generated by the channel linear distortions. The equalizer adaptation is driven by the magnitude of an error signal from a decision device called a slicer, which indicates to the equalizer the direction its coefficients must change in order to more accurately represent the data symbols at the slicer input. An ideal equalizer will have a transfer function inverse to that of the transmission channel. There are three main approaches to solve the equalization problem as follows: (A) linear equalization, (B) decision-feedback equalization, and (C) transmitter precoding. First, let us explain the linear equalization approach.

The linear transversal equalizer (LTE), which represents the simplest equalizer, is constructed using a tapped delay line, a series of tap weights or coefficients $c_n$, and a slicer. The impulse response and the frequency response of such an equalizer can be written as [24]

$$g_k(t) = \sum_{n=-N}^{n=N} c_{nk} \cdot \delta(t - nT) \tag{7.20}$$

$$G_k(t) = \sum_{n=-N}^{n=N} c_{nk}^{'} \cdot e^{-j2\pi nTf} \tag{7.21}$$

where $\{c_{nk}\}$ are the $(2N+1)$ equalizer tap coefficients at the $k$th sampling time. It was also assumed that the LTE taps are symmetrically located around the main tap (zero-time delay), and the delay time between adjacent taps is a symbol period $T$. Such an equalizer is called a symbol-spaced equalizer. In practice, the LTE filter is approximated by a finite-duration impulse response (FIR) filter. If the output of the equalizer is sampled at $1/T < 2W$, where 2W is the channel bandwidth, then the frequencies in the received signal above the folding frequency $1/T$ are aliased into frequencies below $1/T$. The equalizer compensates for the distorted signal caused by the aliasing effect. In contrast, if the time-delay between adjacent taps $\tau$ satisfies the condition $\tau \leq 1/2W \leq T$, then no aliasing occurs, and the equalizer compensates for the "real" channel distortion. Such an equalizer is called a *fractionally spaced equalizer*. Typically, $\tau = T/2$ is selected.

Let us assume that the impulse response of the transmission channel is $h(t)$, and the equalizer output signal $x(t)$ is sampled when $t = mT$; then the equalizer output signal is given by

$$x(mT) = \sum_{n=-N}^{N} c_{nm} \cdot h(mT - n\tau) \tag{7.22}$$

where n = 0, ±1, ....,±$N$, and τ is the time delay between adjacent taps. Now, we may apply the condition that Equation (7.22) is equal to zero for $m \neq 0$, and equal to 1 for $m = 0$. This condition is called the zero-forcing (ZF) condition. Thus, one obtains (2$N$+1) linear equations for the coefficients of the *zero-forcing equalizer* [24]. It should be noted that a finite-length ZF-LTE does not completely eliminate ISI. However, as the number of taps $N$ is increased, the ISI is reduced to an acceptable level. The main drawback of the ZF equalizer is that it ignores the presence of an additive noise, which may result in a noise enhancement. For example, suppose that a deep null caused by multipath reflections occurs in the channel frequency response. The ZF will provide a large gain to equalize the total response, and in the process also will enhance the noise in the channel, which in turn reduces the transmitted SNR in the receiver.

The noise enhancement problem of the ZF equalizer can be minimized using an AE with an adaptation algorithm based on the minimum-mean-square-error (MMSE) criterion. Let us assume that all the signals are wide-sense stationary. This assumption will become clear later. Suppose that the desired sample output of the equalizer at $t = mT$ is the transmitted values $\{d_m\}$. Then, the mean-square error (MSE) between the actual output sample $x(mT)$ and the desired values $\{d_m\}$ is given by

$$MSE = E\left\{\left[x(mT) - d_m\right]^2\right\} = E\left\{\left[\sum_{n=-N}^{N} c_{nm} h(mT - n\tau) - d_m\right]^2\right\}$$

$$= \sum_{n=-N}^{N}\sum_{k=-N}^{N} c_{nm}c_{km} E\left[h(mT - n\tau)h(mT - k\tau)\right] - 2\sum_{k=-N}^{N} c_{km} E\left[h(mT - k\tau)d_m\right] + E(d_m^2) \tag{7.23}$$

where the expectation is taken with respect to the desired values $\{d_m\}$ and the channel noise. The MMSE condition is obtained by requiring that the derivative of Equation (7.23) with respect to the equalizer coefficients $\{c_{nm}\}$ be equal to zero. Thus, the necessary conditions for the MMSE are given by

$$\sum_{n=-N}^{N} c_{nm} E\left[h(mT - n\tau)h(mT - k\tau)\right] = E\left[h(mT - k\tau)d_m\right] \tag{7.24}$$

for $k$ = 0, ±1, ±2, ...,±$N$. Now, we have (2$N$+1) linear equations for the equalizer coefficients that depend on the statistical properties of the channel noise.

In general, the solution for the equalizer coefficients [Equation (7.24)] is usually obtained by an iterative process. However, since the channel characteristics change with time, adaptive channel equalization is needed. Thus, the adaptive equalizer must adapt its coeffi-

cients in response to the channel time variations to reduce ISI. Combining the adaptive equalizer with the *least mean square* (LMS) algorithm provides an approximate solution to Equation (7.24). The LMS algorithm, which is also known as a *stochastic gradient or steepest-descent* algorithm, does not provide an exact solution to the problem of minimizing the MSE [25, 26]. This approximation is the result of not requiring that the channel characteristics be known or stationary. The LMS approach is to substitute time-average estimates for the ensemble estimates in Equation (7.24) and to solve for the equalizer coefficients. The LMS algorithm works as follows: the equalizer output is driven by an error signal, which is given by

$$e_m = d_m - \sum_{n=-N}^{N} c_{nm} h(mT - n\tau) \tag{7.25}$$

Thus, the error signal is simply the difference between the desired and actual equalizer output. The equalizer coefficients are adjusted during each symbol or fraction of a symbol period to track more accurately the data symbols at the slicer input. Since the equalizer output is typically noisy, a threshold decision device, which is known as a slicer, is used to quantize the output and make decisions. Since the MSE as given by Equation (7.23) has a quadratic dependence on the equalizer coefficients, it has a unique global minimum. Thus, the equalizer coefficients are iteratively adjusted to minimize the MSE by steeply descending along the MSE surface. The n equalizer coefficient at the $(m+1)$ sampling time is given by

$$c_{n(m+1)} = c_{nm} - \frac{\mu}{2} \frac{\partial \left[ e_m^2 \right]}{\partial c_{nm}} \tag{7.26}$$

for $n = 0, \pm 1, \ldots, \pm N$, where $\mu$ is the adaptation step size that controls the size of the change in $c_{nm}$ at each update. The division by 2 in Equation (7.26) was done to avoid a factor of 2 in the subsequent adaptation algorithm. Notice from Equation (7.26) that we have taken the gradient with respect to the equalizer coefficients of the error-signal square instead of the MSE of $e_m$. This is because the iterative process rapidly attenuates the channel noise, and thus provides a good approximation. Substituting Equation (7.25) in Equation (7.26), one obtains

$$c_{n(m+1)} = c_{nm} + \mu e_m h^*(mT - n\tau) \tag{7.27}$$

where $h^*$ represents the complex channel impulse response input to the equalizer. Equation (7.27) is the LMS algorithm. Initially, the adaptive equalizer can be trained by a known training sequence over the channel. This allows the equalizer to initially adjust its coefficients. Once the initial adjustment is completed, it can operate in a decision-directed mode,

where the slicer output is sufficiently reliable to allow the equalizer to adapt to the channel characteristics for any input.

The convergence rate of adaptive equalizers is also an important feature in their design. When fast converging algorithms for adaptive filters are needed, the *recursive-least-squares* (RLS) and/or the *Kalman* filter algorithms are typically used [26–28]. These equalizer adaptation algorithms have been recognized to provide the fastest convergence rates. In fact, the RLS algorithm converges (in MSE) in about $2N$ iteration, where $N$ is the number of equalizer taps. This convergence rate of the RLS algorithm is about an order of magnitude faster than that of the LMS algorithm [26]. Several RLS algorithms were developed where the number of operations per iteration grows linearly rather than quadratic with $N$. These efficient algorithms are called fast RLS algorithms [27]. The potential disadvantage of the RLS algorithm over the LMS algorithm is its numerical stability, particularly for fixed-point arithmetic.

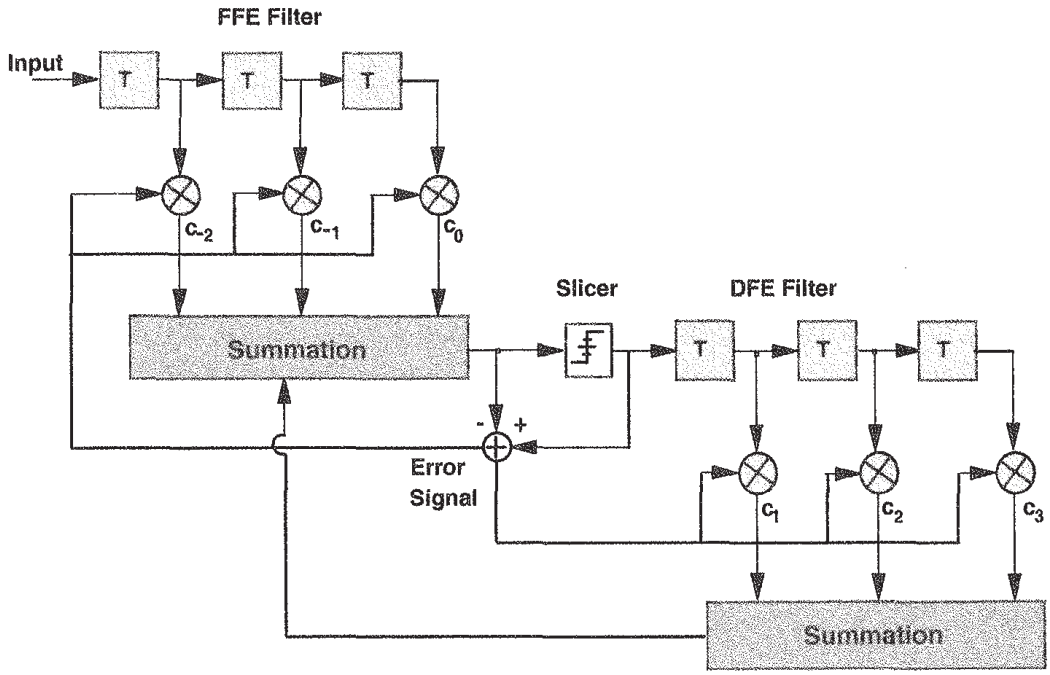



Figure 7.19 Block diagram of adaptive equalizer with feed-forward and decision-feedback filters.

Equalization techniques that do not require a training sequence for the initial adjustment of the equalizer coefficients are called *blind* or *self-recovering* equalization. Blind equalization allows the equalizer to converge without any preliminary carrier phase recovery. The

various blind equalization methods will briefly be described here, and the reader is encouraged to check the literature [28–31]. Three different classes of blind equalization techniques have been developed after the pioneering work by Sato [29]. The first class of algorithms is based on the LMS adaptation algorithm, where the most widely used algorithm in practice is the so-called *Godard* or *constant-modulus algorithm* (CMA) for the adaptive equalizer [30]. The CMA is designed to minimize deviations of the blind equalizer from a constant modulus. The advantage of this algorithm is that it does not require knowing the carrier phase. However, it tends to converge slowly compared with a known training sequence. The second class of algorithms estimates the channel characteristics based on the second- and higher-order statistics of the received signal and then computes the equalizer coefficients [31, 32]. The main drawback of this algorithm class is the large amount of computational complexity involved in the estimation of the higher-order moments of the received signal. The third class of blind equalization algorithms is based on the maximum-likelihood criterion. This algorithm, which jointly estimates the channel impulse response and data sequence similar to the Viterbi algorithm approach, is suitable only for low symbol rates because of its computational complexity.

The adaptive linear equalizers described so far may perform poorly in transmission channels with large linear distortions such as cable TV channels with multipath echoes. A decision-feedback equalizer (DFE) is a nonlinear filter that uses the slicer decisions on the previously detected symbols to eliminate the ISI of the currently input symbol to the slicer. Figure 7.19 shows a block diagram of adaptive equalizer filter with DFE and feed-forward equalizer (FFE) filters. The first filter in Figure 7.19 is an FFE, which is identical to the form of linear equalizer that we have discussed earlier, showing only three coefficients. Each block $T$ represents a symbol-spaced or fractional-spaced time delay between adjacent taps. Notice that the output from all the taps is summed and fed forward to the slicer. In contrast, the output from the DFE is subtracted from the output of the FFE before it is fed to the slicer. For the equalizer shown in Figure 7.19, the output of the slicer is given by

$$x(mT) = \sum_{n=-(N-1)}^{n=0} c_{nm} \cdot h(mT - n\tau) - \sum_{n=1}^{n=M} a_{nm} \cdot \hat{d}_{m-n} \tag{7.28}$$

where the FFE is anticausal with $N$ coefficients $\{c_{nm}\}$ and the DFE is causal with M coefficients $\{a_{nm}\}$. The desired slicer output $d_{m-n}$ has been replaced by the estimated decision outputs. The MSE criterion can be applied to Equation (7.28) and solved for the DFE coefficients, which are given by [24]

$$a_{km} = \sum_{n=-(N-1)}^{n=0} c_{nm} \cdot h(kT - n\tau) \tag{7.29}$$

for $1 \leq k \leq M$. Since the DFE coefficients in Equation (7.29) are expressed in terms of the FFE coefficients, one can substitute this solution in Equation (7.24), and apply the MSE criterion to minimize the $\{c_{nm}\}$ coefficients. A simple extension of the stochastic gradient algorithm can be applied here to enable a joint adaptation of both the FFE and DFE filters. Although DFE is able to combat linear channel distortions, it is not the optimum equalizer since it does not necessarily minimize the probability of error. The ability of the DFE to cancel residual ISI caused by past symbols provides more freedom in the selection of the FFE coefficients. In particular, the FFE does not need to approximate the inverse of the channel impulse response and thus avoids excessive noise enhancement. However, the DFE may be sensitive to an error propagation, namely, there is a greater probability for more incorrect decisions from the slicer following the first one. Fortunately, the error propagation is not usually catastrophic. In practice, both adaptive FFE and DFE are used to combat linear and nonlinear distortions over cable TV channels.

The third approach to the equalization problem is to use transmitter precoding. This approach will be explained later in Chapter 11. The performance of adaptive equalizers used in consumer home terminals such as CMs and digital STBs will be discussed in Chapter 8.

## 7.10 Carrier and Timing Recovery

So far, it has been assumed that both the carrier frequency and phase and the symbol timing are known at the receiver. The QAM receiver performs demodulation of the transmitted signal where its timing reference is usually independent from the transmitter. Extracting the carrier frequency and phase as well as the symbol timing from the received signal are referred to as a carrier and a timing recovery, respectively, which are essential to the operation of the QAM receiver, particularly over cable TV networks. This is because the QAM receiver has to operate in the simultaneous presence of multiple impairment environments, which exist over the cable TV plants. The following subsections briefly explain the operation principles of some of the currently used methods for carrier and timing recovery in QAM receivers.

### 7.10.1 Carrier Recovery in QAM Receivers

In this section we will assume that the symbol timing is known in order to extract the carrier frequency. Carrier recovery is an essential part of the QAM receiver, even if differential coding is employed. When a QAM signal is detected at the receiver, the sampled QAM constellation points are spinning around the origin with a fixed phase error. Thus, the goal of the carrier recovery loop is to minimize the phase error, extract the carrier frequency and phase, and allow the QAM receiver to proceed with the demodulation process.

The carrier recovery methods can be classified into feedforward and feedback methods. In the feedforward method, the carrier recovery circuit generates a signal at four times the carrier frequency ($4f_c$), which is then refined using a bandpass filter (BPF) or a PLL. For QPSK

constellations, the phase error of the detected signal changes only by multiple of $2\pi$, and the signal would contain a discrete component at $4f_c$. This method suffers from self-noise or pattern jitter, and is suitable only for QPSK modulation.

In general, the feedback methods for carrier recovery can further be divided into two main categories: (A) Costas loops [33] and times-N loops [33, 34], and (B) decision-directed loops [35, 36]. Figure 7.20 shows a block diagram of a fourth-order Costas loop for a carrier recovery [33]. Let us assume that $\theta_e$ is the phase error between the received baseband signal and the VCO, which is used to generate the local carrier at $f_c$. The received signal is split four ways, where each branch is multiplied by the VCO signal with 0°, 45°, 90°, and 135° phase shifts, respectively. The LPF at each branch allows one to keep only the lowest order terms in $f_c$. It can be shown (after some manipulations) that the estimated phase error output from the phase detector, which drives the VCO to match the phase of the received signal, is proportional to sin $(4\theta_e)$.

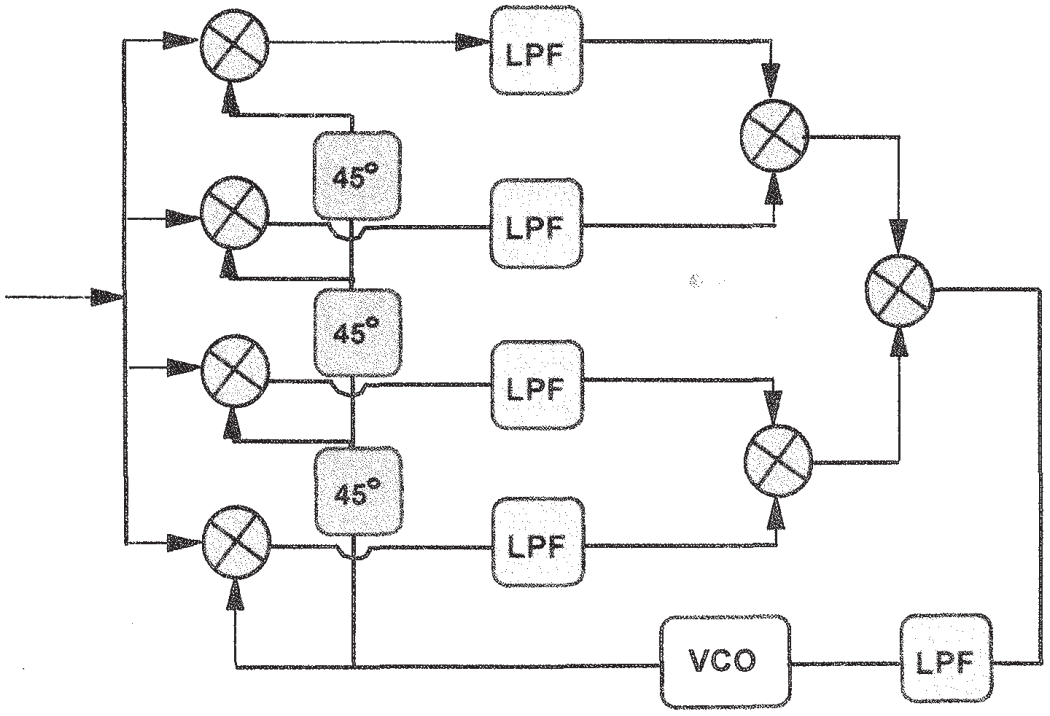



Figure 7.20 Block diagram of a fourth-order Costas loop for M-ary QAM carrier recovery.

Another popular method for a carrier recovery, which is stochastically equivalent to a hard-limited Costas loop, is called the demodulation-remodulation method. In the demodu-

lation part, the two-way split received signal is multiplied by the VCO output, where a phase shift of 90° has been introduced between the two branches. The resultant baseband signals, after passing through the LPF and the slicer, remodulate the locally generated carrier where their phases are compared with the received signal to generate the phase error control signal for the VCO. This method, which uses two-phase detectors, may require a careful balance of the delays between the demodulation and remodulation signal paths. The primary drawback of this method is that it estimates the phase error continuously, and thus the control signal to the VCO can be corrupted by self-noise or pattern noise even with no added noise. For QAM constellations higher than QPSK, the carrier acquisition speed may be a limiting factor, requiring a larger loop bandwidth.

To overcome these issues, let us discuss the second category of feedback loops for carrier recovery, namely, the *decision-directed* carrier recovery loop [35–37]. Figure 7.21 shows a generic block diagram of a decision-directed carrier recovery loop. The received signal is multiplied by the quadrature carriers, which are obtained from the VCO output. The resultant signal is sampled and passed through a slicer, where a decision is made on the symbol every $T$ seconds. The error signal is filtered by the LPF to reject the double-frequency term and to drive the VCO. Notice that the second quadrature component of the sampled signal is

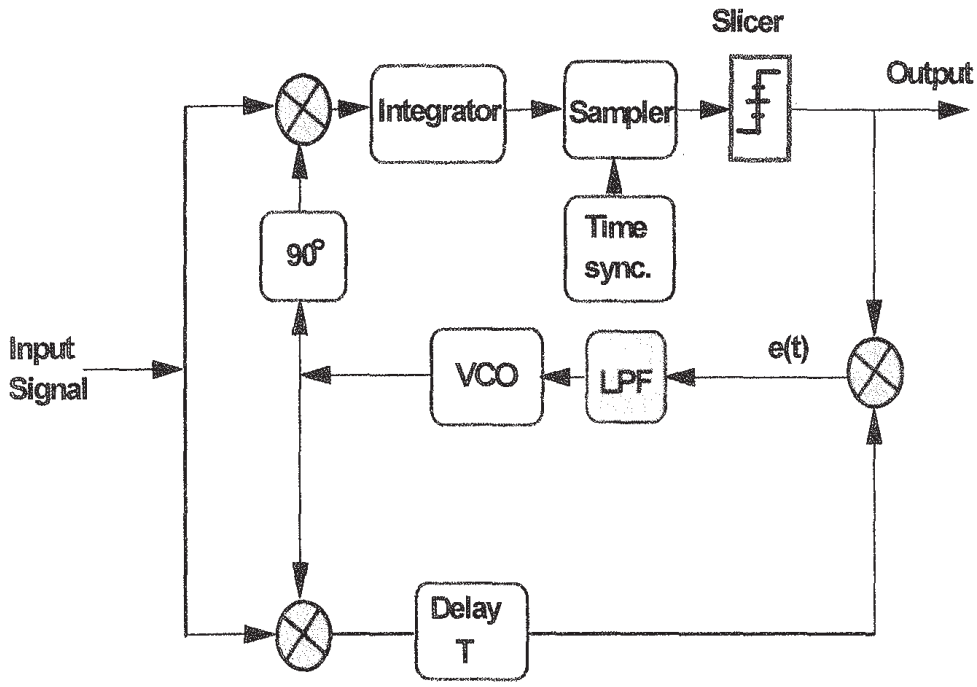



Figure 7.21 Basic block diagram of a decision directed carrier recovery loop.